bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. This can accommodate complex local community and global metacommunity environmental feedbacks owing to inter- and intra-specific interactions, as well as local environmental stochastic fluctuations. For the focal species, we derive a fitness measure for a mutant allele affecting class-specific trait expression. Using classical results from geometric 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. This invariance on allele frequency entails the mutant either dies out or will out-compete the ancestral resident (or 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 random mutation and natural selection in spatially and physiologically class structured metacommunities.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

> Invariance that had appeared in the criterion for altruism with respect to gene frequency in the case of sibs had seemed a gift from God, and I did not expect to see it repeated in the more complex trial cases I had moved on to. So it was with joy and almost with incredulity that I at last found emerging out of acres of my tedious and usually wrong algebra for the case of uncles, and then for the case of cousins, the same invariance as I had found before.
>
> (Hamilton, 1988, p. 189)

1 Introduction

Darwinian evolution – the gradual, step-by-step transformation of traits due to random mutation and non-random cumulative natural selection – is the central 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 fix in the population as a result of natural selection. Each fixation thus displays invariance in the direction of selection with respect to gene frequency, and repeated fixations under the constant influx of mutations enable gradual phenotypic evolution and adaptation. Such invariance with respect to gene frequency is, however, not expected to hold under general biological processes. In particular, social as well as ecological interactions can lead to survival and reproduction being frequency-dependent, and this may result in genetic polymorphisms preventing allelic fixation. Furthermore, in spatially and physiologically structured populations, it is not obvious to identify the conditions under which selective sweeps occur owing to the complications arising from genetic correlations at local spatial scales and across physiological classes.

In this paper we provide sufficient conditions for selective sweeps when evolution occurs in an asexual focal species that resides in a wider metacommunity (Wilson, 1992; Leibold et al., 2004), and where socially interacting individuals are structured into discrete classes that determine their physiological state, such as age and size, as well as their local biotic and abiotic environment. We provide these conditions in terms of an "invasion implies substitution"-principle, which is a formal characterization of the invariance of selection with respect to a population-wide frequency of a mutant allele that affects quantitative traits in the context of density- and frequency-dependent selection (Hamilton, 1964; Rousset, 2004; Meszéna et al., 2005; Dercole and Rinaldi, 2008; Lehmann and Rousset, 2014; Dercole, 2016; Cantrell et al., 2017; Ito et al., 2020; Cai and Geritz, 2020; Priklopil and Lehmann, 2020). The principle 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 (wild)-type, a successful invasion of the mutant allele will generically lead to its fixation. In the present metacommunity model, moreover, the dynamical process for some weighted population-wide mutant frequency in the focal species $p(t) \in [0,1]$ is supposed to take the form of a scalar-valued ordinary differential equation

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

Here, 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 z_{A} denoting the phenotype produced by a carrier of the resident allele in class $\mathrm{A} \in \mathcal{C}$ where $\mathcal{C}$ is the set of relevant classes individuals can be in, and where $\boldsymbol{\eta} = (\eta_{\mathrm{A}})_{\mathrm{A}\in\mathcal{C}}$ is the mutant phenotypic deviation. The expression $\delta S_{\boldsymbol{\eta}}(\boldsymbol{z})$ in (1) is the so-called

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

directional selection coefficient, and because it is frequency-independent the allele frequency affects the evolutionary dynamics only via the term $p(t)(1-p(t))$. Since $S_{\boldsymbol{\eta}}(\boldsymbol{z})$ is independent of time t, a successful invasion of the mutant allele implies its substitution and thus a selective sweep is obtained in the focal species. This in turn implies that in longer timescales recurrent mutation and invasion events will gradually transform the phenotype $\boldsymbol{z}$ of interest until the phenotype reaches the boundary of its feasible values or $S_{\boldsymbol{\eta}}(\boldsymbol{z})$ 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 allele is model independent, and that it can be expressed in terms of an inclusive fitness effect as

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

where all quantities are calculated from the resident population. 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 common ancestor, that is, their lineages coalesce backward in time. 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 inheriting the same allele, 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 $\mathbf{v} = (v_{\mathrm{A}})_{\mathrm{A}\in\mathcal{C}}$ are the right and left dominant eigenvectors of $\mathbf{H}$ and are normalized to give class frequencies and reproductive values, respectively. Expressions taking broadly the form (2) and special cases thereof are routinely applied in evolutionary biology to ascertain the directions of selection and evaluate singular points, be it in the kin selection, quantitative genetics, life-history, demography, and adaptive dynamics literatures (e.g., Charlesworth, 1994; Frank, 1998; Caswell, 2000; Rousset, 2004; Dercole and Rinaldi, 2008; Lynch and Walsh, 2018 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; Lion and Gandon, 2009 for lattice spatial structure and Cantrell et al., 2017 where dispersal follows a diffusion model], and our approach and results here extend and complement 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 class structured. Second, we provide a 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). Third, we pay particular attention to tie up all underlying concepts needed to prove gradual evolution in metacommunities, from the individual based demographic processes to the dynamics of moments of allelic states and the fast-slow analysis of all the relevant dynamical variables. We find that the approximation (1)-(2) holds for small but non-zero δ

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where the perturbation of steady states caused by the invasion is taken into account, a result given without proof previously. More importantly, we sharpen earlier representations of (1) in terms of the reproductive value weighted average frequency (Roze and Rousset, 2004; Rousset and Ronce, 2004; Rousset, 2004; Van Cleve et al., 2010; Priklopil and Lehmann, 2020) by allowing $p(t)$ to be defined as any population-wide average mutant frequency, with special cases being the reproductive value weighted mutant frequency as well as the arithmetic mean mutant frequency. In terms of the arithmetic mean, equation (1) becomes functionally equivalent to the standard and widely applied textbook representation of natural selection under additive gene action in spatially and physiologically homogeneous populations without social interactions nor ecological feedback (e.g., Crow and Kimura, 1970; Gillespie, 2004; Hartl et al., 1997).

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 $\delta S_{\boldsymbol{\eta}}(\mathbf{z})$ as characterised in (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 Outline of the model

We consider a metacommunity consisting of an infinite number of local communities connected to each other by random uniform dispersal. Each local community, which we will generally refer to as a local group, consists of a finite number of interacting asexual species 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 and assume a continuous time demographic process.

At each moment in time, 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, such as of certain age and size, and we denote the set of physiological states of the focal species with $\mathcal{P}$. All the other non-focal species can be similarly structured, and in addition each group is characterized by a set of *abiotic factors* such as the temperature or quality of the habitat. An abiotic factor is defined as any factor outside of the focal and non-focal species, be it a factor that is affected by the species or not, in which case we call it an external factor. Each group of individuals will therefore be in one of a finite number $k_{\mathcal{S}}$ of group-level states, defined as a union of physiological states of each and every individual of every species inhabiting the group, including any abiotic factors. The set of all group states (g-state) is denoted with $\mathcal{S}$. As each element of $\mathcal{S}$ prescribes exactly the size distributions of all species, including the focal, we will use N_{as} to denote the number of individuals $a \in \mathcal{P}$ in the focal species inhabiting a group in g-state $s \in \mathcal{S}$.

Each group in the metacommunity changes its state in $\mathcal{S}$ due to individual processes (i-processes) of the focal and non-focal species, as well as due to processes of abiotic factors such as changes in the temperature or quality of the habitat. The i-processes of the focal species are birth, where offspring of any physiological state is being produced, and dispersal, where each offspring either disperses randomly to other groups or stays in its natal group. Individuals may also undergo physiological development defined as a transition from one physiological state to another, as well as death. Whenever a group $s \in \mathcal{S}$ transitions to $u \in \mathcal{S}$ due to an event where an individual $a \in \mathcal{P}$ is produced into

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

or removed from the group, we have the identities $N_{as} + 1 = N_{au}$ or $N_{as} - 1 = N_{au}$, respectively, and otherwise $N_{as} = N_{au}$. Individuals of the other non-focal species may undergo similar i-process. All the processes are assumed stochastic occurring on a Poissonian basis (e.g., Kingman, 1992), and can depend in an arbitrary way on the physiological state of the individual, the g-state of the group it inhabits as well as the state of all the other non-focal groups.

Because the i-procesess of individuals may differ not only due to their physiological state but also the g-state of the group they inhabit, we will use the notion of a *class* of an individual. This is defined such that whenever any two individuals are identical in terms of their i-processes, 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 $k_{\mathcal{C}} = k_{\mathcal{P}} \times k_{\mathcal{S}}$.

Finally, individuals in the focal species express a genetically evolving phenotype that may affect the i-processes of any individual in the metacommunity. We assume that phenotypic expression is a class-specific function. For example, the phenotype may depend on the physiological state of the individual as well as the abiotic factors of the group that it inhabits. 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 that can result in different phenotypic expression segregate in the focal species, a mutant allele M and a resident allele R. We assume that the mutation rate is exceedingly small so that no other alleles interfere with the mutant-resident dynamics (this assumption can be mildly relaxed, see Section 8). 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); Parvinen et al. (2018); Kuijper and Johnstone (2019); Ohtsuki et al. (2020). No prior model has however considered metacommunities with multiple interacting species where individuals and groups are characterized by arbitrary states and as such all these previous models, as well as models with finite class-structure in panmictic populations, can be conceptually subsumed to the present analysis.

3 Resident dynamics in the metacommunity

The aim here is to describe 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 dynamics of the entire resident metacommunity. We define the state of the resident monomorphic metacommunity as a 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 $\mathcal{S}$, which is the simplex in $\mathbb{R}^{k_{\mathcal{S}}}$. We will represent the frequency distribution of the metacommunity at time $t \in \mathbb{R}$ with the vector $\mathbf{x} := \mathbf{x}(t)$, where $\mathbf{x} = (x_s)_{s \in \mathcal{S}} \in \Delta(\mathcal{S})$ and where x_s is the frequency of groups in state $s \in \mathcal{S}$ with $\sum_{s \in \mathcal{S}} x_s = 1$, and next we track the dynamics of this vector. While the upcoming characterization may at first appear lengthy, we fully work it out as (i) it

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plays a central role in the analysis of selection and (ii) it is of interest in the formulation of purely metacommunity ecological models.

3.1 Individual-level processes

To construct dynamical equations for the metacommunity dynamics, we start from i-processes of the focal species. To that end, we introduce four elementary individual-level rates (i-rates). First, the rate $\gamma_{a\leftarrow bs}$ at which an individual in physiological state $b \in \mathcal{P}$ inhabiting a group $s \in \mathcal{S}$, that is, an individual in class $bs \in \mathcal{C}$, transitions into physiological state $a \neq b \in \mathcal{P}$. By construction, $\sum_{a\neq b\in\mathcal{P}} \gamma_{a\leftarrow bs}$ then stands for the total rate at which an individual in class $bs \in \mathcal{C}$ undergoes a transition and $\gamma_{a\leftarrow bs}$ is undefined as a transition rate for $a = b \in \mathcal{P}$. Second, the rate μ_{bs} at which an individual in class $bs \in \mathcal{C}$ dies. Third, the rate $\beta^{\mathrm{p}}_{a\leftarrow bs}$ at which an individual in class $bs \in \mathcal{C}$ produces by reproduction an offspring individual in physiological state $a \in \mathcal{P}$ that stays in the focal group (hence superscript 'p' for philopatry), and finally, the rate $\beta^{\mathrm{d}}_{a\overset{w}{\leftarrow} bs}$ at which an individual in class $bs \in \mathcal{C}$ produces by reproduction an offspring individual $a \in \mathcal{P}$ into a non-focal group $w \in \mathcal{S}$, conditional on offspring landing in that group (hence superscript 'd' for dispersal). We further assume that dispersed offspring land into a group $w \in \mathcal{S}$ with probability x_w, and this assumption allows us to write $\beta^{\mathrm{d}}_{a\leftarrow bs} = \sum_{w\in\mathcal{S}} x_w \beta^{\mathrm{d}}_{a\overset{w}{\leftarrow} bs}$ giving the total rate at which $bs \in \mathcal{C}$ gives birth to $a \in \mathcal{P}$ into non-focal groups (Figure 1). By convention we assume that whenever $N_{bs} = 0$ then $\gamma_{a\leftarrow bs} = 0$, $\mu_{bs} = 0$, $\beta^{\mathrm{p}}_{a\leftarrow bs} = 0$ and $\beta^{\mathrm{d}}_{a\overset{w}{\leftarrow} bs} = 0$, and if group $s \in \mathcal{S}$ is at its maximum size then $\beta^{\mathrm{p}}_{a\leftarrow bs} = 0$ and $\beta^{\mathrm{d}}_{a\overset{s}{\leftarrow} bw} = 0$ for all $a, b \in \mathcal{P}, s, w \in \mathcal{S}$.

Formally, we can express the four basic rates of the focal species as the mappings

$$\gamma_{a\leftarrow bs},\ \mu_{bs},\ \beta^{\mathrm{p}}_{a\leftarrow bs},\ \beta^{\mathrm{d}}_{a\overset{w}{\leftarrow} bs} : \mathcal{Z}_{\mathcal{C}} \times \mathcal{Z} \times \mathcal{S} \times \mathcal{Z} \times \Delta(\mathcal{S}) \to \mathbb{R}_+ \qquad \text{for all } a, b \in \mathcal{P},\ s \in \mathcal{S}, \quad (3)$$

such that, for instance, the first argument in $\gamma_{a\leftarrow bs}(z_{bs}, \boldsymbol{z}, s, \boldsymbol{z}, \mathbf{x})$ indicates the class-specific phenotype $z_{bs} \in \mathcal{Z}_{\mathcal{C}}$ expressed by the focal individual who undergoes the i-process, and the remaining arguments describe the group of the focal individual and the rest of the metacommunity that affect these i-processes. That is, the second argument $\boldsymbol{z} \in \mathcal{Z}$ indicates the phenotypic profile of resident individuals of the focal species that inhabit the group of the focal individual and the third argument $s \in \mathcal{S}$ gives the state of the entire focal group, namely, the distribution of the focal and non-focal species as well as abiotic factors (Section 2). 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 $\mathbf{x} \in \Delta(\mathcal{S})$. Moreover, in calculations involving the i-rates (3), it will be useful to use the short-hand notation

$$\gamma_{a\leftarrow bs}(\mathbf{x}) = \gamma_{a\leftarrow bs}(z_{bs}, \boldsymbol{z}, s, \boldsymbol{z}, \mathbf{x}) \qquad \text{for all } a, b \in \mathcal{P},\ s \in \mathcal{S}, \quad (4)$$

indicating the dependence on the state of the metacommunity. The arguments and the short-hand notation of functions $\mu_{bs}, \beta^{\mathrm{p}}_{a\leftarrow bs}$ and $\beta^{\mathrm{d}}_{a\overset{w}{\leftarrow} bs}$, as well as other individual-level functions introduced henceforth, are interpreted identically. Note that the mapping for the total birth rate via dispersal $\beta^{\mathrm{d}}_{a\leftarrow bs}$ can be obtained by using $\beta^{\mathrm{d}}_{a\overset{w}{\leftarrow} bs}$.

For ease of presentation and calculation, we will further use a single symbol for all the i-rates (3), and in addition indicate the g-state of the group after the focal individual has undergone the i-process. To this end, and for all $au, bs \in \mathcal{C}$, the class-specific physiological transition rate satisfies

$$\lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) = \begin{cases} \gamma_{a\leftarrow bs}(\mathbf{x}) & \text{for} \quad a \neq b \quad \text{and} \quad N_{au} = N_{as} + 1 \quad \text{and} \quad N_{bu} = N_{bs} - 1 \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the class-specific death rate satisfies

$$\lambda^{\mathrm{D}}_{au,bs}(\mathbf{x}) = \begin{cases} \mu_{bs}(\mathbf{x}) & \text{for} \quad a = b \text{ and } N_{bu} = N_{bs} - 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

and the class-specific philopatric birth rate satisfies

$$\lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) = \begin{cases} \beta^{\mathrm{p}}_{a\leftarrow bs}(\mathbf{x}) & \text{for} \quad N_{au} = N_{as} + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The class-specific conditional birth rate via dispersal satisfies

$$\lambda^{\mathrm{dB}}_{au,w,bs}(\mathbf{x}) = \begin{cases} \beta^{\mathrm{d}}_{a\overset{w}{\leftarrow} bs}(\mathbf{x}) & \text{for} \quad N_{au} = N_{aw} + 1 \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

and so the total birth rate via dispersal is

$$\lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) = \sum_{w\in\mathcal{S}} x_w \lambda^{\mathrm{dB}}_{au,w,bs}(\mathbf{x}). \tag{9}$$

3.2 Group state dynamics

The dynamics of the frequency of g-states in the metacommunity $\mathbf{x} \in \Delta(\mathcal{S})$ can now be expressed as a system of ODE's

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

where the dot represents the time derivative and $\mathbf{Q}(\mathbf{x}) = (q_{u,s}(\mathbf{x}))_{u,s\in\mathcal{S}} \in \mathbb{R}^{k_{\mathcal{S}}\times k_{\mathcal{S}}}$ 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 $u \in \mathcal{S}$. Consistency requires that $\sum_{u\in\mathcal{S}} q_{u,s}(\mathbf{x}) = 0$ for all $s \in \mathcal{S}$ and hence $q_{s,s}(\mathbf{x}) = -\sum_{w\neq s\in\mathcal{S}} q_{w,s}(\mathbf{x})$ gives the rate at which a group transitions from $s \in \mathcal{S}$ to any other g-state (Figure 1). Due to the independence of the i-processes (Section 3.1), we can write

$$q_{u,s}(\mathbf{x}) = q^{\mathrm{p}}_{u,s}(\mathbf{x}) + q^{\mathrm{d}}_{u,s}(\mathbf{x}) + q^{\mathrm{mc}}_{u,s}(\mathbf{x}), \tag{11}$$

where the group transitions due to philopatric i-processes of the focal species as

$$q^{\mathrm{p}}_{u,s}(\mathbf{x}) = \sum_{a,b\in\mathcal{P}} \lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) N_{bs}, \tag{12}$$

with

$$\lambda^{\mathrm{p}}_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{D}}_{au,bs}(\mathbf{x}), \tag{13}$$

and that due to dispersal as

$$q^{\mathrm{d}}_{u,s}(\mathbf{x}) = \sum_{a,c\in\mathcal{P}} \sum_{w\in\mathcal{S}} \lambda^{\mathrm{dB}}_{au,s,cw}(\mathbf{x}) N_{cw} x_w. \tag{14}$$

This sums both over all physiological states of the parents and their possible progeny as well, and recall that N_{as} is a constant specifying the number of individuals $a \in \mathcal{P}$ in group $s \in \mathcal{S}$ (Section 2). 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 (10). First, one can interpret $\mathbf{x}$ as the frequency distribution of an *infinite collection of groups* and which is taken

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as the state of the metacommunity (as discussed above). In this view the ODE in (10) 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 interpret 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 (11) 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 call this the 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 describe the population dynamics of the focal species, which needs to be done in the backdrop of the full metacommunity state dynamics (10). 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 in a monomorphic metacommunity, where $n_{as} = N_{as}x_s$ defines the density of individuals in class $as \in \mathcal{C}$. Because groups are assumed to be well-mixed in the meta-community, the density of individuals refers to the number of individuals per unit space on the scale of groups. The vector $\mathbf{n} \in \mathbb{R}_+^{k_\mathcal{C}}$ satisfies the system of ODEs

$$\dot{\mathbf{n}} = \mathbf{H}(\mathbf{x})\mathbf{n}, \tag{15}$$

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 elements of this matrix are

$$h_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) + \delta_{a,b}\left[q_{u,s}(\mathbf{x}) - \sum_{c\in\mathcal{P}} \lambda^{\mathrm{T}}_{cu,bs}(\mathbf{x}) - \lambda^{\mathrm{D}}_{bu,bs}(\mathbf{x})\right], \tag{16}$$

for all $au, bs \in \mathcal{C}$, which is the rate at which a focal 'parent' individual in class $bs \in \mathcal{C}$ produces or removes an 'offspring' individual of class $au \in \mathcal{C}$ (see for a derivation Appendix A, Section A.1).

Equation (16) gives a representation of the class-specific individual growth-rate, or individual fitness, whose arguments are identically interpreted to those of the i-rates (3)-(4) and it consists of four terms. The first three terms describe the rate at which a parent individual $bs \in \mathcal{C}$ produces offspring individuals $au \in \mathcal{C}$ via dispersal, philopatric birth and physiological transition as detailed in Section 3.1, and in the case of physiological transition the parent and offspring individuals are interpreted as one and the same individual. The final term in (16) describes the 'survival' of the parent individual $bs \in \mathcal{C}$ in a changing local environment, where $\delta_{a,b}$ is the Kronecker delta taking value 1 when $a = b \in \mathcal{P}$ and otherwise 0. More specifically, for $s \neq u \in \mathcal{S}$, the final term in (16) gives the rate at which the parent $a = b \in \mathcal{P}$ survives the group transition from $s \in \mathcal{S}$ to $u \in \mathcal{S}$, and because $q_{u,s}(\mathbf{x})$ contains the i-processes of all individuals including the focal parent, the physiological transition and death of the parent are being subtracted in (16). For $s = u \in \mathcal{S}$, the survival term reduces to $q_{ss}(\mathbf{x}) = -\sum_{w\neq s} q_{w,s}(\mathbf{x})$ and gives the rate at which the parent $au = bs \in \mathcal{C}$ is removed from class $au = bs \in \mathcal{C}$ due to any event that causes the group to transition from $s = u \in \mathcal{S}$ to any other state in $\mathcal{S}$, including events caused by the focal individual (and note that in contrast to other processes dispersal may also contribute to fitness for $s = u \in \mathcal{S}$). The first term in (16) thus contributes to the rate of offspring production through dispersal while the second, third and fourth term together gives the net philopatric fitness. We note that in order to study the dynamics of $\mathbf{n}$, the ODE (15) needs to be coupled with the dynamics for the metacommunity state $\mathbf{x}$ in (10).

The dispersal and net philopatric fitness contributions of individual fitness (16) corresponds to

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.1.1). By using (16), we can also represent the individual fitness for models that have been previously considered in the literature. For instance, supposing that the metacommunity is composed only of the focal species that is physiologically unstructured and where groups are identical in terms of abiotic factors, the individual fitness (16) reduces to $h_{N',N}(\mathbf{x}) = \lambda^{\mathrm{d}}_{N',N}(\mathbf{x}) + \lambda^{\mathrm{pB}}_{N',N}(\mathbf{x}) + \left(q_{N',N}(\mathbf{x}) - \lambda^{\mathrm{D}}_{N,N}(\mathbf{x})\right)$ with $q_{N',N}(\mathbf{x}) = \left(\lambda^{\mathrm{pB}}_{N',N}(\mathbf{x}) + \lambda^{\mathrm{D}}_{N',N}(\mathbf{x})\right) N + \sum_{K\in\mathcal{S}} \lambda^{\mathrm{dB}}_{N',N,K}(\mathbf{x}) K x_K$, for all $N, N' \in \mathcal{S}$, where the set of g-states $\mathcal{S} = \{0, 1, 2, \ldots, N_{\max}\}$ specifies the number of individuals inhabiting a group and where $N_{\max}$ is the maximum group size (Appendix A, Section A.1.1). Thus, the only processes occurring in each group are death, philopatric birth and birth via dispersal of individuals in the focal species, which is conceptually similar to a number of previous models (Metz and Gyllenberg, 2001; Cadet et al., 2003; Rousset and Ronce, 2004; Lehmann et al., 2006; Alizon and Taylor, 2008; Parvinen, 2013). Further, for well-mixed populations where each 'group' is either empty or is inhabited by a single physiologically structured individual of the focal population, the individual fitness reduces to $h_{a,b}(\mathbf{n}) = \lambda^{\mathrm{d}}_{a,b}(\mathbf{n}) + \lambda^{\mathrm{T}}_{a,b}(\mathbf{n}) - \delta_{a,b}\left(\sum_{c\in\mathcal{P}} \lambda^{\mathrm{T}}_{c,b}(\mathbf{n}) + \lambda^{\mathrm{D}}_{b,b}(\mathbf{n})\right)$, for all $a, b \in \mathcal{C}$, where we have $\mathcal{S} = \{0\} \cup \mathcal{P}$ and where the population demographic state can be characterized directly in terms of the vector $\mathbf{n} = (n_a)_{a\in\mathcal{P}}$ of class densities (Appendix A, Section A.1.1). Now, the only processes are physiological transition, death and birth via dispersal because each 'group' can contain only a single individual, and which is similar to that of the spatially well-mixed model of Lion (2018a,b, eq. 1) and Priklopil and Lehmann (2020, eq. 3) and conceptual equivalent to standard discrete models of only physiologically structured populations (e.g., Caswell, 2000). This model reduces to an age structured model with T discrete age groups by setting $\mathcal{P} = \{1, 2, .., T\}$ and where newborns are individuals of age group $1 \in \mathcal{P}$ produced at rate β_b by an individual of age group (class) $b \in \mathcal{P}$ so that $\lambda^{\mathrm{d}}_{a,b}(\mathbf{n}) = \beta_b$ when $a = 1$ and zero otherwise, and where $\lambda^{\mathrm{T}}_{a,b}(\mathbf{n}) = \gamma_b$ is the constant rate at which an individual in age group $b \in \mathcal{P}$ progresses to age group $a \in \mathcal{P}$ and where the (average) unit of time an individual spends in an age class $b \in \mathcal{P}$ is $1/\gamma_b$ (e.g., Li and Brauer, 2008). Setting $\gamma_b = \gamma$ for all $b \in \mathcal{P}$ where age progresses at a steady rate is conceptually similar to the classic discrete-time age-structured models (e.g. Charlesworth, 1994; Caswell, 2000).

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; Lion, 2018b) and use the following population dynamical definition of reproductive values $\mathbf{v}^{\mathbf{n}} = (v^n_{as})_{as\in\mathcal{C}} \in \mathbb{R}^{k_{\mathcal{C}}}_+$:

$$-\dot{\mathbf{v}}^{\mathbf{n}} = \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}), \tag{17}$$

normalized such that $\mathbf{v}^{\mathbf{n}} \cdot \mathbf{n} = 1$ for all t. Because (17) is the adjoint equation (backward in time counterpart) to (15), the superscript in $\mathbf{v}^{\mathbf{n}}$ indicates that the reproductive value is an adjoint variable to $\mathbf{n}$. Adjoint variables are useful in the analysis of dynamical systems because, by definition, the product of a variable with its adjoint variable is constant in time (e.g. Greiner et al., 1994; Inaba, 2017 in the context of age-structured populations) and one can readily check by applying (15) that $\mathbf{v}^{\mathbf{n}} \cdot \mathbf{n}$ indeed stays constant in time. Similarly to (15), in order to study the dynamics of $\mathbf{v}^{\mathbf{n}}$, the ODE (17) needs to be coupled with the dynamics for the metacommunity state $\mathbf{x}$ in (10). We note that the above definition of reproductive values departs slightly from the standard definition (Taylor, 1990; Lion, 2018b) given as an adjoint to the class frequencies, see further discussion in Section 7.2.

The biological interpretation of $\mathbf{v}^{\mathbf{n}}$ is obtained from (17) and is tied to a final value problem where (17) is coupled with a final condition $\mathbf{v}^{\mathbf{n}}(t_{\mathrm{f}})$ where $t_{\mathrm{f}} \in [t, \infty]$ 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 in evolutionary analysis, 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.

3.4 Relatedness dynamics

The remaining variable that appears in (2) and whose dynamics we aim to characterize in the resident population is relatedness. We *define* relatedness between two different individuals $a, b \in \mathcal{P}$ 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. In other words, it is the probability that the ancestral lineages of the two individuals coalesce in a common ancestor, and because we assume an infinite number of groups the probability that two individuals sampled from two different groups share a common ancestor is 0. 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), but we will here derive it from the forward perspective of the demographic process so as to treat all variables in a similar fashion.

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. Note that in a continuous time Poisson-based model with one birth per unit of time, one of the two individuals that share a parent must be the parent through survival. 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 and whose size is denoted with $k_\mathcal{B}$. Then,

$$n^r_{abs} = r_{abs} n_{abs} \tag{18}$$

is the density of pairs of individuals who are related and where

$$n_{abs} = N_{abs} x_s \tag{19}$$

is the total density of pairs $abs \in \mathcal{B}$ in that group. Here,

$$N_{abs} = \begin{cases} N_{as} N_{bs} & \text{for} \quad b \neq a \\ \frac{N_{as}(N_{as}-1)}{2} & \text{for} \quad b = a \end{cases} \tag{20}$$

is 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'. By differentiating (19) we obtain an ODE for the pair dynamics

$$\begin{aligned} \dot{n}_{abu} &= N_{abu} \dot{x}_u = N_{abu} \sum_{s \in \mathcal{S}} q_{u,s}(\mathbf{x}) x_s \\ &= N_{abu} \sum_{s \neq u \in \mathcal{S}} q_{u,s}(\mathbf{x}) x_s - n_{abu} \sum_{w \neq u \in \mathcal{S}} q_{w,u}(\mathbf{x}), \quad \text{for all } abu \in \mathcal{B}, \end{aligned} \tag{21}$$

where we used (10), (19) and the fact that $\sum_{u \in \mathcal{S}} q_{u,s}(\mathbf{x}) = 0$ (Section 3.2). By substituting (11)-(14)

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

into (21) and using the relationship $n_{as} = N_{as}x_s$, we get

$$\begin{aligned}\dot{n}_{abu} = \sum_{s\neq u\in\mathcal{S}} \Big[\sum_{d\in\mathcal{P}} \alpha^{\mathrm{CA}}_{abu,ds}(\mathbf{x})n_{ds} + \sum_{d\in\mathcal{P}} \alpha^{\mathrm{IM}}_{abu,ds}(\mathbf{x})n_{ds} + \sum_{c,d\in\mathcal{P}} \alpha^{\mathrm{DP}}_{abu,cds}(\mathbf{x})n_{cds}\Big] - \\ - n_{abu} \sum_{w\neq u\in\mathcal{S}} q_{w,u}(\mathbf{x}), \quad \text{for all } abu \in \mathcal{B},\end{aligned} \tag{22}$$

which depends on three rates. First,

$$\alpha^{\mathrm{CA}}_{abu,ds}(\mathbf{x}) = \begin{cases} \lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) & \text{for} \quad d = b, \\ \lambda^{\mathrm{pB}}_{bu,as}(\mathbf{x}) & \text{for} \quad d = a, \\ 0 & \text{otherwise,} \end{cases} \tag{23}$$

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,

$$\alpha^{\mathrm{IM}}_{abu,ds}(\mathbf{x}) = \begin{cases} \sum_{w\in\mathcal{S}} \sum_{e\in\mathcal{P}} \lambda^{\mathrm{d}}_{au,s,ew}(\mathbf{x})N_{ew}x_w & \text{for} \quad d = b, \\ \sum_{w\in\mathcal{S}} \sum_{e\in\mathcal{P}} \lambda^{\mathrm{d}}_{bu,s,ew}(\mathbf{x})N_{ew}x_w & \text{for} \quad d = a, \\ 0 & \text{otherwise,} \end{cases} \tag{24}$$

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 $abu \in \mathcal{B}$. Recall from Section 3.1 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,

$$\alpha^{\mathrm{DP}}_{abu,cds}(\mathbf{x}) = \alpha^{\mathrm{pB,T}}_{abu,cds}(\mathbf{x}) + \delta_{ab,cd}\alpha^{\mathrm{S}}_{abu,abs}(\mathbf{x}), \tag{25a}$$

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,

$$\alpha^{\mathrm{pB,T}}_{abu,cds}(\mathbf{x}) = \begin{cases} \lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) & \text{for} \quad d = b \quad \text{and} \quad c \neq b, \\ 2(\lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x})) & \text{for} \quad d = b \quad \text{and} \quad c = b, \\ \lambda^{\mathrm{pB}}_{bu,as}(\mathbf{x}) + \lambda^{\mathrm{T}}_{bu,as}(\mathbf{x}) & \text{for} \quad c = a \quad \text{and} \quad d \neq a, \\ 2(\lambda^{\mathrm{pB}}_{bu,as}(\mathbf{x}) + \lambda^{\mathrm{T}}_{bu,as}(\mathbf{x})) & \text{for} \quad c = a \quad \text{and} \quad d = a, \\ 0 & \text{otherwise,} \end{cases} \tag{25b}$$

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. 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,

$$\alpha^{\mathrm{S}}_{abu,abs}(\mathbf{x}) = q_{u,s}(\mathbf{x}) - [\lambda^{\mathrm{D}}_{au,as}(\mathbf{x}) + \lambda^{\mathrm{D}}_{bu,bs}(\mathbf{x})] - \sum_{c\in\mathcal{P}} [\lambda^{\mathrm{T}}_{cu,as}(\mathbf{x}) + \lambda^{\mathrm{T}}_{cu,bs}(\mathbf{x})], \tag{25c}$$

which is the rate at which each existing pair $abs \in \mathcal{B}$ survives upon any i-process occurring in the group (note the similarity of this expression with the 'survival' term in (16)). 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 rates (5)-(9) at which single individuals undergo demographic i-processes.

The dynamics of n^r_{abs} is obtained from (22) by subtracting the immigration term, because immigrants

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are not related, and by only considering the fraction of local pairs that are related and we get

$$\dot{n}^{\mathrm{r}}_{abu} = \sum_{s\neq u\in\mathcal{S}}\sum_{d\in\mathcal{P}} \alpha^{\mathrm{CA}}_{abu,ds}(\mathbf{x})n_{ds} + \sum_{s\neq u\in\mathcal{S}}\sum_{c,d\in\mathcal{P}} \alpha^{\mathrm{DP}}_{abu,cds}(\mathbf{x})n^{r}_{cds} - n^{r}_{abu}\sum_{w\neq u\in\mathcal{S}} q_{w,u}(\mathbf{x}), \tag{26}$$

for all pairs $abu \in \mathcal{B}$. We see from the first two terms in (26) that the density of related pairs $abu \in \mathcal{B}$ increases due to a single local parent producing pairs anew, or, due to two related local parents producing offspring-pairs. From the last term in (26) we see that the density of related pairs $abu \in \mathcal{B}$ decreases due to any event that changes the state of the focal group $u \in \mathcal{S}$. 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

$$\dot{r}_{abu} = \frac{1}{n_{abu}}\left[\sum_{s\neq u\in\mathcal{S}}\sum_{d\in\mathcal{P}} \alpha^{\mathrm{CA}}_{abu,ds}(\mathbf{x})n_{ds} + \sum_{s\neq u\in\mathcal{S}}\sum_{c,d\in\mathcal{P}} \alpha^{\mathrm{DP}}_{abu,cds}(\mathbf{x})n_{cds}r_{cds} - r_{abu}N_{abu}\sum_{s\neq u\in\mathcal{S}} q_{u,s}(\mathbf{x})x_s\right] \tag{27}$$

for all $abu \in \mathcal{B}$, and where we have used (21), (26) and $n^{r}_{cds} = r_{cds}n_{cds}$. Notice that the rates that appear in (21) and (26) at which pairs $abu \in \mathcal{B}$ are removed due to the group $u \in \mathcal{S}$ changing its state, cancel in (27). 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 (15) and (17), in order to study the dynamics in (27) it needs to be coupled with the dynamics for the metacommunity state $\mathbf{x}$ in (10).

By ignoring physiological structure in the focal species, the dynamics in (27) 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 and where the vector $\mathbf{x}$ plays no role, the continuous time dynamical equation for relatedness (27) is simplified further to

$$\dot{r} = \frac{2}{N(N-1)}\left[\alpha^{\mathrm{CA}}N + \alpha^{\mathrm{DP}}\frac{N(N-1)}{2}r - rq\frac{N(N-1)}{2}\right], \tag{28}$$

where q is the total rate at which events happen in a group. When each birth event is coupled with a death event (i.e., the Moran process, Ewens, 2012), the rates in (28) can be written as

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

where m is the probability that upon a birth event the offspring disperses. Here, α^{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 α^{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 (27) as a discrete time jump-process, with jumps occurring whenever an 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(10) (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 $T_1, T_2, \ldots$, where each jump time T_k is a random variable at which the focal group undergoes an event either due to local or non-local (via immigration) i-processes. The jump process for relatedness is given in vector form as

$$\mathbf{r}' = \mathbf{c}(\mathbf{x}) + \mathbf{D}(\mathbf{x})\mathbf{r}, \tag{30}$$

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 in the same group, and hence a common ancestor. This is the coalescence probability and is computed as $c_{abu}(\mathbf{x}) = \sum_{ds\in\mathcal{C}} C_{abu,ds}(\mathbf{x})$ with

$$C_{abu,ds}(\mathbf{x}) = \frac{\alpha^{\mathrm{CA}}_{abu,ds}(\mathbf{x})N_{ds}}{N_{abu}}\frac{x_s}{q_u}, \tag{31}$$

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 groups transition to state $u \in \mathcal{S}$. Equation (31) was obtained using the rates (23), and q_u is obtained from the first expression on the right hand side of (21). Equation (31) gives the probability, conditional on the occurrence of a group transition to $u \in \mathcal{S}$, that the 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 $ds \in \mathcal{C}$. 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

$$D_{abu,cds}(\mathbf{x}) = \frac{\alpha^{\mathrm{DP}}_{abu,cds}(\mathbf{x})N_{cds}}{N_{abu}}\frac{x_s}{q_u}, \tag{32}$$

for all $s \neq u \in \mathcal{C}$ and 0 otherwise, and gives the probability that, conditional on the occurrence of a group transition to $u \in \mathcal{S}$, that the group was in state $s \in \mathcal{S}$ before the transition and a randomly sampled offspring-pair $abu \in \mathcal{B}$ was produced by a local parent-pair $cds \in \mathcal{B}$ (and where we used (25)). Note that the probabilities (31)–(32) are both defined to be 0 whenever $N_{abu} = 0$.

The recursion (30) along with (31)–(32) reduces to recursion (A.5) of Alizon and Taylor (2008) who considered group size as the only class-structure and hence $\mathcal{C} = \mathcal{S}$. The recursion (30) 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 $\mathcal{C} = \mathcal{P}\times\mathcal{S}$, Ohtsuki et al., 2020, eq. 33 for $\mathcal{C} = \mathcal{S}$). Yet owing to the continuous time nature of our model, expressions (31)–(32) do not exactly match those of the discrete time process since only one individual-level event 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 (30) simplifies to

$$r' = c + Dr, \tag{33}$$

where

$$\begin{aligned} c &= \frac{2\alpha^{\mathrm{CA}}N}{N(N-1)}\frac{1}{q} = \frac{2}{N^2} \\ D &= \frac{\alpha^{\mathrm{DP}}}{q} = 2(1-m)\frac{1}{N}\left(\frac{N-1}{N}\right) + \left(1-\frac{2}{N}\right), \end{aligned} \tag{34}$$

and where the expressions and the interpretation for $\alpha^{\mathrm{CA}}, \alpha^{\mathrm{DP}}$ and q are given in (28)-(29).

3.5 Steady states

In the evolutionary analysis that follows we assume that a non-trivial steady state of the metacommunity state dynamics (10) satisfying $\mathbf{0} = \mathbf{Q}(\hat{\mathbf{x}})\hat{\mathbf{x}}$ exists, and that it is hyperbolically stable. A

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (15) 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 (17) is linear in $\mathbf{v^n}$, we have $\mathbf{v^n}(t) \to \hat{\mathbf{v}}^{\mathbf{n}}$ whenever $\mathbf{x}(t) \to \hat{\mathbf{x}}$ as $t \to \infty$ and the steady state for $\mathbf{v^n}$ 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 (27) or the jump-process in (30). By a direct calculation we get

$$\hat{\mathbf{r}} = (\mathbf{I} - \mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{c}(\hat{\mathbf{x}}), \tag{35}$$

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). 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 $(\mathbf{n}(t), \mathbf{v^n}(t), \mathbf{r}(t))$ converge to their unique steady state $(\hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}, \hat{\mathbf{r}})$ as $t \to \infty$.

As a check of the recurrence equations for relatedness, we note that for the homogeneous island model we obtain $\hat{r} = c/(1-D) = (1-m)/(1-m+Nm)$ from (34), 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.

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, in addition to class, on the genetic state of the individual. 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 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 $\mathcal{I}_{as} = \{i_{as}\}_{0\le i_{as}\le N_{as}}$ 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 abiotic factors, the set $\mathcal{I}$ refers only to the genetic structure of the focal species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 $(s,i) := si \in \Omega$ and 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 $\mathbf{x} = (x_s)_{s\in\mathcal{S}} \in \Delta(\mathcal{S})$. All the upcoming rates describing the mutant-resident metacommunity dynamics are identical to those given for the resident dynamics (Section 3), the only difference is that it takes into account the genetic structure of the group and the metacommunity at large.

4.1 Individual-level processes

The rates at which i-processes of the focal species in the mutant-resident metacommunity occur are defined, similarly to the resident metacommunity in Section 3.1, as mappings

$$\gamma^{\theta}_{a\leftarrow bsi},\ \mu^{\theta}_{bsi},\ \beta^{\theta,\mathrm{p}}_{a\leftarrow bsi},\ \beta^{\theta,\mathrm{d}}_{a\overset{w}{\leftarrow} bsi} : \mathcal{Z}_{\mathcal{C}} \times \mathcal{Z}^2 \times \Omega \times \mathcal{Z}^2 \times \Delta(\Omega) \to \mathbb{R}_+ \quad \text{for all } au, bs \in \mathcal{C},\ i \in \mathcal{I}_s, \tag{36}$$

where $\theta \in \{\mathrm{M}, \mathrm{R}\}$ specifies whether it is a mutant or a resident who undergoes the process. We also use a shorthand notation $\gamma^{\theta}_{a\leftarrow bs}(\mathbf{y}) := \gamma^{\theta}_{a\leftarrow bs}(z_{\theta,bs}, \boldsymbol{z}_{\mathrm{M}}, \boldsymbol{z}_{\mathrm{R}}, si, \boldsymbol{z}_{\mathrm{M}}, \boldsymbol{z}_{\mathrm{R}}, \mathbf{y})$ where the arguments are interpreted similarly to Section 3.1, and the interpretation and shorthand notation for the other rates in (36) is similar. Moreover, we assume that dispersed offspring land into a group $wi \in \Omega$ with a probability x_w, that is, independent of its genetic state, and hence we will write $\beta^{\theta,\mathrm{d}}_{a\leftarrow bsi}(\mathbf{y}) = \sum_{w\in\mathcal{S}} x_w \beta^{\theta,\mathrm{d}}_{a\overset{w}{\leftarrow} bsi}(\mathbf{y})$. The constrains for the i-rates (36) are also analogous to Section 3.1. Finally, and for all $au, bs \in \mathcal{C}$ and $j \in \mathcal{I}_u, i \in \mathcal{I}_s$, the class-specific physiological transition rate satisfies

$$\lambda^{\mathrm{M,T}}_{auj,bsi}(\mathbf{y}) = \begin{cases} \gamma^{\mathrm{M}}_{a\leftarrow bsi}(\mathbf{y}) & \text{for} \quad a \neq b \quad \text{and} \quad j_{au} = i_{as} + 1 \quad \text{and} \quad j_{bu} = i_{bs} - 1 \\ 0 & \text{otherwise,} \end{cases} \tag{37}$$

the class-specific death rate satisfies

$$\lambda^{\mathrm{M,D}}_{auj,bsi}(\mathbf{y}) = \begin{cases} \mu^{\mathrm{M}}_{bs}(\mathbf{y}) & \text{for} \quad a = b \text{ and } j_{bu} = i_{bs} - 1 \\ 0 & \text{otherwise,} \end{cases} \tag{38}$$

and the class-specific philopatric birth rate satisfies

$$\lambda^{\mathrm{M,pB}}_{auj,bsi}(\mathbf{y}) = \begin{cases} \beta^{\mathrm{M,p}}_{a\leftarrow bsi}(\mathbf{y}) & \text{for} \quad i_{au} = i_{as} + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

The class-specific conditional birth rate via dispersal satisfies

$$\lambda^{\mathrm{M,dB}}_{auj,w,bsi}(\mathbf{y}) = \begin{cases} \beta^{\mathrm{M,d}}_{a\overset{w}{\leftarrow} bs}(\mathbf{y}) & \text{for} \quad i_{au} = i_{aw} + 1 \\ 0 & \text{otherwise,} \end{cases} \tag{40}$$

and so the total birth rate via dispersal is

$$\lambda^{\mathrm{M,d}}_{auj,bsi}(\mathbf{y}) = \sum_{w\in\mathcal{S}} x_w \lambda^{\mathrm{M,dB}}_{auj,w,bsi}(\mathbf{y}). \tag{41}$$

The rates for resident individuals are defined similarly, one only needs to replace the number of mutants i_{as} with the number of resident $N_{as} - i_{as}$ for all $as \in \mathcal{C}$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.2 Group state dynamics

The mutant-resident metacommunity dynamics is expressed analogously to the resident metacommunity dynamics (10) as a system of ODEs'

$$\dot{\mathbf{y}} = \mathbf{P}(\mathbf{y})\mathbf{y}, \tag{42}$$

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

$$\pi_{uj,si}(\mathbf{y}) = \pi^{\mathrm{p}}_{uj,si}(\mathbf{y}) + \pi^{\mathrm{d}}_{uj,si}(\mathbf{y}) + \pi^{\mathrm{mc}}_{uj,si}(\mathbf{y}), \tag{43}$$

giving the rate at which a group in state $si \in \Omega$ transitions to state $uj \in \Omega$, and where $\sum_{uj\in\Omega} \pi_{uj,si} = 0$ 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 abiotic 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 (37)-(41) by summing over all physiological states of the parents and their possible progeny, that is, we have

$$\begin{aligned}
\pi^{\mathrm{p}}_{uj,si}(\mathbf{y}) &= \sum_{a,b\in\mathcal{P}} \bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y}) N_{bs} \\
\pi^{\mathrm{d}}_{uj,si}(\mathbf{y}) &= \sum_{a,c\in\mathcal{P}} \sum_{wk\in\Omega} \bar{\lambda}^{\mathrm{d}}_{auj,s,cwk}(\mathbf{y}) N_{cw} y_{wk},
\end{aligned} \tag{44a}$$

for all $au, bs \in \mathcal{C}, j \in \mathcal{I}_u, i \in \mathcal{I}_s$ and $\theta \in \{\mathrm{M}, \mathrm{R}\}$, where

$$\begin{aligned}
\bar{\lambda}^{\mathrm{p}}_{auj,bsi}(\mathbf{y}) &= \lambda^{\mathrm{M,p}}_{auj,bsi}(\mathbf{y}) \frac{i_{bs}}{N_{bs}} + \lambda^{\mathrm{R,p}}_{auj,bsi}(\mathbf{y}) \frac{N_{bs} - i_{bs}}{N_{bs}} \\
\bar{\lambda}^{\mathrm{d}}_{auj,s,cwk}(\mathbf{y}) &= \lambda^{\mathrm{M,d}}_{auj,s,cwk}(\mathbf{y}) \frac{k_{cw}}{N_{cw}} + \lambda^{\mathrm{R,d}}_{auj,s,cwk}(\mathbf{y}) \frac{N_{cw} - k_{cw}}{N_{cw}}
\end{aligned} \tag{44b}$$

with

$$\lambda^{\theta,\mathrm{p}}_{auj,bsi}(\mathbf{y}) = \lambda^{\theta,\mathrm{pB}}_{auj,bsi}(\mathbf{y}) + \lambda^{\theta,\mathrm{T}}_{auj,bsi}(\mathbf{y}) + \lambda^{\theta,\mathrm{D}}_{auj,bsi}(\mathbf{y}) \tag{45}$$

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

$$\begin{aligned}
\dot{\mathbf{n}}_{\mathrm{M}} &= \mathbf{G}_{\mathrm{M}}(\mathbf{y})\mathbf{n}_{\mathrm{M}} \\
\dot{\mathbf{n}}_{\mathrm{R}} &= \mathbf{G}_{\mathrm{R}}(\mathbf{y})\mathbf{n}_{\mathrm{R}},
\end{aligned} \tag{46}$$

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

$$g^{\mathrm{M}}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u} \sum_{i\in\mathcal{I}_s} \Bigg[\lambda^{\mathrm{M,d}}_{auj,bsi}(\mathbf{y}) + \lambda^{\mathrm{M,pB}}_{auj,bsi}(\mathbf{y}) + \lambda^{\mathrm{M,T}}_{auj,bsi}(\mathbf{y}) + \tag{47}$$

$$+ \delta_{a,b} \left(\pi_{uj,si}(\mathbf{y}) - \sum_{c\in\mathcal{P}} \lambda^{\mathrm{M,T}}_{cuj,asi}(\mathbf{y}) - \lambda^{\mathrm{M,D}}_{auj,asi}(\mathbf{y}) \right) \Bigg] y_{\mathrm{M},bsi|\mathrm{M},bs}, \tag{48}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

$$g^{\mathrm{R}}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\Bigg[\lambda^{\mathrm{R,d}}_{auj,bsi}(\mathbf{y}) + \lambda^{\mathrm{R,pB}}_{auj,bsi}(\mathbf{y}) + \lambda^{\mathrm{R,T}}_{auj,bsi}(\mathbf{y}) + \tag{49}$$

$$+\,\delta_{a,b}\left(\pi_{uj,si}(\mathbf{y}) - \sum_{c\in\mathcal{P}}\lambda^{\mathrm{R,T}}_{cuj,asi}(\mathbf{y}) - \lambda^{\mathrm{R,D}}_{auj,asi}(\mathbf{y})\right)\Bigg]y_{\mathrm{R},bsi|\mathrm{R},bs}, \tag{50}$$

giving the rate at which a single resident in class $bs \in \mathcal{C}$ produces residents 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.2.1). Note that similarly to the resident species 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 (46) they need to be coupled with the dynamics of $\mathbf{y}$ in (42).

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 (46)-(49) by using two new vectors. First, with a slight abuse of notation, we use a 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, we use a 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 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 is common in population genetics (e.g., Nagylaki, 1992).

The vector of density dynamics satisfies

$$\dot{\mathbf{n}} = \bar{\mathbf{G}}(\mathbf{y})\mathbf{n}, \tag{51}$$

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 each element

$$\bar{g}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\Bigg[\bar{\lambda}^{\mathrm{d}}_{auj,bsi}(\mathbf{y}) + \bar{\lambda}^{\mathrm{pB}}_{auj,bsi}(\mathbf{y}) + \bar{\lambda}^{\mathrm{T}}_{auj,bsi}(\mathbf{y}) + \tag{52}$$

$$+\,\delta_{a,b}\left(\pi_{uj,si}(\mathbf{y}) - \sum_{c\in\mathcal{P}}\bar{\lambda}^{\mathrm{T}}_{cuj,asi}(\mathbf{y}) - \bar{\lambda}^{\mathrm{D}}_{auj,asi}(\mathbf{y})\right)\Bigg]y_{si|s} \tag{53}$$

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.2.2). The vector of within-class mutant frequency dynamics satisfies

$$\dot{\mathbf{p}} = \mathbf{F}(\mathbf{y})\mathbf{p}, \tag{54}$$

where

$$\mathbf{F} = \mathbf{F}^{\mathrm{sel}}(\mathbf{y}) + \bar{\mathbf{F}}(\mathbf{y}). \tag{55}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

$$f^{\text{sel}}_{au,bs}(\mathbf{y}) = \frac{n_{bs}}{n_{au}}\left[g^{\text{M}}_{au,bs}(\mathbf{y}) - \bar{g}_{au,bs}(\mathbf{y})\right], \tag{56}$$

and

$$\bar{f}_{au,bs}(\mathbf{y}) = \begin{cases} \frac{n_{bs}}{n_{au}}\bar{g}_{au,bs}(\mathbf{y}) & \text{for } bs \neq au \\ -\sum_{cw\neq au}\frac{n_{cs}}{n_{au}}\bar{g}_{au,cw}(\mathbf{y}) & \text{for } bs = au. \end{cases} \tag{57}$$

Because the rate in (56) 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 (57) depends only on the average phenotype it captures effects on allele frequency change due to individuals in different classes having different average growth rates. We have added a bar analogously to (51) 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 $\mathbf{p}$, (51) and/or (57) need to be coupled with the dynamics of $\mathbf{y}$ in (42).

4.5 Average mutant frequency

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

$$p = \frac{n_{\text{M}}}{n}\ (= \frac{\mathbf{1}}{n}\cdot\mathbf{n}_{\text{M}}), \tag{58}$$

where $n_{\text{M}} = \sum_{as\in\mathcal{S}} n_{\text{M},as}$ 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 (58) is obtained by differentiation and by using Sections 4.2-4.4, and it can be written as

$$\begin{aligned} \dot{p} &= \frac{\mathbf{1}}{n}\cdot\left[\mathbf{G}_{\text{M}}(\mathbf{y})\mathbf{n}_{\text{M}} - \bar{\mathbf{G}}(\mathbf{y})\mathbf{n}p\right] \\ &= \frac{1}{n}\sum_{au,bs\in\mathcal{C}}\left[g^{\text{M}}_{au,bs}(\mathbf{y})p_{\text{M},bs} - \bar{g}_{au,bs}(\mathbf{y})p\right]n_{bs}. \end{aligned} \tag{59}$$

In order to study its dynamics it must be coupled with the dynamics of $\mathbf{y}$ in (42).

5 Mutant-resident dynamics for similar phenotypes

We now study mutant-resident dynamics for closely similar phenotypes, with a particular focus on the change in the arithmetic mean allele frequency (58). To that end, we set $\boldsymbol{z}_{\text{M}} = \boldsymbol{z}_{\text{R}} + \delta\boldsymbol{\eta}$ where the vector $\boldsymbol{\eta} = (\eta_{as})_{as\in\mathcal{C}} \in \mathbb{R}^{k_\mathcal{C}}$ gives the direction of the deviation (where $\boldsymbol{z}_{\text{R}} = \boldsymbol{z} \in \mathcal{Z}$) and δ 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 (42), (51) and (54), respectively. To facilitate the Taylor expansion, we use the following consistency relation relating mutant-resident i-processes

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Section 3.1) to the resident i-processes (Section 4.1):

$$\lambda^{\theta,\xi}_{auj,bsi}(\mathbf{y})\Big|_{\delta=0} = \lambda^{\xi}_{au,bs}(\mathbf{x}), \quad \text{for all } au, bs \in \mathcal{C},\ i \in \mathcal{I}_s, j \in \mathcal{I}_u \text{ and } \theta \in \{\mathrm{M}, \mathrm{R}\}, \tag{60}$$

and for $\xi \in \{\mathrm{d}, \mathrm{pB}, \mathrm{T}, \mathrm{D}\}$, and we will also use $\lambda^{\theta,\mathrm{dB}}_{auj,w,bsi}(\mathbf{y})\Big|_{\delta=0} = \lambda^{\mathrm{dB}}_{au,w,bs}(\mathbf{x})$. These consistency relations say that whenever mutants and residents express identical phenotypes, the i-processes are independent of the genetic structure of the metacommunity. From (60), it directly follows that

$$\pi_{uj,si}(\mathbf{y})\Big|_{\delta=0} = q_{u,s}(\mathbf{x}), \quad \text{for all } u, s \in \mathcal{S},\ i \in \mathcal{I}_s, j \in \mathcal{I}_u, \tag{61}$$

and

$$g^{\theta}_{au,bs}(\mathbf{y})\Big|_{\delta=0} = h_{au,bs}(\mathbf{x}), \quad \text{for } \theta \in \{\mathrm{M}, \mathrm{R}\} \text{ and for all } au, bs \in \mathcal{C}, \tag{62}$$

and consequently the growth-rates in (51) and (54) satisfy

$$\begin{aligned} \bar{g}_{au,bs}(\mathbf{y})\Big|_{\delta=0} &= h_{au,bs}(\mathbf{x}) \\ f_{au,bs}(\mathbf{y})\Big|_{\delta=0} &= f^{\circ}_{au,bs}(\mathbf{x}) = \begin{cases} \frac{n_{bs}}{n_{au}} h_{au,bs}(\mathbf{x}) & \text{for } bs \neq au \\ -\sum_{cw \neq au} \frac{n_{cs}}{n_{au}} h_{au,bs}(\mathbf{x}) & \text{for } bs = au \end{cases} \end{aligned} \tag{63}$$

for all $au, bs \in \mathcal{C}$.

It follows from the above relations that the dynamics for $\mathbf{y}, \mathbf{n}$ and $\mathbf{p}$ for small δ can be written as

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{Q}(\mathbf{x})\mathbf{x} + O(\delta) \\ \dot{\mathbf{n}} &= \mathbf{H}(\mathbf{x})\mathbf{n} + O(\delta) \\ \dot{\mathbf{p}} &= \mathbf{F}^{\circ}(\mathbf{x})\mathbf{p} + O(\delta), \end{aligned} \tag{64}$$

where we used (61) and (63) and where $\mathbf{F}^{\circ}(\mathbf{x}) = (f^{\circ}_{au,bs}(\mathbf{x}))_{au,bs \in \mathcal{C}}$ is the relative fitness matrix for mutants (when $\delta = 0$) with elements as defined in (63). 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 (64) up to order $O(1)$ follows from (42), (61) and $x_s = \sum_{i \in \mathcal{I}_s} y_{si}$ (the dependency on $\mathbf{y}$ in (64) comes only via the terms of order $O(\delta)$). The dynamics of the arithmetic mean mutant frequency p in (58) for small δ is obtained by Taylor expansion and by using (62)-(63), in which case (59) reduces to

$$\begin{aligned} \dot{p} &= \frac{\mathbf{1}}{n} \cdot \mathbf{H}(\mathbf{x}) \Big[\mathbf{n}_{\mathrm{M}} - \mathbf{n}p\Big] + O(\delta) \\ &= \frac{1}{n} \sum_{au,bs \in \mathcal{C}} h_{au,bs}(\mathbf{x}) \big[p_{\mathrm{M},bs} - p\big] n_{bs} + O(\delta), \end{aligned} \tag{65}$$

which must be coupled with the dynamics for $\mathbf{p}$ and $\mathbf{x}$ in (64) and where the dependency on $\mathbf{y}$ comes only via $O(\delta)$.

The remaining 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 in a standard form (Fenichel, 1979; Kuehn, 2015; Wechselberger, 2020), which is relatively easy to analyze, is a system of ODEs where a model parameter tunes the rate at which the various dynamical variables operate: for small parameter values δ some variables are fast variables in that their Taylor expansion about $\delta = 0$ is dominated by $O(1)$, while others are slow variables and their Taylor expansion about $\delta = 0$ is dominated by $O(\delta)$. We see from (64), however, that for small δ the dynamics of the vectors $\mathbf{x}, \mathbf{n}$ and $\mathbf{p}$, as well as the arithmetic mean

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mutant frequency p in (65), are all dominated by the terms $O(1)$ thus all having a phase of fast dynamics. As a consequence, the mutant-resident system is not readily in the desired (standard) fast-slow form.

In order to construct an appropriate fast-slow system, 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 can be used, on a suitably chosen slow timescale, as a proxy for the arithmetic mean mutant frequency (Section 5.2). Finally, we study the slow dynamics of this slow variable and calculate all the necessary fast variables (Section 5.3) that define the complete 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 (64), which gives

$$\dot{\mathbf{p}} = \mathbf{F}^{\circ}(\mathbf{x})\mathbf{p}. \tag{66}$$

Because the rows of $\mathbf{F}^{\circ}$ sum to 0 it is well-known that the steady state satisfies

$$\hat{\mathbf{p}} = \mathbf{1}p_{\mathrm{c}} \tag{67}$$

for some constant $p_{\mathrm{c}} \in \mathbb{R}$ (e.g., Lessard and Soares, 2018). This implies 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 (66) thus contains infinite number of steady states (equilibria) and the biologically meaningful values lie on a line $[0, 1]$. This has two important consequences. First, any linear combination of within-class frequencies, at the steady state (67), is equal to p_{c}. More specifically, every arbitrarily weighted average mutant frequency

$$p_{\boldsymbol{\phi}} = \boldsymbol{\phi} \cdot \mathbf{n}_{\mathrm{M}} \ (= \sum_{as \in \mathcal{C}} \phi_{as} n_{as} p_{\mathrm{M},as}), \tag{68}$$

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

$$\hat{p}_{\boldsymbol{\phi}} = p_{\mathrm{c}} \tag{69}$$

at the steady state where $\hat{p}_{\mathrm{M},as} = p_{\mathrm{c}}$ for all $as \in \mathcal{C}$. This is in particular true also for the arithmetic mean frequency p in (58) where $\boldsymbol{\phi} = \mathbf{1}/n$. This implies that 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 (67), 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). Finally, we want to note that so far we have analyzed the fast dynamics of $\mathbf{p}$ and $p_{\boldsymbol{\phi}}$ dominated by $O(1)$ only; in later sections we will see that these variables have also a phase of slow dynamics and this will be analyzed in later sections (see in particular Section 6).

5.2 Reproductive value weighted mutant frequency

In the previous Section 5.1 we found that for $\delta = 0$ and for any weights $\boldsymbol{\phi}$, the average mutant frequency (68) satisfies $p_{\boldsymbol{\phi}}(t) \to p_{\mathrm{c}}$ as $t \to \infty$ with the exact value of p_{c} depending on the initial condition $(\mathbf{x}(t_0), \mathbf{p}(t_0))$. If one were then to find weights $\boldsymbol{\phi}$ for which $p_{\boldsymbol{\phi}}(t) = p_{\mathrm{c}}$ for all $t \geq t_0$,

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 standard fast-slow system). We can achieve this by setting $\boldsymbol{\phi} = \mathbf{v}^{\mathbf{n}}$ where the weights $\mathbf{v}^{\mathbf{n}}$ for non-zero δ 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 reproductive value analogue to the one defined for the resident population in Section 3. Such average individual reproductive values satisfy

$$-\dot{\mathbf{v}}^{\mathbf{n}} = \mathbf{v}^{\mathbf{n}}\bar{\mathbf{G}}(\mathbf{y}), \tag{70}$$

for any δ, with the normalization $\mathbf{v}^{\mathbf{n}} \cdot \mathbf{n} = 1$ where $\mathbf{n}$ is given by (51). By defining a reproductive values weighted mutant frequency (see (68)) as

$$p_{\mathbf{v}} = \mathbf{v}^{\mathbf{n}} \cdot \mathbf{n}_{\mathrm{M}} = \sum_{as\in\mathcal{C}} v^{n}_{as} n_{as} p_{\mathrm{M},as}, \tag{71}$$

its dynamics is obtained by differentiation and satisfies

$$\begin{aligned} \dot{p}_{\mathbf{v}} &= \mathbf{v}^{\mathbf{n}} \cdot \Big[\mathbf{G}_{\mathrm{M}}(\mathbf{y}) - \bar{\mathbf{G}}(\mathbf{y})\Big]\mathbf{n}_{\mathrm{M}} \\ &= \sum_{au,bs\in\mathcal{C}} v^{n}_{au}\Big(g^{\mathrm{M}}_{au,bs}(\mathbf{y}) - \bar{g}_{au,bs}(\mathbf{y})\Big)p_{\mathrm{M},bs}n_{bs}, \end{aligned} \tag{72}$$

for any δ. Indeed, because $\mathbf{G}_{\mathrm{M}}(\mathbf{y}) - \bar{\mathbf{G}}(\mathbf{y}) = \mathbf{0}$ for $\delta = 0$ (62)-(63), the Taylor expansion of (72) 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 since (72) is valid for arbitrary δ and thus arbitrary strength of selection, it must be coupled with the dynamics of $\mathbf{y}$ in (42).

5.3 Slow dynamics of the weighted mutant frequency

We here study the slow dynamics of the weighted mutant frequency $p_{\mathbf{v}}$ in (71)-(72), with a specific aim to identify all the dynamical variables that it depends on in order to construct a complete fast-slow mutant-resident system. To that end, we Taylor expand (72) 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

$$\frac{\mathrm{d}}{\mathrm{d}\tau}p_{\mathbf{v}} = \frac{\mathrm{d}}{\mathrm{d}\delta}\left[\sum_{au,bs\in\mathcal{C}} v^{n}_{au}\Big(g^{\mathrm{M}}_{au,bs}(\mathbf{y}) - \bar{g}_{au,bs}(\mathbf{y})\Big)p_{\mathrm{M},bs}n_{bs}\right]_{\delta=0}, \tag{73}$$

giving the rate at which the reproductive valued weighted frequency changes in a pure slow time τ where $\delta = 0$. Note that in the fast timescale t, the above expression is nothing else than the $O(\delta)$-term of the Taylor expansion.

Next, we take the derivative in (73) 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 individual fitness (16) as

$$h_{au,bs}(\mathbf{x}) = h_{au,bs}(z_{\mathrm{foc},bs}, \boldsymbol{z}_{\mathrm{loc}}, s, \boldsymbol{z}_{\mathrm{pop}}, \mathbf{x}), \tag{74}$$

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)

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

but excluding the focal individual, and $\boldsymbol{z}_{\text{pop}}$ 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 all individuals in the resident metacommunity express the resident phenotype $\boldsymbol{z} \in \mathcal{Z}$ and that these are just labels to distinguish the spatial relationship of individuals to the focal individual according to which the different partial derivatives will be taken in Section 5.3.1. Likewise, we write $g_{au,bs}(\mathbf{y}) = g_{au,bs}(z_{\text{M},bs}, \boldsymbol{z}_{\text{M,loc}}, \boldsymbol{z}_{\text{R,loc}}, si, \boldsymbol{z}_{\text{M,pop}}, \boldsymbol{z}_{\text{R,pop}}, \mathbf{y})$ for a focal mutant and $\bar{g}_{au,bs}(\mathbf{y}) = g_{au,bs}(\boldsymbol{z}_{\text{M,loc}}, \boldsymbol{z}_{\text{R,loc}}, si, \boldsymbol{z}_{\text{M,pop}}, \boldsymbol{z}_{\text{R,pop}}, \mathbf{y})$ for an average individual where the notation and interpretation of $\boldsymbol{z}_{\text{M}_{\text{foc}}}, \boldsymbol{z}_{\text{M}_{\text{loc}}}$ and $\boldsymbol{z}_{\text{M}_{\text{pop}}}$ (and $\boldsymbol{z}_{\text{R}_{\text{foc}}}, \boldsymbol{z}_{\text{R}_{\text{loc}}}$ and $\boldsymbol{z}_{\text{R}_{\text{pop}}}$) for the mutant-resident metacommunity is similar.

5.3.1 Properties of fitness

To take the derivative in (73) we apply a property that relates, for small δ, mutant-resident fitnesses (47)-(49) to the resident fitness (16). This property follows from the so-called generalized law of mass action in situations where weak selection results from small differences in phenotype (Diekmann et al., 2001; Meszéna et al., 2005; Dercole, 2016, for discussion on the different forms of weak selection see Wild and Traulsen, 2007; Lehmann and Rousset, 2014). The generalized law of mass action says that the first-order perturbation of the mutant phenotype of all mutant individuals in a spatially well-mixed mutant-resident species, is equal to the first-order perturbation of the phenotype of all individuals in a purely resident species when multiplied by the frequency of mutant individuals. This is computationally convenient because instead of considering full distributions of individuals only population means need to be calculated. 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 $\mathbf{G}_{\text{M}}(\mathbf{y})$ is related to the resident fitness matrix $\mathbf{H}(\mathbf{x})$ in the following way. First,

$$\frac{\partial g^{\text{M}}_{au,bs}(\mathbf{y})}{\partial z_{\text{M}_{\text{foc}},bs}}\Big|_{\delta=0} = \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\text{foc},bs}} \quad \text{for all } au, bs \in \mathcal{C} \tag{75}$$

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

$$\frac{\partial g^{\text{M}}_{au,bs}(\mathbf{y})}{\partial z_{\text{M}_{\text{loc}},cs}}\Big|_{\delta=0} = \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\text{loc},cs}} p_{\text{M},cs|\text{M},bs} \quad \text{for all } au, bs \in \mathcal{C}, c \in \mathcal{P} \tag{76}$$

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

$$\frac{\partial g^{\text{M}}_{au,bs}(\mathbf{y})}{\partial z_{\text{M}_{\text{pop}},cw}}\Big|_{\delta=0} = \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\text{pop},cw}} p_{\text{M},cw} \quad \text{for all } au, bs, cw \in \mathcal{C}. \tag{77}$$

In (76) we introduced the *conditional* within-group mutant frequency $\mathbf{p}_{\text{M}|\text{M}} = (p_{\text{M},as|\text{M},bs})_{abs\in\mathcal{B}} \in [0,1]^{k_{\mathcal{B}}}$ defined as

$$p_{\text{M},as|\text{M},bs} = \frac{p_{\text{MM},abs}}{p_{\text{M},bs}}, \quad \text{for all } a, b \in \mathcal{P}, s \in \mathcal{S}, \tag{78}$$

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

$$p_{\text{MM},abs} = \frac{n_{\text{MM},abs}}{n_{abs}}, \quad \text{for all } abs \in \mathcal{B}. \tag{79}$$

Here, $n_{\text{MM},abs} = \sum_{i\in\mathcal{I}} i_{as} i_{bs} y_{si}$ 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 $abs \in \mathcal{B}$, that is, the total density of pairs inde-

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pendent 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 (75), 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. In other words, the 'frequency' of a mutant allele in a single mutant individual is 1. The derivative on the right-hand-side measures (when multiplied by δ) the additional offspring the focal individual produces per unit of time due to the focal individual itself expressing the mutation ('direct fitness effect'). In (76), 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 (when multiplied by δ) 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 (77) 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 (75)-(77) 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

$$\begin{aligned}
\frac{\partial \bar{g}_{au,bs}(\mathbf{y})}{\partial z_{\mathrm{M}_{\mathrm{foc}},bs}}\Big|_{\delta=0} &= \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} p_{\mathrm{M},bs} \\
\frac{\partial \bar{g}_{au,bs}(\mathbf{y})}{\partial z_{\mathrm{M}_{\mathrm{loc}},cs}}\Big|_{\delta=0} &= \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} p_{\mathrm{M},cs} \\
\frac{\partial \bar{g}_{au,bs}(\mathbf{y})}{\partial z_{\mathrm{M}_{\mathrm{pop}},cw}}\Big|_{\delta=0} &= \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{pop},cw}} p_{\mathrm{M},cw} \quad \text{for all } au, bs, cw \in \mathcal{C},
\end{aligned} \tag{80}$$

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 (80) 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 (73) and partitioning it according to the different individuals as discussed above, we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau} p_{\mathbf{v}} = \sum_{au,bs \in \mathcal{C}} v_{au}^{n} \Big[&\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} \eta_{bs} p_{\mathrm{M},bs}(1 - p_{\mathrm{M},bs}) + \\
&+ \sum_{cs \in \mathcal{C}} \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} \eta_{cs} (p_{\mathrm{MM},bcs} - p_{\mathrm{M},bs} p_{\mathrm{M},cs}) \Big] n_{bs},
\end{aligned} \tag{81}$$

where we used the definition for directional derivatives, the properties (75)-(77), (80) and (78). We can see from (81) that the dynamics of the slow weighted mutant frequency $p_{\mathbf{v}}$ depends on the dynamics of $\mathbf{x}, \mathbf{n}, \mathbf{p}, \mathbf{v}^{\mathbf{n}}$ and $\mathbf{p}_{\mathrm{MM}}$. This is convenient because instead of analyzing the dynamics of the large system for $\mathbf{y}$ it is enough to study the simpler dynamics of $\mathbf{x}, \mathbf{n}, \mathbf{p}, \mathbf{v}^{\mathbf{n}}$ and $\mathbf{p}_{\mathrm{MM}}$ (and as will be shown below). Because we have already derived dynamical equations for $\mathbf{x}, \mathbf{n}$ and $\mathbf{p}$, and in (64) showed that they are dominated by $O(1)$ and so is $\mathbf{v}^{\mathbf{n}}$ because under the consistency relation (62) its dynamics reduces to (17), the next task is to find the equation for a vector of mutant-pair frequencies $\mathbf{p}_{\mathrm{MM}}$ (Section 5.3.3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5.3.3 Mutant-pair frequencies

In order to see whether $\mathbf{p}_{\mathrm{MM}}$ defined in (79) 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

$$\begin{aligned}\dot{p}_{\mathrm{MM},abu} = &\sum_{s\neq u\in\mathcal{S}}\sum_{d\in\mathcal{P}}\frac{n_{ds}}{n_{abu}}\left[\alpha^{\mathrm{CA}}_{abu,ds}+\alpha^{\mathrm{IM}}_{abu,ds}\right]p_{\mathrm{M},ds}+\sum_{s\neq u\in\mathcal{S}}\sum_{c,d\in\mathcal{P}}\frac{n_{cds}}{n_{abu}}\alpha^{\mathrm{DP}}_{abu,cds}p_{\mathrm{MM},cds}-\\ &-p_{\mathrm{MM},abu}\frac{N_{abu}}{n_{abu}}\sum_{s\neq u\in\mathcal{S}}q_{u,s}x_s \quad \text{for all } abu\in\mathcal{B},\end{aligned} \tag{82}$$

which can be written in vector form as

$$\dot{\mathbf{p}}_{\mathrm{MM}} = \mathbf{E}(\mathbf{p},\mathbf{p}_{\mathrm{MM}},\mathbf{x}). \tag{83}$$

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 dominated by terms of order $O(1)$ and is hence a fast variable (it is to be seen whether it changes also in slow time, Section 6).

To characterise for $\delta = 0$ the steady state of $\mathbf{p}_{\mathrm{MM}}$ we focus for convenience on the equivalent discrete time jump-process dynamics, which is given by

$$\mathbf{p}'_{\mathrm{MM}} = \left[\mathbf{C}(\mathbf{x})+\mathbf{M}(\mathbf{p},\mathbf{x})\right]\mathbf{p}+\mathbf{D}(\mathbf{x})\mathbf{p}_{\mathrm{MM}}, \tag{84a}$$

where $\mathbf{C}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ are as in (30), 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

$$M_{abu,ds,ew}(\mathbf{x}) = \frac{1}{N_{abu}}\frac{\alpha^{\mathrm{IM}}_{abu,ds,ew}(\mathbf{x})N_{ds}x_sN_{ew}x_w}{\sum_{w\neq u\in\mathcal{S}}q_{u,w}(\mathbf{x})x_w} \tag{84b}$$

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 (84)). For all pairs $abu\in\mathcal{B}$ we have

$$\sum_{ds\in\mathcal{C}}C_{abu,ds}(\mathbf{x})+\sum_{ds\in\mathcal{C}}\sum_{ew\in\mathcal{C}}M_{abu,ds,ew}(\mathbf{x})+\sum_{cds\in\mathcal{B}}D_{abu,cds}(\mathbf{x}) = 1, \tag{85}$$

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 $\hat{\mathbf{p}}_{\mathrm{MM}}$ for $\delta = 0$ 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 (84) evaluated at $\hat{\mathbf{p}}$ and $\hat{\mathbf{x}}$, and we get

$$\begin{aligned}\hat{\mathbf{p}}_{\mathrm{MM}} &= (\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{C}(\hat{\mathbf{x}})\hat{\mathbf{p}}+(\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{M}(\hat{\mathbf{p}},\hat{\mathbf{x}})\hat{\mathbf{p}}\\ &= (\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{C}(\hat{\mathbf{x}})\mathbf{1}p_{\mathrm{c}}+(\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}}))^{-1}(\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}})-\mathbf{C}(\hat{\mathbf{x}}))\mathbf{1}p_{\mathrm{c}}^2\\ &= (\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{c}(\hat{\mathbf{x}})p_{\mathrm{c}}+(\mathbf{1}-(\mathbf{I}-\mathbf{D}(\hat{\mathbf{x}}))^{-1}\mathbf{c}(\hat{\mathbf{x}}))p_{\mathrm{c}}^2\\ &= \hat{\mathbf{r}}p_{\mathrm{c}}+(\mathbf{1}-\hat{\mathbf{r}})p_{\mathrm{c}}^2\end{aligned} \tag{86}$$

where we used (67), (85) and (35), 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 (82)-(83)). 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 pair 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in (86) is obtained only as a limit when $t \to \infty$.

6 Mutant-resident dynamics as a fast-slow system

We are now ready to analyse the fast-slow ODE system where mutant and resident individuals have closely similar phenotypes (δ is small). Owing to Section 5, the complete fast-slow system for the mutant-resident dynamics is given in fast time t as

$$\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{Q}(\mathbf{x})\mathbf{x} + O(\delta) \\
\dot{\mathbf{n}} &= \mathbf{H}(\mathbf{x})\mathbf{n} + O(\delta) \\
-\dot{\mathbf{v}}^{\mathbf{n}} &= \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}) + O(\delta) \\
\dot{\mathbf{p}} &= \mathbf{F}^{\circ}(\mathbf{x})\mathbf{p} + O(\delta) \\
\dot{\mathbf{p}}_{\mathrm{MM}} &= \mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}) + O(\delta) \\
\dot{p}_{\mathbf{v}} &= \delta \sum_{au,bs \in \mathcal{C}} v_{au}^{n} \Big[\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} \eta_{bs} p_{\mathrm{M},bs}(1 - p_{\mathrm{M},bs}) + \\
&\qquad + \sum_{cs \in \mathcal{C}} \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} \eta_{cs}(p_{\mathrm{MM},bcs} - p_{\mathrm{M},bs} p_{\mathrm{M},cs}) \Big]_{\delta=0} n_{bs} + O(\delta^2),
\end{aligned} \tag{87}$$

which, on using $\mathrm{d}/\,\mathrm{d}t = \delta\,\mathrm{d}/\,\mathrm{d}\tau$ can be equivalently written in slow time $\tau = \delta t$ as

$$\begin{aligned}
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{x} &= \mathbf{Q}(\mathbf{x})\mathbf{x} + O(\delta) \\
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{n} &= \mathbf{H}(\mathbf{x})\mathbf{n} + O(\delta) \\
-\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{v}^{\mathbf{n}} &= \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}) + O(\delta) \\
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{p} &= \mathbf{F}^{\circ}(\mathbf{x})\mathbf{p} + O(\delta) \\
\delta \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{p}_{\mathrm{MM}} &= \mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}) + O(\delta) \\
\frac{\mathrm{d}}{\mathrm{d}\tau} p_{\mathbf{v}} &= \sum_{au,bs \in \mathcal{C}} v_{au}^{n} \Big[\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} \eta_{bs} p_{\mathrm{M},bs}(1 - p_{\mathrm{M},bs}) + \\
&\qquad + \sum_{cs \in \mathcal{C}} \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} \eta_{cs}(p_{\mathrm{MM},bcs} - p_{\mathrm{M},bs} p_{\mathrm{M},cs}) \Big]_{\delta=0} n_{bs} + O(\delta).
\end{aligned} \tag{88}$$

The two systems (87) and (88) are equivalent, the only difference is the notation. Next we study the fast and slow subsystems of (87) and (88) 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 (87), which yields

$$\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{Q}(\mathbf{x})\mathbf{x} \\
\dot{\mathbf{n}} &= \mathbf{H}(\mathbf{x})\mathbf{n} \\
-\dot{\mathbf{v}}^{\mathbf{n}} &= \mathbf{v}^{\mathbf{n}}\mathbf{H}(\mathbf{x}) \\
\dot{\mathbf{p}} &= \mathbf{F}^{\circ}(\mathbf{x})\mathbf{p} \\
\dot{\mathbf{p}}_{\mathrm{MM}} &= \mathbf{E}(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}, \mathbf{x}) \\
\dot{p}_{\mathbf{v}} &= 0.
\end{aligned} \tag{89}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 purely 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), the variable $p_{\mathbf{v}}$ is a constant in the fast subsystem (89) and we can identify p_{c} with $p_{\mathbf{v}}$. 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

$$\begin{aligned}\mathcal{M}_0 = \{(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}}, p_{\mathbf{v}}) \in \mathbb{R}^{k_{\mathcal{S}}} \times \mathbb{R}^{k_{\mathcal{C}}} \times \mathbb{R}^{k_{\mathcal{C}}} \times \mathbb{R}^{k_{\mathcal{C}}} \times \mathbb{R}^{k_{\mathcal{B}}} \times \mathbb{R} \ : \\ \mathbf{x} = \hat{\mathbf{x}},\ \mathbf{n} = \hat{\mathbf{n}}, \mathbf{v}^{\mathbf{n}} = \hat{\mathbf{v}}^{\mathbf{n}},\ \mathbf{p} = \mathbf{1}p_{\mathbf{v}},\ \mathbf{p}_{\mathrm{MM}} = \hat{\mathbf{r}}p_{\mathbf{v}} + (\mathbf{1} - \hat{\mathbf{r}})p_{\mathbf{v}}^2,\ p_{\mathbf{v}} \in [0, 1]\},\end{aligned} \tag{90}$$

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 (89) and is a curve consisting of an infinite number of steady states (equilibria) parametrized by $p_{\mathbf{v}} \in [0, 1]$ (see Figure 2, panels (a)-(c) for a graphical representation in). 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 fast 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 and can be seen as the state space for the pure slow dynamics where $\delta = 0$ in (88) (studied in detail in Section 6.2). 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 (87)-(88) on and near the slow-manifold is equivalent to the dynamics of $p_{\mathbf{v}}$ on and near $\mathcal{M}_0$ in the slow subsystem (Section 6.2).

6.2 Slow subsystem on the critical manifold

We obtain the slow subsystem by setting $\delta = 0$ in (88). However, motivated by Section 6.1, we are in fact not interested in every possible slow subsystem of (88), only the one associated to $\hat{\mathbf{x}}$ and the critical manifold $\mathcal{M}_0$ (90). 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 (90) and the dynamics of the weighted mutant frequency $p_{\mathbf{v}}$ restricted to this manifold,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}p_{\mathbf{v}} &= \sum_{au,bs \in \mathcal{C}} \Bigg[v_{au}^{n} \Big[\frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{foc},bs}} \eta_{bs} p_{\mathrm{M},bs}(1 - p_{\mathrm{M},bs}) + \\ &\qquad + \sum_{cs \in \mathcal{C}} \frac{\partial h_{au,bs}(\mathbf{x})}{\partial z_{\mathrm{loc},cs}} \eta_{cs}(p_{\mathrm{MM},bcs} - p_{\mathrm{M},bs}p_{\mathrm{M},cs}) \Big] n_{bs} \Bigg]_{\mathcal{M}_0} \\ &= p_{\mathbf{v}}(1 - p_{\mathbf{v}}) S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}),\end{aligned} \tag{91a}$$

where

$$S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) = \sum_{au,bs \in \mathcal{C}} \hat{v}_{au}^{\mathrm{n}} \Big[\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{foc},bs}} \eta_{bs} + \sum_{cs \in \mathcal{C}} \frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{loc},cs}} \hat{r}_{bcs}\ \eta_{cs} \Big] \hat{n}_{bs} \tag{91b}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is evaluated at the critical manifold $\mathcal{M}_0$. To obtain equation (91) we have taken the limits $\delta \to 0$ and $t \to \infty$ and recall that throughout we have assumed that the total population size is infinity. Equation (91) can thus be seen as the first moment of change in the reproductive value weighted allele frequency of the diffusion approximation for slow subsystems (Ethier and Nagylaki, 1980, 1988).

6.3 Solution of the slow subsystem

Given that the dynamics of the weighted mutant frequency $p_{\mathbf{v}}$ can be solved from a single ODE given in (91), we can express this in the context of an initial value problem as follows. First, recall that the fast subsystem (89) 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 let

$$(\mathbf{x}(\tau_0), \mathbf{n}(\tau_0), \mathbf{v}^{\mathbf{n}}(\tau_0)) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}) \tag{92}$$

and

$$(\mathbf{p}(\tau_0), \mathbf{p}_{\mathrm{MM}}(\tau_0)) = (\mathbf{1}p_{\mathbf{v}}(\tau_0), \hat{\mathbf{r}}p_{\mathbf{v}}(\tau_0) + (1 - \hat{\mathbf{r}})p_{\mathbf{v}}(\tau_0)^2). \tag{93}$$

Note that the dependency of (92)-(93) on $(\mathbf{x}(t_0), \mathbf{p}(t_0)))$ comes only through (93). Then, in slow time τ, the weighted mutant frequency $p_{\mathbf{v}}(\tau)$ changes according to (91) for all $\tau \geq \tau_0$, and for the initial condition (92)-(93) its solution is

$$p_{\mathbf{v}}(\tau \mid \mathbf{x}(t_0), \mathbf{p}(t_0)) = \frac{1}{1 + \frac{1-p_{\mathbf{v}}(\tau_0)}{p_{\mathbf{v}}(\tau_0)} \exp[-\tau S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]}, \tag{94}$$

where $S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$ is given in (91b) and where we have indicated that the initial condition $p_{\mathbf{v}}(\tau_0) = p_{\mathrm{c}}$ depends on $(\mathbf{x}(t_0), \mathbf{p}(t_0))$.

It is important to note, that 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 (94). Because $\mathbf{p}$ and $\mathbf{p}_{\mathrm{MM}}$ change also in the fast subsystem (Section 6.1), they are both fast *and* slow variables. Moreover, because every average mutant frequency $p_{\boldsymbol{\phi}}$ in (68) in the fast subsystem converges to p_{c} as $t \to \infty$ (Section 5.1), we will also define

$$p_{\boldsymbol{\phi}}(\tau_0) = p_{\mathbf{v}}(\tau_0). \tag{95}$$

Then, $p_{\boldsymbol{\phi}}(\tau) = p_{\mathbf{v}}(\tau)$ for all $\tau \geq \tau_0$, and the solution of $p_{\boldsymbol{\phi}}$ and thus also the solution of the arithmetic mean mutant frequency p in (58), is identical to the solution (94). 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.

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* (10)*. 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*

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 arithmetic mean mutant frequency p can be approximated as

$$\dot{p} = p(1-p)\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) + O(\delta^2) \tag{96}$$

with solution

$$p(t) = \frac{1}{1+\frac{1-p_c}{p_c}\exp[-t\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]} + O(\delta), \tag{97}$$

where

$$S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) = \sum_{au,bs\in\mathcal{C}} \hat{v}^n_{au}\Big[\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{foc,bs}}\eta_{bs} + \sum_{cs\in\mathcal{C}}\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{loc,cs}}\hat{r}_{bcs}\ \eta_{cs}\Big]\hat{n}_{bs}. \tag{98}$$

The constant p_c in (97) 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 (15). 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).

7.1 Proof of the principle

Here we give a poof of the "invasion implies substitution"-principle stated above. 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 can study the fast-slow dynamics of five fast variables $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}, \mathbf{p}, \mathbf{p}_{\mathrm{MM}})$ dominated by $O(1)$ and a single slow variable $p_{\mathbf{v}}$ dominated by $O(\delta)$. 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 $\mathcal{M}_0$ at which $(\mathbf{x}, \mathbf{n}, \mathbf{v}^{\mathbf{n}}) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}})$ 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 $p_{\mathbf{v}} \in [0,1]$ which is a constant in the fast subsystem (and hence is on $\mathcal{M}_0$ by definition). 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 chosen sufficiently close to $\hat{\mathbf{x}}$.

In the slow subsystem (Section 6.2), we found that the slow variable $p_{\mathbf{v}}$ changes according to an ODE given in (91) and when coupled with an initial condition its solution is given by (94). Consequently, 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 and are thus purely fast variables, and the two moments of allelic states $(\mathbf{p}, \mathbf{p}_{\mathrm{MM}}) = (\mathbf{1}p_{\mathbf{v}}, \hat{\mathbf{r}}p_{\mathbf{v}} + (\mathbf{1}-\hat{\mathbf{r}})p_{\mathbf{v}}^2)$ change due to $p_{\mathbf{v}}$ and are hence both fast and slow variables. 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), which proves the "invasion implies substitution"-principle for small but nonzero δ. To this end, we proceed in two steps. In **step 1** we apply '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 (87)-(88). Property **(i)** says that for any point on $\mathcal{M}_0$ there

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

$$(\mathbf{x}(\tau_0), \mathbf{n}(\tau_0), \mathbf{v}^{\mathbf{n}}(\tau_0))|_{\mathcal{M}_\delta} = (\mathbf{x}(\tau_0), \mathbf{n}(\tau_0), \mathbf{v}^{\mathbf{n}}(\tau_0))|_{\mathcal{M}_0} + \mathbf{O}(\delta) = (\hat{\mathbf{x}}, \hat{\mathbf{n}}, \hat{\mathbf{v}}^{\mathbf{n}}) + \mathbf{O}(\delta) \tag{99a}$$

and

$$(\mathbf{p}(\tau_0), \mathbf{p}_{\mathrm{MM}}(\tau_0))|_{\mathcal{M}_\delta} = (\mathbf{1}p_{\mathbf{v}}(\tau_0), \hat{\mathbf{r}}p_{\mathbf{v}}(\tau_0) + (\mathbf{1} - \hat{\mathbf{r}})p_{\mathbf{v}}(\tau_0)^2)|_{\mathcal{M}_0} + \mathbf{O}(\delta), \tag{99b}$$

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 $\mathcal{M}_0$ denoted as $p_{\mathbf{v}}(\tau)|_{\mathcal{M}_0}$ can be written as an ODE (91) and when coupled with an initial condition its solution is given as in (94) 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

$$\frac{\mathrm{d}}{\mathrm{d}\tau} p_{\mathbf{v}}\Big|_{\mathcal{M}_\delta} = p_{\mathbf{v}}(1 - p_{\mathbf{v}})S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) + O(\delta), \tag{100}$$

and when coupled with an initial condition given in (99), its solution is

$$p_{\mathbf{v}}(\tau)\Big|_{\mathcal{M}_\delta} = \frac{1}{1 + \frac{1-p_{\mathrm{c}}}{p_{\mathrm{c}}} \exp[-\tau S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]} + O(\delta) \tag{101}$$

where $p_{\mathrm{c}} = p_{\mathbf{v}}(\tau_0)$ is the initial weighted mutant frequency in the slow subsystem. The equations (100)-(101) are the first order approximations of the dynamics on $\mathcal{M}_\delta$.

So far we focused on the dynamcis of $p_{\mathbf{v}}(\tau)$. But since on $\mathcal{M}_0$ we have $p_{\mathbf{v}}(\tau) = p_{\boldsymbol{\phi}}(\tau)$ for all $\tau \geq \tau_0$ (Section 6.3), and because δ is nonzero in (100)-(101), we can re-write (100)-(101) in fast time $t = \tau/\delta$ and in terms of $p_{\boldsymbol{\phi}}$ as

$$\frac{\mathrm{d}}{\mathrm{d}t} p_{\boldsymbol{\phi}}\Big|_{\mathcal{M}_\delta} = \delta p_{\boldsymbol{\phi}}(1 - p_{\boldsymbol{\phi}})S_{\boldsymbol{\eta}}(\hat{\mathbf{x}}) + O(\delta^2) \tag{102}$$

and

$$p_{\boldsymbol{\phi}}(t)\Big|_{\mathcal{M}_\delta} = \frac{1}{1 + \frac{1-p_{\mathrm{c}}}{p_{\mathrm{c}}} \exp[-t\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})]} + O(\delta). \tag{103}$$

Equations (100)-(103) thus establish the first order approximation for any average mutant frequency in the fast-slow mutant resident system (87)-(88) 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).

It now remains to investigate whether the first order approximations (100)-(103) on $\mathcal{M}_\delta$ are generic in the sense that any other solution nearby is a small perturbation away, which is a necessary requirement for us to be able to choose initial conditions away from $\mathcal{M}_\delta$. This leads to **step 2** of the proof in which 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 $p_{\boldsymbol{\phi}}(t_0)$ near $\mathcal{M}_\delta$, for some initial time t_0, will become arbitrarily close to (102)-(103) in finite time that depends on δ. See Figure 2 panels (d)-(f) for a graphical representation of the proof. □

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Using the weighted mutant frequency $p_{\mathbf{v}}$ and geometric singular perturbation theory (Fenichel, 1979; Hek, 2010; Kuehn, 2015), we showed that the "invasion implies substitution"-principle holds for our metacommunity model, saying that the dynamics of any average population-wide 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 (96)-(98). Moreover, the rate at which change in mutant frequency occurs in the 'normal' population dynamical time t is characterized by $\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$ giving the number of additional mutant alleles produced per unit of time by the focal mutant allele (see further discussion in Section 7.2). This 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 mutant frequency change 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. (100) and (102)), and thus $S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$ characterized the rate of mutant frequency change in the slow evolutionary time τ. In contrast, the solution of the average mutant frequency $p_{\boldsymbol{\phi}}$ of the fast-slow system where δ is small is $O(\delta)$-distance away from the solution of the slow subsystem (where $\delta = 0$), and this distance is independent of the timescale (compare e.g. (101) and (103)). This implies that the per unit time change in the arithmetic mean mutant frequency is δ-distance away from the asymptotic change in the reproductive value weighted frequency. Moreover, as δ approaches 0 but remains non-zero, the slow subsystem (94) gives an increasingly good approximation to (97).

7.2 Inclusive fitness effect and singular points

An alternative representation of the directional selection coefficient $\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$ as characterised in (98) can be obtained by considering a 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), and recall that 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 (90) 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 $\delta S_{\boldsymbol{\eta}}(\hat{\mathbf{x}})$ characterised in (98) can be written and interpreted as given in (2) (by adding a superscript $\mathbf{u}$ to $\mathbf{v}$). Note that in (2) we write $S_{\boldsymbol{\eta}}(\boldsymbol{z})$ thus using $\boldsymbol{z}$ instead of $\hat{\mathbf{x}}$ as the main argument because it is commonly used as such to study phenotype evolution and we shall follow this notational change henceforth. We emphasise that under both scalings, (2) and (98), 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 (98) where the contribution of group members is accounted for.

We can now further 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

$$S_{\boldsymbol{\eta}}(\boldsymbol{z}) = \widetilde{\mathbf{S}}(\boldsymbol{z}) \cdot \boldsymbol{\eta} = \sum_{bs\in\mathcal{C}} S_{bs}(\boldsymbol{z})\hat{u}_{bs}\eta_{bs}, \tag{104}$$

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

$$S_{bs}(\boldsymbol{z}) = \sum_{au\in\mathcal{C}} \hat{v}^{\mathrm{u}}_{au} \left[\frac{\partial h_{au,bs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{foc},bs}} + \sum_{cs\in\mathcal{C}} \frac{\partial h_{au,cs}(\hat{\mathbf{x}})}{\partial z_{\mathrm{loc},bs}} \frac{N_{cs}}{N_{bs}} \hat{r}_{bcs} \right] \tag{105}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is the selection pressure on phenotypic component $z_{bs} \in \mathcal{Z}_{\mathcal{C}}$. Similarly to (2) and (98), the first term in the brackets $\partial h_{au,bs}(\hat{\mathbf{x}})/\partial z_{\text{foc},bs}$, when multiplied by δ, gives the 'additional' number of gene copies (or offspring number owing to asexual reproduction) in class $au \in \mathcal{C}$ produced per unit of time by an individual of class $bs \in \mathcal{C}$ as a consequence of it bearing the mutant instead of the resident allele. In contrast to (2) and (98), however, the second term $[(\partial h_{au,bs}(\hat{\mathbf{x}})/\partial z_{\text{loc},cs})(N_{cs}/N_{bs})]$ in the second summation, when multiplied by δ, gives the 'additional' number of class $au \in \mathcal{C}$ gene copies produced per unit of time by all *non-focal* local individuals of class $cs \in \mathcal{C}$ (group members of the focal individual) as a consequence of a focal individual of class $bs \in \mathcal{C}$ bearing the mutant instead of the resident allele. The weight $\hat{r}_{bcs}$ ensures that we only count those non-focal individuals $cs \in \mathcal{C}$ who have a common ancestor with the focal individual $bs \in \mathcal{C}$ and are thus likewise mutants. Each such additional mutant gene copy in class $au \in \mathcal{C}$ is then weighted with its reproductive value $\hat{v}^u_{au}$ that gives the expected asymptotic *number* of descendants, and we finally sum over all classes of the offspring $au \in \mathcal{C}$. The directional selection coefficient $\delta S_{\boldsymbol{\eta}}(\boldsymbol{z})$ thus gives the number of additional mutant alleles produced either directly (the first term in (105)) or indirectly (via related individuals, the second term in (105)) per unit of time by a focal mutant allele that is distributed according to $\hat{\mathbf{u}}$. That is, in (104)-(105), the individual fitness 'belongs' to the focal individual in the first term in (105) (as in (2) and (98)) whereas in the second term in (105) it belongs to non-focal but related individuals (in contrast to (2) and (98)) that are affected by the focal's trait(s), whereby the focal individual can be taken as the actor of all fitness changes as according to inclusive fitness theory (Hamilton, 1970). Hence, (104) 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 (105).

It then follows from (104) that a singular point $\boldsymbol{z}^* \in \mathcal{Z}$ needs to satisfy

$$S_{bs}(\boldsymbol{z}^*) = 0 \qquad \text{for all } bs \in \mathcal{C}, \tag{106}$$

since at a singular point no unilateral deviation in phenotype should result in a change in fitness, that is, (104) 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 when the mutation matrix is positive definite (Leimar, 2005, 2009).

8 Discussion

We have proved the "invasion implies substitution"-principle ((96)-(98), Section 7) for the evolution of a multidimensional quantitative trait expressed by the individuals of a focal species that is physiologically class structured e.g. by age, size and stage, and that resides in a metacommunity with finite local community size where reproduction, dispersal, survival, and development are stochastic and frequency- and density-dependent.

8.1 Evolutionary dynamics across multiple timescales

The evolutionary dynamics discussed in this paper are driven by three qualitatively different evolutionary forces (processes): mutation, selection and class-transmission. Mutation-driven changes

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (55)). 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 (96)-(98) (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 and 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 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 arithmetic mean allele frequency change (96)-(98) whose sign is independent of frequency. Our result thus extends earlier work whose representation of (1) was given in terms of a reproductive value weighted mutant frequency and not the arithmetic mean (Roze and Rousset, 2004; Rousset and Ronce, 2004; Rousset, 2004; Van Cleve et al., 2010; Priklopil and Lehmann, 2020).

We must stress that the arithmetic mean allele frequency change (96)-(98) holds not only in the limiting slow subsystem where $\delta \to 0$ 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 (96)-(98) hold (Section 7). In this respect, our results complement 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). In this work, the strength of selection δ is inversely proportional to the system size that tends to infinity while $\delta \to 0$ but keeping their product finite, and it invokes a two timescale method of Ethier and Nagylaki (1980). Our 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, that is, when δ deviates from 0. In the present metacommunity model, this is guaranteed by the assumption of hyperbolicity (ii), and jointly with the assumption on phenotypic similarity (i) and the weighted frequency $p_{\mathbf{v}}$ fully justify the "invasion implies substitution"-principle

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(96)-(98). 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 the mutation processes. For scalar-valued phenotypes, recurrent invasion-substitution events will always either increase or decrease the phenotypic value until the phenotype becomes a singular point and the same should hold for multidimensional traits under mild assumptions on the mutation matrix when the singular point is convergence stable (see section 7.2). Interestingly, a mutation rate that results in several competing mutations in the metacommunity does not interfere with the qualitative nature of this gradual process. This is because any finite number of overlapping mutations perturb the structure of the resident species only to order $O(\delta)$ (Meszéna et al., 2005). Therefore, if the mutation of interest results in a positive selection coefficient $\delta S_{\boldsymbol{\eta}}(\mathbf{z})$ 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 $\delta S_{\boldsymbol{\eta}}(\mathbf{z})$ will go to fixation upon a successful invasion event. A mutation rate that generates several overlapping mutations thus only affects the rate of evolution but not its direction. Furthermore, it is important to emphasise that the "invasion implies substitution"-principle (96)-(98) is valid also for 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 and mutations have small phenotypic effects, the "invasion implies substitution"-principle (1)-(2) gives a 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 frequency- and density-dependent environmental feedbacks both at the group- as well as metacommunity-level. Our model also applies to situations where reproduction, dispersal, survival and physiological development 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, that is, all rates used in this paper 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 ergodic path of environmental states including such environmental change is unlikely to change the generic shape of the selection coefficient, a point illustrated by Lenormand et al. (2009), proven more generally by Lion (2018b, eq. 23) for panmixia, and implied by the analysis of Svardal et al. (2015, eq. B.8) for spatial structure. Moreover, in metacommunity models where the total size of the focal species is finite, our dynamical equations can in principle be used to construct a diffusion approximation with two timescales 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). Here, the infinitesimal mean of the diffusion is given by the mean change (91) and the infinitesimal variance is obtainable from the recurrence equations for relatedness (Rousset, 2004, chapter 9-11). Finally, we formulated our model only in the context of haploid reproduction. Still, 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 (96) and to interpret relatedness in the directional selection coefficient

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(98) 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 of structured natural populations.

Acknowledgment

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A Appendix: Individual fitness

A.1 Resident individual fitness

In this section we detail the steps to obtain individual fitness (16) from the resident group and metacommunity dynamics (10). Using $n_{au} = N_{au}x_u$, we have

$$\begin{aligned}\dot{n}_{au} &= N_{au}\dot{x}_u \\ &= N_{au}\sum_{s\in\mathcal{S}}\left[q^{\mathrm{p}}_{u,s}(\mathbf{x}) + q^{\mathrm{d}}_{u,s}(\mathbf{x}) + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]x_s \\ &= \sum_{s\in\mathcal{S}} N_{au}\left[\sum_{c,b\in\mathcal{P}}\lambda^{\mathrm{p}}_{cu,bs}(\mathbf{x})N_{bs} + \sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\lambda^{\mathrm{dB}}_{cu,s,dw}(\mathbf{x})N_{dw}x_w + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]x_s,\end{aligned} \tag{A.1}$$

where we have applied Section 3.2. Whenever an individual $a \in \mathcal{P}$ is produced or removed from group $s \in \mathcal{S}$ so that this group transitions to $u \in \mathcal{S}$, the number of $a \in \mathcal{P}$ in the group changes and N_{au} is either equal to $N_{as} + 1$ or $N_{as} - 1$, otherwise $N_{au} = N_{as}$. This motivates us to re-write the above as

$$\begin{aligned}\dot{n}_{au} &= \sum_{s\in\mathcal{S}}(N_{au} - N_{as} + N_{as})\left[\sum_{c,b\in\mathcal{P}}\lambda^{\mathrm{p}}_{cu,bs}(\mathbf{x})N_{bs} + \sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\lambda^{\mathrm{dB}}_{cu,s,dw}(\mathbf{x})N_{dw}x_w + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]x_s \\ &= \sum_{s\in\mathcal{S}}(N_{au} - N_{as})\left[\sum_{b\in\mathcal{P}}\lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x})N_{bs} + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x})N_{bs} + \sum_{dw\in\mathcal{C}}\lambda^{\mathrm{dB}}_{au,s,dw}(\mathbf{x})N_{dw}x_w\right]x_s + \\ &+ \sum_{s\in\mathcal{S}}(N_{au} - N_{as})\left[\sum_{c\in\mathcal{P}}\lambda^{\mathrm{T}}_{cu,as}(\mathbf{x})N_{as} + \lambda^{\mathrm{D}}_{au,as}(\mathbf{x})N_{as}\right]x_s + \\ &+ \sum_{s\in\mathcal{S}}N_{as}\left[\sum_{c,b\in\mathcal{P}}\lambda^{\mathrm{p}}_{cu,bs}(\mathbf{x})N_{bs} + \sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\lambda^{\mathrm{dB}}_{cu,s,dw}(\mathbf{x})N_{dw}x_w + q^{\mathrm{mc}}_{u,s}(\mathbf{x})\right]x_s\end{aligned} \tag{A.2}$$

where in the second equality we have applied Sections 3.1-3.2, and by relabeling the indices on the second line in (A.2) we get

$$\begin{aligned}\dot{n}_{au} &= \sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{P}}\left[(N_{au} - N_{as})\left(\lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x})\right) + \right. \\ &\left. + \sum_{w\in\mathcal{S}}(N_{au} - N_{aw})x_w\lambda^{\mathrm{dB}}_{au,w,bs}(\mathbf{x})\right]N_{bs}x_s + \\ &+ \sum_{s\in\mathcal{S}}(N_{au} - N_{as})\left[\sum_{c\in\mathcal{P}}\lambda^{\mathrm{T}}_{cu,as}(\mathbf{x})N_{as} + \lambda^{\mathrm{D}}_{au,as}(\mathbf{x})N_{as}\right]x_s + \\ &+ \sum_{s\in\mathcal{S}}q_{u,s}(\mathbf{x})N_{as}x_s.\end{aligned} \tag{A.3}$$

Now, for all $s \neq u \in \mathcal{S}$, we have $N_{au} - N_{as} = 1$ and $N_{au} - N_{aw} = 1$ on the first and second line in (A.3), respectively, because an offspring $a \in \mathcal{P}$ is produced by birth into a group that transitions to $u \in \mathcal{S}$, and $N_{au} - N_{as} = -1$ on the third line in (A.3) because a parent $a \in \mathcal{P}$ is removed by physiological transition or death from a group that transitions to $u \in \mathcal{S}$. In all other situations where

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$s \in \mathcal{S}$ transitions to $u \in \mathcal{S}$ we have $N_{au} = N_{as}$. We thus get, after some reorganization,

$$\begin{aligned}
\dot{n}_{au} &= \sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{P}} \left[\lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{dB}}_{au,bs}(\mathbf{x})\right] N_{bs}x_s - \\
&\quad - \sum_{s\in\mathcal{S}} \left[\sum_{c\in\mathcal{P}} \lambda^{\mathrm{T}}_{cu,as}(\mathbf{x}) + \lambda^{\mathrm{D}}_{au,as}(\mathbf{x})\right] N_{as}x_s + \sum_{s\in\mathcal{S}} q_{u,s}(\mathbf{x}) N_{as}x_s \\
&= \sum_{bs\in\mathcal{C}} \left[\lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{dB}}_{au,bs}(\mathbf{x})\right] n_{bs} + \\
&\quad + \sum_{s\in\mathcal{S}} \left[q_{u,s}(\mathbf{x}) - \sum_{c\in\mathcal{P}} \lambda^{\mathrm{T}}_{cu,as}(\mathbf{x}) - \lambda^{\mathrm{D}}_{au,as}(\mathbf{x})\right] n_{as} \\
&= \sum_{bs\in\mathcal{C}} h_{au,bs}(\mathbf{x}) n_{bs},
\end{aligned} \tag{A.4}$$

where for $u \neq s \in \mathcal{S}$ we can then write

$$h_{au,bs}(\mathbf{x}) = \begin{cases} \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) & \text{for} \quad a \neq b \\ \lambda^{\mathrm{d}}_{bu,bs}(\mathbf{x}) + \lambda^{\mathrm{pB}}_{bu,bs}(\mathbf{x}) + q_{u,s}(\mathbf{x}) - \sum_{c\neq a\in\mathcal{P}} \lambda^{\mathrm{T}}_{cu,bs}(\mathbf{x}) - \lambda^{\mathrm{D}}_{bu,bs}(\mathbf{x}) & \text{for} \quad a = b, \end{cases} \tag{A.5a}$$

and for $u = s \in \mathcal{S}$ we have

$$h_{as,bs}(\mathbf{x}) = \lambda^{\mathrm{d}}_{as,bs}(\mathbf{x}) + \delta_{a,b} q_{s,s}(\mathbf{x}) = \lambda^{\mathrm{d}}_{as,bs}(\mathbf{x}) - \delta_{a,b} \sum_{w\neq s\in\mathcal{S}} q_{w,s}(\mathbf{x}), \tag{A.5b}$$

for all $a, b \in \mathcal{P}$. Alternatively, we can represent (A.5) with a single equation

$$h_{au,bs}(\mathbf{x}) = \lambda^{\mathrm{d}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) + \delta_{a,b}\left[q_{u,s}(\mathbf{x}) - \sum_{c\in\mathcal{P}} \lambda^{\mathrm{T}}_{cu,as}(\mathbf{x}) - \lambda^{\mathrm{D}}_{au,as}(\mathbf{x})\right], \tag{A.6}$$

for all $au, bs \in \mathcal{C}$, and which is the expression for individual fitness (16) given in Section 3.

A.1.1 Individual fitness: connection to previous models and formulations

Physiologically unstructured metapopulation model. Here we consider a single physiologically unstructured species and suppose that groups are identical in terms of abiotic factors. The set of g-states can then be represented with $\mathcal{S} = \{0, 1, 2, \ldots, N_{\max}\}$ specifying the number of individuals of this focal species inhabiting a group, and the only processes occurring in each group are death, philopatric birth and birth via dispersal, which is conceptually similar to models with local group demographic fluctuations considered previously (Metz and Gyllenberg, 2001; Cadet et al., 2003; Rousset and Ronce, 2004; Lehmann et al., 2006; Alizon and Taylor, 2008; Parvinen, 2013). The individual fitness (16) then reduces to

$$h_{N',N}(\mathbf{x}) = \lambda^{\mathrm{d}}_{N',N}(\mathbf{x}) + \lambda^{\mathrm{pB}}_{N',N}(\mathbf{x}) + \left(q_{N',N}(\mathbf{x}) - \lambda^{\mathrm{D}}_{N,N}(\mathbf{x})\right) \qquad \text{for all} \quad N, N' \in \mathcal{S}, \tag{A.7}$$

where

$$q_{N',N}(\mathbf{x}) = \left(\lambda^{\mathrm{pB}}_{N',N}(\mathbf{x}) + \lambda^{\mathrm{D}}_{N',N}(\mathbf{x})\right) N + \sum_{K\in\mathcal{S}} \lambda^{\mathrm{dB}}_{N',N,K}(\mathbf{x}) K x_K \qquad \text{for all} \quad N, N' \in \mathcal{S},\ N \neq N'. \tag{A.8}$$

Here,

$$\lambda^{\mathrm{D}}_{N',N}(\mathbf{x}) = \begin{cases} \mu_N(\mathbf{x}) & \text{for} \quad N' = N - 1 \\ 0 & \text{otherwise,} \end{cases} \tag{A.9}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and

$$\lambda^{\mathrm{pB}}_{N',N}(\mathbf{x}) = \begin{cases} \beta^{\mathrm{p}}_{N}(\mathbf{x}) & \text{for} \quad N' = N+1 \\ 0 & \text{otherwise,} \end{cases} \tag{A.10}$$

and $\lambda^{\mathrm{d}}_{N',N}(\mathbf{x}) = \sum_{K\in\mathcal{S}} x_K \lambda^{\mathrm{dB}}_{N',K,N}(\mathbf{x})$ where

$$\lambda^{\mathrm{dB}}_{N',K,N}(\mathbf{x}) = \begin{cases} \beta^{\mathrm{d}}_{N' \overset{K}{\leftarrow} N}(\mathbf{x}) & \text{for} \quad N' = K+1 \\ 0 & \text{otherwise.} \end{cases} \tag{A.11}$$

By convention we have $\mu_0(\mathbf{x}) = 0$, $\beta^{\mathrm{p}}_{N_{\max}}(\mathbf{x}) = 0$ and $\beta^{\mathrm{d}}_{N' \overset{N_{\max}}{\leftarrow} N}(\mathbf{x}) = 0$. The interpretation of (A.7) is analogous to the one in (16) except that the i-process for physiological transitions as well as all the other processes of non-focal species and abiotic factors are absent. Note that, if we further have $\beta^{\mathrm{p}}_{N}(\mathbf{x}) = (1-m)\beta_N(\mathbf{x})$ and $\beta^{\mathrm{dB}}_{N' \overset{K}{\leftarrow} N}(\mathbf{x}) = m\beta_N(\mathbf{x})$ for some migration probability m and birth rate $\beta_N(\mathbf{x})$ depending only on the local density of the parent and which is the realm of the aforementioned demographic models, then

$$h_{N',N}(\mathbf{x}) = \begin{cases} \beta_N(\mathbf{x}) + q_{N+1,N}(\mathbf{x}) & \text{for} \quad N' = N+1 \\ x_{N-2} m\beta_N(\mathbf{x}) + \big(q_{N-1,N}(\mathbf{x}) - \mu_N(\mathbf{x})\big) & \text{for} \quad N' = N-1 \\ x_{N-1} m\beta_N(\mathbf{x}) + q_{N,N}(\mathbf{x}) & \text{for} \quad N' = N \\ x_{N'-1} m\beta_N(\mathbf{x}) & \text{for} \quad N' \neq \{N-1, N, N+1\}, \end{cases} \tag{A.12}$$

where

$$q_{N',N}(\mathbf{x}) = \begin{cases} (1-m)\beta_N(\mathbf{x})N + m\sum_{K\in\mathcal{S}} \beta_K(\mathbf{x}) K x_K & \text{for} \quad N' = N+1 \\ \mu_N(\mathbf{x})N & \text{for} \quad N' = N-1 \\ -[q_{N+1,N}(\mathbf{x}) + q_{N-1,N}(\mathbf{x})] & \text{for} \quad N' = N \\ 0 & \text{otherwise.} \end{cases} \tag{A.13}$$

Well-mixed physiologically structured population model. Here we reduce our metacommunity model to a well-mixed population model for a single focal species that is physiologically structured. In so doing, we assume that each 'group' is inhabited by at most one individual and that individuals reproduce via dispersal into an empty uninhabited group. The set of g-states can be represented with $\mathcal{S} = \{0\} \cup \mathcal{P}$ implying that each group is either empty or it is inhabited by a single individual that is in one of the finite number of physiological states. The only processes are birth (via dispersal), physiological transition and death. This model is conceptually similar to the one in Lion (2018a,b) and Priklopil and Lehmann (2020). The individual fitness (16) then reduces to

$$h_{a,b}(\mathbf{x}) = \lambda^{\mathrm{d}}_{a,b}(\mathbf{x}) + \lambda^{\mathrm{T}}_{a,b}(\mathbf{x}) - \delta_{a,b}\left(\sum_{c\in\mathcal{P}} \lambda^{\mathrm{T}}_{c,b} + \lambda^{\mathrm{D}}_{b,b}\right), \qquad \text{for all} \quad a, b \in \mathcal{P}. \tag{A.14}$$

Here, for all $a, b \in \mathcal{P}$, we have $\lambda^{\mathrm{d}}_{a,b}(\mathbf{x}) = x_0 \lambda^{\mathrm{dB}}_{a,0,b}(\mathbf{x})$ where $\lambda^{\mathrm{d}}_{a,0,b}(\mathbf{x}) = \beta^{\mathrm{d}}_{a \overset{0}{\leftarrow} b}(\mathbf{x})$, and $\lambda^{\mathrm{D}}_{b,b}(\mathbf{x}) = \mu_b(\mathbf{x})$ and

$$\lambda^{\mathrm{T}}_{a,b}(\mathbf{x}) = \begin{cases} \gamma_{a\leftarrow b}(\mathbf{x}) & \text{for} \quad a \neq b \\ 0 & \text{otherwise.} \end{cases} \tag{A.15}$$

The interpretation of (A.14) is analogous to the one in (16), except that here each 'group' can

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transition from one state to another only due to the i-processes of the focal parent individual $b \in \mathcal{P}$. Therefore, the 'survival' term in the brackets in (A.14) implies that the focal parent $a = b \in \mathcal{P}$ has either died and is removed from the group or it physiologically transitions to some other physiological state.

Conditional fitness. The concept of individual fitness (16) 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 because 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

$$h_{au,bs}(\mathbf{x}) = h^{\mathrm{d}}_{au,bs}(\mathbf{x}) + h^{\mathrm{p}}_{au,bs}(\mathbf{x}), \tag{A.16a}$$

where

$$h^{\mathrm{d}}_{au,bs}(\mathbf{x}) = \sum_{w \in \mathcal{S}} x_w h^{\mathrm{d}}_{au,w,bs}(\mathbf{x}) \tag{A.16b}$$

and

$$\begin{aligned} h^{\mathrm{d}}_{au,w,bs}(\mathbf{x}) &= \lambda^{\mathrm{dB}}_{au,w,bs}(\mathbf{x}) \\ h^{\mathrm{p}}_{au,bs}(\mathbf{x}) &= \lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x}) + \delta_{a,b}\left[q_{u,s}(\mathbf{x}) - \sum_{c \in \mathcal{P}} \lambda^{\mathrm{T}}_{cu,as}(\mathbf{x}) - \lambda^{\mathrm{D}}_{au,as}(\mathbf{x})\right]. \end{aligned} \tag{A.16c}$$

The *conditional* individual fitness (e.g. Rousset, 2004; Rousset and Ronce, 2004,Lehmann and Rousset, 2010 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

$$w_{au,bs}(\mathbf{x}) = w^{\mathrm{d}}_{au,bs}(\mathbf{x}) + w^{\mathrm{p}}_{au,bs}(\mathbf{x}), \tag{A.17a}$$

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

$$\begin{aligned} w^{\mathrm{d}}_{au,w,bs}(\mathbf{x}) &= \frac{h^{\mathrm{d}}_{au,w,bs}(\mathbf{x})}{q_{u,w}} \\ w^{\mathrm{p}}_{au,bs}(\mathbf{x}) &= \frac{h^{\mathrm{p}}_{au,bs}(\mathbf{x})}{q_{u,s}}. \end{aligned} \tag{A.17b}$$

Note that when $a = b \in \mathcal{P}$, then $w^{\mathrm{p}}_{au,as}(\mathbf{x}) = 1 + \frac{\lambda^{\mathrm{pB}}_{au,as}(\mathbf{x}) - \sum_{c \in \mathcal{P}} \lambda^{\mathrm{T}}_{cu,as}(\mathbf{x}) - \lambda^{\mathrm{D}}_{au,as}(\mathbf{x})}{q_{u,s}}$ and when $a \neq b \in \mathcal{P}$, then $w^{\mathrm{p}}_{au,bs}(\mathbf{x}) = \frac{\lambda^{\mathrm{pB}}_{au,bs}(\mathbf{x}) + \lambda^{\mathrm{T}}_{au,bs}(\mathbf{x})}{q_{u,s}}$. If we now further make the assumptions of the "physiologically unstructured metapopulation model" case where $\mathcal{S} = \{0, 1, 2, \ldots, N_{\max}\}$ and use (A.16)–(A.17) along with (A.12)–(A.13) and noting that $h^{\mathrm{d}}_{N',K,N}(\mathbf{x}) = m\beta_N(\mathbf{x})$ for $N' = K + 1$, then we recover equations (A.43)-(A.48) of Lehmann and Rousset (2010), which hold as the conditional fitnesses for the model of Alizon and Taylor (2008).

A.2 Mutant-resident individual fitnesses

A.2.1 Fitness of mutant and resident individuals

The derivation of fitness functions for mutant-resident dynamics is analogous to the one for resident dynamics (Appendix A, Section A.1). 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

all $a \in \mathcal{P}$, we get

$$\begin{aligned}
\dot{n}_{\mathrm{M},au} &= \sum_{j\in\mathcal{I}_u} j_{au}\dot{y}_{uj} \\
&= \sum_{j\in\mathcal{I}_u} j_{au} \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} \pi_{uj,si}(\mathbf{y})y_{si} \\
&= \sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} (i_{as}+j_{au}-i_{as})\pi_{uj,si}(\mathbf{y})y_{si} \\
&= \sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s} (j_{au}-i_{as})\left[\sum_{b\in\mathcal{P}}\left(\lambda^{\mathrm{M,pB}}_{auj,bsi}(\mathbf{y})+\lambda^{\mathrm{M,T}}_{auj,bsi}(\mathbf{y})\right)i_{bs}+\sum_{k\in\mathcal{I}_w}\sum_{dw\in\mathcal{C}}\lambda^{\mathrm{M,dB}}_{auj,s,dwk}(\mathbf{y})k_{dw}y_{dw}\right]y_{si} \\
&\quad+\sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}(j_{au}-i_{as})\left[\sum_{c\in\mathcal{P}}\lambda^{\mathrm{M,T}}_{cuj,asi}(\mathbf{y})+\lambda^{\mathrm{M,D}}_{auj,asi}(\mathbf{y})\right]i_{as}y_{si} \\
&\quad+\sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\left[\sum_{c,b\in\mathcal{P}}\bar{\lambda}^{\mathrm{p}}_{cu,bsi}(\mathbf{y})N_{bs}+\sum_{c,d\in\mathcal{P}}\sum_{wk\in\Omega}\bar{\lambda}^{\mathrm{d}}_{cu,s,dwk}(\mathbf{y})N_{dw}y_{wk}\right]i_{as}y_{si}+ \\
&\quad+\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})i_{as}y_{si}.
\end{aligned} \tag{A.18}$$

By exchanging the indices and using $y_{\mathrm{M},bsi|\mathrm{M},bs} = \frac{i_{bs}y_{si}}{n_{\mathrm{M},bs}}$, we can re-write the above as

$$\begin{aligned}
\dot{n}_{\mathrm{M},au} &= \sum_{bs\in\mathcal{C}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[(j_{au}-i_{as})\left(\lambda^{\mathrm{M,pB}}_{auj,bsi}(\mathbf{y})+\lambda^{\mathrm{M,T}}_{auj,bsi}(\mathbf{y})\right)i_{bs}+\right. \\
&\quad\left.+\sum_{w\in\mathcal{S}}(j_{au}-i_{aw})x_w\lambda^{\mathrm{M,dB}}_{auj,w,bsi}(\mathbf{y})\right]y_{\mathrm{M},bsi|\mathrm{M},bs}n_{\mathrm{M},bs} \\
&\quad+\sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}(j_{au}-i_{as})\left[\sum_{c\in\mathcal{P}}\lambda^{\mathrm{M,T}}_{cuj,asi}(\mathbf{y})+\lambda^{\mathrm{M,D}}_{auj,asi}(\mathbf{y})\right]y_{\mathrm{M},asi|\mathrm{M},as}n_{\mathrm{M},as} \\
&\quad+\sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[\pi^{\mathrm{p}}_{uj,si}(\mathbf{y})+\pi^{\mathrm{d}}_{uj,si}(\mathbf{y})+\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right]y_{\mathrm{M},asi|\mathrm{M},as}n_{\mathrm{M},as} \\
&= \sum_{bs\in\mathcal{C}} g^{\mathrm{M}}_{au,bs}(\mathbf{y})n_{\mathrm{M},bs},
\end{aligned} \tag{A.19}$$

where

$$g^{\mathrm{M}}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[\lambda^{\mathrm{M,d}}_{auj,bsi}(\mathbf{y})+\lambda^{\mathrm{M,pB}}_{auj,bsi}(\mathbf{y})+\lambda^{\mathrm{M,T}}_{auj,bsi}(\mathbf{y})+\right. \tag{A.20}$$

$$\left.+\delta_{a,b}\left(\pi_{uj,si}(\mathbf{y})-\sum_{c\in\mathcal{P}}\lambda^{\mathrm{M,T}}_{cuj,asi}(\mathbf{y})-\lambda^{\mathrm{M,D}}_{auj,asi}(\mathbf{y})\right)\right]y_{\mathrm{M},bsi|\mathrm{M},bs}, \tag{A.21}$$

which is the partitioning given in (47).

The derivation of the resident fitness for the mutant-resident dynamics (49) is identical to (A.18)-(A.20), except that one needs to replace the numbers of mutants within a group i_{as} with resident 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.2.2 Fitness of an average individual

We can obtain the average individual fitness (51)-(52) by taking an average over mutant and resident individual fitness's (47) and (49). Here, instead, we will provide an alternative derivation that is similar to the the derivation for resident fitness (16): the average individual growth-rate (51)-(52)

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

can be calculated from the mutant-resident group and metacommunity dynamics (42) as

$$
\begin{aligned}
\dot{n}_{au} &= N_{au}\dot{x}_u \\
&= N_{au}\sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\left[\pi^{\mathrm{p}}_{uj,si}(\mathbf{y})+\pi^{\mathrm{d}}_{uj,si}(\mathbf{y})+\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right]y_{si} \\
&= N_{au}\sum_{j\in\mathcal{I}_u}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_s}\left[\sum_{c,b\in\mathcal{P}}\bar{\lambda}^{\mathrm{p}}_{cuj,bsi}(\mathbf{y})N_{bs}+\sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\bar{\lambda}^{\mathrm{dB}}_{cuj,s,dwk}(\mathbf{y})N_{dw}x_w+\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right]y_{si} \\
&= \sum_{bs\in\mathcal{C}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[(N_{au}-N_{as})\left(\bar{\lambda}^{\mathrm{pB}}_{auj,bsi}(\mathbf{y})+\bar{\lambda}^{\mathrm{T}}_{auj,bsi}(\mathbf{y})\right)+\sum_{w\in\mathcal{S}}(N_{au}-N_{aw})x_w\bar{\lambda}^{\mathrm{dB}}_{auj,w,bsi}(\mathbf{y})\right]N_{bs}y_{si}+ \\
&\quad+\sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}(N_{au}-N_{as})\left[\sum_{c\in\mathcal{C}}\bar{\lambda}^{\mathrm{T}}_{cuj,asi}(\mathbf{y})+\bar{\lambda}^{\mathrm{D}}_{auj,asi}(\mathbf{y})\right]N_{as}y_{si}+ \\
&\quad+\sum_{s\in\mathcal{S}}\sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\left[\sum_{c,b\in\mathcal{P}}\bar{\lambda}^{\mathrm{p}}_{cuj,bsi}(\mathbf{y})N_{bs}+\sum_{c\in\mathcal{P}}\sum_{dw\in\mathcal{C}}\bar{\lambda}^{\mathrm{dB}}_{cuj,s,dwk}(\mathbf{y})N_{dw}x_w+\pi^{\mathrm{mc}}_{uj,si}(\mathbf{y})\right]N_{as}y_{si} \\
&= \sum_{bs\in\mathcal{C}}\bar{g}_{au,bs}(\mathbf{y})n_{bs},
\end{aligned}
\tag{A.22}
$$

where

$$
\bar{g}_{au,bs}(\mathbf{y}) = \sum_{j\in\mathcal{I}_u}\sum_{i\in\mathcal{I}_s}\Bigg[\bar{\lambda}^{\mathrm{d}}_{auj,bsi}(\mathbf{y})+\bar{\lambda}^{\mathrm{pB}}_{auj,bsi}(\mathbf{y})+\bar{\lambda}^{\mathrm{T}}_{auj,bsi}(\mathbf{y})+ \tag{A.23}
$$

$$
+\delta_{a,b}\left(\pi_{uj,si}(\mathbf{y})-\sum_{c\in\mathcal{P}}\bar{\lambda}^{\mathrm{T}}_{cuj,asi}(\mathbf{y})-\bar{\lambda}^{\mathrm{D}}_{auj,asi}(\mathbf{y})\right)\Bigg]y_{si|s}, \tag{A.24}
$$

which is the partitioning (52) from the main text.

B Appendix: Relatedness and genetic correlations

B.1 Jump-process approach

To facilitate the derivation of the recursion (30) for relatedness (Section 3.4.1), we represent here the metacommunity dynamics as a discret 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 (10) 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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (30)-(32) as given in the main text.

B.2 Mutant-pair dynamics

Here we derive the mutant-pair dynamics (82)-(83) 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 (21). 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

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

where the above rates are given in terms of i-processes as in (23)-(25) and where we used the consistency relation (61).

We obtain the ODE (82) for the mutant-pair frequencies $\mathbf{p}_{\mathrm{MM}}$ defined in (79) by differentiation and using (21) and (B.2), whereby

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

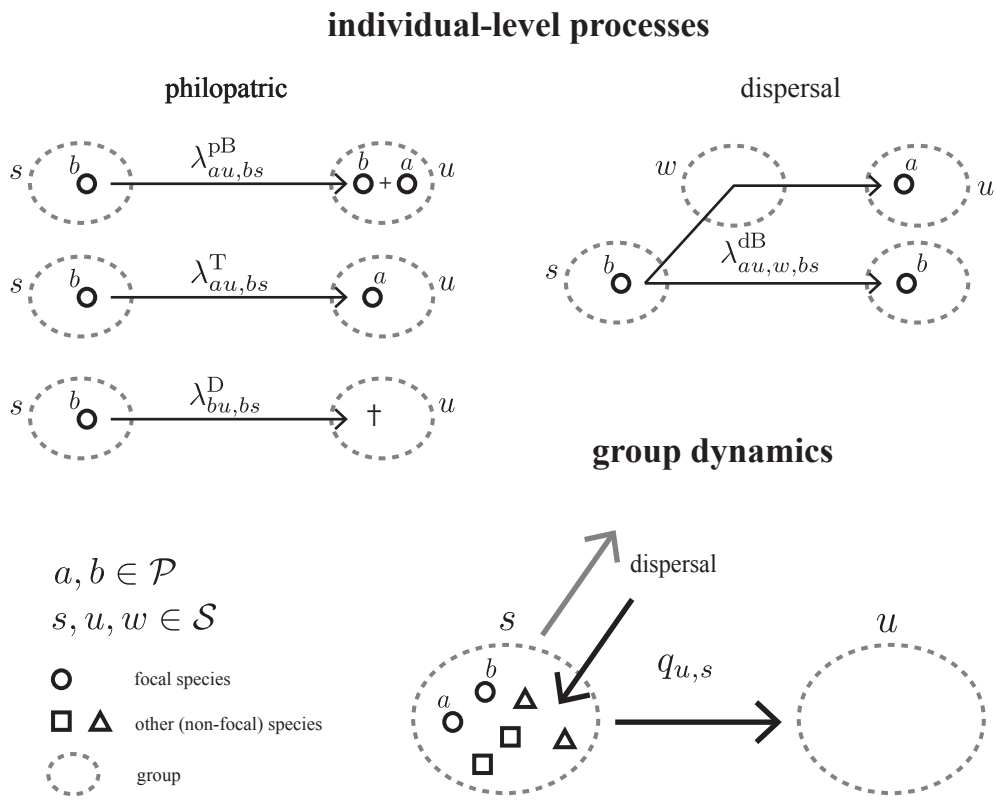

Figure 1: Individual-level processes of the focal species (top panels) and the group dynamics of the metacommunity (bottom panels). **Top panel:** As given in Section 3.1, the focal resident species undergoes philopatric individual-level processes of birth, physiological transition and death (top left 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 (11)-(14) 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 (10).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

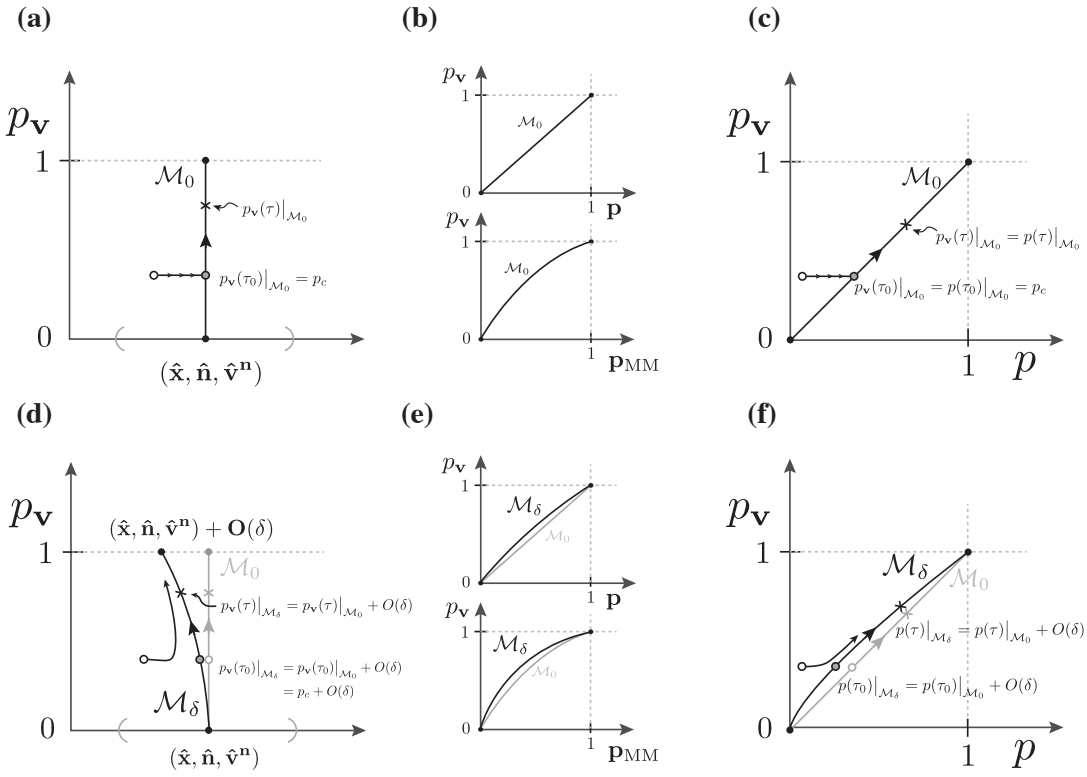

Figure 2: Panels (a)-(c) depict the fast and slow subsystems and the critical manifold $\mathcal{M}_0$ (89)-(90) (where $\delta = 0$, Sections 6.1-6.2), and, panels (d)-(f) depict their singular perturbations and the slow manifold $\mathcal{M}_\delta$ (where δ small but non-zero, Section 7). **(a)** 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 (58) 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 (91) and (94). **(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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

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

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

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.

J. F. Crow and M. Kimura. *An Introduction to Population Genetics Theory*. Harper and Row, New York, 1970.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

J. H. Gillespie. *Population Genetics: a Concise Guide*. Johns Hopkins University Press, Baltimore, Maryland, 2004.

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.

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

W. Hamilton. This week's citation classic. *Current Contents*, 40:16, 1988.

W. D. Hamilton. Selfish and spiteful behavior in an evolutionary model. *Nature*, 228:1218–1220, 1970.

D. L. Hartl, A. G. Clark, and A. G. Clark. *Principles of population genetics*, volume 116. Sinauer associates Sunderland, MA, 1997.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

J. Kingman. *Poisson Processes*. Oxford University Press, Oxford, 1992.

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.

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

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.

J. Li and F. Brauer. Continuous-time age-structured models in population dynamics and epidemiology. In *Mathematical epidemiology*, pages 205–227. Springer, 2008.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

S. Lion and S. Gandon. Habitat saturation and the spatial evolutionary ecology of altruism. *Journal of Evolutionary Biology*, 22(7):1487–1502, 2009. ISSN 1420-9101 (Electronic). doi: 10.1111/j.1420-9101.2009.01769.x.

M. Lynch and B. Walsh. *Evolution and Selection of Quantitative Traits*. Sinauer Associates, New York, 2018.

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.

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.

K. Parvinen, H. Ohtsuki, and J. Y. Wakano. Spatial heterogeneity and evolution of fecundity-affecting traits. *Journal of Theoretical Biology*, 2018.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

H. Svardal, C. Rueffler, and J. Hermisson. A general condition for adaptive genetic polymorphism in temporally and spatially heterogeneous environments. *Theoretical Population Biology*, 99:76–97, 2015.

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.

M. Wechselberger. Geometric singular perturbation theory beyond the standard form. 2020.

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

G. Wild and A. Traulsen. The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology*, 247(2):382–390, 2007.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.15.426821; this version posted August 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D. S. Wilson. Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. *Ecology*, 73(6):1984–2000, 1992.