Eco-evolutionary dynamics under limited dispersal: ecological inheritance, altruism within and spite between species

Laurent Lehmann*, Charles Mullon †

Tuesday 7th November, 2017

Abstract

Understanding selection on traits affecting intra- and inter-specific interactions in meta-communities subject to stochastic demography is a long-standing problem in ecology and evolution. The problem is that eco-evolutionary dynamics are too complicated to make analytical investigations tractable. We here circumvent this problem by approximating the selection gradient on a quantitative trait that influences local community dynamics, assuming that such dynamics are deterministic but taking into account kin selection effects arising from local demographic stochasticity. Selection on a trait is shown to depend on how an individual expressing a change in trait value changes: (1) its own fitness and that of its relatives living in the present; and (2) the fitness of all its relatives living in the future through modifications of local ecological conditions. The latter effects capture feedbacks between ecology and evolution, which impact downstream generations ("ecological inheritance"), and take the form of press-perturbations of community ecology. As an illustration, we analyze the evolution of altruism within and spite between species. We show that altruism within species is more likely when competition is for material resources than for space, and that spite between species can readily evolve to alleviate local competition for future relatives. More broadly, our approximate selection gradient opens an avenue to analyze a variety of eco-evolutionary questions in meta-communities, from selection on mutualism to prey-predator coevolution.

Keywords: ecology, kin selection, population dynamics, adaptive dynamics, species interactions.

*Department of Ecology and Evolution, University of Lausanne, Switzerland.
†Department of Ecology and Evolution, University of Lausanne, Switzerland.

1 Introduction

Interactions within and between species are common in nature and probably connect almost all living organisms to one another. How such intra- and inter-species interactions emerge and are maintained depends on intertwined ecological and evolutionary dynamics, also known as eco-evolutionary dynamics (see Lion, 2017 for a review). One major challenge in understanding these dynamics is due to the spatial structuring of communities, which results from interactions between individuals typically occurring at a finer spatial scale than the global scale of the community. Such "meta-community" structure, in which individuals of different species are divided among small local patches connected to each other by limited dispersal (e.g., Hanski and Gilpin, 1997; Clobert et al., 2001; Hartl and Clark, 2007), leads to local demographic stochasticity and thus spatial heterogeneities in community demography. Local demographic stochasticity, which is the outcome of individual reproductive and survival variance, has two complicating consequences for eco-evolutionary dynamics.

First, local demographic stochasticity causes *genetic stochasticity*, whereby allele frequencies fluctuate within patches, which leads to the build up of genetic relatedness between members of the same species. Genetic relatedness in turn influences selection on social traits, which like helping or aggressiveness are traits that mediate intra-species interactions and affect the reproductive success of both their actor (direct effects) and recipient (indirect effects) (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 patches. As a consequence, multi-species demography in patch structured populations is typically significantly more complicated that in panmictic populations (Chesson, 1978, 1981; Hubbell, 2001; Neuhauser, 2002; Cornell and Ovaskainen, 2008). As genetic and ecological stochasticities are further likely to influence one another through eco-evolutionary feedbacks, understanding selection on a trait under limited dispersal requires a careful account of how phenotypic effects on local population dynamics translate into fitness costs and benefits.

Fundamental insights into selection can be obtained from the long-term adaptive dynamics of quantitative traits, which are the gradual changes of phenotypic 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 major goals of adaptive dynamics is to identify local attractors of evolutionary dynamics, which are phenotypic trait values towards which selection drives a population under gradual evolution (referred to as convergence stable strategies, Eshel, 1983; Taylor, 1989; Christiansen, 1991; Geritz et al., 1998; Rousset, 2004; Leimar, 2009). Convergence stable phenotypes can be identified through 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 in present and past generations changing trait value. The selection gradient allows to

determine the direction of selection and has helped to understand natural selection on many phenotypic traits of broad biological interest, from senescence and life history, to sperm competition and sex-ratio, to altruism, cumulative cultural evolution 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; Kuijper et al., 2012; Akçay and Van Cleve, 2012; Mullon et al., 2014; Wakano and Miura, 2014; Kobayashi et al., 2015).

Gold standard expressions for the selection gradient on traits that influence meta-population dynamics 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). These expressions can be extended to consider multi-species interactions and thus in principle, allow to study convergence stability under eco-evolutionary feedback for many biologically relevant situations. However, even under the simplest model of limited dispersal, which is the island model of dispersal (Wright, 1931), the selection gradient on traits affecting eco-evolutionary dynamics remains dispiritingly complicated (Rousset and Ronce, 2004; Lehmann et al., 2006b; Alizon and Taylor, 2008; Wild, 2011) and is often computed numerically as a derivative without 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 very specific demographic models under limited dispersal with finite patch size have been studied in detail (Comins et al., 1980; Gandon and Michalakis, 1999; Lehmann et al., 2006b; Rodrigues and Gardner, 2012). A biologically intuitive understanding of selection on traits that influence meta-community dynamics, in particular traits that underpin intra- and inter-species interactions, is therefore out of immediate reach when using the exact selection gradient.

To circumvent this problem, various approximations for the selection gradient under limited dispersal have been proposed. The most prominent of those has perhaps been the heuristically-based 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; Dèbarre et al., 2012). However, more general multi-species coevolution scenarios have not been much investigated using pair approximation, presumably because analysis remains difficult under isolation-by-distance.

In this paper, we propose a novel heuristic approximation for the selection gradient on traits that influence eco-evolutionary dynamics under limited dispersal. The crux of our approximation is that it treats local population size dynamic as deterministic (i.e., we ignore ecological stochasticity) with a single ecological attractor. Our approximation is derived for patch-structured populations (i.e., the population genetic island model of dispersal of Wright, 1931, and its ecological counterpart, Chesson, 1981), but it allows to address a wide spectrum of multi-species interactions and provides a biologically meaningful interpretation of selection on traits that influence meta-community evolution.

The rest of the paper is organized as follows. (1) We build a stochastic meta-community eco-evolutionary model. (2) We motivate an approximation of this model that ignores ecological stochasticity but takes genetic stochasticity into account. (3) We derive the selection gradient on a trait that influences the eco-evolutionary dynamics through intra- and inter-species interactions under our approximation, which leads to a biologically intuitive interpretation of selection in terms of the trait's direct, indirect and ecological feedback effects. (4) We use our approximation to study two examples of intra- and inter-species interactions: altruism within species and spite between species. We show that for these examples, our proposed approximation performs well compared to individual-based simulations.

2 Model

2.1 Biological assumptions

2.1.1 Life-cycle with stochastic demography

We consider a multi-species population that is divided among 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 finite number of offspring that follows a Poisson distribution, and then either survives or dies (with a probability that may depend on local interactions). (2) Each offspring either remains in its natal patch, or disperses to another that is randomly chosen among the remaining patches (the probability dispersal is assumed to be non-zero for all species but may differ among species). (3) 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).

2.1.2 Evolving phenotypes and the uninvadable species coalition

Each individual expresses a genetically determined evolving phenotype, or strategy, which can affect any event in the life cycle of any species, such as reproduction, survival, or dispersal. 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. Suppose that the population of each species is monomorphic. The meta-community (i.e., the collection of subdivided populations of each species) can then be described by the vector of strategies $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_S)$, where θ_i is the strategy expressed by each individual of species i (i.e. a monomorphic resident population). We say that $\boldsymbol{\theta}$ is an *uninvadable coalition* if any mutation $\tau_i \in \Theta_i$, which arises in any species i and results in an 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 that of a multi-species evolutionary stable strategy (Brown and Vincent, 1987, p. 68) or an evolutionary stable coalition (Apaloo and Butler, 2009, p. 640).

2.1.3 Adaptive dynamics

Since it is sufficient to consider unilateral deviations in phenotypic expression in single species to determine whether a coalition is uninvadable, we focus our attention on the evolutionary dynamics of a mutant allele τ_i for species i in a resident coalition $\boldsymbol{\theta}$. Owing to the assumptions of the infinite island model of dispersal, the change Δp_i in frequency p_i of a mutant allele τ_i in species i over one demographic time period (one life cycle iteration) can be written as

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

where $\delta_i = \tau_i - \theta_i$ is phenotypic effect of the mutation (Rousset, 2004, pp. 206–207 and Rousset and Ronce, 2004 p. 129, with notations here adapted to introduce species specific allele frequency change). The function $s_i(\boldsymbol{\theta})$, which depends on quantities evaluated only in the resident community $\boldsymbol{\theta}$, is the selection gradient. When selection is weak (so that $\delta \ll 1$ and terms $O(\delta^2)$ can be neglected), the selection gradient gives the direction of selection on the mutant at any allele frequency.

The selection gradient $s_i(\boldsymbol{\theta})$ is then sufficient to evaluate the first-order condition for a coalition to be uninvadable. The coalition $\boldsymbol{\theta}^*$ is uninvadable only if it satisfies

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

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

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

has negative real parts.

From our assumptions on the life cycle, the selection gradient $s_i(\boldsymbol{\theta})$ can be computed as a reproductive value and relatedness weighted change in the fitness of individuals experiencing different demographic states, which here are the different configurations of local patch states. The expression for the selection gradient is given formally by Rousset and Ronce, 2004, eqs. 26-27 (or Lehmann et al., 2016, eqs. E.27-29). Reproductive values reflect the fact that individuals residing in different demographic states contribute differently to the gene pool and thus capture the effects of ecological stochasticity on selection (Rousset

and Ronce, 2004; Lehmann et al., 2016). Relatedness, meanwhile, captures the notion that individuals that interact are more likely to carry alleles that are identical-by-descent, which is a consequence of genetic stochasticity.

In principle, one can use this "exact" selection gradient (Rousset and Ronce, 2004, eqs. 26-27, Lehmann et al., 2016, eqs. E.27-29) to study community evolution under the assumptions of our model. However, doing so would have two drawbacks. The first is that owing to the large number of possible demographic states within patches (here all possible configurations of the number of individuals of the species in interaction), the necessary computations are not straightforward and numerically expensive, even in the absence of inter-specific interactions (e.g., Rousset and Ronce, 2004; Lehmann et al., 2006b; 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 multi-species with dynamic stochastic demography can be found in the literature. The second drawback is that the decomposition of the exact selection gradient is in terms of reproductive value (Rousset and Ronce, 2004, eqs. 26-27, Lehmann et al., 2016, eqs. E.27-29, see Rousset and Ronce, 2004; Lehmann et al., 2006b; Alizon and Taylor, 2008; Wild et al., 2009; Wild, 2011 for applications to explicit demographic models), which is applicable to any type of demographic structuring (e.g., by age, stage, or environment, Frank, 1998; Rousset, 2004; Grafen, 2006). Since the relationship between reproductive values and ecological interactions is still unknown, the decomposition of the exact selection gradient offers little information about the influence of ecological interactions on selection.

The aim of this paper is therefore to give a tractable and biologically informative approximation for the selection gradient $s_i(\boldsymbol{\theta})$. We will show that this approximation, which neglects ecological stochasticity, allows to obtain a relevant interpretation of selection on ecological interactions under limited dispersal, and accurately predicts the direction of selection under a meaningful set of demographic parameters. We now first focus only on resident demographic dynamics and motivate our assumption to treat population dynamics as deterministic.

3 The resident community dynamics

3.1 Deterministic community dynamics

From a stochastic process point of view, our life-cycle assumptions entail that we consider an infinite stochastic model, i.e., an infinite number of interacting multidimensional Markov chains, that each describe the community dynamics on a patch, and that are coupled by dispersal (Chesson, 1981). In such a (complicated) system, considering only the mean number of individuals of each species on each patch can be a good guide to local community dynamics, provided stochastic effects are weak relative

to deterministic ones. This should be the case when the number of individuals of each species on each patch is not too low, and patches are not too disconnected from each other (i.e., dispersal is not too small). The dynamics of the expected number of individuals (abundance) of each species in a patch can then be approximated by the deterministic dynamics of the process (Chesson, 1981), which we set out to do.

Let us thus start by assuming that the meta-community is monomorphic for a resident $\boldsymbol{\theta}$. We denote by $n_{i,t}(\boldsymbol{\theta})$ the abundance of species i in a focal patch at time t (since all patches are identical, we do not need to label patches, e.g., Neuhauser, 2002). This number is assumed to be given deterministically by

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

where F_i is a population dynamic transition function for species i that depends on the community vector $\mathbf{n}_t = (n_{1,t}, n_{2,t}, ..., n_{S,t})$ in the previous time period in the focal patch and the phenotype vector $\boldsymbol{\theta}$ of each species. Importantly, the transition function F_i (eq. 4) is defined such as under the stochastic model, it gives the expected abundance of species i, conditional on the local community state being $\mathbf{n}_{t-1}(\boldsymbol{\theta})$ in the previous generation.

The resident community equilibrium then solves

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

Such equilibrium is locally stable or attracting when the eigenvalues of the community matrix (e.g., Yodzis, 1989; Case, 2000),

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

have absolute value less than one at the equilibrium (i.e., all derivatives in $\mathbf{C}(\boldsymbol{\theta})$ are evaluated at $\mathbf{n}(\boldsymbol{\theta})$ given by eq. 5). We will assume throughout that resident community dynamics (eq. 4) are such that they have a single locally stable equilibrium.

As an illustration of a transition function (eq. 4), consider a meta-community of two semelparous species whose life-cycle is as follows. (1) Each individual of species $i \in \{1, 2\}$ in a patch with $\mathbf{n} = (n_1, n_2)$ individuals produces a mean number $f_i / (1 + \gamma n_1 + \eta n_2)$ of offspring, where f_i is the number of offspring produced in the absence of density-dependent competition (maximal fecundity). The denominator $1 +$

$\gamma n_1 + \eta n_2$ captures density-dependent competition within species (with intensity γ), and between species (with intensity η). This model for offspring production is thus a special case of the Leslie-Gower model of species interaction (Leslie and Gower, 1958, eq. 1.1). (2) All adults die. (3) Each offspring of species i disperses with probability m_i to another patch and all offspring survive to adulthood.

Under assumptions (1)-(4) of the preceding paragraph, the expected number of individuals of species 1 and 2 on a focal patch, conditional on the local community state being $\mathbf{n}_{t-1}(\boldsymbol{\theta})$ in the previous time period can be written as

$$\begin{aligned} n_{1,t}(\boldsymbol{\theta}) &= (1-m_1)\left(\frac{f_1}{1+\gamma n_{1,t-1}(\boldsymbol{\theta})+\eta n_{2,t-1}(\boldsymbol{\theta})}\right) n_{1,t-1}(\boldsymbol{\theta}) + m_1 n_1(\boldsymbol{\theta}) \\ n_{2,t}(\boldsymbol{\theta}) &= (1-m_2)\left(\frac{f_2}{1+\gamma n_{2,t-1}(\boldsymbol{\theta})+\eta n_{1,t-1}(\boldsymbol{\theta})}\right) n_{2,t-1}(\boldsymbol{\theta}) + m_2 n_2(\boldsymbol{\theta}), \end{aligned} \tag{7}$$

The first term in each line of eq. (7) represents the number of settled individuals in the focal patch that were born locally. The second term in each line represents the total number of offspring that immigrate to the focal patch from other patches, which are assumed to be at the resident demographic equilibrium, denoted by $n_1(\boldsymbol{\theta})$ and $n_2(\boldsymbol{\theta})$ for species 1 and 2 respectively. In order to understand better the second term $m_i n_i(\boldsymbol{\theta})$ of each line of eq. (7), consider first that when the population is at the resident demographic equilibrium, which solves eq. (5), an individual must on average produce one offspring (so that the number of individuals remains constant). As a consequence, $n_1(\boldsymbol{\theta})$ and $n_2(\boldsymbol{\theta})$ give the total number of offspring produced of each species in a patch other than the focal one. In turn, this means that $m_1 n_1(\boldsymbol{\theta})$ and $m_2 n_2(\boldsymbol{\theta})$ are the average number of offspring immigrating into the focal patch of each species. Note that this entails that the transition function eq. (4) will in general depend on the resident demographic equilibrium (and could therefore be written more formally as $F_i(\boldsymbol{\theta}, \mathbf{n}_{t-1}(\boldsymbol{\theta}), \mathbf{n}(\boldsymbol{\theta})))$. For ease of presentation, however, we simply write the transition function as $F_i(\boldsymbol{\theta}, \mathbf{n}_{t-1}(\boldsymbol{\theta}))$ (in fact, we write only the arguments of function that will be used for taking derivatives).

For the two species example (eq. 7), the unique equilibrium satisfying eq. (5) that is positive (i.e., $n_1(\boldsymbol{\theta}) > 0$ and $n_2(\boldsymbol{\theta}) > 0$) reads as

$$\begin{aligned} n_1(\boldsymbol{\theta}) &= \frac{(f_1-1)\gamma - (f_2-1)\eta}{\gamma^2-\eta^2} \\ n_2(\boldsymbol{\theta}) &= \frac{(f_2-1)\gamma - (f_1-1)\eta}{\gamma^2-\eta^2}, \end{aligned} \tag{8}$$

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

3.2 Comparing deterministic and stochastic dynamics

In order to assess our proposition to ignore ecological stochasticity, we compared the deterministic community dynamics of the two species Leslie-Gower model introduced above (eq. 7) with individual based simulations of the stochastic model (see Appendix A for a description of the individual-based simulations).

3.2.1 Ecological stochasticity

We find that there is a good qualitative match between the deterministic dynamics given by eq. (7) and the average number of individuals of each species in a group observed under individual based simulations (Fig. 1, top panel). As predicted by theory (Chesson, 1981; Neuhauser, 2002), the deterministic dynamic is different to the observed average number of individuals within patches (Fig. 1, top panel, grey lines). However, this discrepancy is small provided dispersal is not too weak (Fig. 1, middle panel). Why dispersal mitigates the effects of ecological stochasticity can be understood as follows. Local population dynamics depend on both a local process at the patch level, which has a strong stochastic component because patches have few individuals, and a global process, which has a weaker stochastic component when the number of patches is large (as the number of patches grow infinite, patches affect each other in a non-stochastic way, Chesson, 1981). As dispersal increases, local population dynamics increasingly depend on the global process, so that they become less stochastic and more deterministic. The good qualitative match between the deterministic and stochastic dynamics under dispersal is also observed when no stable equilibrium exists, such as when the meta-community consists of preys and predators whose numbers cycle (see Appendix A.2 for a description of this model, Fig. 1, top panel).

3.2.2 Genetic stochasticity

Our simulations suggest that ecological stochasticity can be ignored for dispersal values above 0.1, approximately (Fig. 1, middle panel). Unless patches are large, individuals from the same species may still be intra-specifically related for such dispersal values, which suggests that unlike ecological stochasticity, genetic stochasticity cannot be ignored under limited dispersal. To see this, let us continue with our example of the Leslie-Gower model (eq. 7), and consider the probability that two individuals from the same species, randomly sampled in the same patch, carry an allele that is identical-by-descent when the population is monomorphic and at its demographic equilibrium (eq. 8). This is a standard measure of relatedness, which reflects the degree of genetic stochasticity (Frank, 1998; Rousset, 2004). This probability satisfies the recursion

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

for species $i \in \{1, 2\}$, where $n_i(\boldsymbol{\theta})$ is given by eq. (8) (see Rousset and Ronce, 2004, p. 135 for a formal justification of eq. 9). Eq. (9) can be understood as follows. The factor $(1 - m_i)^2$ is the probability that two randomly sampled individuals of species i are of philopatric origin. Then, with probability $1/n_i(\boldsymbol{\theta})$, they descend from the same parent so their relatedness is one, and with complementary probability $1 - 1/n_i(\boldsymbol{\theta})$, they descend from different parent so their relatedness is $r_i(\boldsymbol{\theta})$.

The solution to the relatedness recursion eq. (9) is

$$r_i(\boldsymbol{\theta}) = \frac{(1 - m_i)^2}{1 - [1 - (1 - m_i)^2]\,(n_i(\boldsymbol{\theta}) - 1)}, \tag{10}$$

which is equivalent to the standard F_{ST} quantity (e.g, Hartl and Clark, 2007; Rousset, 2004) except that here, the number of individuals is endogenously determined by an explicit demographic process. Inspection of eq. (10) shows that relatedness can build up to significant values even when dispersal is in a range in which we can legitimately approximate mean abundance by the deterministic model (Figure 2). In addition, we observe a good match between relatedness calculated from individual based simulations and from the approximation neglecting ecological stochasticity (eq. 10 with eq. 8, Figure 2). Taken together, these observations support the idea to study selection ignoring ecological stochasticity but taking into account genetic stochasticity.

4 Evolutionary analysis

4.1 The mutant deterministic community dynamics

We now derive the selection gradient $s_i(\boldsymbol{\theta})$ on a trait in species i, assuming that the ecological dynamics are deterministic. We assume that in this species i, only two alleles segregate in the population: the mutant with type τ_i and a resident with type θ_i. To describe how local community dynamics depend on the two alleles, we focus on the focal patch and denote by $M_{i,t} \in \{0, 1, 2, ..., n_{i,t}\}$ the number of mutant individuals of species i at demographic time t. Then, the average phenotype in the focal patch for species i at generation t, denoted $\bar{\tau}_i$, is

$$\bar{\tau}_{i,t} = \frac{M_{i,t}}{n_{i,t}}\tau_i + \left(1 - \frac{M_{i,t}}{n_{i,t}}\right)\theta_i. \tag{11}$$

For a deterministic dynamic, the number $n_{i,t}(\boldsymbol{\tau}_i)$ of individuals of species i at time t in a focal group is given by

$$n_{i,t}(\boldsymbol{\tau}_i) = F_i(\bar{\boldsymbol{\tau}}_{i,t-1}, \mathbf{n}_{t-1}(\boldsymbol{\tau}_i)) \qquad \text{for } i = 1, 2, ..., S, \tag{12}$$

where F_i the local population dynamic map (as in eq. 4), which now depends on the vector of average phenotypes $\bar{\boldsymbol{\tau}}_{i,t} = (\theta_1, ..., \theta_{i-1}, \bar{\tau}_{i,t}, \theta_{i+1}, ..., \theta_S)$ at time $t-1$ in the local community and on the community vector $\mathbf{n}_{t-1}(\boldsymbol{\tau}_i)$. The dependence of $n_{i,t}(\boldsymbol{\tau}_i)$ on $\bar{\boldsymbol{\tau}}_{i,t-1}$ captures that local population dynamics depend on the occurrence of both mutants and residents, so that ecology is influenced by evolution. If simultaneously $\bar{\boldsymbol{\tau}}_{i,t-1}$ depends on the community vector at the preceding generation $\mathbf{n}_{t-2}(\boldsymbol{\tau}_i)$, evolution is influenced by ecology, and an eco-evolutionary feedback emerges.

4.2 The inclusive fitness effect for the interactive community

Under the assumption that the local population dynamic (i.e., eq. 12) has a single attractor point for each resident community (see eqs. 5-6), we show in Appendix B that the (approximate) selection gradient on trait θ_i in species i in a $\boldsymbol{\theta}$ community can be decomposed as the sum of two terms

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

where the first term $s_{\mathrm{w},i}(\boldsymbol{\theta})$ captures selection owing to the immediate direct and indirect fitness effects of the trait, and $s_{\mathrm{e},i}(\boldsymbol{\theta})$ capture selection owing to indirect, feedback effects of the trait on the fitness of individuals living in downstream generations, which emerges are 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 terms in the next two sections.

4.2.1 Selection on immediate direct and indirect effects

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

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

(see Appendix B.3.1, eq. B.14 for derivation), where $w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n})$ 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), which is a function of (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); and (3) the focal community vector $\mathbf{n}$ (explicit examples of such a fitness function are given later when we apply our method, see eqs. 21 and 32). The fitness derivatives in eq. (14), which are evaluated in the resident population (i.e., with resident phenotype $\tau_i = \theta_i$ and resident equilibrium community $\mathbf{n} = \mathbf{n}(\boldsymbol{\theta})$), capture different effects of the trait. The first derivative represents the change in the fitness of a focal individual of species i resulting from the focal switching from resident to 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 patch neighbours resulting from the 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 relatedness coefficient $r_i(\boldsymbol{\theta})$, which gives the probability that any neighbours also carries the mutation. As such, $s_{\mathrm{w},i}(\boldsymbol{\theta})$ is the standard selection gradient in a meta-population with constant demography (Taylor and Frank, 1996; Frank, 1998; Rousset, 2004), which is evaluated at the equilibrium resident $\mathbf{n}(\boldsymbol{\theta})$. In our model, however, demography can vary in response to phenotype changes, and selection due to this ecological effect is captured by the second term of the selection gradient eq. (13), as we elaborate in the next section.

4.2.2 Selection on feedback effects due to ecological inheritance

The second term of the selection gradient eq. (13) can be written as

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

(see Appendix B.3.2, eq. B.16 for details), where $\overline{n}_j$ is the expected 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, see eq. B.17 for formal definition). The first derivative in eq. (15) is the effect of an increase in the local abundance of species j on the fitness of an individual from species i (in which the trait under selection is expressed), and the second derivative, the effect of an increase in the trait in species i (due to individuals carrying the mutation) on the local abundance of species j. By multiplying these two effects and summing them over all species of the community, eq. (15) captures how selection depends on the interplay between community ecology and evolution.

In particular, eq. (15) shows that a phenotypic change will be selected when such a change results in local ecological conditions that are favourable for the carriers of the change (i.e., when $(\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n})/\partial n_j)(\partial \overline{n}_j/\partial \tau_i) > 0$). This can occur when the fitness of an individual from species i either (1) increases with the abundance of species j and the phenotypic change in species i increases the abundance of species j; or (2) decreases with the abundance of species j and the phenotypic change in species i decreases the abundance of species j. In case (1), selection favours species i helping species j and in case (2), species i harming species j.

We further find that the effect of a phenotypic change in species i on the equilibrium abundance of

species j can be expressed as

$$\left.\frac{\partial \overline{n}_j}{\partial \tau_i}\right|_{\substack{\tau_i=\theta_i \\ \mathbf{n}=\mathbf{n}(\boldsymbol{\theta})}} = \sum_{k=1}^{S} \Psi_{k,i}(\boldsymbol{\theta}) \underbrace{\left[(1-m_i(\boldsymbol{\theta}))\,\bar{r}_i(\boldsymbol{\theta}) n_i(\boldsymbol{\theta})\right]}_{n_{\mathrm{R},i}(\boldsymbol{\theta})} \Lambda_{jk,i}(\boldsymbol{\theta}). \tag{16}$$

The first term of eq. (16),

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

is the change in the abundance of species k over a single generation due to a single focal individual in species i changing its phenotype, and $\bar{\boldsymbol{\tau}}_i = (\theta_1, ..., \theta_{i-1}, \bar{\tau}_i, \theta_{i+1}, ..., \theta_S)$ is the vector of average phenotypes. The second term of eq. (16), $n_{\mathrm{R},i}(\boldsymbol{\theta})$, is the expected number of relatives (individuals carrying an allele identical-by descent to the focal) in the generation descending from the focal individual, which depends on the relatedness

$$\bar{r}_i(\boldsymbol{\theta}) = \left(\frac{1}{n_i(\boldsymbol{\theta})} + \left(\frac{n_i(\boldsymbol{\theta})-1}{n_i(\boldsymbol{\theta})}\right) r_i(\boldsymbol{\theta})\right) \tag{18}$$

between two individuals sampled with replacement in the same patch. These $n_{\mathrm{R},i}(\boldsymbol{\theta})$ relatives in the descendant generation thus experience the change in abundance of species k. Finally, the third and last term in eq. (16), $\Lambda_{jk,i}(\boldsymbol{\theta})$, is given by the jk element of the matrix

$$\boldsymbol{\Lambda}_i(\boldsymbol{\theta}) = \left(\mathbf{I} - [1-m_i(\boldsymbol{\theta})]\mathbf{C}(\boldsymbol{\theta})\right)^{-1}, \tag{19}$$

and 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. In other words, $\Lambda_{jk,i}(\boldsymbol{\theta})$ reflects that the perturbation $\Psi_{k,i}(\boldsymbol{\theta})$ in species k will influence the abundance of species j in multiple ways by inter-generational effects via intermediary species (for e.g., by increasing the local abundance of one species in one generation, which in turn causes a decrease in the local abundance of another at the next generation, and so on, Figure 3).

Taken together, these considerations (eqs. 17-18) show that eq. (16) can be interpreted as the effect of a focal individual of species i carrying the mutant allele on the abundance of species j experienced by all the focal's relatives living in the future in the focal patch (see Figure 3 for a diagrammatic representation and Appendix B.3.2 for more mathematical details). Hence, the second term $s_{\mathrm{e},i}(\boldsymbol{\theta})$ (eq. 15) of the selection gradient reflects selection on the way ecological changes induced by an individual feedback on the fitness of its relatives, and such feedbacks depend on an entangled web of inter-species interactions that is woven across generations by ecological inheritance (Figure 3).

Interestingly, as dispersal goes to zero in the focal species ($m_i(\boldsymbol{\theta}) \to 0$), the matrix $\boldsymbol{\Lambda}_i(\boldsymbol{\theta})$ (eq. 19) tends

to the well-known matrix of press perturbations of community ecology [i.e., $\mathbf{\Lambda}_i(\boldsymbol{\theta}) \to (\mathbf{I} - \mathbf{C}(\boldsymbol{\theta}))^{-1}$], whose entries measure how a constant and persistent change in the abundance of one species influences the equilibrium abundance of another (e.g., Yodzis, 1989; Case, 2000). This is because as $m_i(\boldsymbol{\theta}) \to 0$, the mutant lineage may persist locally forever. But as dispersal increases ($m_i(\boldsymbol{\theta}) \to 1$), a mutant lineage will spend fewer generations locally, which means that ecological inheritance will take place over fewer generations and hence fewer intergenerational carry-over effects can take place (and so $\mathbf{\Lambda}_i(\boldsymbol{\theta}) \to \mathbf{I}$ approaches the identity matrix).

In summary, the selection gradient, $s_i(\boldsymbol{\theta})$, on a trait that influences meta-community dynamics depends on how a phenotypic change in a focal individual changes (1) its own fitness and that of relatives living in the present generation; and (2) the fitness of all relatives living in the future in the focal patch through heritable modifications of the ecological environment, which is captured by an analogue of a press perturbation (in the case where the demography of the evolving species is constant and affects a single other species, $s_{\mathrm{e},i}(\boldsymbol{\theta})$, reduces to Lehmann, 2008, eq. A39). How selection mould intra- and inter-species interactions in meta-communities can therefore be intuitively understood in terms of a multigenerational inclusive fitness effect.

5 Applications

In the following, we use our approximation of the selection gradient to study two examples of intra- and inter-species interactions. The first is the evolution of altruism within species under limited dispersal, which is a topic that has been extensively investigated in the past, and the second, the evolution of spite between species, which has not been studied so far.

5.1 Altruism within a single species

5.1.1 The biological scenario

We focus on a single species and study the evolution of helping, which is beneficial to patch neighbours but comes at a cost to self. We consider the following life cycle, which is a variant of the example behind eqs. (7). (1) Each adult individual reproduces. 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 (which is measured by γ). This fecundity function $f(\tau_\bullet, \tilde{\tau})$ decreases with the level $\tau_\bullet$ of helping of the focal individual, but increases with the average level $\tilde{\tau}$ of helping among 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 to another patch. (4) All offspring survive to adulthood.

Our assumptions for the life-cycle can be biologically interpreted as individuals competing locally to acquire material resources, but 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 Selection on helping

The selection gradient (eqs. 13–16) on a single isolated trait can be written as $s(\theta) = s_{\mathrm{w}}(\theta) + s_{\mathrm{e}}(\theta)$, where

$$\begin{aligned} s_{\mathrm{w}}(\theta) &= \frac{\partial w(\tau_\bullet, \tilde{\tau}, n)}{\partial \tau_\bullet} + \frac{\partial w(\tau_\bullet, \tilde{\tau}, n)}{\partial \tilde{\tau}} r(\theta) \\ s_{\mathrm{e}}(\theta) &= \frac{\partial w(\tau_\bullet, \tilde{\tau}, n)}{\partial n} \underbrace{\left[1 - (1-m)\frac{\partial F(\bar{\tau}, n)}{\partial n}\right]^{-1}}_{\Lambda} \frac{\partial F(\bar{\tau}, n)}{\partial \bar{\tau}} (1-m)\bar{r}(\theta). \end{aligned} \tag{20}$$

According to our life-cycle assumptions for the model of helping, the fitness of a focal individual is given by

$$w(\tau_\bullet, \tilde{\tau}, n) = \frac{f(\tau_\bullet, \tilde{\tau})}{1 + \gamma n}, \tag{21}$$

and following the argument to derive eq. (7), we find that the population dynamics in the focal patch are given by

$$F(\bar{\tau}, n) = (1-m)\left(\frac{f(\bar{\tau}, \bar{\tau})}{1 + \gamma n}\right) n + mn(\theta). \tag{22}$$

To use the selection gradient (eq. 20), we also need the equilibrium population size $n(\theta)$ in the resident population, which is found by solving $n = F(\theta, n)$ for n from eq. (22). This operation yields

$$n(\theta) = \frac{f(\theta, \theta) - 1}{\gamma}, \tag{23}$$

which further allows us to obtain the pairwise relatedness $r(\theta)$, which is given by eq. (23) substituted into eq. (10), and $\bar{r}(\theta)$, which is obtained from eq. (18).

We first study selection on helping according to its immediate direct and indirect effects (i.e., by looking at $s_{\mathrm{w}}(\theta)$). From eq. (21) and our assumption that helping is individually costly but increases the fecundity of neighbours, the direct and indirect effects of helping are

$$\begin{aligned} \frac{\partial w(\tau_\bullet, \tilde{\tau}, n)}{\partial \tau_\bullet} &= \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} \frac{1}{1 + \gamma n} < 0 \\ \frac{\partial w(\tau_\bullet, \tilde{\tau}, n)}{\partial \tilde{\tau}} &= \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}} \frac{1}{1 + \gamma n} > 0, \end{aligned} \tag{24}$$

negative and positive, respectively. Hence, the helping trait is altruistic *sensu* evolutionary biology (e.g.,

Hamilton, 1964; Rousset, 2004; West and Gardner, 2010). Substituting eqs. (24) into eq. (20) gives

$$s_{\mathrm{w}}(\theta) = \frac{1}{1+\gamma n}\left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}} r(\theta)\right) \tag{25}$$

for the first component of the selection gradient, which shows that relatedness favors altruism.

Second, we study selection on helping due to its effect on demography (i.e., by looking at $s_{\mathrm{e}}(\theta)$). From eqs. (21) and (22), it is immediate that helping increases patch productivity,

$$\frac{\partial F(\bar{\tau}, n)}{\partial \bar{\tau}} > 0, \tag{26}$$

which results in an increase in the equilibrium abundance experienced by mutants in the focal patch (i.e., $\Lambda > 0$, see eq. C.3 of Appendix C), which in turn feedbacks negatively on fitness,

$$\frac{\partial w(\tau_\bullet, \tilde{\tau}, n)}{\partial n} < 0, \tag{27}$$

as greater abundance means increased competition. Overall, we find that the demographic feedback,

$$s_{\mathrm{e}}(\theta) = -\frac{(1-m)^2 \gamma n(\theta)}{\gamma n(\theta) + (2-m)m}\left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}}\right)\bar{r}(\theta) < 0, \tag{28}$$

is negative, provided the fecundity cost of altruism is greater than its benefit (see Appendix C.1 for derivation).

Summing eqs. (25) and eq. (28), we find that the selection gradient is proportional to,

$$\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \kappa(\theta)\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}}, \tag{29}$$

where

$$\kappa(\theta) = \frac{(1-m)^2}{n(\theta) f(\theta, \theta) - (1-m)^2 (n(\theta) - 1)}, \tag{30}$$

is an effective relatedness coefficient, which varies between zero (when $m = 1$) and $\kappa(\theta) = 1/[1 + n(\theta)(f(\theta, \theta) - 1)]$ (when $m = 0$, see Fig. 4). Eq. (29) shows that when dispersal is unlimited ($m = 1$), altruism cannot evolve because individuals never interact with relatives. However, as dispersal decreases ($m \to 0$), altruism is increasingly directed towards relatives and can therefore be favoured, even though it increases competition (eq. 28).

In our to check our analytical predictions with individual-based simulations, we assumed that maximal fecundity is given by

$$f(\tau_\bullet, \tilde{\tau}) = f_{\mathrm{b}} \times (1 + B\tilde{\tau} - C\tau_\bullet^2), \tag{31}$$

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. We find good qualitative agreement between predictions and simulations, as much for the evolution of trait (Fig. 5, top panel) than for the concomitant size of a focal group (Fig. 5, bottom panel).

5.1.3 Connection to previous results on the evolution of altruism

Our finding that the level of altruism decreases as dispersal increases is a standard result of evolutionary biology. However, our model of altruism departs in two crucial ways from the extensive literature on this topic (e.g., Taylor, 1992; Taylor and Irwin, 2000; Gardner and West, 2006; Lehmann et al., 2006b; 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, kin competition strongly inhibits the benefits of interacting among relatives. By contrast, we have assumed here that competition occurs for resources before dispersal among interacting patch members. In this situation, which is biologically plausible, altruism is more moderately inhibited by kin competition because local population size can increase and thus limit the adverse effects of competition (in line with previous results that selection on altruism increases when altruism increases local equilibrium patch size, van Baalen and Rand, 1998; Lehmann et al., 2006b; Lion and Gandon, 2009).

Second, previous models with endogenous patch dynamics have had to rely on extensive numerics to compute the selection gradient on altruism explicitly (Lehmann et al., 2006b; Alizon and Taylor, 2008; Wild et al., 2009; Wild, 2011), making it difficult to understand how selection on altruism varies with demographic parameters (e.g., Lehmann et al., 2006b, eq. 12). Here, we have obtained a simple expression for the selection gradient (eq. 29), which nonetheless fits well with simulation results (Fig. 5). Further, the selection gradient we have derived (eq. 29) applies to any type of social interactions within the life-cycle given in 5.1.1, and can be adjusted to other models by changing the fecundity function eq. (31). Similar expressions to eq. (29) for selection on social behavior, in terms of direct and indirect marginal fecundity effects, have been derived previously for lattice structured populations using pair approximations when competition is for space (Lion and Gandon, 2009, eq. 14, Lion and Gandon, 2010, eq. 19).

Our model here assumes that dispersal is constant. In nature, dispersal probably also evolves, and this evolution influences local demography and selection on social behavior (e.g., Le Galliard et al., 2005; Purcell et al., 2012; Mullon et al., 2016). In order to illustrate how our approximation can take the evolution of dispersal into account, we derived a co-evolutionary model between altruism and dispersal when dispersing offspring survive dispersal with a probability s (see Appendix C.2 for details). We find that dispersal increases as survival during dispersal increases, which then selects for lower altruism

(Fig. 6). Again, our predicted equilibria computed from the approximate selection gradient fit well with stochastic simulations (Fig. 6). This suggests that our proposed approximation of the selection gradient (eqs. 13–16) is not only apt to model dispersal evolution, which has an important influence on local demography, but also the co-evolution of multiple traits.

5.2 Spite between species

5.2.1 The biological scenario

We now study the interaction among two species through the evolution of a trait expressed by species 1, which is costly and harms individuals of species 2, when both species have the following life-cycle. (1) Individuals reproduce. A focal individual of species 1 produces $f_1(\tau_{\bullet,1})/(1+\gamma n_1+\eta n_2)$ offspring, which depends on intra- as well as inter-specific competition (measured by parameters γ and η respectively). 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., $\partial f_1(\tau_{\bullet 1})/\partial\tau_{\bullet 1} < 0$). A focal individual of species 2 produces on average $f_2/(1+\gamma n_2)$ offspring where f_2 is the maximal fecundity of species 2 and γ is the level of intra-specific competition. Only species 1 therefore experiences inter-species competition, which would occur, for instance, because species 2 is a more generalist species than species 1. (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., $\partial D(\bar{\tau}_1)/\partial\bar{\tau}_1 > 0$). (3) All adults of both species die. (4) Surviving offspring of each species disperse with probability m to another patch and all offspring survive to adulthood.

5.2.2 Selection on harming

Under our above assumptions, the fitness of a focal individual from species 1 and from species 2 are

$$\begin{aligned} w_1(\tau_{\bullet,1},\mathbf{n}) &= \frac{f_1(\tau_{\bullet,1})}{1+\gamma n_1+\eta n_2} \\ w_2(\bar{\tau}_1,\mathbf{n}) &= \frac{f_2}{1+\gamma n_2} - D(\bar{\tau}_1)\frac{n_1}{n_2}, \end{aligned} \tag{32}$$

respectively, and the local population dynamics of species 1 and 2 in the focal patch are given by

$$\begin{aligned} F_1(\bar{\tau}_1,\mathbf{n}) &= (1-m)\left(\frac{f_1(\bar{\tau}_1)}{1+\gamma n_1+\eta n_2}\right)n_1 + mn_1(\theta_1) \\ F_2(\bar{\tau}_1,\mathbf{n}) &= (1-m)\left(\frac{f_2}{1+\gamma n_2} - D(\bar{\tau}_1)\frac{n_1}{n_2}\right)n_2 + mn_2(\theta_1). \end{aligned} \tag{33}$$

A general analysis of the selection gradient for this model is more complicated than for altruism within species (see Appendix C.3 for details), but we find that selection for harming can be explained intuitively

as follows. We see from eq. (32), that harming has a direct fitness cost

$$\frac{\partial w_1(\tau_{\bullet,1}, \mathbf{n})}{\partial \tau_{\bullet,1}} < 0, \tag{34}$$

since $\partial f_1(\tau_{\bullet,1})/\partial \tau_{\bullet,1} < 0$, but no immediate indirect fitness effects

$$\frac{\partial w_1(\tau_{\bullet,1}, \mathbf{n})}{\partial \tilde{\tau}_1} = 0. \tag{35}$$

The first component of the selection gradient is therefore negative, $s_{\mathrm{w}}(\theta) < 0$. Harming can then only be favoured if it leads to a positive ecological feedbacks for downstream generations (i.e., $s_{\mathrm{e}}(\theta) > 0$). We find that it is the case when harming decreases the equilibrium abundance of species 2, which alleviates competition for relatives living in downstream generations (Appendix C.3 eq. C.20). In this case, selection for harming, which has a direct cost on recipients of species 2, is driven entirely by multi-generational (indirect) carry-over effects. Such harming behaviour can thus be qualified as spiteful towards species 2 *sensu* evolutionary biology (Lehmann et al., 2006a; West and Gardner, 2010).

Since selection on harming depends on its influence on downstream relatives reaping its benefits, harming should be favoured when dispersal is limited. To test this, we assumed that fecundity in species 1 is given by

$$f_1(\tau_{\bullet,1}) = f_{1,\mathrm{b}} \times \left(1 - C_1 \tau_{\bullet,1}^2\right) \tag{36}$$

where $f_{1,\mathrm{b}}$ is a baseline fecundity in species 1 and C is the individual cost of harming, which increases quadratically according to the investment of the focal into harming. We further assumed that an individual of species 2 which gets harmed has its fecundity decreased by

$$D(\bar{\tau}_1) = \alpha \bar{\tau}_1, \tag{37}$$

where α is a parameter tuning the deleteriousness of harming.

In line with our intuition, we find that the convergence stable level of harming is zero when dispersal is unlimited ($m = 1$, Figure 7, top panel). This is because in that case, the benefit of harming, which is a decrease of inter-species competition for downstream generations, can never go to relatives. However, as dispersal becomes limited, these benefits increasingly go to relatives so that spite evolves, which in turn causes a significant reduction in the abundance of species 2 and an increase of species 1 (Figure 7, bottom panel). We find that the results obtained from our approximation match well those from the individual-based simulations of the exact model (Figure 7), which further supports the goodness of fit of our approximation.

6 Discussion

Most communities of interacting species are spatially structured into meta-communities (e.g., Tilman, 1982). Yet, understanding how selection shapes intra- and inter-species interactions and their influence on community composition has proven challenging when dispersal is limited. Here, we have derived an approximation for the selection gradient on a quantitative trait that influences meta-community dynamics, which neglects ecological stochasticity but takes into account the consequences of genetic stochasticity. We found that it worked very well qualitatively for all models and conditions we studied, but also quantitatively provided dispersal is not too weak (as a rule of thumb, dispersal rate should be no less than 0.1 when patches are small with less than ten individuals, Fig. 1 and Fig. 5-7). The simplicity of our expression for the selection gradient allows to characterize convergence stable species coalitions and intuitively understand long-term multi-species coevolution under limited dispersal.

The selection gradient we derived shows that selection on a trait can be decomposed into direct and indirect fitness effects, where the latter are further decomposed into intra- (eq. 14) and inter-generational effects (eqs. 15–16). Inter-generational effects are important due to the interaction between kin selection and local ecological dynamics: when a focal individual affects local species abundances, these modifications lead to a change in community dynamics due to ecological interactions within and between species, which in turn modifies the conditions of individuals living in downstream generations who potentially carry genes identical-by-descent to the focal (ecological inheritance). The inter-generational effects of a trait thus depend on the extent to which fitness depends on species abundances (eqs. 15), and how the trait influences ecological conditions for individuals living in downstream generations (eq. 16, see Figure 3). As such, our model extends previous models of multi-species evolution in the presence of relatives (Frank, 1994; Foster and Wenseleers, 2006; Wyatt et al., 2013; Akçay, 2017) to incorporating local and global fluctuations in species abundance, including the case when these fluctuations are affected by the evolving traits.

Interestingly, the eco-evolutionary multigenerational feedbacks that emerge under limited dispersal are taken into account by analogues of classical press perturbations of community ecology (eq. 19). Traditionally, press perturbations measure how a persistent change in the abundance of a given species alters the equilibrium abundance of another one due to ecological interactions, possibly through multiple species (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 evolutionary significant for the focal, which is the time the focal's relatives experience it (Figure 3). Ecological inheritance therefore increases as dispersal becomes limited, which ties in the fate of a trait-changing mutation with its carry-over, long-lasting, ecological effects. A broad brush prediction that comes from our analysis is therefore that communities that are strongly limited spatially, like plants, are more likely to carry traits with long-term effects that have shaped by selection.

We applied our model to the evolution of two traits. First, we studied the evolution of altruism within species, a long studied topic in patch-structured populations (e.g., Eshel, 1972; Rogers, 1990; Taylor, 1992; Frank, 1998; Rousset, 2004; Gardner and West, 2006; Lion and Gandon, 2009). By contrast to this literature that usually assumes that competition occurs for space after dispersal, we introduced the notion that competition occurs for material resources before dispersal. In this case, we found that altruism tends to be favored by selection because as altruism increases, patch size also increases (Figure 5). As result, the kin competition effects that usually inhibit the benefits of altruism towards kin (e.g., Taylor, 1992) are mitigated. Our results therefore suggest that altruism is more likely to be found in species in which competition occurs for material resources rather than for space.

Second, we studied the evolution of spite between species. Spite between species occurs when harming an individual from another species comes at a cost to self. We found that spiteful behaviour can evolve when it alleviates inter-species competition for future relatives. Since such evolution depends in multi-generational carry over effects, dispersal must be limited in the two interacting species. Our analysis thus makes the empirical prediction that antagonism is more likely when dispersal is limited (Figure 7). Previous theory has traditionally focused on understanding altruism between species (Frank, 1994; Wyatt et al., 2013), which promotes mutualism. Here, our model reveals that antagonism between species can also emerge, which raises the interesting question of whether mutualistic or antagonistic interactions are more likely to evolve under limited dispersal.

Our approximation of the selection gradient of course cannot be applied to all evolutionary meta-community scenarios and should be supplemented with simulation checks, in particular when dispersal is severely limited and patches are very small, like when populations are close to extinction. Nevertheless, due to its simplicity, it can provide a useful guide to investigate several questions pertaining to how natural selection affects local community dynamics. For instance, it could be applied to study mutualism and host-parasite eco-evolutionary dynamics, or sex-specific local demography under limited dispersal. These and other applications may help to provide a better understanding of how selection moulds intra- and inter-species interactions under limited dipsersal.

Appendix A: Individual-based simulations

A.1 Simulation implementation

Here, we describe how the individual-based simulations for the life cycle given in the section "Life-cycle with stochastic demography" were implemented (using Mathematica 10.0.1.0, Wolfram Research, 2016, the code is available on request).

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 (the exact expression depends on the mode 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) Depending on the underlying life-cycle assumptions, density-dependent competition may occur after dispersal in each patch. If it does, survival follows a Bernoulli trial with mean depending on the number of individuals entering in competition on the patch. (4) A mutation occurs 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 perturbations to the parental phenotypic values that are 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) of each individual and their number in each patch, from which we can evaluate various statistics such as mean population size in each patch, average trait value, and relatedness. To compute relatedness in the neutral model (no natural selection so all vital rates are the same within each species) displayed in Fig. (2), we used the phenotypic values of all individuals and all patches to compute the ratio of the 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, eq. 3.9-3.10).

A.2 Predator-prey model

The predator-prey dynamics shown in the lower panel of Fig. 1 are given by

$$
\begin{aligned}
n_{1,t}(\boldsymbol{\theta}) &= (1-m_1)\left\{\frac{f_1}{1+\gamma n_{1,t-1}(\boldsymbol{\theta})} \times \exp\left(-a n_{2,t-1}(\boldsymbol{\theta})\right)\right\} n_{1,t-1}(\boldsymbol{\theta}) + m_1 n_1(\boldsymbol{\theta}) \\
n_{2,t}(\boldsymbol{\theta}) &= (1-m_2) f_2 \left[1-\exp\Big(-a n_{2,t-1}(\boldsymbol{\theta})\Big)\right] n_{2,t-1}(\boldsymbol{\theta}) + m_2 n_2(\boldsymbol{\theta}),
\end{aligned} \tag{A.1}
$$

where species 1 is the prey and species 2 the predator. Here, $\exp\left(-a n_{2,t-1}(\boldsymbol{\theta})\right)$ can be thought of as the probability that a prey is not caught by a predator per unit of demographic time, where a is the attack rate per predator. Mean fecundity of the prey is then given by the term in curly brackets in the first line of eq. A.1. In the second line of eq. A.1, which gives predator dynamics, the term $1-\exp\left(-a n_{2,t-1}(\boldsymbol{\theta})\right)$ is the probability that a predator gets a prey and f_2 is the conversion of preys into predators. This model (eq. A.1) is very close to the classical Nicholson-Bailey model (Yodzis, 1989), but modified to have density-dependence in the host, and it exhibits prey-predator cycles (see Panel b of Fig. 1).

Appendix B: Evolutionary analysis

In this appendix, we derive the selection gradient on a mutant τ_i in species i (eqs. 13–19 of the main text).

B.1 Allele frequency change and evolutionary invasion analysis

We start by noting that under the full stochastic model (life cycle described in "Life-cycle with stochastic demography"), the selection gradient $s_i(\boldsymbol{\theta})$ can be obtained as the sensitivity of the invasion fitness $\rho_i(\tau_i, \boldsymbol{\theta})$ of a mutant allele τ_i in species i in a resident population $\boldsymbol{\theta}$; namely

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

(Lehmann et al., 2016, Box 2), where 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 the whole mutant lineage descending from a single initial mutant (as long as the mutant remains 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 derive, we will use an invasion fitness *proxy* that is sign equivalent to invasion fitness (e.g., Stearns, 1992; Charlesworth, 1994; Case, 2000; Metz and Gyllenberg, 2001; Lehmann et al., 2016). We will use the basic reproductive number $R_{0,i}(\tau_i, \boldsymbol{\theta})$

for the island model, which is defined as the expected number of successful offspring produced by a randomly sampled lineage member of local mutant lineage of species i (whose individuals carry τ_i) during its (expected) sojourn time in a single patch, when the mutant is overall rare in the population (see Lehmann et al., 2016, eq. 14 for a formal definition of $R_{0,i}(\tau_i, \boldsymbol{\theta})$ and eq. B.3 for a formal definition of $\rho_i(\tau_i, \boldsymbol{\theta})$). Using this invasion proxy here is justified by the observation that

$$s_i(\boldsymbol{\theta}) = \left.\frac{\partial \rho_i(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i=\theta_i} = \left.\frac{\partial R_{0,i}(\tau_i, \boldsymbol{\theta})}{\partial \tau_i}\right|_{\tau_i=\theta_i}, \tag{B.2}$$

where both sensitivities are evaluated in the backdrop of a monomorphic resident population. An algebraic proof of the second equality in eq. B.2 is straightforward but lengthy using Lehmann et al. (2016)'s results. The proof relies on the fact that both derivatives are sensitivities of averages of per capita number of offspring produced and therefore must be equal when the average is evaluated in the backdrop of a monomorphic population.

B.2 Basic reproductive number

We now provide an expression for the basic reproductive number $R_{0,i}(\tau_i, \boldsymbol{\theta})$ when local population dynamics are deterministic and given by eq. (12). To that end, we first let $M_{i,t} \in I_{i,t} = \{0, 1, 2, ..., n_{i,t}\} \subset \mathbb{N}_+$ be a random variable for the number of mutant individuals of species i at generation $t = 0, 1, 2, ...$ in the focal patch where the mutant arose as a single copy a generation $t = 0$; and $\mathbf{n}_t(\boldsymbol{\tau}_i) \in \mathcal{N}_{i,t}$ the random variable for the community vector in the focal patch at time t, where $\mathcal{N}_{i,t} \subset \mathbb{R}_+$ denotes the state space of the community demography at time t when we consider a mutant in species i. We assume that the community vector is bounded (i.e., $n_{i,t} < \infty$ for all i).

The coupled dynamics $\{M_{i,t}, \mathbf{n}_t(\boldsymbol{\tau}_i)\}_{t\geq 0}$ then occur on $\Omega_{i,t} = I_{i,t} \times \mathcal{N}_{i,t}$, which is the state space of the genetic-demographic process in the focal patch at time t, conditional on there being at least one mutant in the patch (conditional on $M_{i,t} \geq 1$). We denote by $\Pr(\boldsymbol{s}_{i,t})$ the probability that the realized state at time t is $\boldsymbol{s}_{i,t} = (M_{i,t}, \mathbf{n}_t(\boldsymbol{\tau}_i)) \in \Omega_{i,t}$.

Next, we observe that in each generation, $M_{i,t}$ has a single absorbing state, which is the local extinction of the lineage ($M_{i,t} = 0$). This is because we have assumed that all offspring have a non-zero dispersal probability and because as long the mutant is rare in the meta-population, there can be no immigration of mutants into the focal group (e.g., Metz and Gyllenberg, 2001; Lehmann et al., 2016). Hence, the lineage of the mutant will eventually go extinct locally $\lim_{t\to\infty} \Pr(M_{i,t} = 0) = 1$, and as a consequence, $\lim_{t\to\infty} \Pr(\boldsymbol{s}_{i,t} = 0) = 1$ (for a product space like $\Omega_{i,t}$, the finite dimensional distribution $\Pr(\boldsymbol{s}_{i,t})$ is defined for any realization $\boldsymbol{s}_{i,t} \in \Omega_{i,t}$, and it can be extended to events in $\Omega_i = \lim_{t\to\infty} \Omega_{i,t}$, Meyn and Tweedie, 2009, p. 60).

With this in mind, we can write the basic reproductive number as

$$R_{0,i}(\tau_i, \boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} w_i(\tau_i, \boldsymbol{s}_t) q_i(\boldsymbol{s}_t), \tag{B.3}$$

where $w_i(\tau_i, \boldsymbol{s}_t)$ 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 self through survival) of species i that carries mutant phenotype τ_i when the patch is in genetic-demographic $\boldsymbol{s}_t$ at time t, and

$$q_i(\boldsymbol{s}_t) = \frac{M_{i,t} \Pr(\boldsymbol{s}_t)}{N_{\mathrm{L}}(\tau_i, \boldsymbol{\theta})}, \tag{B.4}$$

is the probability that a randomly sampled mutant lineage member is sampled exactly at time t on realization $\boldsymbol{s}_t$ of the stochastic process, where $N_{\mathrm{L}}(\tau, \theta)$ is the expected total size of the mutant lineage over its lifetime in a single patch: $N_{\mathrm{L}}(\tau_i, \boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} M_{i,t} \Pr(\boldsymbol{s}_t)$. 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 sum in eq. B.3 always converges.

B.3 First-order condition for local uninvadability

We now use eq. (B.3) to derive the selection gradient eq. (13) presented in the main text. First, we write individual fitness explicitly in terms of the genetic and demographic processes as

$$w_i(\tau_i, \boldsymbol{s}_t) = w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t), \tag{B.5}$$

where $\tilde{\boldsymbol{\tau}}_{i,t} = (\theta_1, ..., \theta_{i-1}, \tilde{\tau}_{i,t}, \theta_{i+1}, ..., \theta_S)$ is the vector of average phenotype among the neighbours of the focal of individual, where

$$\tilde{\tau}_{i,t} = \frac{n_{i,t}(\boldsymbol{\tau}_i)\bar{\tau}_{i,t} - \tau_i}{n_{i,t}(\boldsymbol{\tau}_i) - 1}, \tag{B.6}$$

is the average phenotype among the neighbours of species i of the focal of individual (thus excluding the focal individual). Eqs. (11) and (B.6) give the relationship between the genetic and demographic processes $\boldsymbol{s}_t = (M_{i,t}, \mathbf{n}_t(\boldsymbol{\tau}_i))$ and the average phenotypic vector $\tilde{\boldsymbol{\tau}}_{i,t}$.

Second, we assume that types are real-valued ($\Theta_i = \mathbb{R}$). Substituting eq (B.5) into (B.3), which is in turn substituted into (B.2), we obtain using the chain rule

$$\begin{aligned} s_i(\boldsymbol{\theta}) &= \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \left[\frac{\partial w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t)}{\partial \tau_i} q_i^{\circ}(\boldsymbol{s}_t) + \left(w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t)|_{\tau_i = \theta_i} \right) \frac{\partial q_{i,t}(\boldsymbol{s}_t)}{\partial \tau_i} \right] \\ &= \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \frac{\partial w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t)}{\partial \tau_i} q_i^{\circ}(\boldsymbol{s}_t), \end{aligned} \tag{B.7}$$

where here and throughout all derivatives will be evaluated at the resident value $\boldsymbol{\theta}$ and $q_i^\circ(\boldsymbol{s}_t)$ denote the genetic-demographic state distribution evaluated in the resident population (i.e., in a monomorphic $\boldsymbol{\theta}$ population). The second line in eq. (B.7) follows from two observations: (1) in a resident population at its demographic equilibrium individual fitness is one, i.e., $w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t)|_{\tau_i=\theta_i} = 1$; and (2)

$$\sum_{\boldsymbol{s}_t \in \Omega_t} \frac{\partial q_i(\boldsymbol{s}_t)}{\partial \tau_i} = \frac{\partial}{\partial \tau_i}\left[\sum_{\boldsymbol{s}_t \in \Omega_t} q_i(\boldsymbol{s}_t)\right] = \frac{\partial}{\partial \tau_i}(1) = 0, \tag{B.8}$$

since $q_i(\boldsymbol{s}_t)$ is a probability density function over Ω_t.

Inserting eq. (11) and eq. (B.6) into $w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t)$, which is in turn derived with respect to τ_i, we obtain using the chain rule,

$$\frac{\partial w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t)}{\partial \tau_i} = \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \tilde{\tau}_i} \underbrace{\left(\frac{M_{i,t}-1}{n_{i,t}-1}\right)}_{\partial \tilde{\tau}_{i,t}/\partial \tau_i} + \sum_{j=1}^{S} \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial n_{j,t}} \frac{\partial n_{j,t}}{\partial \tau_i}, \tag{B.9}$$

where $w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t) = w_i(\tau_i, \tilde{\boldsymbol{\tau}}_{i,t}, \mathbf{n}_t)$ but it now refers to 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 neighbors of the focal individual (see also Mullon et al., 2016, Appendix C, pp. 5-6). The term $(M_{i,t}-1)/(n_{i,t}-1)$ is obtained by evaluating $\partial \tilde{\tau}_{i,t}/\partial \tau_i$ using eq. (B.6) and eq. (11). Substituting eq. (B.9) into (B.7), gives

$$s_i(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \left[\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \tilde{\tau}_i}\left(\frac{M_{i,t}-1}{n_{i,t}-1}\right) + \sum_{j=1}^{S} \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial n_{j,t}} \frac{\partial n_{j,t}}{\partial \tau_i}\right] q_i^\circ(\boldsymbol{s}_t), \tag{B.10}$$

for the selection gradient. Next, we decompose eq. (B.10) as $s_i(\boldsymbol{\theta}) = s_{\mathrm{w},i}(\boldsymbol{\theta}) + s_{\mathrm{e},i}(\boldsymbol{\theta})$, where

$$s_{\mathrm{w},i}(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \left[\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \tilde{\tau}_i}\left(\frac{M_{i,t}-1}{n_{i,t}-1}\right)\right] q_i^\circ(\boldsymbol{s}_t) \tag{B.11}$$

$$s_{\mathrm{e},i}(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \left[\sum_{j=1}^{S} \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial n_{j,t}} \frac{\partial n_{j,t}}{\partial \tau_i}\right] q_i^\circ(\boldsymbol{s}_t). \tag{B.12}$$

We will now show that $s_{\mathrm{w},i}(\boldsymbol{\theta})$ as defined by eq. (B.11) reads as eq. (14) of the main text, and $s_{\mathrm{e},i}(\boldsymbol{\theta})$ as defined by eq. (B.12) reads as eq. (15) with (16) of the main text.

B.3.1 Standard direct and direct effect $s_{\mathrm{w},i}(\boldsymbol{\theta})$

To show that eq (B.11) equals eq. (14), we first note that since the derivatives that appear in eq (B.11) and the distribution $q_i^\circ(\boldsymbol{s}_t)$ are evaluated in a monomorphic $\boldsymbol{\theta}$ population, the local community vector $\mathbf{n}_t$ in eq. (B.11) is at its single fixed point $\mathbf{n}(\boldsymbol{\theta})$ (which satisfies eq. 5 of the main text). Substituting $\mathbf{n}_t = \mathbf{n}$ into eq (B.11), we have

$$s_{\mathrm{w},i}(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \left[\frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n})}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n})}{\partial \tilde{\tau}_i} \left(\frac{M_{i,t} - 1}{n_i - 1} \right) \right] q_i^\circ(\boldsymbol{s}_t). \tag{B.13}$$

Using the fact that $\sum_{\boldsymbol{s}_t \in \Omega_t} q_i^\circ(\boldsymbol{s}_t) = 1$, we then have

$$s_{\mathrm{w},i}(\boldsymbol{\theta}) = \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n})}{\partial \tau_{\bullet,i}} + \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n})}{\partial \tilde{\tau}_i} r_i(\boldsymbol{\theta}) \tag{B.14}$$

where by definition (eq. B.4),

$$r_i(\boldsymbol{\theta}) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \frac{M_{i,t} - 1}{n_i - 1} q_i^\circ(\boldsymbol{s}_t) = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \frac{M_{i,t} - 1}{n_i(\boldsymbol{\theta}) - 1} q_i^\circ(\boldsymbol{s}_t) \tag{B.15}$$

is the relatedness between two individuals sampled without replacement from the same patch in the generation (i.e., the probability that two randomly sampled individuals from species i in the same patch carry an allele identical-by-descent). As required, eq. (B.14) corresponds to eq. (14).

B.3.2 Ecological feedback effet $s_{\mathrm{e},i}(\boldsymbol{\theta})$

We now show that eq. (B.12) is equivalent to eq. (15) with (16). We first use the fact that the derivatives are evaluated in a monomorphic $\boldsymbol{\theta}$ population (in which the local community vector is at its single fixed point $\mathbf{n}(\boldsymbol{\theta})$) to write eq. (B.12) as eq. (15) of the main text,

$$s_{\mathrm{e},i}(\boldsymbol{\theta}) = \sum_{j=1}^{S} \frac{\partial w_i(\tau_{\bullet,i}, \tilde{\boldsymbol{\tau}}_i, \mathbf{n})}{\partial n_j} \frac{\partial \overline{n}_j}{\partial \tau_i}, \tag{B.16}$$

where

$$\overline{n}_j = \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} n_{j,t} q_i^\circ(\boldsymbol{s}_t) \tag{B.17}$$

is the expected abundance of species j experienced by a mutant of species i, which is randomly sampled from its local lineage under neutrality (i.e., from $q_i^\circ(\boldsymbol{s}_t)$). Next, we seek an expression for the term $\partial n_{j,t}/\partial \tau_i$, which appears in eq. (B.16), and which is the derivative of the number of individuals in species j in the focal patch at time t with respect to mutant trait variation in species i. To obtain this,

we differentiate both sides of $n_{j,t}(\boldsymbol{\tau}_i) = F_j(\bar{\boldsymbol{\tau}}_{i,t-1}, \mathbf{n}_{t-1}(\boldsymbol{\tau}_i))$ (i.e., eq. 12 of the main text) with respect to τ_i, which in vector form reads as

$$\frac{\partial \mathbf{n}_t}{\partial \tau_i} = \mathbf{f}_{t-1}(\boldsymbol{\theta}) \frac{M_{i,t-1}}{n_{i,t-1}} + \mathbf{C}(\boldsymbol{\theta}) \frac{\partial \mathbf{n}_{t-1}}{\partial \tau_i}, \tag{B.18}$$

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

$$\mathbf{f}_t(\boldsymbol{\theta}) = \left(\frac{\partial F_1(\bar{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \bar{\tau}_i}, \frac{\partial F_2(\bar{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \bar{\tau}_i}, ..., \frac{\partial F_S(\bar{\boldsymbol{\tau}}_i, \mathbf{n}_t)}{\partial \bar{\tau}_i} \right) \tag{B.19}$$

is a vector of S elements (all evaluated at $\boldsymbol{\theta}$) with $\bar{\boldsymbol{\tau}}_i = (\theta_1, ..., \theta_{i-1}, \bar{\tau}_i, \theta_{i+1}, ..., \theta_S)$ where $\bar{\tau}_i$ is the average phenotype in the focal patch.

Note that eq. (B.18) is a linear recurrence equation in $\partial \mathbf{n}_t / \partial \tau_i$ with initial condition $\partial \mathbf{n}_0 / \partial \tau_i = 0$. As the solution to the linear vector recurrence $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{b}_{t-1}$ is $\mathbf{x}_t = \mathbf{A}^t \mathbf{x}_0 + \sum_{h=0}^{t-1} \mathbf{A}^{t-h-1} \mathbf{b}_h$, e.g., (Sydsaeter et al., 2008) the solution of eq. (B.18) is

$$\frac{\partial \mathbf{n}_t}{\partial \tau_i} = \sum_{h=0}^{t-1} \mathbf{C}(\boldsymbol{\theta})^{t-h-1} \mathbf{f}_h(\boldsymbol{\theta}) \frac{M_{i,h}}{n_{i,h}} = \sum_{h=1}^{t} \mathbf{C}(\boldsymbol{\theta})^{h-1} \mathbf{f}_{t-h}(\boldsymbol{\theta}) \frac{M_{i,t-h}}{n_{i,t-h}}. \tag{B.20}$$

Integrating eq. (B.20) over all time and all sample paths gives

$$\begin{aligned} \sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \frac{\partial \mathbf{n}_t}{\partial \tau_i} q_i^{\circ}(\boldsymbol{s}_t) &= \sum_{t=0}^{\infty} \sum_{h=1}^{t} \sum_{\boldsymbol{s}_t \in \Omega_t} \mathbf{C}(\boldsymbol{\theta})^{h-1} \mathbf{f}_{t-h}(\boldsymbol{\theta}) \frac{M_{i,t-h}}{n_{i,t-h}} q_i^{\circ}(\boldsymbol{s}_t) \\ &= \sum_{h=1}^{\infty} \mathbf{C}(\boldsymbol{\theta})^{h-1} \left[\sum_{t=h}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \mathbf{f}_{t-h}(\boldsymbol{\theta}) \frac{M_{i,t-h}}{n_{i,t-h}} q_i^{\circ}(\boldsymbol{s}_t) \right], \end{aligned} \tag{B.21}$$

where the second line is obtained by exchanging dummy variables. We use again the fact that the distribution $q_i^{\circ}(\boldsymbol{s}_t)$ is evaluated in a monomorphic $\boldsymbol{\theta}$ population, so that the local community vector $\mathbf{n}_t$ in eq. (B.21) is at its single fixed point $\mathbf{n}(\boldsymbol{\theta})$. In which case we can substitute $n_{i,t} = n_i(\boldsymbol{\theta})$ and $\mathbf{f}_t(\boldsymbol{\theta}) = \mathbf{f}(\boldsymbol{\theta})$ into eq. (B.21) and obtain

$$\sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \frac{\partial \mathbf{n}_t}{\partial \tau_i} q^{\circ}(\boldsymbol{s}_t) = \sum_{h=1}^{\infty} \mathbf{C}(\boldsymbol{\theta})^{h-1} \mathbf{f}(\boldsymbol{\theta})\, r_{i,h}(\boldsymbol{\theta}), \tag{B.22}$$

where

$$r_{i,h}(\boldsymbol{\theta}) = \sum_{t=h}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \frac{M_{i,t-h}}{n_i(\boldsymbol{\theta})} q_i^{\circ}(\boldsymbol{s}_t), \tag{B.23}$$

is the relatedness between two individuals of species i sampled at h generations apart in a monomorphic resident population ($r_{i,h}(\boldsymbol{\theta})$ sums the probability of sampling two mutants at h generations apart over the whole sojourn time of the lineage in a single group, see eq. B.4). Such relatedness coefficient can be computed using standard identity-by-descent arguments (Malécot, 1973; Epperson, 1999; Lehmann,

2008), and for $h \geq 1$, it takes the form

$$r_{i,h}(\boldsymbol{\theta}) = (1 - m_i(\boldsymbol{\theta}))^h \bar{r}_i(\boldsymbol{\theta}), \tag{B.24}$$

where $m_i(\boldsymbol{\theta})$, which is called the the backward dispersal rate, is the probability that a randomly sampled individual in a patch descends from that patch in the previous generation (in the presence of overlapping generation, $m_i(\boldsymbol{\theta})$ will depend on the survival of individuals) and $\bar{r}_i(\boldsymbol{\theta})$ is the relatedness between two individuals sampled with replacement in the same patch (see eq. 18).

Substituting eq. (B.24) into eq. (B.22), we obtain

$$\begin{aligned}\sum_{t=0}^{\infty} \sum_{\boldsymbol{s}_t \in \Omega_t} \frac{\partial \mathbf{n}_t}{\partial \tau_i} q^{\circ}(\boldsymbol{s}_t) &= \sum_{h=1}^{\infty} \mathbf{C}(\boldsymbol{\theta})^{h-1} (1 - m_i(\boldsymbol{\theta}))^{h-1} \mathbf{f}(\boldsymbol{\theta})(1 - m_i(\boldsymbol{\theta}))\bar{r}_i(\boldsymbol{\theta}) \\ &= \boldsymbol{\Lambda}_i(\boldsymbol{\theta}) \mathbf{f}(\boldsymbol{\theta})(1 - m_i(\boldsymbol{\theta}))\bar{r}_i(\boldsymbol{\theta}),\end{aligned} \tag{B.25}$$

where

$$\boldsymbol{\Lambda}_i(\boldsymbol{\theta}) = [\mathbf{I} - (1 - m_i(\boldsymbol{\theta}))\mathbf{C}(\boldsymbol{\theta})]^{-1} \tag{B.26}$$

comes from the standard matrix summation result $\sum_{h=1}^{\infty} \mathbf{A}^{h-1} = (\mathbf{I} - \mathbf{A})^{-1}$ (Sydsaeter et al., 2008). Using eq. (B.17), the component j of eq. (B.25) then is

$$\frac{\partial \overline{n}_j}{\partial \tau_i} = \left[\sum_{k=1}^{S} \Lambda_{jk,i}(\boldsymbol{\theta}) \frac{\partial F_k(\bar{\boldsymbol{\tau}}_i, \mathbf{n})}{\partial \bar{\tau}_i} \right] (1 - m_i(\boldsymbol{\theta}))\bar{r}_i(\boldsymbol{\theta}), \tag{B.27}$$

where $\Lambda_{jk,i}(\boldsymbol{\theta})$ is the jk element of the matrix $\boldsymbol{\Lambda}_i(\boldsymbol{\theta})$. Finally, substituting eq. (B.27) into (B.16), we obtain

$$s_{\mathrm{e},i}(\boldsymbol{\theta}) = \sum_{j=1}^{S} \frac{\partial w_i(\tau_{\bullet,i}, \boldsymbol{\tau}_{0,i}, \mathbf{n})}{\partial n_j} \left[\sum_{k=1}^{S} \frac{\partial F_k(\bar{\boldsymbol{\tau}}_i, \mathbf{n})}{\partial \bar{\tau}_i} (1 - m_i(\boldsymbol{\theta}))\bar{r}_i(\boldsymbol{\theta}) \Lambda_{jk,i}(\boldsymbol{\theta}) \right], \tag{B.28}$$

which is equivalent to eq. (15) with eqs. (16)–(17) of the main text, as required.

Appendix C: Applications

C.1 Altruism within species

We here derive the first-order condition for the model for the evolution of altruism within species (eq. 29). The different components of the selection gradient for this model (eq. 20) are obtained by using the expressions for individual fitness (eq. 21) and local population dynamics (eq. 22).

The resulting direct and indirect phenotypic effects are given by eq. (24) of the main text, while the change in patch size on fitness is

$$\frac{\partial w(\tau_\bullet, \tilde{\tau}, n)}{\partial n} = -\frac{f(\theta,\theta)\gamma}{(1+\gamma n)^2}. \tag{C.1}$$

This term is negative, and therefore reflect a net cost owing to the fact that increasing the number of competitors decreases one's share of resources. The effects on population dynamic is

$$\frac{\partial F(\bar{\tau}, n)}{\partial \bar{\tau}} = \frac{(1-m)\gamma}{(1+\gamma n)}\left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}}\right) \tag{C.2}$$

and the eco-evolutionary feedback term is

$$\Lambda = \left[1 - (1-m)\frac{\partial F(\bar{\tau}, n)}{\partial n}\right]^{-1} = \frac{f(\theta,\theta)}{f(\theta,\theta) - (1-m)^2}. \tag{C.3}$$

Substituting eqs. (C.1)–(C.3) into the second line of eq. (20), we obtain that

$$s_{\mathrm{e}}(\theta) = -\frac{(1-m)^2\gamma n}{\gamma n + (2-m)m}\left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}}\right)\bar{r}(\theta), \tag{C.4}$$

which is negative, as long as the evolving trait increases patch productivity.

Finally, substituting eq. (24), eq. (C.3), and eq. (9) into the full selection gradient (eq. 20), and rearranging produces

$$s(\theta) = \left(\frac{(\gamma+1)(1-m)^2 + f(\theta,\theta)\left(f(\theta,\theta) + (2-m)m - 2\right)}{(f(\theta,\theta)-1)f(\theta,\theta)\left(f(\theta,\theta) - (1-m)^2\right)}\right)\left(\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tau_\bullet} + \kappa(\theta)\frac{\partial f(\tau_\bullet, \tilde{\tau})}{\partial \tilde{\tau}}\right), \tag{C.5}$$

where the term in the first parenthesis is a constant of proportionality, which does not affect the direction of selection, and is positive if $f(\theta,\theta) - 1 > 0$; a necessary condition for $n(\theta) > 0$ to be asymptotically stable. The term in the second parenthesis in $s(\theta)$ determines the direction of selection on helping (it yields eq. 29 of the main text) and depends on the effective relatedness

$$\kappa(\theta) = \frac{(1-m)^2}{n(\theta)f(\theta,\theta) - (1-m)^2(n(\theta)-1)}. \tag{C.6}$$

C.2 Coevolution of altruism and dispersal within species

We here work out an example where both altruism and dispersal co-evolve within a single species. While this could be done with the model of competition before dispersal (eq. 21), it can then be shown that dispersal will evolve to zero, owing to the fact that it does not alleviate kin competition between individuals. Since kin competition avoidance is the main driver of dispersal in the absence of environmental

extinction and inbreeding avoidance (Hamilton and May, 1977; Clobert et al., 2001), individuals remain philopatric.

In order to allow for the evolution of non-zero dispersal, we assume a semelparous life-cycle where adults interact socially (by way of helping each other) and reproduce (without any density-dependent competition) and die after reproduction. Offspring disperse according to the dispersal trait they have inherited, and survive dispersal with probability s. After dispersal, each offspring arriving in a patch, survives density-dependent regulation with probability $1/(1+\gamma J)$ where J is the number of juveniles entering in competition in a patch. Hence, this model considers that density-dependent competition occurs after dispersal [in stage (4) of the life-cycle described in "Life-cycle with stochastic demography", instead of occurring in stage (2)].

Since we have two traits, we let the phenotype of a focal individual be $\tau_\bullet = (z_\bullet, d_\bullet)$, where $z_\bullet$ is its level of altruism, while $d_\bullet$ is the dispersal probability of each of its offspring. Likewise, we let $\tilde{\tau} = (\tilde{z}, \tilde{d})$, where $\tilde{z}$ is the average level of helping among the neighbours of the focal, and $\tilde{d}$ is the dispersal probability of any of their offspring. For these life-cycle assumption, the special case of the evolution of dispersal (holding z constant) was investigated in Rousset and Ronce (2004), while the special case for the evolution of altruism in a linear public goods game (holding d constant) was investigated in Lehmann et al. (2006b), where these two works used the "exact" selection gradient of the model (Rousset and Ronce, 2004, eqs. 26-27, Lehmann et al., 2006b, eqs. A.18-19), which requires heavy numerics.

In order to evaluate the selection gradient for the approximate model but coevolutionary model, we need the fitness of a focal individual when local population size dynamic is given by eq. 11, which is

$$w(\tau_\bullet, \tilde{\tau}, n) = F(\bar{\tau}, n) \times \left(\frac{(1-d_\bullet) f(z_\bullet, \tilde{z})}{(1-\bar{d}) f(\bar{z}, \bar{z}) n + ds f(z,z) n(\theta)} \right) + n(\theta) \times \left(\frac{d_\bullet s f(z_\bullet, \tilde{z})}{(1-d+ds)\, f(z,z) n(\theta)} \right), \quad \text{(C.7)}$$

where $f(z_\bullet, \tilde{z})$ is the fecundity in the absence of density-dependent competition, 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, which for definiteness is assumed to be given by eq. (31). The term in the first line of eq. (C.7) is the philopatric component of fitness, while the second line in eq. (C.7) gives the dispersal component of fitness.

In order to close the model, we need an expression for $F(\bar{\tau}, n)$, 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 demographic equilibrium. To compute this, we use the shorthand notation

$$\lambda(\bar{\tau}) = (1-\bar{d}) f(\bar{z}, \bar{z}) n + ds f(z,z) n(\theta) \quad \text{(C.8)}$$

to denote the expected number of offspring entering in competition in the focal patch after dispersal. The (random) number of offspring J entering in competition in the focal patch is Poisson distributed with mean λ, and each of which survives competition with probability $1/(1+\gamma J)$. Hence, $F(\bar{\tau}, n)$ is obtained as the mean of the convolution of a Poisson distribution with parameter λ and a Binomial distribution with parameters J and $1/(1+\gamma J)$, which is a complicated expression, but is well approximated by

$$F(\bar{\tau}, n) \approx \frac{\lambda(\bar{\tau})}{1+\gamma\lambda(\bar{\tau})}. \tag{C.9}$$

This can be obtained as first-order delta approximation to the mean (Lynch and Walsh, 1998). Eq. (C.9) implies that the equilibrium patch size in a resident population [obtained by solving $n = F(\theta, n)$ for n where $\theta = (d, z)$] is

$$n(\theta) = \frac{1 - \dfrac{1}{(1-d+ds)\, f(z,z)}}{\gamma}, \tag{C.10}$$

which depends on the level of dispersal d and altruism z in the resident population.

Using eq. (C.7) and we can compute the selection gradient on dispersal, denoted $s_{\mathrm{d}}(\theta)$ (where the subscript means that all derivatives in the gradient are taken with respect to trait d) and on altruism, denoted $s_{\mathrm{z}}(\theta)$ (where the subscript means that all derivatives in the gradient are taken with respect to trait z).

Singular strategies are then found by setting

$$s_{\mathrm{d}}(\theta^*) = 0 \qquad \text{and} \qquad s_{\mathrm{z}}(\theta^*) = 0, \tag{C.11}$$

and solving numerically for the equilibrium pair (d^*, z^*) subject to the constraint on patch size, eq. (C.10). The so obtained singular strategies for specific parameter values are graphed in Fig. (6), which shows that the analytical predictions match qualitatively and quantitatively those of individual-based simulations.

C.3 Spite between species species

We here present the first-order condition for the model of the evolution of harming. According to this model (eqs. 32–33), in which a single trait in species 1 is evolving, the equilibrium resident demography satisfies

$$\begin{aligned} n_1(\theta_1) &= (1-m)\left(\frac{f_{1,\mathrm{b}}\,(1-C_1(\theta_1))}{1+\gamma n_1(\theta_1)+\eta n_2(\theta_1)}\right) n_1(\theta_1) + m n_1(\theta_1) \\ n_2(\theta_1) &= (1-m)\left(\frac{f_2}{1+\gamma n_2(\theta_1)} - D(\theta_1)\frac{n_1(\theta_1)}{n_2(\theta_1)}\right) n_2(\theta_1) + m n_2(\theta_1). \end{aligned} \tag{C.12}$$

The explicit solutions to these equations are too cumbersome to give here (a Mathematica notebook is available on request), but one can check that there is only a single valid interior point (with $n_1(\theta_1) > 0$ and $n_2(\theta_1) > 0$).

From eqs. (13)–(16) of the main text and using the fact that the fitness function for species 1 depends only on the phenotype of the actor (eq. 32), the selection gradient on harming in species 1 is

$$s_1(\theta_1) = \underbrace{\frac{\partial w_1(\tau_{\bullet,1}, \mathbf{n})}{\partial \tau_{\bullet,1}}(\theta)}_{s_{\mathrm{w},1}} + \underbrace{\left[\sum_{k=1}^{2}\sum_{j=1}^{2} \frac{\partial w_1(\tau_{\bullet,1}, \mathbf{n})}{\partial n_k} \Lambda_{kj}(\theta_1) \frac{\partial F_j(\bar{\tau}_1, \mathbf{n})}{\partial \bar{\tau}_1}\right] (1-m_1)\bar{r}_1(\theta_1)}_{s_{\mathrm{e},1}(\theta)}, \tag{C.13}$$

where the Λ_{kj}'s refer to those induced by species 1, since we have only one trait evolving (i.e., $\Lambda_{kj} = \Lambda_{kj,1}$) and $\mathbf{n} = (n_1, n_2)$.

From the perspective of a focal individual in species 1 in which the trait is evolving, expressing harming results in spite between species, since for a focal actor we have from eq. (32) that

$$\frac{\partial w_1(\tau_{\bullet,1}, \mathbf{n})}{\partial \tau_{\bullet,1}} = -\frac{C_1'(\tau_1)}{1 + \gamma n_1 + \eta n_2} < 0 \tag{C.14}$$

is a direct fitness cost, and this hurts individuals of species 2, since from eq. (32) that

$$\frac{\partial w_2(\bar{\tau}_1, \mathbf{n})}{\partial \bar{\tau}_1} = -\frac{\partial D(\bar{\tau}_1)}{\partial \bar{\tau}_1}\frac{n_1}{n_2} < 0. \tag{C.15}$$

The demographic effect of expressing harming is to reduce the abundance of each species locally, namely

$$\begin{aligned} \frac{F_1(\bar{\tau}_1, \mathbf{n})}{\partial \bar{\tau}_1} &= -\frac{(1-m) f_{1,\mathrm{b}} C_1'(\tau_1) n_1}{(1 + \gamma n_1 + \eta n_2)} < 0 \\ \frac{F_2(\bar{\tau}_1, \mathbf{n})}{\partial \bar{\tau}_1} &= -(1-m) n_2 \frac{\partial D(\bar{\tau}_1)}{\partial \bar{\tau}_1} < 0, \end{aligned} \tag{C.16}$$

and the changed in abundance of each species affect the fitness of species 1 according to

$$\begin{aligned} \frac{\partial w_1(\tau_{\bullet,1}, \mathbf{n})}{\partial n_1} &= -\frac{\gamma f_{1,\mathrm{b}}[1 - C_1(\theta_1)]}{(1 + \gamma n_1 + \eta n_2)^2} < 0 \\ \frac{\partial w_1(\tau_{\bullet,1}, \mathbf{n})}{\partial n_2} &= -\frac{\eta f_{1,\mathrm{b}}[1 - C_1(\theta_1)]}{(1 + \gamma n_1 + \eta n_2)^2} < 0. \end{aligned} \tag{C.17}$$

In order to understand the extent to which changes in local demography will feedback on the fitness of individuals living relatives, we need to to evaluate the Λ_{kj}'s appearing in eq. (C.14). Using eq. (6) and

eq. (19) along with eq. (C.16), these are

$$\begin{aligned}
\Lambda_{11} &= \left(1 - \frac{f_2(1-m)^2}{(1+\gamma n_2)^2}\right)\frac{1}{X} \\
\Lambda_{12} &= -\left(\frac{f_{1,\mathrm{b}}(1-C_1(\theta_1))(1-m)^2\xi n_1}{(1+\gamma n_1+\xi n_2)^2}\right)\frac{1}{X} \\
\Lambda_{21} &= -\frac{D(\theta_1)(1-m)^2}{X} \\
\Lambda_{22} &= \left(1 - \frac{f_{1,\mathrm{b}}(1-C_1(\theta_1))(1-m)^2(1+\xi n_2)}{(1+\gamma n_1+\xi n_2)^2}\right)\frac{1}{X},
\end{aligned} \tag{C.18}$$

where

$$\begin{aligned}
X = \left(1 - \frac{f_2(1-m)^2}{(1+\gamma n_2)^2}\right)&\left(1 - \frac{f_{1,\mathrm{b}}(1-C_1(\theta_1))(1-m)^2(1+\xi n_2)}{(1+\gamma n_1+\xi n_2)^2}\right) \\
&- \frac{f_{1,\mathrm{b}}(1-C_1(\theta_1))D(\theta_1)(1-m)^4\xi n_1}{(1+\gamma n_1+\xi n_2)^2}.
\end{aligned} \tag{C.19}$$

Further insights into these expressions can be brought by approximating them when m is close to one,

$$\begin{aligned}
\Lambda_{11} &= 1 + (1-m)^2\frac{f_{1,\mathrm{b}}(1+\xi n_2)(1-C_1(\theta_1))}{(1+\gamma n_1+\xi n_2)^2} + O\left((1-m)^4\right) \\
\Lambda_{12} &= -(1-m)^2\frac{f_{1,\mathrm{b}}\xi n_1(1-C_1(\theta_1))}{(1+\gamma n_1+\xi n_2)^2} + O\left((1-m)^4\right) \\
\Lambda_{21} &= -(1-m)^2 D(\theta_1) + O\left((1-m)^4\right) \\
\Lambda_{22} &= 1 + (1-m)^2\frac{f_2}{(1+\gamma n_2)^2} + O\left((1-m)^4\right),
\end{aligned} \tag{C.20}$$

which says that a constant increase in the number of individuals in one species increases the abundance of that species ($\Lambda_{11} > 0$ and $\Lambda_{22} > 0$), but decreases the abundance of the other ($\Lambda_{12} < 0$ and $\Lambda_{21} < 0$).

The sensitivities (eqs. C.12–C.16) then imply that

$$\begin{aligned}
\frac{\partial w_1(\tau_{\bullet,1},\mathbf{n})}{\partial n_1}\Lambda_{11}\frac{F_1(\bar{\tau}_1,\mathbf{n})}{\partial\bar{\tau}_1} &> 0 \\
\frac{\partial w_1(\tau_{\bullet,1},\mathbf{n})}{\partial n_1}\Lambda_{12}\frac{F_2(\bar{\tau}_1,\mathbf{n})}{\partial\bar{\tau}_1} &< 0 \\
\frac{\partial w_1(\tau_{\bullet,1},\mathbf{n})}{\partial n_2}\Lambda_{21}\frac{F_1(\bar{\tau}_1,\mathbf{n})}{\partial\bar{\tau}_1} &< 0 \\
\frac{\partial w_1(\tau_{\bullet,1},\mathbf{n})}{\partial n_2}\Lambda_{22}\frac{F_2(\bar{\tau}_1,\mathbf{n})}{\partial\bar{\tau}_1} &> 0.
\end{aligned} \tag{C.21}$$

Owing to the complexity of this model, the conditions favoring harming cannot be solved analytically, but substituting eq. (C.12), eq. (C.16), and eq. (C.18) into eq. (C.13), the singular strategies can be

found numerically by setting

$$s_1(\theta_1^*) = 0, \tag{C.22}$$

and solving for the θ_1^* subject to the constraint eq. (C.12). Results of this procedure for specific parameter values are graphed in Fig. (7), which displays the consistency between the predictions obtained from the approximate selection gradients and the qualitative and quantitative observed outcomes in the exact individual based simulations.

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.

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.

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.

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.

Dèbarre, 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. 1972. On the neighbor effect and the evolution of altruistic traits. *Theoretical Population Biology* 11:258–277.

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.

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.

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.

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

Lehmann, L., K. Bargum, and M. Reuter. 2006a. An evolutionary analysis of the relationship between spite and altruism. *Journal of Evolutionary Biology* 19:1507–1523.

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. 2006b. Population demography and the evolution of helping behaviors. *Evolution* 60:1137–1151.

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.

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.

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

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

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

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.

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.

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

Rogers, A. R. 1990. Group selection by selective emigration: the effects of migration and kin structure. *American Naturalist* 135:398–413.

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.

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.

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.

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.

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.

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.

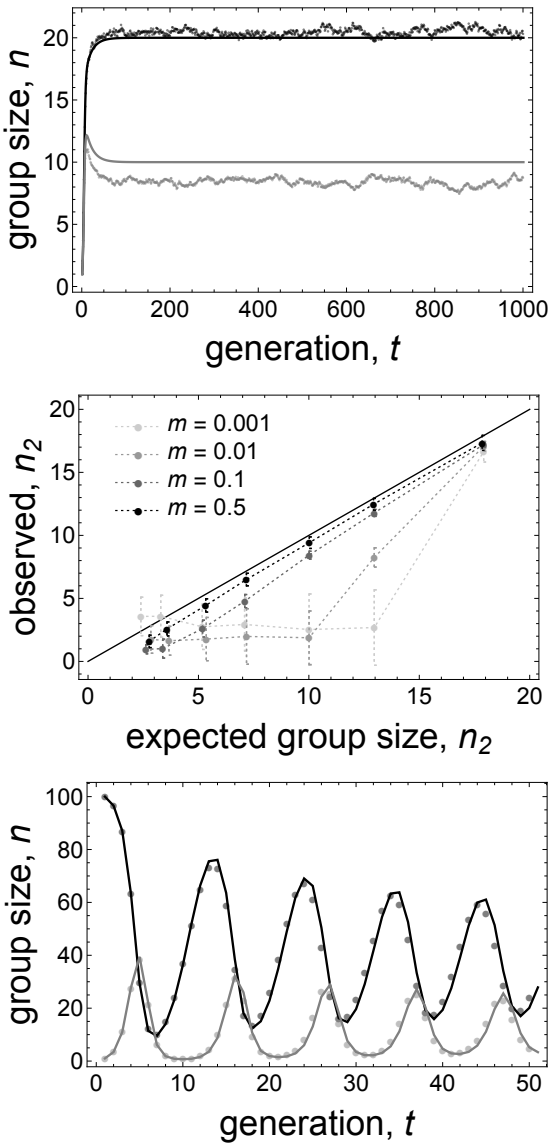

Figure 1: Stochastic dynamics of species local abundance and their deterministic approximation. **Top panel**: deterministic (full line, from eq. 7) and stochastic (dots, with 1000 patches, see Appendix A.1 for details) dynamics for Leslie-Gower model with two species (species 1 in black and 2 in grey, with parameters $f_1 = 2.1$; $f_2 = 2$; $\gamma = 0.04$; $\eta = 0.03$; $m_1 = m_2 = 0.1$, starting with one individual of each species in each patch). **Middle panel**: Comparisons between the deterministic (x-axis, from eq. 7) and stochastic (y-axis with 1000 patches) number individuals of species 2, averaged over 1000 generations starting with one individual of each species in each patch (error bars give the standard deviation of the difference between deterministic and stochastic value, with parameters $f_1 = 2.1$; $f_2 = 2$; $(\gamma, \eta) = (0.05, 0.045)$, $(0.048, 0.043)$, $(0.045, 0.039)$, $(0.043, 0.035)$, $(0.04, 0.03)$, $(0.040, 0.025)$, $(0.05, 0.003)$; $m_1 = m_2 = 0.001, 0.01, 0.1, 0.5$). Departure from the diagonal indicate discrepancies between the exact process and our approximation. Discrepancies occur only for small values of dispersal. **Bottom panel**: deterministic (full line, from eq. A.1) and stochastic (dots, with 1000 patches) dynamics for predator-prey model model (species 1 is the prey and shown in black and species 1 is the predator and shown in grey, with parameters $f_1 = 2$, $f_2 = 1$, $a = 0.035$, $\gamma = 0.01$, $m_1 = m_2 = 0.9$ starting with 100 preys and one predator in each patch, see Appendix A.2 for a description of the predator-prey model).

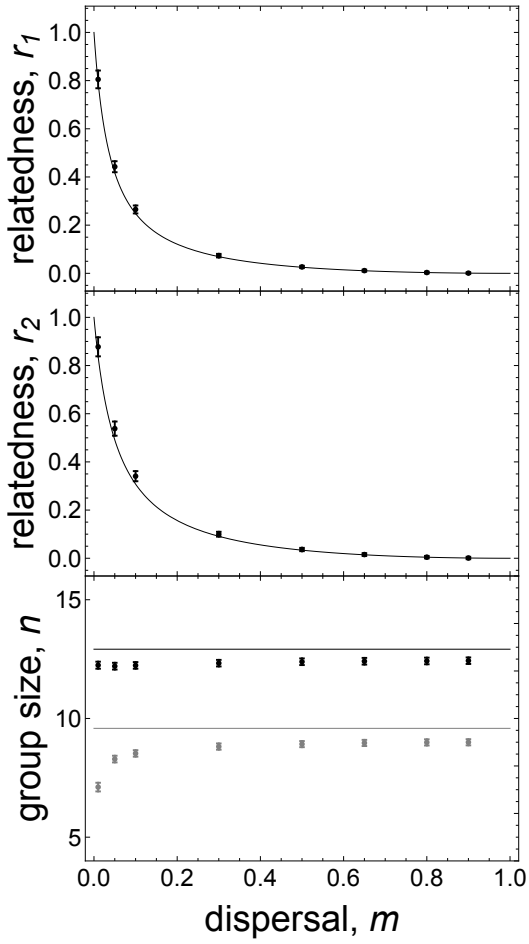

Figure 2: Relatedness and average species local abundance under stochastic individual-based simulations and their deterministic approximation. Full lines give the analytic relatedness in species 1 and 2 (obtained from eq. 10) with Leslie-Gower demographic model (eq. 8, parameter values $f_1 = 2$, $f_2 = 1.8$, $\gamma = 0.07$, and $\eta = 0.01$. Species 1 is shown in black, species 2 in grey. 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.

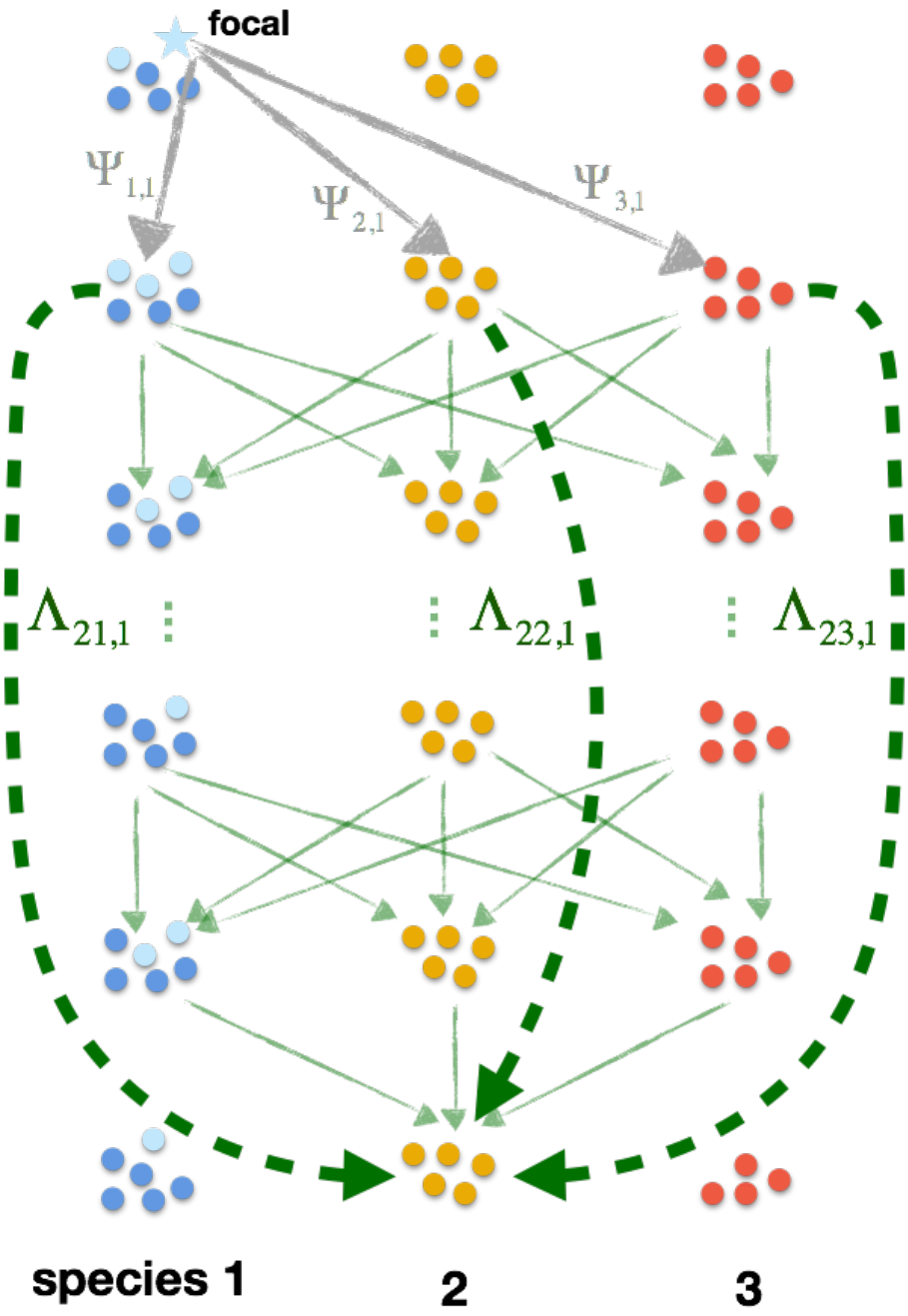

Figure 3: Inter-generational effects of a mutant in one species on the abundance of another. As an example, consider a local community of three species, labelled 1 (blue), 2 (orange) and 3 (red), and let us look at the effect of a mutation in species 1 on the mutant-experienced abundance of species 2, i.e., $\partial\overline{n}_2/\partial\tau_1$. A carrier of a mutation τ_1 in species 1 (denoted by "focal") first influences the population dynamics of species 1, 2 and 3 according to $\Psi_{1,1}(\theta)$, $\Psi_{2,1}(\theta)$ and $\Psi_{3,1}(\theta)$ (grey arrows, eq. 16) in the next generation, and thus the abundance of these species experienced by all its relatives [on average $n_{\mathrm{R},i}(\boldsymbol{\theta})$)] in this generation (carriers of the mutation are denoted by light circle). This change in abundance itself affects the abundance of the other species due to ecological interactions, whose effects ripple downstream to influence other mutation carriers in the focal patch. The change in the abundance of species j experienced by all individuals of species 1 belonging to a lineage of individuals descending from a single ancestor who experienced a change in abundance of species k is given by $\Lambda_{jk,1}(\boldsymbol{\theta})$.

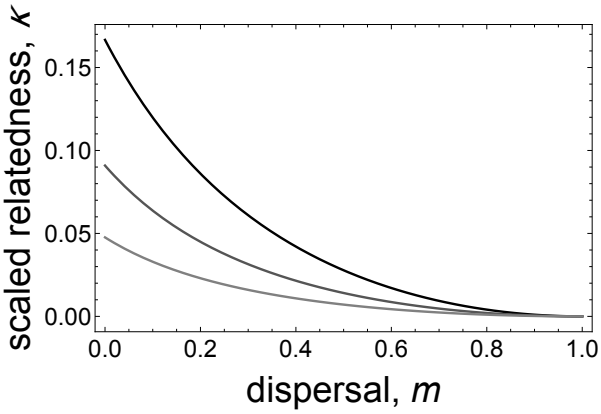

Figure 4: Effective relatedness for the evolution of social interactions within species. Eq. (30) plotted against dispersal with $\gamma = 0.2$ (black), $\gamma = 0.1$ (dark grey) and $\gamma = 0.05$ (light grey). Other parameters: $\theta = 0$ and $f_{\mathrm{b}} = 2$. Effective relatedness therefore decreases with dispersal and patch size (since smaller values of γ lead to larger equilibrium patch size).

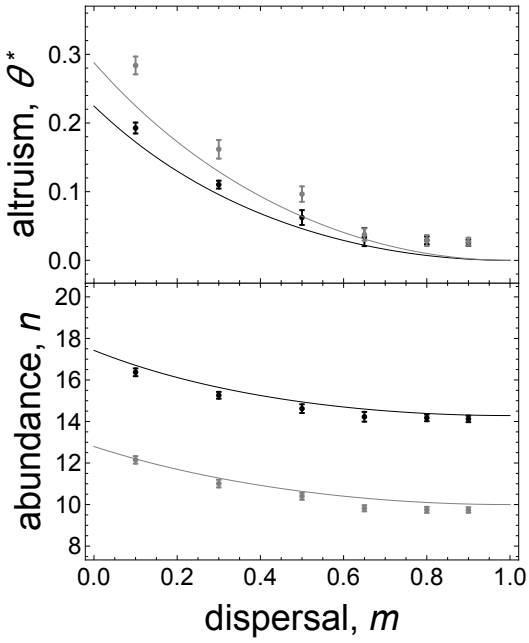

Figure 5: The evolution of altruism within species and concomitant local abundance it generates. Full lines are the singular strategies (top) and concomitant local abundance (bottom) obtained from the selection gradient (obtained by finding the θ^* value solving eq. 29 along with eqs. 24–31) under weak (black, $\gamma = 0.07$) and strong (grey, $\gamma = 0.1$) intra-specific competition (other parameters: $B = 0.05$, $C = 0.5$, $f = 2$). 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.

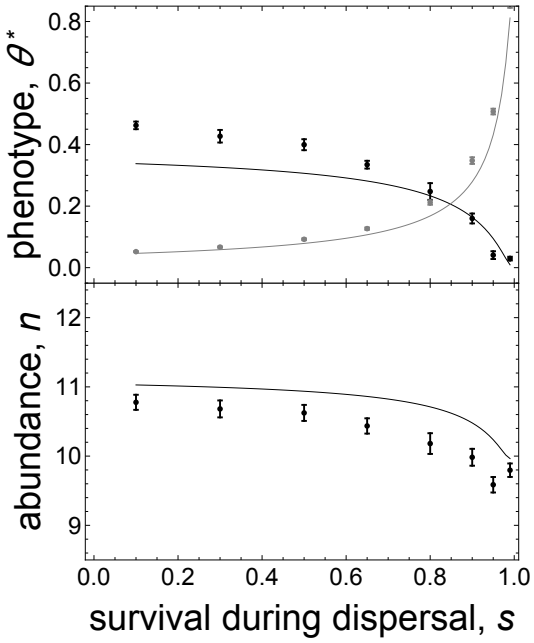

Figure 6: The co-evolution of altruism with dispersal and the concomitant local abundance it generates. Full lines are the singular strategies (top) for altruism (black) and dispersal (grey) and concomitant local abundance (bottom) obtained from the selection gradient (eq. C.11 along with eqs. C.7–C.10, with $B = 0.05$, $C = 0.5$, $f = 2$, $\gamma = 0.05$). 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.

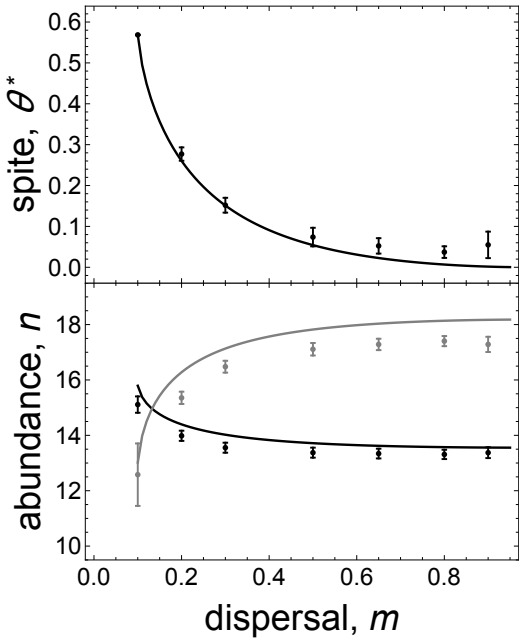

Figure 7: The evolution of spite between species and concomitant local abundance it generates. Full lines are the singular strategies (top) and concomitant local abundance (bottom) in species 1 (black) and 2 (grey) obtained from the selection gradient (eq. C.13 along with eqs. C.14–C.18, with $C_1 = 0.0085$, $f_1 = 2.2$, $f_1 = 2$, $\gamma = 0.055$, $\eta = 0.025$, $\alpha = 0.24$). 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.