Journal of Theoretical Biology 574 (2023) 111612

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Evolution of dispersal under spatio-temporal heterogeneity

Kalle Parvinen [a,b,c,*], Hisashi Ohtsuki [d,e], Joe Yuichiro Wakano [f,g]

[a] Department of Mathematics and Statistics, FI-20014, University of Turku, Finland
[b] Advancing Systems Analysis Program, International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria
[c] Okinawa Institute of Science and Technology, Onna-son, Kunigami-gun, Okinawa, 904-0495, Japan
[d] Department of Evolutionary Studies of Biosystems, School of Advanced Sciences, SOKENDAI (The Graduate University for Advanced Studies), Shonan Village, Hayama, Kanagawa 240-0193, Japan
[e] Research Center for Integrative Evolutionary Science, SOKENDAI (The Graduate University for Advanced Studies), Shonan Village, Hayama, Kanagawa 240-0193, Japan
[f] School of Interdisciplinary Mathematical Sciences, Meiji University, Tokyo 164-8525, Japan
[g] Meiji Institute for Advanced Study of Mathematical Sciences, Tokyo 164-8525, Japan

ARTICLE INFO

Keywords:
Conditional dispersal
Arrival bias
Metapopulation fitness
Evolutionary branching

ABSTRACT

Theoretical studies over the past decades have revealed various factors that favor or disfavor the evolution of dispersal. Among these, environmental heterogeneity is one driving force that can impact dispersal traits, because dispersing individuals can obtain a fitness benefit through finding better environments. Despite this potential benefit, some previous works have shown that the existence of spatial heterogeneity hinders evolution of dispersal. On the other hand, temporal heterogeneity has been shown to promote dispersal through a bet-hedging mechanism. When they are combined in a patch-structured population in which the quality of each patch varies over time independently of the others, it has been shown that spatiotemporal heterogeneity can favor evolution of dispersal. When individuals can use patch quality information so that dispersal decision is conditional, the evolutionary outcome can be different since individuals have options to disperse more/less offspring from bad/good patches. In this paper, we generalize the model and results of previous studies. We find richer dynamics including bistable evolutionary dynamics when there is arrival bias towards high-productivity patches. Then we study the evolution of conditional dispersal strategy in this generalized model. We find a surprising result that no offspring will disperse from a patch whose productivity was low when these offspring were born. In addition to mathematical proofs, we also provide intuition behind this initially counter-intuitive result based on reproductive-value arguments. Dispersal from high-productivity patches can evolve, and its parameter dependence behaves similarly, but not identically, to the case of unconditional dispersal. Our results unveil an importance of whether or not individuals can use patch quality information in dispersal evolution.

1. Introduction

Dispersal plays an important role at various levels of biological systems (Clobert et al., 2001). At the population level, the pattern of dispersal specifies spatial genetic diversity (Wright, 1943; Crow and Kimura, 1970). Dispersal may also contribute to rescuing a population from extinction (Levins, 1969; Higgins and Lynch, 2001). At the community level, the importance of dispersal has been stressed, such as its impact on community structure and biodiversity (Leibold et al., 2004).

From an individual point of view, dispersal is one of the key components in the life-history of organisms, and natural selection has shaped its current form (reviewed in Ronce (2007)). Various mechanisms affecting the evolution of dispersal have been presented (Van Valen, 1971; Hamilton and May, 1977; Comins et al., 1980; Dieckmann et al., 1999; Clobert et al., 2001; Bowler and Benton, 2005; Ronce, 2007). Cost of dispersal is an obvious factor, which disfavors evolution of dispersal in most cases. Dispersal can mitigate kin competition between relatives, so dispersal can evolve through indirect fitness benefits (Hamilton and May, 1977; Frank, 1986; Taylor, 1988). In addition, dispersal may reduce the chance of mating between siblings and hence contribute to avoiding the harm of inbreeding depression (Bengtsson, 1978; Motro, 1991; Gandon, 1999; Perrin and Mazalov, 1999).

One major benefit of dispersal is that migrants may find a better environment than the current one. Intuitively, it would seem beneficial to stay in habitat patches with good living conditions, and disperse to better ones from harsh conditions. Individuals specialized in resources available in specific habitats only can be expected to benefit less from

* Corresponding author at: Department of Mathematics and Statistics, FI-20014, University of Turku, Finland.
E-mail address: kalle.parvinen@utu.fi (K. Parvinen).

https://doi.org/10.1016/j.jtbi.2023.111612
Received 11 February 2023; Received in revised form 30 June 2023; Accepted 22 August 2023
Available online 1 September 2023

dispersal than generalists. However, the ability of individuals to distinguish those conditions may be limited. The relative quality of habitats may change in time and that creates unpredictability. Moreover, the quality of a habitat is determined not only intrinsically but also by a demographic factor; it is affected by other individuals competing for the locally available resources. The benefit of dispersal is, therefore, not obvious in such a game-theoretic situation.

The absence of temporal heterogeneity, in which the local population densities and living conditions in each patch do not change in time (equilibrium dynamics), although they can differ between patches, has been observed to select against dispersal (Hastings, 1983; Holt, 1985; Cohen and Levin, 1991; Parvinen, 1999; Gyllenberg et al., 2002; Parvinen, 2006), and dispersal rate can evolve to zero in the absence of other mechanisms promoting dispersal. Under equilibrium dynamics, there are typically more individuals in patches of better quality, so that dispersal on average causes a dispersing individual to encounter worse living conditions than those it left from. Such effect of spatial heterogeneity selecting against dispersal has been observed also in other models (Parvinen, 2002; Parvinen et al., 2020).

Temporal heterogeneity, on the other hand, has been observed to promote dispersal (Gadgil, 1971; Holt and McPeek, 1996; Doebeli and Ruxton, 1997; Johst et al., 1999; Parvinen, 1999, 2006) and allow divergence of dispersal traits through disruptive selection (evolutionary branching) (Metz et al., 1996; Geritz et al., 1997, 1998). Temporal heterogeneity can be caused, e.g., by cyclic or chaotic local population dynamics, environmental heterogeneity or local catastrophes. Under such circumstances, the living conditions in patches change in time, promoting dispersal through a bet-hedging mechanism. For example, if local catastrophes occasionally take place to wipe out a local population, non-dispersing strategies will eventually be eliminated from the system.

There are some articles, in which the evolution of dispersal under spatiotemporal heterogeneity has been investigated (Cohen and Levin, 1991; McPeek and Holt, 1992; Johst and Brandl, 1997; Parvinen, 2002; Massol and Débarre, 2015). McPeek and Holt (1992) found that under both spatially and temporally heterogeneous environments but without the cost of dispersal, dispersal is favored. Parvinen (2002) studied a continuous-time metapopulation model with infinitely many patches of two different types, in which the local population growth is modeled with an ordinary differential equation, and local catastrophes cause temporal heterogeneity. In that model, dispersal does not evolve without catastrophes. Increasing the catastrophe rate allowed dispersal to evolve, but the evolved dispersal rate was observed to be non-monotonic with respect to the catastrophe rate (Ronce et al., 2000; Gyllenberg et al., 2002). Furthermore, in Parvinen (2002) spatial heterogeneity was observed to disfavor dispersal. Cohen and Levin (1991) (their "saturation model") and Massol and Débarre (2015) studied a discrete-time model, in which the population density of adults in a patch is always at a carrying capacity, but the quality of the patch may change in time, resulting in spatiotemporal heterogeneity. They showed that temporal heterogeneity favors dispersal, and Massol and Débarre (2015) demonstrated the importance of different life-cycle assumptions.

In this article we extend the model of Cohen and Levin (1991) and Massol and Débarre (2015) to a more general setting in order to better understand the combined effect of spatial and temporal heterogeneity on evolution of dispersal. Specifically, we will study the effect of patch-quality dependence of dispersal (Bowler and Benton, 2005; Kokko and López-Sepulcre, 2006). For that purpose the following two factors are newly introduced. First, patches of different quality may not be equally likely reached by immigrants, so we introduce arrival bias as model parameters (Parvinen et al., 2020). We will show below that introduction of such bias has a considerable impact on evolutionary dynamics. Second, organisms may be able to use information cues about the current patch for emigration, so we consider conditional dispersal strategies (Parvinen, 1999; Travis et al., 1999; Poethke and Hovestadt, 2002; Kun and Scheuring, 2006; Poethke et al., 2011; Parvinen et al., 2012; Weigang, 2017). We assume that the dispersal probability can depend on the current quality of the patch. As a consequence, the evolving strategy is vector-valued. We will show that some evolutionary outcomes are strikingly different between unconditional and conditional cases. In particular, we find that, except for some special cases, dispersal probability from low-productivity patches always evolves to zero.

The paper is structured as follows. We describe our theoretical model in Section 2. We then analyze the evolution of unconditional dispersal with arrival bias in Section 3. We study conditional dispersal in Section 4. We summarize our results and have discussions in Section 5.

2. Model

2.1. Island model

We employ the model framework proposed by Massol and Débarre (2015) to study the effects of temporary varying environments on the evolution of dispersal, and make some extensions to it to consider a more general case. Specifically, we extend their model by incorporating (i) different carrying capacities of patches, (ii) arrival bias in immigration, and (iii) the possibility of conditional dispersal on patch qualities, as we will describe in detail below.

We consider a population that is subdivided into infinitely many patches. Each patch is in one of N potential quality states affecting reproduction, immigration and competition. The quality state of a patch is assumed to vary over time (transition). Various events can happen in one life-cycle in a different order, but here we pay attention to the following specific life-cycle. In each season, the following events are assumed to occur in the order described below:

1. **Census**: In census at time t, each patch of quality k is assumed to be fully occupied with population density n_k. Here, we mean by "density" that each patch is occupied with many individuals so that any effects arising from the finiteness of the local population size, such as kin selection and demographic stochasticity, can be neglected.
2. **Reproduction**: Each individual will reproduce with fecundity, i.e., expected number of offspring, of γF_k, where $\gamma \to \infty$. The parameter F_k thus describes relative fecundity in a patch of quality k. All adults die after reproduction.
3. **Emigration**: An individual juvenile with dispersal strategy $\boldsymbol{m} = (m_1, \dots, m_N)$, which is genetically encoded, will emigrate from the natal patch of quality k with probability m_k. Each emigrant will survive dispersal with probability p.
4. **Immigration**: Dispersers will immigrate in patches randomly. The parameters λ_k represent bias to arrive at patches of quality k. If there is no bias ($\lambda_1 = \lambda_2 = \cdots = \lambda_N$), the probability ϕ_k to arrive into a patch of quality k is equal to the proportion π_k of such patches. In such case each patch will receive the same amount of immigrants independent of its quality. Otherwise, $\phi_k = \lambda_k \pi_k / \sum_{l=1}^{N} \lambda_l \pi_l$, and patches of different quality will receive different amounts of immigrants. Note that since there are infinitely many patches, the chance that an immigrant arrives in the natal patch is zero.
5. **Transition**: Each patch will potentially experience transition of patch quality. Let $P(j \leftarrow k)$ denote the probability that a patch with quality k at time t will have quality j at time $t+1$.
6. **Competition**: Only a fraction of the juveniles present in the patch (philopatric and immigrant) survive to the next census, so that the population density in the patch is n_j, i.e., based on the new patch quality j. Note that philopatric juveniles in this patch were produced with relative fecundity F_k.

The life-cycle above can be called "(F, D, E, R) life-cycle" in the terminology of Massol and Débarre (2015), where F stands for reproduction, D for dispersal, E for environmental change, and R for density

regulation. Throughout the paper we use symbols that are consistent with our previous paper (Parvinen et al., 2020) as much as possible. We note that the model reduces to the model of "juvenile dispersal with local density regulation" by Massol and Débarre (2015) if we set our parameter values to: two different patch qualities ($N = 2$), the same patch size ($n_1 = n_2$), no arrival bias ($\lambda_1 = \lambda_2$), and unconditional dispersal ($m_1 = m_2 = m$). We shall also employ two different patch qualities ($N = 2$) in our analysis, but make some remarks on a methodology applicable to the case of general N in Appendix A.

As we noted above, we treat n_k as density, not as the number of adults in each patch, which means that each patch has infinitely many adult individuals. Such a treatment excludes the possibility of any form of kin selection acting in our model, and we employ this formalism to simplify our analysis. We will discuss this point once again in Discussion.

The goal of our paper is to study the evolution of dispersal strategies, $\boldsymbol{m}$. For that purpose we consider the metapopulation fitness of mutants (Gyllenberg and Metz, 2001; Metz and Gyllenberg, 2001) to see whether they can successfully invade the population of residents. Details of this approach are described in Appendices A and B.

2.2. Spatial and temporal heterogeneity

Let us now especially consider the case of two qualities, $q \in \{\mathrm{H}, \mathrm{L}\}$, representing patches with High and Low productivity, respectively. In each patch, the transition of quality occurs independently of the other patches. The transition is governed by a time-homogeneous Markov chain, in which the transition probabilities $P(j \leftarrow i)$ $(i, j \in \{\mathrm{H}, \mathrm{L}\})$ describing the probability that the quality of a patch changes from i to j are

$$\begin{pmatrix} P(\mathrm{H}\leftarrow\mathrm{H}) & P(\mathrm{H}\leftarrow\mathrm{L}) \\ P(\mathrm{L}\leftarrow\mathrm{H}) & P(\mathrm{L}\leftarrow\mathrm{L}) \end{pmatrix} = \begin{pmatrix} 1-\beta & \alpha \\ \beta & 1-\alpha \end{pmatrix} \tag{1}$$

where $0 < \alpha, \beta < 1$. Eq. (1) can be re-written as

$$\begin{pmatrix} P(\mathrm{H}\leftarrow\mathrm{H}) & P(\mathrm{H}\leftarrow\mathrm{L}) \\ P(\mathrm{L}\leftarrow\mathrm{H}) & P(\mathrm{L}\leftarrow\mathrm{L}) \end{pmatrix} = \begin{pmatrix} 1-(1-\pi)(1-\tau) & \pi(1-\tau) \\ (1-\pi)(1-\tau) & 1-\pi(1-\tau) \end{pmatrix}, \tag{2}$$

(see Eq. (2) of Massol and Débarre (2015)) with the following transformation from (α, β)-space to (π, τ)-space:

$$\begin{cases} \pi &= \frac{\alpha}{\alpha+\beta}, \\ \tau &= 1-(\alpha+\beta), \end{cases} \tag{3}$$

which maps the region of feasible parameters, $\{(\alpha, \beta) \mid 0 < \alpha, \beta < 1\}$, to a pentagon-like region (Fig. S.1),

$$\left\{(\pi, \tau) \mid 0 < \pi < 1, \max\left[-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\right] < \tau < 1\right\}. \tag{4}$$

The new parameter π represents the equilibrium proportion of high-productivity patches, and τ represents the temporal autocorrelation of patch quality between two consecutive censuses (see Massol and Débarre (2015); see also our Appendix B for details).

2.3. Model parameters

In Appendix A we show that the metapopulation fitness, and thus the evolution of dispersal, depends on the fecundities F_k and population densities n_k only through their product, $F_k n_k$. In the case of two qualities, this dependence occurs through the productivity ratio $f = F_\mathrm{H} n_\mathrm{H}/(F_\mathrm{L} n_\mathrm{L})$. Without loss of generality we can assume $f > 1$, so that high productivity patch type has a larger product, $F_\mathrm{H} n_\mathrm{H} > F_\mathrm{L} n_\mathrm{L}$.

Concerning arrival-bias parameters λ_k, only their relative magnitude matters (if multiplied with the same constant, the arrival probabilities ϕ_k remain unchanged). In the case of two patch qualities, metapopulation fitness depends on the parameters λ_H and λ_L only through $\lambda = \lambda_\mathrm{H}/\lambda_\mathrm{L}$, which measures arrival bias to high-productivity patches relative to the low-productivity ones.

An explicit expression of the metapopulation fitness of mutants can be derived (details in Appendix C and the electronic appendix). That expression uses the parameters f and λ together with other parameters, listed in Table 1.

Table 1
List of symbols.

Model parameters	
$0 < p \le 1$	Dispersal survival probability
$f > 1$	Productivity ratio
$0 < \pi < 1$	Proportion of high-productivity patches
$-1 < \tau < 1$	Temporal autocorrelation
$\lambda > 0$	Arrival bias ($\lambda > 1$ represents bias towards high-productivity patches)
Evolving strategies	
$0 \le m \le 1$	Dispersal probability of unconditional strategy
$0 \le m_\mathrm{H} \le 1$	Dispersal probability from high-productivity patch of conditional strategy
$0 \le m_\mathrm{L} \le 1$	Dispersal probability from low-productivity patch of conditional strategy

3. Evolution of unconditional dispersal

We will first study the evolution of unconditional dispersal, i.e., assume that the patch quality has no effect on the dispersal strategy of an individual, so that $m_1 = m_2 = m$.

3.1. Fitness gradient, singular strategy, and its properties

In Appendix C we derived an explicit expression for the metapopulation fitness $R((m_{\mathrm{mut,H}}, m_{\mathrm{mut,L}}); (m_\mathrm{H}, m_\mathrm{L}))$ in case of two patch types, $N = 2$. We find that the fitness gradient for unconditional dispersal is given by (see Appendix D)

$$D(m) = \frac{\partial}{\partial m_\mathrm{mut}} R((m_\mathrm{mut}, m_\mathrm{mut}); (m, m))\Big|_{m_\mathrm{mut}=m} = \frac{1}{\tilde{A}}(Z + mY), \tag{5}$$

where

$$\begin{aligned} \tilde{A} &= pm(1-\pi+f\pi)[\lambda pm(1-\pi+f\pi) + (1-\tau)(1-\pi+\lambda\pi)(1-m) \\ &\quad \times (\lambda\pi + f(1-\pi))] \geqslant 0, \\ Z &= (1-\tau)(1-\pi+\lambda\pi)\hat{D}_0, \\ \hat{D}_0 &= -f(1-p)(1-\pi+\pi\lambda) + p\pi(1-\pi)(f-1)(f-\lambda), \\ Y &= f(1-p)(1+(-1+\lambda)\pi)^2(1-\tau) - p(1-p)\lambda(1-\pi+\pi f)^2 \\ &\quad - (\lambda-f)^2 p(1-\pi)\pi. \end{aligned} \tag{6}$$

Since the denominator $\tilde{A}$ is non-negative, the sign of the fitness gradient is determined by the expression $Z + mY = (1-\tau)(1-\pi+\lambda\pi)\hat{D}_0 + mY$. Consequently, the zero-dispersal strategy $m = 0$ is evolutionarily attracting if $Z < 0$, or equivalently $\hat{D}_0 < 0$. The full dispersal strategy $m = 1$ is evolutionarily attracting if $Z + Y > 0$. When the zero-dispersal strategy is evolutionarily attracting and the full-dispersal strategy is evolutionarily repelling, the fitness gradient is negative for all m. In such a situation, assuming small mutational steps, zero-dispersal evolves from all initial strategies (Fig. 1a). In the opposite case full dispersal evolves (Fig. 1de). Fig. 1d illustrates that for some parameter combinations, other scenarios are possible with large mutational steps.

When $\hat{D}_0 > 0$ and $Z + Y < 0$, both boundary strategies are evolutionarily repelling, and there exists an intermediate singular strategy

$$m^* = -\frac{Z}{Y} = -\frac{(1-\tau)(1-\pi+\lambda\pi)\hat{D}_0}{Y}, \tag{7}$$

which is evolutionarily attracting from all initial strategies (Fig. 1bc). For $\lambda = 1$ our Eq. (7) agrees with Eq. (18) of Cohen and Levin (1991) and Eq. (13) of Massol and Débarre (2015). Fig. 2a–d illustrates the dependence of the singular strategy on parameters depicting spatiotemporal heterogeneity, discussed in more detail in Section 3.2. For comparison purposes, Fig. 2e–h show analogous results for conditional dispersal, which will be discussed in Section 4.

Whether the singular strategy is uninvadable (evolutionarily stable strategy, ESS) or not, can be determined based on the second derivative

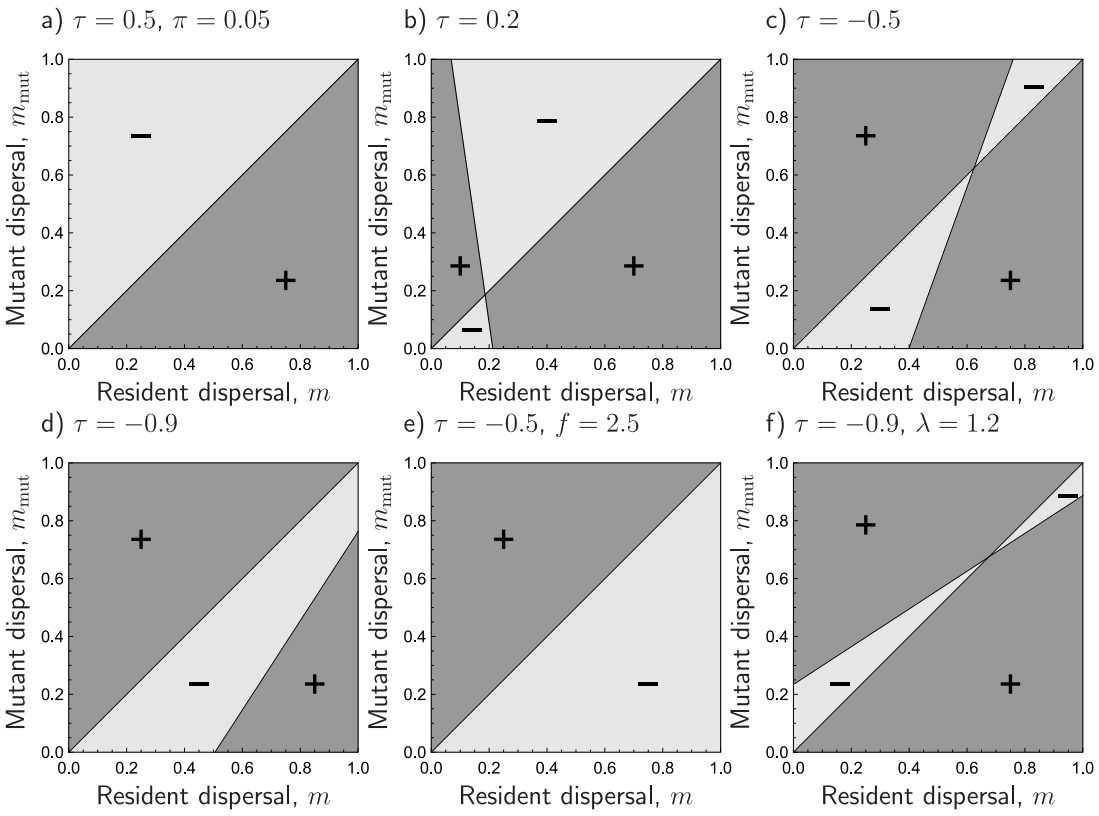

Fig. 1. Pairwise invasibility plots illustrating qualitatively different scenarios of unconditional dispersal evolution. Regions, in which $R((m_{\mathrm{mut}}, m_{\mathrm{mut}}); (m, m)) > 1$, are plotted in dark gray, and regions with $R((m_{\mathrm{mut}}, m_{\mathrm{mut}}); (m, m)) < 1$ in light gray. In all panels $p = 0.95$. Unless otherwise indicated in the panel heading, $f = 1.7$, $\lambda = 1$ and $\pi = 0.5$.

of metapopulation fitness, $\frac{\partial^2}{\partial m_{\mathrm{mut}}^2} R((m_{\mathrm{mut}}, m_{\mathrm{mut}}); (m, m))\Big|_{m_{\mathrm{mut}}=m}$. According to Theorem D.3, the sign of this second derivative is given by the sign of $-\tau$. Therefore, if $\tau > 0$ (positive autocorrelation), the evolutionarily attracting singular strategy is uninvadable, and dispersal evolution will converge to the singular strategy (Fig. 1b). If, $\tau < 0$ (negative autocorrelation), the evolutionarily attracting singular strategy is not uninvadable, and it is thus a branching point (Fig. 1c). In Fig. 2a, the black thick line $\tau = 0$ separates the parameter regions of these two cases. The branching threshold ($\tau = 0$) corresponds to a case when the quality of a patch changes fully randomly (i.e., independent of the previous state). This threshold has already been found by Massol and Débarre (2015) for the case $\lambda = 1$.

When m^* is a branching point (evolutionarily attracting and $\tau < 0$), dispersal evolution is first expected to approach m^*, but then the population becomes dimorphic, and disruptive selection will cause the two strategies present in the population to evolve away from each other. In other words, evolutionary branching will happen. The expression for metapopulation fitness derived in Appendix C gives the metapopulation fitness of a rare mutant in the environment set by a monomorphic resident only. Therefore, in Appendix D.7, we present results from numerical simulations illustrating that evolutionary branching in this model typically leads to the coexistence of highly-dispersing individuals and almost sessile individuals (see Fig. S.2).

When both boundary strategies are evolutionarily attracting, $\hat{D}_0 < 0$ and $Z + Y > 0$, the intermediate singular strategy m^* is evolutionarily repelling, and separates the domains of initial conditions from which strategies converge either to 0 or 1, so there is bistability in the evolutionary dynamics (Fig. 1f). This scenario is possible only if $\lambda > 1$, i.e., when there is arrival bias towards high-productivity patches. Note that Massol and Débarre (2015) showed that bistability is impossible in their model, which assumes $\lambda = 1$.

3.2. Spatiotemporal heterogeneity promotes the evolution of dispersal

According to (5) and (6) (See also Appendix D.2), the zero-dispersal strategy $m = 0$ is evolutionarily repelling, if $\hat{D}_0 = -f(1-p)(1-\pi+\pi\lambda) + p\pi(1-\pi)(f-1)(f-\lambda) > 0$. The first part of this expression is non-positive and the second one is non-negative. Positive dispersal can thus evolve for intermediate π, when p and f are large enough.

When patches have high productivity most of the time ($\pi_{\max} < \pi \leqslant 1$) or when they have low productivity most of the time ($0 \leqslant \pi < \pi_{\min}$), there is not much spatial heterogeneity in the model, and the zero-dispersal strategy is evolutionarily attracting (Fig. 2a, light gray shading). The thresholds $\pi_{\min}$ and $\pi_{\max}$ given by (D.6) in the Appendix do not depend on τ, and thus the boundaries separating the light-gray area of zero-dispersal and positive dispersal in Fig. 2a are vertical lines. As illustrated in Fig. 2b, the singular dispersal strategy typically reaches its maximal value for intermediate π. Spatial heterogeneity, in the form of substantial proportions of both patch types, thus promotes the evolution of dispersal in this model.

The productivity ratio f is another measure of spatial heterogeneity. When $f \approx 1$, there is very little spatial heterogeneity, as patch productivities are then similar. In such a situation we have $\hat{D}_0 < 0$, and positive dispersal does not evolve. Theorem D.4 in Appendix D.6 shows that $\hat{D}_0$ is an increasing function of the productivity ratio f. Therefore, the zero-dispersal strategy is evolutionarily attracting only for $1 < f < \hat{f}$, and positive dispersal can evolve for $f > \hat{f}$, in which the expression for $\hat{f}$ is given in (D.7). This effect is illustrated in Figs. 2d and 3. Increasing the productivity ratio f increases $\hat{D}_0$ and thus promotes the emergence of dispersal.

The effect of f on the singular dispersal strategy m^*, however, is more complicated. The following properties are proved analytically in Theorem D.4: When $\lambda \leqslant 1$, the singular dispersal strategy m^* increases when f increases (Fig. 3ab). The same is true, when $\lambda > 1$ and p is small enough (Fig. 3c). However, when $\lambda > 1$ and $p \approx 1$, m^* can decrease

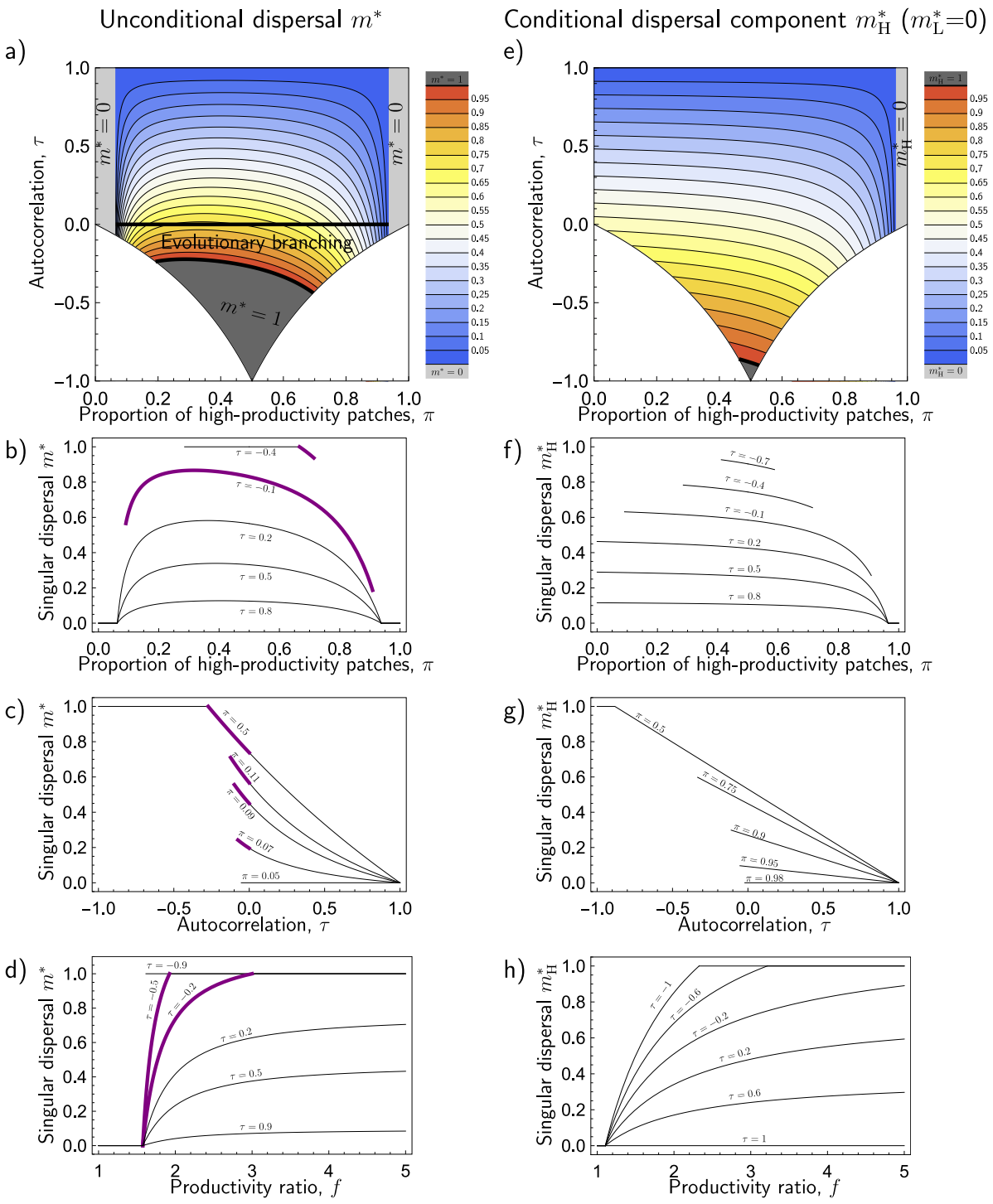

Fig. 2. Spatiotemporal heterogeneity promotes the evolution of unconditional dispersal (left panels). Singular dispersal strategies of unconditional dispersal m^* (left panels) and singular dispersal strategy components m_H^* of conditional dispersal (right panels) (a–c,e–g) with respect to autocorrelation τ and proportion of high-productivity patches π, when $f = 2.5$, (d,h) with respect to τ and productivity ratio f, when $\pi = 0.5$. Other parameters: $p = 0.95$, $\lambda = 1$. Purple curves in (b–d) correspond to branching points, black curves to evolutionarily stable strategies. The conditional dispersal strategy component $m_L^* = 0$.

when f increases. We can thus conclude that spatiotemporal heterogeneity, in the form of substantial differences in the patch qualities, mostly promotes the evolution of dispersal.

The autocorrelation parameter τ can be interpreted to measure temporal heterogeneity (Appendix B). If τ is close to 1, patch qualities change very seldom, so that there is very little temporal heterogeneity. On the other hand, if τ is close to -1, patch qualities change almost every time, which means strong temporal heterogeneity. The quantity $\hat{D}_0$ does not depend on τ. If $\hat{D}_0 > 0$, positive dispersal evolves under our assumption $\tau < 1$. Autocorrelation τ thus does not affect the emergence of positive dispersal, but has an effect on the singular strategy m^*. According to Theorem D.4, m^* is a decreasing function of τ, and $m^* \to 0$ as $\tau \to 1$. Furthermore, there exists such a threshold value τ_{thresh}, that when τ is small enough, $\tau < \tau_{\text{thresh}}$, full dispersal ($m^* = 1$) may evolve (Fig. 2c). These effects of τ on the singular strategy m^* were already found by Massol and Débarre (2015) for $\lambda = 1$. Theorem D.4 thus provides a mathematical proof on this dependence for general λ. The analytical expression for τ_{thresh} is given in (D.10). In Fig. 2a, this is

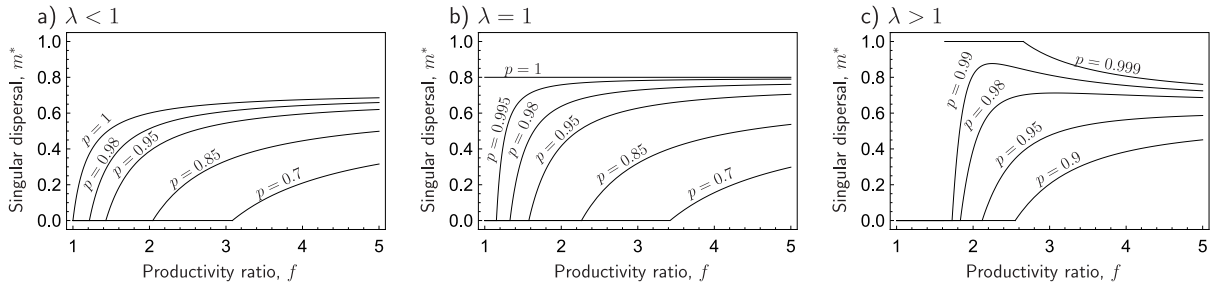

Fig. 3. The singular dispersal strategy m^* with respect to productivity ratio f for different values of the dispersal survival probability p. (a) $\lambda = 0.8$, $\tau = 0.2$ (b) $\lambda = 1$, $\tau = 0.2$ (c) $\lambda = 1.6$, $\tau = 0.5$ Common parameters: $\pi = 0.5$.

the thick black curve separating the areas of positive singular strategy (color shading ranging from blue to red) and full dispersal $m^* = 1$ (dark gray shading). We can thus conclude that temporal heterogeneity promotes the evolution of dispersal.

The parameter plot of the singular dispersal strategy m^* with respect to the proportion of high-productivity patches π and autocorrelation τ in Fig. 2a illustrates that spatiotemporal heterogeneity promotes dispersal. In Appendix D.8 we show that depending on other parameters, such parameter plots may look qualitatively different. Fig. S.3 shows all qualitatively different types of such parameter plots for $p < 1$, and Fig. S.5 for $p = 1$. Furthermore, Fig. S.4 illustrates when each of these cases occurs depending on the choice of the other remaining parameters, productivity ratio f, dispersal survival probability p, and arrival bias λ.

In this section we have observed that spatiotemporal heterogeneity promotes the evolution of dispersal. In particular, the evolutionarily attracting singular dispersal strategy m^* reaches a maximum for intermediate π. This result is strikingly contrasting with Parvinen et al. (2020), in which we observed that spatial heterogeneity disfavors dispersal, so that the singular dispersal strategy reaches a minimum for intermediate π. Note that the model investigated in Parvinen et al. (2020) is otherwise similar to the current one, but no temporal heterogeneity was present and local populations were assumed to be finite, so that kin selection was the only mechanism promoting the evolution of dispersal. In the model studied in the present paper, kin selection is absent due to the large size of local populations, and temporal heterogeneity is the mechanism allowing dispersal evolution. In the absence of temporal heterogeneity in the current model ($\tau \to 1$), dispersal does not evolve, $m^* \to 0$. The results in the present paper and those in Parvinen et al. (2020) thus show that under different circumstances (presence or absence of temporal heterogeneity), spatial heterogeneity can have opposing effects on the evolution of dispersal.

3.3. Larger dispersal survival promotes dispersal

It is relatively easy to see that $\hat{D}_0$ increases with p when $f > \max\{1, \lambda\}$. Therefore, increasing p promotes the emergence of positive dispersal. Furthermore, according to Theorem D.4, the singular strategy m^* increases with dispersal survival p (Fig. 4a), which is again a generalization of a result by Massol and Débarre (2015) for general λ.

3.4. The effect of arrival bias λ may be non-monotonic

According to Theorem D.4, when λ is large, $\lambda > f$, both $\hat{D}_0 < 0$ and $D(1) < 0$ hold, which means that dispersal evolves to zero, $m^* = 0$ from all initial conditions. Using continuity arguments, if $\hat{D}_0 > 0$ for some values of $\lambda < f$, there exists $\lambda^* < f$ at which $\hat{D}_0 = 0$. Furthermore, the singular strategy is decreasing with respect to λ for $\lambda < \lambda^*$ at least when $\lambda \approx \lambda^*$. Numerical explorations illustrate (Fig. 4b) that the singular strategy can either be a decreasing function of λ for all $\lambda < \lambda^*$, or a non-monotonic function of λ (Fig. 4b).

4. Evolution of conditional dispersal

Next, we will study the evolution of conditional dispersal, i.e., assume that the dispersal strategy of a juvenile depends on the quality of the originating patch. The evolving strategy $\boldsymbol{m} = (m_\mathrm{H}, m_\mathrm{L})$ is therefore vector-valued. The explicit expression for the metapopulation fitness $R((m_{\mathrm{mut,H}}, m_{\mathrm{mut,L}}); (m_\mathrm{H}, m_\mathrm{L}))$ is given in Appendix C.

4.1. Fitness gradient

According to Theorem E.1, the fitness gradient for conditional dispersal is given by (see Appendix E.1)

$$\begin{aligned}\boldsymbol{D}(\boldsymbol{m}) &= \left(\frac{\partial}{\partial m_{\mathrm{mut,H}}} R(\boldsymbol{m}_{\mathrm{mut}}; \boldsymbol{m}), \frac{\partial}{\partial m_{\mathrm{mut,L}}} R(\boldsymbol{m}_{\mathrm{mut}}; \boldsymbol{m}) \right) \Bigg|_{\boldsymbol{m}_{\mathrm{mut}} = \boldsymbol{m}} \\ &= (D_\mathrm{H}(m_\mathrm{H}, m_\mathrm{L}), D_\mathrm{L}(m_\mathrm{H}, m_\mathrm{L})), \end{aligned} \tag{8}$$

in which

$$\begin{aligned} D_\mathrm{H}(m_\mathrm{H}, m_\mathrm{L}) &= \pi f (B_0 - B_\mathrm{H} m_\mathrm{H} + B_\mathrm{L} m_\mathrm{L})/A, \\ D_\mathrm{L}(m_\mathrm{H}, m_\mathrm{L}) &= (1-\pi)(-C_0 + C_\mathrm{H} m_\mathrm{H} - C_\mathrm{L} m_\mathrm{L})/A. \end{aligned} \tag{9}$$

The coefficients are given by

$$\begin{aligned} A =& p(m_\mathrm{L}(1-\pi) + f m_\mathrm{H} \pi)[\lambda p (m_\mathrm{L}(1-\pi) + f m_\mathrm{H}\pi) \\ &+ (1-\tau)(1-\pi+\lambda\pi)(\lambda(1-m_\mathrm{L})\pi + f(1-m_\mathrm{H})(1-\pi))] \geqslant 0 \\ B_0 =& (1-\tau)(1-\pi+\pi\lambda)((fp-1)(1-\pi) - \lambda(1-p)\pi) \\ B_\mathrm{H} =& fp(1-\pi+\lambda(1-p)\pi) > 0 \\ B_\mathrm{L} =& \lambda p^2(1-\pi) + (1-p)(1-\tau)(1+(\lambda-1)\pi)^2 > 0 \\ C_0 =& (1-\tau)(1-\pi+\pi\lambda)(f(1-\pi)(1-p) + \lambda\pi(f-p)) \geqslant 0 \\ C_\mathrm{H} =& f(\lambda p^2 \pi + (1-p)(1-\tau)(1+(\lambda-1)\pi)^2) > 0 \\ C_\mathrm{L} =& \lambda p((1-p)(1-\pi) + \lambda\pi) > 0. \end{aligned} \tag{10}$$

4.2. Dispersal from low-productivity patches evolves to zero

The evolutionary dynamics of conditional dispersal can be understood by analyzing phase-plane plots (Fig. 5) showing the isoclines (nullclines) of the components of the fitness gradient,

$$\begin{aligned} D_\mathrm{H}(m_\mathrm{H}, m_\mathrm{L}) = 0 \quad &\Leftrightarrow \quad m_\mathrm{L} = \frac{1}{B_\mathrm{L}}(-B_0 + B_\mathrm{H} m_\mathrm{H}), \\ D_\mathrm{L}(m_\mathrm{H}, m_\mathrm{L}) = 0 \quad &\Leftrightarrow \quad m_\mathrm{L} = \frac{1}{C_\mathrm{L}}(-C_0 + C_\mathrm{H} m_\mathrm{H}), \end{aligned} \tag{11}$$

together with arrows illustrating the direction of the fitness gradient.

According to Theorem E.2 in Appendix E.2, all qualitatively different phase-plane plots are given in Fig. 5. Furthermore, the final outcomes of the evolution of conditional dispersal are as follows:

- If $p = 1$, the two isoclines (11) are identical and dispersal evolution first converges to the isocline but is neutral along it (Fig. 5bc). The isoclines may reside outside the feasible domain, in which case m_H evolves to 1 and m_L evolves to 0 (Fig. 5d).

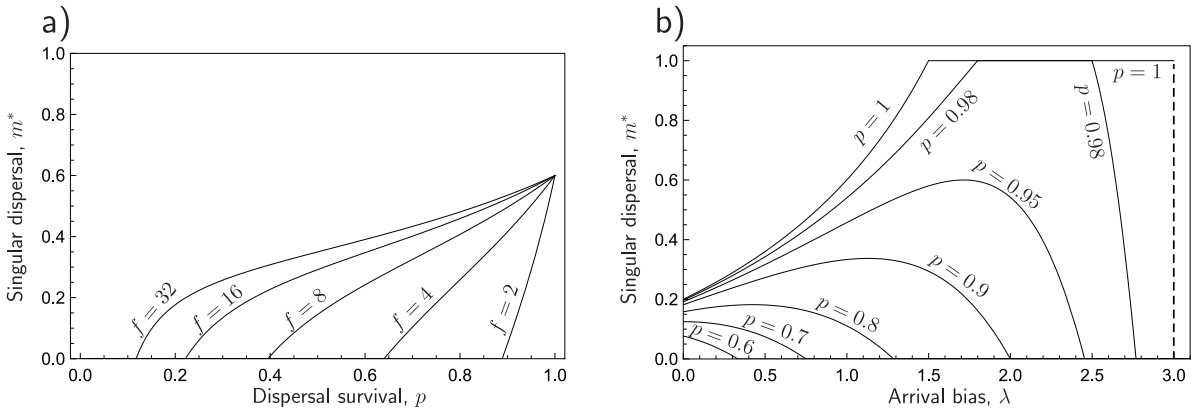

Fig. 4. The singular dispersal strategy m^* (a) with respect to dispersal survival probability p for different values of the productivity ratio f, (b) with respect to the arrival bias λ for different values of the dispersal survival probability p. Parameters: (a) $\lambda = 1$, (b) $f = 3$. Common parameters: $\pi = 0.5$ and $\tau = 0.4$.

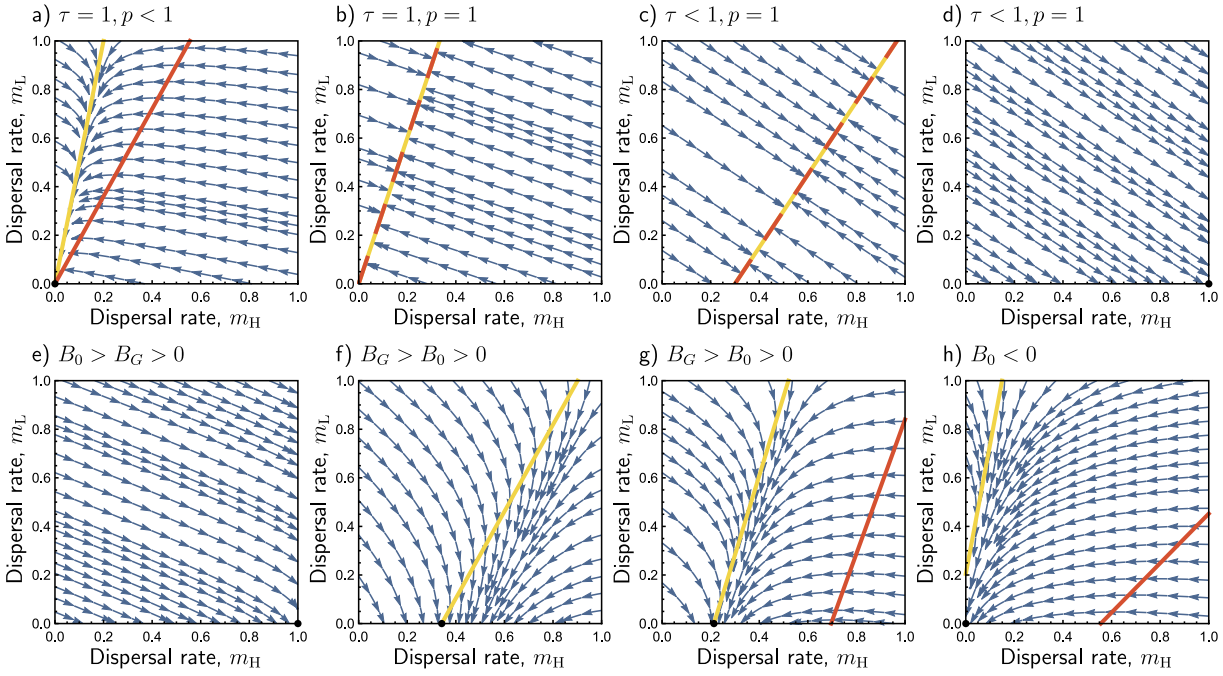

Fig. 5. Qualitatively different phase-plane plots of evolution of conditional dispersal, illustrating the direction of the fitness gradient. The isoclines of the fitness gradient are shown with thick yellow ($D_H = 0$) and red ($D_L = 0$) lines. When $p = 1$, these isoclines coincide, and are therefore illustrated with a dashed line in panels b and c. Curves with arrows illustrate potential trajectories of evolutionary dynamics. When applicable, the unique evolutionary endpoint $(m_H^*, 0)$ is marked with a black dot. Parameters: (a) $\tau = 1$, $p = 0.75$, $\lambda = 1$, (b) $\tau = 1$, $p = 1$, $\lambda = 1$, (c) $\tau = 0.7$, $p = 1$, $\lambda = 2$, (d) $\tau = -0.7$, $p = 1$, $\lambda = 2$, (e) $\tau = -0.75$, $p = 0.95$, $\lambda = 1$, (f) $\tau = -0.5$, $p = 0.65$, $\lambda = 1$, (g) $\tau = 0.4$, $p = 0.75$, $\lambda = 1$, (h) $\tau = 0.9$, $p = 0.4$, $\lambda = 1.5$. All panels $\pi = 0.5$, $f = 3$.

- If $p < 1$, the dispersal strategy component m_L (dispersal from low-productivity patches) always evolves to zero. The other component, m_H, evolves to (Fig. 5e–h)

$$m_H^* = \begin{cases} 0, & \text{if } B_0 \leqslant 0, \\ 1, & \text{if } B_0 \geqslant B_H, \\ \frac{B_0}{B_H}, & \text{otherwise.} \end{cases} \tag{12}$$

According to Appendix E.3, in case $0 < m_H^* < 1$, we have

$$\begin{aligned} R((m_{\text{mut,H}}, m_{\text{mut,L}}); (m_H^*, 0)) < 1 \quad & \text{for } m_{\text{mut,L}} > 0 \text{ and for any } m_{\text{mut,H}} \\ R((m_{\text{mut,H}}, 0); (m_H^*, 0)) = 1 \quad & \text{for any } m_{\text{mut,H}}. \end{aligned} \tag{13}$$

which means that the boundary-singular strategy $(m_H^*, 0)$ is uninvadable against mutants with $m_{\text{mut,L}} > 0$ and neutral against mutants with $m_{\text{mut,L}} = 0$. This means that $(m_H^*, 0)$ is an evolutionary endpoint, and no evolutionary branching can happen for conditional dispersal.

Fecundity F_k measures reproductive success, and population densities n_k alter the probability of surviving competition. Patches with the larger value of $F_k n_k$ have high productivity. Naively one could have expected that it would be better to avoid patches of low productivity, so that m_L would evolve to some positive value and m_H would evolve to zero. However, the local conditions individuals observe consist not only of F_k and n_k, but also of the amount of competitors therein. We will consider the heuristic reasons behind our analytical result in Discussion.

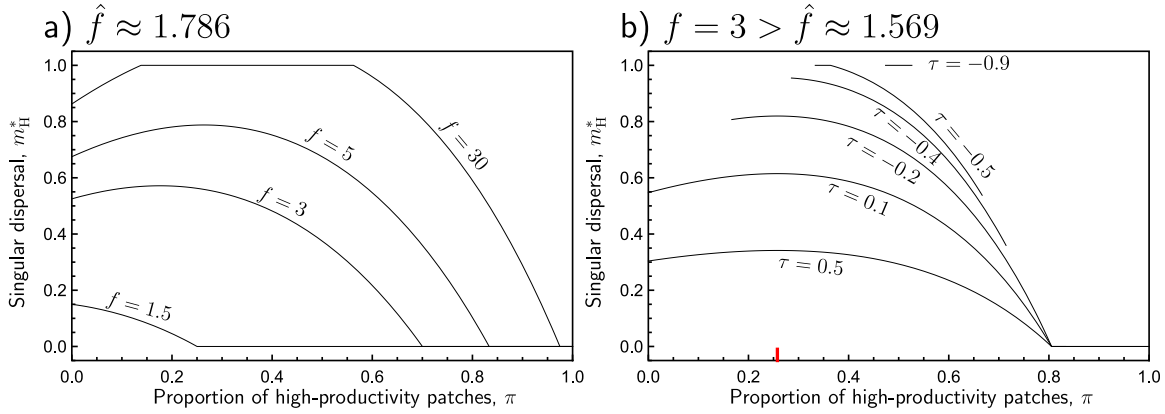

Fig. 6. The singular conditional dispersal strategy component m_H^* with respect to the proportion of high-productivity patches π for (a) different values of productivity ratio f (b) different values of autocorrelation τ. When $\tau \geqslant 0$, the singular strategy m_H^* is a non-monotonic function of π, when $\lambda > 1/p$ and $f > \hat{f}$ (panel a, $f = 3, 5, 30$). Panel (b) illustrates, that when $\tau < 0$, the condition $\lambda > 1/p$ and $f < \hat{f}$ is only a necessary condition for m_H^* to be a non-monotonic function of π. When τ is small, the value of π at which m_H^* would reach its maximum (marked in red) may lie outside of the parameter domain. Parameters (a) $p = 0.8$, $\lambda = 3 > 1/p \approx 1.25$, $\tau = 0.1$, (b) $p = 0.85$, $\lambda = 2.5 > 1/p \approx 1.18$, $f = 3$.

4.3. Spatiotemporal heterogeneity and evolution of conditional dispersal

According to (9) and (12), positive dispersal from high-productivity patches, $m_H^* > 0$ evolves, if $B_0 > 0$, or equivalently $\tau < 1$ and

$$B_0^* = (fp - 1)(1 - \pi) - \lambda(1 - p)\pi > 0. \tag{14}$$

From this expression we see that positive m_H can evolve, if (1) p and f are large enough, and (2) π and λ are not too large. The autocorrelation $\tau < 1$ does not quantitatively affect the emergence of dispersal. If $f < 1/p$, then $B_0^* < 0$ holds and we have $m_H^* = 0$. The condition $fp > 1$ is thus necessary for positive dispersal to evolve.

Substantial spatial heterogeneity is present when π is intermediate and f is large. We observed in Section 3 that such conditions favor unconditional dispersal. In contrast, the evolution of the conditional dispersal component m_H^* is promoted by decreasing π: Assuming $fp > 1$, which is necessary for positive dispersal to emerge, the quantity B_0^* is a decreasing function of π. In particular, $\lim_{\pi\to 0} B_0^* = fp - 1$ holds. Positive dispersal for small π thus evolves, when $f > 1/p$. Furthermore, $\lim_{\pi\to 1} B_0^* = -\lambda(1 - p) \leqslant 0$ holds, which means that $m_H^* = 0$ for $\pi \approx 1$ (Figs. 2ef and 6). More precisely, we have $m_H^* = 0$ for $\tilde{\pi} \leqslant \pi \leqslant 1$, in which

$$\tilde{\pi} = \frac{fp - 1}{fp - 1 + \lambda(1 - p)} < 1. \tag{15}$$

The motivation for an individual to emigrate from a high-productivity patch is the hope to immigrate into a low-productivity patch to avoid competition, which has a very low probability when $\pi \approx 1$. Here we observe that the existence of low-productivity patches together with their possibility to become high-productivity patches promotes the evolution of conditional dispersal.

In contrast with unconditional dispersal m^*, the singular strategy component m_H^* of conditional dispersal can either be a monotonically decreasing or non-monotonic function of π. Detailed conditions for the two cases are given by Theorem E.3: If $\frac{1}{p} < f < \frac{1}{p^2}$ or $f > \frac{1}{p^2}$ and $\lambda < \tilde{\lambda}$, where $\tilde{\lambda} = \frac{fp-1}{fp^2-1} > 1$, m_H^* is a non-increasing function of π (Fig. 2f). For $\tau \geqslant 0$, the singular strategy m_H^* is a non-monotonic function of π, if

$$\left(f > \frac{1}{p^2} \text{ and } \lambda > \tilde{\lambda} = \frac{fp - 1}{fp^2 - 1}\right) \Leftrightarrow \left(\lambda > \frac{1}{p} \text{ and } f > \frac{\lambda - 1}{p(\lambda p - 1)} =: \hat{f}\right). \tag{16}$$

For $\tau < 0$, the condition (16) is only a necessary condition for the singular strategy m_H^* to be a non-monotonic function of π (Fig. 6b).

According to Theorem E.3, zero dispersal $m_H^* = 0$ evolves for low productivity ratios, $1 \leqslant f \leqslant \tilde{f}$, in which $\tilde{f}$ is given in (E.18). When the productivity ratio is large enough, $f > \tilde{f}$, the strategy m_H^* is positive, and increases with f (Fig. 2h). The productivity ratio f can thus have qualitatively different effects on unconditional and conditional dispersal, since the unconditional dispersal strategy m^* was observed to be either increasing or non-monotonic with respect to f (Figs. 2d and 3). Spatiotemporal heterogeneity, in the form of substantial differences in the patch qualities, promotes the evolution of conditional dispersal.

Temporal heterogeneity (small τ) promotes evolution of unconditional dispersal as well as conditional dispersal. According to Theorem E.3, for $\tau = 1$ we have $m_H^* = 0$, and m_H^* increases when τ is decreased. This dependence is linear as long as $m_H^* < 1$. For small enough τ we may have $m_H^* = 1$ (Fig. 2g).

Analogous to the evolution of unconditional dispersal, the parameter plots of the singular dispersal strategy component m_H^* with respect to the proportion of high-productivity patches π and autocorrelation τ (as in Fig. 2e) may have qualitatively different forms depending on the other parameters. Fig. S.7 shows all qualitatively different types of such parameter plots. The details are explained in Appendix E.5. Furthermore, Fig. S.8 illustrates when each of these cases occurs depending on the choice of the other remaining parameters, productivity ratio f, dispersal survival probability p, and arrival bias λ.

4.4. Larger dispersal survival promotes dispersal

Positive dispersal $m_H^* > 0$ evolves if $B_0^* > 0$. The expression B_0 given by (14) is an increasing function of p. Furthermore, the condition $B_0^* > 0$ can be written as

$$p > \frac{1 - \pi + \lambda\pi}{\lambda\pi + f(1 - \pi)}. \tag{17}$$

According to Theorem E.3, the singular dispersal strategy component m_H^* is an increasing function of p (Fig. 7a). (Note, however, the neutrality of dispersal evolution for $p = 1$, explained in Section 4.2, proved in Theorem E.2.) We can thus conclude that increasing dispersal survival p promotes also the evolution of conditional dispersal.

4.5. The effect of arrival bias λ may be non-monotonic

In Section 4.3 we noticed that positive dispersal may evolve, $m_H^* > 0$, if λ is not too large. The expression B_0 given by (14) is a decreasing function of λ. Therefore, we can write the condition $B_0^* > 0$ as

$$\lambda < \frac{(fp - 1)(1 - \pi)}{(1 - p)\pi} \text{ and } fp > 1. \tag{18}$$

Increasing the arrival bias λ thus hinders the emergence of dispersal. Its effect on the singular strategy m_H^* may, however, be non-monotonic. According to Theorem E.3, if $\frac{1}{p} < f < \frac{1}{p^2}$, m_H^* decreases with respect to λ. If $f > \frac{1}{p^2}$, m_H^* is non-monotonic with respect to λ (Fig. 7b).

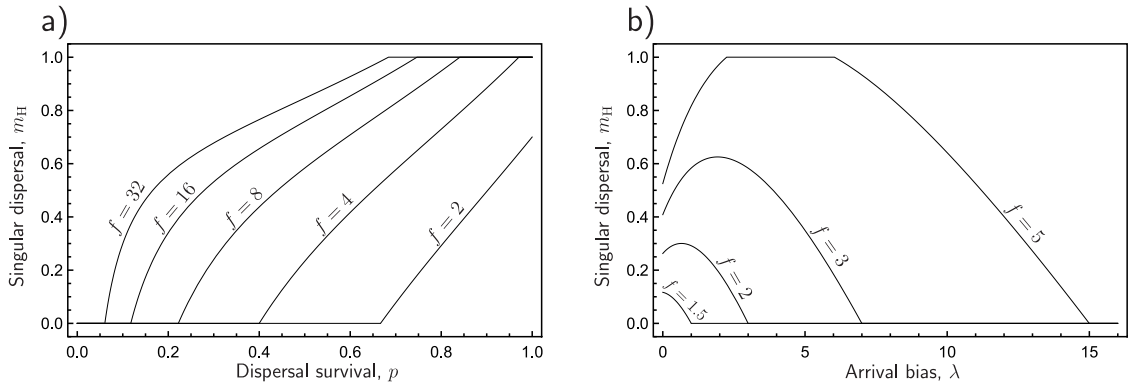

Fig. 7. The singular conditional dispersal strategy component m_H^* with respect to (a) dispersal survival probability p and (b) arrival bias λ. Dispersal evolves to zero ($m_H^* = 0$), when p is low or λ is large. (a) When p is increased, m_H^* increases, until it potentially reaches the strategy boundary $m_H^* = 1$. (b) Dispersal m_H^* is a decreasing function of λ, when $f < 1/p^2$ (= 1.5625). When $f > 1/p^2$, m_H^* is non-monotonic with respect to λ Parameters: (a) $\lambda = 1$ (b) $p = 0.8$. Common parameters in both panels: $\tau = -0.4$, $\pi = 0.5$.

5. Discussion

5.1. Summary

We extended previous models (Cohen and Levin, 1991; Massol and Débarre, 2015) of the evolution of unconditional dispersal in which patch qualities are spatially heterogeneous and temporally fluctuating, to include a case in which arrivals can be biased towards ($\lambda > 1$) or against ($\lambda < 1$) the high-productivity patches. Increasing λ thus makes competition in the high-productivity patches stronger. The extended model shows richer dynamics including evolutionary bistability and non-monotonic parameter dependence with respect to the productivity ratio f. We also studied the evolution of conditional dispersal. Surprisingly, we found that conditional dispersal from low-productivity patches never evolves. Furthermore, evolutionary branching occurs for $\tau < 0$ in the unconditional case, while it never occurs in the conditional case. Phenotypic plasticity can be expected to hinder evolutionary branching. Evolutionary bistability can occur only for unconditional dispersal.

The effects of parameters on the evolutionarily singular dispersal strategies are summarized in Table 2. For some of the parameters, their qualitative effect is similar for both unconditional and conditional dispersal. First, increasing the dispersal survival probability p increases both m^* (unconditional dispersal) and m_H^* (conditional dispersal). This result is rather intuitive, but note that increasing dispersal survival has been shown to decrease dispersal in some models (Comins et al., 1980; Gandon and Michalakis, 1999; Heino and Hanski, 2001). Second, increasing the temporal autocorrelation decreases both m^* and m_H^*. Temporal heterogeneity (decreasing τ) thus promotes dispersal. Third, the singular dispersal strategy either decreases when the arrival bias λ is increased, or is non-monotonic with respect to λ. For other parameters (f and π) there are differences in their qualitative effect.

The productivity ratio f is one of the parameters with qualitatively different effects for unconditional and conditional dispersal: Although in both cases f needs to be large enough for dispersal to emerge, the conditional dispersal strategy m_H^* is always increasing with respect to f, whereas the unconditional dispersal strategy can also be non-monotonic with respect to f. This can happen only when the cost of dispersal is small, $p \approx 1$, and there is arrival bias towards high-productivity patches, $\lambda > 1$.

The proportion of high-productivity patches π has two qualitatively different effects on unconditional and conditional dispersal. First, the unconditional dispersal strategy m^* is always non-monotonic with respect to π, whereas the conditional dispersal strategy m_H^* is either decreasing or non-monotonic with respect to π. Second, the unconditional dispersal strategy $m^* = 0$ for $\pi \approx 0$ or $\pi \approx 1$, whereas if conditional dispersal m_H^* evolves for some π, we have $m_H^* > 0$ for $\pi \approx 0$. The reason for the latter difference is as follows. Positive unconditional dispersal can evolve, if the average benefit of dispersal from high-productivity patches and from low-productivity patches is positive. When one patch type dominates the population, the selection pressure in dominant patches almost exclusively determines m^* (and the rare patches matter very little). Dispersal from the dominant patch type is not beneficial, because one may die during dispersal, and survivors arrive with high probability into a similar patch they left from. This is the reason why no unconditional dispersal evolves when π is close to either 0 or 1 (see Fig. 2b). Such evolution to no dispersal in a homogeneous environment has been observed before (Hastings, 1983; Holt, 1985; Cohen and Levin, 1991; Parvinen, 1999; Gyllenberg et al., 2002; Parvinen, 2006). In case of conditional dispersal, the mechanism described above explains also why $m_H^* = 0$ for $\pi \approx 1$ and $m_L^* = 0$ for $\pi \approx 0$. For conditional dispersal, each strategy component affects dispersal behavior in the corresponding patch type only. Dispersal from the rare patch type may evolve, if the living conditions in dominant patches are better. When high-productivity patches are rare, $\pi \approx 0$, individuals dispersing from high-productivity patches to low-productivity patches experience less competition, for which reason positive dispersal $m_H^* > 0$ can evolve. In our model, however, we always have $m_L^* = 0$, reasons for which are discussed in Section 5.2.

Consider next the effect of parameters on average dispersal. There are at least two different ways to take such an average, either from a patch perspective or from a juvenile perspective. For conditional dispersal, these averages are defined as

$$m_H^*\pi + m_L^*(1-\pi) \qquad \text{(patch-average)} \qquad (19a)$$

$$m_H^*\frac{\pi f}{\pi f + (1-\pi)} + m_L^*\frac{(1-\pi)}{\pi f + (1-\pi)}, \qquad \text{(juvenile-average)} \qquad (19b)$$

respectively, and our results tell us $m_L^* = 0$. For unconditional dispersal $m_H = m_L$, so both expressions simply become m^*. In the conditional dispersal case, the average dispersal rates (19) depend not only on the evolved dispersal rate from high-productivity patches m_H^*, but also on the proportion of high-productivity patches, π. Although m_H^* in Fig. 8a decreases with respect to π, both patch-average and juvenile-average dispersal rates are non-monotonic with respect to π, reaching a maximum at an intermediate proportion of high-productivity patches (Fig. 8). In both unconditional and conditional cases, average dispersal rate is large when both high- and low-productivity patches are present in substantial magnitudes and when the productivity ratio f is large. In this sense, we can say that spatiotemporal heterogeneity promotes dispersal.

As expected, the arrival bias λ has an effect on the evolutionarily singular strategies m^* (unconditional) and m_H^* (conditional), provided that spatiotemporal heterogeneity is present. If we would have $f = 1$ or $\tau = 1$, positive dispersal would not evolve, so that arrival bias ($\lambda \neq 1$)

Table 2
How parameters affect the singular dispersal strategies.

	unconditional dispersal m^*	conditional dispersal m_{H}^*	m_{L}^*
Larger dispersal survival, p	increasing	increasing	0
Larger productivity ratio, f	increasing OR non-monotonic, when $\lambda > 1$ and $p \approx 1$	increasing	0
Larger autocorrelation, τ	decreasing	decreasing	0
Larger proportion of high-productivity patches, π	non-monotonic	decreasing OR non-monotonic, when $f > \frac{1}{p^2}$ and $\lambda > \tilde{\lambda} > 1$	0
Larger arrival bias λ	decreasing OR non-monotonic	decreasing OR non-monotonic	0
Evolutionary branching	when $\tau_{\mathrm{thresh}} < \tau < 0$ and $0 < m^* < 1$	never	
Evolutionary bistability	when $\lambda > 1$, $\tau < \tau_{\mathrm{thresh}}$, and f is intermediate	never	

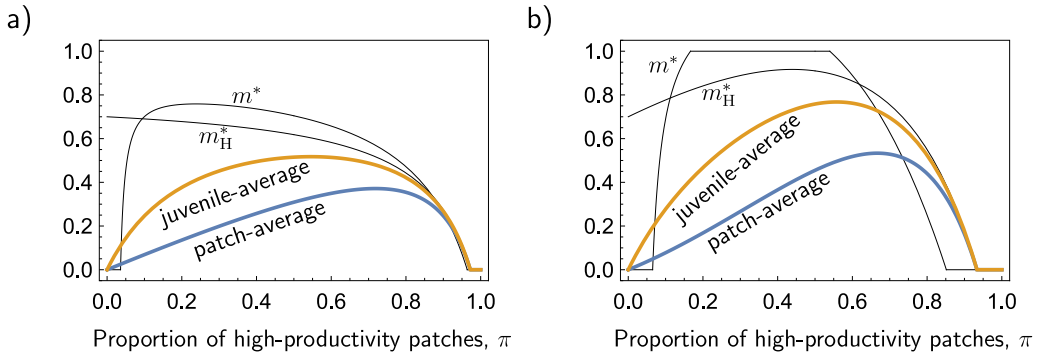

Fig. 8. Average dispersal rates are non-monotonic with respect to π for both unconditional and conditional dispersal. For unconditional dispersal, m^* is both the evolved dispersal strategy and the average dispersal rate (thin curve). For conditional dispersal, the dispersal strategy component m_{H}^* (thin curve) can be decreasing (panel a) with respect to π, but both the patch-average dispersal rate (thick blue curve) and the juvenile-average dispersal rate (thick orange curve), given by (19), are non-monotonic with respect to π. Parameters: (a) $\lambda = 1$ (b) $\lambda = 2.5$. Common parameters: $\tau = 0.1$, $p = 0.9$, $f = 5$.

alone does not promote dispersal. As explained above, m^* and m_{H}^* can be either decreasing or non-monotonic with respect to λ. However, including the arrival bias can also change the qualitative effect of other parameters on m^* and m_{H}^*. First, for $\lambda = 1$, the unconditional dispersal strategy is increasing with respect to f, but for $\lambda > 1$ the effect of f can be non-monotonic. Second, for $\lambda = 1$, the conditional dispersal strategy m_{H}^* is decreasing with respect to π, but for $\lambda > 1$ the effect of π can be non-monotonic (Fig. S.8). Furthermore, evolutionary bistability of unconditional dispersal is possible only for $\lambda > 1$. This is illustrated by Fig. S.4, from which we also observe that the boundaries of the parameter regions (red and purple) allowing for evolutionary bistability cross at $\lambda = 1$. Analogously, the parameter region for which full dispersal never evolves (light blue) has a corner at $\lambda = 1$. Consequently, unbiased dispersal $\lambda = 1$ is a special case, as an infinitesimally small change of λ allows richer evolutionary scenarios of unconditional dispersal evolution.

5.2. Why conditional dispersal from low-productivity patches does not evolve

One of our major findings is $m_{\mathrm{L}}^* = 0$, irrespective of parameter values. One may naively expect that dispersing offspring from a low-productivity patch could be adaptive especially when $\tau \sim 1$, but this is not the case. Here we intuitively explain why.

Let us consider one complete life-cycle starting at the moment immediately after emigration before dispersal survival selection, immigration, and patch-quality transition. At that moment, there are three different types of juveniles in the population; H-juveniles are those juveniles who decided to stay in a currently high-productivity patch, L-juveniles are those juveniles who decided to stay in a currently low-productivity patch, and dispersing juveniles are those juveniles who decided to disperse. Consider a special case when resident dispersal rates are negligible; $(m_{\mathrm{H}}, m_{\mathrm{L}}) \sim (0, 0)$, and hence the former two types of juveniles are mainly present in the population. Before patch-quality transition, in a high-productivity patch, $n_{\mathrm{H}}\gamma F_{\mathrm{H}}$ H-juveniles compete for adult spots. With probability $1-\beta$, the patch quality remains the same and those H-juveniles compete for n_{H} adult spots and each surviving adult produces γF_{H} H-juveniles in the next generation. Thus, the average number of H-juveniles produced by a single H-juvenile after the completion of one life-cycle is

$$\underbrace{\frac{n_{\mathrm{H}}}{n_{\mathrm{H}}\gamma F_{\mathrm{H}}}}_{\text{chance for becoming adult}} \cdot \underbrace{(\gamma F_{\mathrm{H}})}_{\text{production of H-juveniles}} = 1 \tag{20}$$

multiplied with $1-\beta$. On the other hand, with probability β the patch quality changes and H-juveniles compete for n_{L} adult spots and each surviving adult produces γF_{L} L-juveniles in the next generation. The average number of L-juveniles that are produced by a single H-juvenile after the completion of one life-cycle is

$$\underbrace{\frac{n_{\mathrm{L}}}{n_{\mathrm{H}}\gamma F_{\mathrm{H}}}}_{\text{chance for becoming adult}} \cdot \underbrace{(\gamma F_{\mathrm{L}})}_{\text{production of L-juveniles}} = \frac{1}{f} \tag{21}$$

multiplied with β. Similarly, an L-juvenile on average produces f many H-juveniles with probability α, and one L-juvenile with probability $1-\alpha$. Therefore, the resident population dynamics is governed by the matrix

$$\begin{pmatrix} 1-\beta & \alpha f \\ \beta/f & 1-\alpha \end{pmatrix}. \tag{22}$$

Its left eigenvector $(1, f)$ gives (relative) reproductive values of an H-juvenile and an L-juvenile, showing that an L-juvenile is f times more valuable than an H-juvenile. An intuition behind this is simple. Since dispersal is negligible, all descendants stay in the original patch. When one follows a family line of a single L-juvenile, its descendants are, on average, f H-juveniles when the focal patch productivity is high, and are one L-juvenile when the productivity is low. Thus an L-juvenile is f

times more valuable than an H-juvenile. For a more formal description of the argument here, see Appendix F.

If a mutant adult increases m_H (resp., m_L) from zero, he/she gains one dispersing juvenile in exchange of losing one natal H- (resp., L-) juvenile. Whether this is adaptive or not depends on gains/loss in reproductive values. Dispersing juvenile will survive with probability p and become either H- or L- juvenile depending on where it arrives, so its reproductive value is p times an weighted-average of 1 and f (see (F.7) in Appendix), which can be larger than 1, the reproductive value of an H-juvenile. So m_H can evolve from $(m_H, m_L) \sim (0,0)$, but the reproductive value of a dispersing juvenile can never be greater than f, the reproductive value of an L-juvenile, so m_L never evolves from $(m_H, m_L) \sim (0,0)$. Note that this result holds for any $\alpha > 0$ and $\beta > 0$ (or equivalently, for any $\tau < 1$). Even if patch quality rarely changes (i.e., $\tau \sim 1$), dispersal from high-productivity patches can evolve but dispersal from low-productivity patches never evolves from $(m_H, m_L) \sim (0,0)$.

5.3. Comparison with previous literature & future directions

Our model and the result for the unconditional dispersal without arrival bias ($\lambda = 1$) are identical to the model of "juvenile dispersal with local density regulation" by Massol and Débarre (2015). We extended the model to include arrival bias ($\lambda \neq 1$) and found richer dynamics such as evolutionary bistability, as summarized in Section 5.1.

We also studied the evolution of conditional dispersal and found a novel and simple result that dispersal from low-productivity patches never evolves. As Massol and Débarre (2015) have shown for the case of unconditional dispersal, our results for the conditional dispersal may also depend on the life cycle assumptions (see also Johst and Brandl (1997)).

For a case of evolution of conditional dispersal with spatial heterogeneity but without temporal fluctuation, McPeek and Holt (1992) studied a two-patch model and found the selectively neutral combination of (m_H, m_L). An analogous result was found by Parvinen (1999), Fig. 3b therein. These results correspond to our neutral isocline with $\tau = 1$ and $p = 1$ (Fig. 5b). McPeek and Holt (1992) also studied cases with spatiotemporal fluctuation, but the strategy they investigated was assumed to be conditional on the patch index, not on the patch quality. Thus our results with spatiotemporal heterogeneity have no direct correspondence to their results.

We introduced arrival bias λ in immigration as a model parameter. The differences in arrival bias could be caused by active decisions by dispersers when they encounter potential patches to immigrate into. Alternatively, patches with different qualities could possess different vegetation, so that the attachment probabilities of seeds dispersing by wind would be different. Although one might expect arrival bias towards high-productivity patches ($\lambda > 1$) to be advantageous, there is typically less competition in the low-productivity patches. When $\lambda < 1$, dispersers may thus experience less competition. Alternatively, some studies have considered biased immigration as evolutionary traits (Gyllenberg et al., 2016; Parvinen and Brännström, 2016; Nurmi et al., 2018), in which case migrants are able to actively choose certain types of patches for settlement. In our current model, we have considered only patch-quality-dependent emigration strategy, but considering patch-quality-dependent immigration strategy and its coevolution with emigration strategy (i.e., Weigang (2017)) will further enrich our understanding of evolution of dispersal traits.

If we assume the local population sizes to be finite, kin selection will be present, which is expected to promote dispersal (e.g. Hamilton and May, 1977). Concerning conditional dispersal, one would thus expect positive dispersal to evolve from both patch types. That is, our prediction $m_L = 0$ might not hold for models with finite local population sizes. However, depending on parameter values, dispersal from one patch type may evolve to zero, while dispersal from the other patch type evolves to a positive value, analogous to our result in Section 4.

We have studied a particular life-cycle assumption in our model, but we could consider many of its variants. As Massol and Débarre (2015) stressed, the order of events in life-cycle assumptions can dramatically change the evolutionary outcomes, and our prediction here may not be robust against changing those details.

For mathematical simplicity, we have assumed in our model that there are only two different patch qualities and that each patch follows the same transition rule between those quality states. However, for a full understanding of the roles of spatial and temporal heterogeneity, more studies are needed. Deviations from our assumptions to include a more general situation, such as studying a model with $N(> 2)$ quality states, would be more difficult mathematically, but would help us find a more general conclusion about how spatiotemporal heterogeneity contributes to the evolution of dispersal. Furthermore, one could study the evolution of dispersal when evolving species could adapt to local environmental conditions (Balkau and Feldman, 1973; Nurmi and Parvinen, 2011, 2013; Blanquart and Gandon, 2014).

In summary, we have studied evolution of dispersal in a heterogeneous environment with arrival bias. We have found that arrival bias introduces richer evolutionary dynamics including bistability. We have also found that evolutionary outcomes are strikingly different between unconditional and conditional dispersal. We believe that those results shed light on ecological factors that impact dispersal evolution.

CRediT authorship contribution statement

Kalle Parvinen: Conceptualization, Writing, Formal analysis, Visualization. **Hisashi Ohtsuki:** Conceptualization, Writing, Formal analysis. **Joe Yuichiro Wakano:** Conceptualization, Writing, Visualization.

Declaration of competing interest

None

Acknowledgments

HO acknowledges the support by JSPS KAKENHI Grant Number JP20K06812. JYW acknowledges the support by JSPS KAKENHI Grant Number JP21K03357 and by MEXT KAKENHI Grant Number JP16H06412.

Appendix A. Metapopulation fitness for general N

A.1. Resident emigrants

In the initial phase of potential invasion by mutants, practically all patches are occupied by residents only. In patches of type k, the adult population density is n_k, and adults reproduce with fecundity γF_k. Juveniles disperse with probability m_k. Since the proportion of patches of type k is π_k, the average amount of resident emigrants is

$$\sum_{l=1}^{N} \pi_l m_l \underbrace{n_l \gamma F_l}_{\substack{\#\text{ of juveniles produced} \\ \text{in a patch of type } l}} \tag{A.1}$$

A.2. Resident immigrants

Emigrants will survive dispersal with probability p. The parameters λ_k describe the relative attractivity of patches of quality k, so that emigrants arrive into a patch of type k with probability

$$\phi_k = \frac{\lambda_k \pi_k}{\sum_{l=1}^{N} \lambda_l \pi_l} = \hat{\lambda}_k \pi_k \text{ in which } \hat{\lambda}_k = \frac{\lambda_k}{\sum_{l=1}^{N} \lambda_l \pi_l}. \tag{A.2}$$

Note that this definition differs from the one used by Parvinen et al. (2020), who used the arrival probabilities ϕ_k as parameters, which then

determined arrival bias parameters. In our model, the patch proportions are not fixed as such (although they eventually do converge to fixed values), so it is not meaningful to define fixed arrival probabilities ϕ_k.

According to (A.1) and (A.2), the amount of resident immigrants arriving into a patch of type k is

$$p\frac{\phi_k}{\pi_k}\sum_{l=1}^{N}\pi_l m_l n_l \gamma F_l = \gamma\frac{\lambda_k}{\sum_{l=1}^{N}\lambda_l\pi_l}\cdot \underbrace{p\sum_{l=1}^{N}\pi_l m_l n_l F_l}_{\equiv I_{\mathrm{res}}} = \gamma\frac{\lambda_k}{\sum_{l=1}^{N}\lambda_l\pi_l} I_{\mathrm{res}}$$
$$= \gamma\hat{\lambda}_k I_{\mathrm{res}}. \tag{A.3}$$

If all patches are equally attractive ($\lambda_k = 1$), we obtain from (A.2) that $\hat{\lambda}_k = 1$ for all k, and the probability ϕ_k to arrive into a patch of quality k is equal to the proportion π_k of such patches. In such case each patch will receive the same amount of immigrants, γI_{res}, independent of its quality.

According to (A.3), each patch of type k receives $\gamma\hat{\lambda}_k I_{\mathrm{res}}$ immigrant residents, and there are $(1-m_k)\gamma F_k n_k$ natal residents, so the total population density in patches of type k after immigration is

$$(1-m_k)\gamma F_k n_k + \gamma\hat{\lambda}_k I_{\mathrm{res}} \tag{A.4}$$

A.3. Competition

After immigration, each patch experiences the possibility of transition of patch type. Consider a patch that was of type k at time t, and is of type j at time $t+1$. Let $\frac{1}{\gamma}S_{j\leftarrow k}$ denote the settlement probability in such patches. Since n_j is the population density after competition, we have

$$S_{j\leftarrow k} = \frac{n_j}{(1-m_k)F_k n_k + \hat{\lambda}_k I_{\mathrm{res}}} \tag{A.5}$$

The settlement probability thus depends both on the current and previous quality of the patch.

A.4. Metapopulation fitness

In the moment after immigration, but before transition, consider a mutant juvenile present in a patch of type k. This mutant and all its descendants remaining in this patch will be called the mutant colony. Over the years, this mutant colony will send emigrants. Let R_k denote the total amount of emigrants sent by the mutant colony founded by the focal mutant juvenile. The patch type after transition will be j with probability $P(j\leftarrow k)$. In such case, the settlement probability for the mutant juvenile is $\frac{1}{\gamma}S_{j\leftarrow k}$. It will reproduce with fecundity γF_j. The proportion $m_{\mathrm{mut},j}$ of the offspring will disperse and survive dispersal with probability p. On the other hand, the proportion $1-m_{\mathrm{mut},j}$ of the offspring will remain in this patch, and will eventually produce R_j emigrants. We thus obtain

$$\begin{aligned} R_k &= \sum_{j=1}^{N} P(j\leftarrow k)\frac{1}{\gamma}S_{j\leftarrow k}\gamma F_j(m_{\mathrm{mut},j}p + (1-m_{\mathrm{mut},j})R_j) \\ &= \sum_{j=1}^{N} P(j\leftarrow k)S_{j\leftarrow k}F_j(m_{\mathrm{mut},j}p + (1-m_{\mathrm{mut},j})R_j) \\ &= \sum_{j=1}^{N} P(j\leftarrow k)\frac{F_j n_j}{(1-m_k)F_k n_k + \hat{\lambda}_k I_{\mathrm{res}}}(m_{\mathrm{mut},j}p + (1-m_{\mathrm{mut},j})R_j). \end{aligned} \tag{A.6}$$

Note that the scaling factors γ in this equation cancel each other. These relations form a system of linear equations, from which $R_1,\ldots,R_N$ can be solved, at least numerically. A dispersing juvenile will arrive at a patch of type j with probability ϕ_j, and the metapopulation fitness is

$$R = \sum_{j=1}^{N}\phi_j R_j = \sum_{j=1}^{N}\hat{\lambda}_j\pi_j R_j = \frac{1}{\sum_{l=1}^{N}\lambda_l\pi_l}\sum_{j=1}^{N}\lambda_j\pi_j R_j. \tag{A.7}$$

From the last expression of (A.6) we observe that the relative fecundity and local population density always appear as a product, $F_j n_j$ or $F_k n_k$. Therefore, these parameters affect dispersal evolution only through their product.

A.5. Metapopulation fitness from adult perspective

From a viewpoint of an adult, a focal adult mutant in a patch of type k produces F_k juveniles, among which the fraction $1-m_{\mathrm{mut},k}$ stays in the natal patch whose type might change from k to j. Thus, the number of emigrants produced by the focal adult and its descendants is given by

$$\tilde{R}_k = F_k(1-m_{\mathrm{mut},k})\sum_{j=1}^{N}P(j\leftarrow k)S_{j\leftarrow k}\tilde{R}_j + F_k m_{\mathrm{mut},k}p \tag{A.8}$$

To combine them to obtain metapopulation fitness, we again calculate the number of emigrants produced by a mutant disperser that has just landed in a patch of type j. The patch type might change from j to k, and the juvenile settles there to become an adult in a patch of type k. Such an adult will have a metapopulation fitness $\tilde{R}_k$. Thus we have

$$R_j = \sum_{k=1}^{N}P(k\leftarrow j)S_{k\leftarrow j}\tilde{R}_k \tag{A.9}$$

Actually the two methods yield the identical expression of R, irrespective of the explicit forms of $P(j\leftarrow k)$ and $S_{j\leftarrow k}$.

Appendix B. Spatial and temporal heterogeneity for two patch qualities

Let us now especially consider the case of two qualities, $q\in\{1,2\}$. In the main text, label 1 corresponds to High productivity, and label 2 corresponds to Low productivity, but the following calculations hold for any two patch-type situations. The transition is governed by a time-homogeneous Markov chain, in which the transition probabilities $P(j\leftarrow i)$ $(i,j\in\{1,2\})$ describing the probability that the quality of a patch changes from i to j are

$$\begin{aligned} &P(1\leftarrow 1) = 1-\beta, \quad && P(1\leftarrow 2) = \alpha, \\ &P(2\leftarrow 1) = \beta, && P(2\leftarrow 2) = 1-\alpha, \end{aligned} \tag{B.1}$$

where $0<\alpha,\beta<1$. Let the random variable $\mathbf{X}(t)$ denote the quality of a focal patch at time t, and $p_i(t) = P(\mathbf{X}(t)=i)$ the probability that the quality of the focal patch is i at time t. These probabilities satisfy

$$\begin{pmatrix} p_1(t+1) \\ p_2(t+1)\end{pmatrix} = \begin{pmatrix} 1-\beta & \alpha \\ \beta & 1-\alpha\end{pmatrix}\begin{pmatrix} p_1(t) \\ p_2(t)\end{pmatrix}. \tag{B.2}$$

Since the eigenvalues of the transition matrix are 1 and $1-\alpha-\beta$, we obtain

$$\begin{pmatrix} p_1(t) \\ p_2(t)\end{pmatrix} = \begin{pmatrix} 1-\beta & \alpha \\ \beta & 1-\alpha\end{pmatrix}^t\begin{pmatrix} p_1(0) \\ p_2(0)\end{pmatrix} = \begin{pmatrix} \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta}\end{pmatrix} + c(1-\alpha-\beta)^t\begin{pmatrix} 1 \\ -1\end{pmatrix}, \tag{B.3}$$

in which $c = p_1(0)-\alpha/(\alpha+\beta)$. Since we assume $0<\alpha,\beta<1$, we have $-1<1-\alpha-\beta<1$, so that $(1-\alpha-\beta)^t\to 0$ as $t\to\infty$. Therefore, after some transient, the probability distribution of quality of each patch becomes independent of its initial condition and of that of other patches as well. Consequently, the probabilities that a given patch at a given census time is in quality 1 and 2 respectively converges to π_1 and π_2, where they satisfy

$$\begin{pmatrix} \pi_1 \\ \pi_2\end{pmatrix} = \begin{pmatrix} P(1\leftarrow 1) & P(1\leftarrow 2) \\ P(2\leftarrow 1) & P(2\leftarrow 2)\end{pmatrix}\begin{pmatrix} \pi_1 \\ \pi_2\end{pmatrix} = \begin{pmatrix} 1-\beta & \alpha \\ \beta & 1-\alpha\end{pmatrix}\begin{pmatrix} \pi_1 \\ \pi_2\end{pmatrix} \tag{B.4}$$

with $\pi_1+\pi_2 = 1$. The solution is

$$\pi_1 = \frac{\alpha}{\alpha+\beta}, \tag{B.5}$$

which we call π in the main text.

The meaning of parameter τ in Eq. (3) in the main text can be explained by following the argument by Rodrigues and Gardner (2012), as follows. Let us introduce a dummy variable S_t, which is equal to 1 if the patch quality at tth census time is 1 and is equal to 0 if the quality is 2. Then, the correlation coefficient between S_t and S_{t+1} of the same patch is deemed as a quantity that measures how similar/different

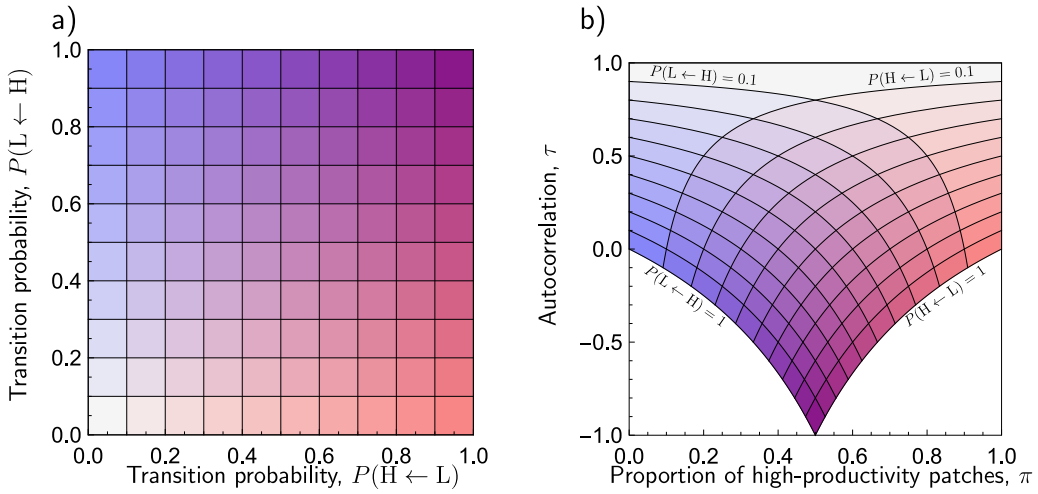

Fig. S.1. Dependence of τ and π on α and β.

the patch quality is between two adjacent census times, which takes a value between -1 (perfect negative correlation) and 1 (perfect positive correlation). Let τ denote this correlation coefficient. By definition we have

$$\tau = \frac{\mathrm{Cov}[S_t, S_{t+1}]}{\sqrt{\mathrm{Var}[S_t]}\sqrt{\mathrm{Var}[S_{t+1}]}} \tag{B.6}$$

Each term is evaluated as

$$\begin{aligned}\mathrm{Cov}[S_t, S_{t+1}] &= \mathrm{E}[S_t S_{t+1}] - \mathrm{E}[S_t]\cdot\mathrm{E}[S_{t+1}] \\ &= \Pr[S_t = 1]\cdot\Pr[S_{t+1} = 1 \mid S_t = 1] - \Pr[S_t = 1] \\ &\quad\cdot \Pr[S_{t+1} = 1] \\ &= (1-\beta)\pi_1 - \pi_1^2\end{aligned} \tag{B.7}$$

and

$$\begin{aligned}\big(\mathrm{Var}[S_{t+1}] =\big)\,\mathrm{Var}[S_t] &= \mathrm{E}[S_t^2] - \mathrm{E}[S_t]^2 = \Pr[S_t = 1] - \Pr[S_t = 1]^2 \\ &= \pi_1 - \pi_1^2,\end{aligned} \tag{B.8}$$

so we have

$$\tau = \frac{(1-\beta)\pi_1 - \pi_1^2}{\pi_1 - \pi_1^2} = \frac{(1-\beta) - \alpha/(\alpha+\beta)}{1 - \alpha/(\alpha+\beta)} = 1 - (\alpha+\beta), \tag{B.9}$$

where we used (B.5). This expression makes intuitive sense because larger α and β mean that the patch quality frequently changes and hence that the stability of patch quality τ is low. By solving Eqs. (B.5), (B.9) with respect to α and β, we obtain

$$\alpha = \pi_1(1-\tau), \beta = (1-\pi_1)(1-\tau). \tag{B.10}$$

Therefore, our model is parameterized either by (α, β) or by (π_1, τ). In the latter case, its domain is a curved pentagon-like region represented as

$$\left\{(\pi_1, \tau) \mid 0 < \pi_1 < 1, \max\left[-\frac{\pi_1}{1-\pi_1}, -\frac{1-\pi_1}{\pi_1}\right] < \tau < 1\right\}, \tag{B.11a}$$

or, equivalently,

$$\left\{(\pi_1, \tau) \mid -1 < \tau < 1, \begin{array}{ll} 0 < \pi_1 < 1, & \text{for } 0 \leqslant \tau < 1 \\ -\tau/(1-\tau) < \pi_1 < 1/(1-\tau), & \text{for } -1 < \tau < 0 \end{array}\right\}, \tag{B.11b}$$

as illustrated in Fig. S.1.

The limit of no temporal heterogeneity, $\tau \to 1$, can be obtained by letting the transition probabilities α and β tend to zero, while keeping their relative magnitudes constant: $\alpha = \epsilon\pi$ and $\beta = \epsilon(1-\pi)$ and $\epsilon \to 0$. Alternatively, we can assume that the patch qualities do not change, $\alpha = \beta = 0$, and assume that the patch proportions are π and $1-\pi$. Both approaches result in the same expression of metapopulation fitness, which means that our results for $\tau \to 1$ are the same as for $\tau = 1$.

The limit $\tau \to -1$, can be obtained by letting $\alpha \to 1$ and $\beta \to 1$, so that $\pi \to 1/2$. That limit, however, is not consistent with starting with $\alpha = \beta = 1$. A Markov chain with such a transition matrix has the equilibrium $(1/2, 1/2)$, but it is not asymptotically stable. From any other initial condition $(x, 1-x)$, the Markov chain fluctuates between the states $(x, 1-x)$ and $(1-x, x)$. In other words, it is not meaningful to consider the limit $\tau \to -1$.

Appendix C. Explicit expression for the metapopulation fitness for $N = 2$

C.1. Arrival probabilities

Concerning the relative attractivity parameters λ_k, only their relative magnitude matters. According to (A.2), we have

$$\hat{\lambda}_1 = \frac{\lambda_1}{\lambda_1\pi_1 + \lambda_2\pi_2} = \frac{\lambda_1/\lambda_2}{(\lambda_1/\lambda_2)\pi + (1-\pi)} = \frac{\lambda}{\lambda\pi + (1-\pi)}. \tag{C.1}$$

Analogously, we have

$$\hat{\lambda}_2 = \frac{1}{\lambda\pi + (1-\pi)}. \tag{C.2}$$

Therefore, the arrival probabilities are

$$\phi_1 = \hat{\lambda}_1\pi_1 = \frac{\lambda\pi}{\lambda\pi + (1-\pi)} \text{ and } \phi_2 = \hat{\lambda}_2\pi_2 = \frac{1-\pi}{\lambda\pi + (1-\pi)}. \tag{C.3}$$

In the case of two patch types, metapopulation fitness thus depends on the parameters λ_1 and λ_2 only through $\lambda = \frac{\lambda_1}{\lambda_2}$, which measures the attractivity of high-productivity patches relative to the low-productivity patches.

C.2. Pair of linear equations for the metapopulation fitness components

The system of Eqs. (A.6) in the case of two patch types is

$$\begin{aligned}R_1 &= \frac{P(1\leftarrow 1)F_1 n_1}{(1-m_1)F_1 n_1 + \hat{\lambda}_1 I_{\mathrm{res}}}(m_{\mathrm{mut},1}p + (1-m_{\mathrm{mut},1})R_1) \\ &\quad + \frac{P(2\leftarrow 1)F_2 n_2}{(1-m_1)F_1 n_1 + \hat{\lambda}_1 I_{\mathrm{res}}}(m_{\mathrm{mut},2}p + (1-m_{\mathrm{mut},2})R_2), \\ R_2 &= \frac{P(1\leftarrow 2)F_1 n_1}{(1-m_2)F_2 n_2 + \hat{\lambda}_2 I_{\mathrm{res}}}(m_{\mathrm{mut},1}p + (1-m_{\mathrm{mut},1})R_1) \\ &\quad + \frac{P(2\leftarrow 2)F_2 n_2}{(1-m_2)F_2 n_2 + \hat{\lambda}_2 I_{\mathrm{res}}}(m_{\mathrm{mut},2}p + (1-m_{\mathrm{mut},2})R_2),\end{aligned} \tag{C.4}$$

in which $I_{\mathrm{res}} = p(\pi m_1 n_1 F_1 + (1-\pi)m_2 n_2 F_2)$. In Appendix A.4 we observed from the last expression of (A.6) that the metapopulation

fitness, and thus the evolution of dispersal, depends on the fecundities F_k and population densities n_k only through their product, $F_k n_k$. Here we observe that such dependence in (C.4) occurs only through $f = \frac{F_1 n_1}{F_2 n_2}$ describing the productivity ratio of patches of quality 1. Without loss of generality we can assume $f > 1$, so that patch quality 1 corresponds to higher productivity. Furthermore, with the help of (B.1) and (B.10), (C.4) becomes

$$\begin{aligned} R_1 &= \frac{(\pi+\tau(1-\pi))f(m_{\mathrm{mut},1}p+(1-m_{\mathrm{mut},1})R_1)}{(1-m_1)f+\hat{\lambda}_1 p(\pi m_1 f+(1-\pi)m_2)} \\ &\quad + \frac{(1-\pi)(1-\tau)(m_{\mathrm{mut},2}p+(1-m_{\mathrm{mut},2})R_2)}{(1-m_1)f+\hat{\lambda}_1 p(\pi m_1 f+(1-\pi)m_2)}, \\ R_2 &= \frac{\pi(1-\tau)f(m_{\mathrm{mut},1}p+(1-m_{\mathrm{mut},1})R_1)}{(1-m_2)+\hat{\lambda}_2 p(\pi m_1 f+(1-\pi)m_2)} \\ &\quad + \frac{(1-\pi+\pi\tau)(m_{\mathrm{mut},2}p+(1-m_{\mathrm{mut},2})R_2)}{(1-m_2)+\hat{\lambda}_2 p(\pi m_1 f+(1-\pi)m_2)}. \end{aligned} \tag{C.5}$$

C.3. Metapopulation fitness

The pair of linear equations (C.5) can be solved explicitly. Substituting the solution into (A.7), and applying (C.1), (C.2) and (C.3) we obtain

$$R((m_{\mathrm{mut},1},m_{\mathrm{mut},2});(m_1,m_2)) = \phi_1 R_1 + \phi_2 R_2. \tag{C.6}$$

The electronic appendix illustrates how this expression is obtained with Mathematica.

Appendix D. Evolution of unconditional dispersal

We will first study the evolution of unconditional dispersal, i.e, assume that the patch quality has no effect on the dispersal strategy of an individual, so that $m_1 = m_2 = m$ and $m_{\mathrm{mut},1} = m_{\mathrm{mut},2} = m_{\mathrm{mut}}$.

D.1. Fitness gradient

The fitness gradient is the first derivative of the metapopulation fitness (C.6) with respect to the mutant dispersal strategy, evaluated when the mutant dispersal strategy is equal to that of the resident.

$$D(m) = \frac{\partial}{\partial m_{\mathrm{mut}}} R((m_{\mathrm{mut}},m_{\mathrm{mut}});(m,m))\Big|_{m_{\mathrm{mut}}=m}. \tag{D.1}$$

This calculation leads to the following theorem:

Theorem D.1. *The fitness gradient for unconditional dispersal is*

$$D(m) = \frac{1}{\tilde{A}}(Z+mY), \tag{D.2}$$

in which

$$\begin{aligned} \tilde{A} &= pm(1-\pi+f\pi)[\lambda pm(1-\pi+f\pi)+(1-\tau)(1-\pi+\lambda\pi)(1-m) \\ &\quad \times(\lambda\pi+f(1-\pi))] \geqslant 0, \\ Z &= (1-\tau)(1-\pi+\lambda\pi)\hat{D}_0, \\ \hat{D}_0 &= -f(1-p)(1-\pi+\pi\lambda)+p\pi(1-\pi)(f-1)(f-\lambda), \\ Y &= f(1-p)(1+(-1+\lambda)\pi)^2(1-\tau)-p(1-p)\lambda(1-\pi+\pi f)^2 \\ &\quad -(\lambda-f)^2 p(1-\pi)\pi. \end{aligned} \tag{D.3}$$

Since the coefficient $\tilde{A}$ is positive, the direction of selection is completely determined by $Z+mY$, which is linear with respect to m. Therefore, there exists at most one singular strategy $m^* = -Z/Y$, provided that $0 < m^* < 1$. Before investigating the properties of the singular strategy, consider the attractivity of the boundaries of the strategy space.

D.2. Attractivity of zero dispersal

The fitness gradient $D(0)$ is not defined, but we can determine the direction of selection at zero dispersal by calculating the limit of (D.2) when $m \to 0$,

$$\lim_{m\to 0} mD(m) = \frac{\hat{D}_0}{p(f\pi+1-\pi)(f(1-\pi)+\lambda\pi)}. \tag{D.4}$$

The denominator of (D.4) is clearly positive. The sign is thus determined by the numerator $\hat{D}_0$ given by (6):

$$\hat{D}_0 = \underbrace{-f(1-p)(1-\pi+\pi\lambda)}_{\leqslant 0} + \underbrace{p\pi(1-\pi)(f-1)}_{\geqslant 0}(f-\lambda) \tag{D.5}$$

The first part of $\hat{D}_0$ is either zero, when $p = 1$, and negative otherwise. The second part of $\hat{D}_0$ has the multiplier $\pi(1-\pi)$, which is zero when there is no spatial heterogeneity, $\pi = 0$ or $\pi = 1$. This multiplier reaches its maximum for $\pi = \frac{1}{2}$. The sign of the second term is determined by $f-\lambda$. If $f > \lambda$, the second term is positive for intermediate π, and zero dispersal may be repelling. If $1 < f < \lambda$, zero dispersal is attracting for all π.

Fig. 2a illustrates the evolution of unconditional dispersal with respect to the proportion of high-productivity patches π and autocorrelation τ. As expected, dispersal does not evolve, when $\pi \approx 0$ or $\pi \approx 1$ (areas with light-gray shading). Dispersal evolves to a positive singular strategy for intermediate values of π (color shading ranging from blue to red). The curves separating these two areas are straight vertical lines, because the expression $\hat{D}_0$ does not depend on autocorrelation τ. By solving $\hat{D}_0 = 0$ for π we obtain

$$\begin{aligned} &\pi_{\mathrm{min,max}} \\ &= \frac{1}{2} + \frac{(\lambda-1)f(1-p) \pm \sqrt{[p(f-1)(f-\lambda)+(\lambda-1)f(1-p)]^2 - 4p(f-1)(f-\lambda)f(1-p)}}{2p(f-1)(f-\lambda)} \\ &= \frac{1}{2} \pm \frac{\sqrt{p^2(f-1)^2-4pf(1-p)}}{2p(f-1)}, \text{ for } \lambda = 1 \end{aligned} \tag{D.6}$$

By solving $\hat{D}_0 = 0$ for f we obtain

$$\begin{aligned} \hat{f} &= \frac{\lambda+1}{2} + \frac{(1-p)(1-\pi+\pi\lambda) \pm \sqrt{[(\lambda+1)p\pi(1-\pi)+(1-p)(1-\pi+\pi\lambda)]^2 - 4\lambda p^2\pi^2(1-\pi)^2}}{2p\pi(1-\pi)} \\ &= 1 + \frac{(1-p) \pm \sqrt{[2p\pi(1-\pi)+(1-p)]^2 - 4p^2\pi^2(1-\pi)^2}}{2p\pi(1-\pi)}, \text{ for } \lambda = 1 \end{aligned} \tag{D.7}$$

D.3. Attractivity of full dispersal

According to (D.2), fitness gradient at $m = 1$ is given by

$$D(1) = \frac{Z+Y}{\lambda p^2(1-\pi+f\pi)^2}, \tag{D.8}$$

and its sign is determined by

$$\begin{aligned} Z+Y = &\underbrace{(1-\tau)(1-\pi+\lambda\pi)p\pi(1-\pi)(f-1)}_{\geqslant 0}(f-\lambda) \\ &\underbrace{-p(1-p)\lambda(1-\pi+\pi f)^2-(f-\lambda)^2 p\pi(1-\pi)}_{\leqslant 0}. \end{aligned} \tag{D.9}$$

When $f \leqslant \lambda$ we have $Z+Y \leqslant 0$. Strict equality holds only in special circumstances ($p = 1$ and $\pi(1-\pi) = 0$), otherwise the full-dispersal strategy is evolutionarily repelling. When $f > \lambda$, $D(1) \leqslant 0$ when $\tau \geqslant \tau_{\mathrm{thresh}}$, and $D(1) > 0$ when $\tau < \tau_{\mathrm{thresh}}$, in which

$$\begin{aligned} \tau_{\mathrm{thresh}} &= 1 - \frac{(1-p)\lambda(1-\pi+\pi f)^2+(f-\lambda)^2\pi(1-\pi)}{(1-\pi+\lambda\pi)\pi(1-\pi)(f-1)(f-\lambda)} \\ &= 1 - \frac{(1-p)\lambda(1-\pi+\pi f)^2}{(1-\pi+\lambda\pi)\pi(1-\pi)(f-1)(f-\lambda)} - \frac{(f-\lambda)}{(1-\pi+\lambda\pi)(f-1)}. \end{aligned} \tag{D.10}$$

In Fig. 2a, this is the curve separating the areas of positive singular strategy (color shading ranging from blue to red) and full dispersal $m^* = 1$ (dark gray shading).

D.4. Outcomes of dispersal evolution

Based on the results above, there are four qualitatively different scenarios of monomorphic evolution of unconditional dispersal. When the potential of evolutionary branching is included (see Theorem D.3), there are five scenarios listed below and illustrated in Fig. 1.

Theorem D.2.

1. *If $\hat{D}_0 < 0$ and $Z + Y < 0$, the zero-dispersal strategy $m = 0$ is evolutionarily attracting, and the full-dispersal strategy $m = 1$ is evolutionarily repelling. Evolution converges to $m = 0$ from all initial conditions (Fig. 1a).*
2. *If $\hat{D}_0 > 0$ and $Z + Y > 0$, the zero-dispersal strategy $m = 0$ is evolutionarily repelling, and the full-dispersal strategy $m = 1$ is evolutionarily attracting. Evolution converges to $m = 1$ from all initial conditions (Fig. 1d).*
3. *If $\hat{D}_0 < 0$ and $Z + Y > 0$, the zero-dispersal strategy $m = 0$ is locally evolutionarily attracting, and the full-dispersal strategy $m = 1$ is locally evolutionarily attracting. A unique evolutionarily singular strategy m^* exists, and it is evolutionarily repelling. If $m < m^*$, evolution converges to $m = 0$, and if $m > m^*$, evolution converges to $m = 1$ (Fig. 1f).*
4. *If $\hat{D}_0 > 0$ and $Z + Y < 0$, the zero-dispersal strategy $m = 0$ is evolutionarily repelling, and the full-dispersal strategy $m = 1$ is evolutionarily repelling. A unique evolutionarily singular strategy m^* exists, and it is evolutionarily attracting. Evolution converges to m^* from all initial conditions. If $\tau > 0$, the singular strategy is uninvadable (Fig. 1b). If $\tau < 0$, evolutionary branching is expected to occur (Fig. 1c).*

Consequently, potential evolutionary endpoints are

$$-\frac{Z}{Y} = \begin{cases} 0, & \text{if } \hat{D}_0 < 0 \\ -\dfrac{(1-\tau)(1-\pi+\lambda\pi)\hat{D}_0}{Y}, & \text{if } Z > 0 \text{ and } Z + Y < 0 \\ 1, & \text{if } Z + Y > 0. \end{cases} \tag{D.11}$$

D.5. Evolutionary branching

Theorem D.3. *If the singular strategy m^* exists ($0 < m^* < 1$) and is evolutionarily attracting, evolutionary branching occurs if and only if $\tau < 0$.*

Proof. The second derivative at singularity is

$$\frac{\partial^2}{\partial m_{\text{mut}}^2} R\bigg|_{m_{\text{mut}}=m=m^*} = -2f(1-p)\tau \frac{C^2}{(f-\lambda)\hat{D}_0 B_1 B_2}, \tag{D.12}$$

where the components C, B_1 and B_2 are listed below.

The component $\hat{D}_0$ is the numerator of the scaled selection gradient for zero dispersal, given in (D.4). It is positive, when the zero-dispersal strategy is evolutionarily repelling, which is a necessary condition for the singular strategy to be evolutionarily attracting. Furthermore, $f > \lambda$ is a necessary condition for the zero-dispersal strategy is evolutionarily repelling. Therefore, assuming that the singular strategy is evolutionarily attracting, we have $(f-\lambda)\hat{D}_0 > 0$.

The component

$$\begin{aligned} C =& f^2 p\pi(\lambda(p-1)\pi + \pi - 1) \\ & - f\left(2\lambda p^2(\pi-1)\pi - p(\tau-1)((\lambda-1)\pi+1)^2 + (\tau-1)((\lambda-1)\pi+1)^2\right) \\ & + \lambda p(\pi-1)((\lambda-1)\pi + p(\pi-1)+1) \end{aligned} \tag{D.13}$$

does not affect the sign of (D.12).

The component

$$B_1 = p^2(1-\pi)\pi(\tau-1)^2(f\pi+1-\pi)(\lambda\pi+1-\pi) \geqslant 0 \tag{D.14}$$

is clearly non-negative. It is zero, when $p = 0$, $\pi = 0$, $\pi = 1$, or $\tau = 1$, which all are such special cases that dispersal does not evolve.

Finally,

$$B_2 = \underbrace{\pi(1-\tau)(f-\lambda)(f-1)}_{\geqslant 0} + \underbrace{(f(f-1)}_{\geqslant 0}\tau + \underbrace{(f-\lambda)}_{>0}). \tag{D.15}$$

Since $f > \lambda$ when the singular strategy is evolutionarily attracting, the expression B_2 is certainly positive, when $\tau \geqslant 0$. The expression is increasing with respect to τ. Therefore it is enough to prove that $B_2 > 0$ when $\tau = \tau_{\text{thresh}}$ defined in (D.10). We obtain

$$B_2\Big|_{\tau=\tau_{\text{thresh}}} = \underbrace{\frac{\lambda(1-\pi+f\pi)}{(f-\lambda)(1-\pi)\pi(1-\pi+\lambda\pi)}}_{>0}\hat{D}_0, \tag{D.16}$$

which is positive, when the zero-dispersal strategy is repelling. □

D.6. The qualitative effect of parameters on the singular strategy m^*

Theorem D.4. *Provided that the singular strategy m^* is evolutionarily attracting and $0 < m^* < 1$, the qualitative effect of parameters on the singular strategy is as follows:*

- *The singular strategy decreases with autocorrelation τ (Fig. 2c).*
- *The singular strategy increases, when dispersal survival p is increased (Fig. 4a).*
- *Increasing the relative fecundity f increases $\hat{D}_0$ and thus promotes the emergence of dispersal. The singular strategy m^* increases with f when $\lambda \leqslant 1$ (Fig. 3bc). For $\lambda > 1$ the singular strategy m^* mostly increases when f increases, but m^* decreases when f increases when p is sufficiently close to 1 (Fig. 3a).*
- *When λ is large, both $\hat{D}_0 < 0$ and $D(1) < 0$, which means that dispersal evolves to zero, $m^* = 0$. Numerical explorations illustrate that the singular strategy can be decreasing or non-monotonic with respect to λ (Fig. 4b).*

Proof. In order for dispersal to evolve, we must have $\hat{D}_0 > 0$, for which a necessary condition is $f > \lambda$. Assuming $\hat{D}_0 > 0$ and $Y < 0$ we have the following effects of parameters on the expression $-\frac{Z}{Y}$:

- Autocorrelation τ: When $\tau = 1$, we have $Z = 0$ and $Y < 0$, so that dispersal evolves to zero. It is easy to see that

$$\frac{\partial}{\partial\tau} Z = -(1-\pi+\lambda\pi)\hat{D}_0 < 0 \tag{D.17}$$

and

$$\frac{\partial}{\partial\tau} Y = -f(1-p)(1+-\pi+\pi\lambda)^2 < 0 \tag{D.18}$$

Therefore,

$$\frac{\partial}{\partial\tau}\left(-\frac{Z}{Y}\right) = \frac{\overbrace{(-Y)}^{>0}\overbrace{\frac{\partial}{\partial\tau}Z}^{<0} + \overbrace{Z}^{>0}\overbrace{\frac{\partial}{\partial\tau}Y}^{<0}}{Y^2} < 0, \tag{D.19}$$

which means that when $0 < m^* < 1$, the singular strategy decreases with autocorrelation τ.

- Dispersal survival p:

$$\frac{\partial}{\partial p} Z = (1-\tau)(1-\pi+\lambda\pi)[f(1-\pi+\pi\lambda)+\pi(1-\pi)(f-1)(f-\lambda)] > 0, \tag{D.20}$$

which means that increasing p promotes the emergence of dispersal. Furthermore,

$$\frac{\partial}{\partial p}m^* = \frac{(1-\tau)(1-\pi+\lambda\pi)}{Y^2}\underbrace{\left[-Y\left(\frac{\partial}{\partial p}\hat{D}_0\right)+\hat{D}_0\left(\frac{\partial}{\partial p}Y\right)\right]}_{U+V\tau}, \quad \text{(D.21)}$$

in which

$$V = f(f-\lambda)(f-1)\pi(1-\pi)(1-\pi+\lambda\pi)^2 > 0. \quad \text{(D.22)}$$

The component determining the sign of (D.21) thus decreases when τ is decreased. Substituting the lower bound $\tau = \tau_{\text{thresh}}$ from (D.10) results in

$$U + V\tau_{\text{thresh}} = \lambda p(1-\pi+\pi f)^2\hat{D}_0, \quad \text{(D.23)}$$

which is positive, when the singular strategy is evolutionarily attracting. Therefore, the singular strategy m^* increases with p.

- Productivity ratio f: With the help of $\hat{D}_0 > 0$ we obtain

$$\begin{aligned}\frac{\partial}{\partial f}Z &= (1-\tau)(1-\pi+\lambda\pi)[-(1-p)(1-\pi+\pi\lambda)+p\pi(1-\pi)(2f-1-\lambda)]\\ &> (1-\tau)(1-\pi+\lambda\pi)[-(1-p)(1-\pi+\pi\lambda)\\ &\quad+\frac{f(1-p)(1-\pi+\pi\lambda)}{(f-1)(f-\lambda)}(2f-1-\lambda)]\\ &= (1-\tau)(1-p)(1-\pi+\lambda\pi)^2\frac{f^2-\lambda}{(f-1)(f-\lambda)} \geqslant 0,\end{aligned} \quad \text{(D.24)}$$

which means that increasing f promotes the emergence of dispersal.

$$\frac{\partial}{\partial f}m^* = \frac{(1-\tau)(1-\pi+\lambda\pi)}{Y^2}\underbrace{\left[-Y\left(\frac{\partial}{\partial f}\hat{D}_0\right)+\hat{D}_0\left(\frac{\partial}{\partial f}Y\right)\right]}_{\tilde{U}+\tilde{V}\tau}, \quad \text{(D.25)}$$

in which

$$\tilde{V} = p(1-p)\pi(1-\pi)(f^2-\lambda)(1-\pi+\lambda\pi)^2 > 0 \quad \text{(D.26)}$$

The component determining the sign of (D.25) thus decreases when τ is decreased. Substituting the lower bound $\tau = \tau_{\text{thresh}}$ from (D.10) results in

$$\tilde{U} + \tilde{V}\tau_{\text{thresh}} = \frac{p\hat{D}_0}{(f-1)(f-\lambda)}(\tilde{W}_0 - p\tilde{W}_1), \quad \text{(D.27)}$$

in which $\tilde{W}_0$ does not depend on p and

$$\tilde{W}_1 = (\lambda*(1-\pi+f\pi))((f-1)(\lambda-1)\pi+(f-1)+(f-\lambda)) > 0 \quad \text{(D.28)}$$

Therefore $\tilde{W}_0 - p\tilde{W}_1$ decreases when p is increased. Substituting $p = 1$ results in

$$\tilde{W}_0 - \tilde{W}_1 = (1-\lambda)(f-\lambda)^2\pi(1-\pi) \quad \text{(D.29)}$$

We thus reach the following conclusion: If $\lambda \leqslant 1$ then $\tilde{W}_0 - \tilde{W}_1 \geqslant 0$, so that the singular dispersal strategy increases when f increases. If $\lambda > 1$ then the singular dispersal strategy usually increases when f increases, but can decrease, when $p \approx 1$. In particular, for $p = 1$ we have

$$m^*|_{p=1} = \frac{(f-1)(1-\tau)(1-\pi+\lambda\pi)}{f-\lambda}, \quad \text{(D.30)}$$

so that

$$\frac{\partial}{\partial f}m^*|_{p=1} = \frac{(1-\lambda)(1-\tau)(1-\pi+\lambda\pi)}{(f-\lambda)^2}, \quad \text{(D.31)}$$

which is positive for $\lambda < 1$ and negative for $\lambda > 1$.

In the special case

$$\tilde{W}|_{\lambda=1} = 2(f-1)(1-p)(1-\pi+\pi f) > 0. \quad \text{(D.32)}$$

Therefore, the singular strategy m^* increases with f at least when $\lambda = 1$. □

D.7. Exploring the consequences of evolutionary branching

To confirm our analytic result and study evolutionary dynamics after branching, we performed simulations. An initially unimodal distribution first evolved towards the singular strategy. The predicted branching threshold is $\tau = 0$. At $\tau = 0.4$, the distribution stayed unimodal around the analytically predicted m^* value. At $\tau = -0.4$, a clear evolutionary branching was observed (see Fig. S.2).

D.8. Qualitatively different parameter plots with respect to π and τ

As discussed in Section 3.2, the parameter plot of the singular dispersal strategy m^* with respect to the proportion of high-productivity patches π and autocorrelation τ in Fig. 2a illustrates that spatiotemporal heterogeneity promotes dispersal. However, depending on other parameters, such parameter plots may look qualitatively different. Fig. S.3 shows all qualitatively different types of such parameter plots for $p < 1$. Below we will explain the details of all such cases ranging from a–g. Furthermore, Fig. S.4 illustrates when each of these cases occurs depending on the choice of the other remaining parameters, productivity ratio f, dispersal survival probability p, and arrival bias λ.

(a) Dispersal evolves to zero ($m^* = 0$) for all feasible combinations of π and τ (Fig. S.3a). This occurs at least when the productivity ratio is smaller than the arrival bias, $f < \lambda$ (gray area in Fig. S.4), and the parameter range becomes wider when the dispersal survival probability is decreased.

(b) Bistability: dispersal evolves either to zero or to full dispersal ($m^* = 0$ or $m^* = 1$) in Fig. S.3b. The zero-dispersal strategy $m^* = 0$ is at least locally evolutionarily attracting for all feasible combinations of π and τ, and full-dispersal strategy $m^* = 1$ is locally evolutionarily attracting when τ is small enough, $\tau < \tau_{\text{thresh}}$, in which the analytical expression for τ_{thresh} is given in (D.10). This occurs (purple area in Fig. S.4), when $\lambda > 1$ and f is increased enough from the case a. More precisely, by substituting $\pi = 1/2$ into $D(1)$ given by (5) and (6) and taking the limit $\tau \to -1$, we obtain that $D(1) > 0$ in the bottom corner of Fig. S.3, when

$$f > \frac{1}{2\lambda p}\left(1+2(1-p)\lambda+\lambda^2+(1+\lambda)\sqrt{(1+\lambda)^2-4\lambda p}\right), \quad \text{(D.33)}$$

providing analytically the boundary curve between the gray and purple areas of Fig. S.4, as well as the green and blue areas, which will be explained below.

(c,d) Bistability or intermediate singularities: for intermediate proportions of π, a positive singular dispersal exists for large enough autocorrelation, $\tau > \tau_{\text{thresh}}$, while full dispersal evolves for $\tau < \tau_{\text{thresh}}$. On both sides of this region, a region of bistability exists. For $\pi \approx 0$ and $\pi \approx 1$ only zero-dispersal evolves (Fig. S.3cd), red area in Fig. S.4. Necessarily $\lambda > 1$. Fig. S.3c and d differ in that respect that in Fig. S.3c $\tau_{\text{thresh}} < 0$ for all π, whereas in Fig. S.3d $\tau_{\text{thresh}} > 0$ for some π. The parameter regions in which cases c and d occur are plotted in red in Fig. S.4, with a dashed curve separating the cases c and d. By substituting $\tau = 0$ into $D(1)$ and finding its maximal value with respect to π, we observe that case d occurs when

$$f > \frac{\lambda(x+(-1+\lambda)\lambda(-1+p))(x+(-1+\lambda)(-1+p)p)}{(x+(-1+\lambda)(-1+p))(x+(-1+\lambda)\lambda(-1+p)p)} \text{ in which}$$
$$x = (1-\lambda)(1-p)\sqrt{\lambda p}. \quad \text{(D.34)}$$

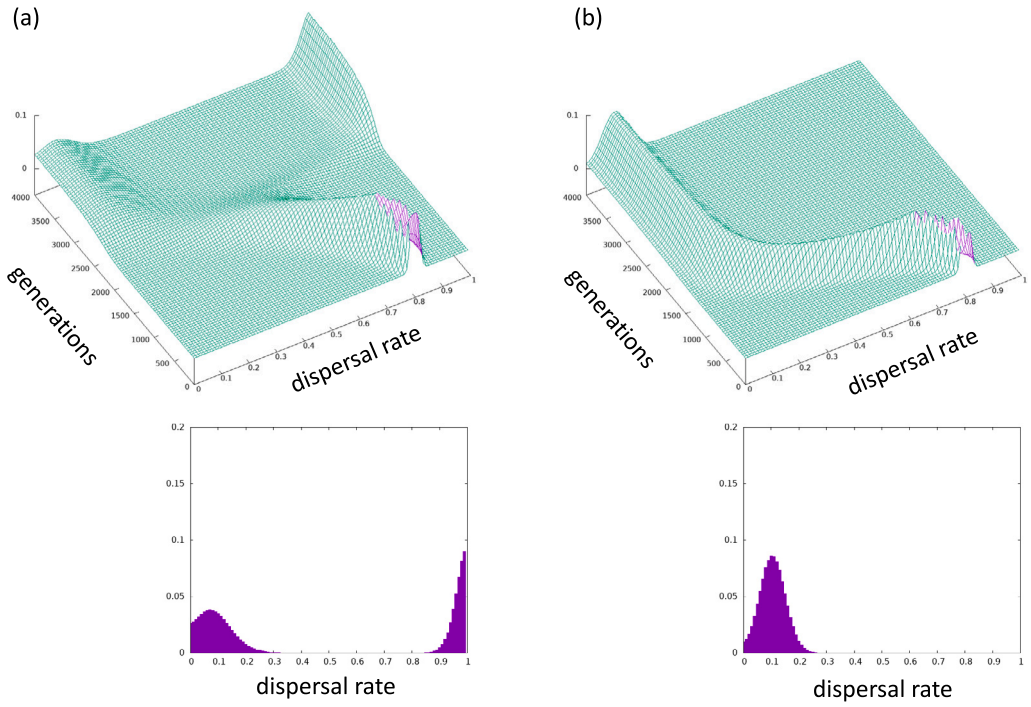

Fig. S.2. Simulation results for 1000 patches. The bottom rows show the distribution at the end of the simulation. There are 100 strategies ($m = 0.01, 0.02, \ldots, 1.00$). The fraction of each strategy at each patch takes a continuous value and changes over time. A fraction $\mu = 0.01$ and another fraction μ mutates to have $+1/100$ and $-1/100$ in their dispersal probability, respectively. $f = 4$, $\pi = 0.5$, and $p = 0.7$. (a) $\tau = -0.4$, (b) $\tau = 0.4$.

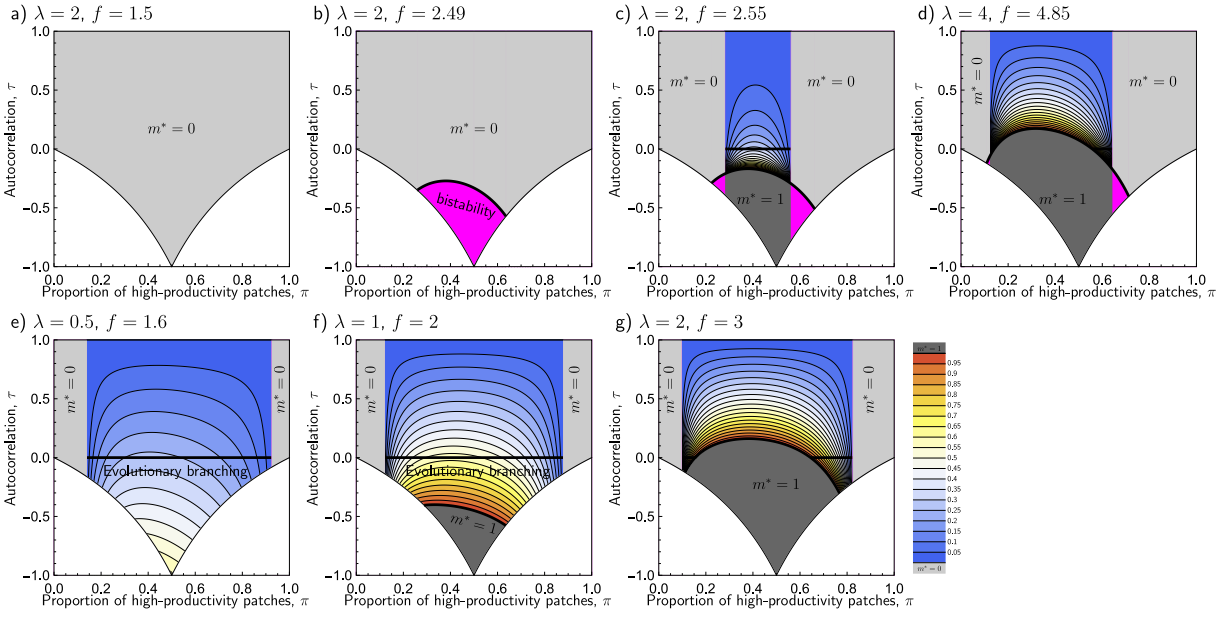

Fig. S.3. Qualitatively different parameter plots with respect to autocorrelation τ and proportion of high-productivity patches π, when $p < 1$. (a) Dispersal evolves to zero ($m^* = 0$) for all combinations of (τ, π), (b) For large τ, dispersal evolves to zero ($m^* = 0$), for τ close to -1 there is bistability: depending on the initial strategy, dispersal evolves either to $m^* = 0$ or $m^* = 1$, (f–g) The zero-dispersal strategy is attracting only when $\pi \approx 0$ or $\pi \approx 1$. For intermediate π, a singular strategy $0 < m^* < 1$ exists, when τ is large enough. For τ close to -1 full dispersal evolves ($m^* = 1$), (e) As in (f), but full dispersal does not evolve for τ close to -1 (c–d) As in (f–g), but with a region of bistability as in (b). In c and f the boundary between $m^* = 1$ is below $\tau = 0$, while in d and g, a part of it is above $\tau = 0$.

(e) Positive dispersal evolves for intermediate π, but full dispersal never evolves ($m^* < 1$), Fig. S.3e. This case occurs, when $\lambda < 1$ and f is increased from the case a, shown as a blue region in Fig. S.4. By solving when the discriminant in (D.6) is positive, we

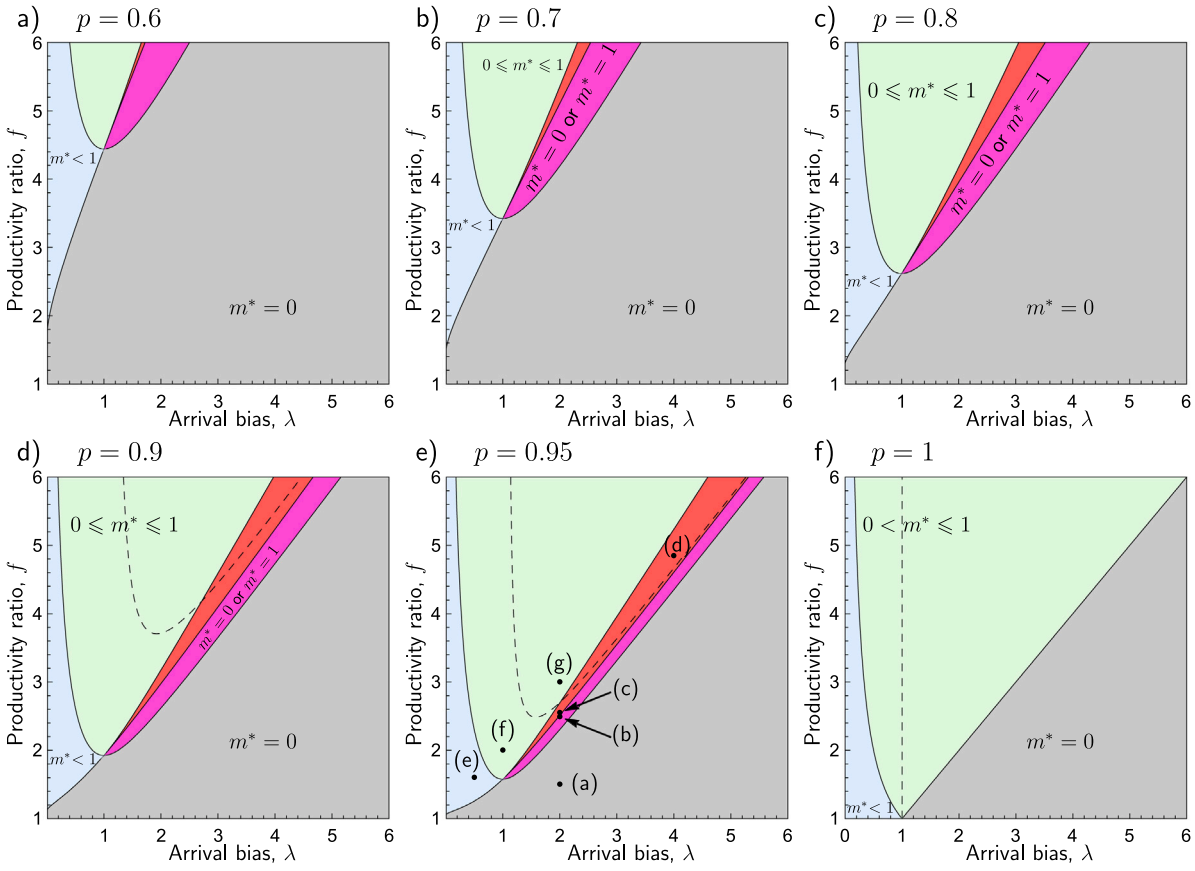

Fig. S.4. Regions of cases (a–g) of Fig. S.3 with respect to λ and f for different values of p. The dots in panel (e) correspond to parameter combinations shown in Fig. S.3. For the meanings of colors, please refer to the text in Appendix D.8.

obtain that positive singular strategies exist for intermediate π, when

$$f > \frac{1}{2p}\left(1+\lambda+2(1-p)\sqrt{\lambda}+\sqrt{1+6\lambda+\lambda^2+4(1-p)(1+\lambda)\sqrt{\lambda}-8\lambda p}\right), \tag{D.35}$$

providing analytically the curve between the gray (a) and blue (e) areas ($\lambda < 1$) and purple (b) and red (c) areas.

(f,g) For intermediate proportions of π, a positive singular dispersal strategy exists for large enough autocorrelation, $\tau > \tau_{\text{thresh}}$, while full dispersal evolves for $\tau < \tau_{\text{thresh}}$. For $\pi \approx 0$ and $\pi \approx 1$ only zero-dispersal evolves, $0 \leqslant m^* \leqslant 1$, Fig. S.3fg. Fig. S.3f and g differ in the same respect as Fig. S.3cd: in Fig. S.3f $\tau_{\text{thresh}} < 0$ for all π, whereas in Fig. S.3g $\tau_{\text{thresh}} > 0$ for some π. The parameter regions in which cases f and g occur are plotted in green in Fig. S.4, with a dashed curve separating the cases f and g. The curve separating the red and green areas is obtained by solving numerically when τ_{thresh} meets the boundary of the feasible parameter region (B.11) at $\pi_{\min}$ or at $\pi_{\max}$ given by (D.6). According to our numerical explorations, these two conditions occur at the same time.

For $p = 1$ the evolutionary scenarios are to some extent different from those presented in Fig. S.3. For $p = 1$, bistability is not possible. Furthermore, either $m^* = 0$ evolves for all feasible combinations of π and τ (Fig. S.5a), or positive dispersal evolves for all feasible combinations of π and τ (Fig. S.5b–e). For $\lambda > 1, \lambda = 1$ and $\lambda < 1$, the threshold τ_{thresh} satisfies $\tau_{\text{thresh}} > 0$ (Fig. S.5b), $\tau_{\text{thresh}} = 0$ (Fig. S.5c), and $\tau_{\text{thresh}} < 0$ (Fig. S.5d), respectively.

Appendix E. Evolution of conditional dispersal

E.1. Fitness gradient

The fitness gradient is the vector of first derivatives of the metapopulation fitness (C.6) with respect to the mutant dispersal strategy components, evaluated when the mutant dispersal strategy is equal to that of the resident.

$$\boldsymbol{D}(\boldsymbol{m}) = \left(\frac{\partial}{\partial m_{\text{mut},1}}R(\boldsymbol{m}_{\text{mut}};\boldsymbol{m}), \frac{\partial}{\partial m_{\text{mut},2}}R(\boldsymbol{m}_{\text{mut}};\boldsymbol{m})\right)\Bigg|_{\boldsymbol{m}_{\text{mut}}=\boldsymbol{m}}. \tag{E.1}$$

This calculation leads to the following theorem:

Theorem E.1. *The fitness gradient for conditional dispersal*

$$\boldsymbol{D}(m_H, m_L) = \frac{1}{A}(\pi f(B_0 - B_H m_H + B_L m_L), (1-\pi)(-C_0 + C_H m_H - C_L m_L)) \tag{E.2}$$

in which

$$\begin{aligned}
A =& p(m_L(1-\pi)+fm_H\pi)[\lambda p(m_L(1-\pi)+fm_H\pi)\\
&+(1-\tau)(1-\pi+\lambda\pi)(\lambda(1-m_L)\pi+f(1-m_H)(1-\pi))] \geqslant 0\\
B_0 =& -(1-\tau)(1-\pi+\pi\lambda)(1+fp(-1+\pi)+(-1+\lambda-\lambda p)\pi)\\
B_H =& fp(1-\pi+\lambda(1-p)\pi) > 0\\
B_L =& \lambda p^2(1-\pi)+(1-p)(1-\tau)(1+(\lambda-1)\pi)^2 > 0\\
C_0 =& (1-\tau)(1-\pi+\pi\lambda)(f(1-\pi)(1-p)+\lambda\pi(f-p)) \geqslant 0\\
C_H =& f(\lambda p^2\pi+(1-p)(1-\tau)(1+(\lambda-1)\pi)^2) > 0\\
C_L =& \lambda p((1-p)(1-\pi)+\lambda\pi) > 0.
\end{aligned} \tag{E.3}$$

Earlier we obtained the fitness gradient for unconditional dispersal (D.2) by differentiating (C.6) when $m_1 = m_2 = m$ and $m_{\text{mut},1} =$

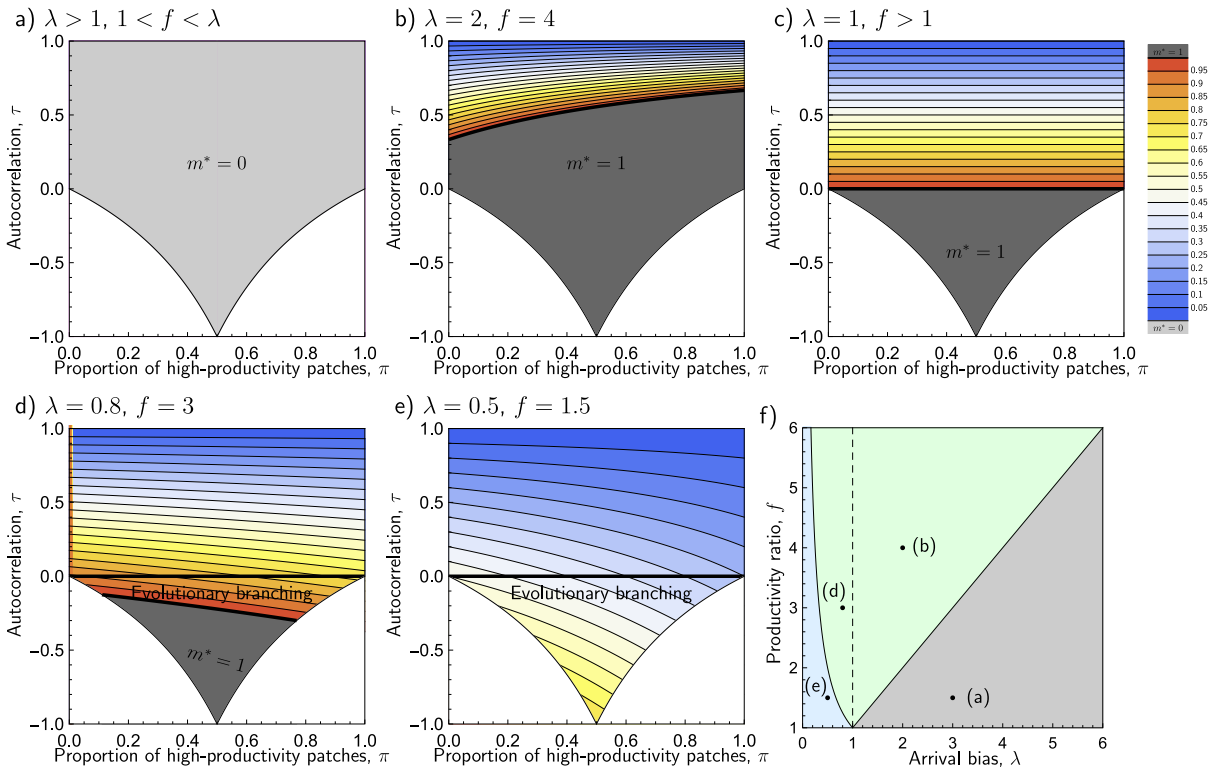

Fig. S.5. Qualitatively different parameter plots with respect to autocorrelation τ and proportion of high-productivity patches π when $p = 1$. (a) Dispersal evolves to zero ($m^* = 0$) for all combinations of (τ, π), (b–d) A singular strategy $0 < m^* < 1$ exists, when τ is large enough. For τ close to -1 full dispersal evolves ($m^* = 1$), (e) As in (d), but full dispersal does not evolve. Panels b–d differ in respect of whether evolutionary branching is possible or not. Panel f is the same as Fig. S.4f, but here the dots correspond to parameter combinations in panels a–e.

$m_{\text{mut},2} = m_{\text{mut}}$. It can also be obtained from the fitness gradient for conditional dispersal (E.3), by adding its components together and setting $m_1 = m_2 = m$. Therefore, the connection between (6) and (E.3) is the following:

$$\begin{aligned} \tilde{A} &= A\big|_{m_{\text{H}}=m_{\text{L}}=m} \\ Z &= \pi f B_0 + (1-\pi)(-C_0) \\ Y &= \pi f (B_{\text{L}} - B_{\text{H}}) + (1-\pi)(C_{\text{H}} - C_{\text{L}}). \end{aligned} \tag{E.4}$$

E.2. Outcomes of conditional dispersal evolution

According to Theorem E.1, the fitness gradient isoclines are straight lines (Eq. (11) in the main text).

$$\begin{aligned} D_{\text{H}}(m_{\text{H}}, m_{\text{L}}) = 0 \text{ for } m_{\text{L}} &= \frac{1}{B_{\text{L}}}(-B_0 + B_{\text{H}} m_{\text{H}}) \\ D_{\text{L}}(m_{\text{H}}, m_{\text{L}}) = 0 \text{ for } m_{\text{L}} &= \frac{1}{C_{\text{L}}}(-C_0 + C_{\text{H}} m_{\text{H}}). \end{aligned} \tag{E.5}$$

Next we show that all qualitatively different types of phase-plane plots of conditional dispersal evolution are as shown in Fig. 5. Consequently, there will be no dispersal from less productive patches, i.e., m_{L} evolves to zero.

Theorem E.2. *If $p = 1$, the fitness gradient isoclines (E.5) are identical, in which case dispersal evolution first converges to the isocline, but is neutral along it. Otherwise dispersal evolution converges to $(m_H^*, 0)$, where*

$$m_H^* = \begin{cases} 0, & \text{if } B_0 \leqslant 0 \\ 1, & \text{if } B_0 \geqslant B_H \\ \frac{B_0}{B_H}, & \text{otherwise.} \end{cases} \tag{E.6}$$

Proof. All qualitatively different types of phase-plane plots are shown in Fig. 5, illustrating that conditional dispersal converges to $(m_{\text{H}}^*, 0)$, where m_{H}^* is given in (E.6). Next we go through all cases.

- When $p = 1$, we obtain from (E.3)

$$\begin{aligned} B_0 &= (1-\tau)(1-\pi+\pi\lambda)(1-\pi)(f-1) \geqslant 0 \\ B_{\text{H}} &= f(1-\pi) > 0 \\ B_{\text{L}} &= \lambda(1-\pi) > 0 \\ C_0 &= (1-\tau)(1-\pi+\pi\lambda)\lambda\pi(f-1) \geqslant 0 \\ C_{\text{H}} &= f\lambda\pi > 0 \\ C_{\text{L}} &= \lambda^2\pi > 0. \end{aligned} \tag{E.7}$$

  In this case, the coefficients of the isoclines are

$$\frac{B_0}{B_{\text{L}}} = \frac{(1-\tau)(1-\pi+\pi\lambda)(f-1)}{\lambda} = \frac{C_0}{C_{\text{L}}} \geqslant 0, \quad \frac{B_{\text{H}}}{B_{\text{L}}} = \frac{f}{\lambda} = \frac{C_{\text{H}}}{C_{\text{L}}} > 0, \tag{E.8}$$

  which means that the isoclines (E.5) are equal. According to (E.2) and because the coefficients (E.7) are non-negative, it is clear that strategies converge towards the isocline. At the isocline, both components of the fitness gradient are zero, so dispersal evolution is neutral along it (Fig. 5bc).
- Consider the remaining case $p < 1$. To complete the proof concerning convergence, we show that the isocline for $D_{\text{H}} = 0$ lies above the isocline for $D_{\text{L}} = 0$ which means that singular strategies in the interior of the strategy space do not exist. Together with the fact that the isocline for $D_{\text{L}} = 0$ lies below the origin, the result then follows (Fig. 5defg).

  Isoclines at $m_{\text{H}} = 0$: The isocline for $D_{\text{H}} = 0$ lies above the isocline for $D_{\text{L}} = 0$ at $m_{\text{H}} = 0$, if $-\frac{B_0}{B_{\text{L}}} > -\frac{C_0}{C_{\text{L}}}$, which is equivalent with $-B_0 C_{\text{L}} + C_0 B_{\text{L}} > 0$. This result holds, because under the

assumptions

$$
\begin{aligned}
&-B_0C_\mathrm{L}+C_0B_\mathrm{L}\\
&=(1-p)(1-\tau)(1-\pi+\lambda\pi)^3\left[\lambda p+(1-\tau)(f(1-p)(1-\pi)+\lambda(f-p)\pi)\right]\\
&>0.
\end{aligned}
\tag{E.9}
$$

Isoclines at $m_\mathrm{H}=1$: The isocline for $D_\mathrm{H}=0$ lies above the isocline for $D_\mathrm{L}=0$ at $m_\mathrm{H}=1$, if $\frac{B_\mathrm{H}-B_0}{B_\mathrm{L}}>\frac{C_\mathrm{H}-C_0}{C_\mathrm{L}}$, which is equivalent with $(B_\mathrm{H}-B_0)C_\mathrm{L}+(C_0-C_\mathrm{H})B_\mathrm{L}>0$. We have

$$
\begin{aligned}
(B_\mathrm{H}-B_0)C_\mathrm{L}+(C_0-C_\mathrm{H})B_\mathrm{L}&=\lambda(1-p)p(1-\pi+\lambda\pi)^2[\overbrace{f p\tau}^{*}\\
&+(1-\tau)(1-\pi)+\lambda(1-\tau)\pi(1+\pi(1-\tau)(f-1))\\
&+(f-1)(1-\pi)\pi(1-\tau)^2].
\end{aligned}
\tag{E.10}
$$

The expression (E.10) is clearly positive, if $0\leqslant\tau<1$ and $p<1$. However, the expression $f p\tau$ (marked with *) is negative, when $\tau<0$. Consider therefore the case $\tau<0$ with $p<1$ in more detail. In such case, we use $\frac{-\tau}{1-\tau}\leqslant\pi\leqslant\frac{1}{1-\tau}$ (recall the stationary fraction of high-productivity patches satisfies $\pi=\frac{\alpha}{\alpha+\beta}=\frac{\alpha}{1-\tau}=\frac{1-\tau-\beta}{1-\tau}$). From $\pi\leqslant\frac{1}{1-\tau}$ we obtain $(1-\tau)(1-\pi)\geqslant-\tau$. Furthermore, $\pi(1-\pi)$ reaches its maximum at $\pi=\frac{1}{2}$, and its minimum at $\pi=\frac{-\tau}{1-\tau}$ or $\pi=\frac{1}{1-\tau}$. Therefore

$$
\pi(1-\pi)\geqslant\min\left\{\frac{1}{1-\tau}\left(1-\frac{1}{1-\tau}\right),\frac{-\tau}{1-\tau}\left(1-\frac{-\tau}{1-\tau}\right)\right\}=\frac{-\tau}{(1-\tau)^2},
\tag{E.11}
$$

so that $(1-\pi)\pi(1-\tau)^2\geqslant-\tau$. As a result, the expression in square brackets satisfies

$$
\begin{aligned}
&[\underbrace{f p\tau}_{\geqslant f\tau}+\underbrace{(1-\tau)(1-\pi)}_{>-\tau}+\underbrace{\lambda(1-\tau)\pi(1+\pi(1-\tau)(f-1))}_{\geqslant 0}\\
&\quad+(f-1)\underbrace{(1-\pi)\pi(1-\tau)^2}_{\geqslant-\tau}]\\
&>f\tau-\tau+0+(f-1)(-\tau)=0
\end{aligned}
\tag{E.12}
$$

We have shown that the isocline for $D_\mathrm{H}=0$ lies above the isocline for $D_\mathrm{L}=0$ both at $m_\mathrm{H}=0$ and at $m_\mathrm{H}=1$. Since the isoclines are straight lines, the isocline for $D_\mathrm{H}=0$ lies above the isocline for $D_\mathrm{L}=0$ for $0\leqslant m_\mathrm{H}\leqslant 1$, which completes the proof concerning convergence. □

When $\tau\to 1$, according to (E.3) we have $B_0\to 0$ and $C_0\to 0$, which means that both isoclines go through the origin. The slope of the isocline D_H is greater than the slope of isocline D_L, when $B_\mathrm{H}C_\mathrm{L}-B_\mathrm{L}C_\mathrm{H}>0$. As $\tau\to 1$, we have

$$
B_\mathrm{H}C_\mathrm{L}-B_\mathrm{L}C_\mathrm{H}\to f\lambda(1-p)p^2(1-\pi+\lambda\pi)^2\geqslant 0.
\tag{E.13}
$$

Equality in (E.13) holds only in the special case $p=1$ investigated above (Fig. 5b). When $p<1$, (E.13) holds with strict inequality. The isoclines cross only at the origin, and conditional dispersal evolves to zero (Fig. 5a), so that also (E.6) holds in this case.

E.3. No evolutionary branching for conditional dispersal

Finally, we investigate uninvadability (evolutionary stability) of the strategy $(m_\mathrm{H}^*,0)$. When $0<m_\mathrm{H}^*<1$, we could in principle observe disruptive selection in the direction of m_H. However, when the resident strategy is $(m_\mathrm{H}^*,0)$, the metapopulation fitness of the mutant is equal to 1 for all mutant strategies of form $(m_{\mathrm{H,mut}},0)$. In other words,

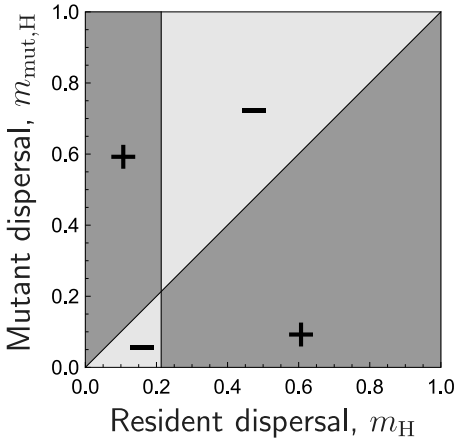

Fig. S.6. Pairwise invasibility plot with respect to the strategy component m_H of conditional dispersal. Parameters are as in Fig. 5g, i.e., $p=0.75$, $f=3$, $\lambda=1$, $\tau=0.4$, $\pi=0.5$.

$$
\begin{aligned}
R((m_{\mathrm{mut,H}},m_{\mathrm{mut,L}});(m_\mathrm{H}^*,0))<1 &\quad\text{for } m_{\mathrm{mut,L}}>0 \text{ and for any } m_{\mathrm{mut,H}}\\
R((m_{\mathrm{mut,H}},0);(m_\mathrm{H}^*,0))=1 &\quad\text{for any } m_{\mathrm{mut,H}}.
\end{aligned}
\tag{E.14}
$$

The neutral contour line of a pairwise invasibility plot (PIP) is thus a vertical line. This is the boundary case between strict ESS and branching, meaning that evolutionary branching does not happen (see Fig. S.6).

E.4. The qualitative effect of parameters on the singular strategy $(m_H^*,0)$

Theorem E.3. *The qualitative effect of parameters on the strategy* $(m_H^*,0)$ *is as follows*

- *Dispersal survival probability* p*:* m_H^* *is a non-decreasing function of* p *(Fig. 7a).*
- *Productivity ratio* f*:* $m_H^*=0$ *for* $1\leqslant f\leqslant\tilde f$*, in which* $\tilde f$ *is given in* (E.18)*. For* $f>\tilde f$ *the strategy* m_H^* *increases with* f *(Fig. 2h).*
- *Relative attractivity* λ*: If* $\frac{1}{p}<f<\frac{1}{p^2}$*,* m_H^* *decreases with respect to* λ*. If* $f>\frac{1}{p^2}$*,* m_H^* *is non-monotonic with respect to* λ *(Fig. 7b)*
- *Autocorrelation* τ*: For* $\tau=1$ *we have* $m_H^*=0$*, and* m_H^* *is linear with respect to* τ *around* $m_H^*=0$*. For small enough* τ *we may have* $m_H^*=1$ *(Fig. 2g).*
- *Proportion of high-productivity patches* π*: If* $\frac{1}{p}<f<\frac{1}{p^2}$ *or* $f>\frac{1}{p^2}$ *and* $\lambda<\tilde\lambda$*, where* $\tilde\lambda=\frac{fp-1}{fp^2-1}>1$*,* m_H^* *is a non-increasing function of* π *(Fig. 6ab). For* $\tau\geqslant 0$*, the singular strategy* m_H^* *is a non-monotonic function of* π*, if*

$$
f>\frac{1}{p^2}\ \text{and}\ \lambda>\tilde\lambda=\frac{fp-1}{fp^2-1},
\tag{E.15a}
$$

which is equivalent with the condition (Fig. 6c).

$$
\lambda>\frac{1}{p}\ \text{and}\ f>\frac{\lambda-1}{p(\lambda p-1)}=:\hat f.
\tag{E.15b}
$$

Furthermore, $m_H^*=0$ *for* $\pi\geqslant\tilde\pi$*, where* $\tilde\pi=\frac{fp-1}{fp-1+\lambda(1-p)}<1$*. For* $\tau<0$*, the condition* (E.15) *is only a necessary condition for the singular strategy* m_H^* *to be a non-monotonic function of* π*.*

Proof.

- The effect of dispersal survival probability p: By differentiation we obtain

$$
\frac{\partial}{\partial p}\frac{B_0}{B_\mathrm{H}}=\frac{(1-\tau)(1-\pi+\lambda\pi)}{fp^2(1-\pi+\lambda(1-p)\pi)^2}\left[(1-\pi(1+\lambda(1-p)))^2\right.
$$

$$+\lambda\pi(1-\pi)(4(1-p)+fp^2)\big] \geqslant 0. \tag{E.16}$$

Therefore, m_{H}^* is a non-decreasing function of p (Fig. 7a).

- The effect of productivity ratio f:
  The coefficient $B_0 \leqslant 0$ for $1 \leqslant f \leqslant \tilde{f}$ so that $m_{\mathrm{H}}^* = 0$. This can be seen from

$$B_0\Big|_{f=1} = -(1-p)(1-\tau)(1-\pi+\lambda\pi)^2 \leqslant 0 \tag{E.17}$$

and by solving $B_0 = 0$ for f we obtain

$$\tilde{f} = \begin{cases} \frac{1-\pi+\lambda\pi(1-p)}{p(1-\pi)} \geqslant \frac{1}{p}, & \tau < 1 \\ \infty, & \tau = 1 \end{cases} \tag{E.18}$$

By differentiation we obtain

$$\frac{\partial}{\partial f}\frac{B_0}{B_{\mathrm{H}}} = \frac{(1-\tau)(1-\pi+\lambda\pi)}{(f^2 p)} \geqslant 0. \tag{E.19}$$

For $f > \tilde{f}$ the strategy m_{H}^* thus increases with f (Fig. 2h).

- Arrival bias λ: By differentiation, we obtain

$$\frac{\partial}{\partial\lambda}\frac{B_0}{B_{\mathrm{H}}} = \frac{\pi(1-\tau)}{fp(1-\pi+\lambda(1-p)\pi)^2}\underbrace{\left[fp^2(1-\pi)^2-(1-\pi+\lambda(1-p)\pi)^2\right]}_{=Z(\lambda)} \tag{E.20}$$

The sign of the derivative (E.20) is determined by the term $Z(\lambda)$ in square brackets. It is a second-order polynomial with respect to λ, and the sign of the second-order term is negative. Therefore, $Z(\lambda)$ is negative for large λ. Furthermore,

$$Z'(0) = -2(1-p)\pi(1-\pi) \leqslant 0. \tag{E.21}$$

We have thus two potential cases: If $Z(0) < 0$, $Z(\lambda)$ is negative for all $\lambda \geqslant 0$. If $Z(0) > 0$, $Z(\lambda)$ is positive for small λ, and negative for large λ. By calculating

$$Z(0) = (fp^2-1)(1-\pi)^2 \tag{E.22}$$

we observe that for $f > \frac{1}{p^2}$, we have $Z(0) > 0$, in which case m_{H}^* is non-monotonic with respect to λ. If $f < \frac{1}{p^2}$ we have $Z(0) < 0$. However, f needs to be large enough to dispersal evolve in the first place, $f > \tilde{f}$. In the limit $\lambda \to 0$ the condition becomes $f > \frac{1}{p}$. To conclude, if $\frac{1}{p} < f < \frac{1}{p^2}$, m_{H}^* decreases with λ.

- Autocorrelation τ: The denominator B_{H} does not depend on τ, and the numerator B_0 includes τ only in the form of multiplier $(1-\tau)$, from which the result follows: For $\tau = 1$ we have $m_{\mathrm{H}}^* = 0$, and m_{H}^* is linear with respect to τ around $m_{\mathrm{H}}^* = 0$. For small enough τ we may have $m_{\mathrm{H}}^* = 1$ (Fig. 2g).
- Proportion of high-productivity patches π: By differentiation, we obtain

$$\frac{\partial}{\partial\pi}\frac{B_0}{B_{\mathrm{H}}} = \frac{(1-\tau)}{fp(1-\pi+\lambda(1-p)\pi)^2}H(\pi), \tag{E.23}$$

in which

$$\begin{aligned} H(\pi) = &-(-1+fp)(-1+\pi)^2 - \lambda^3(-1+p)^2\pi^2 \\ &+ \lambda^2(-1+p)\pi(2+(-3+p+fp)\pi) \\ &- \lambda(-1+\pi)(-1+(3-2p)\pi+fp(p-2\pi+p\pi)) \end{aligned} \tag{E.24}$$

The sign of the derivative (E.23) is determined by the term $H(\pi)$ in (E.24). Based on above, it is enough to consider $f > \frac{1}{p}$. We obtain

$$\begin{aligned} H(0) &= 1-fp+\lambda(-1+fp^2) \\ H(1) &= \lambda^2(1-p)[(1-p)(1-\lambda)-fp] \leqslant 0 \text{ for } f > \frac{1}{p} \\ H'(\pi) &= 2(1-\lambda)\underbrace{(\lambda(1-p)+fp-1)}_{>0 \text{ for } f>\frac{1}{p}}\underbrace{(1-\pi+\lambda(1-p)\pi)}_{\geqslant 0} \end{aligned} \tag{E.25}$$

For $f > \frac{1}{p}$ we have $H(1) \leqslant 0$, because $(1-p)(1-\lambda)-fp \leqslant (1-p)-1 = -p \leqslant 0$. According to (E.25), the derivative $H'(\pi)$ does not change sign in the interval $0 \leqslant \pi \leqslant 1$. Therefore, we get the following classification (Fig. 6):

- If $\frac{1}{p} < f < \frac{1}{p^2}$, we have $H(0) < 0$. Since the derivative does not change sign and $H(1) \leqslant 0$, we have $H(\pi) \leqslant 0$ for $0 \leqslant \pi \leqslant 1$. The singular strategy m_{H}^* is a non-increasing function of π.
- If $f > \frac{1}{p^2}$ and $\lambda < \tilde{\lambda}$, where $\tilde{\lambda} = \frac{fp-1}{fp^2-1} > 1$, we have $H(0) < 0$. Again, the singular strategy m_{H}^* is a non-increasing function of π.
- If $f > \frac{1}{p^2}$ and $\lambda > \tilde{\lambda}$, we have $H(0) > 0$. Furthermore, the derivative $H'(\pi) \leqslant 0$. Therefore, $H(\pi)$ changes sign only once, and is positive for small π and negative for large π. When $\tau \geqslant 0$, valid parameter range is $0 \leqslant \pi \leqslant 1$, and the singular strategy m_{H}^* is a non-monotonic function of π. When $\tau < 0$, the valid parameter range is $\frac{-\tau}{1-\tau} \leqslant \pi \leqslant \frac{1}{1-\tau}$, and $H(0) > 0$ is a necessary, but not a sufficient condition for m_{H}^* to be non-monotonic with respect to π.

Finally, by investigating B_0 we obtain $m_{\mathrm{H}}^* = 0$ for $\pi \geqslant \tilde{\pi}$. □

E.5. Qualitatively different parameter plots with respect to π and τ

Analogous to the evolution of unconditional dispersal, the parameter plots of the singular dispersal strategy component m_{H}^* with respect to the proportion of high-productivity patches π and autocorrelation τ (as in Fig. 2e) may have qualitatively different forms depending on the other parameters. Fig. S.7 shows all qualitatively different types of such parameter plots. Below we will explain the details of all such cases ranging from a–f. Furthermore, Fig. S.8 illustrates when each of these cases occurs depending on the choice of the other remaining parameters, productivity ratio f, dispersal survival probability p, and arrival bias λ.

(a) Dispersal evolves to zero ($m_{\mathrm{H}}^* = 0$) for all feasible combinations of π and τ (Fig. S.7a). Based on results presented in Section 4.3, this scenario occurs when $f < 1/p$ (gray area in Fig. S.8). Again, this parameter range becomes wider when the dispersal survival probability is decreased, but it is considerably smaller than for unconditional dispersal (Fig. S.4).

(b) Positive dispersal can evolve, $0 \leqslant m_{\mathrm{H}}^* < 1$ and m_{H}^* is a non-monotonic function of π (Fig. S.7b): Dispersal evolves to zero, $m_{\mathrm{H}}^* = 0$ for $\tilde{\pi} \leqslant \pi \leqslant 1$, in which $\tilde{\pi}$ is given by (15). Positive dispersal evolves for $0 < \pi < \tilde{\pi}$, but full dispersal does not evolve. This scenario occurs, when $f > 1/p$, but f is not large enough for scenarios d-f to occur, i.e., (E.26) and (E.27) do not hold, and (16) holds, so that the contours of m_{H}^* in Fig. S.7b are non-monotonic at least for $\tau > 0$ (light blue area in Fig. S.8).

(c) Positive dispersal can evolve, $0 \leqslant m_{\mathrm{H}}^* < 1$ and m_{H}^* is decreasing with respect to π (Fig. S.7c). The case (c) is otherwise the same as (b), but (16) does not hold, so that the contours of m_{H}^* in Fig. S.7b are decreasing with respect to π (dark blue area in Fig. S.8).

(d,e) Positive dispersal and full dispersal can evolve, $0 \leqslant m_{\mathrm{H}}^* \leqslant 1$, and m_{H}^* is a non-monotonic function of π at least for $\tau > 0$ (Fig. S.7de): The cases (d) and (e) are similar to case (b), but also full dispersal can evolve for $\pi < \tilde{\pi}$. The cases (d) and (e) differ in the positioning of the parameter combinations of (π, τ) for which full dispersal ($m_{\mathrm{H}}^* = 1$) evolves (dark gray region in Fig. S.7). In case (d), this region is connected only to the bottom left boundary of the feasible parameter combinations ($\tau = -\pi/(1-\pi)$ for $\pi < 1/2$), whereas in case (d) it is connected to both bottom boundaries.
By substituting $\pi = 1/2$ into $D_{\mathrm{H}}(1,0)$ given by (9) and taking the limit $\tau \to -1$, we obtain that $D_{\mathrm{H}}(1,0) > 0$ in the bottom corner of

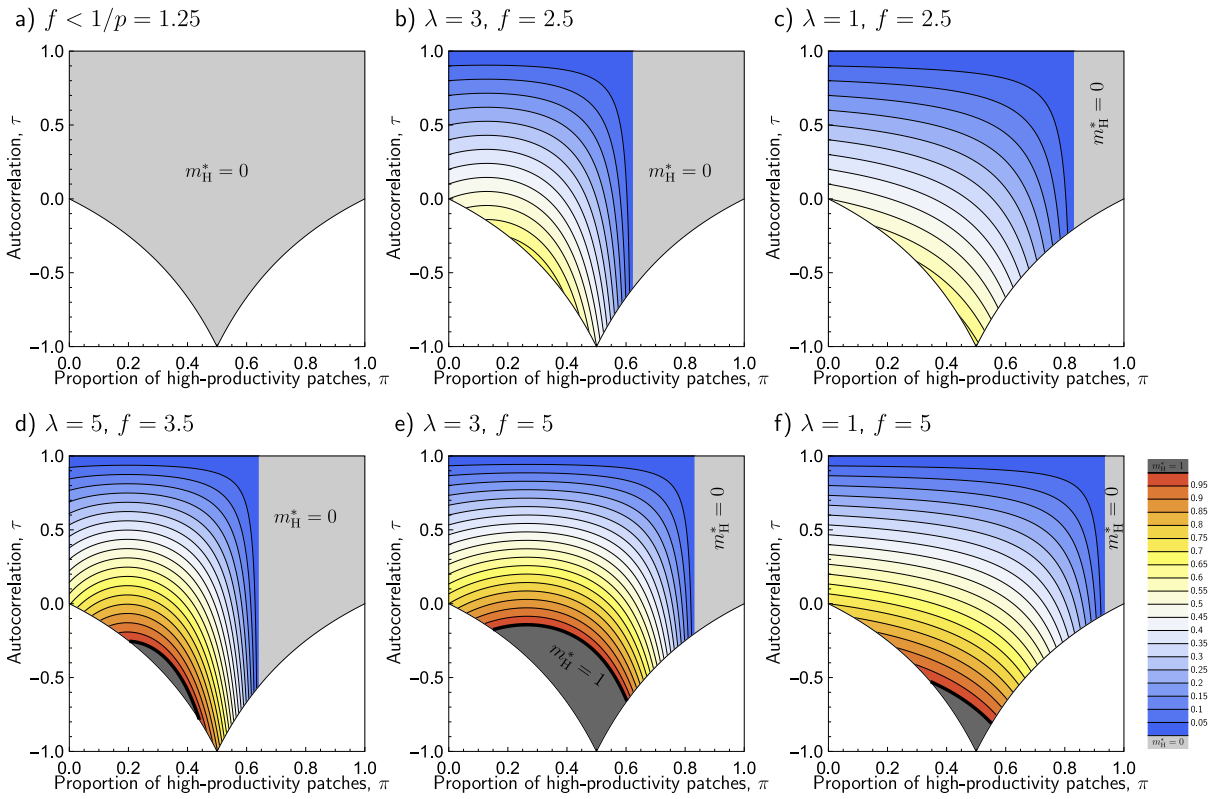

Fig. S.7. Qualitatively different parameter plots with respect to autocorrelation τ and proportion of high-productivity patches π, when $p < 1$. (a) Dispersal evolves to zero ($m_H^* = 0$) for all combinations of (τ, π) (b–f) Positive dispersal evolves for $0 < \pi < \tilde{\pi}$. In b, d and e, m_H^* is non-monotonic with respect to π (at least for $\tau > 0$), whereas in c and f, m_H^* is decreasing with respect to π. In b and c full dispersal does not evolve, whereas it evolves in d–f. Panel d differs from e and f in the positioning of the dark gray region corresponding to the parameter combinations of (π, τ) for which full dispersal evolves. Parameters: $p = 0.8$.

Fig. S.7, when

$$f > \frac{(1+\lambda)(1+\lambda(1-p))}{\lambda p^2}. \tag{E.26}$$

Case (e) thus occurs when (16) and (E.26) hold (light green area above the dashed curve in Fig. S.8).
Case (d) occurs when (E.26) does not hold, but (E.27) given below holds (light green area below the dashed curve in Fig. S.8). These conditions together guarantee that (16) holds.

$$\lambda > \frac{1}{\sqrt{1-p}} \text{ and } f > \frac{(1+\sqrt{1-p})^2}{p^2}. \tag{E.27}$$

(f) Positive dispersal and full dispersal can evolve, $0 \leqslant m_H^* \leqslant 1$, and m_H^* is a decreasing function of π (Fig. S.7f). As in case (e) (E.26) holds, but in contrast (16) does not hold. (dark green area in Fig. S.8).

Appendix F. Analysis of conditional dispersal based on reproductive values

F.1. Overview

Here we provide an argument based on reproductive values and show how the direction of selection can be predicted. We will describe the case for conditional dispersal below.

In contrast to metapopulation fitness approach, where we consider the total "reproduction" by a single patch and where we count the cumulative "reproductive success" of the patch from present to future, arguments based on reproductive values consider only immediate reproduction by a single individual, but each offspring produced shall be counted with "appropriate" weights and those weights reflect different ability of those different offsprings to contribute to a future gene pool.

There is a freedom of choice in how to normalize reproductive values, but we herein assume that at any moment of the time the total reproductive value of all individuals present in the population is equal to the same constant.

Mathematically rigid proofs of how to proceed computation based reproductive values are not discussed here because the current paper is not the place where we develop those arguments from scratch, so interested readers should refer to existing literature (e.g. Taylor, 1990; Taylor and Frank, 1996; Lehmann et al., 2016; Ohtsuki et al., 2020; Avila and Mullon, 2023). Instead, here we demonstrate how to apply the methodology developed by those frameworks. A basic principle of reproductive values normalized as above in a stationary (that is, neither growing or shrinking) population is that one's reproductive value carries over to the aggregate of its offspring. For example, if the reproductive value of a single parent is $v = 0.06$, and if he dies but leaves two offspring who have the same quality as each other, the reproductive value of each of those two offspring is $v' = 0.03$, because v should be equal to $2v'$. For another example, imagine a parent with reproductive value of $v = 0.06$, and suppose that he dies and leaves exactly one offspring and that this offspring will either survive or die with equal probabilities. Then, the offspring immediately after birth inherits its parent's reproductive value, which is $v' = 0.06$. In contrast, if this offspring avoids mortality, the reproductive value of the offspring is raised to $v'' = 0.12$, because one can alternatively interpret this mortality risk as that the offspring immediately after birth will eventually leave on average $1/2$ many dying individual whose reproductive value is zero, and leaves on average $1/2$ surviving individual whose reproductive value is v'', and therefore from the preservation law of reproductive values the relation $v' = (1/2) \cdot 0 + (1/2) \cdot v''$ should hold, leading to $v'' = 0.12$.

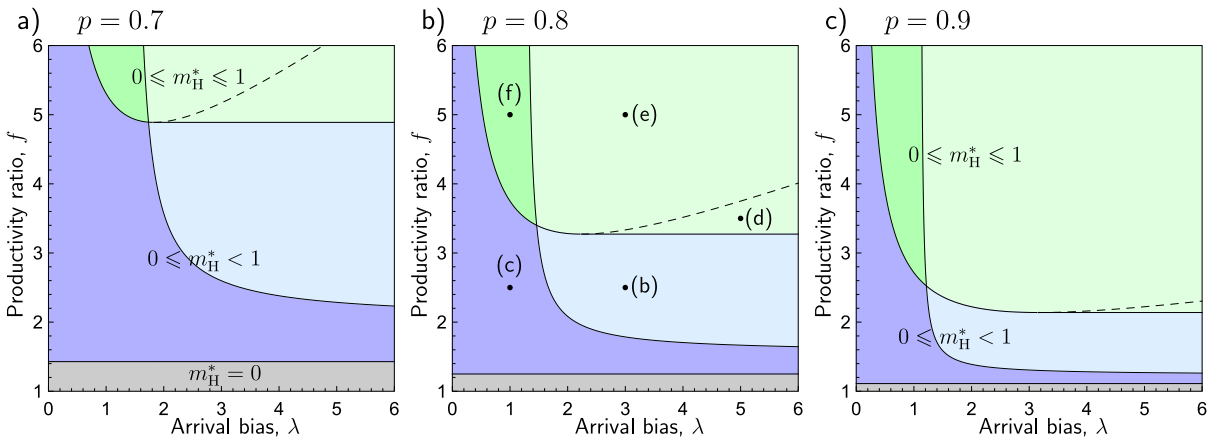

Fig. S.8. Regions of cases (a–g) of Fig. S.7 with respect to λ and f for different values of p. The dots in panel (b) correspond to parameter combinations shown in Fig. S.7. For the meanings of colors, please refer to the text in Appendix E.5.

F.2. Recursions on reproductive values

Now, suppose that everyone in the population adopts resident dispersal strategy $\boldsymbol{m} = (m_H, m_L)$.

Recall the life-cycle assumption described in Section 2.1 in the main text and imagine the population immediately after step "2. Reproduction" but before step "3. Emigration". There are only two types of juveniles in the population, and they are those juveniles who are born in a currently high-productivity patch (which we call "H-born-juveniles") and those juveniles who are born in a currently low-productivity patch (which we call "L-born-juveniles"). We write the reproductive value of a single H-born-juvenile as v_{HB} and that of a single L-born-juvenile as v_{LB}.

Next, imagine the moment in the middle of step "3. Emigration" in Section 2.1 when the decision of whether each juvenile remains in the natal patch or disperses to a random patch is already made but the mortality in dispersal has not produced any victims yet. There are four different juveniles at that moment: those juveniles who are in a currently high-productivity patch and will not disperse (which we call "H-juveniles"), those juveniles who are in a currently low-productivity patch and will not disperse (which we call "L-juveniles"), those juveniles who are in a currently high-productivity patch and will disperse (which we call "H-dispersing-juveniles"), and those juveniles who are in a currently low-productivity patch and will disperse (which we call "L-dispersing-juveniles"). However, an H- and L-dispersing-juvenile leave the same number of progeny in the future because survival probability p is independent of juvenile's place of origin, and hence we call both of them simply "dispersing juveniles". Let v_H, v_L, v_D be reproductive values of a single H-juvenile, of a single L-juvenile, and of a single dispersing juvenile, respectively. Since the fraction $1 - m_H$ of H-born-juveniles become H-juveniles and the fraction m_H of them become dispersing juveniles, and since the fraction $1 - m_L$ of L-born-juveniles become L-juveniles and the fraction m_L of them become dispersing juveniles, from the preservation law of reproductive values it follows that

$$\begin{aligned} v_{HB} &= (1 - m_H)v_H + m_H v_D \\ v_{LB} &= (1 - m_L)v_L + m_L v_D. \end{aligned} \tag{F.1}$$

Next, consider the moment immediately after step "3. Emigration" but before step "4. Immigration" in Section 2.1. There are those juvenile who survived mortality in dispersal, and we call them "surviving dispersing juveniles", the reproductive value of each of which is denoted by v_S. From the preservation law of reproductive values, it follows that

$$v_D = p v_S. \tag{F.2}$$

Next, consider the moment immediately after step "4. Immigration" but before step "5. Transition" in Section 2.1. There are only two types of those juvenile. They are those juveniles who are in a currently high-productivity patch (which we again call "H-juveniles") and those juveniles who are in a currently low-productivity patch (which we again call "L-juveniles"). When we consider a "surviving dispersing juvenile", the chance that he arrives at a currently high-productivity patch to become an H-juvenile is $\lambda_H\pi/(\lambda_H\pi + \lambda_L(1-\pi)) = \lambda\pi/(\lambda\pi + (1-\pi))$, and the chance that he arrives at a currently low-productivity patch to become an L-juvenile is $(1-\pi)/(\lambda\pi+(1-\pi))$, so from the preservation law of reproductive values we have

$$v_S = \frac{\lambda\pi}{\lambda\pi + (1-\pi)} v_H + \frac{1-\pi}{\lambda\pi + (1-\pi)} v_L. \tag{F.3}$$

We next consider the moment immediately after step "5. Transition" but before step "6. Competition" in Section 2.1. After a transition of patch productivity, there are four types of juveniles in the population. They are, those juveniles who are in a patch whose productivity was high before transition and is high after transition (which we call "HH-juveniles"), those juveniles who are in a patch whose productivity was high before transition and is low after transition (which we call "HL-juveniles"), those juveniles who are in a patch whose productivity was low before transition and is high after transition (which we call "LH-juveniles"), and those juveniles who are in a patch whose productivity was low before transition and is low after transition (which we call "LL-juveniles"). Reproductive values of each of those juveniles, denoted by v_{HH}, v_{HL}, v_{LH} and v_{LL}, satisfy

$$\begin{aligned} v_H &= (1-\beta)v_{HH} + \beta v_{HL} \\ &= (1 - (1-\pi)(1-\tau))v_{HH} + (1-\pi)(1-\tau)v_{HL} \\ v_L &= \alpha v_{LH} + (1-\alpha)v_{LL} \\ &= \pi(1-\tau)v_{LH} + (1 - \pi(1-\tau))v_{LL}. \end{aligned} \tag{F.4}$$

Finally, we consider the moment immediately after step "2. Reproduction" but before step "3. Emigration" once again and consider the relationship between v_{HH}, v_{HL}, v_{LH}, v_{LL} and v_{HB}, v_{LB} to complete a series of recursions on reproductive values. For that purpose we need to count how many H-born-juveniles a single HH/LH-juvenile bears on average, and count how many L-born-juveniles a single HL/LL-juvenile bears on average. Those quantities have already essentially been calculated in Appendix C, and by using the preservation law of reproductive values

we have

$$\begin{aligned} v_{\mathrm{HH}} &= \frac{f}{(1-m_{\mathrm{H}})f + \frac{\lambda}{\lambda\pi+(1-\pi)}p(\pi m_{\mathrm{H}} f + (1-\pi)m_{\mathrm{L}})} v_{\mathrm{HB}} \\ v_{\mathrm{HL}} &= \frac{1}{(1-m_{\mathrm{H}})f + \frac{\lambda}{\lambda\pi+(1-\pi)}p(\pi m_{\mathrm{H}} f + (1-\pi)m_{\mathrm{L}})} v_{\mathrm{LB}} \\ v_{\mathrm{LH}} &= \frac{f}{(1-m_{\mathrm{L}}) + \frac{1}{\lambda\pi+(1-\pi)}p(\pi m_{\mathrm{H}} f + (1-\pi)m_{\mathrm{L}})} v_{\mathrm{HB}} \\ v_{\mathrm{LL}} &= \frac{1}{(1-m_{\mathrm{L}}) + \frac{1}{\lambda\pi+(1-\pi)}p(\pi m_{\mathrm{H}} f + (1-\pi)m_{\mathrm{L}})} v_{\mathrm{LB}}. \end{aligned} \tag{F.5}$$

F.3. Direction of selection

A set of recursions, Eqs. ((F.1), (F.2), (F.3), (F.4), (F.5)), specify the relations that the ten different reproductive values, $v_{\mathrm{HB}}, v_{\mathrm{LB}}, v_{\mathrm{H}}, v_{\mathrm{L}}, v_{\mathrm{D}}, v_{\mathrm{S}}, v_{\mathrm{HH}}, v_{\mathrm{HL}}, v_{\mathrm{LH}}$, and v_{LL}, must satisfy. It is easy to see that if we find a solution, then C times the solution, where C is some constant, is also a solution. To determine values of v's, therefore, we must resort to the normalization condition that we have already mentioned above. However, even without such a normalization, we can obtain the ratio of any two of the reproductive values, such as $v_{\mathrm{HB}}/v_{\mathrm{LB}}$ or $v_{\mathrm{H}}/v_{\mathrm{L}}$, and they are functions of m_{H} and m_{L}.

Given these solutions, we consider whether a mutant in dispersal strategy can invade the resident or not. Below we suppose that mutant and resident strategies are close. Regarding dispersal from a high-productivity patch, if an H-born-juvenile does not disperse then this juvenile obtains the reproductive value of v_{H}, and if an H-born-juvenile disperses then this juvenile obtains the reproductive value of v_{D}, so if those two reproductive values are not equal, a mutant strategy that increases the proportion of whichever juvenile that has the larger reproductive value than the other is more adaptive than the resident strategy. In other words, if $v_{\mathrm{H}}/v_{\mathrm{D}} > 1$, then a mutant with $m_{\mathrm{mut,H}}$ being greater than m_{H} is advantageous, and vice versa. By a similar reasoning, if $v_{\mathrm{L}}/v_{\mathrm{D}} > 1$, then a mutant with $m_{\mathrm{mut,L}}$ being greater than m_{L} is advantageous, and vice versa. In fact, we can confirm that the direction of selection derived in this manner matches that in the previous section. In particular, the argument here can precisely predict the position of two isoclines derived by the metapopulation-based argument, which is given in Eq. (E.5).

F.4. When $(m_H, m_L) \to (0, 0)$

For a special case of $(m_{\mathrm{H}}, m_{\mathrm{L}}) \to (0, 0)$, it is shown that

$$\frac{v_{\mathrm{L}}}{v_{\mathrm{H}}} = f \tag{F.6}$$

$$\frac{v_{\mathrm{D}}}{v_{\mathrm{H}}} = p \underbrace{\left(\frac{\lambda\pi}{\lambda\pi+(1-\pi)} 1 + \frac{1-\pi}{\lambda\pi+(1-\pi)} f \right)}_{\text{(weighted average of 1 and } f)} \tag{F.7}$$

holds, suggesting that the relative reproductive value of an L-juvenile to that of an H-juvenile is given by (F.6) and that the relative reproductive value of a dispersing juvenile to that of an H-juvenile is given by (F.7), which leads to the argument in Section 5.2 in the main text.

Appendix G. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jtbi.2023.111612.

References

Avila, P., Mullon, C., 2023. Evolutionary game theory and the adaptive dynamics approach: adaptation where individuals interact. Philos. Trans. R. Soc. B 378 (1876), 20210502.

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

Bengtsson, B., 1978. Avoiding inbreeding: at what cost? J. Theoret. Biol. 73, 439–444.

Blanquart, F., Gandon, S., 2014. On the evolution of migration in heterogeneous environments. Evolution 68 (6), 1617–1628.

Bowler, D.E., Benton, T.G., 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biol. Rev. 80 (2), 205–225.

Clobert, J., Danchin, E., Dhondt, A.A., Nichols, J.D. (Eds.), 2001. Dispersal. Oxford University Press.

Cohen, D., Levin, S.A., 1991. Dispersal in patchy environments - the effects of temporal and spatial structure. Theor. Popul. Biol. 39, 63–99.

Comins, H.N., Hamilton, W.D., May, R.M., 1980. Evolutionarily stable dispersal strategies. J. Theoret. Biol. 82, 205–230.

Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory. Burgess Pub. Co..

Dieckmann, U., O'Hara, B., Weisser, W., 1999. The evolutionary ecology of dispersal. Trends Ecol. Evol. 14, 88–90.

Doebeli, M., Ruxton, G.D., 1997. Evolution of dispersal rates in metapopulation models: branching and cyclic dynamics in phenotype space. Evolution 51, 1730–1741.

Frank, S.A., 1986. Dispersal polymorphisms in subdivided populations. J. Theoret. Biol. 122, 303–309.

Gadgil, M., 1971. Dispersal: population consequences and evolution. Ecology 52 (2), 253–261.

Gandon, S., 1999. Kin competition, the cost of inbreeding and the evolution of dispersal. J. Theoret. Biol. 200, 245–364.

Gandon, S., Michalakis, Y., 1999. Evolutionarily stable dispersal rate in a metapopulation with extinctions and kin competition. J. Theoret. Biol. 199, 275–290.

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

Geritz, S.A.H., Metz, J.A.J., Kisdi, É., Meszéna, G., 1997. Dynamics of adaptation and evolutionary branching. Phys. Rev. Lett. 78, 2024–2027.

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

Gyllenberg, M., Metz, J.A.J., 2001. On fitness in structured metapopulations. J. Math. Biol. 43, 545–560.

Gyllenberg, M., Parvinen, K., Dieckmann, U., 2002. Evolutionary suicide and evolution of dispersal in structured metapopulations. J. Math. Biol. 45, 79–105.

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

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

Heino, M., Hanski, I., 2001. Evolution of migration rate in a spatially realistic metapopulation model. Am. Nat. 157, 495–511.

Higgins, K., Lynch, M., 2001. Metapopulation extinction caused by mutation accumulation. Proc. Natl. Acad. Sci. USA 98 (5), 2928–2933.

Holt, R.D., 1985. Population dynamics in two-patch environment: some anomalous consequences of an optimal habitat distribution. Theor. Popul. Biol. 28, 181–208.

Holt, R.D., McPeek, M., 1996. Chaotic population dynamics favors the evolution of dispersal. Am. Nat. 148, 709–718.

Johst, K., Brandl, R., 1997. Evolution of dispersal: the importance of the temporal order of reproduction and dispersal. Proc. R. Soc. Lond. Ser. B 264 (1378), 23–30.

Johst, K., Doebeli, M., Brandl, R., 1999. Evolution of complex dynamics in spatially structured populations. Proc. R. Soc. Lond. Ser. B 266, 1147–1154.

Kokko, H., López-Sepulcre, A., 2006. From individual dispersal to species ranges: perspectives for a changing world. Science 313 (5788), 789–791.

Kun, Á., Scheuring, I., 2006. The evolution of density-dependent dispersal in a noisy spatial population model. Oikos 115 (2), 308–320.

Lehmann, L., Mullon, C., Akçay, E., Van Cleve, J., 2016. Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. Evolution 70, 1689–1702.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., et al., 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecol. Lett. 7 (7), 601–613.

Levins, R., 1969. Some demographic and genetic consequenses of environmental heterogeneity for biological control. Bull. Entomol. Soc. Am. 15, 237–240.

Massol, F., Débarre, F., 2015. Evolution of dispersal in spatially and temporally variable environments: The importance of life cycles. Evolution 69, 1925–1937.

McPeek, M.A., Holt, R.D., 1992. The evolution of dispersal in spatially and temporally varying environments. Am. Nat. 140, 1010–1027.

Metz, J.A.J., Geritz, S.A.H., Meszéna, G., Jacobs, F.J.A., van Heerwaarden, J.S., 1996. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In: van Strien, S.J., Verduyn Lunel, S.M. (Eds.), Stochastic and Spatial Structures of Dynamical Systems. North-Holland, Amsterdam, pp. 183–231.

Metz, J.A.J., Gyllenberg, M., 2001. How should we define fitness in structured metapopulation models? including an application to the calculation of ES dispersal strategies. Proc. R. Soc. Lond. Ser. B 268, 499–508.

Motro, U., 1991. Avoiding inbreeding and sibling competition: the evolution of sexual dimorphism for dispersal. Am. Nat. 137, 108–115.

Nurmi, T., Parvinen, K., 2011. Joint evolution of specialization and dispersal in structured metapopulations. J. Theoret. Biol. 275, 78–92.
Nurmi, T., Parvinen, K., 2013. Evolution of specialization under non-equilibrium population dynamics. J. Theoret. Biol. 321, 63–77.
Nurmi, T., Parvinen, K., Selonen, V., 2018. Joint evolution of dispersal propensity and site selection in structured metapopulation models. J. Theoret. Biol. 444, 50–72.
Ohtsuki, H., Rueffler, C., Wakano, J.Y., Parvinen, K., Lehmann, L., 2020. The components of directional and disruptive selection in heterogeneous group-structured populations. J. Theoret. Biol. 507, 110449.
Parvinen, K., 1999. Evolution of migration in a metapopulation. Bull. Math. Biol. 61, 531–550.
Parvinen, K., 2002. Evolutionary branching of dispersal strategies in structured metapopulations. J. Math. Biol. 45, 106–124.
Parvinen, K., 2006. Evolution of dispersal in a structured metapopulation model in discrete time. Bull. Math. Biol. 68, 655–678.
Parvinen, K., Brännström, Å., 2016. Evolution of site-selection stabilizes population dynamics, promotes even distribution of individuals, and occasionally causes evolutionary suicide. Bull. Math. Biol. 78, 1749–1772.
Parvinen, K., Ohtsuki, H., Wakano, J.Y., 2020. Evolution of dispersal in a spatially heterogeneous population with finite patch sizes. Proc. Natl. Acad. Sci. USA 117, 7290–7295.
Parvinen, K., Seppänen, A., Nagy, J.D., 2012. Evolution of complex density-dependent dispersal strategies. Bull. Math. Biol. 74, 2622–2649.
Perrin, N., Mazalov, V., 1999. Dispersal and inbreeding avoidance. Am. Nat. 154, 282–292.
Poethke, H.J., Gros, A., Hovestadt, T., 2011. The ability of individuals to assess population density influences the evolution of emigration propensity and dispersal distance. J. Theoret. Biol. 282 (1), 93–99.
Poethke, H.J., Hovestadt, T., 2002. Evolution of density–and patch–size–dependent dispersal rates. Proc. R. Soc. Lond. Ser. B 269 (1491), 637–645.
Rodrigues, A.M.M., Gardner, A., 2012. Evolution of helping and harming in heterogeneous populations. Evolution 66–7, 2065–2079.
Ronce, O., 2007. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. Annu. Rev. Ecol. Evol. Syst. 38, 231–257.
Ronce, O., Perret, F., Olivieri, I., 2000. Evolutionarily stable dispersal rates do not always increase with local extinction rates. Am. Nat. 155, 485–496.
Taylor, P.D., 1988. An inclusive fitness model for dispersal of offspring. J. Theoret. Biol. 130, 363–378.
Taylor, P.D., 1990. Allele-frequency change in a class-structured population. Am. Nat. 135, 95–106.
Taylor, P.D., Frank, S.A., 1996. How to make a kin selection model. J. Theoret. Biol. 180, 27–37.
Travis, J.M.J., Murrell, D.J., Dytham, C., 1999. The evolution of density-dependent dispersal. Proc. R. Soc. Lond. Ser. B 266, 1837–1842.
Van Valen, L., 1971. Group selection and the evolution of dispersal. Evolution 25, 591–598.
Weigang, H.C., 2017. Coevolution of patch-type dependent emigration and patch-type dependent immigration. J. Theoret. Biol. 426, 140–151.
Wright, S., 1943. Isolation by distance. Genetics 28 (2), 114.