Eco-evolutionary dynamics in metacommunities: ecological inheritance, helping within- and harming between-species

Charles Mullon * [1] and Laurent Lehmann [1]

[1] *Department of Ecology and Evolution, University of Lausanne, 1004 Lausanne, Switzerland*

Abstract

Understanding selection on intra- and inter-specific interactions that take place in dispersal-limited communities is a challenge for ecology and evolutionary biology. The problem is that local demographic stochasticity generates eco-evolutionary dynamics that are generally too complicated to make tractable analytical investigations. Here, we circumvent this problem by approximating the selection gradient on a quantitative trait that influences local community dynamics, assuming that such dynamics are deterministic with a stable fixed point. We nonetheless incorporate unavoidable kin selection effects arising from demographic stochasticity. Our approximation reveals that selection depends on how an individual expressing a trait-change influences: (1) its own fitness and the fitness of its current relatives; and (2) the fitness of its downstream relatives through modifications of local ecological conditions (i.e., through ecological inheritance). Mathematically, the effects of ecological inheritance on selection are captured by dispersal-limited versions of press-perturbations of community ecology. We use our approximation to investigate the evolution of helping within- and harming between-species when these behaviours influence demography. We find that individually costly helping evolves more readily when intra-specific competition is for material resources rather than for space because in this case, the costs of kin competition are paid by downstream relatives. Similarly, individually costly harming between species evolves when it alleviates downstream relatives from inter-specific competition. Beyond these examples, our approximation can help better understand the influence of ecological inheritance on a variety of eco-evolutionary dynamics in metacommunities, from consumer-resource and predator-prey coevolution to selection on mating systems with demographic feedbacks.

Keywords: kin selection, adaptive dynamics, metacommunity, altruism, mutualism, niche construction.

*Electronic address: `charles.mullon@unil.ch`

1 Introduction

Interactions within and between species are extremely common in nature and probably connect almost all living organisms to one another. How such intra- and inter-specific interactions evolve depends on interdependent ecological and evolutionary processes (Urban, 2011, Wagner et al., 2017, terHorst et al., 2018), also known as eco-evolutionary dynamics (Lion, 2017 for review). One major difficulty in understanding these dynamics is due to the spatial structuring of communities, which emerges from the physical limitations of movements and interactions. This spatial structure is captured by the notion of a "metacommunity", in which individuals of different species are divided among small local patches connected to each other by dispersal (e.g., Hanski and Gilpin, 1997, Clobert et al., 2001, Urban et al., 2008, Leibold and Chase, 2017).

When dispersal among the patches of a metacommunity is limited, individual reproductive and survival variance generate local demographic stochasticity. This has two complicating consequences for eco-evolutionary dynamics. First, it causes genetic stochasticity, whereby allele frequencies fluctuate within and between patches. These fluctuations lead to the build up of genetic relatedness between members of the same species. Genetic relatedness then influences selection on traits, in particular social traits, which like helping, are traits that affect the reproductive success of both their actor (direct effects) and recipient (indirect effects) (e.g., Hamilton, 1971, Hamilton and May, 1977, Taylor, 1994, Taylor and Frank, 1996, Frank, 1998, Rousset, 2004, West et al., 2007, Lion and van Baalen, 2007b, Van Cleve, 2015). Second, local demographic stochasticity results in ecological stochasticity, whereby the abundance of different species fluctuate within and between patches. As a consequence, multi-species demography in patch structured populations is significantly more complicated that in panmictic populations (Chesson, 1978, 1981, Hubbell, 2001, Neuhauser, 2002, Cornell and Ovaskainen, 2008). As genetic and ecological stochasticity are further likely to influence one another through eco-evolutionary feedbacks, understanding selection on traits mediating ecological interactions is a major challenge when dispersal is limited.

In complicated demographic scenarios, fundamental insights into selection can be obtained from the long-term adaptive dynamics of quantitative traits. These dynamics follow the gradual changes of traits displayed by a population under the constant but limited influx of mutations (e.g., Eshel, 1983, Parker and Maynard Smith, 1990, Christiansen, 1991, Grafen, 1991, Abrams et al., 1993, Metz et al., 1996, Eshel, 1996, Geritz et al., 1998, Rousset, 2004). One of the main goals of evolutionary analysis is to identify local attractors of such adaptive dynamics. These attractors are trait values towards which selection drives a population under gradual evolution (referred to as convergence stable phenotypes, Eshel, 1983, Taylor, 1989, Christiansen, 1991, Geritz et al., 1998, Rousset, 2004, Leimar, 2009). Convergence stable phenotypes can be identified from the selection gradient on a trait, which is the marginal change in the fitness of an individual due to this individual and all its relatives

changing trait value. Such analysis has helped understand how natural selection moulds phenotypic traits of broad biological interest, from senescence, dispersal, mate choice, life history, sperm competition, sex-ratio, to altruism, cumulative cultural evolution, bet hedging and optimal foraging (e.g., Hamilton, 1966, Charnov, 1976, Schaffer, 1982, Taylor, 1988b, Parker, 1990, Taylor, 1988a, Frank, 1998, Gardner and West, 2004, Foster, 2004, Kisdi and Priklopil, 2010, Kuijper et al., 2012, Akçay and Van Cleve, 2012, Mullon et al., 2014, Wakano and Miura, 2014, Kobayashi et al., 2015, Weigang and Kisdi, 2015, Nurmi et al., 2018).

Gold standard expressions for the selection gradient on traits that influence the demography of a single species, where all consequences of genetic and ecological stochasticity for natural selection are taken into account, have been worked out long ago (Rousset and Ronce, 2004, eqs. 23-24, Rousset, 2004, chapter 11). In principle, these expressions can be directly extended to consider multi-species interactions. However, even under the simplest model of limited dispersal, which is the island model of dispersal (Wright, 1931), the evaluation of the selection gradient on traits affecting multi-species eco-evolutionary dynamics remains dispiritingly complicated (Rousset and Ronce, 2004, Lehmann et al., 2006, Alizon and Taylor, 2008, Wild, 2011). As a result, the selection gradient is most often computed numerically as the derivative of an invasion fitness measure, without the provision of any biological interpretation of selection on the trait under focus (Metz and Gyllenberg, 2001, Cadet et al., 2003, Parvinen et al., 2003, Parvinen and Metz, 2008). Only specific demographic models under limited dispersal with finite patch size have been studied analytically in detail (Comins et al., 1980, Gandon and Michalakis, 1999, Lehmann et al., 2006, Rodrigues and Gardner, 2012). A biologically intuitive understanding of selection on traits that influence metacommunity dynamics is therefore out of immediate reach when using the exact selection gradient.

To circumvent the difficulty of computing the selection gradient under limited dispersal, various approximations have been proposed. The most prominent is perhaps the heuristic pair approximation, which has been used to study intra-specific social evolution and host-parasite coevolution in lattice structured populations (e.g., Nakamaru et al., 1997, van Baalen and Rand, 1998, Le Galliard et al., 2003, 2005, Nakamaru and Iwasa, 2005, Lion and van Baalen, 2007a, Lion and Gandon, 2009, 2010, Débarre et al., 2012). However, more general multi-species coevolution scenarios have not been much investigated using pair approximation. This is presumably because analytical explorations remain complicated in lattice structured populations due to the effects of isolation-by-distance.

In this paper, we present a novel heuristic approximation for the selection gradient on traits that influence eco-evolutionary dynamics in patch-structured populations, which do not experience isolation-by-distance or heterogeneities in abiotic factors (i.e., the population is structured according to the homogeneous island model of dispersal of Wright, 1931, and see Chesson, 1981, for its ecological counterpart). The crux of this approxi-

mation is that it assumes that local population size dynamics are deterministic with a stable fixed point (i.e., we ignore ecological stochasticity, and periodic or chaotic population dynamics). This assumption allows us to reach key analytical insights, which can be applied to understand a wide spectrum of multi-species interactions. Importantly, our approximation provides a biologically meaningful interpretation of selection on traits that influence ecological interactions under limited dispersal. The rest of the paper is organized as follows. (1) We describe a stochastic metacommunity eco-evolutionary model. (2) We motivate an approximation of our model that ignores ecological stochasticity. (3) Under this approximation, we derive analytically the selection gradient on a trait that influences eco-evolutionary dynamics through intra- and inter-specific interactions. (4) We use our approximation to study two examples of social and ecological interactions: evolution of helping within- and harming between-species when these behaviours influence demography. We show that for these examples, our approximation performs well compared to individual-based simulations in predicting equilibrium strategies and the species abundances these strategies generate.

2 Model

2.1 Metacommunity structure

We consider an infinite number of patches that are connected by uniform dispersal (Wright's 1931 infinite island model of dispersal). On each patch, a community of up to S species may coexist. The life cycle events of each species $i \in \{1, 2, ..., S\}$ are as follows. (1) Each adult produces a number of offspring that follows a Poisson distribution (with a mean that may depend on local interactions within and between species). (2) Each adult then either survives or dies (with a probability that may depend on local interactions within and between species). (3) Independently of one another, each offspring either remains in its natal patch or disperses to another randomly chosen one (the dispersal probability is assumed to be non-zero for all species but may differ among species). (4) Each offspring either dies or survives to adulthood (with a probability that may depend on local population numbers, for instance if space on each patch is a limiting factor).

2.2 Evolving phenotypes and the uninvadable species coalition

Each individual expresses a genetically determined evolving phenotype, or strategy, which can affect any event, such as reproduction, survival, or dispersal, in the life cycle of any species. We assume that the expression of a strategy and its effects are independent of age (i.e., no age-structure). We denote by Θ_i the set of feasible strategies for species i (this set is either the set or a subset of the real numbers, $\Theta_i \subset \mathbb{R}$). When the population of each

species is monomorphic, the metacommunity (i.e., the collection of subdivided populations of each species) is described by a vector of strategies $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_S)$, where θ_i is the strategy expressed by all individuals of species i (i.e. $\boldsymbol{\theta}$ denotes a monomorphic resident population).

We define $\boldsymbol{\theta}$ as an uninvadable coalition if any mutation $\tau_i \in \Theta_i$, which arises in any species i and which results in a unilateral deviation $\boldsymbol{\tau}_i = (\theta_1, ..., \theta_{i-1}, \tau_i, \theta_{i+1}, ..., \theta_S)$ of the resident vector, goes extinct. The concept of an uninvadable coalition is the same as the concepts of a multi-species evolutionary stable strategy (Brown and Vincent, 1987, p. 68) and of an evolutionary stable coalition (Apaloo and Butler, 2009, p. 640).[1]

2.3 Adaptive dynamics

Under the above definition of uninvadability, it is sufficient to consider the fate of a unilateral phenotypic deviation in one species at a time in order to determine whether a coalition is uninvadable. We can therefore focus our attention on the evolutionary dynamics of a mutant allele, which arises in species i and codes for phenotype τ_i, when the resident community expresses $\boldsymbol{\theta}$. In the infinite island model of dispersal, the change Δp_i in frequency p_i of such a mutant allele over one demographic time period (one life cycle iteration) can be written as,

$$\Delta p_i = p_i(1 - p_i)\delta_i \mathcal{S}_i(\boldsymbol{\theta}) + O(\delta_i^2), \tag{1}$$

where $\delta_i = \tau_i - \theta_i$ is the phenotypic effect of the mutation (Rousset, 2004, pp. 206–207 and Rousset and Ronce, 2004 p. 129, with notation here adapted to introduce species specific allele frequency change). The function $\mathcal{S}_i(\boldsymbol{\theta})$, which depends on quantities evaluated only in the resident community $\boldsymbol{\theta}$, is the selection gradient on the trait in species i. When selection is weak (so that $|\delta| \ll 1$ and terms $O(\delta_i^2)$ can be neglected), the selection gradient $\mathcal{S}_i(\boldsymbol{\theta})$ and phenotypic effect δ_i give the direction of selection on the mutant at any allele frequency: selection favours the invasion and fixation of the mutation when $\delta_i \mathcal{S}_i(\boldsymbol{\theta}) > 0$, and conversely, extinction when $\delta_i \mathcal{S}_i(\boldsymbol{\theta}) < 0$. The selection gradient thus captures directional selection.

The selection gradient $\mathcal{S}_i(\boldsymbol{\theta})$ is useful to derive necessary conditions for a coalition to be uninvadable. First, a necessary condition for a coalition $\boldsymbol{\theta}^*$ that lies in the interior of the state space to be uninvadable is that,

$$\mathcal{S}_i(\boldsymbol{\theta}^*) = 0 \qquad \forall i. \tag{2}$$

Such a coalition $\boldsymbol{\theta}^*$ is said to be singular. The allele frequency change eq. (1) also informs us whether a singular coalition $\boldsymbol{\theta}^*$ will be approached by gradual evolution from its neighbourhood under the constant but limited in-

[1] We here refrain of using the terminology "evolutionary stability" as it subsumes that such strategies are attractor of the evolutionary dynamics (Maynard Smith, 1982), which is not covered by the concept of uninvadability.

flux of mutations, i.e., if it is convergence stable. A singular coalition is convergence stable when the eigenvalues of the $S \times S$ Jacobian matrix, $\mathbf{J}(\boldsymbol{\theta}^*)$, with (i, j) entry,

$$\left(\mathbf{J}(\boldsymbol{\theta}^*)\right)_{ij} = \left.\frac{\partial \mathcal{S}_i(\boldsymbol{\theta})}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}, \tag{3}$$

all have negative real parts (e.g., Débarre et al., 2014, eq. 7a).

When at most two alleles can ever segregate in a given species, a convergence stable strategy is also locally uninvadable (Débarre and Otto, 2016). In that case, the collection of selection gradients acting in each species is sufficient to establish whether a coalition is uninvadable. When more that two alleles can segregate at a locus, establishing local uninvadability requires looking into the second-order effects of selection (i.e., terms of $O(\delta_i^2)$ in eq. 1, Taylor, 1989, Geritz et al., 1998, Rousset, 2004, Leimar, 2009). These effects, which capture disruptive selection, are difficult to characterize analytically under limited dispersal (e.g., Ajar, 2003, Mullon et al., 2016). We will therefore focus on the effects of directional selection in this paper.

2.4 The selection gradient in metacommunities

The selection gradient, $\mathcal{S}_i(\boldsymbol{\theta})$, in a given species i in the island model can be written as a change in the fitness of individuals experiencing different demographic states, weighted jointly by reproductive values and relatedness coefficients (eqs. 26-27 of Rousset and Ronce, 2004 for a single-species demography dynamics, or eqs. E.27-29 of Lehmann et al., 2016 for arbitrary demographic states). Reproductive values reflect the fact that individuals residing in different demographic states contribute differently to the future gene pool (Rousset and Ronce, 2004, Lehmann et al., 2016). Reproductive values thus capture the effects of ecological stochasticity on selection. Relatedness, meanwhile, captures the fact that individuals from the same species that reside in the same patch are more likely to carry alleles identical-by-descent than randomly sampled individuals in the population (see Nagylaki, 1992 and Rousset, 2004 for textbook treatments). The relatedness coefficients in the selection gradient thus reflect the consequences of genetic stochasticity on selection.

In spite of the insights brought by the exact selection gradient on ecological and genetic stochasticity, its usage to study community evolution under the assumptions of our model presents two caveats. The first is that owing to the large number of possible demographic states within patches (all possible configurations of the number of individuals of all species on a patch), the necessary computations are not straightforward and would be extremely expensive numerically (as shown by the computations necessary even in the absence of inter-specific interactions, Rousset and Ronce, 2004, Lehmann et al., 2006, Alizon and Taylor, 2008, Wild et al., 2009, Wild, 2011). It is possibly due to this computational hurdle that no application of the exact selection gradient to the coevolution

of multiple species that experience stochastic demography and limited dispersal can be found in the literature. The second caveat is that the expression of selection in terms of reproductive values applies to any type of demographic structuring (e.g., by age, stage, sex, or environment, Frank, 1998, Rousset, 2004, Grafen, 2006). As a consequence, without solving reproductive values explicitly in terms of model parameters, the exact selection gradient carries little biological information about how local ecological interactions influence selection.

The goal of this paper is to provide a tractable and biologically informative approximation for the selection gradient in a metacommunity. We propose to achieve this by assuming that changes in local population size are deterministic with a stable fixed point when the metacommunity is monomorphic for $\boldsymbol{\theta}$. Resident patches will therefore experience neither stochastic ecological fluctuations nor periodic/chaotic dynamics. As a consequence, it will no longer be necessary to consider all the possible demographic states that a focal patch can transit between. Before deriving this approximation, let us first study resident demographic dynamics (i.e., in a metacommunity monomorphic for the resident $\boldsymbol{\theta}$) in order to investigate how and when these dynamics can be assumed to be deterministic.

3 Resident community dynamics

3.1 Deterministic resident community dynamics

Our life-cycle assumptions (section 2.1) entail that we are considering an infinite stochastic system, i.e., we have an infinite number of interacting multidimensional Markov chains (each chain describes the community dynamics on a single patch and chains interact with one another through dispersal, Chesson, 1981, 1984). For this system, let us consider the expected number of individuals (or abundance) of each species in a focal patch at a given demographic time t, which we denote by $\boldsymbol{n}_t(\boldsymbol{\theta}) = (n_{1,t}(\boldsymbol{\theta}), n_{2,t}(\boldsymbol{\theta}), \ldots, n_{S,t}(\boldsymbol{\theta}))$ where $n_{i,t}(\boldsymbol{\theta})$ is the expected abundance of species i (note that we do not need to label patches since they are identical on average, e.g., Neuhauser, 2002). This expected abundance can be written as $n_{i,t}(\boldsymbol{\theta}) = F_i\big(\boldsymbol{\theta}, \boldsymbol{N}_{t-1}, \overline{\boldsymbol{n}}_{t-1}(\boldsymbol{\theta})\big)$, for $i = 1, 2, ..., S$, where the population dynamic transition map F_i gives the expected number of individuals of species i, given the local community state $\boldsymbol{N}_{t-1} = (N_{1,t-1}, N_{2,t-1}, ..., N_{S,t-1})$ in the previous demographic time period (i.e., $N_{i,t-1}$ denotes the random variable for the number of individuals of species i in the focal patch at time $t-1$), and when the global average community state is $\overline{\boldsymbol{n}}_{t-1}(\boldsymbol{\theta}) = (\overline{n}_{1,t-1}(\boldsymbol{\theta}), \overline{n}_{2,t-1}(\boldsymbol{\theta}), ..., \overline{n}_{S,t-1}(\boldsymbol{\theta}))$ (i.e., $\overline{n}_{i,t-1}(\boldsymbol{\theta})$ is the average number of individuals of species i across all patches at $t-1$, note that $\overline{\boldsymbol{n}}_{t-1}(\boldsymbol{\theta})$ is not a random variable because there is an infinite number of patches). We have written the transition map $F_i\big(\boldsymbol{\theta}, \boldsymbol{N}_{t-1}, \overline{\boldsymbol{n}}_{t-1}(\boldsymbol{\theta})\big)$ such that it also depends explicitly on the vector of phenotypes $\boldsymbol{\theta}$ of each species in the focal patch in the previous time period, which will be useful when we introduce selection.

The basis of our approximation is to assume that the dynamics of the abundance of each species in a patch are deterministic (in an abuse of notation, $\boldsymbol{N}_t \approx \mathrm{E}[\boldsymbol{N}_t] = \boldsymbol{n}_t(\boldsymbol{\theta})$), so that the ecological dynamics on a focal patch, which are no longer stochastic, are given by

$$n_{i,t}(\boldsymbol{\theta}) = F_i\big(\boldsymbol{\theta}, \boldsymbol{n}_{t-1}(\boldsymbol{\theta}), \overline{\boldsymbol{n}}_{t-1}(\boldsymbol{\theta})\big) \qquad \text{for } i = 1,2,...,S \tag{4}$$

(Chesson, 1981). We further assume that these deterministic ecological dynamics are such that abundances of all species converge to a stable fixed point. From the dynamical eq. (4), this ecological fixed point, which we denote as $\hat{\boldsymbol{n}}(\boldsymbol{\theta}) = (\hat{n}_1(\boldsymbol{\theta}), \hat{n}_2(\boldsymbol{\theta}), \ldots, \hat{n}_S(\boldsymbol{\theta}))$, solves

$$\hat{n}_i(\boldsymbol{\theta}) = F_i(\boldsymbol{\theta}, \hat{\boldsymbol{n}}(\boldsymbol{\theta}), \hat{\boldsymbol{n}}(\boldsymbol{\theta})) \qquad \text{for } i = 1,2,...,S. \tag{5}$$

Local stability of the fixed point $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$ entails that it is such that the local community matrix (e.g., Yodzis, 1989, Case, 2000)

$$\mathbf{C}(\boldsymbol{\theta}) = \begin{pmatrix} \left.\frac{\partial F_1(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_1}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} & \left.\frac{\partial F_1(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_2}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} & \cdots & \left.\frac{\partial F_1(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_S}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} \\ \left.\frac{\partial F_2(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_1}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} & \left.\frac{\partial F_2(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_2}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} & \cdots & \left.\frac{\partial F_2(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_S}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} \\ \vdots & \vdots & \ddots & \vdots \\ \left.\frac{\partial F_S(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_1}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} & \left.\frac{\partial F_S(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_2}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} & \cdots & \left.\frac{\partial F_S(\boldsymbol{\theta},\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_S}\right|_{\boldsymbol{n}=\hat{\boldsymbol{n}}} \end{pmatrix}, \tag{6}$$

whose $(i\,j)$-entry measures the influence of the abundance of species j on the abundance of species i over one demographic time period, has eigenvalues all with absolute value less than one.

3.2 Illustrating example

In order to illustrate the transition function (F_i, eq. 4) and provide a basis for later comparisons with individual-based simulations, consider two semelparous species whose life-cycle is as follows. (1) Each individual of species $i \in \{1,2\}$ in a patch with $\boldsymbol{n} = (n_1, n_2)$ individuals produces a mean number $f_i / \big(1 + \gamma n_i + \eta n_j\big)$ of offspring (with $j \in \{1,2\}$ and $j \neq i$), where f_i is the number of offspring produced in the absence of density-dependent competition. The denominator $1 + \gamma n_i + \eta n_j$ captures density-dependent competition within species (with intensity γ), and between species (with intensity η). This model for offspring production can be seen as a special case of the Leslie-Gower model of species interaction (Leslie and Gower, 1958, eq. 1.1). (2) All adults die. (3) Independently of one another, each offspring of species i disperses with probability $m_i(\boldsymbol{\theta})$. (4) Finally, all offspring survive to adulthood.

According to this life cycle, the abundance of species 1 and 2 in the focal patch, conditional on the abundance

being $\boldsymbol{n}_{t-1}(\boldsymbol{\theta})$ in the previous time period on the focal patch and on the abundance in other patches being at a stable equilibrium $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$, can be written as

$$\begin{aligned} n_{1,t}(\boldsymbol{\theta}) &= (1-m_1)\left(\frac{f_1}{1+\gamma n_{1,t-1}(\boldsymbol{\theta})+\eta n_{2,t-1}(\boldsymbol{\theta})}\right) n_{1,t-1}(\boldsymbol{\theta}) + m_1 \hat{n}_1(\boldsymbol{\theta}) \\ n_{2,t}(\boldsymbol{\theta}) &= (1-m_2)\left(\frac{f_2}{1+\gamma n_{2,t-1}(\boldsymbol{\theta})+\eta n_{1,t-1}(\boldsymbol{\theta})}\right) n_{2,t-1}(\boldsymbol{\theta}) + m_2 \hat{n}_2(\boldsymbol{\theta}). \end{aligned} \tag{7}$$

Eq. (7) is an example of a so-called coupled map lattice with implicit space (e.g., eq. 3.11 of Ranta et al., 2006). The first summand in each line of eq. (7) is the number of settled individuals in the focal patch that are born locally. The second summand in each line is the total number of offspring that immigrate into the focal patch from other patches. In order to understand better this second summand, consider that when the population is at the resident demographic equilibrium, an individual on average produces one offspring (so that the total number of individuals remains constant). As a consequence, $\hat{n}_1(\boldsymbol{\theta})$ and $\hat{n}_2(\boldsymbol{\theta})$ give the total number of offspring produced of species 1 and 2 respectively, in any patch other than the focal one. Therefore, $m_1(\boldsymbol{\theta})\hat{n}_1(\boldsymbol{\theta})$ and $m_2(\boldsymbol{\theta})\hat{n}_2(\boldsymbol{\theta})$ in eq. (7) give the average number of offspring immigrating into the focal patch of species 1 and 2, respectively.

The equilibrium abundance of both species is found by substituting eq. (7) in eq. (5) (i.e., putting $n_{1,t}(\boldsymbol{\theta}) = n_{1,t-1}(\boldsymbol{\theta}) = \hat{n}_1(\boldsymbol{\theta})$ and $n_{2,t}(\boldsymbol{\theta}) = n_{2,t-1}(\boldsymbol{\theta}) = \hat{n}_2(\boldsymbol{\theta})$) and solving for $\hat{n}_1(\boldsymbol{\theta})$ and $\hat{n}_2(\boldsymbol{\theta})$ simultaneously. Doing so, we find that the unique positive equilibrium (i.e., $\hat{n}_1(\boldsymbol{\theta}) > 0$ and $\hat{n}_2(\boldsymbol{\theta}) > 0$) reads as

$$\begin{aligned} \hat{n}_1(\boldsymbol{\theta}) &= \frac{(f_1-1)\gamma-(f_2-1)\eta}{\gamma^2-\eta^2} \\ \hat{n}_2(\boldsymbol{\theta}) &= \frac{(f_2-1)\gamma-(f_1-1)\eta}{\gamma^2-\eta^2}. \end{aligned} \tag{8}$$

This reveals that for the two species to co-exist, it is necessary for intra-specific competition to be stronger than inter-specific competition ($\gamma > \eta$), which is a classical result (Case, 2000).

3.3 Comparing deterministic and stochastic dynamics

In order to assess when ecological stochasticity can be ignored (i.e., when eq. 4 accurately reflects the true stochastic dynamics), we compared the deterministic community dynamics of the Leslie-Gower model (eq. 7) with individual based simulations of the full stochastic model (see Appendix A for a description of the simulation procedure).

3.3.1 Ecological stochasticity

We find that there is a good qualitative match between the deterministic equilibrium abundance given by eq. (8), and the average number of individuals of each species in a group observed in stochastic individual based simulations (Fig. 1, top row). As predicted by theory (Chesson, 1981, Neuhauser, 2002), the deterministic equilibrium deviates systematically from the observed average (Fig. 1, top row). However, these deviations are small provided dispersal is not too weak (roughly no less than 0.1) and competition is such that deterministic abundance on a patch is greater or equal to 10 individuals (Fig. 1, middle row). This suggests that under such demographic situations, ecological stochasticity can be ignored.

When fecundity is low and competition is weak, one might expect that ecological stochasticity is too important to be ignored as forces stabilising the ecological equilibrium are weaker. Strikingly, even when fecundity is extremely low (barely above one, which is the threshold for the population to be maintained), it is still true that as long as dispersal is greater or equal to 0.1 and competition is such that deterministic abundance on a patch is greater or equal to 10 individuals, deterministic ecological dynamics are a very good approximation of the stochastic process (Fig. 1, left column). The mitigating effects of dispersal on ecological stochasticity is further illustrated by the observation that in stochastic simulations, variation in abundance among patches rapidly becomes vanishingly small as dispersal increases (Fig. 1, bottom row).

Why dispersal mitigates the effects of ecological stochasticity can be understood as follows. Local population dynamics depend on the balance between two processes: 1) a local process at the patch level (i.e., dependence on $\boldsymbol{N}_{t-1}$), which has a strong stochastic component when patches have few individuals; and 2) a global process at the metacommunity level (i.e., dependence on $\overline{\boldsymbol{n}}_{t-1}(\boldsymbol{\theta})$), which has a weak stochastic component when the number of patches is large (in fact, as the number of patches grow infinite, patches affect each other deterministically, Chesson, 1981). As dispersal increases, local population dynamics depend increasingly on the global process and less on the local one. As a result, local population dynamics become increasingly deterministic.

3.3.2 Genetic stochasticity

The above analysis suggests that ecological stochasticity can be ignored when dispersal values are roughly above 0.1 and demographic parameters are such that equilibrium abundance within species is greater or equal to 10 (Fig. 1). This raises the question of whether genetic stochasticity can also be ignored for such parameter values. The consequence of genetic stochasticity for selection can be ignored when relatedness coefficients are very small. The standard relatedness coefficient in the island model is the probability $r_i(\boldsymbol{\theta})$ that two individuals from the same species i, which are randomly sampled in the same patch, carry an allele that is identical-by-descent

when the population is monomorphic for $\boldsymbol{\theta}$ (also referred to as pairwise relatedness, Frank, 1998, Rousset, 2004). Let us consider this probability when the community has reached its (deterministic) demographic equilibrium $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$ (given by eq. 8). Owing to our assumption that fecundity is Poisson distributed, pairwise relatedness satisfies the relationship

$$r_i(\boldsymbol{\theta}) = (1 - m_i(\boldsymbol{\theta}))^2 \left(\frac{1}{\hat{n}_i(\boldsymbol{\theta})} + \left(1 - \frac{1}{\hat{n}_i(\boldsymbol{\theta})}\right) r_i(\boldsymbol{\theta}) \right), \tag{9}$$

which can be understood as follows. With probability $(1 - m_i(\boldsymbol{\theta}))^2$, two randomly sampled individuals of species i are both of philopatric origin (i.e., they were born in the focal patch). Then, with probability $1/\hat{n}_i(\boldsymbol{\theta})$, these individuals descend from the same parent so their relatedness is one. With complementary probability $1 - 1/\hat{n}_i(\boldsymbol{\theta})$, they descend from different parent so their relatedness is $r_i(\boldsymbol{\theta})$. The solution to eq. (9) is

$$r_i(\boldsymbol{\theta}) = \frac{(1 - m_i(\boldsymbol{\theta}))^2}{1 + \left[1 - (1 - m_i(\boldsymbol{\theta}))^2\right](\hat{n}_i(\boldsymbol{\theta}) - 1)}, \tag{10}$$

which is equivalent to the standard F_{ST} quantity (when individuals are sampled without replacement, e.g, Rousset, 2004, Hartl and Clark, 2007). Note, however, that in contrast to most mathematical treatments of F_{ST}, the number of individuals $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$ here is endogenously determined by an explicit demographic process (eqs. 7-8).

Inspection of eq. (10) reveals that relatedness can build up to significant values, even when dispersal is within a range under which we can legitimately approximate mean abundance by the deterministic model (e.g., local population size must be greater than 422 for relatedness to be less than 0.01 when dispersal is equal to 0.1). This shows that there exists a demographic regime under which ecological stochasticity can be neglected, but genetic stochasticity cannot (which is in line with the fact that genetic stochasticity can lead to significant levels of relatedness even when patch size is constant and there is no ecological stochasticity, Rousset, 2004, Hartl and Clark, 2007). We will therefore take into account the effects of genetic stochasticity when deriving our approximation for the selection gradient. It is noteworthy that we find an excellent match between pairwise relatedness observed in individual based simulations, and pairwise relatedness calculated from the deterministic ecological approximation (i.e., eq. 10 with eq. 8, Figure 2). This lends further support to the usefulness of the deterministic ecological approximation to study populations at ecological equilibrium (eq. 5).

4 Evolutionary analysis

We now specify the (approximate) selection gradient on a trait expressed in species i, when the effects of ecological stochasticity are neglected. First, we characterise ecological dynamics when they can be influenced by the presence of genetic mutants.

4.1 Mutant community dynamics

We now assume that two alleles segregate in the focal species i: a rare mutant that codes for phenotype τ_i and a resident for θ_i. We focus on a focal patch in which both alleles are present, while other patches are considered to be monomorphic for the resident $\boldsymbol{\theta}$, and at their ecological equilibrium, $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$ (eq. 5). In this focal patch, we assume that the number of individuals, $n_{j,t}(\boldsymbol{\tau}_i)$, of species j at time t is given by

$$n_{j,t}(\boldsymbol{\tau}_i) = F_j\big(\bar{\boldsymbol{\tau}}_{i,t-1}, \boldsymbol{n}_{t-1}(\boldsymbol{\tau}_i), \hat{\boldsymbol{n}}(\boldsymbol{\theta})\big) \qquad \text{for } j = 1,2,...,S, \tag{11}$$

where F_j is the map introduced in section 3.1 (eq. 4) but the first argument of this map now is $\bar{\boldsymbol{\tau}}_{i,t-1} = (\theta_1,...,\theta_{i-1},\bar{\tau}_{i,t-1},\theta_{i+1},...,\theta_S)$, which is a vector collecting the average phenotypes expressed in each species in the focal patch at demographic time $t-1$ (in species $j \neq i$ other than the focal, this average is simply the resident θ_j; in the focal species i, this average is denoted by $\bar{\tau}_{i,t}$). Since the average phenotype in the focal species, $\bar{\tau}_{i,t}$, depends on the number of genetic mutants, the first argument of F_j in eq. (11) captures the effect of the genetic state on local abundance. The dependence of F_j on the average phenotype approximates possibly more complicated relationships between genetic state and abundance to the first order, which is sufficient to evaluate the selection gradient (Rousset, 2004, p. 95). The map F_j also depends on $\boldsymbol{n}_{t-1}(\boldsymbol{\tau}_i) = (n_{1,t-1}(\boldsymbol{\tau}_i),...,n_{S,t-1}(\boldsymbol{\tau}_i))$, which is the ecological state of the focal patch at time $t-1$, and on the equilibrium $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$, which is the ecological state of other patches.

Because we take genetic stochasticity into account, the number of mutants, and hence the average phenotype in the focal species, $\bar{\tau}_{i,t}$, fluctuate randomly and should be considered as random variables. As a result, the abundance $n_{j,t}(\boldsymbol{\tau}_i)$ at time t given by eq. (11) is also a random variable. Importantly, this stochasticity in abundance is only due to genetic stochasticity in our approximation. When the average phenotype $\bar{\tau}_{i,t}$ is fixed (for instance for the resident, $\bar{\tau}_{i,t} = \theta_i$), ecological dynamics are fully deterministic and given by the recurrence eq. (11). In other words, our approximation ignores the influence of ecological stochasticity on ecological dynamics.

4.2 The inclusive fitness effect for the interactive community

In order to derive our approximation for the selection gradient, we use the basic reproductive number as an invasion fitness proxy (e.g., Stearns, 1992, Charlesworth, 1994, Case, 2000, Metz and Gyllenberg, 2001, Lehmann et al., 2016). This allows us to drastically simplify our calculations and to equivalently characterise directional selection (i.e., the first-order effects of selection on allele frequency change). These points are further detailed in Appendix B, where we show that the selection gradient on an evolving trait in species i in a $\boldsymbol{\theta}$ community can be

approximated as

$$\mathcal{S}_i(\boldsymbol{\theta}) \approx \epsilon(\boldsymbol{\theta}) s_i(\boldsymbol{\theta}), \tag{12}$$

where $\epsilon(\boldsymbol{\theta}) > 0$ is a factor of proportionality that depends on $\boldsymbol{\theta}$ only (see eqs. B.3 - B.19), and

$$s_i(\boldsymbol{\theta}) = s_{\mathrm{w},i}(\boldsymbol{\theta}) + s_{\mathrm{e},i}(\boldsymbol{\theta}) \tag{13}$$

is the approximate selection gradient. Since $\epsilon(\boldsymbol{\theta}) > 0$, $s_i(\boldsymbol{\theta})$ is sufficient to ascertain singular trait values and their convergence stability (by replacing $\mathcal{S}_i(\boldsymbol{\theta})$ with $s_i(\boldsymbol{\theta})$ in eqs. 2 and 3). The approximate selection gradient, $s_i(\boldsymbol{\theta})$, consists of two terms: $s_{\mathrm{w},i}(\boldsymbol{\theta})$, which captures selection owing to the trait's intra-temporal effects (effects within a demographic period); and $s_{\mathrm{e},i}(\boldsymbol{\theta})$, which captures selection owing to the trait's inter-temporal effects (effects between demographic periods) that emerge as a result of ecological inheritance (i.e., modified environmental conditions passed down to descendants, Odling-Smee et al., 2003, Bonduriansky, 2012). We detail these two components of selection in the next two sections.

4.2.1 Selection on intra-temporal effects

The first term of eq. (13) can be expressed as

$$s_{\mathrm{w},i}(\boldsymbol{\theta}) = \left.\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tau_{\bullet,i}}\right|_{\substack{\tau_i=\theta_i \\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} + \left.\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tilde{\tau}_i}\right|_{\substack{\tau_i=\theta_i \\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} \times r_i(\boldsymbol{\theta}) \tag{14}$$

(see Appendix B.3.4 for derivation), where w_i is the individual fitness of a focal individual of species i (i.e., the expected number of successful offspring produced over one life cycle iteration by the focal, including itself if it survives). Individual fitness, $w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})$, is written as a function of four variables: (1) the phenotype $\tau_{\bullet,i}$ of the focal individual; (2) the vector $\tilde{\boldsymbol{\tau}}_i = (\theta_1, ..., \theta_{i-1}, \tilde{\tau}_i, \theta_{i+1}, ..., \theta_S)$ of average phenotypes of neighbours in the focal patch (where $\tilde{\tau}_i$ is the average phenotype among the neighbors of species i of the focal individual); (3) the vector of abundances in the focal community $\boldsymbol{n}$; and (4) the vector of average abundance across the metacommunity, which is at its equilibrium $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$ (explicit examples of such a fitness function are given later when we apply our method, see eqs. 23, 36 and C.1 in Appendix C). Note that individual fitness may also depend on the phenotype expressed in patches other than the focal, which is the resident $\boldsymbol{\theta}$, but we have chosen not to write this dependency explicitly.

The two derivatives in eq. (14), which are evaluated in the resident population (i.e., with resident phenotype $\tau_i = \theta_i$ and resident ecological equilibrium $\boldsymbol{n} = \hat{\boldsymbol{n}}(\boldsymbol{\theta})$), capture different fitness effects of the trait. The first derivative represents the change in the fitness of a focal individual of species i resulting from this individual switching from the resident to the mutant phenotype (i.e., the direct effect of the trait). The second derivative

can be interpreted as the change in the fitness of the whole set of same-species patch neighbours resulting from the focal individual switching from the resident to mutant phenotype (i.e., the indirect effect of the trait). This second derivative is weighted by the neutral relatedness coefficient, $r_i(\boldsymbol{\theta})$, which gives the probability that any same-species neighbour also carries the mutation in the monomorphic resident.

4.2.2 Selection on inter-temporal feedback effects due to ecological inheritance

The feedback between local ecology and evolution. We find that the second term of the selection gradient eq. (13) can be written as

$$s_{\mathrm{e},i}(\boldsymbol{\theta}) = \sum_{j=1}^{S} \left.\frac{\partial \overline{n}_j}{\partial \tau_i}\right|_{\substack{\tau_i=\theta_i \\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} \times \left.\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial n_j}\right|_{\substack{\tau_i=\theta_i \\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} \tag{15}$$

(see Appendix B.3.5, eq. B.28 for details), where $\partial \overline{n}_j / \partial \tau_i$ is the effect of the mutation on the local abundance of species j experienced by a mutant of species i that is randomly sampled from its local lineage (i.e., the lineage of carriers of the mutant trait τ_i that reside in the focal patch in which the mutation first appeared). The second derivative in eq. (15) is the effect that this abundance change of species j has on the fitness of a focal individual of species i. By multiplying these two effects and summing them over all species j of the community, eq. (15) therefore captures how selection depends on the feedback between local community ecology and evolution.

A lineage-centred perspective on the ecological influence of a trait. The feedback effect captured by eq. (15) reveals that a phenotypic change will be selected when such a change results in local ecological conditions that are favourable for the lineage of those that express the change (i.e., when $\partial \overline{n}_j / \partial \tau_i \times \partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}}) / \partial n_j > 0$). This brings us to the question of what is the nature of the influence of a local lineage on its own ecology, which is captured by the derivative $\partial \overline{n}_j / \partial \tau_i$. We find that this derivative can be expressed as

$$\left.\frac{\partial \overline{n}_j}{\partial \tau_i}\right|_{\substack{\tau_i=\theta_i \\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} = \sum_{k=1}^{S} \Psi_{k,i}(\boldsymbol{\theta}) \sum_{h=1}^{\infty} \mathrm{K}_{jk,h}(\boldsymbol{\theta}) \left[\hat{n}_i(\boldsymbol{\theta}) r_{i,h}(\boldsymbol{\theta})\right] \tag{16}$$

(see Appendix B.3.5, eq. B.36 for details). In order to understand eq. (16), consider a focal individual from species i expressing the mutant allele, who lives at a demographic time that we arbitrarily label as time zero ($t = 0$, see blue star Fig. 3). The first term of eq. (16),

$$\Psi_{k,i}(\boldsymbol{\theta}) = \frac{1}{\hat{n}_i(\boldsymbol{\theta})} \left.\frac{\partial F_k(\bar{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \bar{\tau}_i}\right|_{\substack{\tau_i=\theta_i \\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} \tag{17}$$

is the effect that a trait change in the focal individual has on the abundance of species k in the focal patch at the next demographic time, i.e., at $t = 1$ (grey arrows, Fig. 3).

The second term in eq. (16), $\mathrm{K}_{jk,h}(\boldsymbol{\theta})$, is given by the (jk) element of the matrix

$$\mathbf{K}_h(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})^{h-1}, \tag{18}$$

where $\mathbf{C}(\boldsymbol{\theta})$ is the community matrix given in eq. (6). Eq. (18) reveals that $\mathrm{K}_{jk,h}(\boldsymbol{\theta})$ in eq. (16) is the effect that a change in the abundance of species k at time $t = 1$ has on the abundance of species j at time $t = h$. Importantly, this effect takes into account the influence that species have on one anothers' abundance, cumulated over $h-1$ demographic time periods (as indicated by the exponent $h-1$ in eq. 18, see green arrows in Fig. 3 for e.g.).

Finally, the term in square brackets in eq. (16) can be interpreted as the expected number of carriers of identical-by-descent copies of the mutant allele (number of "relatives" of the focal individual) that live in the focal patch at time $t = h \geq 1$ in the future (and which therefore experience the mutant-modified ecological conditions at time h, light blue disks, Fig. 3). Indeed, this term consists of the product of the equilibrium abundance, $\hat{n}_i(\boldsymbol{\theta})$, with

$$r_{i,h}(\boldsymbol{\theta}) = \left(1 - m_i(\boldsymbol{\theta})\right)^h \underbrace{\left(\frac{1}{\hat{n}_i(\boldsymbol{\theta})} + \frac{\hat{n}_i(\boldsymbol{\theta}) - 1}{\hat{n}_i(\boldsymbol{\theta})} r_i(\boldsymbol{\theta})\right)}_{\bar{r}_i(\boldsymbol{\theta})}, \tag{19}$$

which is the relatedness between two individuals of species i that are sampled h demographic time periods apart in the focal patch in the resident population (i.e., the probability that these two individuals share a common ancestor that resided in the focal patch, see also eq. B.35 for a formal definition). Note that in general, $m_i(\boldsymbol{\theta})$ is the backward probability of dispersal, which is defined as the probability that a randomly sampled individual of species i in the resident population is an immigrant. This will of course be influenced by dispersal behaviour, but also on other organismal aspects depending on the life-cycle (e.g., on adult survival from one time period to the next).

The above considerations (eqs. 17-18) show that the influence of a local lineage on the abundance of its own or another species j (eq. 16) can be intuitively understood as the effect of a trait change in a focal individual on the abundance of species j, which is experienced by all its downstream relatives residing in the focal patch (see Figure 3 for a diagram and Appendix B.3.5 for mathematical details).

Evolutionary press perturbations. In order to evaluate eq. (16) explicitly, we can use the fact that under our assumptions that patches are not totally isolated from one another (i.e., $m_i(\boldsymbol{\theta}) > 0$) and that the resident community is at a stable fixed point (i.e., $\mathbf{C}(\boldsymbol{\theta})$ has eigenvalues with absolute value less than one), the infinite sum in eq. (16) converges. This leads to the following expression (see Appendix B.3.5, eq. B.37 for details),

$$\left.\frac{\partial \overline{n}_j}{\partial \tau_i}\right|_{\substack{\tau_i=\theta_i \\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} = \sum_{k=1}^{S} \Psi_{k,i}(\boldsymbol{\theta})\hat{n}_i(\boldsymbol{\theta})\Lambda_{jk,i}(\boldsymbol{\theta})\left(1 - m_i(\boldsymbol{\theta})\right)\bar{r}_i(\boldsymbol{\theta}), \tag{20}$$

where $\bar{r}_i(\boldsymbol{\theta})$ is the relatedness between two individuals sampled with replacement in the same patch (see eq. 19 for definition), and $\Lambda_{jk,i}(\boldsymbol{\theta})$ is given by the (jk)-entry of the matrix

$$\Lambda_i(\boldsymbol{\theta}) = \left(\mathbf{I} - [1 - m_i(\boldsymbol{\theta})]\mathbf{C}(\boldsymbol{\theta})\right)^{-1}. \tag{21}$$

The term $\Lambda_{jk,i}(\boldsymbol{\theta})$ in eq. (20) captures the effect of a change in the abundance of species k on the abundance of species j, experienced by all individuals of species i descending from a single ancestor in the focal patch. Interestingly, as dispersal goes to zero in the focal species ($m_i(\boldsymbol{\theta}) \to 0$), the matrix $\Lambda_i(\boldsymbol{\theta})$ (eq. 21) tends to the matrix of press perturbations of community ecology [i.e., $\Lambda_i(\boldsymbol{\theta}) \to (\mathbf{I} - \mathbf{C}(\boldsymbol{\theta}))^{-1}$]. The entries of this matrix measure how a constant and persistent change in the abundance of one species influences the equilibrium abundance of another through multi-species interactions (e.g., Yodzis, 1989, Case, 2000). The correspondence between eq. (21) and press perturbation matrices reflects that as $m_i(\boldsymbol{\theta}) \to 0$, the mutant lineage may persist locally forever and thus experience persistent changes in the abundance of other species. But as dispersal $m_i(\boldsymbol{\theta})$ increases, the mutant lineage will spend fewer time periods locally, which means that its experience of changes in local species abundance will last fewer time periods (and so $\Lambda_i(\boldsymbol{\theta})$ approaches the identity matrix as dispersal becomes complete, i.e., $\Lambda_i(\boldsymbol{\theta}) \to \mathbf{I}$ as $m_i(\boldsymbol{\theta}) \to 1$).

4.2.3 Connections with previous results on selection gradients and ecological feedback

The selection gradient we have derived is closely connected to existing gradients in the literature. To see these connections, consider first the case when dispersal is complete ($m_i(\boldsymbol{\theta}) = 1$ so that $r_i(\boldsymbol{\theta}) = 0$). In this case, the selection gradient reduces to

$$s_i(\boldsymbol{\theta}) = \left.\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tau_{\bullet,i}}\right|_{\substack{\tau_i = \theta_i \\ \boldsymbol{n} = \hat{\boldsymbol{n}}}}, \tag{22}$$

which embodies the classical ecological feedback considered in evolutionary analyses (e.g., Michod, 1979, Charlesworth, 1994, and see in particular eq. 29 of Lion, 2017): the invasion of a rare mutant depends on resident-set ecological conditions only (i.e., on $\boldsymbol{\theta}$ and $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$ only), and if the mutant invades, it becomes the resident and thereby modifies these conditions. The simplicity of eq. (22) reflects that when dispersal is complete, a globally rare mutant is also always locally rare. As a consequence, the selection gradient depends only on the effect that a mutant carrier has on its own individual fitness.

When dispersal is limited ($m_i(\boldsymbol{\theta}) < 1$), however, a globally rare mutant may become locally common and remain so over multiple demographic time periods. This has two implications that are important for the way selection targets this mutant. First, mutants living in the same time period interact directly with one another. This effect is captured by the relatedness-weighted fitness effect of neighbours in $s_{\mathrm{w},i}(\boldsymbol{\theta})$ (i.e., the second summand of

eq. 14). In fact, $s_{\mathrm{w},i}(\boldsymbol{\theta})$ (eq. 14) is equivalent to the standard selection gradient in the island model with constant demography (Taylor and Frank, 1996, Frank, 1998, Rousset, 2004). But in contrast to the selection gradient under constant demography, abundance in $s_{\mathrm{w},i}(\boldsymbol{\theta})$ (eq. 14) is endogenously determined and evaluated at the resident ecological equilibrium, $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$. As such, $s_{\mathrm{w},i}(\boldsymbol{\theta})$ approximates the exact selection gradient in a demographically-structured population when the trait has no demographic effect (denoted as S_{f}, eq. 26 of Rousset and Ronce, 2004, eq. E-28 of Lehmann et al., 2016). The main difference between the approximation and the exact expression is that the latter depends on reproductive values while the approximation $s_{\mathrm{w},i}(\boldsymbol{\theta})$ (eq. 14) does not. This is because we ignore stochastic demographic fluctuations here, and thus do not need to consider fitness effects in all possible demographic states.

The second implication of limited dispersal for the way selection targets a mutant is that a rare mutant can modify the demographic/ecological conditions experienced by its own lineage. Put differently, mutants living at different time periods interact indirectly, through heritable ecological modifications. Selection due to these indirect interactions is captured by the second term of the selection gradient, $s_{\mathrm{e},i}(\boldsymbol{\theta})$ (eq. 13). With constant demography in the focal evolving species, the ecological inheritance term $s_{\mathrm{e},i}(\boldsymbol{\theta})$ is consistent with the selection gradient on inter-temporal within-species altruism (Lehmann, 2007, eq. 9, Sozou, 2009, eq. 4.11) and niche construction traits affecting the abundance of a local abiotic resource (Lehmann, 2008, eq. A39 in the absence of isolation-by-distance). With fluctuating demography, $s_{\mathrm{e},i}(\boldsymbol{\theta})$ (eq. 13) approximates the part of the exact selection gradient that captures selection on a trait due to its demographic effect on the focal species only (sometimes denoted as S_{Pr}, eq. 27 of Rousset and Ronce, 2004, eq. E-29 of Lehmann et al., 2016).

4.2.4 Summary

In summary, we have shown that the selection gradient on a trait, $s_i(\boldsymbol{\theta})$, depends on how a trait-change in a focal individual affects: (1) its own fitness and the fitness of its current relatives through intra-temporal interactions ($s_{\mathrm{w},i}(\boldsymbol{\theta})$, eq. 15); and (2) the fitness of its downstream relatives living in the focal patch through heritable modifications of the ecological environment ($s_{\mathrm{e},i}(\boldsymbol{\theta})$, eqs. 16-17). This reveals that under limited dispersal, selection on intra- and inter-specific interactions can generally be interpreted in terms of inter-temporal inclusive fitness effects, i.e., in terms of the effect that a trait change in a focal individual has on the fitness of this focal and of all its relatives (current and downstream). Such a perspective allows for an intuitive understanding of selection on ecological interactions that take place in dispersal limited communities. In particular, our approximation highlights the nature of inter-temporal effects and their roles in the moulding of functional traits. We illustrate more concretely the potential importance of inter-temporal effects when we apply our approximation to specific models in the next section.

5 Applications

Here, we use our approximation to study the evolution of two traits that underlie intra- and inter-specific interactions under limited dispersal. The first is the evolution of helping within species, which has received considerable attention. This will allow us to contextualise our approach to study intra-specific interactions, when such interactions influence demography. The second example is the evolution of harming between species, which has so far not been investigated under limited dispersal. Analytical calculations checks of these examples, as well as the codes for the associated individual-based simulations, are available in Supplemental Material: Mathematica Notebook.

5.1 Helping within species

5.1.1 The biological scenario

We focus on a single species and study the evolution of a social trait or behavior that increases the fitness of patch neighbours, but comes at a fitness cost to self. We consider the following life cycle. (1) Adults reproduce. A focal individual has mean fecundity $f(\tau_\bullet, \tilde{\tau})/(1+\gamma n)$, where $f(\tau_\bullet, \tilde{\tau})$ is its fecundity in the absence of density-dependent competition. The latter has intensity γ. Maximal fecundity $f(\tau_\bullet, \tilde{\tau})$ decreases with the level $\tau_\bullet$ of helping of the focal individual, but increases with the average level $\tilde{\tau}$ of helping among its neighbours in the focal patch ($\partial f(\tau_\bullet, \tilde{\tau})/\partial \tau_\bullet < 0$ and $\partial f(\tau_\bullet, \tilde{\tau})/\partial \tilde{\tau} > 0$). (2) All the adults die. (3) Each offspring independently disperses with a probability m. (4) All offspring survive to adulthood (i.e., no competition for space among offspring).

Our assumptions for the life-cycle can be biologically interpreted as individuals competing locally to acquire material resources, and that the transformation of these resources into offspring depends on the level of helping within the patch (for instance because individuals share resources).

5.1.2 Necessary components

We first specify the components necessary to compute the selection gradient (i.e., the terms that appear in eqs. 13–20). According to the life-cycle assumptions for the model of helping, the fitness of a focal individual that expresses a level of helping $\tau_\bullet$ in a patch of size n, when its average neighbour expresses level $\tilde{\tau}$, is

$$w(\tau_\bullet, \tilde{\tau}, n, \hat{n}(\theta)) = \frac{f(\tau_\bullet, \tilde{\tau})}{1+\gamma n}. \tag{23}$$

Note that here, fitness does not depend on species abundance in patches other than the focal ($\hat{n}(\theta)$). This is because we have assumed that competition occurs locally for material resources (see eq. C.1 in the appendix for an example of a fitness function that depends on $\hat{n}(\theta)$). Following the same argument used to derive eq. (7), we find that the population dynamic (i.e., the abundance in the focal patch after one iteration of the life-cycle, given that the average level of helping in the patch is $\bar{\tau}$, that abundance at the previous time period was n and that other patches are at equilibrium $\hat{n}(\theta)$) can be written as

$$F(\bar{\tau}, n, \hat{n}(\theta)) = (1-m)\left(\frac{f(\bar{\tau},\bar{\tau})}{1+\gamma n}\right) n + m\hat{n}(\theta). \tag{24}$$

The equilibrium population size $\hat{n}(\theta)$ in the resident population is found by solving $\hat{n} = F(\theta, \hat{n}, \hat{n})$ for $\hat{n}$, which yields

$$\hat{n}(\theta) = \frac{f(\theta,\theta)-1}{\gamma}. \tag{25}$$

This equilibrium population size further allows us to obtain the pairwise relatedness $r(\theta)$, which is given by substituting eq. (25) into eq. (10), i.e.,

$$r(\theta) = \frac{(1-m)^2}{1+\left[1-(1-m)^2\right]\left[(f(\theta,\theta)-1)/\gamma-1\right]}. \tag{26}$$

This shows that pairwise relatedness increases as intra-specific competition γ increases because this leads to smaller patch size (eq. 25). As expected, relatedness increases as dispersal becomes limited ($m \to 0$). From here, it is straightforward to obtain the other necessary relatedness coefficient, $\bar{r}(\theta)$ (see eq. 19 for definition).

5.1.3 Selection on helping

We now proceed to calculate the selection gradient on helping under our scenario. Note that the selection gradient on a single trait in a single species can be written as $s(\theta) = s_{\mathrm{w}}(\theta) + s_{\mathrm{e}}(\theta)$, where $s_{\mathrm{w}}(\theta)$ captures the intra-, and $s_{\mathrm{e}}(\theta)$, the inter-temporal effects.

Intra-temporal effects of helping. Let us first study selection on helping according to its intra-temporal effects (i.e., by looking at $s_{\mathrm{w}}(\theta)$, eq. 14). These effects can be expressed as

$$\begin{aligned} s_{\mathrm{w}}(\theta) &= \frac{\partial w(\tau_\bullet, \tilde{\tau}, n, \hat{n})}{\partial \tau_\bullet} + \frac{\partial w(\tau_\bullet, \tilde{\tau}, n, \hat{n})}{\partial \tilde{\tau}} r(\theta) \\ &= \frac{1}{f(\theta,\theta)}\left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}} r(\theta)\right), \end{aligned} \tag{27}$$

where we used eqs. (23) and (25). Note that since helping is individually costly but increases the fecundity of neighbours, the direct and indirect fitness effects of helping are negative and positive, respectively (i.e., $\partial w/\partial \tau_\bullet < 0$ and $\partial w/\partial \tilde{\tau} > 0$). Hence, the helping trait in our model is altruistic *sensu* evolutionary biology (e.g., Hamilton, 1964, Rousset, 2004, West and Gardner, 2010). Eq. (27) shows that altruistic helping is favoured by high relatedness. From the relatedness eq. (26), we therefore expect limited dispersal and intra-specific competition to favour the evolution of helping, owing to its intra-temporal effects. However, selection on helping also depends on its inter-temporal effects, which we investigate in the next paragraph.

Inter-temporal effects of helping. When a single species is under scrutiny, selection on inter-temporal effects (i.e., $s_{\mathrm{e}}(\theta)$, eq. 15) can be expressed as

$$s_{\mathrm{e}}(\theta) = \frac{\partial \overline{n}}{\partial \tau} \times \frac{\partial w(\tau_\bullet, \tilde{\tau}, n, \hat{n})}{\partial n}. \tag{28}$$

Using eq. (20), the effect of helping on the lineage-experienced equilibrium abundance can be written as

$$\begin{aligned} \frac{\partial \overline{n}}{\partial \tau} &= \frac{\partial F(\bar{\tau}, n, \hat{n})}{\partial \bar{\tau}} \left[1 - (1-m) \frac{\partial F(\bar{\tau}, n, \hat{n})}{\partial n} \right]^{-1} (1-m)\bar{r}(\theta) \\ &= r(\theta) \frac{[f(\theta,\theta)-1]/\gamma}{f(\theta,\theta)-(1-m)^2} \left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}} \right). \end{aligned} \tag{29}$$

We will assume that $\partial f(\tau_\bullet, \tilde{\tau})/\partial \tau_\bullet + \partial f(\tau_\bullet, \tilde{\tau})/\partial \tilde{\tau} > 0$, so that helping increases equilibrium abundance (i.e., $\partial \overline{n}/\partial \tau > 0$). In turn, this increase in abundance feedbacks negatively on the fitness of downstream individuals according to

$$\frac{\partial w(\tau_\bullet, \tilde{\tau}, n, \hat{n})}{\partial n} = -\frac{\gamma}{f(\theta,\theta)} < 0 \tag{30}$$

(using eq. 23). This is because greater abundance leads to stronger intra-specific competition (according to γ). As a result, the selective inter-temporal fitness effects of helping,

$$s_{\mathrm{e}}(\theta) = -\frac{r(\theta)}{f(\theta,\theta)} \underbrace{\left(\frac{f(\theta,\theta)-1}{f(\theta,\theta)-(1-m)^2} \right)}_{=\beta(\theta)>0} \left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}} \right) < 0, \tag{31}$$

are negative (found by substituting eqs. 29-30 into 28).

Balance between intra- and inter-temporal effects. Summing eqs. (27) and (31), we find that the selection gradient is proportional to

$$s(\theta) \propto \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \kappa(\theta) \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}}, \tag{32}$$

where

$$\kappa(\theta) = \frac{(1-m)^2}{\hat{n}(\theta) f(\theta,\theta) - (1-m)^2(\hat{n}(\theta)-1)} = r(\theta)\frac{1-\beta(\theta)}{1-r(\theta)\beta(\theta)}, \tag{33}$$

is a scaled relatedness coefficient, which decreases with dispersal (m, see Fig. 4 and eq. 31 for the definition of $\beta(\theta) > 0$). This scaled relatedness can be understood by looking at the right hand side of eq. (33). There, the relatedness coefficient $r(\theta)$ in the numerator reflects selection on helping due to its positive intra-temporal indirect effects (eq. 27). This positive effect, however, is discounted by a factor $[1-\beta(\theta)]/[1-r(\theta)\beta(\theta)] < 1$, due to the negative inter-temporal indirect effects of helping (eq. 31). Scaled relatedness coefficient $\kappa(\theta)$ thus reflect how selection on helping depends on the balance between the positive intra-temporal indirect effects of helping, and its negative inter-temporal indirect effects owing to increased competition.

To understand the balance between intra- and inter-temporal effects better, it is noteworthy that relatedness among individuals decreases with the number of generations that separate them (eq. 19). As a result, selection on fitness effects becomes increasingly weak over generations. This is reflected in the fact that scaled relatedness coefficient is non-negative (i.e., $\kappa(\theta) \geq 0$, eq. 33, Fig. 4). In fact, provided $r(\theta) > 0$ (so that $\kappa(\theta) > 0$), altruistic helping can evolve in our model. This can be seen more explicitly if we further assume that maximal fecundity is given by

$$f(\tau_\bullet, \tilde{\tau}) = f_\mathrm{b} \times (1 - C\tau_\bullet^2 + B\tilde{\tau}), \tag{34}$$

where f_b is a baseline fecundity, C is the cost of helping, which increases quadratically with the investment of the focal into helping, and B is the benefit of helping, which results from one unit invested into helping. Substituting eq. (34) into eqs. (32)-(33) and solving $s(\theta^*) = 0$ allows us to find the singular strategy θ^*. When both C and B are small (of the order order of a parameter $\epsilon \ll 1$), the singular strategy θ^* can be found by solving a first-order Taylor expansion of the selection gradient about $\epsilon = 0$. Doing so, we obtain a simple expression for the convergence stable strategy,

$$\theta^* = \frac{B}{2C}\kappa(0) = \frac{B}{2C} \times \frac{1}{1 + \hat{n}(0)\left(\frac{f_\mathrm{b}}{(1-m)^2} - 1\right)} \geq 0. \tag{35}$$

Eq. (35) makes it straightforward to see that helping can evolve in spite of its negative inter-temporal indirect effects. It further shows how the equilibrium level of helping, θ^*, decreases with dispersal m and local abundance in the absence of helping, $\hat{n}(0) = (f_\mathrm{b} - 1)/\gamma$.

More generally, by solving eqs. (32) and (33) with eq. (34) numerically, we find that predictions generated from our approximation fits qualitatively and quantitatively well with observations from individual-based simulations, as much for the value of the convergence stable level of helping θ^* (Fig. 5, top panel) as for the concomitant equilibrium group size $\hat{n}(\theta^*)$ this generates (Fig. 5, bottom panel).

5.1.4 Connections to previous results on altruism evolution

Our finding that altruism decreases as dispersal and local abundance increase is a standard result of evolutionary biology. However, our model of altruism and its results depart in two ways from the literature on this topic (e.g., Taylor, 1992, van Baalen and Rand, 1998, Taylor and Irwin, 2000, Gardner and West, 2006, Lehmann et al., 2006, El Mouden and Gardner, 2008, Lion and Gandon, 2009, 2010, Rodrigues and Gardner, 2012, Johnstone and Cant, 2008, Wild, 2011, Bao and Wild, 2012, Johnstone et al., 2012, Kuijper and Johnstone, 2012). First, the vast majority of previous analyses assumes that density-dependent competition occurs for space after dispersal (i.e., space or "open sites" is the limiting factor, e.g., Tilman, 1982, Chapter 8). In this case, intra-temporal kin competition effects strongly inhibit the benefits of interacting among relatives. Here, we have assumed that competition occurs for resources before dispersal. In this situation, we found that intra-temporal kin competition effects do not abate the selective advantage of interacting with relatives (this can be seen from eq. 27, which only depends on pairwise relatedness). Rather, by increasing abundance, altruism increases kin competition for future generations (eq. 31). This also hinders the evolution of altruism but only moderately so, because relatedness between individuals of different generations is on average lower than individuals of the same generation.

A second important difference between our and previous models of social evolution with endogenous patch dynamics is that the latter had to rely exclusively on numerical approaches to compute the selection gradient (in the island model of dispersal, e.g., Lehmann et al., 2006, Alizon and Taylor, 2008, Wild et al., 2009, Wild, 2011 for models of altruism evolution, and Metz and Gyllenberg, 2001, Cadet et al., 2003, Parvinen et al., 2003, Rousset and Ronce, 2004 for models of dispersal evolution). This reliance on numerical analysis makes it more difficult to understand how scaled relatedness κ and selection on altruism vary with demographic parameters (e.g., Lehmann et al., 2006, eq. 12). Here, our approximation yields a simple and intuitive expression for the selection gradient (eq. 32), which nonetheless fits well with simulation results (Fig. 5).

It is noteworthy that the selection gradient we have derived for this example (eq. 32) applies to any type of social interactions within the life-cycle given in 5.1.1. In fact, the selection gradient eq. (32) can be adjusted to study other social behaviours simply by changing the fecundity function (e.g., eq. 34). Such selection gradient written in terms of marginal fecundity effects of behaviour have also been derived for lattice structured populations using the pair approximation (Lion and Gandon, 2009, eq. 14, Lion and Gandon, 2010, eq. 19). Comparing these expressions with ours would be interesting, in particular to investigate the effects of isolation-by-distance (which are ignored here).

5.1.5 Coevolution of helping and dispersal

Our social evolution model assumes that dispersal is fixed. Dispersal, however, is likely to be an evolving trait. Because dispersal determines whether individuals interact and compete with relatives, dispersal evolution is important for selection on social behaviour (e.g., Le Galliard et al., 2005, Purcell et al., 2012, Mullon et al., 2018). Dispersal evolution can also influence demography, in particular when individuals compete for space (e.g., when offspring survival after dispersal depends on local abundance, Metz and Gyllenberg, 2001, Cadet et al., 2003, Parvinen et al., 2003, Rousset and Ronce, 2004).

In order to test whether our approximation could capture the interplay between social behaviour, dispersal and demography, we used it to study a model of the coevolution between altruistic helping and dispersal when offspring compete for space following dispersal. We assumed that dispersal is costly, with offspring surviving dispersal with a probability $s < 1$. Details on this model and analysis of selection are given in Appendix C.

We find that dispersal increases as survival during dispersal, s, increases (Fig. 6, top panel, grey curve). This in turn selects for lower levels of helping (Fig. 6, top panel, black curve), in line with previous models of helping-dispersal coevolution that assume that demography is constant (e.g., Mullon et al., 2018). Here, we further find that as survival during dispersal s increases, the resulting collapse in helping and increase in dispersal, leads to fewer individuals populating each patch (Fig. 6, bottom panel). The predictions derived from our approximation agree well with observations we made from individual-based simulations, for the equilibria of the two traits and the concomitant abundance these equilibria generate (Fig. 6). This supports that the approximate selection gradient (eqs. 13–16) can be used to model dispersal evolution, in particular when local demography, genetic structure, and social traits feedback on one another. Our approximation, however, cannot be used to investigate disruptive selection, which can emerge when helping and dispersal coevolve (e.g., Purcell et al., 2012, Mullon et al., 2018). Such an investigation would require studying the second-order effects of selection, which is beyond the scope of this paper.

5.2 Harming between species

5.2.1 The biological scenario

In order to illustrate how the selection gradient can be applied to study ecological interactions among species, we now model the evolution of antagonistic interactions among two species, species 1 and species 2. Specifically, we model the evolution of a trait in species 1 that is costly to express and that harms individuals of species 2. Our two species go through the following life-cycle. (1) Individuals reproduce. A focal individual of species 1 has

mean fecundity $f_1(\tau_{\bullet,1})/(1+\gamma n_1+\eta n_2)$, which decreases with intra- and inter-specific competition (respectively measured by parameters γ and η). The maximal fecundity of a focal individual of species 1, $f_1(\tau_{\bullet,1})$, decreases with its investment $\tau_{\bullet,1}$ into harming (i.e., $f_1'(\tau_{\bullet,1})<0$). A focal individual of species 2 has mean fecundity $f_2/(1+\gamma n_2)$, where f_2 is the maximal fecundity of species 2 and γ is the level of intra-specific competition. Note that only species 1 experiences inter-specific competition. This would occur, for instance, because species 2 is a generalist consumer while species 1 a specialist. (2) Adult individuals of species 1 kill offspring of species 2 in amount $D(\bar{\tau}_1)$ per capita, which increases with the average $\bar{\tau}_1$ level of harming in the focal patch (i.e., $D'(\bar{\tau}_1)>0$). Costly inter-specific harming could for instance occur through the release of chemicals into the environment that suppress the growth and establishment of offspring (i.e. through allelopathy, Lankau, 2008). (3) All adults of both species die. (4) Surviving offspring of species 1 and 2 disperse with probability m_1 and m_2, respectively. (5) All offspring survive to adulthood.

5.2.2 Necessary components

We first specify the components necessary for deriving the selection gradient on inter-specific harming. According to the above, a focal individual from species 1 that invests $\tau_{\bullet,1}$ into harming in a patch with $\boldsymbol{n}=(n_1,n_2)$ individuals of species 1 and 2 respectively, has fitness

$$w_1(\tau_{\bullet,1},\boldsymbol{n},\hat{\boldsymbol{n}}(\boldsymbol{\theta}))=\frac{f_1(\tau_{\bullet,1})}{1+\gamma n_1+\eta n_2}. \tag{36}$$

The abundances of both species in the focal patch after one iteration of the life-cycle, given that: (1) the average level of harming in the patch is $\bar{\tau}_1$; (2) the abundance at the previous generation was $\boldsymbol{n}=(n_1,n_2)$; and (3) other patches are at equilibrium $\hat{\boldsymbol{n}}(\boldsymbol{\theta})=(\hat{n}_1(\theta_1),\hat{n}_2(\theta_1))$, are given by

$$\begin{aligned}F_1(\bar{\tau}_1,\boldsymbol{n},\hat{\boldsymbol{n}}(\boldsymbol{\theta}))&=(1-m_1)\left(\frac{f_1(\bar{\tau}_1)}{1+\gamma n_1+\eta n_2}\right)n_1+m_1\hat{n}_1(\theta_1)\\F_2(\bar{\tau}_1,\boldsymbol{n},\hat{\boldsymbol{n}}(\boldsymbol{\theta}))&=(1-m_2)\left(\frac{f_2}{1+\gamma n_2}-D(\bar{\tau}_1)\frac{n_1}{n_2}\right)n_2+m_2\hat{n}_2(\theta_1).\end{aligned} \tag{37}$$

The resident ecological equilibrium, which is found by solving $\hat{\boldsymbol{n}}(\boldsymbol{\theta})=(F_1(\theta_1,\hat{\boldsymbol{n}}(\boldsymbol{\theta}),\hat{\boldsymbol{n}}(\boldsymbol{\theta})),F_2(\theta_1,\hat{\boldsymbol{n}}(\boldsymbol{\theta}),\hat{\boldsymbol{n}}(\boldsymbol{\theta})))$ simultaneously, is too complicated to be presented here for the general case. Note, however, that when the resident level θ_1 of harming is small, a first-order Taylor expansion of the resident ecological equilibrium around

$\theta_1 = 0$ gives

$$\begin{aligned}
\hat{n}_1(\theta_1) &= \hat{n}_1(0) + \frac{\theta_1}{\gamma}\left(\frac{f_2}{f_2-1}\hat{n}_1(0)\eta D'(0) - \underbrace{c_{\mathrm{h}}(0)}_{-f_1'(0)}\right) + O\left(\theta_1^2\right) \\
\hat{n}_2(\theta_1) &= \hat{n}_2(0) - \theta_1\left(\frac{f_2}{f_2-1}\hat{n}_1(0)D'(0)\right) + O\left(\theta_1^2\right),
\end{aligned} \tag{38}$$

where

$$\begin{aligned}
\hat{n}_1(0) &= \frac{f_1(0)-1}{\gamma} - \frac{\eta}{\gamma}\hat{n}_2(0) \\
\hat{n}_2(0) &= \frac{f_2-1}{\gamma},
\end{aligned} \tag{39}$$

are the abundances in the absence of harming (assuming that $D(0) = 0$). Eq. (38) reveals that harming of species 2 reduces its abundance (i.e., $\hat{n}_2(\theta_1) \leq \hat{n}_2(0)$). The abundance of species 1, $\hat{n}_1(\theta_1)$, depends on the balance between two opposite effects of harming. On one hand, by reducing the abundance of species 2, harming increases the abundance of species 1 due to inter-specific competition (this is captured by the first summand within brackets on the first line of eq. 38). On the other hand, abundance decreases with the cost of harming (which is captured by $c_{\mathrm{h}}(0) = -f_1'(0) > 0$ in eq. 38). The final component necessary to derive the selection gradient is the relatedness coefficient for species 1. It is given by eq. (10), with the resident ecological equilibrium for species 1, $\hat{n}_1(\theta_1)$, which is defined by eq. (37).

5.2.3 Selection on harming

Intra-temporal effects. Substituting eq. (36) into eq. (14), we obtain that selection on the intra-temporal effects of harming,

$$s_{\mathrm{w},1}(\theta_1) = -\frac{c_{\mathrm{h}}(\theta_1)}{f_1(\theta_1)} < 0, \tag{40}$$

where $c_{\mathrm{h}}(\theta_1) = -f_1'(\theta_1) > 0$, are always negative. This is because harming is intra-temporally costly to express at the individual level and does not provide any intra-temporal indirect fitness benefits. Hence, the only way for harming to evolve in this model is if this cost is compensated by future benefits received by downstream relatives, which we investigate below.

Inter-temporal effects. Selection on harming due to its effects on the fitness of downstream relatives is captured by the inter-temporal part of the selection gradient eq. (15). Substituting eqs. (36)-(38) into eqs. (15) and

(20), we find that selection on harming due to its inter-temporal effects is given by

$$s_{\mathrm{e},1}(\theta_1) = \frac{1-m_1}{f_1(\theta_1)}\left((1-m_1)\frac{c_{\mathrm{h}}(\theta_1)}{f_1(\theta_1)}\gamma + (1-m_2)D'(\theta_1)\eta\right) + O\left(e^4\right), \tag{41}$$

where e is such that $O(1-m_1) \sim O(1-m_2) \sim O(e)$. From eq. (41), we see that overall, inter-temporal effects favour the evolution of inter-specific harming (i.e., $s_{\mathrm{e},1}(\theta_1) > 0$ since $c_{\mathrm{h}}(\theta_1) = -f_1'(\theta_1) > 0$ and $D'(\theta_1) > 0$). This is due to two inter-temporal effects of harming. First, as captured by the first summand within brackets of eq. (41), harming benefits downstream relatives that remain philopatric because by paying the cost of harming, a focal individual produces less offspring and thus diminishes local intra-specific competition. Accordingly, this first effect scales with the tendency to remain philopatric in the harming species, $1-m_1$, the cost of harming, $c_{\mathrm{h}}(\theta_1)$, and the strength of intra-specific competition, γ. The second way that harming benefits downstream relatives is by reducing the local abundance of the harmed species (species 2, as captured by the second summand within brackets of eq. 41), which reduces inter-specific competition. In line with this, harming is favoured when the effect of harming, $D'(\theta_1)$, and the intensity of inter-specific competition, η, are large, and when dispersal is limited in both species (in species 1 to ensure that relatives benefit from the reduction of inter-specific competition, and in species 2 since otherwise, the local abundance of species 2 in downstream generations depends only on the process of immigration and not on local harming, eq. 37).

Convergence stable equilibrium of harming. To test explicitly the effect of limited dispersal on the evolution of harming, we assumed that the fecundity of an individual of species 1 that expresses a level $\tau_{\bullet,1}$ of harming is,

$$f_1(\tau_{\bullet,1}) = f_{1,\mathrm{b}} \times \left(1 - C\tau_{\bullet,1}^2\right) \tag{42}$$

where $f_{1,\mathrm{b}}$ is a baseline fecundity in species 1 and C is the individual cost of harming. We further assumed that an individual of species 2 that is in a patch in which the average harming level is $\bar{\tau}_1$ suffers a fecundity cost given by,

$$D(\bar{\tau}_1) = \alpha\bar{\tau}_1, \tag{43}$$

where α is a parameter tuning the deleteriousness of harming. The convergence stable level of harming, which is found by solving $s_1(\theta_1^*) = s_{\mathrm{w},1}(\theta_1^*) + s_{\mathrm{e},1}(\theta_1^*) = 0$ for θ_1^*, are shown in Figure 7 as a function of dispersal.

In line with eq. (41), we find that individually costly harming does not evolve when dispersal is complete (Figure 7, top and middle panel). This is because in that case, downstream relatives can never benefit from a decrease of inter-specific competition owing to harming. As dispersal becomes limited in both species, this inter-temporal benefit increasingly goes to relatives so that harming evolves (Figure 7, top and middle panel). This

evolution, in turn, causes a significant reduction in the abundance of species 2 and an increase of species 1 (Figure 7, bottom panel). These results were confirmed using individual-based simulations, further supporting the goodness of fit of our approximation (Figure 7, middle and bottom panel).

6 Discussion

Due to the physical limitations of movement, a community of species is typically structured in space to form a metacommunity (e.g., Tilman, 1982, Clobert et al., 2001, Urban et al., 2008, Leibold and Chase, 2017). Understanding selection in such a metacommunity is challenging due to the feedback between local ecology and trait composition, which emerges when dispersal is limited and local demography is stochastic. In order to better understand these eco-evolutionary dynamics, we here derived an approximation for the selection gradient on a quantitative trait that influences local ecology in the island model of dispersal.

The basis of our approximation is to neglect ecological stochasticity and to assume that the resulting deterministic ecological dynamics have a single fixed point (i.e., we do not consider periodic or chaotic dynamics). We nonetheless take into account the consequences of genetic stochasticity for selection. We found that this approximation works well qualitatively for all models and conditions we studied. We further found that it is quantitatively accurate in predicting ecological and evolutionary dynamics as long as dispersal is not excessively weak. As a rule of thumb, effective dispersal rate should be no less than 0.1 when patches are small, with fewer than ten individuals (Fig. 1 and Figs. 5-7). Such demographic regime leads to an F_{ST} well within the range of F_{ST} values that have been estimated across a wide spectrum of taxa (when $m_i(\boldsymbol{\theta}) = 0.1$ and $\hat{n}_i(\boldsymbol{\theta}) = 10$, eq. 10 gives $F_{\mathrm{ST}} = 0.30$ for haploids; for diploids, eq. 6.23 of Hartl and Clark, 2007, gives $F_{\mathrm{ST}} = 0.20$; equivalently, this regime entails one migrant per demographic time period, i.e., $\hat{n}_i(\boldsymbol{\theta})m_i(\boldsymbol{\theta}) = 1$; for empirical estimates, see Barton, 2001, p. 334; Hartl and Clark, 2007, p. 302). This suggests that our approximation takes into account dispersal levels that are relevant to many species (Bohonak, 1999).

The simplicity of our approximate selection gradient allows to investigate convergence stable species coalitions, and to intuitively understand community evolution under limited dispersal. In particular, our selection gradient reveals that selection can be decomposed into intra- (eq. 14) and inter-temporal effects (eqs. 15–16). Inter-temporal effects reflect the interaction between kin selection and local eco-evolutionary dynamics. This interaction can be understood by considering that when a focal individual perturbs species abundance locally, this perturbation leads to changes in community composition in the future due to ecological interactions (eq. 16, Figure 3). These changes then feedback on the fitness individuals living in the future, who potentially carry genes that are identical-by-descent to the focal (i.e., who are relatives). In other words, inter-temporal effects emerge

because individuals inherit not only their genes, but also an ecological environment that has been transformed by their ancestors (Odling-Smee et al., 2003, Bonduriansky, 2012). By considering the effects of such ecological inheritance on multi-species interactions, our model generalises previous models of local ecological interactions in the presence of relatives that ignored trait-induced changes in abundance, either altogether (Frank, 1994, Foster and Wenseleers, 2006, Wyatt et al., 2013, Akçay, 2017) or in the evolving species (Lehmann, 2008).

Interestingly, the eco-evolutionary, inter-temporal, feedbacks that emerge in our model are captured mathematically by analogues of press perturbations (eqs. 20-21). These are central notions of classical community ecology. Press perturbations traditionally measure how a persistent change in the abundance of a given species alters the equilibrium abundance of another due to ecological interactions (Yodzis, 1989, Case, 2000). Here, the change in abundance is initiated by a phenotypic change in a focal individual, and its persistence is measured over the time it is evolutionarily significant from the perspective of the focal, which is the time the focal's lineage of relatives experiences it (Figure 3). Because this time increases as dispersal becomes more limited, inter-temporal effects are more important for selection when individuals remain in philopatry (i.e., when dispersal is limited). This can also be understood from a gene-centred perspective: because limited dispersal ties in the fate of a trait-changing mutation with its inter-temporal ecological effects, these effects become more important for how selection targets this mutation.

We applied our model to the evolution of two behavioural traits with demographic and ecological consequences. First, we studied the evolution of altruistic helping within species. Our model follows a rich literature on this topic (see section "Connections to previous results on altruism evolution"), which traditionally assumes that local patch size remains constant and that competition occurs for space after dispersal. In this case, the benefits of helping relatives are offset or partially offset by the intra-temporal cost of kin competition (e.g., Taylor, 1992, Gardner and West, 2006, Lion and Gandon, 2009). By contrast, we assumed that competition occurs for material resources before dispersal and that helping can influence patch size. In this case, the costs of kin competition are delayed and paid by downstream relatives. Because these inter-temporal costs are weaker than intra-temporal ones, we found that selection favours intra-specific altruism for a large range of parameters, in particular when dispersal is limited (Figure 5). Our results therefore suggest that altruistic traits are more likely to be found in species in which competition occurs for material resources rather than for space.

Second, we studied the evolution of individually costly harming between species. We found that harming evolution strongly depends on inter-temporal effects. Specifically, we found that harming evolves when it alleviates inter-specific competition for downstream relatives by reducing the other species' abundance, which requires that dispersal is limited in both interacting species. Our analysis thus makes the empirical prediction that antagonism is more likely when dispersal is limited (Figure 7). Previous theory has focused on understanding how

"altruism" or mutualism between species can evolve in the presence of relatives (Frank, 1994, Foster and Wenseleers, 2006, Wyatt et al., 2013, Akçay, 2017). These studies have highlighted that the evolution of mutualism is facilitated by among-species genetic correlations, which emerge when dispersal is limited. Here, our model reveals that antagonism between species can also evolve in this case, which raises the interesting question of whether mutualistic or antagonistic interactions are more likely to evolve under limited dispersal. Presumably, this would depend on the degree of inter-specific competition.

Beyond the examples presented here, our analysis helps identify the conditions under which selection on a trait depends on inter-temporal effects arising from that trait changing the demographic dynamics of other species. First, when expressed in a focal individual, a change in this trait should have a large influence on the local community over one demographic time period (i.e., $\Psi_{k,i}(\boldsymbol{\theta})$ should be large, eq. 17). We expect this to be the case for traits that are directly involved in inter-specific interactions – such as defences against predators, traits that attract mutualists, or resource extraction strategies – especially when expressed by keystone or dominant species (i.e., species with large effects on communities). Our analysis further reveals that there are more opportunities for selection on inter-temporal effects when the local abundances of the different species that are part of the community are inter-dependent (so that: (1) fitness in the focal species depends on community composition, $\partial w_i/\partial n_j \neq 0$; (2) the community matrix $\mathbf{C}(\boldsymbol{\theta})$ is non-sparse; and (3) evolutionary press perturbations $\Lambda_i(\boldsymbol{\theta})$ are large, eq. 21). This ensures that the ecological perturbation initiated by a focal individual has multiple downstream effects through indirect ecological interactions (e.g., focal trait in species 1 increases the abundance of species 2, who is a competitor of species 3, who itself is a competitor of species 1, e.g., terHorst et al., 2018). Multiple downstream effects of a perturbation by a focal individual then increase the likelihood that this perturbation feedbacks on the fitness of downstream relatives of the focal. This leads to the broad prediction that communities that are tightly inter-connected are more likely to show traits whose inter-temporal ecological effects are under selection.

One crucial condition for inter-temporal ecological effects to be under selection is that dispersal is limited. This needs to be the case in the focal species to ensure that relatives experience trait-driven ecological changes, but also in other species of the community so that the effect of local interactions on abundance is not swamped by immigration. In our example of harming between species, for instance, we found that harming did not evolve if either the harming or recipient species showed full dispersal (Fig. 7, top). Plant communities would be ideal to test the notion that traits within spatially-limited communities are more likely to have inter-temporal effects that have been shaped by selection. For instance, many plants are engaged in inter-specific chemical warfare, with lasting effects on soil composition (Inderjit et al., 2011, for review). In the light of our results, it would be interesting to study how these inter-temporal effects of allelopathy vary with the degree of dispersal (or gene flow, which can be estimated from Fst values). In particular, we expect allelopathy to be most adapted among

competitors that show inter-specific genetic correlations.

Of course, the approximate selection gradient derived here cannot be applied to all evolutionary scenarios. It should generally be supplemented with simulation checks, in particular when dispersal is severely limited and patches are very small (like when populations are close to extinction). In fact, it would be useful to analyse our model with greater mathematical rigour in order to obtain a sharper understanding of the conditions under which ecological stochasticity can be neglected (for instance by generalizing the results of Chesson, 1981). One major limitation to our approach is that it relies on the assumption that ecological dynamics converge to a fixed point. This assumption, which allowed us to improve the understanding of selection on traits affecting meta-community stochastic demography, precludes the consideration of limit cycles or spatio-temporal fluctuations in abundance, which are thought to be prevalent in many ecological systems (e.g., Yodzis, 1989, Case, 2000). It would therefore be very relevant to extend our approach to derive the selection gradient under more complicated ecological dynamics. Another assumption we have made is that reproduction occurs as a discrete time process. It would thus be relevant to derive the selection gradient under continuous time, but this is unlikely to change our main qualitative results (as this essentially requires replacing sums by integrals, individual fitness by individual growth rates, and calculating inter-temporal relatedness coefficients in continuous time, e.g., Sozou, 2009).

To conclude, our heuristic approximation is a step further towards the integration of multi-species ecological theory and kin selection theory. Owing to its simplicity and intuitive interpretation, the approximate selection gradient we have derived can provide a useful guide to answer questions that lie at the intersection of ecology and evolution. In particular, it can be straightforwardly applied to study plant-pollinator, host-parasite or predator-prey coevolution under limited dispersal, or the eco-evolutionary dynamics of sex-specific dispersal. These and other applications should help understand how selection moulds intra- and inter-specific interactions when dispersal is limited.

Acknowledgements

We thank James Rodger for a useful discussion on dispersal evolution. We also thank Scott Nuismer, Minus Van Baalen and an anonymous reviewer for helpful comments on a previous version of this paper. CM is funded by Swiss NSF grant PP00P3-123344 to LL.

Appendix A: Individual-based simulations

Here, we describe how we implemented the individual-based simulations for the life cycle described in section 2.1 (using Mathematica 10.0.1.0, Wolfram Research, 2016, the code can be found in Supplemental Material: Mathematica Notebook).

The simulated population is divided among a finite number $n_{\mathrm{d}} = 1000$ of patches (instead of an infinite number of patches like in the analytical model). At the beginning of a generation, each individual j of species i is characterized by its phenotype θ_{ji} taken from the type space Θ_i. Calculations proceed as follows. (1) We first calculate the mean fecundity of each individual according to its phenotype and possibly that of its neighbours (depending on the model of interaction). Each individual then produces a Poisson distributed number of offspring according to its mean fecundity. (2) Each offspring disperses according to a Bernoulli trial (whose mean may depend on phenotype if dispersal evolves). During dispersal, death occurs according to a Bernoulli trial (whose mean may depend on evolving phenotypes). If an offspring disperses and survives dispersal, it is allocated to a non-natal patch according to discrete uniform distribution. (3) If local density-dependent competition takes place, each offspring survives according to a Bernoulli trial with mean depending on the number of individuals entering competition in the patch. (4) Mutation occurs in each offpsring with probability μ (Bernouilli trial). If no mutation occurs, an offspring has the same phenotypic values as its parent. If a mutations occurs, we add a perturbation to the parental phenotypic value that is sampled from a multivariate Normal distribution with mean zero for each trait, variance σ^2 and no covariance among traits. The resulting phenotypic values are controlled to remain in a given range if necessary.

The model tracks the phenotype(s) θ_{ji} of each individual of all species, and the number of individuals in each patch. From these data, we can evaluate various statistics such as mean patch size, average trait value, and relatedness. Relatedness under the neutral model (i.e., when vital rates are the same within each species, see Fig. 2) is calculated as the ratio of average phenotypic covariance among individuals of the same patch (averaged over all patches) to the total phenotypic variance in the population. This gives the probability of identity-by-descent within patches (Frank, 1998, eqs. 3.9-3.10).

Appendix B: Evolutionary analysis

In this appendix, we derive the approximate selection gradient on a mutant τ_i in species i (eqs. 13–21 of the main text).

B.1 Local evolutionary invasion analysis

Under the full stochastic model (life cycle described in section 2.1), the selection gradient $\mathcal{S}_i(\boldsymbol{\theta})$ is equal to the sensitivity of the invasion fitness $\rho_i(\tau_i, \boldsymbol{\theta})$ of a mutant allele with phenotype $\tau_i \in \mathbb{R}$ in species i in a resident population $\boldsymbol{\theta} \in \mathbb{R}$; namely

$$\mathcal{S}_i(\boldsymbol{\theta}) = \left.\frac{\partial \rho_i(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i = \theta_i} \tag{B.1}$$

(Lehmann et al., 2016, Box 2). Invasion fitness $\rho_i(\tau_i, \boldsymbol{\theta})$ is defined as the per capita number of mutant copies produced asymptotically over a time step of the reproductive process by a whole mutant lineage descending from a single initial mutant (and when the mutant is rare in the population, i.e., $\rho_i(\tau_i, \boldsymbol{\theta})$ is the geometric growth rate of the mutant, Cohen, 1979, Tuljapurkar, 1989, Caswell, 2000, Tuljapurkar et al., 2003).

Instead of using invasion fitness, $\rho_i(\tau_i, \boldsymbol{\theta})$, which is complicated to manipulate, we will use an invasion fitness proxy (i.e., a quantity that is sign equivalent to invasion fitness, e.g., Stearns, 1992, Charlesworth, 1994, Case, 2000, Metz and Gyllenberg, 2001, Lehmann et al., 2016). In the homogeneous island model (i.e., no exogenous differences among islands), one invasion fitness proxy is the basic reproductive number, $R_{0,i}(\tau_i, \boldsymbol{\theta})$, which is the expected number of successful offspring produced by a randomly sampled individual from a local mutant lineage of species i during the (expected) sojourn time of this lineage in a single patch (and when the mutant is rare in the population, see Lehmann et al., 2016, eq. 14 for a formal definition of $R_{0,i}(\tau_i, \boldsymbol{\theta})$).

To see that the selection gradient can be inferred from $R_{0,i}(\tau_i, \boldsymbol{\theta})$, consider a Taylor expansion of $\rho_i(\tau_i, \boldsymbol{\theta})$ and $R_{0,i}(\tau_i, \boldsymbol{\theta})$ around $\delta_i = \tau_i - \theta_i = 0$,

$$\begin{aligned} \rho_i(\tau_i, \boldsymbol{\theta}) &= 1 + \delta_i \left.\frac{\partial \rho_i(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i = \theta_i} + O_\rho\left(\delta_i^2\right) \\ R_{0,i}(\tau_i, \boldsymbol{\theta}) &= 1 + \delta_i \left.\frac{\partial R_{0,i}(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i = \theta_i} + O_{R_0}\left(\delta_i^2\right), \end{aligned} \tag{B.2}$$

where $O_\rho\left(\delta_i^2\right)$ and $O_{R_0}\left(\delta_i^2\right)$ are second-order remainders. Since $\rho_i(\tau_i, \boldsymbol{\theta})$ and $R_{0,i}(\tau_i, \boldsymbol{\theta})$ are sign equivalent around one (i.e., $\rho_i(\tau_i, \boldsymbol{\theta}) \leq 1 \iff R_{0,i}(\tau_i, \boldsymbol{\theta}) \leq 1$, Lehmann et al., 2016, eq. 15), eq. (B.2) reveals that the sensitivities of $\rho_i(\tau_i, \boldsymbol{\theta})$ and $R_{0,i}(\tau_i, \boldsymbol{\theta})$, are proportional, i.e.,

$$\mathcal{S}_i(\boldsymbol{\theta}) = \left.\frac{\partial \rho_i(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i = \theta_i} = \epsilon(\boldsymbol{\theta}) \left.\frac{\partial R_{0,i}(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i = \theta_i}, \tag{B.3}$$

where $\epsilon(\boldsymbol{\theta}) > 0$ is a positive factor of proportionality that depends only on the resident trait $\boldsymbol{\theta}$ (since the left and right hand side of eq. B.3 must only depend on $\boldsymbol{\theta}$). The sensitivity of $R_{0,i}(\tau_i, \boldsymbol{\theta})$ is therefore sufficient to determine the sign of the selection gradient, in line with previous results derived for class-structured population in continuous time (Metz and de Kovel, 2013, differentiating their eq. 2.4 produces our eq. B.3, see also Metz

and Leimar, 2011, p.175; but note that these authors use the average lifespan of an individual as a timescale to evaluate fitness, whereas we use one life cycle iteration).

B.2 Basic reproductive number

We now specify the basic reproductive number for our stochastic metacommunity model (life cycle described in section "metacommunity structure"). To this aim, we first introduce some notation. We let $M_{i,t} \in \{0,1,2,...,\mathcal{N}_{i,t}\} \subset \mathbb{Z}_+$ be the random variable for the number of mutant individuals of species i at demographic time period $t = 0,1,2,...$ in the focal patch (assuming that the mutant arose as a single copy in that patch a time $t = 0$ and is overall rare). Here, $\mathcal{N}_{i,t}$ is the random variable for the local abundance of species i in the focal patch at time t. We collect these variables for all species in the vector $\boldsymbol{\mathcal{N}}_{i,t} \subset \mathbb{Z}_+^S$. We assume that this community vector is bounded (i.e., $\mathcal{N}_{i,t} < \infty$ for all i). With this and owing to our life cycle assumptions, the sequence of couples $\{(M_{i,t}, \boldsymbol{\mathcal{N}}_{i,t})\}_{t\geq 0}$ is a countable Markov chain (Meyn and Tweedie, 2009) with value $(M_{i,t}, \boldsymbol{\mathcal{N}}_{i,t})$ at time t taken from the state space $\Omega_i = \mathbb{Z}_+ \times \mathbb{Z}_+^S$, which is the state space of the genetic-ecological process in the focal patch at time t. In an abuse of notation, we denote by $\boldsymbol{\omega}_t = (M_{i,t}, \boldsymbol{\mathcal{N}}_{i,t}) \in \Omega_i$ the realisation of state $(M_{i,t}, \boldsymbol{\mathcal{N}}_{i,t})$ at time t, and $\Pr(\boldsymbol{\omega}_t)$ as the probability of this realised genetic-ecological state.

Next, we observe that in each generation, $M_{i,t}$, has a single absorbing state: the local extinction of the lineage ($M_{i,t} = 0$). This stems from two of our assumptions: (1) that all offspring have a non-zero dispersal probability; and (2) the mutant is overall rare so there can be no mutant immigration back into the focal patch (e.g., Metz and Gyllenberg, 2001, Lehmann et al., 2016). Hence, the lineage of the mutant will eventually go extinct locally, i.e., $\lim_{t\to\infty} \Pr(M_{i,t} = 0) = 1$.

With the above definitions, we can write the basic reproductive number for our metacommunity model as

$$R_{0,i}(\tau_i, \boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} w_i(\tau_i, \boldsymbol{\omega}_t) q_i(\boldsymbol{\omega}_t), \tag{B.4}$$

where $w_i(\tau_i, \boldsymbol{\omega}_t)$ is the individual fitness of an individual (i.e., the expected number of successful offspring produced by this individual over one life cycle iteration, possibly including itself through survival) of species i that carries mutant phenotype τ_i, when the patch is in a genetic-ecological state $\boldsymbol{\omega}_t$ at time t. Further,

$$q_i(\boldsymbol{\omega}_t) = \frac{M_{i,t} \Pr(\boldsymbol{\omega}_t)}{N_{\mathrm{L}}(\tau_i, \boldsymbol{\theta})}, \tag{B.5}$$

is defined as the probability that a randomly sampled mutant lineage member is sampled exactly at time t on realisation $\boldsymbol{\omega}_t$ of the stochastic process ($N_{\mathrm{L}}(\tau, \theta)$ is the expected total size of the mutant lineage over its lifetime

in a single patch: $N_{\mathrm{L}}(\tau_i, \boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} M_{i,t} \Pr(\boldsymbol{\omega}_t))$. Note that owing to the fact that extinction of the mutant lineage in the focal patch is certain ($\lim_{t\to\infty} \Pr(M_{i,t} = 0) = 1$), the infinite sum in eq. (B.4) converges.

B.3 Approximate selection gradient

Using eq. (B.4), we now approximate heuristically the selection gradient (eq. B.3) by neglecting ecological stochasticity; i.e., assuming that the dynamics in local species abundance in a monomorphic population are deterministic (given by the recursion eq. 4) and have a stable fixed point $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$ (given by eq. 5), and that the change in abundance over one time step in a mutant-resident population is deterministic given the genetic-ecological state at the previous time step (see eq. 11).

B.3.1 Mutant effects on the distribution of genetic-ecological states cancel out

We start by noting that under the assumption that local abundance dynamics in a monomorphic population are deterministic and converge to the fixed point $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$, all patches other than the focal one are at the resident ecological equilibrium, $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$. We can therefore write the fitness of a focal individual in the focal patch as

$$w_i(\tau_i, \boldsymbol{\omega}_t) = w_i\left(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \boldsymbol{\mathcal{N}}_{i,t}, \hat{\boldsymbol{n}}(\boldsymbol{\theta})\right), \tag{B.6}$$

a function

$$w_i : \mathbb{R} \times \mathbb{R}^S \times \mathbb{R}_+^S \times \mathbb{R}_+^S \to \mathbb{R}_+, \tag{B.7}$$

of four arguments, (1) the phenotype τ_i of the focal individual; (2) the vector, $\tilde{\boldsymbol{\tau}}_{i,t} = (\theta_1, ..., \theta_{i-1}, \tilde{\tau}_{i,t}, \theta_{i+1}, ..., \theta_S)$, of average phenotypic values among the neighbours of the focal individual; (3) the realised local abundance vector $\boldsymbol{\mathcal{N}}_{i,t}$ (which from now on is assumed to be taken from the set $\mathbb{R}_+^S$, in order to allow us later to evaluate fitness at the resident equilibrium $\hat{\boldsymbol{n}}(\boldsymbol{\theta}) \in \mathbb{R}_+^S$ and to take the derivative of fitness with respect to local abundance); and (4) the abundance in patches other than the focal, which are at resident ecological equilibrium, $\hat{\boldsymbol{n}}(\boldsymbol{\theta})$. In the vector $\tilde{\boldsymbol{\tau}}_{i,t}$, the average phenotype among the neighbours of the focal individual of the same species i at time t is given by

$$\tilde{\tau}_{i,t} = \frac{\mathcal{N}_{i,t}\bar{\tau}_{i,t} - \tau_i}{\mathcal{N}_{i,t} - 1}, \tag{B.8}$$

i.e., the average phenotype in species i excluding the focal, with

$$\bar{\tau}_{i,t} = \frac{M_{i,t}}{\mathcal{N}_{i,t}}\tau_i + \left(1 - \frac{M_{i,t}}{\mathcal{N}_{i,t}}\right)\theta_i \tag{B.9}$$

being the average phenotype on the focal patch including the focal individual.

Substituting eq. (B.6) into eq. (B.4) and using the chain rule we get

$$\left.\frac{\partial R_{0,i}(\tau_i,\boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i=\theta_i} = \sum_{t=0}^{\infty}\sum_{\boldsymbol{\omega}_t\in\Omega_i}\left[\frac{\partial w_i\left(\tau_i,\tilde{\boldsymbol{\tau}}_{i,t},\boldsymbol{\mathcal{N}}_{i,t},\hat{\boldsymbol{n}}\right)}{\partial \tau_i} q_i^{\circ}(\boldsymbol{\omega}_t) + w_i(\theta_i,\boldsymbol{\theta},\hat{\boldsymbol{n}},\hat{\boldsymbol{n}})\frac{\partial q_{i,t}(\boldsymbol{\omega}_t)}{\partial \tau_i}\right], \tag{B.10}$$

where here and throughout all derivatives are evaluated at the resident phenotypic value $\boldsymbol{\theta}$ and the deterministic ecological equilibrium $\hat{\boldsymbol{n}} = \hat{\boldsymbol{n}}(\boldsymbol{\theta})$ (for which we drop the explicit dependency on $\boldsymbol{\theta}$ for the sake of clarity in the forthcoming equations). We use $q_i^{\circ}(\boldsymbol{\omega}_t)$ to denote the genetic-ecological state distribution in the resident population (i.e., in a monomorphic $\boldsymbol{\theta}$ population).

The second summand of eq. (B.10), which captures the mutant effects on the distribution of genetic-ecological states in fact, vanishes. This follows from the observation that,

$$\sum_{\boldsymbol{\omega}_t\in\Omega_i}\frac{\partial q_i(\boldsymbol{\omega}_t)}{\partial \tau_i} = \frac{\partial}{\partial \tau_i}\left[\sum_{\boldsymbol{\omega}_t\in\Omega_i} q_i(\boldsymbol{\omega}_t)\right] = \frac{\partial}{\partial \tau_i}(1) = 0, \tag{B.11}$$

since $q_i(\boldsymbol{\omega}_t)$ is a probability density function over Ω_i, whereby we obtain

$$\left.\frac{\partial R_{0,i}(\tau_i,\boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i=\theta_i} = \sum_{t=0}^{\infty}\sum_{\boldsymbol{\omega}_t\in\Omega_i}\frac{\partial w_i\left(\tau_i,\tilde{\boldsymbol{\tau}}_{i,t},\boldsymbol{\mathcal{N}}_{i,t},\hat{\boldsymbol{n}}\right)}{\partial \tau_i} q_i^{\circ}(\boldsymbol{\omega}_t). \tag{B.12}$$

B.3.2 Mutant effects on ecologically induced stochasticity are negligible

Here, we further simplify the fitness derivative appearing in eq. (B.12). To do this, we express the fitness function as the Taylor series

$$w_i(\tau_i,\tilde{\boldsymbol{\tau}}_{i,t},\boldsymbol{\mathcal{N}}_{i,t},\hat{\boldsymbol{n}}) = w_i(\tau_i,\tilde{\boldsymbol{\tau}}_{i,t},\boldsymbol{n}_t(\boldsymbol{\tau}_i),\hat{\boldsymbol{n}}) + \sum_{j=1}^{S}\sum_{k=1}^{\infty}\frac{\left(\mathcal{N}_{j,t}-n_{j,t}(\boldsymbol{\tau}_i)\right)^k}{k!}\frac{\partial^k w_i(\tau_i,\tilde{\boldsymbol{\tau}}_{i,t},\boldsymbol{n},\hat{\boldsymbol{n}})}{(\partial n_j)^k}, \tag{B.13}$$

around the community vector $\boldsymbol{n}_t(\boldsymbol{\tau}_i)$, which has entry j determined by

$$n_{j,t}(\boldsymbol{\tau}_i) = F_j\left(\bar{\boldsymbol{\tau}}_{i,t-1},\boldsymbol{n}_{t-1}(\boldsymbol{\tau}_i),\hat{\boldsymbol{n}}(\boldsymbol{\theta})\right) \tag{B.14}$$

(eq. 11 of the main text).

There are two important points to keep in mind about the community vector $n_{j,t}(\boldsymbol{\tau}_i)$ as given by eq. (B.14). The first is that the trajectory $\boldsymbol{n}_t(\boldsymbol{\tau}_i)$ (for $t = 0,1,\ldots$) is stochastic. This is because the abundance at time step t depends on the average trait value at the previous time step, $\bar{\boldsymbol{\tau}}_{i,t-1}$ (the first argument of F_j, eq. B.14), which

itself depends on the random number of mutants in the patch $M_{i,t-1}$ (eq. B.9). Hence, even if the change in abundance over one time step given the genetic-ecological state at the previous time step (i.e., given $\bar{\boldsymbol{\tau}}_{i,t-1}$ and $\boldsymbol{n}_{t-1}(\boldsymbol{\tau}_i)$) is deterministic, genetic stochasticity generate stochasticity in abundance. Note that genetic stochasticity is the only source of stochasticity in abundance under our assumption that the abundance dynamics in a monomorphic population are deterministic. Indeed, eq. (B.14) is a deterministic recursion when $\bar{\boldsymbol{\tau}}_{i,t-1}$ is held fixed.

The second important point to keep in mind about the community vector $n_{j,t}(\boldsymbol{\tau}_i)$ is that it can be treated as a function of the mutant phenotype, which is captured by $\boldsymbol{\tau}_i$. To understand this dependency, consider from eq. (B.14) that $n_{j,t}(\boldsymbol{\tau}_i)$ nests the mutant effects since its initial appearance in the focal patch at $t = 0$, i.e., we can write eq. (B.14) equivalently as

$$\underbrace{G_{j,t}\left(\bar{\boldsymbol{\tau}}_{i,0}, \bar{\boldsymbol{\tau}}_{i,1}, ..., \bar{\boldsymbol{\tau}}_{i,t-1}, \hat{\boldsymbol{n}}(\boldsymbol{\theta})\right)}_{\equiv n_{j,t}(\boldsymbol{\tau}_i)} = F_j\left(\bar{\boldsymbol{\tau}}_{i,t-1}, \underbrace{\boldsymbol{G}_{t-1}\left(\bar{\boldsymbol{\tau}}_{i,0}, \bar{\boldsymbol{\tau}}_{i,1}, ..., \bar{\boldsymbol{\tau}}_{i,t-2}, \hat{\boldsymbol{n}}(\boldsymbol{\theta})\right)}_{\boldsymbol{n}_{t-1}(\boldsymbol{\tau}_i)}, \hat{\boldsymbol{n}}(\boldsymbol{\theta})\right) \tag{B.15}$$

where $\boldsymbol{G}_t = (G_{1,t}, ..., G_{S,t})$. Eq. (B.15) highlights how $n_{j,t}(\boldsymbol{\tau}_i)$ depends on the whole history of mutant allele frequencies (i.e., it depends on $\bar{\boldsymbol{\tau}}_{i,0}, \bar{\boldsymbol{\tau}}_{i,1}, ..., \bar{\boldsymbol{\tau}}_{i,t-1}$). Then, from eq. (B.9), we see that each time step t of this history depends on $M_{i,t}$, which depends on $\boldsymbol{\tau}_i$ (through τ_i). Hence, $n_{j,t}(\boldsymbol{\tau}_i)$ can be written as a function of $\boldsymbol{\tau}_i$. We will assume that as such, $n_{j,t}(\boldsymbol{\tau}_i)$ is differentiable with respect to τ_i, which is a key property that we use below.

We now substitute eq. (B.13) in eq. (B.12) to obtain

$$\begin{aligned}\left.\frac{\partial R_{0,i}(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i=\theta_i} &= \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \frac{\partial w_i\left(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \boldsymbol{n}_t(\boldsymbol{\tau}_i), \hat{\boldsymbol{n}}\right)}{\partial \tau_i} q_i^{\circ}(\boldsymbol{\omega}_t) \\ &+ \sum_{j=1}^{S} \sum_{k=1}^{\infty} \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \frac{\partial^k w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \boldsymbol{n}, \hat{\boldsymbol{n}})}{(\partial n_j)^k} \frac{1}{k!} \left(\mathcal{N}_{j,t} - n_{j,t}(\boldsymbol{\tau}_i)\right)^k q_i^{\circ}(\boldsymbol{\omega}_t).\end{aligned} \tag{B.16}$$

Then, since $q_i^{\circ}(\boldsymbol{\omega}_t)$ captures the joint genetic-ecological process under neutrality (i.e., when $\boldsymbol{\tau}_i = \boldsymbol{\theta}$ is equal to the resident and local abundance vector is $\hat{\boldsymbol{n}} = (\hat{n}_1, ..., \hat{n}_S)$ at resident equilibrium), we have

$$\sum_{\boldsymbol{\omega}_t \in \Omega_i} \left(\mathcal{N}_{j,t} - n_{j,t}(\boldsymbol{\tau}_i)\right)^k q_i^{\circ}(\boldsymbol{\omega}_t) = \sum_{\boldsymbol{\omega}_t \in \Omega_i} \left(\hat{n}_j - n_{j,t}(\boldsymbol{\tau}_i)\right)^k q_i^{\circ}(\boldsymbol{\omega}_t) = O(\delta_i) \qquad \forall j \, \forall k \, \forall t. \tag{B.17}$$

The first equality in eq. (B.17) follows from the fact that according to our assumption that resident ecologcial dynamics are deterministic, $\mathcal{N}_{j,t}$ does not fluctuate under the neutral distribution $q_i^{\circ}(\boldsymbol{\omega}_t)$ (and so $\mathcal{N}_{j,t} = \hat{n}_j \forall t$). The second equality in eq. (B.17) follows from the fact that the difference between $\hat{n}_j$ and the average of $n_{j,t}(\boldsymbol{\tau}_i)$ over $q_i^{\circ}(\boldsymbol{\omega}_t)$ is of order δ_i (this derives from Taylor expanding $n_{j,t}(\boldsymbol{\tau}_i)$ around $\delta = 0$, which according to eqs. 4 and 5, gives $n_{j,t}(\boldsymbol{\tau}_i) = n_{j,t}(\boldsymbol{\theta}) + O(\delta_i) = \hat{n}_j + O(\delta_i)$).

Substituting eq. (B.17) into (B.16), the sensitivity of the growth rate can be written as

$$\left.\frac{\partial R_{0,i}(\tau_i,\boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i=\theta_i} = s_i(\boldsymbol{\theta}) + O(\delta_i), \tag{B.18}$$

where we refer to

$$s_i(\boldsymbol{\theta}) = \sum_{t=0}^{\infty}\sum_{\boldsymbol{\omega}_t\in\Omega_i} \frac{\partial w_i(\tau_i,\tilde{\boldsymbol{\tau}}_{i,t},\boldsymbol{n}_t(\boldsymbol{\tau}_i),\hat{\boldsymbol{n}})}{\partial \tau_i} q_i^{\circ}(\boldsymbol{\omega}_t) \tag{B.19}$$

as the approximate selection gradient. Substituting eq. (B.18) into eq. (B.3), and then into the change of allele frequency $\Delta p_i = p_i(1-p_i)\delta_i \mathcal{S}_i(\boldsymbol{\theta}) + O(\delta_i^2)$ (eq. 1 of the main text), we see that the remainder $O(\delta_i)$ in eq. (B.18) will contribute an effect of order $\delta_i O(\delta_i) = O(\delta_i^2)$ on allele frequency change and can thus be neglected to the first order. This shows that selection on a mutant allele in species i can be approximated as eq. (12) of the main text, i.e., as

$$\mathcal{S}_i(\boldsymbol{\theta}) \approx \epsilon(\boldsymbol{\theta}) s_i(\boldsymbol{\theta}), \tag{B.20}$$

where $s_i(\boldsymbol{\theta})$ is given by eq. (B.19).

B.3.3 Intra- and inter-temporal effects

It now remains to decompose $s_i(\boldsymbol{\theta})$, eq. (B.19), into intra- and inter-temporal effects of selection. To that end, we evaluate the fitness derivative in eq. (B.19) using the chain rule to obtain

$$\frac{\partial w_i(\tau_i,\tilde{\boldsymbol{\tau}}_{i,t},\boldsymbol{n}_t(\boldsymbol{\tau}_i),\hat{\boldsymbol{n}})}{\partial \tau_i} = \frac{\partial w_i(\tau_{\bullet,i},\tilde{\boldsymbol{\tau}}_i,\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i},\tilde{\boldsymbol{\tau}}_i,\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial \tilde{\tau}_i}\frac{\partial \tilde{\tau}_{i,t}}{\partial \tau_i} + \sum_{j=1}^{S}\frac{\partial w_i(\tau_{\bullet,i},\tilde{\boldsymbol{\tau}}_i,\boldsymbol{n},\hat{\boldsymbol{n}})}{\partial n_j}\frac{\partial n_{j,t}}{\partial \tau_i}, \tag{B.21}$$

where $w_i(\tau_{\bullet,i},\tilde{\boldsymbol{\tau}}_i,\boldsymbol{n},\hat{\boldsymbol{n}}(\boldsymbol{\theta}))$ is the fitness of a focal individual in species i with phenotype $\tau_{\bullet,i}$, when its neighbours have average phenotype given by the vector $\tilde{\boldsymbol{\tau}}_i = (\theta_1,...,\theta_{i-1},\tilde{\tau}_i,\theta_{i+1},...,\theta_S)$ (where $\tilde{\tau}_i$ is the average phenotype in species i among the neighbours of the focal individual). The fitness function on the right hand side of eq. (B.21) no longer depends explicitly on time, in contrast to the function on the left hand side (which depends on time through $\tilde{\boldsymbol{\tau}}_{i,t}$ and $\boldsymbol{n}_t$). This is owing to our use of the chain rule, which allows us to capture time dependence only in the derivatives of $\tilde{\boldsymbol{\tau}}_{i,t}$ and $\boldsymbol{n}_t$ with respect to τ_i in the right hand side of eq. (B.21). In this equation, the derivative of $\tilde{\boldsymbol{\tau}}_{i,t}$ can be specified by inserting eq. (B.9) into eq. (B.8), to get

$$\left.\frac{\partial \tilde{\tau}_{i,t}}{\partial \tau_i}\right|_{\substack{\tau_i=\theta_i\\ \boldsymbol{n}=\hat{\boldsymbol{n}}}} = \frac{M_{i,t}-1}{\hat{n}_i-1}. \tag{B.22}$$

Substituting eq. (B.22) into eq. (B.21), which is in turn substituted into eq. (B.19), gives

$$s_i(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \left[\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tilde{\tau}_i} \left(\frac{M_{i,t}-1}{\hat{n}_i - 1} \right) + \sum_{j=1}^{S} \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial n_j} \frac{\partial n_{j,t}}{\partial \tau_i} \right] q_i^{\circ}(\boldsymbol{\omega}_t), \quad \text{(B.23)}$$

for the selection gradient. Next, we decompose eq. (B.23) as $s_i(\boldsymbol{\theta}) = s_{\mathrm{w},i}(\boldsymbol{\theta}) + s_{\mathrm{e},i}(\boldsymbol{\theta})$, where

$$s_{\mathrm{w},i}(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \left[\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tilde{\tau}_i} \left(\frac{M_{i,t}-1}{\hat{n}_i - 1} \right) \right] q_i^{\circ}(\boldsymbol{\omega}_t) \quad \text{(B.24)}$$

$$s_{\mathrm{e},i}(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \left[\sum_{j=1}^{S} \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial n_j} \frac{\partial n_{j,t}}{\partial \tau_i} \right] q_i^{\circ}(\boldsymbol{\omega}_t). \quad \text{(B.25)}$$

We will now show that $s_{\mathrm{w},i}(\boldsymbol{\theta})$ as defined by eq. (B.24) reads as eq. (14) of the main text (which captures the intra-temporal effects), and $s_{\mathrm{e},i}(\boldsymbol{\theta})$ as defined by eq. (B.25) reads as eqs. (15)–(21) of the main text (which captures the inter-temporal effects).

B.3.4 Intra-temporal effects, $s_{\mathrm{w},i}(\boldsymbol{\theta})$

To show that eq (B.24) equals eq. (14), we simply use the fact that $\sum_{\boldsymbol{\omega}_t \in \Omega_t} q_i^{\circ}(\boldsymbol{\omega}_t) = 1$, so that

$$s_{\mathrm{w},i}(\boldsymbol{\theta}) = \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \tilde{\tau}_i} r_i(\boldsymbol{\theta}) \quad \text{(B.26)}$$

where

$$r_i(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \frac{M_{i,t}-1}{\hat{n}_i - 1} q_i^{\circ}(\boldsymbol{\omega}_t). \quad \text{(B.27)}$$

From the definition of $q_i^{\circ}(\boldsymbol{\omega}_t)$ (eq. B.5), $r_i(\boldsymbol{\theta})$ as defined by eq. (B.27) is the probability that two randomly sampled individuals from species i in the same patch carry an allele that is identical-by-descent, i.e., the pairwise coefficient of relatedness. Hence, eq. (B.26) corresponds to eq. (14), as required.

B.3.5 Inter-temporal effects, $s_{\mathrm{e},i}(\boldsymbol{\theta})$

Here, we derive eqs. (15)–(21), which underpin selection on a trait due to its inter-temporal effects.

Lineage-experienced abundance. We first write eq. (B.25) as eq. (15) of the main text,

$$s_{\mathrm{e},i}(\boldsymbol{\theta}) = \sum_{j=1}^{S} \frac{\partial \overline{n}_j}{\partial \tau_i} \times \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial n_j}, \quad \text{(B.28)}$$

where

$$\frac{\partial \overline{n}_j}{\partial \tau_i} = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \frac{\partial n_{j,t}}{\partial \tau_i} q_i^{\circ}(\boldsymbol{\omega}_t), \tag{B.29}$$

is the effect of the mutation on the local abundance of species j experienced by a mutant of species i that is randomly sampled from its local lineage (i.e., the lineage of carriers of the mutant trait τ_i that reside in the focal patch in which the mutation first appeared).

Inter-temporal effects of a mutant on its own ecology. We now derive eqs. (16)–(18) of the main text, which reveal the inter-temporal nature of the effect of a mutant on its lineage-experienced abundance. We begin by writing the effect of a mutant on lineage-experienced abundance (eq. B.29) in vector form,

$$\frac{\partial \overline{\boldsymbol{n}}}{\partial \tau_i} = \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \frac{\partial \boldsymbol{n}_t}{\partial \tau_i} q_i^{\circ}(\boldsymbol{\omega}_t), \tag{B.30}$$

where $\overline{\boldsymbol{n}} = (\overline{n}_1, \overline{n}_2, \ldots, \overline{n}_S)$. The derivative on the right hand side of eq. (B.30) can be evaluated by differentiating both sides of eq. (B.14) with respect to τ_i using the chain rule. The result of this operation can be written in vector form as

$$\frac{\partial \boldsymbol{n}_t}{\partial \tau_i} = \mathbf{C}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{n}_{t-1}}{\partial \tau_i} + \boldsymbol{\Psi}_i(\boldsymbol{\theta}) \hat{n}_i \frac{M_{i,t-1}}{\hat{n}_i}, \tag{B.31}$$

where $\mathbf{C}(\boldsymbol{\theta})$ is the community matrix (eq. 6), and

$$\boldsymbol{\Psi}_i(\boldsymbol{\theta}) = \frac{1}{\hat{n}_i} \left(\frac{\partial F_1(\bar{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \bar{\tau}_i}, \frac{\partial F_2(\bar{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \bar{\tau}_i}, \ldots, \frac{\partial F_S(\bar{\boldsymbol{\tau}}_i, \boldsymbol{n}, \hat{\boldsymbol{n}})}{\partial \bar{\tau}_i} \right), \tag{B.32}$$

is a vector of S elements (all evaluated at $\boldsymbol{\theta}$) with $\bar{\boldsymbol{\tau}}_i = (\theta_1, \ldots, \theta_{i-1}, \bar{\tau}_i, \theta_{i+1}, \ldots, \theta_S)$ where $\bar{\tau}_i$ is the average phenotype in the focal patch (i.e., the k-th element of $\boldsymbol{\Psi}_i(\boldsymbol{\theta})$ is given by eq. 17 of the main text).

Eq. (B.31) is a linear recurrence equation in $\partial \boldsymbol{n}_t / \partial \tau_i$ with initial condition $\partial \boldsymbol{n}_0 / \partial \tau_i = \mathbf{0}$ (as the first mutant appears at time 0). The solution to a linear vector recurrence of the form $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{b}_{t-1}$, with initial condition $\mathbf{x}_0 = \mathbf{0}$, is $\mathbf{x}_t = \sum_{h=0}^{t-1} \mathbf{b}_h \mathbf{A}^{t-h-1}$ (e.g., Sydsaeter et al., 2008). Hence, the solution of eq. (B.31) is

$$\frac{\partial \boldsymbol{n}_t}{\partial \tau_i} = \sum_{h=0}^{t-1} \boldsymbol{\Psi}_i(\boldsymbol{\theta}) \hat{n}_i \frac{M_{i,h}}{\hat{n}_i} \mathbf{C}(\boldsymbol{\theta})^{t-h-1} = \boldsymbol{\Psi}_i(\boldsymbol{\theta}) \sum_{h=1}^{t} \mathbf{K}_h(\boldsymbol{\theta}) \hat{n}_i \frac{M_{i,t-h}}{\hat{n}_i}, \tag{B.33}$$

where recall $\mathbf{K}_h(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})^{h-1}$ (see eq. 18). Substituting the right hand side of eq. (B.33) into eq. (B.30) gives

$$\begin{aligned} \frac{\partial \overline{\boldsymbol{n}}}{\partial \tau_i} &= \sum_{t=0}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \boldsymbol{\Psi}_i(\boldsymbol{\theta}) \sum_{h=1}^{t} \mathbf{K}_h(\boldsymbol{\theta}) \hat{n}_i \frac{M_{i,t-h}}{\hat{n}_i} q_i^{\circ}(\boldsymbol{\omega}_t) \\ &= \boldsymbol{\Psi}_i(\boldsymbol{\theta}) \sum_{h=1}^{\infty} \mathbf{K}_h(\boldsymbol{\theta}) \hat{n}_i \left[\sum_{t=h}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_t} \frac{M_{i,t-h}}{\hat{n}_i} q_i^{\circ}(\boldsymbol{\omega}_t) \right], \end{aligned} \tag{B.34}$$

where the second line is obtained by exchanging dummy variables in the first line.

Next, we define

$$r_{i,h}(\boldsymbol{\theta}) = \sum_{t=h}^{\infty} \sum_{\boldsymbol{\omega}_t \in \Omega_i} \frac{M_{i,t-h}}{\hat{n}_i} q_i^{\circ}(\boldsymbol{\omega}_t). \tag{B.35}$$

According to this definition, $r_{i,h}(\boldsymbol{\theta})$ is the relatedness coefficient between two individuals of species i sampled h time periods apart in the monomorphic resident population ($r_{i,h}(\boldsymbol{\theta})$ sums the probability of sampling two mutants at h time periods apart over the whole sojourn time of the lineage in a single group, see eq. B.5). Such relatedness coefficients can be computed using standard identity-by-descent arguments (Malécot, 1973, Epperson, 1999, Lehmann, 2007, 2008), and for $h \geq 1$, it takes the form given in eq. (19) of the main text. If we substitute eq. (B.35) into eq. (B.34), we obtain

$$\frac{\partial \overline{\boldsymbol{n}}}{\partial \tau_i} = \boldsymbol{\Psi}_i(\boldsymbol{\theta}) \sum_{h=1}^{\infty} \mathbf{K}_h(\boldsymbol{\theta}) \left[\hat{n}_i r_{i,h}(\boldsymbol{\theta}) \right], \tag{B.36}$$

which is eq. (16) of the main text in vector form, as required.

Dispersal-limited versions of press perturbations. Finally, we derive eqs. (20)–(21) of the main text. Substituting eqs. (19) and (18) into eq. (B.36), we have

$$\begin{aligned} \frac{\partial \overline{\boldsymbol{n}}}{\partial \tau_i} &= \boldsymbol{\Psi}_i(\boldsymbol{\theta})(1 - m_i(\boldsymbol{\theta}))\hat{n}_i \bar{r}_i(\boldsymbol{\theta}) \sum_{h=1}^{\infty} (1 - m_i(\boldsymbol{\theta}))^{h-1} \mathbf{C}(\boldsymbol{\theta})^{h-1} \\ &= \boldsymbol{\Psi}_i(\boldsymbol{\theta})(1 - m_i(\boldsymbol{\theta}))\hat{n}_i \bar{r}_i(\boldsymbol{\theta}) \Lambda_i(\boldsymbol{\theta}) \\ &= \boldsymbol{\Psi}_i(\boldsymbol{\theta})\hat{n}_i \Lambda_i(\boldsymbol{\theta})(1 - m_i(\boldsymbol{\theta}))\bar{r}_i(\boldsymbol{\theta}), \end{aligned} \tag{B.37}$$

where

$$\Lambda_i(\boldsymbol{\theta}) = [\mathbf{I} - (1 - m_i(\boldsymbol{\theta}))\mathbf{C}(\boldsymbol{\theta})]^{-1} \tag{B.38}$$

comes from the standard matrix summation result: $\sum_{h=1}^{\infty} \mathbf{A}^{h-1} = (\mathbf{I} - \mathbf{A})^{-1}$ (Sydsaeter et al., 2008), and reveals the dispersal-limited versions of press perturbations. As required, eq. (B.37) is eq. (20) in vector form.

Appendix C: Coevolution of helping and dispersal

C.1 Life-cycle

Here, we work out the example on the coevolution of helping and dispersal that is described in section 5.1.5. First, let us give more details on the life-cycle considered in this model. We assume a semelparous life-cycle whereby adults first interact socially (i.e., help each other), reproduce (in the absence of density-dependent

competition), and then die. Offspring disperse according to a dispersal probability that is inheritable. Those that disperse survive dispersal with probability s. After dispersal, each offspring survives density-dependent regulation with probability $1/(1+\gamma J)$ where J is the total number of juveniles in a patch.

We consider the evolution of two traits: the level of helping and the probability of dispersing at birth. To capture this, we write the phenotype of a focal individual as $\tau_\bullet = (z_\bullet, d_\bullet)$, where $z_\bullet$ is its level of helping and $d_\bullet$ the dispersal probability of its own and of its offspring. Likewise, we write the average phenotype among neighbours of the focal as $\tilde{\tau} = (\tilde{z}, \tilde{d})$, where $\tilde{z}$ is the average level of helping among the neighbours of the focal, and $\tilde{d}$ is the dispersal probability of their offspring.

Two remarks are worth making about this model before proceeding to its analysis. First, our life cycle assumptions integrate those of Rousset and Ronce (2004), who studied the evolution of dispersal only, and those of Lehmann et al. (2006), who studied the evolution of helping only. Both of these works used the exact version of the selection gradient (Rousset and Ronce, 2004, eqs. 26-27, Lehmann et al., 2006, eqs. A.18-19), which required extensive numerics to compute singular strategies. Second, competition in this model occurs after dispersal, rather than before dispersal (like in the model of helping presented in the main text, section 5.1.1). In that latter case, selection in fact drives dispersal to zero (result not shown). This is because when competition occurs before dispersal, dispersal does not alleviate kin competition, which is the only driver of dispersal evolution here (i.e., in the absence of environmental extinction and inbreeding cost, Hamilton and May, 1977, Clobert et al., 2001).

C.2 Necessary components

In order to evaluate the approximate selection gradient, we need the fitness of a focal individual in a mutant patch with n individuals, when other patches are monomorphic for the resident traits $\theta = (z, d)$ and at ecological equilibrium, $\hat{n}(\theta)$. According to our assumptions, this is

$$w(\tau_\bullet, \tilde{\tau}, n, \hat{n}(\theta)) = F(\bar{\tau}, n, \hat{n}(\theta)) \left(\frac{(1-d_\bullet) f(z_\bullet, \tilde{z})}{\lambda(\bar{\tau}, n, \hat{n}(\theta))} \right) + \hat{n}(\theta) \left(\frac{d_\bullet s f(z_\bullet, \tilde{z})}{(1-d+ds) f(z,z) \hat{n}(\theta)} \right), \tag{C.1}$$

where $f(z_\bullet, \tilde{z})$ is the fecundity of the focal individual, which depends on the focal's own level of helping $z_\bullet$ and the average level of helping $\tilde{z}$ among its neighbors in its group (given by eq. 34), and

$$\lambda(\bar{\tau}, n, \hat{n}(\theta)) = (1-\bar{d}) f(\bar{z}, \bar{z}) n + ds f(z,z) \hat{n}(\theta), \tag{C.2}$$

is the expected number of offspring in the focal patch after dispersal (but before competition). The first summand of eq. (C.1) is thus the philopatric component of fitness (i.e., the expected number of offspring that settle in the focal patch), and the second, the dispersal component (i.e., the expected number of offspring that settle in other patches).

Our analysis of selection also requires specifying the population map, $F(\bar{\tau}, n, \hat{n}(\theta))$, which is the expected number of offspring reaching adulthood in the focal patch, conditional on the focal patch being of size n in the parental generation and other patches being at the resident ecological equilibrium, $\hat{n}(\theta)$. To compute $F(\bar{\tau}, n, \hat{n}(\theta))$, note first that from our assumption of Poisson fecundity (section 2.1), the (random) number of offspring J entering competition in the focal patch is Poisson distributed with mean $\lambda(\bar{\tau}, n, \hat{n}(\theta))$ (eq. C.2). Second, note that each of these J offspring independently survives competition with probability $1/(1+\gamma J)$. Hence, the expected number of offspring surviving competition is obtained as the expected value of the Binomially distributed number of surviving offspring (with parameters J and $1/(1+\gamma J)$), over the Poisson distribution of J with parameter λ, i.e.,

$$F(\bar{\tau}, n, \hat{n}(\theta)) = \sum_{J=0}^{\infty} \frac{J}{1+\gamma J} \frac{\lambda(\bar{\tau}, n, \hat{n}(\theta))^J \exp\Big(-\lambda(\bar{\tau}, n, \hat{n}(\theta))\Big)}{J!}. \tag{C.3}$$

This is a complicated expression, but it is well approximated by a first-order delta approximation to the mean (Lynch and Walsh, 1998),

$$F(\bar{\tau}, n, \hat{n}(\theta)) \approx \frac{\lambda(\bar{\tau}, n, \hat{n}(\theta))}{1+\gamma\lambda(\bar{\tau}, n, \hat{n}(\theta))}. \tag{C.4}$$

From eq. (C.4), the equilibrium patch size in a resident population (obtained by solving $\hat{n} = F(\theta, \hat{n}, \hat{n})$ for $\hat{n}$) is

$$\hat{n}(\theta) = \frac{1}{\gamma}\left(1 - \frac{1}{(1-d+ds)\, f(z,z)}\right), \tag{C.5}$$

which depends on the level of helping z (through fecundity, $f(z,z)$) and decreases with dispersal d. The final piece of information that is necessary to compute the selection gradient is the coefficient of relatedness. This is given by eq. (10) of the main text, with resident equilibrium $\hat{n}$ given by eq. (C.5) and

$$m(\theta) = \frac{ds}{1-d+ds}, \tag{C.6}$$

as the backward probability of dispersal in the resident population.

C.3 Analysis of selection

From eqs. (13)–(21) of the main text, the selection gradient on trait $u \in \{z, d\}$ (respectively helping and dispersal) is given by

$$s_u(\theta) = \frac{\partial w(\tau_\bullet, \tilde{\tau}, n, \hat{n})}{\partial u_\bullet} + \frac{\partial w(\tau_\bullet, \tilde{\tau}, n, \hat{n})}{\partial \tilde{u}} r(\theta) + \frac{\partial \overline{n}}{\partial u} \times \frac{\partial w(\tau_\bullet, \tilde{\tau}, n, \hat{n})}{\partial n} \tag{C.7}$$

where

$$\frac{\partial \overline{n}}{\partial u} = \frac{\partial F(\bar{\tau}, n, \hat{n})}{\partial \overline{u}} \left[1 - (1 - m(\theta)) \frac{\partial F(\bar{\tau}, n, \hat{n})}{\partial n} \right]^{-1} (1 - m(\theta)) \bar{r}(\theta). \tag{C.8}$$

Singular helping and dispersal strategies are found by substituting eqs. (C.1)–(C.6) into eq. (C.7) and solving

$$s_z(\theta^*) = 0 \qquad \text{and} \qquad s_d(\theta^*) = 0, \tag{C.9}$$

numerically fo $\theta^* = (d^*, z^*)$ (subject to the constraint that patch size is positive, $\hat{n}(\theta) > 0$, see eq. C.5). The singular strategies for specific parameter values are shown in Fig. (6) (which also shows the results of individual-based simulations for this model).

References

Abrams, P. A., H. Matsuda, and Y. Harada. 1993. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. *Evolutionary Ecology* 7:465–487.

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

Akçay, E. 2017. Population structure reduces benefits from partner choice in mutualistic symbiosis. *Proceedings of the Royal Society B-Biological Sciences* 284:1–10.

Akçay, E. and J. Van Cleve. 2012. Behavioral responses in structured populations pave the way to group optimality. *American Naturalist* 179:257–269.

Alizon, S. and P. D. Taylor. 2008. Empty sites can promote altruistic behavior. *Evolution* 62:1335–1344.

Apaloo, J. and S. Butler. 2009. Evolutionary stabilities in multidimensional-traits and several-species models. *Evolutionary Ecology Research* 11:637–650.

Bao, M. and G. Wild. 2012. Reproductive skew can provide a net advantage in both conditional and unconditional social interactions. *Theoretical Population Biology* 82:200–208.

Barton, N. H. 2001. The evolutionary consequences of gene flow and local adaptation: future approaches. In

Clobert, J., E. Danchin, A. A. Dhondt, and J. D. Nichols (eds.), *Dispersal*, chap. 23, pp. 329–340. Oxford University Press.

Bohonak, A. J. 1999. Dispersal, Gene Flow, and Population Structure. *The Quarterly Review of Biology* 74:21–45.

Bonduriansky, R. 2012. Rethinking heretidy, again. *Trends in Ecology and Evolution* 27:330–6.

Brown, J. S. and T. L. Vincent. 1987. Coevolution as an evolutionary game. *Evolution* 41:66–79.

Cadet, C., R. Ferrière, J. A. J. Metz, and M. van Baalen. 2003. The evolution of dispersal under demographic stochasticity. *American Naturalist* 162:427–41.

Case, T. J. 2000. An Illustrated Guide to Theoretical Ecology. Oxford University Press, Oxford.

Caswell, H. 2000. Matrix Population Models. Sinauer Associates, Massachusetts.

Charlesworth, B. 1994. Evolution in Age-Structured Populations. Cambridge University Press, Cambridge, 2th edn.

Charnov, E. L. 1976. Optimal foraging, the marginal value theorem. *Theoretical Population Biology* 9:129–36.

Chesson, P. L. 1978. Predator-prey theory and variability. *Annual Review of Ecology and Systematics* 9:323–47.

Chesson, P. L. 1981. Models for spatially distributed populations: the effect of within-patch variability. *Theoretical Population Biology* 19:288–325.

Chesson, P. L. 1984. Persistence of a Markovian population in a patchy environment. *Zeitschrift für Wahrscheinlichkeitstheorie* 66:97–107.

Christiansen, F. B. 1991. On conditions for evolutionary stability for a continuously varying character. *American Naturalist* 138:37–50.

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

Cohen, J. E. 1979. Long-run growth rates of discrete multiplicative processes in Markovian environments. *Journal of Mathematical Analysis and Applications* 69:243–251.

Comins, H. N., W. D. Hamilton, and R. M. May. 1980. Evolutionarily stable dispersal strategies. *Journal ot Theoretical Biology* 82:205–230.

Cornell, S. J. and O. Ovaskainen. 2008. Exact asymptotic analysis for metapopulation dynamics on correlated dynamic landscapes. *Theoretical Population Biology* 74:209–25.

Débarre, F., S. L. Nuismer, and M. Doebeli. 2014. Multidimensional (co)evolutionary stability. *Am. Nat.* 184:158–171.

Débarre, F. and S. P. Otto. 2016. Evolutionary dynamics of a quantitative trait in a finite asexual population. *Theoretical Population Biology* 108:75–88.

Débarre, F., L. S., M. van Baalen, and S. Gandon. 2012. Evolution of host life-history traits in a spatially structured host-parasite system. *American Naturalist* 179.

El Mouden, C. and A. Gardner. 2008. Nice natives and mean migrants: the evolution of dispersal-dependent social behaviour in viscous populations. *Journal of Evolutionary Biology* 21:1480–1491.

Epperson, B. K. 1999. Gene genealogies in geographically structured populations. *Genetics* 152:797–806.

Eshel, I. 1983. Evolutionary and continuous stability. *Journal ot Theoretical Biology* 103:99–111.

Eshel, I. 1996. On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *Journal of Mathematical Biology* 34:485–510.

Foster, K. R. 2004. Diminishing returns in social evolution: the not-so-tragic commons. *Journal of Evolutionary Biology* 17:1058–1072.

Foster, K. R. and T. Wenseleers. 2006. A general model for the evolution of mutualisms. *Journal of Evolutionary Biology* 19:1283–1293.

Frank, S. A. 1994. Genetics of mutualism: the evolution of altruism between species. *Journal of Theoretical Biology* 170:393–400.

Frank, S. A. 1998. Foundations of Social Evolution. Princeton University Press, Princeton, NJ.

Gandon, S. and Y. Michalakis. 1999. Evolutionary stable dispersal rate in a metapopulation with extinction and kin competition. *Journal of Theoretical Biology* 199:275–290.

Gardner, A. and S. A. West. 2004. Cooperation and punishment, especially in humans. *American Naturalist* 164:753–764.

Gardner, A. and S. A. West. 2006. Demography, altruism, and the benefits of budding. *Journal of Evolutionary Biology* 19:1707–1716.

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

Grafen, A. 1991. Modeling in behavioural ecology. In Krebs, J. R. and N. Davies (eds.), *Behavioural Ecology*, pp. 5–31. Blackwell Scientific Publications, Oxford.

Grafen, A. 2006. A theory of Fisher's reproductive value. *Journal of Mathematical Biology* 53:15–60.

Hamilton, W. D. 1964. The genetical evolution of social behaviour, 1. *Journal of Theoretical Biology* 7:1–16.

Hamilton, W. D. 1966. The moulding of senescence by natural selection. *Journal of Theoretical Biology* 12:12–45.

Hamilton, W. D. 1971. Selection of selfish and altruistic behaviour in some extreme models. In Eisenberg, J. and W. Dillon (eds.), *Man and Beast: Comparative Social Behavior*, pp. 59–91. Smithsonian Institutions Press, Washington, DC.

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

Hanski, I. and M. Gilpin. 1997. Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press.

Hartl, D. and A. G. Clark. 2007. Principles of Population Genetics. Sinauer, Massachusetts, 4th edn.

Hubbell, S. P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton.

Inderjit, D. A. Wardle, R. Karban, and R. M. Callaway. 2011. The ecosystem and evolutionary contexts of allelopathy. *Trends in Ecology and Evolution* 26:655–662.

Johnstone, R. A., M. Cant, and F. J. 2012. Sex-biased dispersal, haplodiploidy and the evolution of helping in social insects. *Proceedings of the Royal Society B: Biological SciencesBiological Sciences* 279:787–793.

Johnstone, R. A. and M. A. Cant. 2008. Sex differences in dispersal and the evolution of helping and harming. *American Naturalist* 172:318–330.

Kisdi, E. and T. Priklopil. 2010. Evolutionary branching of a magic trait. *J. Math. Biol.* 63:361–397.

Kobayashi, Y., J. Y. Wakano, and H. Ohtsuki. 2015. A paradox of cumulative culture. *Journal of Theoretical Biology* 379:79–88.

Kuijper, B. and R. A. Johnstone. 2012. How dispersal influences parent–offspring conflict over investment. *Behavioral Ecology* .

Kuijper, B., I. Pen, and F. J. Weissing. 2012. A guide to sexual selection theory. *Annual Review of Ecology, Evolution, and Systematics* 43:287–311.

Lankau, R. 2008. A Chemical Trait Creates a Genetic Trade-Off Between Intra- and Interspecific Competitive Ability. *Ecology* 89:1181–1187.

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

Le Galliard, J., R. Ferrière, and U. Dieckmann. 2005. Adaptive Evolution of Social Traits: Origin, Trajectories. *American Naturalist* 165:206–224.

Lehmann, L. 2007. The evolution of trans-generational altruism: kin selection meets niche construction. *Journal of Evolutionary Biology* 20:181–189.

Lehmann, L. 2008. The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes. *Evolution* 62:549–566.

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

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

Leibold, M. A. and J. M. Chase. 2017. Metacommunity Ecology, Volume 59. Princeton University Press.

Leimar, O. 2009. Multidimensional convergence stability. *Evolutionary Ecology Research* 11:191–208.

Leslie, P. H. and J. C. Gower. 1958. The properties of a stochastic model for two competing species. *Biometrika* 45:316–330.

Lion, S. 2017. Theoretical approaches in evolutionary ecology: environmental feedback as a unifying perspective. *American Naturalist* In press.

Lion, S. and S. Gandon. 2009. Habitat saturation and the spatial evolutionary ecology of altruism. *Journal of Evolutionary Biology* 22:1487–1502.

Lion, S. and S. Gandon. 2010. Life history, habitat saturation and the evolution of fecundity and survival altruism. *Evolution* 64:1594–606.

Lion, S. and M. van Baalen. 2007a. From infanticide to parental care: Why spatial structure can help adults be good parents. *American Naturalist* 170:E26–E46.

Lion, S. and M. van Baalen. 2007b. Self-structuring in spatial evolutionary ecology. *Ecology Letters* 11:277–295.

Lynch, M. and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer, Massachusetts.

Malécot, G. 1973. Génétique des population diploïdes naturelle dans le cas d'un seul locus. III-Parenté, mutations et migration. *Annale de Génétique et de Selection Animale* 5:333–361.

Maynard Smith, J. 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Metz, J. and M. Gyllenberg. 2001. How should we define fitness in structured metapopulation models? including an application to the calculation of evolutionary stable dispersal strategies. *Proceedings of the Royal Society of London Series B-Biological Sciences* 268:499–508.

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

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

Metz, J. J. and C. G. F. de Kovel. 2013. The canonical equation of adaptive dynamics for Mendelian diploids and haplo-diploids. *Interface Focus* 3:20130025–20130025.

Meyn, S. and R. L. Tweedie. 2009. Markov Chains and Stochastic Stability. Cambridge University Press, Cambridge, 2th edn.

Michod, R. E. 1979. Evolution of Life Histories in Response to Age-Specific Mortality Factors. *The American Naturalist* 113:531–550.

Mullon, C., L. Keller, and L. Lehmann. 2016. Evolutionary stability of jointly evolving traits in subdivided populations. *American Naturalist* 188:175–195.

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

Mullon, C., M. Reuter, and L. Lehmann. 2014. The evolution and consequences of sex-specific reproductive variance. *Genetics* 196:235–252.

Nagylaki, T. 1992. Introduction to population genetics. Springer-Verlag, Heidelberg.

Nakamaru, M. and Y. Iwasa. 2005. The evolution of altruism by costly punishment in lattice-structured populations: score-dependent viability versus score-dependent fertility. *Evolutionary Ecology Research* 7:853–870.

Nakamaru, M., H. Matsuda, and Y. Iwasa. 1997. The evolution of cooperation in a lattice-structured population. *Journal of Theoretical Biology* 184:65–81.

Neuhauser, C. 2002. Effects of local interactions and local migration on stability. *Theoretical Population Biology* 62:297–308.

Nurmi, T., K. Parvinen, and V. Selonen. 2018. Joint evolution of dispersal propensity and site selection in structured metapopulation models. *Journal of Theoretical Biology* 444:50–72.

Odling-Smee, F. J., K. L. Laland, and M. W. Feldman. 2003. Niche Construction. Princeton University Press, Princeton, NJ.

Parker, G. A. 1990. Sperm competition games: raffles and roles. *Proceedings of the Royal Society B-Biological Sciences* 242:120–126.

Parker, G. A. and J. Maynard Smith. 1990. Optimality theory in evolutionary biology. *Science* 349:27–33.

Parvinen, K., U. Dieckmann, M. Gyllenberg, and J. A. J. Metz. 2003. Evolution of dispersal in metapopulations with local density dependence and demographic stochasticity. *Journal of Evolutionary Biology* 16:143–53.

Parvinen, K. and J. A. J. Metz. 2008. A novel fitness proxy in structured locally finite metapopulations with diploid genetics, with an application to dispersal evolution. *Theoretical Population Biology* 73:517–528.

Purcell, J., A. Brelsford, and L. Avilés. 2012. Co-evolution between sociality and dispersal: The role of synergistic cooperative benefits. *Journal of Theoretical Biology* 312:44–54.

Ranta, E., P. Lundberg, and V. Kaitala. 2006. Ecology of Populations. Cambridge University Press, Cambridge, UK.

Rodrigues, A. M. M. and A. Gardner. 2012. Evolution of helping and harming in heterogeneous populations. *Evolultion* 66:2065–2079.

Rousset, F. 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press, Princeton, NJ.

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

Schaffer, W. M. 1982. The application of optimal control theory to the general life history problem. *American Naturalist* 121:418–431.

Sozou, P. D. 2009. Individual and social discounting in a viscous population. *Proc. R. Soc. B Biol. Sci.* 276:2955–2962.

Stearns, S. 1992. The Evolution of Life Histories. Oxford University Press, Oxford.

Sydsaeter, K., P. Hammond, A. Seierstad, and A. Strøm. 2008. Further Mathematics for Economic Analysis. Prentice Hall, Essex, 2th edn.

Taylor, P. 1988a. Inclusive fitness models with two sexes. *Theoretical Population Biology* 34:145–168.

Taylor, P. D. 1988b. An inclusive fitness model for dispersal of offspring. *Journal of Theoretical Biology* 130:363–378.

Taylor, P. D. 1989. Evolutionary stability in one-parameter models under weak selection. *Theoretical Population Biology* 36:125–143.

Taylor, P. D. 1992. Altruism in viscous populations - an inclusive fitness model. *Evolutionary Ecology* 6:352–356.

Taylor, P. D. 1994. Sex ratio in a stepping-stone population with sex-specific dispersal. *Theoretical Population Biology* 45:203–218.

Taylor, P. D. and S. A. Frank. 1996. How to make a kin selection model. *Journal of Theoretical Biology* 180:27–37.

Taylor, P. D. and A. J. Irwin. 2000. Overlapping generations can promote altruistic behavior. *Evolution* 54:1135–1141.

terHorst, C. P., P. C. Zee, K. D. Heath, T. E. Miller, A. I. Pastore, S. Patel, S. J. Schreiber, M. J. Wade, and M. R. Walsh. 2018. Evolution in a Community Context: Trait Responses to Multiple Species Interactions. *Am. Nat* 191:368–380.

Tilman, D. 1982. Resource Competition and Community Structure. Princeton University Press, New Jersey.

Tuljapurkar, S. 1989. An uncertain life: demography in random environments. *Theoretical Population Biology* 35:227–94.

Tuljapurkar, S., C. C. Horvitz, and J. B. Pascarella. 2003. The many growth rates and elasticities of populations in random environments. *American Naturalist* 162:489–502.

Urban, M. C. 2011. The evolution of species interactions across natural landscapes. *Ecology Letters* 14:723–732.

Urban, M. C., M. A. Leibold, P. Amarasekare, L. De Meester, R. Gomulkiewicz, M. E. Hochberg, C. A. Klausmeier, N. Loeuille, C. de Mazancourt, J. Norberg, J. H. Pantel, S. Y. Strauss, M. Vellend, and M. J. Wade. 2008. The evolutionary ecology of metacommunities. *Trends in Ecology and Evolution* 23:311–317.

van Baalen, M. and A. Rand. 1998. The unit of selection in viscous populations and the evolution of altruism. *Journal of Theoretical Biology* 193:631–648.

Van Cleve, J. 2015. Social evolution and genetic interactions in the short and long term. *Theoretical Population Biology* 103:2–26.

Wagner, C. E., R. J. Best, L. J. Harmon, and B. Matthews. 2017. Evolution in a Community Context: On Integrating Ecological Interactions and Macroevolution. *Trends in Ecology and Evolution* 32:291–304.

Wakano, J. Y. and C. Miura. 2014. Trade-off between learning and exploitation: The Pareto-optimal versus evolutionarily stable learning schedule in cumulative cultural evolution. *Theoretical Population Biology* 91:37–43.

Weigang, H. C. and E. Kisdi. 2015. Evolution of dispersal under a fecundity-dispersal trade-off. *Journal of Theoretical Biology* 371:145–153.

West, S. A. and A. Gardner. 2010. Altruism, spite, and greenbeards. *Science* 327:1341–1344.

West, S. A., A. S. Griffin, and A. Gardner. 2007. Evolutionary explanations for cooperation. *Current Biology* 17:661–672.

Wild, G. 2011. Inclusive fitness from multitype branching processes. *Bulletin of Mathematical Biology* 73:1028–51.

Wild, G., A. Gardner, and S. A. West. 2009. Adaptation and the evolution of parasite virulence in a connected world. *Nature* 459:983–986.

Wolfram Research, I. 2016. Mathematica. Wolfram Research, Inc., Champaign, Illinois.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

Wyatt, G. A., S. A. West, and A. Gardner. 2013. Can natural selection favour altruism between species? *Journal of Evolutionary Biology* 26:1854–65.

Yodzis, P. 1989. Introduction to Theoretical Ecology. Harper & Row, New York.

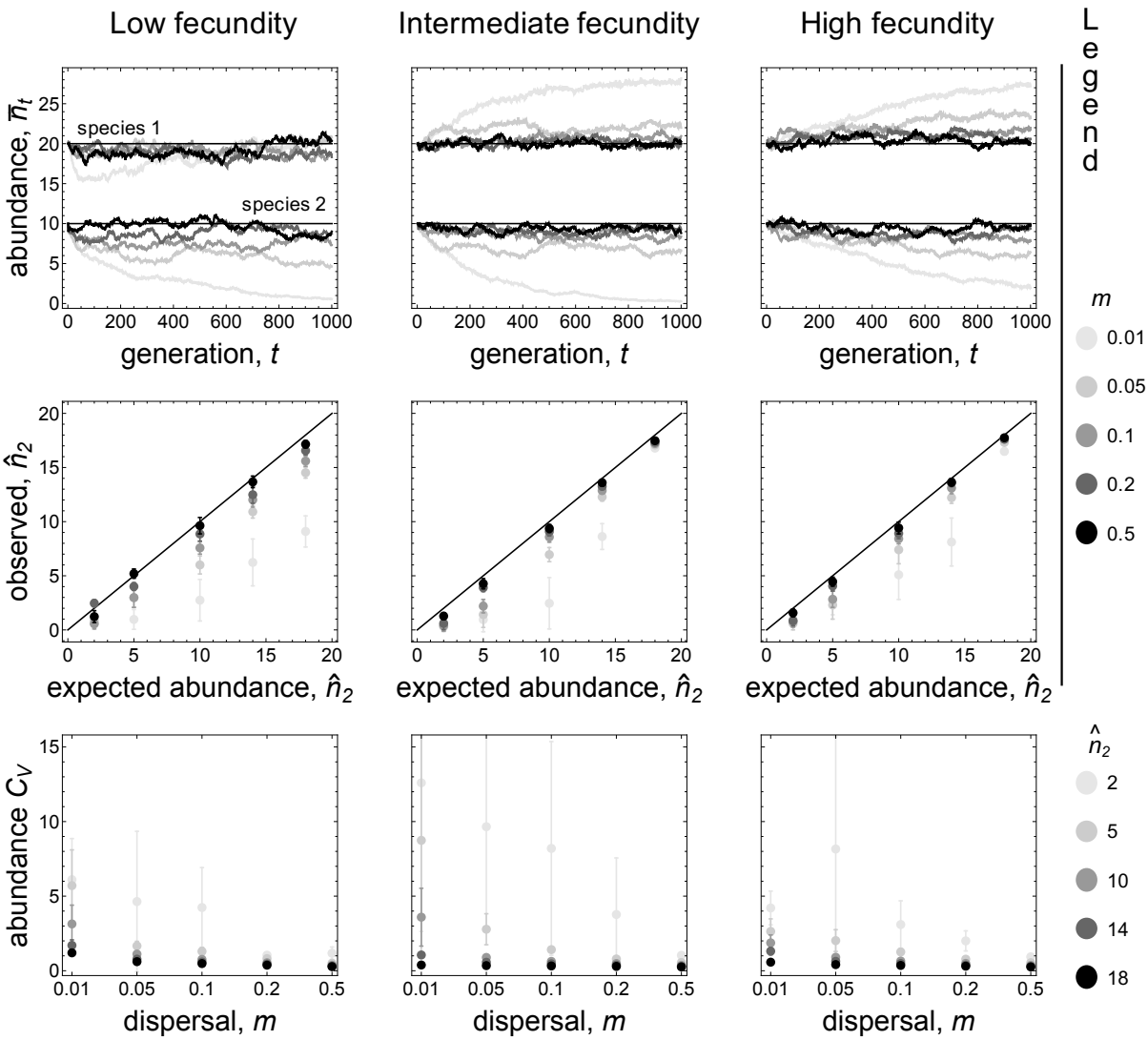

Figure 1: Stochastic dynamics of local abundance and their deterministic approximation (eq. 7) for Leslie-Gower model with two species. Different columns correspond to different levels of baseline fecundity (i.e., fecundity in the absence of density of competition). The first column has extremely low baseline fecundity ($f_1 = 1.1$, $f_2 = 1.095$), the second, intermediate fecundity ($f_1 = 2$, $f_2 = 1.955$), and the third, high fecundity ($f_1 = 5$, $f_2 = 4.955$). **Top row**: deterministic (straight line, determined from eq. 7, with $\hat{\boldsymbol{n}} = \boldsymbol{n}_{t-1}$) and stochastic dynamics (dots, with 1000 patches, see Appendix A.1 for details on simulations, with $m_1 = m_2 = 0.01, 0.05, 0.1, 0.2, 0.5$ – see legend on right hand side – and competition parameters γ and η chosen so that the deterministic equilibrium given by eq. 8 is $\hat{n}_1 = 20$ and $\hat{n}_2 = 10$ – see Supplementary Table 1 for values). **Middle row**: Comparisons between the deterministic (x-axis, from eq. 8) and stochastic (y-axis) number of individuals of species 2, averaged over 1000 generations starting at deterministic equilibrium in each patch (error bars give the standard deviation of the stochastic value, with $m_1 = m_2 = 0.001, 0.01, 0.1, 0.5$ – see legend on right hand side – and competition parameters (γ, η) chosen so that the deterministic equilibrium given by eq. 8 is $\hat{n}_1 = 20$ and $\hat{n}_2 = 2, 5, 10, 14, 18$ – in general, competition parameters decrease as fecundity decreases and patch size increases; see Supplementary Table 1 for the parameter values we used, which span two orders of magnitude, from 0.001 to 0.1). Departures from the diagonal indicate deviations between the exact process and the deterministic approximation. **Bottom row**: From the same simulations described in the middle row, these graphs show the coefficient of variation C_v of abundance of species 2 across patches according to dispersal level (i.e., C_v is the ratio of the average number of individuals of species 2 per patch to its standard deviation, averaged over 1000 generations, shown here for different values of $\hat{n}_2$ – see legend on right hand side – error bars give the standard deviation of C_v over 1000 generations).

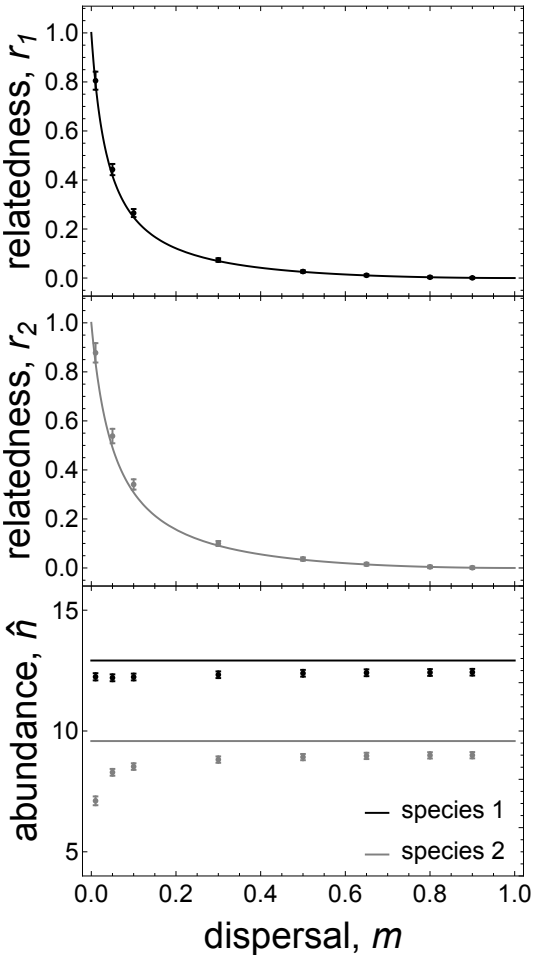

Figure 2: Relatedness and average local abundance under stochastic individual-based simulations and their deterministic approximation. The top two graphs show the relatedness for species 1 and 2, respectively, under the Leslie-Gower demographic model (section 3.2). Relatedness values obtained from the deterministic approximation are shown in full lines (i.e., obtained from eq. 10 with the ecological equilibrium derived from the deterministic approximation, eq. 8, parameter values $f_1 = 2$, $f_2 = 1.8$, $\gamma = 0.07$, and $\eta = 0.01$). Relatedness computed from individual based simulations are shown as points with standard deviation shown by error bars (time average of population mean over 5000 generations after burnin of 5000 generations with 1000 patches, with $m = 0.01, 0.05, 0.1, 0.3, 0.5, 0.8, 0.9$, see Appendix A.1 for details on calculations of relatedness). The bottom graph displays the average local abundance of species 1 (in black) and 2 (in gray). Full lines are for the deterministic approximation (eq. 8). Points are results obtained from stochastic individual-based simulations (time average of population mean over 5000 generations after burnin of 5000 generations, with $m = 0.01, 0.05, 0.1, 0.3, 0.5, 0.8, 0.9$). Error bars show standard deviation.

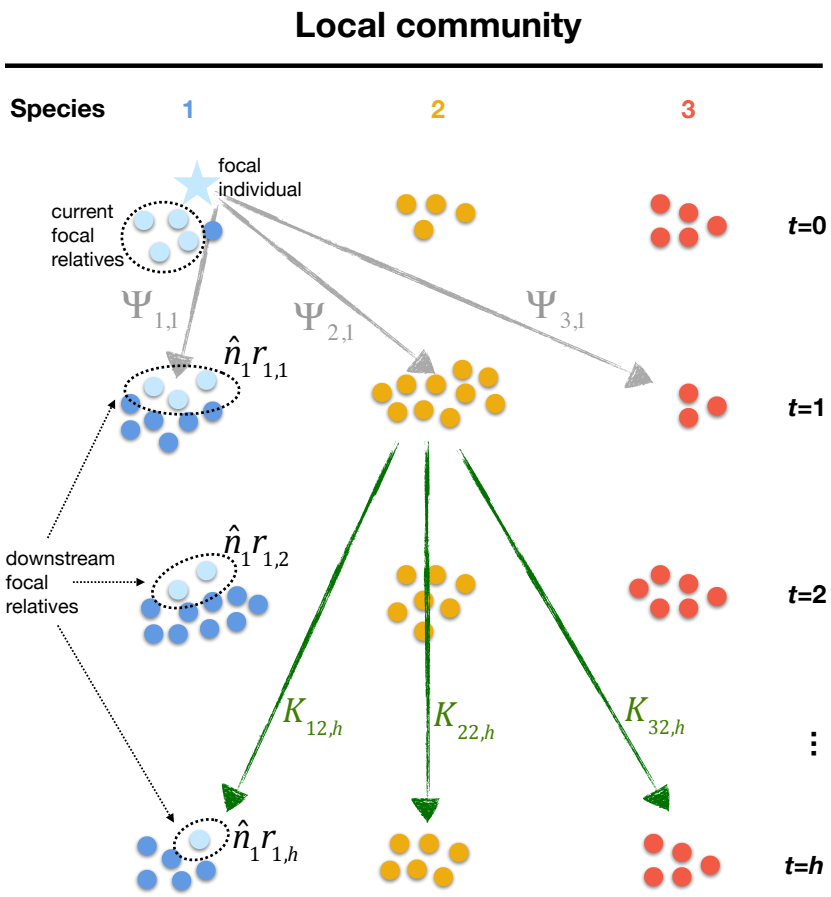

Figure 3: Inter-temporal effects of a focal individual. As an example, consider a local community of three species, labelled 1 (blue), 2 (orange) and 3 (red). We look at the effect of a mutation in species 1. A focal carrier of a mutation τ_1 in species 1 living at time $t = 0$ (denoted by a blue star) first directly influences the population dynamics of species 1, 2 and 3 at time $t = 1$ according to $\Psi_{1,1}(\theta)$, $\Psi_{2,1}(\theta)$ and $\Psi_{3,1}(\theta)$ (grey arrows, eq. 17). This change in abundance at time $t = 1$ affects the abundance of the other species through time due to ecological interactions. For e.g., a change in abundance of species 2 at time $t = 1$ influences the abundance of species 1, 2 and 3 at time $t = h$ according to $\mathrm{K}_{12,h}(\boldsymbol{\theta})$, $\mathrm{K}_{22,h}(\boldsymbol{\theta})$ and $\mathrm{K}_{32,h}(\boldsymbol{\theta})$, respectively (eq. 18). These changes are experienced by $\hat{n}_1(\boldsymbol{\theta})r_{1,h}(\boldsymbol{\theta})$ relatives of the focal individual (eq. 19).

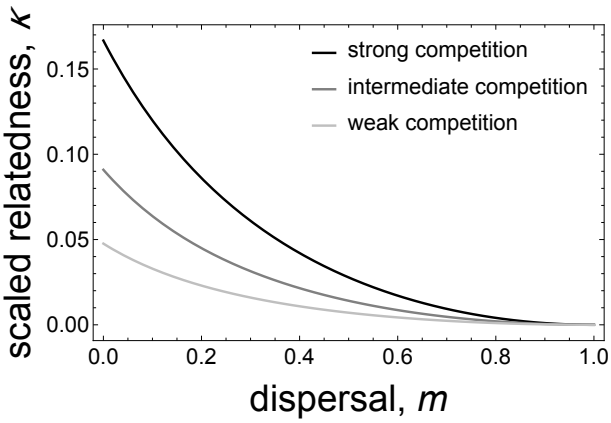

Figure 4: Scaled relatedness for the evolution of social interactions within species. Eq. (33) with eq. (25) plotted against dispersal with strong ($\gamma = 0.2$, black), intermediate ($\gamma = 0.1$, dark grey) and weak ($\gamma = 0.05$, light grey) levels of competition. Other parameters: $f(\theta,\theta) = 2$. Scaled relatedness therefore decreases with dispersal and patch size (since smaller values of γ lead to larger equilibrium patch size).

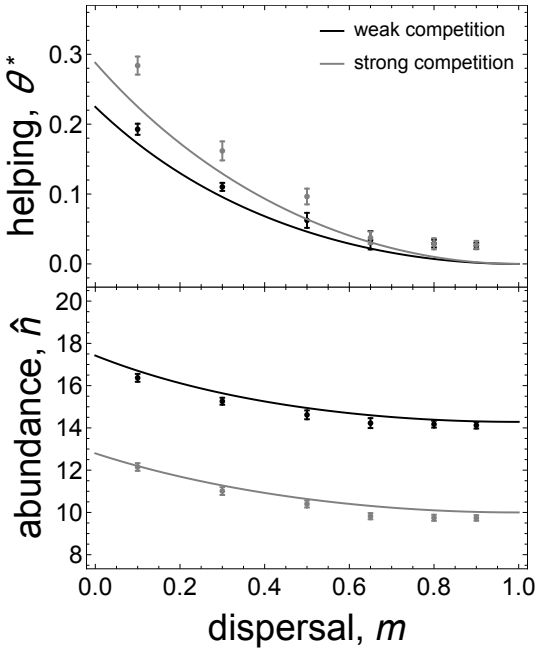

Figure 5: Convergence stable level of within-species helping and the concomitant local abundance it generates. Full lines are the convergence stable strategies (top) and concomitant local abundance (bottom) obtained from the selection gradient (obtained by finding the θ^* value solving eq. 32) under weak (black, $\gamma = 0.07$) and strong (grey, $\gamma = 0.1$) intra-specific competition (other parameters: $B = 0.5$, $C = 0.05$, $f = 2$). Points are the results obtained from individual based simulations (time average over 50000 generation after 50000 generations of evolution, error bars show standard deviation). Parameters for simulations: 1000 patches, $m = 0.1, 0.3, 0.5, 0.8, 0.9$, probability of a mutation = 0.01, standard deviation of the quantitative effect of a mutation = 0.005 (see Appendix A.1 for details on simulations).

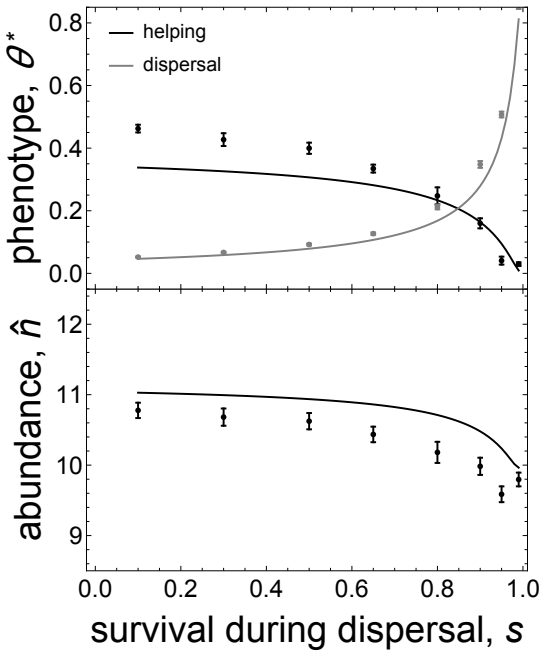

Figure 6: Co-evolutionary convergence stable level of helping and dispersal, and the concomitant local abundance it generates. Full lines are the convergence stable strategies (top) for helping (black) and dispersal (grey) and concomitant local abundance (bottom) obtained from the selection gradient (eq. C.7, with $B = 0.5$, $C = 0.05$, $f = 2$, $\gamma = 0.05$). Points are the results obtained from individual based simulations (time average over 50000 generation after 50000 generations of evolution, error bars show standard deviation). Parameters for simulations: 1000 patches, survival during dispersal $s = 0.1, 0.3, 0.5, 0.8, 0.9, 0.95, 0.99$, probability of a mutation = 0.01, standard deviation of the quantitative effect of a mutation on each trait = 0.005, and no covariance (see Appendix A.1 for details on simulations).

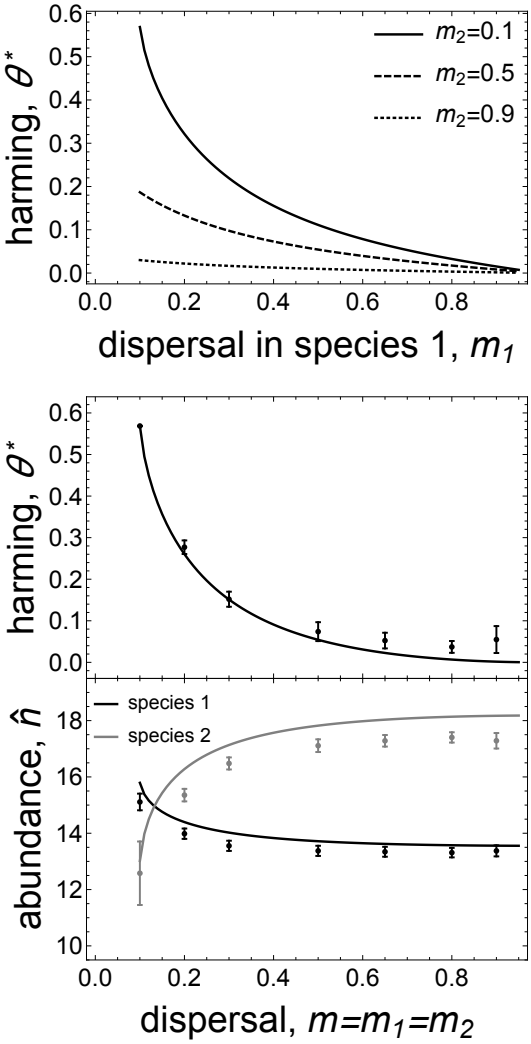

Figure 7: Convergence stable level of between-species harming and concomitant local abundance it generates. Full lines are the convergence stable strategies (top and middle) and concomitant local abundance (bottom) in species 1 (black) and 2 (grey) obtained from the selection gradient (eqs. 13–20 along with eqs. 36–43, with $C_1 = 0.0085$, $f_1 = 2.2$, $f_2 = 2$, $\gamma = 0.055$, $\eta = 0.025$, $\alpha = 0.24$; Top: dispersal in species 2, m_2 =0.1 (full), 0.5 (dashed), 0.9; Middle and bottom: dispersal in species $m_2 = m_1 = m$. Points are the results obtained from individual based simulations (time average over 50000 generation after 50000 generations of evolution, error bars show standard deviation). Parameters for simulations: 1000 patches, $m = 0.1, 0.2, 0.3, 0.5, 0.65, 0.8, 0.9$, probability of a mutation = 0.01, standard deviation of the quantitative effect of a mutation = 0.005.

Supplementary Table

Equilibrium abundance of species 2 – $\hat{n}_2$		**Low Fecundity** $(f_1, f_2) = (1.1, 1.095)$	**Intermediate Fecundity** $(f_1, f_2) = (2, 1.955)$	**High Fecundity** $(f_1, f_2) = (5, 4.955)$
2	γ	0.00457071	0.0456818	0.182045
	η	0.00429293	0.0431818	0.179545
5	γ	0.00406667	0.0406	0.1606
	η	0.00373333	0.0376	0.1576
10	γ	0.0035	0.0348333	0.134833
	η	0.003	0.0303333	0.130333
14	γ	0.00328431	0.0325	0.120735
	η	0.00245098	0.025	0.113235
18	γ	0.00381579	0.0369737	0.115921
	η	0.00131579	0.0144737	0.0934211

Table 1: **Competition parameters used in simulations to generate Figure 1.** Values are found by solving eq. (8) of the main text for γ and η with $\hat{n}_1 = 20$, $\hat{n}_2$ given in the left column and fecundities, f_1 and f_2, given in the top row.