Metacommunities, fitness and gradual evolution

Tadeas Priklopil*[1] and Laurent Lehmann[1]

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

Abstract

We analyze the evolution of a multidimensional quantitative trait in a class structured focal species interacting with other species in a wider metacommunity. The evolutionary dynamics in the focal species as well as the ecological dynamics of the whole metacommunity is described as a continuous-time process with birth, physiological development, dispersal, and death given as rates that can depend on the state of the whole metacommunity and can thus accommodate complex local (community) and global (metacommunity) environmental feedbacks (owing to inter- and intra-specific interactions) as well as stochastic fluctuations. For the focal species, we derive a fitness measure for a mutant allele affecting class-specific trait expression. Using classical results from singular perturbation theory, we provide a detailed proof that if the effect of the mutation on phenotypic expression is small ("weak selection"), the large system of dynamical equations needed to describe selection on the mutant allele in the metacommunity can be reduced to a single ordinary differential equation on the arithmetic mean mutant allele frequency that is of constant sign. The mutant will therefore either die out or will out-compete the ancestral (wild)-type. Moreover, the directional selection coefficient driving arithmetic mean allele frequency can be expressed as an inclusive fitness effect calculated from the resident metacommunity alone, and depends, as expected, on individual fitness differentials, relatedness, and reproductive values. This formalizes the Darwinian process of gradual evolution driven by mutation and selection in spatially and class structured metacommunities.

Keywords: Evolutionary biology, invasion implies substitution, fitness, selection gradient, structured populations, relatedness

1 Introduction

Darwinian evolution – the gradual transformation of traits due to random mutation and non-random cumulative natural selection – is the main mechanism of adaptation (Dawkins, 1986, 1997). A proof of principle of Darwinian evolution is given by selective sweeps, in that mutations that increase in frequency also tend to become fixed as a result of natural selection. Repeated fixations under the constant influx of mutations then enables gradual phenotypic evolution. This raises the question about the generality of conditions under which selective sweeps can occur.

In this paper we provide sufficient conditions for selective sweeps in terms of an "invasion implies substitution"-principle when evolution occurs in an asexual focal species that resides in a wider metacommunity where individuals are structured according to a finite number of classes determining the physiological state they can be in (age, size etc.) as well as their local biotic and abiotic environment. The "invasion implies substitution"-principle is a formal characterization of the spread of mutations affecting quantitative traits in the context of density and frequency-dependent selection (Rousset,

*E-mail: tadeas.priklopil@unil.ch

2004; Meszéna et al., 2005; Dercole and Rinaldi, 2008; Lehmann and Rousset, 2014; Dercole, 2016; Cantrell et al., 2017; Ito et al., 2020), and it consists of two propositions (Priklopil and Lehmann, 2020). The first says that if a mutation in the focal species results in a phenotype with a closely similar expression to its ancestral (resident) form, a successful invasion of the mutant allele will generically lead to its fixation. In particular, in the present metacommunity model, the dynamical process for the proportion (frequency) $p(t) \in [0,1]$ of the mutant allele in the focal species should take the form of a scalar-valued ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}p(t) = p(t)(1-p(t))\delta S_{\boldsymbol{\eta}}(\boldsymbol{z}) + O(\delta^2), \tag{1}$$

where $S_{\boldsymbol{\eta}}(\boldsymbol{z})$ is a frequency-independent directional selection coefficient; δ is a small parameter characterizing the distance between the mutant phenotype $\boldsymbol{z} + \delta\boldsymbol{\eta}$ and the resident phenotype $\boldsymbol{z} = (z_{\mathrm{A}})_{\mathrm{A}\in\mathcal{C}}$, with $\mathcal{C}$ being a set of relevant classes, z_{A} the phenotype produced by a resident individual in class $\mathrm{A} \in \mathcal{C}$, and $\boldsymbol{\eta} = (\eta_{\mathrm{A}})_{\mathrm{A}\in\mathcal{C}}$ the mutant phenotypic deviations. Because $S_{\boldsymbol{\eta}}(\boldsymbol{z})$ is a constant in (1), a successful invasion of the mutant allele implies its substitution (fixation in the focal species). This in turn implies that recurrent mutation and invasion events will gradually transform the phenotype $\boldsymbol{z}$ of interest until $S_{\boldsymbol{\eta}}(\boldsymbol{z})$ becomes zero, which allows to identify singular points where the selection on $\boldsymbol{z}$ changes from directional to either stabilizing or disruptive (e.g. Geritz et al., 1998; Rousset, 2004).

The second proposition says that the rate of substitution of the invading mutant phenotype is model independent, and that it can be expressed in terms of an inclusive fitness effect as

$$S_{\boldsymbol{\eta}}(\boldsymbol{z}) = \sum_{\mathrm{A},\mathrm{B}\in\mathcal{C}} v^u_{\mathrm{A}} \left[\frac{\partial h_{\mathrm{AB}}}{\partial z_{\mathrm{B}}}\eta_{\mathrm{B}} + \sum_{\mathrm{C}\in\mathcal{C}} \frac{\partial h_{\mathrm{AB}}}{\partial z_{\mathrm{C}}} r_{\mathrm{BC}}\eta_{\mathrm{C}} \right] u_{\mathrm{B}}, \tag{2}$$

where all quantities are calculated under phenotypic equality with $\delta = 0$. Here, h_{BA} is an element of a resident growth-rate matrix $\mathbf{H}$ giving the rate at which a single individual in class $\mathrm{A} \in \mathcal{C}$ that expresses phenotype z_{A} produces individuals in class $\mathrm{B} \in \mathcal{C}$. The first partial derivative inside the brackets is taken with respect to the phenotype of the focal individual whose growth-rate we are considering and measures the (direct fitness) effect of the mutation on its own growth-rate, where the mutational effect is of magnitude η_{B}. The second partial derivative is taken with respect to the phenotype of an individual of class C inhabiting the same group as the 'focal' individual, and measures the (indirect fitness) effect that mutant individuals in class C have on the growth-rate of the focal individual. This derivative is weighted with the relatedness coefficient r_{BC} giving the probability that an individual in class $\mathrm{C} \in \mathcal{C}$ and the (focal) individual in class $\mathrm{B} \in \mathcal{C}$ have a (recent) common ancestor. The vector $\mathbf{r} = (r_{\mathrm{BC}})_{\mathrm{B},\mathrm{C}\in\mathcal{C}}$ collecting all these relatedness coefficients thus gives the conditional probability of group members having the mutatant allele given the focal individual is a mutant (under phenotypic equality), and captures the effect of limited genetic mixing on the evolutionary dynamics ($\mathbf{r} = \mathbf{0}$ in spatially well-mixed species). Finally, the vectors $\mathbf{u} = (u_{\mathrm{A}})_{\mathrm{A}\in\mathcal{C}}$ and $\mathbf{v}^{\mathbf{u}} = (v^u_{\mathrm{A}})_{\mathrm{A}\in\mathcal{C}}$ are the right and left dominant eigenvectors of $\mathbf{H}$ and are normalized to give the class-specific frequencies (proportion of individuals in each class) and reproductive values (the expected number of offspring of an individual in each class), respectively. The superscript indicates that $\mathbf{v}^{\mathbf{u}}$ is the reproductive value corresponding to $\mathbf{u}$. The directional selection coefficient $S_{\boldsymbol{\eta}}$ as expressed in (2) is thus calculated from a purely resident population state in the focal species and is interpreted as the asymptotic average excess in the rate of producing offspring if a randomly sampled individual from the resident species were to express the mutant instead of the resident phenotype. Selection coefficients taking broadly the form (2) and special cases thereof are routinely applied in evolutionary biology to ascertain singular points (e.g., Frank, 1998; Rousset, 2004; Dercole and Rinaldi, 2008 for textbook treatments).

In this paper, we extend our earlier result on the "invasion implies substitution"-principle (1)-(2) derived for well-mixed ecological communities (Priklopil and Lehmann, 2020) to spatially structured metacommunities where local communities are of finite size and follow an island model of dispersal. For such spatial structure, the "invasion implies substitution"-principle has been previously considered in Roze and Rousset (2003, 2004); Rousset and Ronce (2004); Rousset (2004); Van Cleve et al. (2010, see also closely related results in Rousset, 2006; Cantrell et al., 2017), and our approach and results here complement and extend this earlier work in various aspects. First, we formulate our model in continuous-time and for several interacting and stochastically reproducing species that are physiologically structured (age, stage etc.), and allow simultaneously for local stochastic fluctuations in external factors (e.g. fluctuations in the quality of the local habitat). Second, we provide a detailed and step-by-step proof of (1)-(2) by singularly perturbing the limiting metacommunity model where $\delta = 0$ (Fenichel, 1979; Hek, 2010; Kuehn, 2015; Priklopil and Lehmann, 2020) and pay particular attention to formulate the dynamics of relatedness. We find that the approximation (1)-(2) holds for small but non-zero δ where the perturbation of steady states caused by the invasion is taken into account (a result given without proof previously), and, for all times $t \geq t^*$ where t^* is some finite time that depends on δ. Moreover, and finally, we sharpen earlier representations of (1) for class structured populations given in terms of a reproductive value weighted mutant frequency (Roze and Rousset, 2004; Rousset and Ronce, 2004; Rousset, 2004; Van Cleve et al., 2010; Priklopil and Lehmann, 2020), by allowing to take any average for the mutant frequency with a special case of the arithmetic mean.

The rest of the paper is organized as follows. In Section 2 we introduce the main modeling assumptions and in Section 3 we detail the specifics of the purely resident (ecological) metacommunity model. All the species dynamical variables and growth-rates that are present in the directional selection coefficient (2) are derived in this section. The remainder of the paper is then dedicated in proving the "invasion implies substitution"-principle (1)-(2) for the metacommunity model. To do this, we first present in Section 4 a mutant-resident metacommunity model for arbitrary mutant and resident phenotypes and thus arbitrary strength of selection. Then, in Sections 5-6, we will derive several results for the limiting metacommunity processes where $\delta = 0$, and the formal proof of the "invasion implies substitution"-principle where δ is small but non-zero is given in Section 7. In Section 8 we will conclude the paper by discussing the results and a particular focus is given in relating our results to the theory of adaptation and gradual evolution.

2 Main assumptions

We consider a metacommunity consisting of a large (ideally infinite) number of local communities connected to each other by random uniform dispersal (i.e., an island model of dispersal). Each local community, which we will generally refer to as a local group, consists of a finite number of interacting asexual species that are subject to density-dependent regulations and we assume each local community to be bounded in size. We focus on the ecological and evolutionary dynamics occurring in some focal species of interest in this metacommunity. Each individual in the focal species is assumed to be in one of a finite number $k_{\mathcal{P}}$ of individual-level (i-level) physiological states. We denote the set of all physiological states with $\mathcal{P}$ (which could be a combination of age and stage structure). We assume that individuals are allowed to transition stochastically between physiological states and we call this the physiological transition process (physiological development or maturation). In addition to physiological transitions, individuals give birth, disperse and die stochastically and this holds also for all other species in the metacommunity.

We assume continuous time, and at each moment in time each group is assumed to be in a given g(roup)-state $s \in \mathcal{S}$ (where $\mathcal{S}$ is the state space of groups) that specifies the number of individuals

in each physiological state of every species inhabiting the group, including any external factor (e.g. quality of the habitat), all of which can affect the survival, reproduction, and dispersal of community members. Because each element $s \in \mathcal{S}$ determines exactly the size distributions in the focal species (and the other species) in a group, we write $\{N_{as}\}_{a\in\mathcal{P}}$ to denote the size distribution of the focal species for a group that is in state $s \in \mathcal{S}$, where N_{as} is the number of individuals that are in physiological state $a \in \mathcal{P}$ and that inhabit a group in state $s \in \mathcal{S}$ (and groups can transition between different states, see below). We assume that the size (number of elements) of the space $\mathcal{S}$ is finite with $k_{\mathcal{S}}$ elements.

Individuals in the focal species express a genetically evolving phenotype that may affect the survival, reproduction, and dispersal of any individual in the metacommunity (including self). To specify how the phenotype is expressed we use the notion of the *class* of an individual in the focal species. This is defined such that whenever any two individuals that express the same phenotype are identical in terms of their survival, reproduction and dispersal, they belong to the same class (Taylor, 1990; Rousset, 2004; Grafen, 2015). In the present model, individuals are in the same class whenever they are in the same physiological state *and* inhabit a group that is in the same g-state. The class space is denoted with $\mathcal{C} := \mathcal{P} \times \mathcal{S}$ and it is of size $k_{\mathcal{C}} = k_{\mathcal{P}} \times k_{\mathcal{S}}$. We assume that phenotypic expression is class-specific. For example, the phenotype may depend on the physiological state of the individual (e.g. age or size) as well as the external factors of the group that it inhabits (e.g. quality of the habitat). An individual in class $as \in \mathcal{C}$ expresses phenotype $z_{as} \in \mathbb{R}$ such that $\boldsymbol{z} = (z_{as})_{as\in\mathcal{C}} \in \mathcal{Z}$ denotes the vector-valued phenotypic profile across classes and where $\mathcal{Z} \subset \mathbb{R}^{k_{\mathcal{C}}}$ is the phenotype space.

We assume that offspring inherit faithfully the phenotype of their parent and that two alleles resulting in different phenotypic expression segregate in the focal species, a mutant allele M and a resident allele R. Characterizing the spread of the mutant allele is the focus of our analysis.

We note that many closely connected ecological or evolutionary models allowing for various heterogeneities within and/or between groups have been considered before for both continuous or discrete time processes e.g. by Chesson (1981, 1985); Grey et al. (1995); Frank (1998); Gandon and Michalakis (1999); Ronce et al. (2000); Metz and Gyllenberg (2001); Arrigoni (2003); Cadet et al. (2003); Barbour and Pugliese (2004); Rousset (2004); Rousset and Ronce (2004); Martcheva and Thieme (2005); Lehmann et al. (2006); Martcheva and Thieme (2006); Martcheva et al. (2006); Alizon and Taylor (2008); Wild et al. (2009); Ronce and Promislow (2010); Wild (2011); Peña (2011); Rodrigues and Gardner (2012); Parvinen (2013); Massol and Débarre (2015); Lehmann et al. (2016); Rodrigues (2018); Kuijper and Johnstone (2019). However, no prior model has considered metacommunities with multiple interacting species where individuals and groups are characterized by arbitrary states and as such all these previous models can in principle be conceptually subsumed to the present analysis (as well as are models with finite class-structure in panmictic populations).

3 Resident dynamics in the metacommunity

We start by describing the population dynamics (ecology) of the focal species in the metacommunity assuming that the focal species is monomorphic for the resident allele. For this, we need to characterize the full metacommunity dynamics. While this characterization may at first appear lengthy, we fully work it out as (i) it plays a central role in the analysis of selection and (ii) it is of interest in the formulation of purely metacommunity ecological models. The g-state of a resident monomorphic metacommunity is described by the discrete probability (frequency) distribution over the state space of groups $\mathcal{S}$. The metacommunity state space is therefore the space of frequency distributions $\Delta(\mathcal{S})$ on $\mathcal{S}$ (i.e. a simplex in $\mathbb{R}^{k_{\mathcal{S}}}$). We will represent the frequency distribution of the metacommunity at

time $t \in \mathbb{R}$ with a vector $\mathbf{x} := \mathbf{x}(t) \in \mathbb{R}_+^{k_\mathcal{S}}$, where $\mathbf{x} = (x_s)_{s\in\mathcal{S}} \in \Delta(\mathcal{S})$ and where x_s is the frequency of groups in state $s \in \mathcal{S}$ (hence $\sum_{s\in\mathcal{S}} x_s = 1$), and next we track the dynamics of this vector.

3.1 Individual-level processes

To construct the dynamics of the frequency distribution of the metacommunity, we start from individual processes of the focal species and introduce the function

$$\lambda_{au\leftarrow bs} : \mathcal{Z}_\mathcal{C} \times \mathcal{Z} \times \mathcal{S} \times \mathcal{Z} \times \Delta(\mathcal{S}) \to \mathbb{R}_+ \qquad \text{for all } au, bs \in \mathcal{C}, \tag{3}$$

which gives the rate at which a focal individual in class $bs \in \mathcal{C}$ (i.e., an individual in physiological state $b \in \mathcal{P}$ inhabits a group in g-state $s \in \mathcal{S}$) either *produces* by reproduction or physiological transition, or *removes* by death an 'offspring' individual in class $au \in \mathcal{C}$. The notation $au \leftarrow bs$ indicates who produced or removed whom, and in the case of physiological transitions and death the 'parent' and 'offspring' individuals are interpreted as one and the same individual. The rate function (3) is defined such that the first argument in $\lambda_{au\leftarrow bs}(z_{bs}, \boldsymbol{z}, s, \boldsymbol{z}, \mathbf{x})$ indicates the phenotype $z_{bs} \in \mathcal{Z}_\mathcal{C}$ expressed by the focal individual who undergoes the i-process, and the remaining arguments describe the rest of the metacommunity that affect the i-processes of the focal individual: the second argument $\boldsymbol{z} \in \mathcal{Z}$ indicates the phenotypic profile of resident individuals that inhabit the group of the focal individual and the third argument $s \in \mathcal{S}$ gives the state of the entire focal group (the distribution of all species as well as 'external factors'). Likewise, the fourth argument $\boldsymbol{z} \in \mathcal{Z}$ indicates the phenotypic profile of resident individuals in the focal species outside of the group of the focal individual and the final argument gives the g-state distribution of the entire metacommunity $\mathbf{x} \in \Delta(\mathcal{S})$.

The rate function (3) can take into account arbitrary (group and metacommunity) environmental feedbacks and for ease of presentation we use the short-hand notation

$$\lambda_{au,bs}(\mathbf{x}) = \lambda_{au\leftarrow bs}(z_{bs}, \boldsymbol{z}, s, \boldsymbol{z}, \mathbf{x}). \tag{4}$$

This can be partitioned into philopatric and dispersal components as

$$\lambda_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) \tag{5a}$$

(hence the superscripts 'p' and 'd'). Here,

$$\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{p,B}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{D}}_{au,bs}(\mathbf{x}), \tag{5b}$$

is the rate at which the focal individual in class $bs \in \mathcal{C}$ undergoes an i-process producing or removing an individual in class $au \in \mathcal{C}$ in the focal group (philopatric events), which is further decomposed into philopatric birth (superscripts 'p' and 'B'), physiological transition (superscript 'T') and death (superscript 'D'). In turn,

$$\lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) = \sum_{w\in\mathcal{S}} x_w \lambda^{\mathrm{d,B}}_{au,w,bs}(\mathbf{x}), \tag{5c}$$

where $\lambda^{\mathrm{d,B}}_{au,w,bs}(\mathbf{x})$ is the rate at which an individual in class $bs \in \mathcal{C}$ gives birth and sends offspring in physiological state $a \in \mathcal{P}$ away (hence superscript for dispersal 'd' and birth 'B'), and the offspring successfully establish themselves in a group that has transitioned due to this event to a state $u \in \mathcal{S}$, given the offspring landed in group $w \in \mathcal{S}$ with probability x_w (Figure 1, see further details in Appendix A, Section A.1).

3.2 Group state dynamics

The dynamics of the frequency of group states in the metacommunity can now be expressed as a system of ODE's

$$\dot{\mathbf{x}} = \mathbf{Q}(\mathbf{x})\mathbf{x}, \tag{6}$$

where $\mathbf{Q}(\mathbf{x}) = (q_{u,s}(\mathbf{x}))_{u,s\in\mathcal{S}} \in \mathbb{R}^{k_\mathcal{S}\times k_\mathcal{S}}$ is a group transition-rate matrix where $q_{u,s}(\mathbf{x})$ is the rate at which a group in state $s \in \mathcal{S}$ transitions to state $u \in \mathcal{S}$ (and hence $\sum_{u\in\mathcal{S}} q_{u,s}(\mathbf{x}) = 0$ for all $s \in \mathcal{S}$, Figure 1). We partition this rate as

$$q_{u,s}(\mathbf{x}) = q^{\mathrm{p}}_{u,s}(\mathbf{x}) + q^{\mathrm{d}}_{u,s}(\mathbf{x}) + q^{\mathrm{mc}}_{u,s}(\mathbf{x}) \tag{7}$$

into group transitions due to philopatric transitions

$$q^{\mathrm{p}}_{u,s}(\mathbf{x}) = \sum_{a,b\in\mathcal{P}} \lambda^{\mathrm{p}}_{au,bs}(\mathbf{x})N_{bs}, \tag{8}$$

and dispersal transitions

$$q^{\mathrm{d}}_{u,s}(\mathbf{x}) = \sum_{a,c\in\mathcal{P}} \sum_{w\in\mathcal{S}} \lambda^{\mathrm{d,B}}_{au,s,cw}(\mathbf{x})N_{cw}x_w, \tag{9}$$

which both sums over all physiological states of the parents and their possible progeny as well. Finally, all other non-focal species processes affecting the metacommunity state dynamics are captured by the rate $q^{\mathrm{mc}}_{u,s}(\mathbf{x})$.

We note that there are two biologically relevant perspectives on the process generated by (6). First, one can interpret $\mathbf{x}$ as the frequency distribution of an *infinite collection of groups* and which is taken as the state of the metacommunity (as discussed above). In this view the ODE in (6) describes a deterministic processes on how the state of an infinite collection of groups changes in time. We will call it the deterministic perspective. Second, one can look at a *single group* and interprets the vector $\mathbf{x} = \mathbf{x}(t)$ as a probability distribution over the state space $\mathcal{S}$, each element giving the probability of the state the group is in at time t. The rates (7) are then interpreted as the rates at which the group transitions from one state in $\mathcal{S}$ to another state in $\mathcal{S}$. Taking this perspective, all groups in the metacommunity are assumed to undergo a non-homogeneous continuous-time Markov chain (Iosifescu, 2007, chap. 7) with transition matrix $\mathbf{Q}(\mathbf{x})$ that is a function of the state $\mathbf{x}$ of the metacommunity (hence the non-homogeneity). We will call it a probabilistic perspective. In the main text, we follow for ease of presentation the deterministic perspective and for the final results in Section 7 we discuss the probabilistic perspective.

3.3 Resident population and reproductive value dynamics

We now characterize the dynamics of the number of individuals in each class of the focal species, which needs to be done in the backdrop of the full metacommunity state dynamics (6). Let $\mathbf{n} = (n_{as})_{as\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}}_+$ denote the density vector of individuals of the focal species (monomorphic population), where $n_{as} = N_{as}x_s$ denotes the density of individuals (number of individuals per unit space) in class $as \in \mathcal{C}$ (and recall that N_{as} is a constant, Section 2). This vector satisfies the system of ODEs

$$\dot{\mathbf{n}} = \mathbf{H}(\mathbf{x})\mathbf{n}, \tag{10}$$

where $\mathbf{H}(\mathbf{x}) = (h_{au,bs}(\mathbf{x}))_{au,bs\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}\times k_\mathcal{C}}$ is the resident individual growth-rate matrix. The element $h_{au,bs}(\mathbf{x})$ of this matrix (see for a derivation Appendix A, Section A.2) is

$$h_{au\leftarrow bs}(z_{bs}, \boldsymbol{z}, s, \boldsymbol{z}, \mathbf{x}) = \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) + (N_{au} - N_{as})\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) + \delta_{a,b}q_{u,s}(\mathbf{x}), \tag{11}$$

which is the rate at which a focal individual in class $bs \in \mathcal{C}$ with phenotype $\boldsymbol{z} \in \mathcal{Z}$ produces an 'offspring' individual of class $au \in \mathcal{C}$ (where the 'parent' and 'offspring' may be one and the same individual e.g. due to physiological transition and survival, Section 3.1) and thus gives a representation of the class-specific individual growth-rate [or individual fitness where $(N_{au} - N_{bs})$ can take values ± 1 or 0 and whose arguments are identically interpreted to those of $\lambda_{au\leftarrow bs}$ (3)]. The (individual) fitness (11) is also partitioned into the i-processes of the focal individual (the first two terms on the right hand side of (11)) and the rates at which the group of the focal individual changes its state (the final term on the right hand side of (11)), and which can be interpreted as the 'survival' of the focal individual in a changing local environment, where the kronecker delta $\delta_{a,b}$ takes value 1 when $a = b \in \mathcal{P}$, and otherwise 0. More precisely, the first term on the right hand side of (11) gives the rate of producing offspring via dispersal, and the second term the rate at which a parent individual $bs \in \mathcal{C}$ either produces a new individual $a \in \mathcal{P}$ into $u \in \mathcal{S}$ by philopatric *reproduction* or *physiological transition* (i.e. $N_{au} - N_{as} = 1$ whenever a parent $bs \in \mathcal{C}$ transitions or gives birth to $au \in \mathcal{C}$), or removes an individual $a \in \mathcal{P}$ from $s \in \mathcal{S}$ by *death* or *physiological transition* (this term is non-zero only for $a = b \in \mathcal{P}$ in which case $N_{au} - N_{as} = -1$), while turning the group $s \in \mathcal{S}$ into $u \in \mathcal{S}$. The final term in (11) is the survival component of the growth-rate giving the rate at which the parent individual $b \in \mathcal{P}$ transitions from inhabiting a group in state $s \in \mathcal{S}$ into $u \in \mathcal{S}$. This rate is given in terms of all the processes in the entire metacommunity, including the i-processes of the parent $bs \in \mathcal{C}$ itself (and then e.g. upon its own death this term is subtracted by the first term in (11)). The second and third term thus contribute to philopatric fitness, while the first term in (11) gives the rate of offspring production through dispersal. We note that in order to study the dynamics of $\mathbf{n}$, the ODE (10) needs to be coupled with the dynamics for the metacommunity state $\mathbf{x}$ in (6).

The dispersal and net philopatric fitness contributions of individual fitness (11) corresponds to those of previous discrete time models, which are not expressed in terms of individual and group transition rates but directly in terms of the two net fitness contributions (Lehmann et al., 2016, eq. 6, Ohtsuki et al., 2020, eq. 38, and see Appendix A, Section A.2.2 for a further decomposition of the individual fitness). In the absence of limited dispersal and group structure the first and third terms in (11) vanish, and the population demographic state can be characterized in terms of vector $\mathbf{n}$ only so that $\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{p}}_{au,bs}(\mathbf{n})$, in which case fitness reduces to that of the spatially well-mixed model in (Priklopil and Lehmann, 2020, eq. 3).

Because we are eventually interested in the asymptotic resident species (population) dynamics, we will utilize the concept of a reproductive value (Fisher, 1930; Taylor, 1990; Rousset, 2004; Grafen, 2006) defined here as being the the adjoint variable corresponding to $\mathbf{n}$ (we note that this departs slightly from the standard definition of reproductive values Taylor, 1990, see further discussion in Section 7.1). Adjoint variables are useful in the analysis of dynamical systems because, by definition, the product of a variable with its adjoint is constant in time (we see its practical use in later sections, and see its previous use e.g. in the context of age-structured populations in Greiner et al., 1994; Inaba, 2017). The population dynamical (non-asymptotic) definition of reproductive values $\mathbf{v}^{\mathbf{n}} = (v^n_{as})_{as\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}}_+$ is obtained from

$$-\dot{\mathbf{v}}^{\mathbf{n}} = \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}), \tag{12}$$

and normalized such that $\mathbf{v}^{\mathbf{n}} \cdot \mathbf{n} = 1$ for all t (where the superscript in $\mathbf{v}^{\mathbf{n}}$ indicates that it is an adjoint to $\mathbf{n}$, and one can readily check by applying (10) that $\mathbf{v}^{\mathbf{n}} \cdot \mathbf{n}$ indeed stays constant in time).

The dynamical interpretation of reproductive values is obtained from (12) and is tied to a final value problem where (12) is coupled with a final condition $\mathbf{v^n}(t_\mathrm{f})$ where $t \leq t_\mathrm{f} \leq \infty$ is some final time: the reproductive value $v^n_{as}(t)$ is the probability that a randomly sampled individual at t_f descends from a single individual of class $as \in \mathcal{C}$ at time t (or alternatively, considering an individual of class $as \in \mathcal{C}$ at time t, $v^n_{as}(t)$ is the *frequency* of individuals at the final time t_f who are its descendants). In most previous use of reproductive value (as well as in the present paper later on), the final time is usually the distant future ($t_\mathrm{f} \to \infty$) so that $v^n_{as}(t)$ is interpreted as an asymptotic contribution to fitness as $t \to \infty$. Similarly to (10), in order to study the dynamics of $\mathbf{v^n}$, the ODE (12) needs to be coupled with the dynamics for the metacommunity state $\mathbf{x}$ in (6).

3.4 Relatedness dynamics

The remaining variable that appears in the directional selection coefficient (2) and whose dynamics we aim to characterize in the resident population is relatedness. We *define* relatedness between two randomly sampled individuals $a, b \in \mathcal{P}$ without replacement and inhabiting the same group $s \in \mathcal{S}$ at time t, as the probability $r_{abs} = r_{abs}(t)$ that the two individuals have a common ancestor; namely the probability that the ancestral lineages of the two individuals coalesce in a common ancestor (and this could be calculated by studying the purely resident i-processes backwards in time, see Michod and Hamilton, 1980 for the original definition and Rousset, 2004 for an analysis of relatedness under a large class of different discrete time scenarios in the island model).

Two individuals that inhabit the same group are related if they either share a parent, who is their common ancestor, or if any of their ancestors shared a parent (and note that if two individuals share a parent in a continuous-time model with one birth per unit of time, one of the two individuals must be the parent through survival). Because there are infinite number of groups in the metacommunity, the probability that two individuals in two different groups share a parent is 0. To find a dynamical equation for relatedness we thus only need to keep track of the ancestors of local parents and to that end we focus on densities of local pairs. We denote a pair of individuals in physiological states $a, b \in \mathcal{P}$ that inhabit the same group that is in state $s \in \mathcal{S}$ with $(as, bs) := (abs) \in \mathcal{B}$, where $\mathcal{B} := \mathcal{P}^2 \times \mathcal{S}$ is the space of local pairs (whose size is denoted $k_\mathcal{B}$). Then,

$$n^r_{abs} = r_{abs} n_{abs} \tag{13}$$

is the density of pairs of individuals who are related with

$$n_{abs} = N_{abs} x_s, \tag{14}$$

being the density of pairs $abs \in \mathcal{B}$ in that group, and where

$$N_{abs} = \begin{cases} N_{as} N_{bs} & \text{if} \quad b \neq a \\ \frac{N_{as}(N_{as}-1)}{2} & \text{if} \quad b = a, \end{cases} \tag{15}$$

is the the number of pairs $abs \in \mathcal{B}$.

We start by deriving ODEs' for n_{abs} and n^r_{abs}, which in turn will allow us to determine r_{abs}. Because we are interested in the ancestry of each pair we will pay special attention to the state and location of 'parent' individuals associated to each 'offspring' (this is in contrast to deriving individual growth-rates in Section 3.3 where we focused on the state and location of offspring associated to each parent).

By differentiating (14) we obtain an ODE for the pair dynamics

$$\begin{aligned}\dot{n}_{abu} &= N_{abu}\dot{x}_u = N_{abu}\sum_{s\in\mathcal{S}} q_{u,s}(\mathbf{x})x_s \\ &= N_{abu}\sum_{s\neq u\in\mathcal{S}} q_{u,s}(\mathbf{x})x_s - n_{abu}\sum_{w\neq u} q_{w,u}(\mathbf{x}), \quad \text{for all } abu\in\mathcal{B},\end{aligned} \tag{16}$$

where we used (6), (14) and the fact that $\sum_{u\in\mathcal{S}} q_{u,s}(\mathbf{x}) = 0$ (Section 3.2), and by substituting (7)-(9) into (16) we get

$$\begin{aligned}\dot{n}_{abu} = &\sum_{s\neq u\in\mathcal{S}}\Big[\sum_{d\in\mathcal{P}}\beta^{\mathrm{CA}}_{abu,ds}(\mathbf{x})n_{ds} + \sum_{d\in\mathcal{P}}\beta^{\mathrm{IM}}_{abu,ds}(\mathbf{x})n_{ds} + \sum_{c,d\in\mathcal{P}}\beta^{\mathrm{DP}}_{abu,cds}(\mathbf{x})n_{cds}\Big] - \\ &- n_{abu}\sum_{w\neq u} q_{w,u}(\mathbf{x}), \quad \text{for all } abu\in\mathcal{B},\end{aligned} \tag{17}$$

which depends on three rates. First,

$$\beta^{\mathrm{CA}}_{abu,ds}(\mathbf{x}) = \begin{cases}\lambda^{\mathrm{p,B}}_{au,bs}(\mathbf{x}) & \text{if} \quad d=b,\\ \lambda^{\mathrm{p,B}}_{bu,as}(\mathbf{x}) & \text{if} \quad d=a,\\ 0 & \text{otherwise,}\end{cases} \tag{18}$$

which is the rate at which a single parent individual $ds\in\mathcal{C}$ produces offspring pairs $abu\in\mathcal{B}$ by philopatric birth and its own survival and is thus said to be the common ancestor to these pairs (hence the superscript). Second,

$$\beta^{\mathrm{IM}}_{abu,ds}(\mathbf{x}) = \begin{cases}\sum_{w\in\mathcal{S}}\sum_{e\in\mathcal{P}}\lambda^{\mathrm{d}}_{au,s,ew}(\mathbf{x})N_{ew}x_w & \text{if} \quad d=b,\\ \sum_{w\in\mathcal{S}}\sum_{e\in\mathcal{P}}\lambda^{\mathrm{d}}_{bu,s,ew}(\mathbf{x})N_{ew}x_w & \text{if} \quad d=a,\\ 0 & \text{otherwise,}\end{cases} \tag{19}$$

which is the rate at which a single local parent individual $ds\in\mathcal{C}$ is being paired up with an immigrant offspring (hence the superscript) to produce an offspring pair $abu\in\mathcal{B}$ (and recall that the probability for an immigrant offspring to land in the group of the individual $ds\in\mathcal{C}$ depends on the state of the entire group $s\in\mathcal{S}$). Finally,

$$\beta^{\mathrm{DP}}_{abu,cds}(\mathbf{x}) = \beta^{\mathrm{BT}}_{abu,cds}(\mathbf{x}) + \delta_{ab,cd}\beta^{\mathrm{S}}_{abu,abs}(\mathbf{x}), \tag{20a}$$

which is the rate at which a local parent pair $cds\in\mathcal{B}$ (two different parents, hence the superscript) produces an offspring pair $abu\in\mathcal{B}$ and this depends on two processes. First,

$$\beta^{\mathrm{BT}}_{abu,cds}(\mathbf{x}) = \begin{cases}\lambda^{\mathrm{p,B}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) & \text{if} \quad d=b \quad \text{and} \quad c\neq b,\\ 2(\lambda^{\mathrm{p,B}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x})) & \text{if} \quad d=b \quad \text{and} \quad c=b,\\ \lambda^{\mathrm{p,B}}_{bu,as}(\mathbf{x}) + \lambda^{\mathrm{T}}_{bu,as}(\mathbf{x}) & \text{if} \quad c=a \quad \text{and} \quad d\neq a,\\ 2(\lambda^{\mathrm{p,B}}_{bu,as}(\mathbf{x}) + \lambda^{\mathrm{T}}_{bu,as}(\mathbf{x})) & \text{if} \quad c=a \quad \text{and} \quad d=a,\\ 0 & \text{otherwise,}\end{cases} \tag{20b}$$

which is the rate at which the local parent pair $cds\in\mathcal{B}$ produces an offspring pair $abu\in\mathcal{B}$ by birth and physiological transition, and because these rates are given for parent pairs, a factor 2 appears whenever both parents can undergo the i-processes that contribute to the birth of the considered offspring pair. Second,

$$\beta^{\mathrm{S}}_{abu,abs}(\mathbf{x}) = q_{u,s}(\mathbf{x}) - [\lambda^{\mathrm{p,D}}_{au,as}(\mathbf{x}) + \lambda^{\mathrm{p,D}}_{bu,bs}(\mathbf{x})] - [\lambda^{\mathrm{p,T}}_{au,as}(\mathbf{x}) + \lambda^{\mathrm{p,T}}_{bu,bs}(\mathbf{x})], \tag{20c}$$

which is the rate at which each existing pair $abs \in \mathcal{B}$ survives upon any i-process occurring in the group. The minus term appears here because parents $a, b \in \mathcal{P}$ that die or transition contribute negatively to the formation of offspring pairs $abu \in \mathcal{B}$. Note that all these rates for production of offspring pairs depend on the original i-rates (5b)-(5c) of offspring production by single individuals.

The dynamics of n^r_{abs} is obtained from (17) by subtracting the immigration term and by only considering the fraction of pairs that are related and we get

$$\dot{n}^{\mathrm{r}}_{abu} = \sum_{s \neq u \in \mathcal{S}} \sum_{d \in \mathcal{P}} \beta^{\mathrm{CA}}_{abu,ds}(\mathbf{x}) n_{ds} + \sum_{s \neq u \in \mathcal{S}} \sum_{c,d \in \mathcal{P}} \beta^{\mathrm{DP}}_{abu,cds}(\mathbf{x}) n^r_{cds} - n^r_{abu} \sum_{w \neq u} q_{w,u}(\mathbf{x}), \tag{21}$$

for all pairs $abu \in \mathcal{B}$. The density of pairs $abu \in \mathcal{B}$ that are related increases due to a single local parent producing pairs anew (the first term in (21)) or due to two related local parents producing pairs (the second term in (21)), and it decreases due to any event that changes the state of the group (the last term in (21)). The ODE for the vector of relatedness $\mathbf{r} = (r_{abs})_{abs \in \mathcal{B}} \in [0,1]^{k_{\mathcal{B}}}$ can be given in terms of its elements $r_{abs} = n^r_{abs}/n_{abs}$ and is obtained by differentiation

$$\dot{r}_{abu} = \frac{1}{n_{abu}} \left[\sum_{s \neq u \in \mathcal{S}} \sum_{d \in \mathcal{P}} \beta^{\mathrm{CA}}_{abu,ds}(\mathbf{x}) n_{ds} + \sum_{s \neq u \in \mathcal{S}} \sum_{c,d \in \mathcal{P}} \beta^{\mathrm{DP}}_{abu,cds}(\mathbf{x}) n_{cds} r_{cds} - r_{abu} N_{abu} \sum_{s \neq u \in \mathcal{S}} q_{u,s}(\mathbf{x}) x_s \right] \tag{22}$$

for all $abu \in \mathcal{B}$, and where we have used (16), (21) and $n^r_{cds} = r_{cds} n_{cds}$. Notice that the rates that appear in (16) and (21) at which pairs $abu \in \mathcal{B}$ are removed due to the group $u \in \mathcal{S}$ changing its state, cancel in (22). Hence, the rate at which relatedness changes is the relative rate of producing pairs $abu \in \mathcal{B}$ by a common parent or by parents that share an ancestor, relative to an average rate of producing pairs $abu \in \mathcal{B}$. We note that similarly to (10) and (12), in order to study the dynamics in (22) it needs to be coupled with the dynamics for the metacommunity state $\mathbf{x}$ in (6).

By ignoring physiological structure in the focal species, the dynamics in (22) reduces to the expression given in Wild et al. (2009, Supplementary Information, equation S4 and Table S1). Moreover, assuming no class-structure so that all groups consists only of the focal species of constant size N (the vector $\mathbf{x}$ plays no role), the continuous time dynamical equation for relatedness (22) is simplified further to

$$\dot{r} = \frac{2}{N(N-1)} \left[\beta^{\mathrm{CA}} N + \beta^{\mathrm{DP}} \frac{N(N-1)}{2} r - rq \frac{N(N-1)}{2} \right], \tag{23}$$

where q is the total rate at which events happen in a group. Note that each term is a rate of production of offspring pairs to the total number of offspring pairs in a group. When each birth event is coupled with a death event (i.e., the Moran process, Ewens, 2012), the rates in (23) can be written as

$$\begin{aligned} \beta^{\mathrm{CA}} &= (1-m) \frac{1}{N} \left(\frac{N-1}{N} \right) q \\ \beta^{\mathrm{DP}} &= \left[2(1-m) \frac{1}{N} \left(\frac{N-1}{N} \right) + \left(1 - \frac{2}{N} \right) \right] q, \end{aligned} \tag{24}$$

where m is the probability that upon a birth event the offspring does not disperse. Here, β^{CA} is interpreted as the rate at which events happen in the group q times the probability that the parent individual reproduces locally $(1-m)\frac{1}{N}$ and does not die $\frac{N-1}{N}$. The rate β^{DP} is interpreted as the rate at which events happen in the group q times the probability that either individual in the pair reproduces locally and does not die $2(1-m)\frac{1}{N}\frac{N-1}{N}$, or, the probability that neither individual in the pair dies $1 - \frac{2}{N}$.

3.4.1 The jump process for relatedness

Next we represent the continuous-time dynamics of relatedness (22) as a discrete-time jump-process, with jumps occurring whenever an (i-level) event changes the state of the group. This representation is useful when only the steady state is considered and also in connecting our result to previous models that have mostly been formulated in discrete-time. We thus consider a jump chain (or embedded Markov chain, e.g., Iosifescu, 2007, chapter 8.3) associated to the continuous-time Markov chain given in (6) (see further details in Appendix B, Section B.1). An embedded Markov chain is a discrete-time Markov chain that records state transitions upon jump times $T_1, T_2, \dots,$ where each jump time T_k is a random variable at which the focal group undergoes an event either due to local (philopatric) or non-local (via immigration) i-processes. The jump process for relatedness is given in a vector form as

$$\mathbf{r}' = \mathbf{c}(\mathbf{x}) + \mathbf{D}(\mathbf{x})\mathbf{r}, \tag{25}$$

where $\mathbf{r} = (r_{abs})_{abs\in\mathcal{B}} \in [0,1]^{k_{\mathcal{B}}}$ and where $\mathbf{r}' = \mathbf{r}(T_{k+1})$ and $\mathbf{r} = \mathbf{r}(T_k)$. Here, $\mathbf{c}(\mathbf{x}) = (c_{abs}(\mathbf{x}))_{abs\in\mathcal{B}} \in [0,1]^{k_{\mathcal{B}}}$ is a vector where entry $c_{abs}(\mathbf{x})$ is the probability that a randomly sampled offspring-pair $abs \in \mathcal{B}$ has a common parent (ancestor) in the same group. This is the coalescence probability (backward in time) and is computed as $c_{abu}(\mathbf{x}) = \sum_{ds\in\mathcal{C}} C_{abu,ds}(\mathbf{x})$ with

$$C_{abu,ds}(\mathbf{x}) = \frac{\beta^{\mathrm{CA}}_{abu,ds}(\mathbf{x})N_{ds}}{N_{abu}}\frac{x_s}{q_u}, \tag{26}$$

for $s \neq u \in \mathcal{C}$ and 0 otherwise, and where $q_u = \sum_{w\neq u\in\mathcal{S}} q_{u,w}(\mathbf{x})x_w$ is the rate at which group transitions to state u occur. Equation (26) gives the probability that, conditional on the occurrence of a group transition event to $u \in \mathcal{S}$, that group was in state $s \in \mathcal{S}$ before the transition and a randomly sampled offspring-pair $abu \in \mathcal{B}$ was produced anew by a local parent individual in class $ds \in \mathcal{C}$ ((26) was obtained using the rates (18), and q_u is obtained from the first expression on the right hand side of (16)). The matrix $\mathbf{D}(\mathbf{x}) = (D_{abu,cds}(\mathbf{x}))_{abu,cds\in\mathcal{B}} \in \mathbb{R}^{k_{\mathcal{B}}\times k_{\mathcal{B}}}$ has entries

$$D_{abu,cds}(\mathbf{x}) = \frac{\beta^{\mathrm{DP}}_{abu,cds}(\mathbf{x})N_{cds}}{N_{abu}}\frac{x_s}{q_u}, \tag{27}$$

for all $s \neq u \in \mathcal{C}$ and 0 otherwise, and gives the probability that, conditional on the occurrence of a group transition event to $u \in \mathcal{S}$, that group was in state $s \in \mathcal{S}$ before the transition and a randomly sampled offspring-pair $abu \in \mathcal{B}$ was produced (by philopatric birth, physiological transition or survival) by a local parent-pair $cds \in \mathcal{B}$ (and where we used (20)). Note that the probabilities (26)–(27) are both defined to be 0 whenever $N_{abu} = 0$.

The recursion (25) along with (26)–(27) reduces to recursion (A.5) of Alizon and Taylor (2008) who considered only $\mathcal{S}$ class-structure with states being directly group size. The recursion (25) is also conceptually analogous to the recursion for relatedness in the presence of class-structure in the discrete time island model (e.g., Ronce et al., 2000, A.9 for $\mathcal{P}$ and $\mathcal{S}$ class-structures, Ohtsuki et al., 2020, eq. 33 for $\mathcal{S}$ class-structure). Yet owing to the continuous time nature of our model, expressions (26)–(27) do not exactly match those of the discrete time process since only one event (e.g. birth or death) can occur per unit of time in a continuous time process, while two or more events can occur in discrete time models. Finally, note that in a model with no class-structure and thus no dependence on $\mathbf{x}$, the recursion (25) simplifies to

$$r' = c + Dr, \tag{28}$$

where

$$
\begin{aligned}
c &= \frac{2\beta^{\mathrm{CA}}N}{N(N-1)}\frac{1}{q} = \frac{2}{N^2} \\
D &= \frac{\beta^{\mathrm{DP}}}{q} = 2(1-m)\frac{1}{N}\left(\frac{N-1}{N}\right) + \left(1-\frac{2}{N}\right),
\end{aligned}
\tag{29}
$$

and where the expressions and the interpretation for $\beta^{\mathrm{CA}}, \beta^{\mathrm{DP}}$ and q are given in (23)-(24).

3.5 Steady states

We assume that a non-trivial steady state of the metacommunity state dynamics (6) satisfying $\mathbf{0} = \mathbf{Q}(\hat{\mathbf{x}})\hat{\mathbf{x}}$ exists, and that it is hyperbolically stable (we note that for (6) no general results on the existence and stability of the steady states are known). A steady state is said to be hyperbolically stable if the real part of the dominant eigenvalue of the linearized version of the dynamical system, evaluated at the steady state, is negative and bounded away from zero (Hirsch et al., 1974). This implies that there exists a neighborhood of $\hat{\mathbf{x}}$ such that for any initial condition $\mathbf{x}(t_0)$ in this neighborhood where t_0 is some initial time, the vector $\mathbf{x}(t) \to \hat{\mathbf{x}}$ as $t \to \infty$.

Because in (10) the population densities of the focal species satisfy $n_{as} = N_{as}x_s$ for all $as \in \mathcal{C}$ where N_{as} is a constant, then $\mathbf{n}(t) \to \hat{\mathbf{n}}$ whenever $\mathbf{x}(t) \to \hat{\mathbf{x}}$ as $t \to \infty$. Since $\hat{\mathbf{n}}$ must also satisfy $\mathbf{0} = \mathbf{H}(\hat{\mathbf{x}})\hat{\mathbf{n}}$, the vector $\hat{\mathbf{n}}$ is the dominant right eigenvector of $\mathbf{H}(\hat{\mathbf{x}})$. Moreover, because (12) is linear in $\mathbf{v}^{\mathbf{n}}$, we have $\mathbf{v}^{\mathbf{n}}(t) \to \hat{\mathbf{v}}^{\mathbf{n}}$ whenever $\mathbf{x}(t) \to \hat{\mathbf{x}}$ as $t \to \infty$ and the steady state for $\mathbf{v}^{\mathbf{n}}$ can be found by solving $\mathbf{0} = \hat{\mathbf{v}}^{\mathbf{n}}\mathbf{H}(\hat{\mathbf{x}})$. Due to the constraint $\hat{\mathbf{v}}^{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$, the vector $\hat{\mathbf{v}}^{\mathbf{n}}$ is the (unique) dominant left eigenvector of $\mathbf{H}(\hat{\mathbf{x}})$. Finally, the steady-state for relatedness $\mathbf{r}$ can be found by either solving the system of ODEs (22) or the jump-process in (25). By a direct calculation we get

$$
\hat{\mathbf{r}} = (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{c}(\hat{\mathbf{x}}),
\tag{30}
$$

where $\mathbf{I}$ is the identity matrix and $(\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}}))^{-1}$ is the inverse matrix of $\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}})$ evaluated at $\hat{\mathbf{x}}$ thus allowing to compute relatedness explicitly in terms of resident i-processes (Section 3.1)[1]. Because its growth-rate is at most linear in $\mathbf{r}$ and the vector $\mathbf{c}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ depend only on $\mathbf{x}$, we have that $\mathbf{r}(t) \to \hat{\mathbf{r}}$ whenever $\mathbf{x}(t) \to \hat{\mathbf{x}}$ as $t \to \infty$. In summary, we have that whenever $\mathbf{x}(t)$ converges to the hyperbolically stable steady state of interest $\hat{\mathbf{x}}$ as $t \to \infty$, the vectors $(\mathbf{n}(t), \mathbf{v}^{\mathbf{n}}(t), \mathbf{r}(t))$ converge to their unique steady state $(\hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}, \hat{\mathbf{r}})$ as $t \to \infty$.

4 Mutant-resident dynamics in the metacommunity

We now study the full mutant-resident system and to that end we need to consider that phenotypic expression depends on the genetic state (in addition to class), and we denote with $\boldsymbol{z}_\theta \in \mathcal{Z}$ the phenotypic profile of a carrier of allele $\theta \in \{\mathrm{M}, \mathrm{R}\}$. Whenever allele θ is in an individual in class $as \in \mathcal{C}$ the expression of the phenotype is $z_{\theta,as} \in \mathcal{Z}_{\mathcal{C}} \subset \mathbb{R}$. We thus need to expand the group state space of the metacommunity since the different alleles can be distributed in any way across groups. Following the construction of Section 2, it will be convenient to define the state space of groups for the the mutant-resident system as $\Omega = \mathcal{S} \times \mathcal{I}$. The set $\mathcal{I} = \{\mathcal{I}_s\}_{s\in\mathcal{S}}$ gives the set of size distributions of mutants of the focal species over all g-states, where each element $\mathcal{I}_s \in \mathcal{I}$ denotes the set of all possible size distributions in a group $s \in \mathcal{S}$. This is defined as $\mathcal{I}_s = \{\mathcal{I}_{as}\}_{a\in\mathcal{P}}$ with

[1] As a check of the recurrence equations for relatedness, we note that for the homogeneous island model (which is independent of $\mathbf{x}$) and by using (29) we obtain $\hat{r} = c/(1-D) = (1-m)/(1-m+Nm)$, which recovers, as it should, the relatedness for the discrete time Moran process in the island model (Lehmann et al., 2015, eq. 8-e). Note that at the steady state the relation $c = (1-D)\hat{r}$ must hold, saying that the probability c that two randomly sampled offspring have a common parent must be equal to the probability $1-D$ that two randomly sampled individuals have two different parents times the probability $\hat{r}$ that they are related.

$\mathcal{I}_{as} = \{i_{as}\}_{0 \leq i_{as} \leq N_{as}}$ where i_{as} is the number of mutants in physiological state $a \in \mathcal{P}$ that inhabit a group $s \in \mathcal{S}$. We note that an element $i \in \mathcal{I}_s$ is a *vector* giving the number of mutants in each physiological state in a group $s \in \mathcal{S}$ (the set $\mathcal{I}_s$ has potentially many elements because a group $s \in \mathcal{S}$ can be in different genetic states). The number of residents in a group can be obtained simply by subtracting the number of mutants from the total number of individuals e.g. $N_{as} - i_{as}$ is the number of residents $a \in \mathcal{P}$ in a group $s \in \mathcal{S}$. We note that while $\mathcal{S}$ gives the demographic structure of the entire group including non-focal species and external factors, the set $\mathcal{I}$ refers only to the genetic structure of the focal species.

Similarly to Section 3, we need to track the dynamics of mutant-resident group state frequencies in order to characterize the mutant-resident dynamics. We represent the mutant-resident state of the metacommunity with a vector $\mathbf{y} = (y_{si})_{si \in \Omega} \in \Delta(\Omega)$ of length k_Ω, where y_{si} is the frequency of groups in mutant-resident state $(s, i) := si \in \Omega$ (hence $\sum_{si \in \Omega} y_{si} = 1$ and $x_s = \sum_{i \in \mathcal{I}_s} y_{si}$ for all $s \in \mathcal{S}$ where $\mathbf{x} = (x_s)_{s \in \mathcal{S}} \in \Delta(\mathcal{S})$).

4.1 Individual-level processes

To construct the dynamics of $\mathbf{y}$, we use the genetically explicit function for the focal species $\lambda^{\theta}_{auj \leftarrow bsi} : \mathcal{Z}_{\mathcal{C}} \times \mathcal{Z}^2 \times \Omega \times \mathcal{Z}^2 \times \Delta(\Omega) \to \mathbb{R}_+$ for all $au, bs \in \mathcal{C}$ and $j \in \mathcal{I}_u, i \in \mathcal{I}_s$, where $\theta \in \{\mathrm{M}, \mathrm{R}\}$ and where

$$\lambda^{\theta}_{auj,bsi}(\mathbf{y}) = \lambda^{\theta}_{auj \leftarrow bsi}(z_{\theta,bs}, \boldsymbol{z}_{\mathrm{M}}, \boldsymbol{z}_{\mathrm{R}}, si, \boldsymbol{z}_{\mathrm{M}}, \boldsymbol{z}_{\mathrm{R}}, \mathbf{y}) \tag{31}$$

gives the rate at which an individual with allele $\theta \in \{\mathrm{M}, \mathrm{R}\}$ in class $bs \in \mathcal{C}$ in $i \in \mathcal{I}_s$ (alternatively, in physiological state $b \in \mathcal{P}$ in group $si \in \Omega$) and who expresses a phenotype $z_{\theta,bs}$ either *produces* by reproduction or physiological transition, or *removes* by death or physiological transition, an 'offspring' individual $au \in \mathcal{C}$ in $j \in \mathcal{I}_u$. The interpretation of the arguments in (31) are identically interpreted to those of $\lambda_{au \leftarrow bs}$ (3), the only difference is that it takes into account the genetic structure of the group and the metacommunity at large. As in the pure resident population, we can partition (31) into philopatric and dispersal components

$$\lambda^{\theta}_{auj,bsi}(\mathbf{y}) = \lambda^{\theta,\mathrm{p}}_{auj,bsi}(\mathbf{y}) + \lambda^{\theta,\mathrm{d}}_{auj,bsi}(\mathbf{y}), \tag{32a}$$

with

$$\lambda^{\theta,\mathrm{p}}_{auj,bsi}(\mathbf{y}) = \lambda^{\theta,\mathrm{p,B}}_{auj,bsi}(\mathbf{y}) + \lambda^{\theta,\mathrm{T}}_{auj,bsi}(\mathbf{y}) + \lambda^{\theta,\mathrm{D}}_{auj,bsi}(\mathbf{y}) \tag{32b}$$

and

$$\lambda^{\theta,\mathrm{d}}_{auj,bsi}(\mathbf{y}) = \sum_{w \in \mathcal{S}} x_w \lambda^{\theta,\mathrm{d}}_{auj,w,bsi}(\mathbf{y}), \tag{32c}$$

for all $au, bs \in \mathcal{C}, j \in \mathcal{I}_u, i \in \mathcal{I}_s$ and $\theta \in \{\mathrm{M}, \mathrm{R}\}$. Note that the probability that an offspring lands into a group is assumed to be independent of the genetic structure of that group.

4.2 Group state dynamics

The mutant-resident metacommunity dynamics is expressed analogously to the resident metacommunity dynamics (6) as a system of ODEs'

$$\dot{\mathbf{y}} = \mathbf{P}(\mathbf{y})\mathbf{y}, \tag{33}$$

where $\mathbf{P}(\mathbf{y}) = (\pi_{uj,si})_{uj,si\in\Omega}(\mathbf{y}) \in \mathbb{R}^{k_\Omega \times k_\Omega}$ is the group transition-rate matrix with

$$\pi_{uj,si}(\mathbf{y}) = \pi^{\mathrm{p}}_{uj,si}(\mathbf{y}) + \pi^{\mathrm{d}}_{uj,si}(\mathbf{y}) + \pi^{\mathrm{mc}}_{uj,si}(\mathbf{y}), \tag{34}$$

giving the rate at which a group in state $si \in \Omega$ transitions to state $uj \in \Omega$ (we have $\sum_{uj\in\Omega} \pi_{uj,si} = 0$ for all $si \in \Omega$). Here, $\pi^{\mathrm{p}}_{uj,si}(\mathbf{y})$ are the philopatric and $\pi^{\mathrm{d}}_{uj,si}(\mathbf{y})$ the dispersal group transition rates of the focal species, and where all other transition rates due to other species and external factors are given in $\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})$. The transition rates due to the focal species can be expressed in terms of i-processes given in (31) by summing over all physiological states of the parents and their possible progeny, that is, we can write

$$\begin{aligned}
\pi^{\mathrm{p}}_{uj,si}(\mathbf{y}) &= \sum_{a,b\in\mathcal{P}} \bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y}) N_{bs} \\
\pi^{\mathrm{d}}_{uj,si}(\mathbf{y}) &= \sum_{a,c\in\mathcal{P}} \sum_{wk\in\Omega} \bar{\lambda}^{\mathrm{d}}_{auj,s,cwk}(\mathbf{y}) N_{cw} y_{wk},
\end{aligned} \tag{35a}$$

where

$$\begin{aligned}
\bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y}) &= \lambda^{\mathrm{M,p}}_{auj,bsi}(\mathbf{y}) \frac{i_{bs}}{N_{bs}} + \lambda^{\mathrm{R,p}}_{auj,bsi}(\mathbf{y}) \frac{N_{bs} - i_{bs}}{N_{bs}} \\
\bar{\lambda}^{\mathrm{d}}_{auj,s,cwk}(\mathbf{y}) &= \lambda^{\mathrm{M,d}}_{auj,s,cwk}(\mathbf{y}) \frac{k_{cw}}{N_{cw}} + \lambda^{\mathrm{R,d}}_{auj,s,cwk}(\mathbf{y}) \frac{N_{cw} - k_{cw}}{N_{cw}}
\end{aligned} \tag{35b}$$

are the rates of an average individual where the average is taken over the allelic states M and R.

4.3 Mutant-resident population dynamics

Consider the focal species in the metacommunity and let $\mathbf{n}_{\mathrm{M}} = (n_{\mathrm{M},as})_{as\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}}_+$ denote the mutant density vector where $n_{\mathrm{M},as} = \sum_{i\in\mathcal{I}_s} i_{as} y_{si}$ is the density of mutant individuals in class $as \in \mathcal{C}$, and let $\mathbf{n}_{\mathrm{R}} = (n_{\mathrm{R},as})_{as\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}}_+$ denote the resident density vector where $n_{\mathrm{R},as} = \sum_{i\in\mathcal{I}_s} (N_{as} - i_{as}) y_{si}$ is the density of resident individuals in class $as \in \mathcal{C}$. The mutant-resident density dynamics is given by

$$\begin{aligned}
\dot{\mathbf{n}}_{\mathrm{M}} &= \mathbf{G}_{\mathrm{M}}(\mathbf{y}) \mathbf{n}_{\mathrm{M}} \\
\dot{\mathbf{n}}_{\mathrm{R}} &= \mathbf{G}_{\mathrm{R}}(\mathbf{y}) \mathbf{n}_{\mathrm{R}},
\end{aligned} \tag{36}$$

where $\mathbf{G}_{\mathrm{M}}(\mathbf{y}) = (g^{\mathrm{M}}_{au,bs}(\mathbf{y}))_{au,bs\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C} \times k_\mathcal{C}}$ is the mutant fitness matrix with elements

$$g^{\mathrm{M}}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u} \sum_{i\in\mathcal{I}_s} \left[\lambda^{\mathrm{M,d}}_{auj,bsi}(\mathbf{y}) + (j_{au} - i_{as}) \lambda^{\mathrm{M,p}}_{auj,bsi}(\mathbf{y}) + \delta_{a,b} \pi_{uj,si}(\mathbf{y}) \right] y_{\mathrm{M},bsi|\mathrm{M},bs}, \tag{37}$$

giving the rate at which a single mutant in class $bs \in \mathcal{C}$ produces mutants in class $au \in \mathcal{C}$, and where $\mathbf{G}_{\mathrm{R}}(\mathbf{y}) = (g^{\mathrm{R}}_{au,bs}(\mathbf{y}))_{au,bs\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C} \times k_\mathcal{C}}$ is the resident fitness matrix with elements

$$g^{\mathrm{R}}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u} \sum_{i\in\mathcal{I}_s} \left[\lambda^{\mathrm{R,d}}_{auj,bsi}(\mathbf{y}) + \big((N_{au} - j_{au}) - (N_{as} - i_{as})\big) \lambda^{\mathrm{R,p}}_{auj,bsi}(\mathbf{y}) + \delta_{a,b} \pi_{uj,si}(\mathbf{y}) \right] y_{\mathrm{R},bsi|\mathrm{R},bs}, \tag{38}$$

giving the rate at which a single resident in class $bs \in \mathcal{C}$ produces resident in class $au \in \mathcal{C}$. Here, $y_{\mathrm{M},bsi|\mathrm{M},bs} = (i_{bs} y_{si})/n_{\mathrm{M},bs}$ denotes the probability that a mutant $bs \in \mathcal{C}$ inhabits a group in genetic state $i \in \mathcal{I}_s$ and $y_{\mathrm{R},bsi|\mathrm{R},bs} = ((N_{bs} - i_{bs}) y_{si})/n_{\mathrm{R},bs}$ denotes the probability that a resident $bs \in \mathcal{C}$ inhabits a group in genetic state $i \in \mathcal{I}_s$ (Appendix A, Section A.3.1). Note that similarly to the resident population dynamics, all the densities of all species are functions of the mutant-resident metacommunity state $\mathbf{y}$ and hence in order to study the dynamics of $\mathbf{n}_{\mathrm{M}}$ and/or $\mathbf{n}_{\mathrm{R}}$ in (36) they need to be coupled with the dynamics of $\mathbf{y}$ in (33).

4.4 Relative mutant-resident population dynamics

To analyse the spread of the mutant allele, it is convenient to rewrite the mutant-resident population dynamics (36)-(38) by introducing two new vectors. First, (with slight abuse of notation) the vector $\mathbf{n} = (n_{as})_{as\in\mathcal{C}} \in \mathbb{R}_+^{k_\mathcal{C}}$ of total densities, where $n_{as} = N_{as}x_s$ for all $as \in \mathcal{C}$ is the total density of mutant and resident individuals in class $as \in \mathcal{C}$. Second, the vector $\mathbf{p} = (p_{\mathrm{M},as})_{as\in\mathcal{C}} \in [0,1]^{k_\mathcal{C}}$ to describe the within-class mutant frequencies where $p_{\mathrm{M},as} = n_{\mathrm{M},as}/n_{as}$ for all $as \in \mathcal{C}$ is the proportion (frequency) of mutants within a class $as \in \mathcal{C}$. Note that the pair of vectors $(\mathbf{n}_\mathrm{M}, \mathbf{n}_\mathrm{R})$ can be expressed in terms of $(\mathbf{n}, \mathbf{p})$ (and *vice versa*) because $(n_{\mathrm{M},as}, n_{\mathrm{R},as}) = (p_{\mathrm{M},as}n_{as}, (1-p_{\mathrm{M},as})n_{as})$ for all $as \in \mathcal{C}$ (and $(n_{as}, p_{\mathrm{M},as}) = (n_{\mathrm{M},as} + n_{\mathrm{R},as}, n_{\mathrm{M},as}/(n_{\mathrm{M},as} + n_{\mathrm{R},as}))$ for all $as \in \mathcal{C}$). The latter representation being common in population genetics (e.g., Nagylaki, 1992).

The vector of density dynamics satisfies

$$\dot{\mathbf{n}} = \bar{\mathbf{G}}(\mathbf{y})\mathbf{n}, \tag{39}$$

where $\bar{\mathbf{G}}(\mathbf{y}) = (\bar{g}_{au,bs}(\mathbf{y}))_{au,bs\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}\times k_\mathcal{C}}$ is the average mutant-resident fitness matrix where element

$$\bar{g}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[\bar{\lambda}^{\mathrm{d}}_{auj,bsi}(\mathbf{y}) + (N_{au} - N_{as})\bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y}) + \delta_{a,b}\pi_{uj,si}(\mathbf{y})\right] y_{si|s} \tag{40}$$

gives the rate at which an average individual in class $bs \in \mathcal{C}$ produces individuals of class $au \in \mathcal{C}$. The average is taken over the allelic states M, R and over all the genetic states the group it inhabits can be in, so that $y_{si|s} = y_{si}/x_s$ is the probability that the group $s \in \mathcal{C}$ that the average individual inhabits is in genetic state $i \in \mathcal{I}_s$ (see Appendix A, Section A.3.2). The vector of within-class mutant frequency dynamics satisfies

$$\dot{\mathbf{p}} = \mathbf{F}(\mathbf{y})\mathbf{p}, \tag{41}$$

where

$$\mathbf{F} = \mathbf{F}^{\mathrm{sel}}(\mathbf{y}) + \bar{\mathbf{F}}(\mathbf{y}), \tag{42}$$

and where $\mathbf{F}^{\mathrm{sel}}(\mathbf{y}) = (f^{\mathrm{sel}}_{au,bs}(\mathbf{y}))_{au,bs\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}\times k_\mathcal{C}}$ and $\bar{\mathbf{F}}(\mathbf{y}) = (\bar{f}_{au,bs}(\mathbf{y}))_{au,bs\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}\times k_\mathcal{C}}$ with elements, respectively, given by

$$f^{\mathrm{sel}}_{au,bs}(\mathbf{y}) = \frac{n_{bs}}{n_{au}}\left[g^{\mathrm{M}}_{au,bs}(\mathbf{y}) - \bar{g}_{au,bs}(\mathbf{y})\right], \tag{43}$$

and

$$\bar{f}_{au,bs}(\mathbf{y}) = \begin{cases} \frac{n_{bs}}{n_{au}}\bar{g}_{au,bs}(\mathbf{y}) & \text{when } bs \neq au \\ -\sum_{cw\neq au}\frac{n_{cs}}{n_{au}}\bar{g}_{au,cw}(\mathbf{y}), & \text{when } bs = au. \end{cases} \tag{44}$$

Because the rate in (43) describes differential growth due to different individuals having different phenotype it captures the effect of selection on allele frequency change (hence the superscript), and because the rate in (44) depends only on the average phenotype it captures effects on allele frequency change due to individuals in different classes having different (average) growth rates (and we added a bar analogously to (39) as it depends only on the average phenotype where the average is taken over the allelic states, see further discussion in Section 8). A similar partitioning of $\mathbf{F}$ was given for well-mixed populations in Lion (2018b, Appendix 3), and Priklopil and Lehmann (2020, eq. 8a) and the changes due to non-selective forces have been referred to as changes due to "transmission" in population genetics (Kirkpatrick et al., 2002). We note that because $\mathbf{n}$ and $\mathbf{p}$ are functions of

the state of the mutant-resident metacommunity $\mathbf{y}$, in order to study the relative mutant-resident population dynamics of $\mathbf{n}$ and/or $\mathbf{p}$, (39) and/or (44) need to be coupled with the dynamics of $\mathbf{y}$ in (33).

4.5 Average mutant frequency

The total proportion (frequency) of mutants in the metacommunity is given as an arithmetic mean mutant frequency, which is defined as

$$p = \frac{n_{\mathrm{M}}}{n} \; (= \frac{\mathbf{1}}{n} \cdot \mathbf{n}_{\mathrm{M}}) \tag{45}$$

where $n_{\mathrm{M}} = \sum_{as \in \mathcal{S}} n_{\mathrm{M},as}$ is the total mutant density, $n = \sum_{as \in \mathcal{S}} n_{as}$ is the total (mutant and resident) density in the focal species (with a slight abuse of notation) and $\mathbf{1}$ is a vector of all 1 of length $k_{\mathcal{C}}$. The dynamics of p in (45) is obtained by differentiation and by using Sections 4.2-4.4, and it can be written as

$$\begin{aligned} \dot{p} &= \frac{\mathbf{1}}{n} \cdot \left[\mathbf{G}_{\mathrm{M}}(\mathbf{y})\mathbf{n}_{\mathrm{M}} - \bar{\mathbf{G}}(\mathbf{y})\mathbf{n}p\right] \\ &= \frac{1}{n} \sum_{au,bs \in \mathcal{C}} \left[g^{\mathrm{M}}_{au,bs}(\mathbf{y})p_{\mathrm{M},bs} - \bar{g}_{au,bs}(\mathbf{y})p\right] n_{bs}, \end{aligned} \tag{46}$$

and in order to study its dynamics it must be coupled with the dynamics of $\mathbf{y}$ in (33).

5 Mutant-resident dynamics for similar phenotypes

We now study mutant-resident dynamics, and in particular average allele frequency change, for closely similar phenotypes. To that end, we set $\boldsymbol{z}_{\mathrm{M}} = \boldsymbol{z}_{\mathrm{R}} + \delta\boldsymbol{\eta}$ where the vector $\boldsymbol{\eta} = (\eta_{as})_{as \in \mathcal{C}} \in \mathbb{R}^{k_{\mathcal{C}}}$ gives the direction of the deviation (where $\boldsymbol{z}_{\mathrm{R}} = \boldsymbol{z} \in \mathcal{Z}$) and where $0 \leq |\delta| << 1$ is a small parameter. To evaluate the mutant-resident dynamics for small δ, we Taylor expand about $\delta = 0$ the dynamical equations for $\mathbf{y}, \mathbf{n}$ and $\mathbf{p}$ in (33), (39) and (41), respectively. To facilitate the Taylor expansion, we use the following consistency relation relating mutant-resident i-processes (31) to the resident i-processes (3):

$$\lambda^{\theta}_{auj,bsi}(\mathbf{y})\Big|_{\delta=0} = \lambda_{au,bs}(\mathbf{x}), \quad \text{for all } au, bs \in \mathcal{C},\ i \in \mathcal{I}_s, j \in \mathcal{I}_u \text{ and } \theta \in \{\mathrm{M}, \mathrm{R}\}, \tag{47}$$

saying that whenever mutants and residents express identical phenotypes, the i-processes are independent of the genetic structure of the metacommunity. From (47) directly follows that

$$\pi_{uj,si}(\mathbf{y})\Big|_{\delta=0} = q_{u,s}(\mathbf{x}), \quad \text{for all } u, s \in \mathcal{S},\ i \in \mathcal{I}_s, j \in \mathcal{I}_u, \tag{48}$$

and

$$g^{\theta}_{au,bs}(\mathbf{y})\Big|_{\delta=0} = h_{au,bs}(\mathbf{x}), \quad \text{for } \theta \in \{\mathrm{M}, \mathrm{R}\} \text{ and for all } au, bs \in \mathcal{C}, \tag{49}$$

and consequently the growth-rates in (39) and (41) satisfy

$$\begin{aligned} \bar{g}_{au,bs}(\mathbf{y})\Big|_{\delta=0} &= h_{au,bs}(\mathbf{x}) \\ f_{au,bs}(\mathbf{y})\Big|_{\delta=0} &= \widetilde{f}_{au,bs}(\mathbf{x}) = \begin{cases} \frac{n_{bs}}{n_{au}} h_{au,bs}(\mathbf{x}) & \text{when } bs \neq au \\ -\sum_{cw \neq au} \frac{n_{cs}}{n_{au}} h_{au,bs}(\mathbf{x}), & \text{when } bs = au \end{cases} \end{aligned} \tag{50}$$

for all $au, bs \in \mathcal{C}$. It follows from these relations, that the dynamics for $\mathbf{y}, \mathbf{n}$ and $\mathbf{p}$ for small δ can be written as

$$\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{Q}(\mathbf{x})\mathbf{x} + O(\delta) \\
\dot{\mathbf{n}} &= \mathbf{H}(\mathbf{x})\mathbf{n} + O(\delta) \\
\dot{\mathbf{p}} &= \widetilde{\mathbf{F}}(\mathbf{x})\mathbf{p} + O(\delta)
\end{aligned} \tag{51}$$

where we used (48) and (50) and where $\widetilde{\mathbf{F}}(\mathbf{x}) = (\widetilde{f}_{au,bs}(\mathbf{x}))_{au,bs\in\mathcal{C}}$ is the relative fitness matrix for mutants (when $\delta = 0$) with elements as defined in (50). Note that because the dynamics of $\mathbf{n}$ and $\mathbf{p}$ depend only on $\mathbf{x}$ up to order $O(1)$, we have represented the dynamics of the genetically explicit vector $\mathbf{y}$ by using the genetically implicit vector $\mathbf{x}$ whose dynamics in (51) up to order $O(1)$ follows from (33), (48) and $x_s = \sum_{i\in\mathcal{I}_s} y_{si}$. The dynamics of the mean mutant frequency p in (45) for small δ is obtained by Taylor expansion and by using (49)-(50), in which case (46) reduces to

$$\begin{aligned}
\dot{p} &= \frac{\mathbf{1}}{n} \cdot \mathbf{H}(\mathbf{x})\Big[\mathbf{n}_{\mathrm{M}} - \mathbf{n}p\Big] + O(\delta) \\
&= \frac{1}{n} \sum_{au,bs\in\mathcal{C}} h_{au,bs}(\mathbf{x})\big[p_{\mathrm{M},bs} - p\big]n_{bs} + O(\delta),
\end{aligned} \tag{52}$$

which must be coupled with the dynamics for $\mathbf{p}$ and $\mathbf{x}$ in (51).

The primary aim of this paper is to prove the "invasion implies substitution"-principle by way of applying geometric singular perturbation theory developed for *fast-slow ODE systems* in Fenichel (1979). A fast-slow ODE system (Fenichel, 1979; Hek, 2010; Kuehn, 2015) is a system of ODEs where a model parameter tunes the rate at which the dynamical variables operate, such that for small parameter values some variables are fast variables (i.e., their Taylor expansion about this model parameter is dominated by $O(1)$) while others are slow variables (i.e., their Taylor expansion about this model parameter is dominated by $O(\delta)$). We see from (51), however, that for small δ, the dynamics of the vectors $\mathbf{x}, \mathbf{n}$ and $\mathbf{p}$ are all dominated by the terms $O(1)$ and are thus all fast variables. The mutant-resident dynamics (51) is thus not readily in the desired fast-slow form. As a consequence, the mean mutant frequency p in (52) is also a fast variable, and the dynamics (52) does not, perhaps counter-intuitively, reduce to (1). In order to construct an appropriate fast-slow system eventually leading to (52) we proceed in three steps. First, we study the purely fast dynamics of the within-class mutant frequency vector $\mathbf{p}$ where $\delta = 0$, and find its steady state $\hat{\mathbf{p}}$ (Section 5.1). Second, by using the properties of $\hat{\mathbf{p}}$ we find a slow variable that is a proxy for average mutant frequency (Section 5.2). Finally, we study the slow dynamics of this slow variable in order to find all the necessary fast variables that this slower variable depends on (Section 5.3), and, that defines the mutant-resident fast-slow system which will be then studied in Section 6.

5.1 Fast dynamics of the within-class mutant frequency

To study the fast dynamics of the vector $\mathbf{p}$ we set $\delta = 0$ in the third equation in (51), which gives

$$\dot{\mathbf{p}} = \widetilde{\mathbf{F}}(\mathbf{x})\mathbf{p}. \tag{53}$$

We immediately see that because the rows of $\widetilde{\mathbf{F}}$ sum to 0, the steady state solution to (53) is the dominant left eigenvector of the transpose of $\widetilde{\mathbf{F}}$ (which is a transition matrix), and thus satisfies

$$\hat{\mathbf{p}} = \mathbf{1}p_{\mathrm{c}} \tag{54}$$

for some constant $p_{\mathrm{c}} \in \mathbb{R}$, implying that the within-class mutant frequencies at the steady state are all equal $\hat{p}_{\mathrm{M},as} = p_{\mathrm{c}}$ for all $as \in \mathcal{C}$. For biologically meaningful values we need $p_{\mathrm{c}} \in [0, 1]$, and the exact value of p_{c} that $p_{\mathrm{M},as}(t)$ for all $as \in \mathcal{C}$ converges to as $t \to \infty$ depends on the initial condition

$(\mathbf{x}(t_0), \mathbf{p}(t_0))$ for some initial time t_0. The system (53) thus contains infinite number of steady states (equilibria) and the biologically meaningful values lie on a line $[0, 1]$.

The above result has two important consequences. First, because all within-class frequencies are equal $\hat{p}_{\mathrm{M},as} = p_{\mathrm{c}}$ for all $as \in \mathcal{C}$, *every* average mutant frequency, defined for all t as

$$p_{\boldsymbol{\phi}} = \boldsymbol{\phi} \cdot \mathbf{n}_{\mathrm{M}} \; (= \sum_{as \in \mathcal{C}} \phi_{as} n_{as} p_{\mathrm{M},as}) \tag{55}$$

where $\boldsymbol{\phi} = (\phi_{as})_{as \in \mathcal{C}}$ is some arbitrary vector of weights normalized such that $\boldsymbol{\phi} \cdot \mathbf{n} = 1$, satisfies

$$\hat{p}_{\boldsymbol{\phi}} = p_{\mathrm{c}} \tag{56}$$

at the steady state. This is in particular true also for the arithmetic mean frequency p in (45) (where $\boldsymbol{\phi} = \mathbf{1}/n$). That is, every average mutant frequency $p_{\boldsymbol{\phi}}(t) \to p_{\mathrm{c}}$ as $t \to \infty$, and where the exact value of p_{c} depends on the initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$. Second, the $k_{\mathcal{C}}$-dimensional vector $\mathbf{p}$ is, at the steady state, a 1-dimensional vector $\hat{\mathbf{p}} = \mathbf{1}p_{\mathrm{c}}$ because it can be spanned by using a single vector $\mathbf{1}p_{\mathrm{c}}$ (Hirsch et al., 1974). This will play an important role in the proof of the "invasion implies substitution"-principle (Section 7).

5.2 Reproductive value weighted mutant frequency

In the previous Section 5.1 we found that for any weights $\boldsymbol{\phi}$, the average mutant frequency (55) satisfies $p_{\boldsymbol{\phi}}(t) \to p_{\mathrm{c}}$ as $t \to \infty$ (and where the exact value of p_{c} depends on the initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$). Motivated by this, if one were to find weights $\boldsymbol{\phi}$ for which $p_{\boldsymbol{\phi}}(t) = p_{\mathrm{c}}$ for all $t \geq t_0$, then such an average frequency would stay constant in the fast dynamics where $\delta = 0$, and hence would be a slow variable for a system where δ is small but non-zero (a pre-requisite to construct a fast-slow system). We can achieve this by setting $\boldsymbol{\phi} = \mathbf{v}^{\mathbf{n}}$ where the weights $\mathbf{v}^{\mathbf{n}}$ are (with a slight abuse of notation) defined as average reproductive values (where the average is taken over allelic states, Priklopil and Lehmann, 2020, see also Lion, 2018a,b) and is thus the mutant-resident analogue to the one defined for the resident population in Section 3. Such average individual reproductive values satisfy

$$-\dot{\mathbf{v}}^{\mathbf{n}} = \mathbf{v}^{\mathbf{n}} \bar{\mathbf{G}}(\mathbf{y}), \tag{57}$$

with a scaling $\mathbf{v}^{\mathbf{n}} \cdot \mathbf{n} = 1$ (where $\mathbf{n}$ is as in (39)), and by defining a reproductive values weighted mutant frequency as

$$p_{\mathbf{v}} = \mathbf{v}^{\mathbf{n}} \cdot \mathbf{n}_{\mathrm{M}} \; (= \sum_{as \in \mathcal{C}} v^n_{as} n_{as} p_{\mathrm{M},as}), \tag{58}$$

its dynamics is obtained by differentiation and satisfies

$$\begin{aligned} \dot{p}_{\mathbf{v}} &= \mathbf{v}^{\mathbf{n}} \cdot \Big[\mathbf{G}_{\mathrm{M}}(\mathbf{y}) - \bar{\mathbf{G}}(\mathbf{y}) \Big] \mathbf{n}_{\mathrm{M}} \\ &= \sum_{au,bs \in \mathcal{C}} v^n_{au} \Big(g^{\mathrm{M}}_{au,bs}(\mathbf{y}) - \bar{g}_{au,bs}(\mathbf{y}) \Big) p_{\mathrm{M},bs} n_{bs}. \end{aligned} \tag{59}$$

Indeed, because $\mathbf{G}_{\mathrm{M}}(\mathbf{y}) - \bar{\mathbf{G}}(\mathbf{y}) = \mathbf{0}$ for $\delta = 0$ (49)-(50), the Taylor expansion of (59) about $\delta = 0$ is dominated by $O(\delta)$ and hence the reproductive value weighted average mutant frequency $p_{\mathbf{v}}$ is a slow variable. We note that (59) is valid for arbitrary δ (and thus arbitrary strength of selection) and in order to study its dynamics it must be coupled with the dynamics of $\mathbf{y}$ in (33).

5.3 Slow dynamics of the weighted mutant frequency

Here, we study the slow dynamics of the weighted mutant frequency $p_{\mathbf{v}}$ in (58)-(59), with a specific aim to identify all the variables its dynamics depends on in order to construct a fast-slow mutant-resident system. We study the slow dynamics of $p_{\mathbf{v}}$ by Taylor expanding (59) about $\delta = 0$ and then change the timescale to slow time while letting δ go to 0. To this end, we introduce a slow time variable τ, set $\tau = \delta t$ from which we get the relation $\frac{\mathrm{d}}{\mathrm{d}t} = \delta \frac{\mathrm{d}}{\mathrm{d}\tau}$, and by letting δ go to 0 we get

$$\frac{\mathrm{d}}{\mathrm{d}\tau} p_{\mathbf{v}} = \frac{\mathrm{d}}{\mathrm{d}\delta} \left[\sum_{au,bs \in \mathcal{C}} v_{au}^{n} \Big(g_{au,bs}^{\mathrm{M}}(\mathbf{y}) - \bar{g}_{au,bs}(\mathbf{y}) \Big) p_{\mathrm{M},bs} n_{bs} \right]_{\delta=0}. \tag{60}$$

Next, we take the derivative in (60) in such a way that it allows us to arrive at an expression for the directional selection coefficient as given in (2). To do this, we must take separate (partial) derivatives with respect to phenotypes of individuals according to the 'spatial' relationship between the focal individual and other individuals from the focal species who affect its fitness either via interactions or directly (the effect on itself). For ease of presentation, we label the phenotypes and write the fitness (11) as

$$h_{au,bs}(\mathbf{x}) = h_{au \leftarrow bs}(z_{\mathrm{foc},bs}, \boldsymbol{z}_{\mathrm{loc}}, s, \boldsymbol{z}_{\mathrm{pop}}, \mathbf{x}), \tag{61}$$

where $z_{\mathrm{foc},bs}$ (and $\boldsymbol{z}_{\mathrm{foc}}$) indicates the phenotype of the focal individual (hence the subscript), $\boldsymbol{z}_{\mathrm{loc}}$ indicates the phenotype of the local group members of the focal individual (hence the subscript) but excluding the focal individual, and $\boldsymbol{z}_{\mathrm{pop}}$ indicates the phenotype of individuals outside of the group of the focal individual (i.e. population/species at large, hence the subscript). We emphasise that these are just labels to distinguish the different partial derivatives we will be taking in Section 5.3.1 (indeed, all individuals in the resident metacommunity express the resident phenotype $\boldsymbol{z} \in \mathcal{Z}$). Likewise, we write $g_{au,bs}(\mathbf{y}) = g_{au \leftarrow bs}(z_{\mathrm{M},bs}, \boldsymbol{z}_{\mathrm{M,loc}}, \boldsymbol{z}_{\mathrm{R,loc}}, si, \boldsymbol{z}_{\mathrm{M,pop}}, \boldsymbol{z}_{\mathrm{R,pop}}, \mathbf{y})$ for a focal mutant and $\bar{g}_{au,bs}(\mathbf{y}) = g_{au \leftarrow bs}(\boldsymbol{z}_{\mathrm{M,loc}}, \boldsymbol{z}_{\mathrm{R,loc}}, si, \boldsymbol{z}_{\mathrm{M,pop}}, \boldsymbol{z}_{\mathrm{R,pop}}, \mathbf{y})$ for an average individual where the notation and interpretation of $\boldsymbol{z}_{\mathrm{M_{foc}}}, \boldsymbol{z}_{\mathrm{M_{loc}}}$ and $\boldsymbol{z}_{\mathrm{M_{pop}}}$ (and $\boldsymbol{z}_{\mathrm{R_{foc}}}, \boldsymbol{z}_{\mathrm{R_{loc}}}$ and $\boldsymbol{z}_{\mathrm{R_{pop}}}$) for the mutant-resident metacommunity is similar (see Appendix C for more details).

5.3.1 Properties of fitness

To take the derivative in (60) we apply a property that relates, for small δ, mutant-resident fitnesses (37)-(38) to the resident fitness (11). This property, which follows from the so-called generalized law of mass action (Diekmann et al., 2001; Meszéna et al., 2005; Dercole, 2016), says that the perturbation of the mutant phenotype of all mutant individuals in a spatially well-mixed mutant-resident species, is equal to the perturbation of the phenotype of all individuals in a purely resident species when multiplied by the frequency of mutant individuals. This property holds for non-linear environmental feedbacks and is secured by assuming that all individuals undergo pairwise interactions in continuous-time and that individuals with the same phenotype are exchangeable (Dercole, 2016). Below we will provide an analogue of this property for the present metacommunity model.

For small but non-zero δ, the mutant fitness matrix $\mathbf{G}_{\mathrm{M}}(\mathbf{y})$ is related to the resident fitness matrix $\mathbf{H}(\mathbf{x})$ in the following way (see Appendix C for more details). First,

$$\frac{\partial g_{au,bs}^{\mathrm{M}}(\mathbf{y})}{\partial z_{\mathrm{M_{foc}},bs}}\Big|_{\delta=0} = \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} \quad \text{for all } au, bs \in \mathcal{C} \tag{62}$$

and where all other partial derivatives with respect to $z_{\mathrm{M_{foc}},cw}$ for $cw \neq bs \in \mathcal{C}$ are 0. Second,

$$\frac{\partial g_{au,bs}^{\mathrm{M}}(\mathbf{y})}{\partial z_{\mathrm{M_{loc}},cs}}\Big|_{\delta=0} = \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} p_{\mathrm{M},cs|\mathrm{M},bs} \quad \text{for all } au, bs \in \mathcal{C}, c \in \mathcal{P} \tag{63}$$

where all other partial derivatives with respect to $z_{\mathrm{M}_{\mathrm{loc}},cw}$ for $w \neq s \in \mathcal{S}$ are 0, and finally,

$$\frac{\partial g^{\mathrm{M}}_{au,bs}(\mathbf{y})}{\partial z_{\mathrm{M}_{\mathrm{pop}},cw}}\Big|_{\delta=0} = \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{pop},cw}} p_{\mathrm{M},cw} \quad \text{for all } au, bs, cw \in \mathcal{C}. \tag{64}$$

In (63) we introduced the *conditional* within-group mutant frequency $\mathbf{p}_{\mathrm{M}|\mathrm{M}} = (p_{\mathrm{M},as|\mathrm{M},bs})_{abs\in\mathcal{B}} \in [0,1]^{k_{\mathcal{B}}}$ defined as

$$p_{\mathrm{M},as|\mathrm{M},bs} = \frac{p_{\mathrm{MM},abs}}{p_{\mathrm{M},bs}}, \quad \text{for all } a, b \in \mathcal{P}, s \in \mathcal{S}, \tag{65}$$

where $p_{\mathrm{MM},abs}$ is an element of the vector of (within-group) mutant-pair frequencies $\mathbf{p}_{\mathrm{MM}} = (p_{\mathrm{MM},abs})_{abs\in\mathcal{B}} \in [0,1]^{k_{\mathcal{B}}}$ defined as

$$p_{\mathrm{MM},abs} = \frac{n_{\mathrm{MM},abs}}{n_{abs}}, \quad \text{for all } abs \in \mathcal{B}. \tag{66}$$

Here, $n_{\mathrm{MM},abs} = \sum_{i\in\mathcal{I}} i_{as} i_{bs} y_{si}$ denotes the density of mutant-pairs $abs \in \mathcal{B}$ and n_{abs} is the total density (with a slight abuse of notation) of pairs $abs \in \mathcal{B}$ (i.e. the total density of pairs independent of their phenotype). The conditional mutant frequencies $\mathbf{p}_{\mathrm{M}|\mathrm{M}}$ thus give the frequency of mutants conditional that another (randomly sampled) individual in the group is a mutant.

In (62), the partial derivative is taken with respect to the focal individual itself and because any mutant individual is (trivially) a mutant with probability 1, the proportionality coefficient is simply 1 (i.e., the frequency of a mutant allele in a mutant individual is 1). The derivative on the right-hand-side measures the 'additional' offspring the focal individual produces (per unit of time) due to the focal individual itself expressing the mutation ('direct fitness effect'). In (63), the partial derivative with respect to all local individuals is proportional to the conditional within-group mutant frequency in class $cs \in \mathcal{C}$, $p_{\mathrm{M},cs|\mathrm{M},bs}$, conditional on the focal individual being a mutant $bs \in \mathcal{C}$. This property reflects the fact that genetic correlations between conspecifics build up within groups of finite size, and the derivative on the right-hand-side measures the 'additional' offspring the focal individual produces (per unit of time) due to interactions with all local individuals that express the mutation ('indirect fitness effect'). Finally, the partial derivative with respect to non-local individuals in (64) is proportional to the (unconditional) within-class mutant frequency because no genetic correlations build up between members in different groups in the infinite island model (see, e.g., Rousset, 2004 for the finite island model).

Using (62)-(64) and analogues properties for (mutant-resident) resident fitness (not shown here), the average fitness matrix $\bar{\mathbf{G}}(\mathbf{y})$ is related to the (pure resident) resident fitness matrix $\mathbf{H}(\mathbf{x})$ as

$$\begin{aligned}
\frac{\partial \bar{g}_{au,bs}(\mathbf{y})}{\partial z_{\mathrm{M}_{\mathrm{foc}},bs}}\Big|_{\delta=0} &= \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} p_{\mathrm{M},bs} \\
\frac{\partial \bar{g}_{au,bs}(\mathbf{y})}{\partial z_{\mathrm{M}_{\mathrm{loc}},cs}}\Big|_{\delta=0} &= \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} p_{\mathrm{M},cs} \\
\frac{\partial \bar{g}_{au,bs}(\mathbf{y})}{\partial z_{\mathrm{M}_{\mathrm{pop}},cw}}\Big|_{\delta=0} &= \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{pop},cw}} p_{\mathrm{M},cw} \quad \text{for all } au, bs, cw \in \mathcal{C},
\end{aligned} \tag{67}$$

and where all other partial derivatives with respect to $z_{\mathrm{M}_{\mathrm{foc}},cw}$ for $cw \neq bs \in \mathcal{C}$ and with respect to $z_{\mathrm{M}_{\mathrm{loc}},cw}$ for $w \neq s \in \mathcal{S}$ are all 0. Notice that all proportionality coefficients are (unconditional) within-class mutant frequencies. In the first relation in (67) this is because an average individual is a mutant with probability given by the within-class mutant frequency, in the second relation this is because no genetic correlations arise for an average individual and in the final relation this is because no genetic correlations arise between non-group members.

5.3.2 Slow dynamics of the weighted mutant frequency and the fast variables

Taking the derivative in (60) and partitioning it according to the different individuals as discussed above, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\tau}p_{\mathbf{v}} = \sum_{au,bs\in\mathcal{C}} v^n_{au}\Big[\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}}\eta_{bs}p_{\mathrm{M},bs}(1-p_{\mathrm{M},bs})+ \\ + \sum_{cs\in\mathcal{C}}\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}}\eta_{cs}(p_{\mathrm{MM},bcs}-p_{\mathrm{M},bs}p_{\mathrm{M},cs})\Big]n_{bs}, \tag{68}$$

where we used the definition for directional derivatives, the properties (62)-(64), (67) and (65). We can see from (68) that the dynamics of the weighted mutant frequency $p_{\mathbf{v}}$ (slow variable) depends on the dynamics of $\mathbf{x}, \mathbf{n}, \mathbf{p}, \mathbf{v^n}$ and $\mathbf{p}_{\mathrm{MM}}$. Because $\mathbf{x}, \mathbf{n}, \mathbf{p}$ are fast variables (as discussed earlier in this section) and $\mathbf{v^n}$ is also fast because under consistency relation (49) its dynamics reduces to (12), the next task is to find the equation for a vector of mutant-pair frequencies $\mathbf{p}_{\mathrm{MM}}$ (Section 5.3.3), with a particular aim to investigate whether it is a fast or a slow variable.

5.3.3 Mutant-pair frequencies

In order to see whether $\mathbf{p}_{\mathrm{MM}}$ defined in (66) is a fast or slow variable, we study its dynamical equation for $\delta = 0$ (Appendix B, Section B.2). The dynamics of $\mathbf{p}_{\mathrm{MM}} = (p_{\mathrm{MM},abs})_{abu\in\mathcal{B}}$ is given as

$$\dot{p}_{\mathrm{MM},abu} = \sum_{s\neq u\in\mathcal{S}}\sum_{d\in\mathcal{P}}\frac{n_{ds}}{n_{abu}}\Big[\beta^{\mathrm{CA}}_{abu,ds}+\beta^{\mathrm{IM}}_{abu,ds}\Big]p_{\mathrm{M},ds} + \sum_{s\neq u\in\mathcal{S}}\sum_{c,d\in\mathcal{P}}\frac{n_{cds}}{n_{abu}}\beta^{\mathrm{DP}}_{abu,cds}p_{\mathrm{MM},cds} - \\ - p_{\mathrm{MM},abu}\frac{N_{abu}}{n_{abu}}\sum_{s\neq u\in\mathcal{S}}q_{u,s}x_s \quad \text{for all } abu\in\mathcal{B}, \tag{69}$$

which can be written in vector form as

$$\dot{\mathbf{p}}_{\mathrm{MM}} = \mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}). \tag{70}$$

Here, $\mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}) \in \mathbb{R}^{k_{\mathcal{B}}}$ is a vector of rates at which $\mathbf{p}_{\mathrm{MM}}$ changes, and because the elements are non-zero the vector of mutant-pair frequencies $\mathbf{p}_{\mathrm{MM}}$ is a fast variable (i.e., the term $O(1)$ in the Taylor expansion does not vanish).

To characterise the steady state of $\mathbf{p}_{\mathrm{MM}}$ we focus for convenience on the discrete-time jump-process dynamics, which is given by

$$\mathbf{p}'_{\mathrm{MM}} = \big[\mathbf{C}(\mathbf{x}) + \mathbf{M}(\mathbf{p}, \mathbf{x})\big]\mathbf{p} + \mathbf{D}(\mathbf{x})\mathbf{p}_{\mathrm{MM}}, \tag{71a}$$

where $\mathbf{C}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ are as in (25), and $\mathbf{M}(\mathbf{p}, \mathbf{x}) = (M_{abu,ds}(\mathbf{p}, \mathbf{x}))_{abu\in\mathcal{B},ds\in\mathcal{C}} \in \mathbb{R}^{k_{\mathcal{B}}\times k_{\mathcal{C}}}$ where $M_{abu,ds}(\mathbf{p}, \mathbf{x}) = \sum_{ew\in\mathcal{C}} M_{abu,ds,ew}(\mathbf{x})p_{\mathrm{M},ew}$ and

$$M_{abu,ds,ew}(\mathbf{x}) = \frac{1}{N_{abu}}\frac{\beta^{\mathrm{IM}}_{abu,ds,ew}(\mathbf{x})N_{ds}x_sN_{ew}x_w}{\sum_{w\neq u\in\mathcal{S}} q_{u,w}(\mathbf{x})x_w} \tag{71b}$$

gives the conditional probability that a randomly sampled mutant-pair $abu \in \mathcal{B}$ was produced when a local mutant parent $bs \in \mathcal{C}$ paired up with a new immigrant mutant offspring $ew \in \mathcal{C}$ (note the dependency on the mutant frequency in (71)). For all pairs $abu \in \mathcal{B}$ we have

$$\sum_{ds\in\mathcal{C}} C_{abu,ds}(\mathbf{x}) + \sum_{ds\in\mathcal{C}}\sum_{ew\in\mathcal{C}} M_{abu,ds,ew}(\mathbf{x}) + \sum_{cds\in\mathcal{B}} D_{abu,cds}(\mathbf{x}) = 1, \tag{72}$$

saying that with probability 1 each pair of individuals has either a common parent, different local parents, or non-local parents (the rows of the matrices $\mathbf{C}(\mathbf{x})$, $\mathbf{D}(\mathbf{x})$ and $\mathbf{M}(\mathbf{p}, \mathbf{x})$ sum to 1).

The steady state solution $\hat{\mathbf{p}}_{\mathrm{MM}}$ is obtained by solving $\hat{\mathbf{p}}_{\mathrm{MM}} = [\mathbf{C}(\hat{\mathbf{x}}) + \mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{x}})]\hat{\mathbf{p}} + \mathbf{D}(\hat{\mathbf{x}})\hat{\mathbf{p}}_{\mathrm{MM}}$ in (71) evaluated at $\hat{\mathbf{p}}$ and $\hat{\mathbf{x}}$, and we get

$$\begin{aligned}
\hat{\mathbf{p}}_{\mathrm{MM}} &= (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{C}(\hat{\mathbf{x}})\hat{\mathbf{p}} + (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{x}})\hat{\mathbf{p}} \\
&= (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{C}(\hat{\mathbf{x}})\mathbf{1}p_{\mathrm{c}} + (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}(\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}) - \mathbf{C}(\hat{\mathbf{x}}))\mathbf{1}p_{\mathrm{c}}^2 \\
&= (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{c}(\hat{\mathbf{x}})p_{\mathrm{c}} + (\mathbf{1} - (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{c}(\hat{\mathbf{x}}))p_{\mathrm{c}}^2 \\
&= \hat{\mathbf{r}}p_{\mathrm{c}} + (\mathbf{1} - \hat{\mathbf{r}})p_{\mathrm{c}}^2
\end{aligned} \tag{73}$$

where we used (54), (72) and (30), respectively, and where $p_{\mathrm{c}} \in [0, 1]$ is a constant that depends on the initial condition as discussed in Section 5.1 (we can also obtain this solution by solving (69)-(70)). We have thus obtained that whenever $\mathbf{x}(t) \to \hat{\mathbf{x}}$ as $t \to \infty$ (Section 3.5), then $\mathbf{p}_{\mathrm{MM}}(t) \to \hat{\mathbf{p}}_{\mathrm{MM}} = \hat{\mathbf{r}}p_{\mathrm{c}} + (\mathbf{1} - \hat{\mathbf{r}})p_{\mathrm{c}}^2$ as $t \to \infty$. Moreover, $\hat{\mathbf{p}}_{\mathrm{MM}}$ can be expressed in terms of the steady state value for relatedness $\hat{\mathbf{r}}$ (Section 3.5) and p_{c} (Section 5.1), and hence its general solution consist of infinite (curve of) steady states (equilibria) because it is a quadratic function of $p_{\mathrm{c}} \in [0, 1]$. Finally, we note that we have made no assumptions on the relationship between within-group mutant frequencies $\mathbf{p}_{\mathrm{MM}}(t)$, relatedness $\mathbf{r}(t)$ and the average mutant frequency $p_{\boldsymbol{\phi}}(t)$ for $t < \infty$: their relationship given in (73) is obtained only as a limit when $t \to \infty$.

6 Mutant-resident dynamics as a fast-slow system

Here, we use Section 5 to set up and analyse a fast-slow ODE system where mutant and resident individuals have closely similar phenotypes ($0 \leq |\delta| << 1$). In Section 5 we found, in particular, that the reproductive value weighted mutant frequency $p_{\mathbf{v}}$ in (58) is a slow variable and that its dynamics depends on the fast variables $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}})$. The fast-slow system for the mutant-resident dynamics is thus given in fast time t as

$$\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{Q}(\mathbf{x})\mathbf{x} + O(\delta) \\
\dot{\mathbf{n}} &= \mathbf{H}(\mathbf{x})\mathbf{n} + O(\delta) \\
-\dot{\mathbf{v}}^{\mathbf{n}} &= \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}) + O(\delta) \\
\dot{\mathbf{p}} &= \widetilde{\mathbf{F}}(\mathbf{x})\mathbf{p} + O(\delta) \\
\dot{\mathbf{p}}_{\mathrm{MM}} &= \mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}) + O(\delta) \\
\dot{p}_{\mathbf{v}} &= \delta \sum_{au,bs \in \mathcal{C}} v_{au}^{n} \Big[\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} \eta_{bs} p_{\mathrm{M},bs}(1 - p_{\mathrm{M},bs}) + \\
&\qquad + \sum_{cs \in \mathcal{C}} \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} \eta_{cs}(p_{\mathrm{MM},bcs} - p_{\mathrm{M},bs}p_{\mathrm{M},cs}) \Big]_{\delta=0} n_{bs} + O(\delta^2),
\end{aligned} \tag{74}$$

which can be equivalently written in slow time $\tau = \delta t$ as

$$
\begin{aligned}
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{x} &= \mathbf{Q}(\mathbf{x})\mathbf{x} + O(\delta) \\
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{n} &= \mathbf{H}(\mathbf{x})\mathbf{n} + O(\delta) \\
-\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{v}^{\mathbf{n}} &= \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}) + O(\delta) \\
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{p} &= \widetilde{\mathbf{F}}(\mathbf{x})\mathbf{p} + O(\delta) \\
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{p}_{\mathrm{MM}} &= \mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}) + O(\delta) \\
\frac{\mathrm{d}}{\mathrm{d}\tau}p_{\mathbf{v}} &= \sum_{au,bs\in\mathcal{C}} v^n_{au}\Big[\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}}\eta_{bs}p_{\mathrm{M},bs}(1-p_{\mathrm{M},bs}) + \\
&\qquad + \sum_{cs\in\mathcal{C}} \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}}\eta_{cs}(p_{\mathrm{MM},bcs} - p_{\mathrm{M},bs}p_{\mathrm{M},cs})\Big]_{\delta=0} n_{bs} + O(\delta),
\end{aligned}
\tag{75}
$$

where we used $\frac{\mathrm{d}}{\mathrm{d}t} = \delta\frac{\mathrm{d}}{\mathrm{d}\tau}$. The two systems (74) and (75) are equivalent, the only difference is the notation. The above confirms that the variable $p_{\mathbf{v}}$ is indeed a slow variable because in fast time (74) it depends on the terms $O(\delta)$, and, the dynamics of $p_{\mathbf{v}}$ depends on the fast variables $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}})$ whose dynamics depends on $O(1)$. Next we study the fast and slow subsystems of (74) and (75) by setting $\delta = 0$, respectively, and then in Section 7 we will join the two subsystems together.

6.1 Fast subsystem and the critical manifold

We obtain the fast subsystem by setting $\delta = 0$ in (74), which yields

$$
\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{Q}(\mathbf{x})\mathbf{x} \\
\dot{\mathbf{n}} &= \mathbf{H}(\mathbf{x})\mathbf{n} \\
-\dot{\mathbf{v}}^{\mathbf{n}} &= \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}) \\
\dot{\mathbf{p}} &= \widetilde{\mathbf{F}}(\mathbf{x})\mathbf{p} \\
\dot{\mathbf{p}}_{\mathrm{MM}} &= \mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}) \\
\dot{p}_{\mathbf{v}} &= 0.
\end{aligned}
\tag{76}
$$

The fast variables $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}})$ thus change in the fast subsystem while the slow variable $p_{\mathbf{v}}$ stays constant. More specifically, whenever $\mathbf{x}(t) \to \hat{\mathbf{x}}$ as $t \to \infty$ (Section 3.5), then also $(\mathbf{n}(t), \mathbf{v}^{\mathbf{n}}(t), \mathbf{p}(t), \mathbf{p}_{\mathrm{MM}}(t)) \to (\hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}, \hat{\mathbf{p}}, \hat{\mathbf{p}}_{\mathrm{MM}})$. Recall that the vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{v}}^{\mathbf{n}}$ describe the unique steady state of the size and class structure (Section 3.5), and $\hat{\mathbf{p}} = \mathbf{1}p_{\mathrm{c}}$ and $\hat{\mathbf{p}}_{\mathrm{MM}} = \hat{\mathbf{r}}p_{\mathrm{c}} + (\mathbf{1}-\hat{\mathbf{r}})p_{\mathrm{c}}^2$ give its genetic structure at the steady state of the fast subsystem. Because $p_{\mathbf{v}}(t) = p_{\mathrm{c}}$ for all t (Section 5.1), we can *identify* the steady state p_{c} with the value of $p_{\mathbf{v}}$ in the fast subsystem (76). Thus, as $t \to \infty$, the general solution of $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}}, p_{\mathbf{v}})$ converges to a curve $(\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}, \hat{\mathbf{p}}, \hat{\mathbf{p}}_{\mathrm{MM}}, p_{\mathbf{v}})$ that is parametrized by $p_{\mathbf{v}} \in [0, 1]$, and a specific 'initial value problem' with an initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$ for some initial time t_0 converges to a single point on this curve.

We collect all the steady states of the fast subsystem into a single set

$$
\begin{aligned}
\mathcal{M}_0 = \{&(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}}, p_{\mathbf{v}}) \in \mathbb{R}^{k_{\mathcal{S}}} \times \mathbb{R}^{k_{\mathcal{C}}} \times \mathbb{R}^{k_{\mathcal{C}}} \times \mathbb{R}^{k_{\mathcal{C}}} \times \mathbb{R}^{k_{\mathcal{B}}} \times \mathbb{R} \ : \\
&\mathbf{x} = \hat{\mathbf{x}},\ \mathbf{n} = \hat{\mathbf{n}}, \mathbf{v}^{\mathbf{n}} = \hat{\mathbf{v}}^{\mathbf{n}},\ \mathbf{p} = \mathbf{1}p_{\mathbf{v}},\ \mathbf{p}_{\mathrm{MM}} = \hat{\mathbf{r}}p_{\mathbf{v}} + (\mathbf{1}-\hat{\mathbf{r}})p_{\mathbf{v}}^2,\ p_{\mathbf{v}} \in [0,1]\},
\end{aligned}
\tag{77}
$$

which defines the critical manifold of the system (e.g., Kuehn, 2015 for general considerations and Priklopil and Lehmann, 2020 in the context of the "invasion implies substitution"-principle). The critical manifold thus gives the steady states of the *general* solution of (76) and is a curve consisting of an infinite number of steady states (equilibria) parametrized by $p_{\mathbf{v}}$ (as discussed in the above

paragraph). The exact steady state point on $\mathcal{M}_0$ that the fast subsystem converges to depends on the given initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$ (see for a graphical representation in Figure 2, panels (a)-(c)). Because $\hat{\mathbf{x}}$ is assumed to be hyperbolically stable (Section 3.5) the critical manifold is said to be a normally hyperbolically stable invariant manifold. For such manifolds the dynamics near the manifold is dominated by the hyperbolicity condition (i.e. in the normal direction to the manifold) while the dynamics on the manifold is constant (in the fast subsystem). Indeed, all the variables are constant on $\mathcal{M}_0$ and because $p_{\mathbf{v}}$ is defined on an interval $[0, 1]$ the manifold is a 1-dimensional manifold in a $(3k_{\mathcal{C}} + k_{\mathcal{S}} + k_{\mathcal{B}} + 1)$-dimensional space that characterizes the state space of all the variables of the fast-slow mutant-resident system (74)-(75). This manifold plays an important role in the proof of the "invasion implies substitution"-principle: in the fast-slow mutant-resident system where δ is small but non-zero the critical manifold persists as a so-called slow manifold (Fenichel, 1979), and, the dynamics of the fast-slow mutant-resident system (74)-(75) on and near the slow-manifold is equivalent to the dynamics of $p_{\mathbf{v}}$ on and near $\mathcal{M}_0$ in the fast and slow subsystems (slow subsystem is studied in Section 6.2).

6.2 Slow subsystem

We obtain the slow subsystem by setting $\delta = 0$ in (75). However, motivated by Section 6.1, we are in fact not interested in every possible slow subsystem of (75), only the one associated to $\hat{\mathbf{x}}$ and the critical manifold $\mathcal{M}_0$ (77) (indeed, there may be other steady states satisfying $\mathbf{0} = \mathbf{Q}(\mathbf{x})\mathbf{x}$ and hence other critical manifolds and slow subsystems). The slow subsystem of interest is then given jointly by the critical manifold $\mathcal{M}_0$ in (77) and the dynamics of the weighted mutant frequency $p_{\mathbf{v}}$ restricted to this manifold,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}p_{\mathbf{v}} &= \sum_{au,bs\in\mathcal{C}}\Bigg[v^{n}_{au}\Big[\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}}\eta_{bs}p_{\mathrm{M},bs}(1-p_{\mathrm{M},bs})+ \\ &\qquad + \sum_{cs\in\mathcal{C}}\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}}\eta_{cs}(p_{\mathrm{MM},bcs}-p_{\mathrm{M},bs}p_{\mathrm{M},cs})\Big]n_{bs}\Bigg]_{\mathcal{M}_0} \\ &= p_{\mathbf{v}}(1-p_{\mathbf{v}})S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}),\end{aligned} \tag{78a}$$

where

$$S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) = \sum_{au,bs\in\mathcal{C}}\hat{v}^{\mathrm{n}}_{au}\Big[\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{foc},bs}}\eta_{bs} + \sum_{cs\in\mathcal{C}}\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{loc},cs}}\hat{r}_{bcs}\ \eta_{cs}\Big]\hat{n}_{bs} \tag{78b}$$

is the directional selection coefficient evaluated at the critical manifold $\mathcal{M}_0$. Recall that since $\mathcal{M}_0$ is a curve parametrized by $p_{\mathbf{v}} \in [0, 1]$ (Section 6.1), the critical manifold $\mathcal{M}_0$ can be seen as the state space for the dynamics of (78). Also, note that to obtain equation (78) we have taken the limits $\delta \to 0$ and $t \to \infty$ (and throughout we have assumed that the total population size $K = \infty$). Equation (78) can thus be seen as the first moment of change in the diffusion approximation for slow subsystems (Ethier and Nagylaki, 1980, 1988).

6.2.1 Solution of the slow subsystem

We found that in the slow subsystem the dynamics of the weighted mutant frequency $p_{\mathbf{v}}$ can be solved from a single ODE given in (78). We can express this in the context of an initial value problem as follows. First, recall that the fast subsystem (76) converges to $\mathcal{M}_0$ as $t \to \infty$, and that the exact point on $\mathcal{M}_0$ that the fast subsystem converges to depends on the given initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$ which determines $p_{\mathbf{v}} = p_{\mathrm{c}} \in [0, 1]$ that parametrizes $\mathcal{M}_0$ (Section 6.1). We will assume that this point is an initial condition for the slow subsystem, and to this end we define $p_{\mathbf{v}}(\tau_0) = p_{\mathrm{c}}$

for some initial slow time τ_0, and

$$(\mathbf{x}(\tau_0), \mathbf{n}(\tau_0), \mathbf{v}^{\mathbf{n}}(\tau_0)) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}) \tag{79}$$

and

$$(\mathbf{p}(\tau_0), \mathbf{p}_{\mathrm{MM}}(\tau_0)) = (\mathbf{1}p_{\mathbf{v}}(\tau_0), \hat{\mathbf{r}}p_{\mathbf{v}}(\tau_0) + (1-\hat{\mathbf{r}})p_{\mathbf{v}}(\tau_0)^2). \tag{80}$$

Note that the dependency of (79)-(80) on $(\mathbf{x}(t_0), \mathbf{p}(t_0)))$ comes only through (80). Moreover, because every average mutant frequency $p_{\boldsymbol{\phi}}$ in (55) in the fast subsystem converges to p_{c} as $t \to \infty$ (Section 5.1), we will also define

$$p_{\boldsymbol{\phi}}(\tau_0) = p_{\mathbf{v}}(\tau_0). \tag{81}$$

Then, as $\tau \to \infty$, the weighted mutant frequency $p_{\mathbf{v}}(\tau)$ changes according to (78) for all $\tau \geq \tau_0$, and for the initial condition (79)-(80) its solution is

$$p_{\mathbf{v}}(\tau \mid \mathbf{x}(t_0), \mathbf{p}(t_0)) = \frac{1}{1 + \frac{1-p_{\mathbf{v}}(\tau_0)}{p_{\mathbf{v}}(\tau_0)} \exp[-\tau S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]}, \tag{82}$$

where $S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$ is given in (78b) and where we have indicated that the initial condition $p_{\mathbf{v}}(\tau_0) = p_{\mathrm{c}}$ depends on $(\mathbf{x}(t_0), \mathbf{p}(t_0))$. Moreover, while $(\mathbf{x}(\tau), \mathbf{n}(\tau), \mathbf{v}^{\mathbf{n}}(\tau)) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}})$ stay constant for all $\tau \geq \tau_0$, the genetic structure changes according to $\mathbf{p}(\tau) = \mathbf{1}p_{\mathbf{v}}(\tau)$ and $\mathbf{p}_{\mathrm{MM}}(\tau) = \hat{\mathbf{r}}p_{\mathbf{v}}(\tau) + (1-\hat{\mathbf{r}})p_{\mathbf{v}}(\tau)^2$ for all $\tau \geq \tau_0$ and where $p_{\mathbf{v}}(\tau) = p_{\mathbf{v}}(\tau \mid \mathbf{x}(t_0), \mathbf{p}(t_0))$ is the solution in (82). Because $\mathbf{p}$ and $\mathbf{p}_{\mathrm{MM}}$ change also in the fast subsystem (Section 6.1), they are both fast *and* slow variables. Finally, because $p_{\boldsymbol{\phi}}(\tau) = p_{\mathbf{v}}(\tau)$ for all $\tau \geq \tau_0$, the solution of $p_{\boldsymbol{\phi}}$ and thus also the solution of the mean mutant frequency p in (45), is identical to the solution (82). See Figure 2, panels (a)-(c) for a graphical representation.

7 Invasion implies substitution

We are now ready to state and prove the "invasion implies substitution"-principle for the metacommunity model presented in this paper.

Invasion implies substitution - principle. *Suppose that resident individuals express phenotype $\boldsymbol{z} \in \mathcal{Z}$ and that the resident dynamics of the metacommunity is given as in* (6)*. Suppose further that the metacommunity is near a hyperbolically stable steady state $\hat{\mathbf{x}}$ that satisfies $\mathbf{0} = \mathbf{Q}(\hat{\mathbf{x}})\hat{\mathbf{x}}$, and that a mutation generates a mutant phenotype $\boldsymbol{z} + \delta\boldsymbol{\eta}$ where $\boldsymbol{\eta}$ gives the direction of the deviation and δ characterizes the distance between mutant and resident phenotypes. Then, for sufficiently small δ and/or large time $t \geq t_0$ where t_0 is some initial time, the dynamics of the mean mutant frequency p can be approximated as*

$$\dot{p} = p(1-p)\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) + O(\delta^2) \tag{83}$$

with a solution

$$p(t) = \frac{1}{1 + \frac{1-p_c}{p_c} \exp[-t\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]} + O(\delta), \tag{84}$$

where

$$S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) = \sum_{au,bs \in \mathcal{C}} \hat{v}^n_{au} \Big[\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{foc,bs}} \eta_{bs} + \sum_{cs \in \mathcal{C}} \frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{loc,cs}} \hat{r}_{bcs}\ \eta_{cs} \Big] \hat{n}_{bs} \tag{85}$$

is the directional selection coefficient. The constant p_c in (84) *is the asymptotic mean mutant frequency of the fast subsystem where $\delta = 0$ and it depends on the initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$ where $\mathbf{p}(t_0)$ is the initial distribution of within-class mutant frequencies. In $S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$, the density distribution vector $\hat{\mathbf{n}}$ and the reproductive value vector $\hat{\mathbf{v}}^{\mathbf{n}}$ are the right and left dominant eigenvectors of $\mathbf{H}(\hat{\mathbf{x}})$ with a scaling $\hat{\mathbf{v}}^{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$ where $\mathbf{H}(\hat{\mathbf{x}})$ is the resident fitness matrix in* (10)*. The relatedness vector $\hat{\mathbf{r}}$ satisfies $\hat{\mathbf{r}} = \mathbf{c}(\hat{\mathbf{x}}) + \mathbf{D}(\hat{\mathbf{x}})\hat{\mathbf{r}}$ where $\mathbf{c}(\hat{\mathbf{x}})$ is a vector giving the probabilities that a pair of individuals have a common parent and $\mathbf{D}(\hat{\mathbf{x}})$ is a matrix giving the probabilities that this pair has two distinct parents within the same group (Section 3.4.1).*

Proof. First, recall that in Section 4 we set up a mutant-resident metacommunity model for arbitrary δ, and in Section 5 we showed that in order to study the spread of a mutant allele for small δ we must study the fast-slow dynamics of the fast variables $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}})$ and a single slow variable $p_{\mathbf{v}}$. To do this, in Section 6 we set up a fast-slow mutant-resident system for $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}}, p_{\mathbf{v}})$ and then analyzed its fast and slow subsystems where $\delta = 0$: in the fast subsystem (Section 6.1), we found that whenever $\mathbf{x}$ converges to $\hat{\mathbf{x}}$, all the remaining fast variables $(\mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}})$ of interest converge to the critical manifold $\mathcal{M}_0$ given in (77) at which, in particular, $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}})$ are isolated steady state points and $(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}) = (\mathbf{1}p_{\mathbf{v}}, \hat{\mathbf{r}}p_{\mathbf{v}} + (1 - \hat{\mathbf{r}})p_{\mathbf{v}}^2)$ take values on a curve parametrized by $p_{\mathbf{v}} \in [0, 1]$ that is a constant in the fast subsystem (and hence is on $\mathcal{M}_0$ by definition), and where $\hat{\mathbf{r}}$ is the unique steady state value for relatedness (Section 3.5). The exact point $p_{\mathbf{v}} = p_{\mathrm{c}} \in [0, 1]$ on $\mathcal{M}_0$ that the fast subsystem converges to depends on the initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$ (where $\mathbf{x}(t_0)$ is sufficiently close to $\hat{\mathbf{x}}$). In the slow subsystem (Section 6.2), where all the variables are on the critical manifold $\mathcal{M}_0$, we found that the slow variable $p_{\mathbf{v}}$ changes according to an ODE given in (78) and when coupled with an initial condition its solution can be given as in (82). Consequently, while the fast variables $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}})$ stay constant in the slow subsystem, the genetic structure $(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}) = (\mathbf{1}p_{\mathbf{v}}, \hat{\mathbf{r}}p_{\mathbf{v}} + (\mathbf{1} - \hat{\mathbf{r}})p_{\mathbf{v}}^2)$ changes due to $p_{\mathbf{v}}$ and are hence fast and slow variables. Moreover, because on $\mathcal{M}_0$ every average mutant frequency $p_{\boldsymbol{\phi}}$ (55) is equal to $p_{\mathbf{v}}$, its dynamics is also determined by an ODE in (78) and when coupled with an initial condition its solution can also be given as in (82). See a graphical representation in Figure 2, panels (a)-(c). Next, we perturb the fast and slow subsystems using classical results from geometric singular perturbation theory (Fenichel, 1979; Hek, 2010; Kuehn, 2015; Priklopil and Lehmann, 2020, see also Nipp and Stoffer, 2013), which proves the "invasion implies substitution"-principle for small but nonzero δ.

To perturb the fast and slow subsystems we proceed in two steps. In **step 1** we apply the 'Fenichel's first theorem' (e.g. Priklopil and Lehmann, 2020, Appendix A.4., Fenichel's invariant manifold theorem 1) saying that because $\mathcal{M}_0$ is a normally hyperbolic invariant manifold, for small but non-zero δ there exists a so-called slow manifold $\mathcal{M}_\delta$ that is **(i)** $O(\delta)$-close and diffeomorphic to $\mathcal{M}_0$, and, **(ii)** invariant under the fast-slow mutant resident dynamics (74)-(75). Property **(i)** says that for any point on $\mathcal{M}_0$ there is a point on $\mathcal{M}_\delta$ that is $O(\delta)$-close. We will state this in terms of an initial condition of the slow subsystem as

$$(\mathbf{x}(\tau_0), \mathbf{n}(\tau_0), \mathbf{v}^{\mathbf{n}}(\tau_0))|_{\mathcal{M}_\delta} = (\mathbf{x}(\tau_0), \mathbf{n}(\tau_0), \mathbf{v}^{\mathbf{n}}(\tau_0))|_{\mathcal{M}_0} + \mathbf{O}(\delta) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}) + \mathbf{O}(\delta) \tag{86a}$$

and

$$(\mathbf{p}(\tau_0), \mathbf{p}_{\mathrm{MM}}(\tau_0))|_{\mathcal{M}_\delta} = (\mathbf{1}p_{\mathbf{v}}(\tau_0), \hat{\mathbf{r}}p_{\mathbf{v}}(\tau_0) + (\mathbf{1} - \hat{\mathbf{r}})p_{\mathbf{v}}(\tau_0)^2)|_{\mathcal{M}_0} + \mathbf{O}(\delta), \tag{86b}$$

and which is valid for any initial condition (i.e., any point on $\mathcal{M}_0$) and where $\mathbf{O}(\delta)$ is a vector of $O(\delta)$. The property **(ii)** says that for all $\tau \geq \tau_0$, the solutions starting on $\mathcal{M}_\delta$ will remain on $\mathcal{M}_\delta$. Importantly, the two properties **(i)** and **(ii)** together imply that for all $\tau \geq \tau_0$ the ODE's (or vector fields) and their solutions on $\mathcal{M}_\delta$ are small perturbation of the ODE's (vector fields) and solutions on $\mathcal{M}_0$. More precisely, because the dynamics of $p_{\mathbf{v}}(\tau)$ on $\mathcal{M}_0$ (i.e., $p_{\mathbf{v}}(\tau)|_{\mathcal{M}_0}$) can be written as an

ODE (78) and when coupled with an initial condition its solution is given as in (82) for all $\tau \geq \tau_0$ (Section 6.2), the dynamics of $p_{\mathbf{v}}$ on $\mathcal{M}_\delta$ can be written as an ODE

$$\frac{\mathrm{d}}{\mathrm{d}\tau} p_{\mathbf{v}}\Big|_{\mathcal{M}_\delta} = p_{\mathbf{v}}(1-p_{\mathbf{v}}) S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) + O(\delta), \tag{87}$$

and when coupled with an initial condition given in (86), its solution is

$$p_{\mathbf{v}}(\tau)\Big|_{\mathcal{M}_\delta} = \frac{1}{1+\frac{1-p_{\mathrm{c}}}{p_{\mathrm{c}}}\exp[-\tau S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]} + O(\delta) \tag{88}$$

where $p_{\mathrm{c}} = p_{\mathbf{v}}(\tau_0)$ is the initial weighted mutant frequency in the slow subsystem (and which is equal to every other average mutant frequency $p_{\boldsymbol{\phi}}$). The equations (87)-(88) are the first order approximations of the dynamics on $\mathcal{M}_\delta$. Because on $\mathcal{M}_0$ we have $p_{\mathbf{v}}(\tau) = p_{\boldsymbol{\phi}}(\tau)$ for all $\tau \geq \tau_0$ (Section 6.2), and because δ is nonzero in (87)-(88), we can re-write (87)-(88) in fast time $t = \tau/\delta$ and in terms of $p_{\boldsymbol{\phi}}$ as

$$\frac{\mathrm{d}}{\mathrm{d}t} p_{\boldsymbol{\phi}}\Big|_{\mathcal{M}_\delta} = \delta p_{\boldsymbol{\phi}}(1-p_{\boldsymbol{\phi}}) S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) + O(\delta^2) \tag{89}$$

and

$$p_{\boldsymbol{\phi}}(t)\Big|_{\mathcal{M}_\delta} = \frac{1}{1+\frac{1-p_{\mathrm{c}}}{p_{\mathrm{c}}}\exp[-t\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]} + O(\delta). \tag{90}$$

Equations (87)-(90) thus establish the first order approximation for the average mutant frequency in the fast-slow mutant resident system (74)-(75) restricted to the slow manifold $\mathcal{M}_\delta$. We note that while the solutions $p_{\mathbf{v}}$ and $p_{\boldsymbol{\phi}}$ on $\mathcal{M}_\delta$ are both $O(\delta)$-distance away from their solutions on $\mathcal{M}_0$, they are also $O(\delta)$-distance away from each other (Figure 2 panel (f), see further discussion below).

In order to investigate whether the the first order approximations (87)-(90) on $\mathcal{M}_\delta$ are generic in the sense that any other solution nearby is a small perturbation away (a necessary requirement for us to be able to choose initial conditions away from $\mathcal{M}_\delta$), in **step 2** we apply Fenichel's results on the persistence of the stability properties of manifolds (e.g. Priklopil and Lehmann, 2020, Appendix A.4., Fenichel's invariant manifold theorems 2 and 3). These results say that because $\mathcal{M}_0$ is attracting so is $\mathcal{M}_\delta$, and moreover, all the variables near $\mathcal{M}_\delta$ converge to $\mathcal{M}_\delta$ exponentially fast (the exact rates can be estimated as given in Priklopil and Lehmann, 2020, Appendix A.4., Corollaries 1 and 2). Therefore, a solution $p_{\boldsymbol{\phi}}(t)$ with an initial condition $p_{\boldsymbol{\phi}}(t_0)$ that is sufficiently close to p_{c} for some initial time t_0, will become arbitrarily close to (89)-(90) in finite time that depends on δ. See Figure 2 panels (d)-(f) for a graphical representation of the proof. □

Using the weighted mutant frequency $p_{\mathbf{v}}$ and geometric singular perturbation theory (Fenichel, 1979; Hek, 2010; Kuehn, 2015), we showed that for our metacommunity model the "invasion implies substitution"-principle holds, saying that the dynamics of any mutant frequency $p_{\boldsymbol{\phi}}$ can in fact be approximated for sufficiently small δ and/or large fast time t by a *single* scalar valued dynamical equation (83)-(85). Moreover, the rate of substitution is characterized by the directional selection coefficient $\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$ that can be expressed purely in terms of quantities obtained from the resident metacommunity dynamics evaluated at their steady states, and are thus constant over the relevant timescales. We note that because the rate of substitution (change in mutant frequency) is measured per unit of time, this rate is of order $O(\delta)$ slower in fast time t than in slow time τ because $1/t = \delta/\tau$ (compare e.g. (87) and (89)). In contrast, the solution of the average mutant frequency $p_{\boldsymbol{\phi}}$ of the fast-slow system (where δ is small but nonzero) is $O(\delta)$-distance away from the solution of the slow subsystem (where $\delta = 0$), and this distance is independent of the timescale (i.e., the distance between points is independent in which units of time we choose to count the time in, compare e.g. (88) and (90)). Moreover, as δ approaches 0 (but remains non-zero), the slow subsystem (82) (the 'sigmoid

term' in (84)) gives an increasingly good approximation to (84).

7.1 Directional selection coefficients and singular points

Alternative representations of the directional selection coefficient (85) are obtained by considering the class-frequency vector $\mathbf{u} = (u_{as})_{as\in\mathcal{C}} = \mathbf{n}/n$ where $n = \sum_{as\in\mathcal{C}} n_{as}$ is the total density of the focal species. Because each individual can be identified with the allele it carries, we will interpret n as the total number of (mutant and resident) alleles in the focal species and $\mathbf{u}$ as the distribution of the *context* they find themselves in (Kirkpatrick et al., 2002, the distribution of mutants and residents is given by $\mathbf{p}$). In other words, given we randomly sample a carrier of the mutant allele (or resident allele) at the steady state for a process where $\delta = 0$, $\hat{u}_{as}$ is the probability that this allele finds itself in class $as \in \mathcal{C}$. Then, by re-scaling the critical manifold (77) as $\hat{\mathbf{n}} = \hat{n}\hat{\mathbf{u}}$ and $\hat{n}\hat{\mathbf{v}}^{\mathbf{n}} = \hat{\mathbf{v}}^{\mathbf{u}}$ such that $\hat{\mathbf{v}}^{\mathbf{u}} \cdot \hat{\mathbf{u}} = 1$ where $\mathbf{v}^{\mathbf{u}}$ is the adjoint variable associated to $\mathbf{u}$ (and corresponds to the standard scaling of reproductive value, e.g., Taylor, 1990), the directional selection coefficient (85) can be written and interpreted as given in (2) (and where we write $\boldsymbol{z}$ as its argument because it is commonly used to study phenotype evolution and we shall follow this notational change henceforth). We emphasise that under both scalings, (2) and (85), the directional selection coefficient measures the 'additional' mutant alleles produced by a randomly sampled carrier of the mutant allele, i.e., a focal mutant individual. This is true in particular in the second summation in (2) and (85) where the contribution of group members is accounted for.

We can change the perspective on the contribution of group members by swapping indices in the second summation in (2), and using $u_{cs} = (N_{cs}/N_{bs})u_{bs}$ for all $bs, cs \in \mathcal{C}$, we then obtain the representation

$$S_{\boldsymbol{\eta}}(\boldsymbol{z}) = \widetilde{\mathbf{S}}(\boldsymbol{z}) \cdot \boldsymbol{\eta} = \sum_{bs\in\mathcal{C}} S_{bs}(\boldsymbol{z})\hat{u}_{bs}\eta_{bs}, \tag{91}$$

with selection gradient $\widetilde{\mathbf{S}}(\boldsymbol{z}) = (S_{bs}(\boldsymbol{z})\hat{u}_{bs})_{bs\in\mathcal{C}}$ where

$$S_{bs}(\boldsymbol{z}) = \sum_{au\in\mathcal{C}} \hat{v}^{\mathrm{u}}_{au} \left[\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{foc},bs}} + \sum_{cs\in\mathcal{C}} \frac{\partial h_{au,cs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{loc},bs}} \frac{N_{cs}}{N_{bs}} \hat{r}_{bcs} \right] \tag{92}$$

is the selection gradient on phenotypic component $z_{bs} \in \mathcal{Z}_{\mathcal{C}}$, which can be interpreted as follows. Similarly to (2) and (85), the first term in the brackets $\partial h_{au,bs}(\hat{\mathbf{x}})/\partial z_{\mathrm{foc},bs}$ gives the 'additional' mutant offspring alleles $au \in \mathcal{C}$ produced per unit of time by a carrier of the mutant allele $bs \in \mathcal{C}$ as a consequence of it expressing the mutant phenotype. In contrast to (2) and (85), however, the second term $[(\partial h_{au,bs}(\hat{\mathbf{x}})/\partial z_{\mathrm{loc},cs})(N_{cs}/N_{bs})]$ in the second summation gives the 'additional' mutant offspring alleles $au \in \mathcal{C}$ produced per unit of time by all *non-focal* local mutants $cs \in \mathcal{C}$ (group members of the focal individual) upon interacting with a single mutant $bs \in \mathcal{C}$ (the focal allele), and where the weight $\hat{r}_{bcs}$ ensures that we only count those non-focal parent alleles $cs \in \mathcal{C}$ who have a common ancestor with the focal allele $bs \in \mathcal{C}$ and are thus mutants. Each additional mutant offspring allele $au \in \mathcal{C}$ is then weighted with its reproductive value $\hat{v}^{u}_{au}$ that gives the expected asymptotic *number* of its descendants, and we finally sum over all classes of the offspring $au \in \mathcal{C}$. The directional selection coefficient (91) thus gives the asymptotic number of additional mutant alleles produced either directly (the first term in (92)) or indirectly (via related individuals, the second term in (92)) per unit of time by the focal mutant allele that is distributed according to $\hat{\mathbf{u}}$. That is, in (91)-(92), the individual fitness 'belongs' to the focal individual in the first term in (92) (as in (2) and (85)) whereas in the second term in (92) it belongs to non-focal but related individuals (in contrast to (2) and (85)) that are affected by the focal's trait(s). Hence, (91) separates mutational step size from structure of selection by associating all fitness consequences of an allele that would induce an

unilateral phenotypic change η_{bs} into $S_{bs}(\boldsymbol{z}^*)$ for all $bs \in \mathcal{C}$ in (92).

It then follows from (91) that a singular point $\boldsymbol{z}^* \in \mathcal{Z}$ needs to satisfy

$$S_{bs}(\boldsymbol{z}^*) = 0 \quad \text{for all } bs \in \mathcal{C}, \tag{93}$$

since at a singular point no unilateral deviation in phenotype should result in a change in fitness (i.e., (91) needs to be 0 for all $\boldsymbol{\eta}$). Whether such a singular point will be approached by gradual evolution from within its neighbourhood hinges on the concept of convergence stability (Eshel, 1983; Lessard, 1990; Christiansen, 1991; Leimar, 2009). However, because convergence stability along with other properties of singular points (e.g. evolutionary branching Geritz et al., 1998; Ajar, 2003) is a 'second order' property we leave such further analysis for future work. Here we simply state that while in the context of multidimensional phenotypic evolution the direction of change of phenotype can depend on the mutation process, the covariance of the class-specific mutational increments are unlikely to interfere with the convergence stability of singular points (Leimar, 2005, 2009).

8 Discussion

We have proved the "invasion implies substitution"-principle (1)-(2) for the evolution of a multidimensional quantitative trait expressed by the individuals of a focal species that is class structured (by age, size, stage, etc..), and that resides in a metacommunity with finite local community size were reproduction, survival, and growth are stochastic. We proved that the principle holds under the following two assumptions, i) the expression of the mutant and its ancestral resident phenotype are similar, and ii) the pure ecological dynamics result in a metacommunity steady state that is hyperbolically stable. Under these assumptions we found that the directional selection coefficient (2) is frequency-independent and hence mutant frequency (1) either monotonically increases or decreases in time. Therefore, whenever $\delta S_{\boldsymbol{\eta}}$ is positive, a mutant (multidimensional) phenotype that is successful in invading into to the metacommunity will also out-compete its ancestral resident phenotype and go to fixation.

8.1 Evolutionary dynamics across multiple timescales

The evolutionary dynamics discussed in this paper is driven by three qualitatively different evolutionary forces (processes): mutation, selection and class-transmission ('neutral population dynamics'). Mutation-driven changes in the metacommunity are caused by the apparently random modification of the expression of the phenotype. And while selection is (classically) driven by the differential survival and reproduction due to differences in phenotypes among individuals, class-transmission is driven by the differential survival and reproduction due to individuals residing in different classes. Mutant frequency can thus change not only due to selection but *also* transmission, even in the absence of selection (see (42)). In the following we will examine all these processes in detail and how they contribute to the adaptive dynamics of a quantitative trait.

In the mutant-resident metacommunity model with arbitrary phenotypic effects (Section 4), both selection and class-transmission are arbitrarily strong and intertwined. They together drive the change in mutant allele frequency at all times, a change that can involve complex density-dependent and frequency-dependent dynamics potentially leading to allelic coexistence (polymorphism). In Section 5, however, we showed (as expected) that under the assumption (i) of phenotypic closeness, the effect of selection on the mutant frequency dynamics is much weaker than the effect of class-transmission. Hence, the large system of mutant frequency dynamics can be partitioned into two simpler limiting subsystems (Section 6), one that is driven only by class-transmission on a fast population dynamical timescale (fast subsystem, Section 6.1) and the other driven only by selection

on a slower micro-evolutionary timescale (slow subsystem, Section 6.2). One can further analyze these subsystems in isolation and join them together using geometric singular perturbation methods (Fenichel, 1979; Hek, 2010; Kuehn, 2015) eventually leading to the "invasion implies substitution"-principle (1)-(2) (Section 7). In order to follow these steps, however, we had to overcome the difficulty that in class structured populations there does not exist *a priori* a purely slow variable (driven by selection only) that can be used as a proxy for mutant frequency (e.g. Rousset, 2004; Priklopil and Lehmann, 2020). Such a variable can nevertheless be constructed by weighting the density (or frequency) distribution of mutants by reproductive values (Taylor, 1990; Rousset, 2004; Grafen, 2015; Lion, 2018b; Priklopil and Lehmann, 2020), because by definition such weights scale away any changes due to class-transmission. Hence, the assumption of small phenotypic differences and the use of the weighted frequency *together* lead to the separation of timescales and the consequent separation of evolutionary dynamics into frequency changes caused by class-transmission and selection, which in turn underlies mean allele frequency change (1)-(2) whose sign is independent of frequency.

We must stress that the mean allele frequency change (1)-(2) holds not only in the limiting slow subsystem where δ tends to 0 (and where the population dynamical time $t \to \infty$, Section 6.2), but also for small but non-zero δ in the 'normal' population dynamical time $t^* \leq t \leq \infty$ (where t^* is some finite time after which (1)-(2) hold, Section 7). Our result thus extends earlier work on the mutant allele dynamics in slow subsystems (e.g. Wakeley, 2003; Roze and Rousset, 2003; Rousset, 2004; Wakeley and Takahashi, 2004; Lessard, 2009, and see for well-mixed but age and/or class structured populations Lessard and Soares, 2018; Soares and Lessard, 2019, 2020), where the dynamics of the average mutant frequency is obtained in the limit $t \to \infty$ and where $\delta \to 0$ while the total population size in the focal species $K \to \infty$ (but keeping $K\delta$ finite) and by invoking a two timescale method of Ethier and Nagylaki (1980). The result (1)-(2) for small but nonzero δ holds whenever one can show that the metacommunity is 'structurally stable' in the sense that the steady state of the metacommunity and its stability properties persist under the small perturbation caused by the invasion of the mutant (where δ deviates from 0). In the present metacommunity model, this is guaranteed by the minor technical assumption of hyperbolicity (ii), and jointly with phenotypic similarity (assumption i)) and the weighted frequency $p_{\mathbf{v}}$ fully justify the "invasion implies substitution"-principle (1)-(2). Recurrent invasion-substitution events thus cause the seemingly gradual changes in the structure of the metacommunity and the gradual evolution of the phenotype of interest.

The final question is whether and how is the process of gradual evolution affected by mutation-driven processes. That is, do the apparently random changes in the phenotypic expression and the rate of mutations contribute to the qualitative nature of the evolutionary process? For scalar-valued phenotypes (a special case of the present class-specific multidimensional phenotypes), recurrent invasion-substitution events will always either increase or decrease the phenotypic value until the directional selection coefficient vanishes (or due to some external constraints on the phenotype), and the phenotype becomes a singular point (Geritz et al., 1998; Rousset, 2004). Interestingly, the rate of mutations does not interfere with the qualitative nature of this gradual process. This is because (any number of) overlapping mutations perturb the structure of the resident species only to order $O(\delta)$ (see for a formal argument in Meszéna et al., 2005). Therefore, if the mutation of interest results in a positive selection coefficient δS in a resident metacommunity, the selection coefficient remains positive upon the arrival of any new mutations, provided the assumption (i) for small phenotypic effects holds. Naturally, this is also true for the other ('overlapping') mutations: *any* mutation that results in a positive selection coefficient δS will go to fixation upon a successful invasion event. The mutation rate thus only controls the rate of evolution but not its direction.

The above discussion on mutation rates can be directly generalized to multidimensional class-specific phenotypes considered in this paper, but now the 'direction' of evolution is affected by the mutation process in the multi-dimensional phenotype space $\mathcal{Z}$ (Section 7.1 and see e.g. Leimar, 2009 for

examples). Nevertheless, the process of mutation-invasion-substitution continues until the phenotype converges to a singular point, after which one is required to study its 'second-order' properties that determine whether selection becomes stabilizing, in which case the singular point is an ESS, or disruptive, in which case the species becomes polymorphic consisting of two coexisting resident alleles with large phenotypic effects (Geritz et al., 1998; Ajar, 2003; Rousset, 2004). Because by definition the "invasion implies substitution"-principle (1)-(2) is not valid at singular points (Section 7.1), we have left this analysis for further work. However, it is important to emphasise that the "invasion implies substitution"-principle (1)-(2) and the process of gradual evolution is valid also for such polymorphic resident metacommunities (Priklopil and Lehmann, 2020). In such a case the "invasion implies substitution"-principle holds along each "morph" of the polymorphic resident species requiring as many directional selection coefficients as there are morphs. Therefore, provided we stay away from singular points, the "invasion implies substitution"-principle (1)-(2) gives the complete description of the gradual evolutionary process.

8.2 Evolutionary dynamics across multiple fluctuations

Because we have formulated the metacommunity dynamics as a stochastic processes that can depend on the state of the group and the metacommunity at large, our model captures complex frequency- and density-dependent environmental feedbacks both at the group- as well as metacommunity-level. Our model also applies to situations where reproduction, survival and physiological transition depends on locally fluctuating external factors, defined as any factors that are not contained in the environmental feedback, and can thus generate the effect of environmental stochasticity on a local (group) scale (and whenever the external factors can be expressed as discrete variables). Extending our model to capture globally fluctuating external factors would require to introduce an additional "environmental" state space $\mathcal{E}$ whose elements would affect all quantities used in this paper but not change their functional form (i.e., all rates in (3) and (31) would also be functions of $\mathcal{E}$; in the context of this present paper but for well-mixed populations see a recent approach in Cai and Geritz, 2020). Yet, we expect that by taking an appropriate average over the path of environmental states including such environmental change is unlikely to change the generic shape of the selection coefficient as has previously been pointed out (Lenormand et al., 2009). Likewise, in metacommunity models where the total size of the focal species is finite ($K < \infty$), our dynamical equations can in principle be used to construct a diffusion approximation with two time scales to analyze the effect of genetic drift on the evolutionary dynamics (e.g., Wakeley, 2003; Roze and Rousset, 2004; Rousset, 2004; Soares and Lessard, 2020), with the infinitesimal mean of the diffusion being given by the mean change (78) and the infinitesimal variance obtainable from the recurrence equations for relatedness (Rousset, 2004, chapter 9-11). Finally, we formulated our model only in the context of haploid reproduction, yet allowing for diploidy and sexual reproduction under additive gene action within individuals should only require changing the scaling of the mean change in allele frequency (83) and to interpret relatedness in the directional selection coefficient (85) in a diploid context (as per simpler models with diploidy and class structure, Roze and Rousset, 2004; Van Cleve et al., 2010). In conclusion, while we have left several biological questions to be settled in future work, we believe that the present model contributes to the understanding of adaptive evolution in a changing environment across multiple timescales of natural populations.

Acknowledgment

TP was supported by the Swiss NSF grant PP00P3-123344 to LL

A Appendix: Individual fitness

A.1 Individual-level processes: properties, consistency relations and demographic consequences

In this section we provide the properties and consistency relations of the i-process rates appearing in (5) as well as discuss their demographic consequences. First, upon a birth event, whether philopatric or via dispersal, a single offspring is added into the group. Therefore, upon a philopatric birth event where (any) parent individual in group $s \in \mathcal{S}$ gives birth to $a \in \mathcal{P}$, due to which the group transitions to $u \in \mathcal{S}$, we must have $N_{au} - N_{as} = 1$ (and otherwise 0). Similarly, upon a dispersal event where (any) parent individual in group $s \in \mathcal{S}$ gives birth to a state $a \in \mathcal{P}$ offspring and who lands into (any) group $w \in \mathcal{S}$ that transitions to $u \in \mathcal{S}$ due to this event, we have $N_{au} - N_{aw} = 1$.

Second, the physiological transition rate in (5b) is assumed to satisfy the consistency relation

$$\lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) = \begin{cases} \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}), & \text{if} \quad a \neq b \\ \sum_{c \neq a \in \mathcal{P}} \lambda^{\mathrm{T}}_{cu,as}(\mathbf{x}), & \text{if} \quad a = b. \end{cases} \tag{A.1}$$

Here, for $a \neq b \in \mathcal{P}$, (A.1) gives the rate at which $b \in \mathcal{P}$ in $s \in \mathcal{S}$ transitions to $a \in \mathcal{P}$ due to which the group transitions to $u \in \mathcal{S}$, and for $a = b \in \mathcal{P}$, (A.1) gives the rate at which an individual $a \in \mathcal{P}$ in $s \in \mathcal{S}$ transitions into any (other) physiological state due to which the group transitions to $u \in \mathcal{S}$. If we focus on the change in the number of 'offspring' individuals $a \in \mathcal{P}$ due to this event, for $a \neq b \in \mathcal{P}$ we have $N_{au} - N_{as} = 1$ because $a \in \mathcal{P}$ is *added* into the group, and for $a = b \in \mathcal{P}$ we have $N_{au} - N_{as} = -1$ because $a \in \mathcal{P}$ is *removed* from the group.

Finally, death rate in (5b) is assumed to satisfy

$$\lambda^{\mathrm{D}}_{au,bs}(\mathbf{x}) = \begin{cases} \lambda^{\mathrm{D}}_{au}(\mathbf{x}), & \text{if} \quad a = b \text{ and } s \backslash \{a\} = u \\ 0, & \text{otherwise}, \end{cases} \tag{A.2}$$

where we use the fact that upon a death of $a \in \mathcal{P}$ due to which the group transitions to $u \in \mathcal{S}$, it is unambiguous in which state the group was before the transition: this rate is non-zero only if by removing $a \in \mathcal{P}$ from $s \in \mathcal{S}$ the group transitions to $u \in \mathcal{S}$. Hence, we write $\lambda^{\mathrm{D}}_{au} = \lambda^{\mathrm{D}}_{au,as}$ (which is non-zero) whenever $s \backslash \{a\} = u$ for all $u, s \in \mathcal{S}$. If we focus on the change in the number of 'offspring' individuals $a \in \mathcal{P}$ due to this event, we have $N_{au} - N_{as} = -1$ because upon a death of $a \in \mathcal{P}$ in $s \in \mathcal{S}$ an individual is *removed* from the group.

A.2 Resident individual fitness

In this section we detail the steps to obtain the resident individual fitness (growth-rate) matrix (10) from the resident group and metacommunity dynamics (6). First, note that whenever a group transitions from state $s \in \mathcal{S}$ to $u \in \mathcal{S}$, the change in the number of individuals in physiological state $a \in \mathcal{P}$ is $N_{au} - N_{as}$, for all $a \in \mathcal{P}$. The value of $N_{au} - N_{as}$ is non-zero (and either $+1$ or -1) whenever an individual in physiological state $a \in \mathcal{P}$ was involved in the event that caused the transition of the group $s \in \mathcal{S}$ to $u \in \mathcal{S}$, and otherwise $N_{au} - N_{as} = 0$ (see further details in Section A.1). Using

$n_{au} = N_{au}x_u$, we get by direct calculation

$$
\begin{aligned}
\dot{n}_{au} &= N_{au}\dot{x}_u \\
&= N_{au}\sum_{s\in\mathcal{S}}\left[q^{\mathrm{p}}_{u,s}(\mathbf{x}) + q^{\mathrm{d}}_{u,s}(\mathbf{x}) + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]x_s \\
&= N_{au}\sum_{s\in\mathcal{S}}\left[\sum_{c,b\in\mathcal{P}}\lambda^{\mathrm{p}}_{cu,bs}(\mathbf{x})N_{bs} + \sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\lambda^{\mathrm{d,B}}_{cu,s,dw}(\mathbf{x})N_{dw}x_w + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]x_s \\
&= \sum_{s\in\mathcal{S}}(N_{as} + N_{au} - N_{as})\left[\sum_{c,b\in\mathcal{P}}\lambda^{\mathrm{p}}_{cu,bs}(\mathbf{x})N_{bs} + \sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\lambda^{\mathrm{d,B}}_{cu,s,dw}(\mathbf{x})N_{dw}x_w + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]x_s \\
&= \sum_{s\in\mathcal{S}}(N_{au} - N_{as})\left[\sum_{b\in\mathcal{P}}\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x})N_{bs} + \sum_{dw\in\mathcal{C}}\lambda^{\mathrm{d,B}}_{au,s,dw}(\mathbf{x})N_{dw}x_w\right]x_s + \\
&\quad + \sum_{s\in\mathcal{S}}\left[\sum_{c,b\in\mathcal{P}}\lambda^{\mathrm{p}}_{cu,bs}(\mathbf{x})N_{bs} + \sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\lambda^{\mathrm{d,B}}_{cu,s,dw}(\mathbf{x})N_{dw}x_w + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]N_{as}x_s \\
&= \sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{P}}\left[(N_{au} - N_{as})\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) + \sum_{w\in\mathcal{S}}(N_{au} - N_{aw})x_w\lambda^{\mathrm{d,B}}_{au,w,bs}(\mathbf{x})\right]n_{bs} + \sum_{s\in\mathcal{S}}q_{u,s}(\mathbf{x})n_{as} \\
&= \sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{P}}\left[(N_{au} - N_{as})\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x})\right]n_{bs} + \sum_{s\in\mathcal{S}}q_{u,s}(\mathbf{x})n_{as} \\
&= \sum_{bs\in\mathcal{C}}h_{au,bs}(\mathbf{x})n_{bs},
\end{aligned}
\tag{A.3}
$$

where we applied (6)-(8); in the fifth equality we used the fact that either the offspring $a \in \mathcal{P}$ in $u \in \mathcal{S}$ is produced anew or is removed by death in which case $N_{au} - N_{as}$ is non-zero (the first line after the equality, and the second line after the equality where in the sum $c = a \in \mathcal{P}$), or, that the offspring $a \in \mathcal{P}$ is a surviving individual in which case $N_{au} = N_{as}$ (the second line after the equality where in the sum $c \neq a \in \mathcal{P}$), in the sixth equality we exchanged indices and in the seventh equality we used the fact that $N_{au} - N_{aw} = 1$ for all $w \in \mathcal{S}$ because upon any dispersal event an individual is added into the group (Section A.1). We have thus obtained

$$
h_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) + (N_{au} - N_{as})\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) + \delta_{a,b}q_{u,s}(\mathbf{x}), \tag{A.4}
$$

where $(N_{au} - N_{bs}) \in \{-1, 0, 1\}$ and which is the partitioning (11) as presented in Section 3.

A.2.1 Individual fitness: connection to previous formulations

The concept of individual fitness (A.4) ((11) of the main text) has been considered in the previous literature for discrete time processes (Lehmann et al., 2016; Ohtsuki et al., 2020) and could generically be called *unconditional* individual fitness (as it is evaluated unconditionally on whether or not a group transition occurs). We can distinguish its dispersal and philopatric components, as discussed in Section 3.3, by writing

$$
h_{au,bs}(\mathbf{x}) = h^{\mathrm{d}}_{au,bs}(\mathbf{x}) + h^{\mathrm{p}}_{au,bs}(\mathbf{x}), \tag{A.5a}
$$

where $h^{\mathrm{d}}_{au,bs}(\mathbf{x}) = \sum_{w\in\mathcal{S}} x_w h^{\mathrm{d}}_{au,w,bs}(\mathbf{x})$ and where

$$
\begin{aligned}
h^{\mathrm{d}}_{au,w,bs}(\mathbf{x}) &= \lambda^{\mathrm{d,B}}_{au,w,bs}(\mathbf{x}) \\
h^{\mathrm{p}}_{au,bs}(\mathbf{x}) &= (N_{au} - N_{as})\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) + \delta_{a,b}q_{u,s}(\mathbf{x}).
\end{aligned}
\tag{A.5b}
$$

The *conditional* individual fitness (e.g. Rousset, 2004; Rousset and Ronce, 2004, Lehmann et al., 2016, Appendix E) is defined as the fitness of an individual given a group-transition (it is conditional

on the realization of the state after the transition), and this can be written as

$$w_{au,bs}(\mathbf{x}) = w^{\mathrm{d}}_{au,bs}(\mathbf{x}) + w^{\mathrm{p}}_{au,bs}(\mathbf{x}), \tag{A.6a}$$

where $w^{\mathrm{d}}_{au,bs}(\mathbf{x}) = \sum_{w\in\mathcal{S}} x_w w^{\mathrm{d}}_{au,w,bs}(\mathbf{x})$ and where

$$\begin{aligned} w^{\mathrm{d}}_{au,w,bs}(\mathbf{x}) &= \frac{h^{\mathrm{d}}_{au,w,bs}(\mathbf{x})}{q_{u,w}} \\ w^{\mathrm{p}}_{au,bs}(\mathbf{x}) &= \frac{h^{\mathrm{p}}_{au,bs}(\mathbf{x})}{q_{u,s}}. \end{aligned} \tag{A.6b}$$

Note that when $a = b \in \mathcal{P}$, then $w^{\mathrm{p}}_{au,as}(\mathbf{x}) = 1 + \frac{(N_{au}-N_{bs})\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x})}{q_{u,s}}$ and when $a \neq b \in \mathcal{P}$, then $w^{\mathrm{p}}_{au,bs}(\mathbf{x}) = \frac{(N_{au}-N_{bs})\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x})}{q_{u,s}}$.

We finally note that several previous works considered explicit single species demographic scenarios that do not experience physiological structure in which case $\mathcal{S}$ represents the set of various group sizes of the single 'focal' species (giving the total number of individuals in a group, e.g., Metz and Gyllenberg, 2001; Cadet et al., 2003; Rousset and Ronce, 2004; Lehmann et al., 2006; Alizon and Taylor, 2008; Parvinen, 2013) or some local environmental state (e.g., Wild et al., 2009; Rodrigues, 2018). For these cases, individuals undergo only birth (philopatric or via dispersal) and death, and the individual (unconditional) fitness (A.5) reduces to

$$h_{u,s}(\mathbf{x}) = h^{\mathrm{d}}_{u,s}(\mathbf{x}) + h^{\mathrm{p}}_{u,s}(\mathbf{x}), \tag{A.7a}$$

where $h^{\mathrm{d}}_{u,s}(\mathbf{x}) = \sum_{w\in\mathcal{S}} x_w h^{\mathrm{d}}_{u,w,s}(\mathbf{x})$ and where

$$\begin{aligned} h^{\mathrm{d}}_{u,w,s}(\mathbf{x}) &= \lambda^{\mathrm{d}}_{u,w,s}(\mathbf{x}) \\ h^{\mathrm{p}}_{u,s}(\mathbf{x}) &= (N_u - N_s)\left(\lambda^{\mathrm{p,B}}_{u,s}(\mathbf{x}) + \lambda^{\mathrm{D}}_{u,s}(\mathbf{x})\right) + q_{u,s}(\mathbf{x}). \end{aligned} \tag{A.7b}$$

Likewise, the conditional individual fitness (A.6) then reduces to

$$w_{u,s}(\mathbf{x}) = w^{\mathrm{d}}_{u,s}(\mathbf{x}) + w^{\mathrm{p}}_{u,s}(\mathbf{x}), \tag{A.8a}$$

where $w^{\mathrm{d}}_{u,s}(\mathbf{x}) = \sum_{w\in\mathcal{S}} x_w w^{\mathrm{d}}_{u,w,s}(\mathbf{x})$ and where

$$\begin{aligned} w^{\mathrm{d}}_{u,w,s}(\mathbf{x}) &= \frac{h^{\mathrm{d}}_{u,w,s}(\mathbf{x})}{q_{u,w}} \\ w^{\mathrm{p}}_{u,s}(\mathbf{x}) &= 1 + \frac{(N_u - N_s)h^{\mathrm{p}}_{u,s}(\mathbf{x})}{q_{u,s}} \end{aligned} \tag{A.8b}$$

(for concrete examples, see Rousset and Ronce, 2004 for discrete time processes and Lehmann and Rousset, 2010, Supplementary material for continuous time process that squarely fits within the present framework).

A.2.2 Individual fitness: further partitioning

For completeness, we here give an explicit representation of the individual fitness (11) in terms of the different i-process rates (5) (see also Section A.1) as they can be seen as the most basic ingredients of our dynamical metacommunity model (Section 3). Let us denote with

$$q^{\mathrm{S}}_{au,bs}(\mathbf{x}) = \delta_{a,b} q_{u,s}(\mathbf{x}), \qquad \text{for all} \quad au, bs \in \mathcal{C}, \tag{A.9}$$

the survival rate (i.e. the rate of change of the local environment) of an individual $bs \in \mathcal{C}$. Then, using (A.1)-(A.9), the individual fitness matrix $\mathbf{H}(\mathbf{x})$ of the resident species dynamics (10) can be

partitioned as

$$\mathbf{H}(\mathbf{x}) = \mathbf{\Lambda}^{\mathrm{d,B}}(\mathbf{x}) + \mathbf{\Lambda}^{\mathrm{p,B}}(\mathbf{x}) + \mathbf{\Lambda}^{\mathrm{T}}(\mathbf{x}) - \mathbf{\Lambda}^{\mathrm{D}}(\mathbf{x}) + \mathbf{Q}^{\mathrm{S}}(\mathbf{x}), \tag{A.10}$$

where $\mathbf{\Lambda}^{\mathrm{d,B}}(\mathbf{x}) = (\lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}))_{au,bs\in\mathcal{C}}$ is a birth-rate matrix via dispersal, $\mathbf{\Lambda}^{\mathrm{p,B}}(\mathbf{x}) = (\lambda^{\mathrm{B}}_{au,bs}(\mathbf{x}))_{au,bs\in\mathcal{C}}$ is a birth-rate matrix via philopatric reproduction, $\mathbf{\Lambda}^{\mathrm{T}}(\mathbf{x}) = ((N_{au} - N_{as})\lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}))_{au,bs\in\mathcal{C}}$ is a physiological transition-rate matrix, $\mathbf{\Lambda}^{\mathrm{D}}(\mathbf{x}) = (\lambda^{\mathrm{D}}_{au,bs}(\mathbf{x}))_{au,bs\in\mathcal{C}}$ is a death-rate matrix, and $\mathbf{Q}^{\mathrm{S}}(\mathbf{x}) = (q^{\mathrm{S}}_{au,bs}(\mathbf{x}))_{au,bs\in\mathcal{C}}$ is a survival-rate matrix. The elements of (A.10) are given as

$$h_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{p,B}}_{au,bs}(\mathbf{x}) + (N_{au} - N_{as})\lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) - \lambda^{\mathrm{D}}_{au,bs}(\mathbf{x}) + q^{\mathrm{S}}_{au,bs}(\mathbf{x}), \tag{A.11}$$

for all $au, bs \in \mathcal{C}$. By grouping appropriately the various i-processes we obtain the representation presented in Section 3.3.

A.3 Mutant-resident individual fitnesses

A.3.1 Fitness of mutant and resident individuals

Using $n_{\mathrm{M},au} = \sum_{j\in\mathcal{I}_u} j_{au} y_{uj}$ and the fact that upon a transition from $si \in \Omega$ to $uj \in \Omega$ the number of mutants that are added or removed is $j_{au} - i_{as}$ for all $a \in \mathcal{P}$, we get

$$\begin{aligned}
\dot{n}_{\mathrm{M},au} &= \sum_{j\in\mathcal{I}_u} j_{au}\dot{y}_{uj} \\
&= \sum_{j\in\mathcal{I}_u} j_{au} \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} \pi_{uj,si}(\mathbf{y}) y_{si} \\
&= \sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} (i_{as} + j_{au} - i_{as})\pi_{uj,si}(\mathbf{y}) y_{si} \\
&= \sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} (j_{au} - i_{as}) \left[\sum_{b\in\mathcal{P}} \lambda^{\mathrm{M,p}}_{auj,bsi}(\mathbf{y}) i_{bs} + \sum_{k\in\mathcal{I}_w}\sum_{dw\in\mathcal{C}} \lambda^{\mathrm{M,d}}_{auj,s,dwk}(\mathbf{y}) k_{dw} y_{dw}\right] y_{si} \\
&\quad + \sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} \left[\sum_{c,b\in\mathcal{P}} \bar{\lambda}^{\mathrm{p}}_{cu,bsi}(\mathbf{y}) N_{bs} + \sum_{c,d\in\mathcal{P}}\sum_{wk\in\Omega} \bar{\lambda}^{\mathrm{d}}_{cu,s,dwk}(\mathbf{y}) N_{dw} y_{wk}\right] i_{as} y_{si} + \\
&\quad + \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} \pi^{\mathrm{mc}}_{uj,si}(\mathbf{y}) i_{as} y_{si}.
\end{aligned} \tag{A.12}$$

By exchanging the indices and using the notation $y_{\mathrm{M},bsi|\mathrm{M},bs} = \frac{i_{bs} y_{si}}{n_{\mathrm{M},bs}}$ we get

$$\begin{aligned}
\dot{n}_{\mathrm{M},au} &= \sum_{bs\in\mathcal{C}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s} \left[(j_{au} - i_{as})\lambda^{\mathrm{M,p}}_{auj,bsi}(\mathbf{y}) + \sum_{w\in\mathcal{S}} (j_{au} - i_{aw}) x_w \lambda^{\mathrm{M,d}}_{auj,w,bsi}(\mathbf{y})\right] y_{\mathrm{M},bsi|\mathrm{M},bs} n_{\mathrm{M},bs} \\
&\quad + \sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s} \left[\pi^{\mathrm{p}}_{uj,si}(\mathbf{y}) + \pi^{\mathrm{d}}_{uj,si}(\mathbf{y}) + \pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right] y_{\mathrm{M},asi|\mathrm{M},as} n_{\mathrm{M},as} \\
&= \sum_{bs\in\mathcal{C}} g^{\mathrm{M}}_{au,bs}(\mathbf{y}) n_{\mathrm{M},bs},
\end{aligned} \tag{A.13}$$

where

$$g^{\mathrm{M}}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s} \left[\lambda^{\mathrm{M,d}}_{auj,bsi}(\mathbf{y}) + (j_{au} - i_{as})\lambda^{\mathrm{M,p}}_{auj,bsi}(\mathbf{y}) + \delta_{a,b}\pi_{uj,si}(\mathbf{y})\right] y_{\mathrm{M},bsi|\mathrm{M},bs}, \tag{A.14}$$

which is the partitioning given in (37).

The derivation for the resident fitness (38) is identical to (A.12)-(A.14), except that one needs to replace the numbers of mutants within a group i_{as} with residents numbers $N_{as} - i_{as}$, for all $as \in \mathcal{C}$,

class-specific densities for mutants $n_{\mathrm{M},as}$ with resident densities $n_{\mathrm{R},as}$, for all $as \in \mathcal{C}$, as well as conditional group frequencies $y_{\mathrm{M},bsi|\mathrm{M},bs}$ with $y_{\mathrm{R},bsi|\mathrm{R},bs}$, for all $b \in \mathcal{P}, si \in \Omega$.

A.3.2 Fitness of an average individual

We can obtain the average individual fitness (39)-(40) by taking an average over mutant and resident individual fitness's (37) and (38). Here, instead, we will provide an alternative derivation that is similar to the the derivation for resident fitness (11) (and (A.4)): the average individual growth-rate (39)-(40) can be calculated from the mutant-resident group and metacommunity dynamics (33) as

$$\begin{aligned}
\dot{n}_{au} &= N_{au}\dot{x}_u \\
&= N_{au}\sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\left[\pi^{\mathrm{p}}_{uj,si}(\mathbf{y})+\pi^{\mathrm{d}}_{uj,si}(\mathbf{y})+\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right]y_{si} \\
&= N_{au}\sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\left[\sum_{c,b\in\mathcal{P}}\bar{\lambda}^{\mathrm{p}}_{cuj,bsi}(\mathbf{y})N_{bs}+\sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\bar{\lambda}^{\mathrm{d}}_{cuj,s,dwk}(\mathbf{y})N_{dw}x_w+\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right]y_{si} \\
&= \sum_{bs\in\mathcal{C}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[(N_{au}-N_{as})\bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y})N_{bs}+\sum_{w\in\mathcal{S}}(N_{au}-N_{aw})x_w\bar{\lambda}^{\mathrm{d}}_{auj,w,bsi}(\mathbf{y})N_{bs}\right]y_{si}+ \\
&\quad+\sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[\sum_{c,b\in\mathcal{P}}\bar{\lambda}^{\mathrm{p}}_{cuj,bsi}(\mathbf{y})N_{bs}+\sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\bar{\lambda}^{\mathrm{d}}_{cuj,s,dwk}(\mathbf{y})N_{dw}x_w+\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right]N_{as}y_{si} \\
&= \sum_{bs\in\mathcal{C}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[(N_{au}-N_{as})\bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y})+\bar{\lambda}^{\mathrm{d}}_{auj,bsi}(\mathbf{y})\right]y_{si|s}n_{bs}+\sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\pi_{uj,si}(\mathbf{y})y_{si|s}n_{as} \\
&= \sum_{bs\in\mathcal{C}}\bar{g}_{au,bs}(\mathbf{y})n_{bs},
\end{aligned} \tag{A.15}$$

where we used the notation $y_{si|s} = \frac{y_{si}}{x_s}(= \frac{N_{bs}y_{si}}{n_{bs}}$ for all $b \in \mathcal{P})$ and where

$$\bar{g}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[\bar{\lambda}^{\mathrm{d}}_{auj,bsi}(\mathbf{y})+(N_{au}-N_{as})\bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y})+\delta_{a,b}\pi_{uj,si}(\mathbf{y})\right]y_{si|s}, \tag{A.16}$$

which is the partitioning (40) from the main text.

B Appendix: Relatedness and genetic correlations

B.1 Jump-process approach

To facilitate the derivation of the recursion (25) for relatedness (Section 3.4.1), we represent here the metacommunity dynamics as a discrete-time jump-process. To this end, we focus on a single group and take a probabilistic perspective on the group dynamics by viewing the vector $\mathbf{x}(t) := \mathbf{x}$ as the probability distribution of the state of some focal group at time t (Section 3). The state of the focal group is considered as a random variable which we denote with $S(t)$ for all t. In the continuous-time formulation discussed in the main text, the focal group (as well as all other groups in the metacommunity) can thus be seen as undergoing a continuous-time (non-homogeneous) Markov chain $\{S(t) \mid t \in [t_0, t_f]\}$ on the state space $\mathcal{S}$ for some initial time t_0 and final time t_f (and recall that the non-homogeneity is implicit and comes via the group transition matrix (6) being a function of $\mathbf{x}(t)$). Now, because we want to consider a jump chain (or embedded Markov chain, e.g., Iosifescu, 2007) associated to the above continuous-time Markov chain, we are interested in the probabilities at which the i-processes happen. Let $\Pi_k = S(T_k)$ denote the random variable of this jump process giving the state of the group at the kth jump starting at time t_0, where each jump time T_k is (also) a random variable. The resident group dynamics of this discrete-time Markov chain can then be

written as

$$\mathbf{x}(T_{k+1}) = \mathbf{A}\mathbf{x}(T_k), \tag{B.1a}$$

where each element

$$a_{u,s} = \begin{cases} -\frac{q_{u,s}}{q_{s,s}} = \frac{q_{u,s}}{\sum_{w\neq s\in\mathcal{S}} q_{w,s}} & u \neq s \\ 0 & u = s \end{cases} \tag{B.1b}$$

gives the probability at which a group in state $s \in \mathcal{S}$ transitions to state $u \in \mathcal{S}$ (because no state is absorbing). Using the discrete-time group dynamics (B.1) we obtain a jump-process for relatedness (25)-(27) as given in the main text.

B.2 Mutant-pair dynamics

Here we derive the mutant-pair dynamics (69)-(70) when $\delta = 0$ (Section 5.3.3). The ODE for the total (or average) density of local pairs n_{abu} for all $abu \in \mathcal{B}$ is given, with a slight abuse of notation, identically to (16). By defining $n_{\mathrm{MM},abs} = \sum_{i\in\mathcal{I}} i_{as} i_{bs} y_{si}$ as the density of mutant-pairs $abs \in \mathcal{B}$, the ODE for mutant-pair dynamics is given as

$$\begin{aligned} \dot{n}_{\mathrm{MM},abu} &= \sum_{j\in\mathcal{I}_u} j_{au} j_{bu} \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{I}_s} \pi_{uj,si}(\mathbf{y}) y_{si} \\ &= \sum_{s\neq u\in\mathcal{S}} \Big[\sum_{d\in\mathcal{P}} \beta^{\mathrm{CA}}_{abu,ds}(\mathbf{x}) n_{\mathrm{M},ds} + \sum_{d\in\mathcal{P}} \beta^{\mathrm{IM}}_{abu,ds}(\mathbf{x}) n_{\mathrm{M},ds} + \sum_{c,d\in\mathcal{P}} \beta^{\mathrm{DP}}_{abu,cds}(\mathbf{x}) n_{\mathrm{MM},cds} \Big] \\ &\quad - n_{\mathrm{MM},abu} \sum_{w\neq u} q_{w,u}(\mathbf{x}), \quad \text{for all } abu \in \mathcal{B}, \end{aligned} \tag{B.2}$$

where the above rates are given in terms of i-processes as in (18)-(20) and where we used the consistency relation (48).

We obtain the ODE (69) for the mutant-pair frequencies $\mathbf{p}_{\mathrm{MM}}$ defined in (66) by differentiation and using (16) and (B.2), whereby

$$\dot{p}_{\mathrm{MM},abu} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{n_{\mathrm{MM},abu}}{n_{abu}}\right) = \frac{\dot{n}_{\mathrm{MM},abu}}{n_{abu}} - p_{\mathrm{MM},abu}\frac{\dot{n}_{abu}}{n_{abu}} \quad \text{for all } abu \in \mathcal{B}. \tag{B.3}$$

C Appendix: Properties of growth-rates for similar phenotypes

We here provide further details on the fitness properties discussed in Section 5.3.1. Following the main text (Section 5.3), we refer to some representative individual of interest (mutant, resident, average) in the focal species as the focal individual, and then refer to its group members as local individuals and non-group members as non-local (or 'population') individuals. Using the notation introduced in Section 5.3 and the definition for the i-processes (3) and (31), we can label the phenotypes of individuals in the resident metacommunity acting on the fitness of a focal individual with trait $z_{\mathrm{foc},bs}$ according to the following arguments on the right-hand side of

$$\lambda_{au,bs}(\mathbf{x}) = \lambda_{au\leftarrow bs}(z_{\mathrm{foc},bs}, \boldsymbol{z}_{\mathrm{loc}}, s, \boldsymbol{z}_{\mathrm{pop}}, \mathbf{x}),, \quad \text{for all } a, b \in \mathcal{P},\ s, u \in \mathcal{S}. \tag{C.1}$$

For focal mutant and resident individuals in a mutant-resident metacommunity we write the phenotypes of actors as

$$\begin{aligned}\lambda^{\mathrm{M}}_{auj,bsi}(\mathbf{y}) &= \lambda_{auj\leftarrow bsi}(z_{\mathrm{M_{foc}},bs}, \boldsymbol{z}_{\mathrm{M_{loc}}}, \boldsymbol{z}_{\mathrm{R_{loc}}}, si, \boldsymbol{z}_{\mathrm{M_{pop}}}, \boldsymbol{z}_{\mathrm{R_{pop}}}, \mathbf{y}) \\ \lambda^{\mathrm{R}}_{auj,bsi}(\mathbf{y}) &= \lambda_{auj\leftarrow bsi}(z_{\mathrm{R_{foc}},bs}, \boldsymbol{z}_{\mathrm{M_{loc}}}, \boldsymbol{z}_{\mathrm{R_{loc}}}, si, \boldsymbol{z}_{\mathrm{M_{pop}}}, \boldsymbol{z}_{\mathrm{R_{pop}}}, \mathbf{y}), \quad \text{for all } a, b \in \mathcal{P},\ si, uj \in \Omega.\end{aligned} \tag{C.2}$$

Then, the following properties are a consequence of the so-called generalized law of mass action (Section 5.3; Diekmann et al., 2001; Meszéna et al., 2005; Dercole, 2016; Priklopil and Lehmann, 2020) when applied to the present metacommunity model with local and global environments. We have,

$$\left.\frac{\partial \lambda^{\mathrm{M}}_{auj,bsi}(\mathbf{y})}{\partial z_{\mathrm{M_{foc}},bs}}\right|_{\delta=0} = \frac{\partial \lambda_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}}, \qquad \left.\frac{\partial \lambda^{\mathrm{R}}_{auj,bsi}(\mathbf{y})}{\partial z_{\mathrm{M_{foc}},bs}}\right|_{\delta=0} = 0, \tag{C.3}$$

for all $a, b \in \mathcal{P}$ and $si, uj \in \Omega$ and where both partial derivatives are 0 with respect to other choices of $z_{\mathrm{M_{foc}},cw}$ for $cw \neq bs \in \mathcal{C}$ (because the focal is $bs \in \mathcal{C}$), and

$$\left.\frac{\partial \lambda^{\mathrm{M}}_{auj,bsi}(\mathbf{y})}{\partial z_{\mathrm{M_{loc}},cs}}\right|_{\delta=0} = \frac{i_{cs} - \delta_{c,b}}{N_{cs} - \delta_{c,b}} \frac{\partial \lambda_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}}, \quad \left.\frac{\partial \lambda^{\mathrm{R}}_{auj,bsi}(\mathbf{y})}{\partial z_{\mathrm{M_{loc}},cs}}\right|_{\delta=0} = \frac{i_{cs}}{N_{cs} - \delta_{c,b}} \frac{\partial \lambda_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} \tag{C.4}$$

for all $a, b, c \in \mathcal{P}$ and $si, uj \in \Omega$, where both partial derivatives are 0 with respect to other choices of $z_{\mathrm{M_{loc}},cw}$ for $w \neq s \in \mathcal{S}$ (because local individuals inhabit group $s \in \mathcal{S}$) and where $\delta_{c,b}$ is the kronecker delta, and

$$\left.\frac{\partial \lambda^{\mathrm{M}}_{auj,bsi}(\mathbf{y})}{\partial z_{\mathrm{M_{pop}},cw}}\right|_{\delta=0} = p_{\mathrm{M},cw} \frac{\partial \lambda_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{pop},cw}}, \qquad \left.\frac{\partial \lambda^{\mathrm{R}}_{auj,bsi}(\mathbf{y})}{\partial z_{\mathrm{M_{pop}},cw}}\right|_{\delta=0} = p_{\mathrm{M},cw} \frac{\partial \lambda_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{pop},cw}}, \tag{C.5}$$

for all $a, b \in \mathcal{P}$, $si, uj, \in \Omega$ and $cw \in \mathcal{C}$. The partial derivative in (C.3) is taken with respect to the phenotype of the focal mutant individual $bs \in \mathcal{C}$, and must hence be (trivially) 0 for all other focal mutants $cw \neq bs \in \mathcal{C}$ and for any focal resident individuals. The partial derivative in (C.4) is taken with respect to the phenotype of mutant individuals that inhabit the group of the focal individual. Hence, the coefficient in front of the mutant and resident i-processes differs, because in any given group a mutant $bs \in \mathcal{C}$ interacts with one less mutant of class $bs \in \mathcal{C}$ than a resident does. Naturally, these derivatives are zero whenever they are taken with respect to mutant individuals outside the focal's group (these derivatives are given in (C.5)). Finally, because both mutants and residents experience the exact same interactions with members of non-local groups, this justifies the partial derivatives in (C.5).

To obtain the properties (62)-(64), we first apply (C.3)-(C.5). To do this, however, we must be careful in labelling individual fitness functions (11) and (40)-(37): while the first two terms in individual fitness describe the i-processes of the focal individual (and hence the labelling is identical to (C.1)-(C.2)), in the the 'survival' term (that tracks changes in the group) the individuals who undergo the i-processes are mostly local or non-local individuals and thus *not only* the focal individual as written in (C.1)-(C.2)). In such situations we must appropriately assign the labels according to who undergoes the i-process. After taking the derivatives, the second step to obtain properties (62)-(64) is to sum over the genetic states of groups by applying the definition for within-class (Section 4.4) and mutant-pair (66) frequencies. As these calculations are straightforward (but lengthy) we will not present them here, but to mention that all above obtained properties are fully consistent with previous derivations of first order fitness perturbations (e.g., Rousset, 2004, chapter 6 and 11, Rousset and Ronce, 2004).

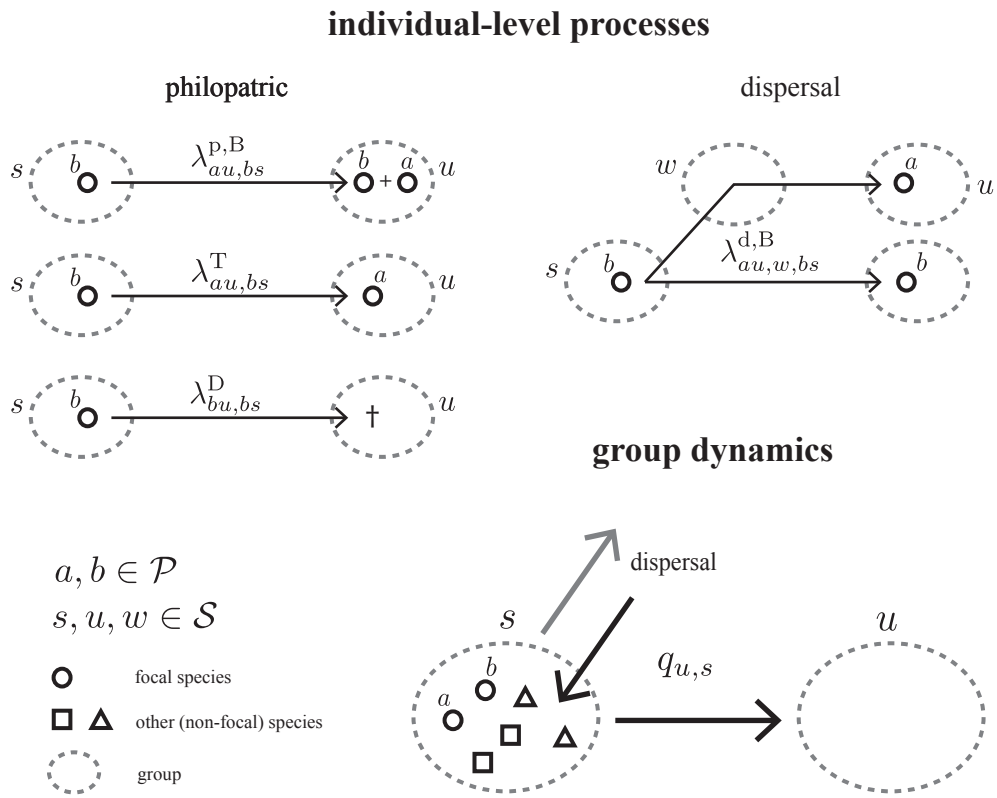

Figure 1: Individual-level processes of the focal species (top panels) and the group dynamics of the metacommunity (bottom panels). **Top panel:** As given in (5), the focal resident species undergoes philopatric individual-level processes of birth, physiological transition and death (top left panel), as well as birth and dispersal (top right panel). The dashed ovals represent groups and circles represent individuals of the focal species. In all processes the parent individual is in physiological state $b \in \mathcal{P}$ and inhabits a group in state $s \in \mathcal{S}$, and the arrow indicates who was being produced or removed as a result of this process. **Bottom panel:** Group transitions (7)-(9) are caused by the individual-level processes of the focal and all non-focal species as well as fluctuations in external factors. Group transitions determine the change in the composition of groups and the metacommunity at large as given in (6).

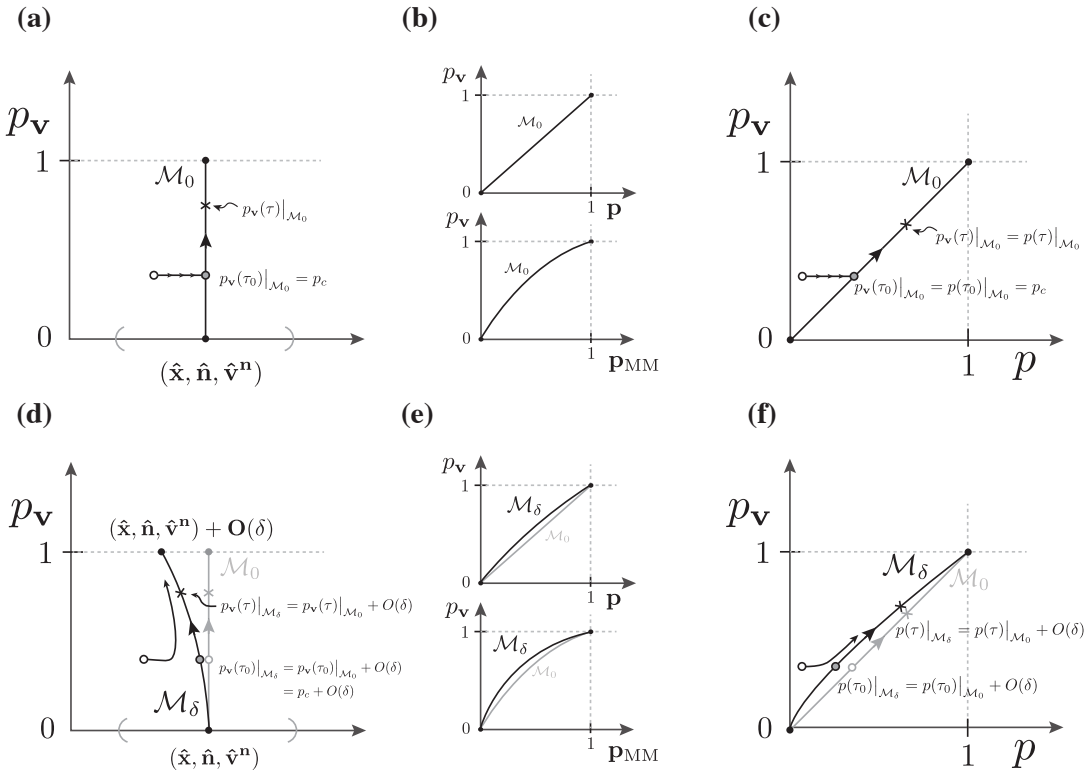

Figure 2: Panels (a)-(c) depict the fast and slow subsystems and the critical manifold $\mathcal{M}_0$ (76)-(77) (where $\delta = 0$, Sections 6.1-6.2), and, panels (d)-(f) depict their singular perturbations and the slow manifold $\mathcal{M}_\delta$ (where δ small but non-zero, Section 7). **(a)** The phase-plane for the fast variables $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}})$ (the "x-axis") and the slow variable $p_{\mathbf{v}}$. In the fast subsystem (Section 6.1) the slow variable $p_{\mathbf{v}}$ stays constant while the fast variables converge to $(\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}})$ whenever $\mathbf{x}(t_0)$ is in the neighborhood (indicated by gray semi-circles) of $\hat{\mathbf{x}}$ for some initial time t_0. A specific solution for the fast subsystem is depicted with small arrows and where the empty circle shows the initial condition for t_0. The slow subsystem (Section 6.2) is constrained on $\mathcal{M}_0$, and a particular solution is constructed by taking the steady state of the fast subsystem (where $t \to \infty$) as the initial condition for the slow subsystem $(\mathbf{x}(\tau_0), \mathbf{n}(\tau_0), \mathbf{v}^{\mathbf{n}}(\tau_0), p_{\mathbf{v}}(\tau_0)) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}, p_c)$. In this example $p_{\mathbf{v}}(\tau) \to 1$ as $\tau \to \infty$. **(b)** The shape of $\mathcal{M}_0$ on the $(\mathbf{p}, p_{\mathbf{v}})$-plane where $\hat{\mathbf{p}} = \mathbf{1} p_{\mathbf{v}}$ (top panel) and $(\mathbf{p}_{\mathrm{MM}}, p_{\mathbf{v}})$-plane where $\hat{\mathbf{p}}_{\mathrm{MM}} = \hat{\mathbf{r}} p_{\mathbf{v}} + (\mathbf{1} - \hat{\mathbf{r}}) p_{\mathbf{v}}^2$ (bottom panel). Note that the shape of $\mathcal{M}_0$ in the bottom panel depends on the steady state for relatedness $\hat{\mathbf{r}}$. **(c)** Phase-plane for the fast (and slow) arithmetic mutant frequency p (45) and the slow reproductive value weighted mutant frequency $p_{\mathbf{v}}$ (the notation is identical to panel (a)). Note that $p = p_{\mathbf{v}}$ on $\mathcal{M}_0$ and that their dynamics on $\mathcal{M}_0$ is as given in (78) and (82). **(d)** The phase plane for $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, p_{\mathbf{v}})$ as a singular perturbation of (a) where, in particular, the slow manifold $\mathcal{M}_\delta$ is $O(\delta)$-close to $\mathcal{M}_0$ as is the initial condition (gray circle) and the solution on $\mathcal{M}_\delta$. Moreover, because $\mathcal{M}_\delta$ is a stable manifold, the solutions in the neighborhood of $\mathcal{M}_\delta$ approach the solutions on $\mathcal{M}_\delta$. **(e)** The shape of $\mathcal{M}_\delta$ as a perturbation of $\mathcal{M}_0$ discussed in (b). **(f)** The phase plane for $(p, p_{\mathbf{v}})$ as a singular perturbation of (c). Note that the solutions $p(\tau)|_{\mathcal{M}_\delta}$ and $p_{\mathbf{v}}(\tau)|_{\mathcal{M}_\delta}$ are $O(\delta)$-distance away from each other because $\mathcal{M}_\delta$ lies off-diagonal. The interpretation of the solutions are identical to (d).

References

É. Ajar. Analysis of disruptive selection in subdivided populations. *BMC Evolutionary Biology*, 3 (1):22, 2003.

S. Alizon and P. Taylor. Empty sites can promote altruistic behavior. *Evolution: International Journal of Organic Evolution*, 62(6):1335–1344, 2008.

F. Arrigoni. Deterministic approximation of a stochastic metapopulation model. *Advances in Applied Probability*, 35(3):691–720, 2003.

A. D. Barbour and A. Pugliese. Convergence of a structured metapopulation model to levins's model. *Journal of Mathematical Biology*, 49(5):468–500, 2004.

C. Cadet, R. Ferrière, J. A. J. Metz, and M. van Baalen. The evolution of dispersal under demographic stochasticity. *American Naturalist*, 162(4):427–41, Oct 2003.

Y. Cai and S. A. H. Geritz. Resident-invader dynamics of similar strategies in fluctuating environments. *Journal of mathematical biology*, pages 1–53, 2020.

R. S. Cantrell, C. Cosner, and K. Lam. Resident-invader dynamics in infinite dimensional systems. *Journal of Differential Equations*, 263(8):4565–4616, 2017.

P. L. Chesson. Models for spatially distributed populations: the effect of within-patch variability. *Theoretical Population Biology*, 19(3):288–325, 1981.

P. L. Chesson. Coexistence of competitors in spatially and temporally varying environments: a look at the combined effects of different sorts of variability. *Theoretical Population Biology*, 28(3): 263–287, 1985.

F. B. Christiansen. On conditions for evolutionary stability for a continuously varying character. *The American Naturalist*, 138(1):37–50, 1991.

R. Dawkins. *The Blind Watchmaker*. Longman, Harlow, 1986.

R. Dawkins. *Climbing mount improbable*. WW Norton & Company, 1997.

F. Dercole. The ecology of asexual pairwise interactions: The generalized law of mass action. *Theoretical Ecology*, 9(3):299–321, 2016.

F. Dercole and S. Rinaldi. *Analysis of evolutionary processes: the adaptive dynamics approach and its applications*. Princeton University Press, 2008.

O. Diekmann, M. Gyllenberg, H. Huang, M. Kirkilionis, J. A. J. Metz, and H. R. Thieme. On the formulation and analysis of general deterministic structured population models II. nonlinear theory. *Journal of Mathematical Biology*, 43(2):157–189, 2001.

I. Eshel. Evolutionary and continuous stability. *Journal of Theoretical Biology*, 103(1):99–111, 1983.

S. Ethier and T. Nagylaki. Diffusion approximations of markov chains with two time scales and applications to population genetics. *Advances in Applied Probability*, 12(1):14–49, 1980.

S. Ethier and T. Nagylaki. Diffusion approximations of markov chains with two time scales and applications to population genetics, ii. *Advances in Applied Probability*, 20(3):525–545, 1988.

W. J. Ewens. *Mathematical population genetics 1: Theoretical introduction*, volume 27. Springer Science & Business Media, 2012.

N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *Journal of Differential Equations*, 31(1):53–98, 1979.

R. A. Fisher. *The genetical theory of natural selection.* OxfordClarendon Press, 1930.

S. A. Frank. *Foundations of social evolution*, volume 2. Princeton University Press, 1998.

S. Gandon and Y. Michalakis. Evolutionarily stable dispersal rate in a metapopulation with extinctions and kin competition. *Journal of Theoretical Biology*, 199(3):275–290, 1999.

S. A. H. Geritz, É. Kisdi, G. Meszéna, and J. A. J. Metz. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, 12(1):35–57, 1998.

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

A. Grafen. Biological fitness and the price equation in class-structured populations. *Journal of Theoretical Biology*, 373:62–72, 2015.

G. Greiner, J. A. P. Heesterbeek, and J. A. J. Metz. A singular perturbation theorem for evolution equations and time-scale arguments for structured population models. *Canadian Applied Mathematics Quarterly*, 3(4):435–459, 1994.

D. Grey, V. Hutson, and E. Szathmary. A re-examination of the stochastic corrector model. *Proceedings of the Royal Society B: Biological Sciences*, 262:29–35, 1995.

G. Hek. Geometric singular perturbation theory in biological practice. *Journal of Mathematical Biology*, 60(3):347–386, 2010.

M. Hirsch, R. Devaney, and S. Smale. *Differential equations, dynamical systems, and linear algebra*, volume 60. Academic press, 1974.

H. Inaba. *Age-structured population dynamics in demography and epidemiology.* Springer, 2017.

M. Iosifescu. *Finite Markov Processes and Their Applications.* Dover, New York, 2007.

H. C. Ito, U. Dieckmann, and J. A. J. Metz. Lotka–volterra approximations for evolutionary trait-substitution processes. *Journal of Mathematical Biology*, 80(7):2141, 2020.

M. Kirkpatrick, T. Johnson, and N. H. Barton. General models of multilocus evolution. *Genetics*, 161(4):1727–1750, 2002.

C. Kuehn. *Multiple time scale dynamics*, volume 191. Springer, 2015.

B. Kuijper and R. A. Johnstone. The evolution of early-life effects on social behaviour-why should social adversity carry over to the future? *Philos Trans R Soc Lond B Biol Sci*, 374(1770):20180111, 04 2019. doi: 10.1098/rstb.2018.0111.

L. Lehmann and F. Rousset. How life history and demography promote or inhibit the evolution of helping behaviours. *Philosophical transactions of the Royal Society B: biological sciences*, 365 (1553):2599–2617, 2010.

L. Lehmann and F. Rousset. The genetical theory of social behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642):20130357, 2014.

L. Lehmann, N. Perrin, and F. Rousset. Population demography and the evolution of helping behaviors. *Evolution*, 60(6):1137–1151, 2006.

L. Lehmann, I. Alger, and J. Weibull. Does evolution lead to maximizing behavior? *Evolution*, 69 (7):1858–1873, 2015.

L. Lehmann, C. Mullon, E. Akcay, and J. Van Cleve. Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution*, 70(8):1689–1702, 2016.

O. Leimar. The evolution of phenotypic polymorphism: randomized strategies versus evolutionary branching. *The American Naturalist*, 165(6):669–681, 2005.

O. Leimar. Multidimensional convergence stability. *Evolutionary Ecology Research*, 11(2):191–208, 2009.

T. Lenormand, D. Roze, and F. Rousset. Stochasticity in evolution. *Trends in Ecology and Evolution*, 24:157–165, 2009.

S. Lessard. Evolutionary stability: one concept, several meanings. *Theoretical population biology*, 37 (1):159–170, 1990.

S. Lessard. Diffusion approximations for one-locus multi-allele kin selection, mutation and random drift in group-structured populations: a unifying approach to selection models in population genetics. *Journal of mathematical biology*, 59(5):659–696, 2009.

S. Lessard and C. D. Soares. Frequency-dependent growth in class-structured populations: continuous dynamics in the limit of weak selection. *Journal of mathematical biology*, 77(1):229–259, 2018.

S. Lion. From the price equation to the selection gradient in class-structured populations: a quasi-equilibrium route. *Journal of Theoretical Biology*, 447:178–189, 2018a.

S. Lion. Class structure, demography, and selection: reproductive-value weighting in nonequilibrium, polymorphic populations. *The American Naturalist*, 191(5):620–637, 2018b.

M. Martcheva and H. R. Thieme. A metapopulation model with discrete size structure. *Natural Resource Modeling*, 18(4):379–413, 2005.

M. Martcheva and H. R. Thieme. Infinite ode systems modeling size-structured meta-populations and macroparasitic diseases. *Mathematical biology and epidemiology, Springer, Berlin Heidelberg New York*, 2006.

M. Martcheva, H. R. Thieme, and T. Dhirasakdanon. Kolmogorov's differential equations and positive semigroups on first moment sequence spaces. *Journal of Mathematical Biology*, 53(4):642–671, 2006.

F. Massol and F. Débarre. Evolution of dispersal in spatially and temporally variable environments: the importance of life cycles. *Evolution*, 69(7):1925–1937, 2015.

G. Meszéna, M. Gyllenberg, F. J. A. Jacobs, and J. A. J. Metz. Link between population dynamics and dynamics of darwinian evolution. *Physical Review Letters*, 95(7):078105, 2005.

J. A. J. Metz and M. Gyllenberg. How should we define fitness in structured metapopulation models? including an application to the calculation of evolutionarily stable dispersal strategies. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1466):499–508, 2001.

R. E. Michod and W. D. Hamilton. Coefficients of relatedness in sociobiology. *Nature*, 288:694–697, 1980.

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

K. Nipp and D. Stoffer. *Invariant manifolds in discrete and continuous dynamical systems.* EMS, 2013.

H. Ohtsuki, C. Rueffler, J. Y. Wakano, K. Parvinen, and L. Lehmann. The components of directional and disruptive selection in heterogeneous group-structured populations. *Journal of Theoretical Biology*, 507:1–16, 2020.

K. Parvinen. Joint evolution of altruistic cooperation and dispersal in a metapopulation of small local populations. *Theoretical Population Biology*, 85(1):12–19, 2013.

J. Peña. Group-size diversity in public goods games. *Evolution*, 66:623–636, 2011.

T. Priklopil and L. Lehmann. Invasion implies substitution in ecological communities with class-structured populations. *Journal of Theoretical Biology*, 134(134):36–52, 2020.

A. M. M. Rodrigues. Demography, life history and the evolution of age-dependent social behaviour. *J Evol Biol*, 31(9):1340–1353, 09 2018.

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

O. Ronce and D. Promislow. Kin competition, natal dispersal and the moulding of senescence by natural selection. *Proceedings of the Royal Society B-Biological Sciences*, 277(1700):3659–67, Dec 2010.

O. Ronce, S. Gandon, and F. Rousset. Kin selection and natal dispersal in an age-structured population. *Theoretical Population Biology*, 58(2):143–159, 2000.

F. Rousset. *Genetic structure and selection in subdivided populations*. Princeton Univ. Press, Princeton, 2004.

F. Rousset. Separation of time scales, fixation probabilities and convergence to evolutionarily stable states under isolation by distance. *Theoretical Population Biology*, 69(2):165–179, 2006.

F. Rousset and O. Ronce. Inclusive fitness for traits affecting metapopulation demography. *Theoretical Population Biology*, 65(2):127–141, 2004.

D. Roze and F. Rousset. Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics*, 165(4):2153–2166, 2003.

D. Roze and F. Rousset. The robustness of Hamilton's rule with inbreeding and dominance: kin selection and fixation probabilities under partial sib mating. *The American Naturalist*, 164(2): 214–231, 2004.

C. D. Soares and S. Lessard. Diffusion approximation for an age-class-structured population under viability and fertility selection with application to fixation probability of an advantageous mutant. *Journal of Mathematical Biology*, 79(6-7):2069–2110, 2019.

C. D. Soares and S. Lessard. First-order effect of frequency-dependent selection on fixation probability in an age-structured population with application to a public goods game. *Theoretical population biology*, 133:80–96, 2020.

P. Taylor. Allele-frequency change in a class-structured population. *The American Naturalist*, 135 (1):95–106, 1990.

J. Van Cleve, M. Feldman, and L. Lehmann. How demography, life history, and kinship shape the evolution of genomic imprinting. *The American Naturalist*, 176(4):440–455, 2010.

J. Wakeley. Polymorphism and divergence for island-model species. *Genetics*, 163(1):411–420, 2003.

J. Wakeley and T. Takahashi. The many-demes limit for selection and drift in a subdivided population. *Theoretical Population Biology*, 66(2):83–91, 2004.

G. Wild. Inclusive fitness from multitype branching processes. *Bulletin of Mathematical Biology*, 73 (5):1028–51, May 2011. ISSN 1522-9602.

G. Wild, A. Gardner, and S. A. West. Adaptation and the evolution of parasite virulence in a connected world. *Nature*, 459(7249):983–986, 2009.