Mathematical Biology

Lotka–Volterra approximations for evolutionary trait-substitution processes

Hiroshi C. Ito[1,2] · Ulf Dieckmann[1,2] · Johan A. J. Metz[1,3,4]

Abstract

A set of axioms is formulated characterizing ecologically plausible community dynamics. Using these axioms, it is proved that the transients following an invasion into a sufficiently stable equilibrium community by a mutant phenotype similar to one of the community's finitely many resident phenotypes can always be approximated by means of an appropriately chosen Lotka–Volterra model. To this end, the assumption is made that similar phenotypes in the community form clusters that are well-separated from each other, as is expected to be generally the case when evolution proceeds through small mutational steps. Each phenotypic cluster is represented by a single phenotype, which we call an approximate phenotype and assign the cluster's total population density. We present our results in three steps. First, for a set of approximate phenotypes with arbitrary equilibrium population densities before the invasion, the Lotka–Volterra approximation is proved to apply if the changes of the population densities of these phenotypes are sufficiently small during the transient following the invasion. Second, quantitative conditions for such small changes of population densities are derived as a relationship between within-cluster differences and the leading eigenvalue of the community's Jacobian matrix evaluated at the equilibrium population densities before the invasion. Third, to demonstrate the utility of our results, the 'invasion implies substitution' result for monomorphic populations is extended to arbitrarily polymorphic populations consisting of well-recognizable and -separated clusters.

✉ Hiroshi C. Ito
   hiroshibeetle@gmail.com

1   Evolution and Ecology Program, International Institute for Applied Systems Analysis,
    Schlossplatz 1, 2361 Laxenburg, Austria

2   Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced
    Studies (Sokendai), Hayama 240-0193, Kanagawa, Japan

3   Mathematical Institute and Institute of Biology, Leiden University, P.O. Box 9512, 2300
    RA Leiden, The Netherlands

4   Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA Leiden, The Netherlands

🖄 Springer

1 Introduction

Ecological interactions create selection pressures that may change those very interactions. Such eco-evolutionary feedback can induce rich coevolutionary dynamics including cyclic coevolution (e.g., Dieckmann et al. 1995; Dieckmann and Law 1996), adaptive radiation (e.g., Ackermann and Doebeli 2004; Egas et al. 2005), adaptive speciation (e.g., Dieckmann and Doebeli 1999; Dieckmann et al. 2004; Rundle and Nosil 2004), taxon cycles (e.g., Kisdi et al. 2001; Ito and Dieckmann 2007), and community formation (e.g., Loeuille and Loreau 2005; Dieckmann et al. 2007; Ito et al. 2009; Takahashi et al. 2013). To arrive at tractable descriptions of such evolutionary dynamics, the assumption is often made that mutation rates are low relative to the timescale of population dynamics. This assumption reduces the evolutionary dynamics to a trait-substitution sequence resulting from repeated mutant invasions (Metz et al. 1992, 1996; Dieckmann and Law 1996). Such invasions potentially bring about various outcomes: most often, (1) extinction of only the resident that is parental to the mutant, and more rarely, (2) coexistence of the mutant with all residents, or (3) other combinations of extinctions of the parental resident, non-parental residents, and mutant.

It has been proved that when for all residents all potentially invading mutants are subject to directional selection and the resulting perturbations to the system are sufficiently weak, as measured by the product of fitness gradients and mutational step sizes relative to the return rate to their population-dynamical equilibrium before the invasion, invading mutants replace their parental residents—a statement referred to as the invasion–implies–substitution theorem (Geritz 2005; Dercole and Rinaldi 2008). The resulting trait-substitution sequences describe directional coevolution, characterized well by a set of ordinary differential equations called the canonical equations of adaptive dynamics theory (Dieckmann and Law 1996), which have a form similar to Lande's equations of quantitative genetics theory (Lande 1979).

Eventually, directional coevolution may take some residents to the neighborhood of peaks, troughs, or saddles of the community's fitness landscape, which means that those populations experience very weak directional selection. Here, an invading mutant may coexist with its parental resident, which may be followed by diversifying evolution of the two morphs, called evolutionary branching (Metz et al. 1996). If the community has a one-dimensional trait space and a single resident, necessary and sufficient conditions for its evolutionary branching into two distinct residents have been obtained (Metz et al. 1996; Geritz et al. 1998).

On the other hand, for higher-dimensional traits or more than one resident, obtaining formal conditions for the occurrence of evolutionary branching is difficult (but see Ito and Dieckmann (2014) for a special case). This is largely because in these more complex community dynamics it is not easy to analyze the outcomes of mutant

invasions (Metz et al. 1996). This difficulty may be reduced when the population dynamics can be approximated by Lotka–Volterra (LV) models, which are analytically more tractable and have been studied well (e.g., Zeeman 1993; Hofbauer and Sigmund 1998). The LV-approximation is possible when all existing residents and the mutant are similar to each other, so that they form a single phenotypic cluster (Meszéna et al. 2005; Durinx et al. 2008), which yields an expression for the invasion-fitness function in terms of resident and mutant phenotypes that is given by a rational function. By using this rational form, considerable progress in deriving conditions for multidimensional evolutionary branching has recently been made (Geritz et al. 2016; Sect. 9.3).

Dercole and Rinaldi (2008) proved that the LV-approximation holds also when all of the existing residents are not similar to each other, i.e., when every cluster has only a single resident, and their initial equilibrium population densities are not small. (Although such limiting assumptions for residents are not made in their proof, these assumptions are required when we consider trait-substitution sequences, as explained in Sect. 4.4). Thus, the remaining cases to be analyzed are (a) only some residents are similar to each other and (b) the population densities of some residents are very small so that they may go extinct as a result of the invasion. Both cases are likely to occur in multispecies coevolution, including processes involving multiple evolutionary branching and taxon cycles, commonly observed in numerical simulations of trait-mediated community dynamics (e.g., Doebeli and Dieckmann 2000; Ito and Dieckmann 2007). Therefore, the goal of the present paper is to obtain formal conditions for ensuring the LV-approximation for an arbitrary set of residents, including the aforementioned two cases. Based on the obtained conditions, the invasion–implies–substitution theorem can be extended to a mutant with an arbitrary set of residents.

The next section, Sect. 2, formulates a set of axioms that are expected to hold for ecologically plausible differential equations describing trait-mediated community dynamics. Section 3 derives a condition for ensuring the LV-approximation. Sections 4 and 5 derive sufficient conditions for satisfying this condition, in terms of properties of the fitness-generating function and mutational step sizes. Section 6 explains how the thresholds for the obtained sufficient conditions can be improved further. Section 7 shows how to examine the obtained sufficient conditions for a specific ecological model. Section 8 extends the invasion–implies–substitution theorem.

2 Framework and assumptions

2.1 Axioms for fitness-generating functions

We consider community dynamics written as

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = n_i \, F(s_i; \mathbf{s}; \mathbf{n}) \tag{2.1}$$

with population densities n_i for $i = 1, \ldots, N$.

We denote by $\mathcal{S} \subset \mathbb{R}^Z$ a compact Z-dimensional trait space, by $\mathbf{s} = (s_1, \ldots, s_N)^T \in \mathcal{S}^N$ an N-dimensional vector of trait values of the phenotypes present in the community, and by $\mathbf{n} = (n_1, \ldots, n_N)^T \in \mathbb{R}_+^N$ the vector of their population densities. The fitness-generating function

$$F : \mathcal{S} \times \bigcup_{N=1}^{\infty} \left(\mathcal{S}^N \times \mathbb{R}^N \right) \to \mathbb{R} : (s', \mathbf{s}, \mathbf{n}) \mapsto F(s'; \mathbf{s}; \mathbf{n}) \qquad (2.2)$$

describes the instantaneous per capita growth rate of an arbitrary phenotype s' with an infinitesimally small population density in the instantaneous environment produced by resident community composition $(\mathbf{s}, \mathbf{n})$ (Brown and Vincent 1987; Cohen et al. 1999). The fitness-generating function provides a fitness landscape for each community composition $(\mathbf{s}, \mathbf{n})$. We assume that it satisfies the following axioms:

(i)   *Smoothness*: F is smooth on each component of its domain $\mathcal{S} \times \mathcal{S}^N \times \mathbb{R}^N$.
(ii) *Symmetry*: $F(s'; \sigma\mathbf{s}; \sigma\mathbf{n}) = F(s'; \mathbf{s}; \mathbf{n})$ for all permutations σ operating on the indices of $(\mathbf{s}; \mathbf{n})$.
(iii) *Reducibility*: $\quad F(s'; (s_1, \ldots, s_N)^T; (n_1, \ldots, n_{N-1}, 0)^T) \quad = \quad F(s'; (s_1, \ldots, s_{N-1})^T; (n_1, \ldots, n_{N-1})^T)$.
(iv) *Exchangeability*: If $s_N = s_{N-1}$, then $F(s'; (s_1, \ldots, s_N)^T; (n_1, \ldots, n_N)^T) = F(s'; (s_1, \ldots, s_{N-1})^T; (n_1, \ldots, n_{N-1} + n_N)^T)$.
(v)   *Bounded world*: There exists an upper bound $\eta > 0$ for the community's total population density, i.e., Eq. (2.1) eventually restricts the population densities to $\left\{ (n_1, \ldots, n_N) \in \mathbb{R}_+^N \middle| \sum_{i=1}^N n_i \leq \eta \right\}$.

Below, we restrict the community's space of population densities to $[0, \eta]^N$.

The smoothness axiom (i) follows from the assumption that the population-dynamical behavior of individuals depends smoothly on their traits and that all ecological interactions are instantaneous. The latter assumption is implicit in the assumption that the per capita growth rate depends only on the arguments s' and $(\mathbf{s}, \mathbf{n})$. Axioms (ii) to (iv) are consistency conditions that go with representing the behaviour of large collectives of individuals by differential equations for their densities. Axiom (ii) follows from the arbitrariness of the ordering of the trait N-tuples, and axiom (iv) from the fact that individuals with the same trait values are assumed to be indistinguishable. The consequent additivity for identical phenotypes mechanistically lies at the heart of the LV-approximability. The bounded-world axiom (v) is just what it says: there necessarily is a limit to the biomass that a patch of world can support. Models that do not acknowledge this may on occasion be good approximations for specific purposes, but when we run into results contradicting the bounded-world assumption, we have to start modifying the model.

To keep the exposition simple, we assume from now on a one-dimensional trait space $\mathcal{S} \subset \mathbb{R}$. The results are generalized to higher-dimensional trait spaces in Sect. 5.4.

2.2 Population dynamics triggered by mutant invasion

We assume that the community is at a locally stable equilibrium $\widehat{n} = (\widehat{n}_1, \ldots, \widehat{n}_N)^{\mathrm{T}}$, determined by $F(s_i; \mathbf{s}; \widehat{n}) = 0$ for all $i = 1, \ldots, N$. When an invasion by a mutant $s' = s_{N+1}$ with $|s_{N+1} - s_N| = \varepsilon_\mu$ has occurred, the combined population dynamics can be written as

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = n_i \, F(s_i; \mathbf{s}'; \mathbf{n}') \tag{2.3}$$

for $i = 1, \ldots, N + 1$, where $\mathbf{s}' = (s_1, \ldots, s_N, s_{N+1})^{\mathrm{T}}$ and $\mathbf{n}' = (n_1, \ldots, n_N, n_{N+1})^{\mathrm{T}}$, starting from $\mathbf{n}' = (\widehat{n}_1, \ldots, \widehat{n}_N, n_{N+1})^{\mathrm{T}}$ with very small n_{N+1}, which means that $\mathbf{n}'$ is almost identical to the equilibrium before the invasion, $\widehat{n}' = (\widehat{n}_1, \ldots, \widehat{n}_N, 0)^{\mathrm{T}}$.

Please notice that here we have introduced the notational convention, to which we adhere throughout this paper, that vectors of dimension $N + 1$ directly corresponding to vectors of dimension N are denoted by an added prime, as in $\mathbf{s}'$, $\mathbf{n}'$, and $\widehat{n}'$.

Proposition 1 *For a sufficiently small mutational step size ε_μ, the fitness-generating function during the transient following mutant invasion can be approximated by a linear function of $\mathbf{n}'$,*

$$F(s_i; \mathbf{s}'; \mathbf{n}') = \widehat{F}_i + \sum_{j=1}^{N+1} a_{ij}(n_j - \widehat{n}_j),$$

$$\widehat{F}_i := F(s_i; \mathbf{s}'; \widehat{n}'),$$

$$a_{ij} := \left. \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} \right|_{\mathbf{n}' = \widehat{n}'}, \tag{2.4}$$

which upon substitution into Eq. (2.3) gives the approximating Lotka–Volterra model,

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = n_i \left[\gamma_i + \sum_{j=1}^{N+1} a_{ij} n_j \right] \tag{2.5}$$

with $\gamma_i = \widehat{F}_i - \sum_{j=1}^{N+1} a_{ij}\widehat{n}_j$.

The remainder of this paper is devoted to making precise the, very general, conditions under which this proposition holds, and to calculating the corresponding error bounds. Important variables and parameters used in our analysis are shown in Table 1.

Table 1 List of notation for parameters, phenotypes, population densities, fitness functions, and other quantities

Parameter	Formula	Explanation	Location
Important parameters			
η		Upper bound for the total population density a community can sustain	Axiom (v)
ε_μ		Phenotypic distance of a mutant phenotype from its parental resident phenotype	Above Eq. (2.3)
ε	$= \rho_\mu \varepsilon_\mu > \varepsilon_\mu$	Threshold phenotypic distance for clustering phenotypes	Above Eq. (3.1a)
ρ_m		Threshold for treating population densities less than $\rho_m \varepsilon$ as small	Above Eq. (5.2a)
N		Number of resident phenotypes before mutant invasion	Eq. (2.1)
M	$< N$	Number of phenotypic clusters described by approximate phenotypes	Above Eq. (3.1a)
L	$\leq M$	Number of approximate phenotypes with not-small population densities	Above Eq. (5.2a)
K	$= M - L \leq M$	Number of approximate phenotypes with small population densities	Below Eq. (5.8c)
d		Positive constant in Lyapunov function	Eq. (5.5c)

Phenotype	Formula	Explanation	Location
Phenotypes			
s	$= (s_1, \ldots, s_N)^{\mathrm{T}}$	Phenotypes before mutant invasion	
s'	$= s_{N+1} = s_N + \varepsilon_\mu$	Mutant phenotype	
s′	$= (s_1, \ldots, s_N, s_{N+1})^{\mathrm{T}}$	Phenotypes after mutant invasion	
s$_a$	$= (s_1, \ldots, s_M)^{\mathrm{T}}$	Approximate phenotypes, which serve as the representative phenotypes of clusters	
ρ_j	$= (s_j - s_{\mathrm{cid}(j)})/\varepsilon < 1$	Difference of the jth phenotype from its cluster's approximate phenotype, scaled by ε	
ρ	$= (0, \ldots, 0, \rho_{M+1}, \ldots, \rho_{N+1})^{\mathrm{T}}$	Vector of the differences of all phenotypes from their clusters' approximate phenotypes, scaled by ε	
cid(j)	$\in \{1, \ldots, M\}$	Identity of cluster to which the jth phenotype belongs	
com(i)	$= \{ j \,\vert\, \mathrm{cid}(j) = i, j = 1, \ldots, N+1\}$	Set of identities of phenotypes that belong to the ith cluster	

Densities	Formula	Explanation	Location
Population densities			
n	$= (n_1, \ldots, n_N)^{\mathrm{T}}$	Population densities before mutant invasion	Eq. (2.1)

Table 1 continued

Densities	Formula	Explanation	Location
$\widehat{n}$	$=(\widehat{n}_1,\ldots,\widehat{n}_N)^{\mathrm{T}}$	Equilibrium population densities before mutant invasion	Below Eq. (2.3)
n_{N+1}		Mutant population density	Below Eq. (2.3)
$\mathbf{n}'$	$=(n_1,\ldots,n_N,n_{N+1})^{\mathrm{T}}$	Population densities after mutant invasion	Below Eq. (2.3)
$\widehat{n}'$	$=(\widehat{n}_1,\ldots,\widehat{n}_N,0)^{\mathrm{T}}$		Below Eq. (2.3)
$\mathbf{m}$	$=(m_1,\ldots,m_M)^{\mathrm{T}}$	Population densities of approximate phenotypes $\mathbf{s}_{\mathrm{a}}$	Above Eq. (3.1a)
m_i	$=\displaystyle\sum_{j\in\mathrm{com}(i)}n_j$	Total population density of ith cluster, assigned to the ith approximate phenotype	Eq. (3.1a)
$(m_{M+1},\ldots,m_{N+1})$	$=(\varepsilon n_{M+1},\ldots,\varepsilon n_{N+1})$		Eq. (3.1b)
$\mathbf{m}'$	$=(m_1,\ldots,m_M,m_{M+1},\ldots,m_{N+1})^{\mathrm{T}}$		Above Eq. (3.3a)
$\widehat{m}'$	$=(\widehat{m}_1,\ldots,\widehat{m}_{N+1})^{\mathrm{T}}$		Above Eq. (3.3a)
	$=(\widehat{m}_1,\ldots,\widehat{m}_M,\varepsilon\widehat{n}_{M+1},\ldots,\varepsilon\widehat{n}_N,0)^{\mathrm{T}}$		
$\widehat{m}$	$=(\widehat{m}_1,\ldots,\widehat{m}_M)^{\mathrm{T}}$		Eq. (3.4a)
$\mathbf{x}$ (in Sect. 4)	$=(x_1,\ldots,x_M)^{\mathrm{T}}=\mathbf{P}(\mathbf{m}-\widehat{m})$	Transformed vectors of $\mathbf{m}$ for stability analysis	Eq. (4.4)
$\mathbf{m}_{\mathrm{x}}$	$=(m_1,\ldots,m_L)^{\mathrm{T}}$	Approximate phenotypes with not-small initial population densities	Eq. (5.2a)
$\mathbf{m}_{\mathrm{y}}$	$=(m_{L+1},\ldots,m_M)^{\mathrm{T}}$	Approximate phenotypes with small initial population densities	Eq. (5.2a)
$\tilde{\mathbf{m}}_{\mathrm{x}}(\mathbf{m}_{\mathrm{y}})$	$=\widehat{m}_{\mathrm{x}}-\mathbf{B}_{\mathrm{xx}}^{-1}\mathbf{B}_{\mathrm{xy}}\mathbf{m}_{\mathrm{y}}$	Center manifold	Eq. (5.3b)
$\widehat{m}$	$=(\widehat{m}_{\mathrm{x}},\widehat{m}_{\mathrm{y}})^{\mathrm{T}}=(\widehat{m}_{\mathrm{x}},\mathbf{0})^{\mathrm{T}}$	Modified equilibrium	Above Eq. (5.4a)
$\mathbf{x}$ (in Sect. 5)	$=(x_1,\ldots,x_L)^{\mathrm{T}}$ $=\mathbf{P}(\mathbf{m}_{\mathrm{x}}-\tilde{\mathbf{m}}_{\mathrm{x}}(\mathbf{m}_{\mathrm{y}}))$	Transformed vectors of $\mathbf{m}_{\mathrm{x}}-\tilde{\mathbf{m}}_{\mathrm{x}}$ for stability analysis	Eq. (5.4a)
$\mathbf{y}$	$=\mathbf{m}_{\mathrm{y}}$		Eq. (5.4a)
$\mathbf{w}$	$=(\mathbf{x},\mathbf{y})^{\mathrm{T}}$		Eq. (5.4a)

Fitness	Formula	Explanation	Location
Fitness functions			
$F(s_i;\mathbf{s}';\mathbf{n}')$	$=\frac{1}{n_i}\frac{\mathrm{d}n_i}{\mathrm{d}t}$	Per capita growth rate of phenotype s_i in the environment determined by phenotypes $\mathbf{s}'$ with population densities $\mathbf{n}'$	Eqs. (2.1), (2.3)
$\breve{F}(s_i;\mathbf{s}';\mathbf{m}')$	$=F(s_i;\mathbf{s}';\mathbf{n}')$		Eq. (3.2)

Table 1 continued

Fitness	Formula	Explanation	Location
$f(s_i; \mathbf{s}'; \mathbf{n}')$	$= \sum\limits_{j \in \text{com}(i)} \frac{n_j}{m_i} F(s_i; \mathbf{s}'; \mathbf{n}')$	Per capita growth rate of approximate phenotype s_i in the environment determined by phenotypes $\mathbf{s}'$ with population densities $\mathbf{n}'$	Eq. (4.1a)

Quantity	Formula or explanation	Location
Other quantities		
$\widehat{F}_i$	$= F(s_i; \mathbf{s}'; \widehat{n'})$	Eq. (2.4)
a_{ij}	$= \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n'_j}$	Eq. (2.4)
$\mathbf{a}_i^{\mathrm{T}}$	$= \left(a_{i,1}, \ldots, a_{i,N+1}\right)$	Eq. (3.5a)
$\mathbf{W}$	$= \frac{\partial \mathbf{m}'}{\partial \mathbf{n}'}$	Eq. (3.5a)
b'_{ij}	$= \left. \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} \right|_{\mathbf{m}' = \widehat{m'}}$	Eq. (3.3c)
$\mathbf{b}_i^{\prime\mathrm{T}}$	$= \left(b'_{i1}, \ldots, b'_{i\,N+1}\right)$	Eq. (3.3c)
b_{ij}	$\left. \frac{\partial \breve{F}(s_i; \mathbf{s_a}; \mathbf{m})}{\partial m_j} \right|_{\mathbf{m} = \widehat{m}}$	Eq. (4.3b)
$\mathbf{B}$	$= \begin{pmatrix} b_{11} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{M1} & \cdots & b_{MM} \end{pmatrix}$	Eq. (4.3b)
$\mathbf{J}$	$= \text{diag}(\widehat{m})\mathbf{B}$	Eq. (4.3b)
$\mathbf{A}$	$= \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$	Eq. (4.4)
$\mathbf{P}$ (in Sect. 4)	Matrix for transforming $\mathbf{J}$ for Lemma 4	Eq. (4.4)
$\lambda_{\max}$	Leading eigenvalue of $\mathbf{J}$, possibly adjusted for repeated eigenvalues	Eq. 4.5a)
$\mathbf{B}_{xx}, \mathbf{B}_{xy}, \mathbf{B}_{yx}, \mathbf{B}_{yy}$	$\begin{pmatrix} \mathbf{B}_{xx} & \mathbf{B}_{xy} \\ \mathbf{B}_{yx} & \mathbf{B}_{yy} \end{pmatrix} = \mathbf{B}$	Eq. (5.2c)
$\mathbf{A}_x$	$= \mathbf{P}\mathbf{J}_x\mathbf{P}^{-1} = \mathbf{P}\text{diag}(\widehat{m}_x)\mathbf{B}_{xx}\mathbf{P}^{-1}$	Eq. (5.4c)
$\mathbf{P}$ (in Sect. 5)	Matrix for transforming $\mathbf{J}_x$ for Lemma 4 with $\mathbf{A} = \mathbf{A}_x$	Eq. (5.4c)
$\mathbf{U}$	$= \mathbf{B}_{yx}\mathbf{P}^{-1}$	Eq. (5.4c)
$\mathbf{J}_y$	$= \mathbf{B}_{yy} - \mathbf{B}_{yx}\mathbf{B}_{xx}^{-1}\mathbf{B}_{xy}$	Eq. (5.4c)
$\tilde{\mathbf{A}}$	$= \begin{pmatrix} \mathbf{A}_x & 0 \\ d\mathbf{U} & d\mathbf{J}_y \end{pmatrix}$	Eq. (5.5a)
$\tilde{\lambda}_{\max}$	Leading eigenvalue of $\frac{1}{2}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^{\mathrm{T}})$	Eq. (5.5b)
$\mathbf{Q}$	$= \begin{pmatrix} \mathbf{P_x} & \mathbf{P_x}\mathbf{B}_{xx}^{-1}\mathbf{B}_{xy} \\ \mathbf{0} & \mathbf{I_y} \end{pmatrix}$	Eq. (5.9b)

3 Linear approximation of the fitness-generating function

3.1 Basic idea

The root of the LV-approximability is the exchangeability axiom (iv) combined with the smoothness axiom (i). Under the exchangeability axiom (iv), the fitness-generating function does not distinguish individuals with identical phenotypes. Hence, the function responds only to the sum of their densities. Under the smoothness axiom (i), this property is approximately inherited by slightly different phenotypes; the fitness-generating function responds primarily to the sum of their densities. In the remainder of this paper, we will work out how to lowest order of approximation the fitness-generating function responds linearly to the separate contributions to this sum, leading to the LV-approximation.

To get a more specific picture, we first suppose that there exist only two phenotypes, a resident phenotype s_1 and a mutant phenotype s_2, with population densities n_1 and n_2, respectively, and with their phenotypic difference given by the mutational step size, $|s_2 - s_1| = \varepsilon_\mu$, with ε_μ being small. Proposition 1 trivially holds when the deviations of n_1 and n_2 from their initial states $\widehat{n'} = (\widehat{n}_1, \widehat{n}_2)^{\mathrm{T}}$ are both small during the transient following mutant invasion. In many cases, however, those changes are large, resulting in the exclusion of the resident (Dercole and Rinaldi 2008, Appendix B). In the latter case, it is not obvious whether a linear approximation of the fitness-generating function in $\mathbf{n'} = (n_1, n_2)^{\mathrm{T}}$ is valid.

On the other hand, as the mutant is similar to the resident, due to the smoothness and exchangeability property of the fitness-generating function, they act almost like a single phenotype in their effect on the environment. Thus, invasion by the mutant in many cases causes only a slight change in their total population density $n_1 + n_2$, and only their fractions may change substantially, but will do so slowly (Dercole and Rinaldi 2008, Appendix B; Meszéna et al. 2005; Durinx et al. 2008). In other words, the fitness-generating function is not sensitive to even large changes of n_1 and n_2, as long as $n_1 + n_2$ is kept almost constant. As shown later, this implies that the change of $F(s_i; \mathbf{s'}; \mathbf{n'})$ induced by a large change of n_2, keeping $n_1 + n_2$ constant, is slight, so that $F(s_i; \mathbf{s'}; \mathbf{n'})$ can be expanded with respect to $\mathbf{m'} = (m_1, m_2)^{\mathrm{T}} = (n_1 + n_2, \varepsilon_\mu n_2)^{\mathrm{T}}$, even for $\varepsilon_\mu \to 0$. The linear relationship between $\mathbf{m'}$ and $\mathbf{n'}$ then makes Proposition 1 hold: as the change of $m_2 = \varepsilon_\mu n_2$ is always small because of the smallness of ε_μ, this is the case whenever the change of the population density $m_1 = n_1 + n_2$ is small. Below, we introduce the notion of approximate phenotypes, so we can abbreviate the preceding condition by stating that the change in the population density $m_1 = n_1 + n_2$ of the approximate phenotype ($s_a = s_1$ or $s_a = s_2$) is small.

The strategy above is readily extended to multiple residents $s_1, \ldots, s_N$ and a mutant s_{N+1} emerged from the parental phenotype s_N with $|s_{N+1} - s_N| = \varepsilon_\mu$, by choosing an approximate phenotype from each of the existing phenotypic clusters, so that density changes of those approximate phenotypes can be kept small during the transient following mutant invasion, and thus an LV-approximation can be warranted (Sect. 3). We can gauge the smallness of their density changes from the leading eigenvalue of the community's Jacobian matrix evaluated at the equilibrium population densities of

the approximate phenotypes before the invasion (Sect. 4). However, this linear stability analysis does not work well when some approximate phenotypes have very small initial equilibrium densities, because those small densities inevitably cause the leading eigenvalue of the community's Jacobian matrix to be close to zero. To overcome this difficulty, we analyze not only the linear terms but also the quadratic terms of the transient dynamics around the initial equilibrium (Sect. 5). In the remainder of this section, we show how we can easily find approximate phenotypes for a set of phenotypes $\mathbf{s}' = (s_1, \ldots, s_{N+1})^T$, such that Proposition 1 holds when the changes of the population densities of these approximate phenotypes are sufficiently small.

3.2 Approximate phenotypes

We consider an arbitrary set of residents together with a mutant, $\mathbf{s}' = (s_1, \ldots, s_N, s_{N+1})^T$. We choose phenotypic clusters so that within-cluster phenotypic differences do not exceed $\varepsilon = \rho_\mu \varepsilon_\mu$ (Fig. 1a), with an arbitrarily chosen constant ρ_μ larger than 1 (but not too large, so that the clustering is meaningful, i.e., the error estimates to be derived below are small). We assume that those phenotypic clusters are well-recognizable and well-separated from each other, so that we can find an ε that is much smaller than the smallest distance among the approximate phenotypes. Generally, this assumption is warranted in evolutionary dynamics with small mutational step sizes (as explained in Sect. 9.2) by the principle of limiting similarity. Notice that in any case the mutant s_{N+1} and its parental phenotype s_N form a cluster. Any resident not similar to any other phenotype forms a cluster by itself. Thus, the number of clusters, denoted by M, satisfies $1 \leq M \leq N$. From each cluster, we arbitrarily pick one phenotype as its representative. Then, by the symmetry axiom (ii), we can permute $\mathbf{s}' = (s_1, \ldots, s_N, s_{N+1})^T$ so that those representatives come first as $s_1, \ldots, s_M$, followed by the other phenotypes, i.e., $\mathbf{s}' = (s_1, \ldots, s_M, s_{M+1}, \ldots, s_{N+1})^T$ (Fig. 1b). We refer to those representatives as approximate phenotypes $\mathbf{s}_a = (s_1, \ldots, s_M)^T$.

We introduce the cluster-identifying function cid, such that $\text{cid}(j) = i$ means that phenotype s_j belongs to the ith cluster, with $s_{\text{cid}(j)}$ as the representative—i.e., approximate—phenotype of that cluster, and $\text{cid}(j) = j$ for $j \leq M$. We also introduce the component-identifying function com, which returns the set of indices of the phenotypes comprising the i-th cluster, i.e., $\text{com}(i) = \{ j | \text{cid}(j) = i \}$. Then, the population densities of these clusters are given by a vector $\mathbf{m} = (m_1, \ldots, m_M)^T$, with the population densities

$$m_i := \sum_{j \in \text{com}(i)} n_j \qquad (3.1a)$$

for $i = 1, \ldots, M$ treated as belonging to the approximate phenotypes $\mathbf{s}_a = (s_1, \ldots, s_M)^T$ (Fig. 1c). While the approximate phenotype of the ith cluster is identical to the representative phenotype of that cluster, the population densities of the former and latter are different and given by m_i and n_i, respectively.

Notice that the number M of approximate phenotypes is less than the number $N+1$ of phenotypes in the original community dynamics. Thus, for expanding the fitness-
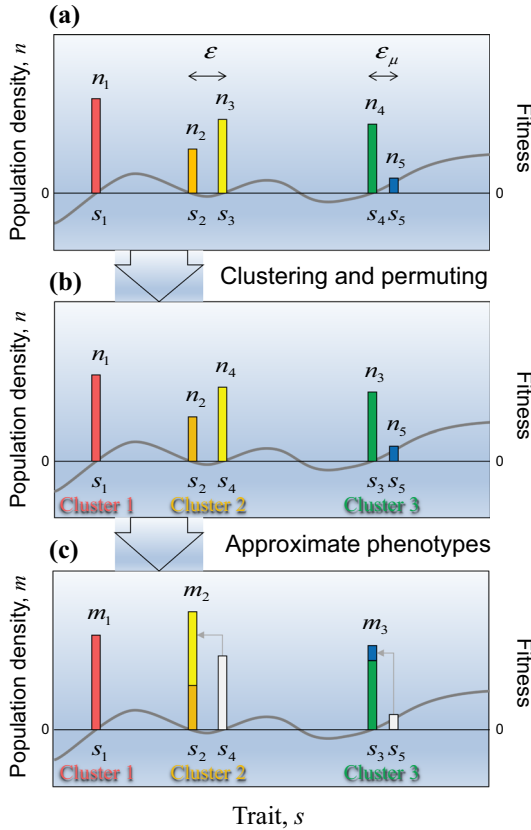
Fig. 1 Construction of approximate phenotypes and their population densities. **a** The population densities $\mathbf{n}' = (n_1, n_2, n_3, n_4, n_5)^{\mathrm{T}}$ of existing phenotypes $\mathbf{s}' = (s_1, s_2, s_3, s_4, s_5)^{\mathrm{T}}$—comprising four residents s_1, s_2, s_3, s_4, and a mutant s_5—are indicated by colored histogram bars. The thick gray curve shows the fitness landscape, which passes through 0 at the resident phenotypes. The existing phenotypes are clustered so that within-cluster phenotypic differences do not exceed the threshold ε, which is chosen to be larger than the mutational step size ε_μ, so that the mutant s_5 and its parental resident s_4 are guaranteed to be part of the same cluster. **b** The existing phenotypes are permuted so that the approximate phenotypes s_1, s_2, s_3 come first. **c** Within each cluster, an approximate phenotype is chosen to represent the cluster. The population densities m_1, m_2, m_3 of the approximate phenotypes are assigned so that each equals the total population density of the corresponding cluster (color figure online)

generating function, we need to define the other $(N - M + 1)$ variables in such a way that their changes stay small during the transient following mutant invasion. As long as the population densities of the approximate phenotypes are kept almost constant, the fitness-generating function is expected to be insensitive to n_i for all $i = 1, \ldots, N+1$. Thus, we describe the remaining degrees of freedom, $m_{M+1}, \ldots, m_{N+1}$, by

$$m_i := \varepsilon n_i \tag{3.1b}$$

for $i = M + 1, \ldots, N + 1$. Combining Eqs. (3.1a) and (3.1b), we write $\mathbf{m}' = (m_1, \ldots, m_M, m_{M+1}, \ldots, m_{N+1})^{\mathrm{T}}$, which has the same dimension as $\mathbf{n}'$. Then, by the smoothness axiom (i), the exchangeability axiom (iv), and the bounded-world axiom (v), we have

Lemma 1 *With $\breve{F}(s_i; \mathbf{s}'; \mathbf{m}') := F(s_i; \mathbf{s}'; \mathbf{n}')$, for sufficiently small ε there exists a constant C'_{Fm} such that*

$$\left| \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} \right| \leq C'_{\mathrm{Fm}} \tag{3.2}$$

for all $i, j = 1, \ldots, N + 1$, $\mathbf{n}' \in [0, \eta]^{N+1}$, and any $\mathbf{s}'$ such that $\left| s_j - s_{\mathrm{cid}(j)} \right| \leq \varepsilon$.

See Appendix A for the proof. Although $\breve{F}(s_i; \mathbf{s}'; \mathbf{m}')$ differs from $F(s_i; \mathbf{s}'; \mathbf{n}')$ as a mathematical object, their biological meaning is the same. Lemma 1 thus ensures the expandability of $F(s_i; \mathbf{s}'; \mathbf{n}')$ in terms of $\mathbf{m}'$. The estimate C'_{Fm} still depends on $\mathbf{s}'$, but is positive and uniformly bounded away from 0 and ∞.

As we did for C'_{Fm}, below we will introduce bounds for other important variables and functions in the form of expressions $C_{\cdot}$ that are independent of population densities (but may be functions of other model parameters). Please notice that here we have introduced the notational convention, to which we adhere throughout this paper, that $C_{F\cdot}$ denotes the upper bound for the absolute value (or norm) of $\cdot$ the derivative of the fitness function with respect to $\cdot$, while $C_{\cdot}$ denotes the upper bound for the absolute value (or norm) of $\cdot$ or for the derivative of the first symbol in $\cdot$ with respect to the subsequent symbol(s). All $C_{F\cdot}$ and $C_{\cdot}$ are positive and uniformly bounded away from 0 and ∞. In the propositions below, we just indicate that such constants exist. Expressions for determining their values are derived in the associated appendices and are shown in Table 2.

3.3 Taylor expansion in the population densities of the approximate phenotypes

We now expand the fitness-generating function in $\mathbf{m}'$. We denote by $\widehat{m}' = (\widehat{m}_1, \ldots, \widehat{m}_{N+1})^{\mathrm{T}}$ the initial state $\widehat{n}' = (\widehat{n}_1, \ldots, \widehat{n}_{N+1})$ expressed in terms of $\mathbf{m}'$, with $\widehat{m}_i = \sum_{j \in \mathrm{com}(i)} \widehat{n}_j$ for $i = 1, \ldots, M$ and $\widehat{m}_i = \varepsilon \widehat{n}_i$ for $i = M + 1, \ldots, N + 1$. Lemma 1 allows $F(s_i; \mathbf{s}'; \mathbf{n}')$ to be expanded in $\mathbf{m}'$ around $\widehat{n}'$ as

$$F(s_i; \mathbf{s}'; \mathbf{n}') = \widehat{F}_i + \mathbf{b}_i'^{\mathrm{T}}(\mathbf{m}' - \widehat{m}') + R_i, \tag{3.3a}$$

where

$$\widehat{F}_i := F(s_i; \mathbf{s}'; \widehat{n}'), \tag{3.3b}$$

Table 2 List of constants describing the upper bounds for variables dependent on population densities and (or) phenotypes. Notation: C_F denotes the upper bound for the absolute value (or norm) of the derivative of the fitness function with respect to $\cdot$, while C denotes the upper bound for the absolute value (or norm) of $\cdot$ or for the derivative of the first symbol in $\cdot$ with respect to the subsequent symbol(s)

Constant and formula	Location
Constants used in formulas below[a]	
$C'_{Fn} = \max\left\{ \left\| \dfrac{\partial F(s_j;\mathbf{s}';\mathbf{n}')}{\partial n_j} \right\| \;\middle|\; i,j=1,\ldots,N+1,\ \mathbf{n}' \in [0,\eta]^{N+1} \right\}$	Eq. (A.2)[b]
$C'_{Fsn} = \max\left\{ \left\| \dfrac{\partial^2 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_j \partial n_j} - \dfrac{\partial^2 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_j \partial n_{\mathrm{cid}(j)}} \right\|_{s_j = s_{jT}} \;\middle|\; \begin{array}{l} i=1,\ldots,N+1, \\ j=M+1,\ldots,N+1, \\ \mathbf{n}' \in [0,\eta]^{N+1}, \\ s_{jT} \in [s_j, s_{\mathrm{cid}(j)}] \end{array} \right\}$	Eq. (A.12)
$\widehat{C}'_{Fz} = \max\left\{ \left\| \dfrac{\widehat{\partial F(z_j;\mathbf{s}';\mathbf{n}')}}{\partial z_j} \right\|_{z_j = s_{jT}} \;\middle|\; j=M+1,\ldots,N+1,\ s_{jT} \in [s_j, s_{\mathrm{cid}(j)}] \right\}$	Eq. (B.12)
$C'_{Fnn} = \max\left\{ \left\| \dfrac{\partial F(s_j;\mathbf{s}';\mathbf{n}')}{\partial n_k \partial n_j} \right\| \;\middle|\; i,j,k=1,\ldots,N+1,\ \mathbf{n}' \in [0,\eta]^{N+1} \right\}$	Eq. (B.17)
$C'_{Fsnn} = \max\left\{ \left\| \dfrac{\partial^3 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_j \partial n_k \partial n_j} - \dfrac{\partial^3 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_j \partial n_k \partial n_{\mathrm{cid}(j)}} \right\|_{s_j = s_{jT}} \;\middle|\; \begin{array}{l} i=1,\ldots,N+1, \\ j=M+1,\ldots,N+1, \\ k=1,\ldots,M, \\ \mathbf{n}' \in [0,\eta]^{N+1}, \\ s_{jT} \in [s_j, s_{\mathrm{cid}(j)}] \end{array} \right\}$	Eq. (B.26)
$C'_{Fssnn} = \max\left\{ \left\| \dfrac{\partial^4 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_k \partial s_j \partial n_k \partial n_j} - \dfrac{\partial^4 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_k \partial s_j \partial n_k \partial n_{\mathrm{cid}(j)}} - \dfrac{\partial^4 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_k \partial s_j \partial n_{\mathrm{cid}(k)} \partial n_j} + \dfrac{\partial^4 F(s_j;\mathbf{s}';\mathbf{n}')}{\partial s_k \partial s_j \partial n_{\mathrm{cid}(k)} \partial n_{\mathrm{cid}(j)}} \right\|_{s_k = s_{kT}, s_j = s_{jT}} \;\middle|\; \begin{array}{l} i,j,k=1,\ldots,N+1, \\ \mathbf{n}' \in [0,\eta]^{N+1}, \\ s_{jT} \in [s_j, s_{\mathrm{cid}(j)}], \\ s_{kT} \in [s_k, s_{\mathrm{cid}(k)}] \end{array} \right\}$	Eq. (B.28)
$C'_{F\varepsilon} = \max\left\{ \left\| \dfrac{\partial F(s_{\mathrm{cid}(j)} + \varepsilon \rho_j; \mathbf{s}_a + \varepsilon \boldsymbol{\rho}'; \mathbf{n}')}{\partial \varepsilon} \right\|_{\varepsilon = \varepsilon_{Tj}} \;\middle|\; \begin{array}{l} j=1,\ldots,N+1, \\ \mathbf{n}' \in [0,\eta]^{N+1}, \\ \varepsilon_{Tj} \in [0,\varepsilon] \end{array} \right\}$	Eq. (C.3)

Table 2 continued

Constant and formula	Location
$C_{\mathbf{Fmm}} = \max \left\{ \left\| \dfrac{\partial^2 F(s_i; \mathbf{s_a}; \mathbf{m})}{\partial \mathbf{m}\, \partial \mathbf{m}^{\mathsf{T}}} \right\|_Q \;\middle|\; i = 1, \ldots, M,\ \mathbf{m} \in [0, \eta]^M \right\}$	Eq. (C.12)
$C_{F\varepsilon} = \max \left\{ \dfrac{\partial F(s_{\mathrm{cid}(j)} + \varepsilon\rho_j; \mathbf{s_a}' + \varepsilon\boldsymbol{\rho}'; \mathbf{n}')}{\partial \varepsilon} \bigg|_{\varepsilon=\varepsilon_{\mathrm{T}j}} \;\middle|\; \begin{array}{l} j = 1, \ldots, M, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \\ \varepsilon_{\mathrm{T}j} \in [0, \varepsilon] \end{array} \right\}$	Eq. (J.21)
$C_{\mathbf{F}\varepsilon\mathbf{m}} = \max \left\{ \left\| \dfrac{\partial \mathbf{F}_\varepsilon'}{\partial \mathbf{m}} \right\| \;\middle|\; \begin{array}{l} \varepsilon_{\mathrm{T}i} \in [0, \varepsilon], \\ m_1, \ldots, m_M \in [0, \eta], \\ m_{M+1}, \ldots, m_{N+1} \in [0, \varepsilon\eta] \end{array} \right\}$	Eq. (J.22)
with $\mathbf{F}_\varepsilon' = \begin{pmatrix} F_{\varepsilon 1} \\ \ldots \\ F_{\varepsilon, N+1} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \tilde{F}(s_{\mathrm{cid}(1)} + \varepsilon\rho_1; \mathbf{s_a}' + \varepsilon\boldsymbol{\rho}'; \mathbf{m}')}{\partial \varepsilon} \\ \ldots \\ \dfrac{\partial \tilde{F}(s_{\mathrm{cid}(N+1)} + \varepsilon\rho_{N+1}; \mathbf{s_a}' + \varepsilon\boldsymbol{\rho}'; \mathbf{m}')}{\partial \varepsilon} \end{pmatrix}$	
$C_{Fzz}' = \max \left\{ \left\| \dfrac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z^2} \right|_{z=s_{j\mathrm{T}}} \;\middle|\; \begin{array}{l} |\Delta\mathbf{m}'| \in [0, \varepsilon C_{\mathbf{m}}'], \\ j = M+1, \ldots, N+1, \\ s_{j\mathrm{T}} \in [s_j, s_{\mathrm{cid}(j)}] \end{array} \right\}$	Eq. (K.6)
$C_{Fzn}' = \max \left\{ \left\| \dfrac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z\, \partial n_j} \right|_{z=s_{j\mathrm{T}}} \;\middle|\; \begin{array}{l} j = 1, \ldots, N+1, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \\ s_{j\mathrm{T}} \in [s_{\mathrm{cid}(j)}, s_j] \end{array} \right\}$	Eq. (K.11)
$C_{Fzsn}' = \max \left\{ \left\| \dfrac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z\, \partial s_j\, \partial n_j} - \dfrac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z\, \partial s_j\, \partial n_{\mathrm{cid}(j)}} \right|_{z=s_{j\mathrm{T}},\, s_j=s_{j\mathrm{TT}}} \;\middle|\; \begin{array}{l} j = 1, \ldots, N+1, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \\ s_{j\mathrm{T}} \in [s_{\mathrm{cid}(j)}, s_j], \\ s_{j\mathrm{TT}} \in [s_{\mathrm{cid}(j)}, s_j] \end{array} \right\}$	Eq. (K.13)

Table 2 continued

Constant and formula	Location and proof
Constants defined based on definitions above	
$C'_{\mathbf{Fm}} = \max\left\{C'_{\mathrm{Fn}}, C'_{\mathrm{Fsn}}\right\}$	Eq. (3.2) Appendix A
$C'_{\mathbf{Fmm}} = \max\left\{C'_{\mathrm{Fnn}}, C'_{\mathrm{Fsnn}}, C'_{\mathrm{Fssnn}}\right\}$	Eq. (B.29) Appendix B.4
$C'_{\mathbf{Fmm}} = (N+1)C'_{\mathbf{Fmm}}$	Eq. (B.31) Appendix B.4
$C_{\mathbf{m}} = \dfrac{2\left\|\mathbf{P}^{-1}\right\| C_{\mathbf{h}}}{\|\lambda_{\max}\|}$ (in Theorem 2)	Eqs. (3.4a), (4.9a) Lemma 6
$C_{\mathbf{m}} = \dfrac{2\alpha\left\|\mathbf{Q}^{-1}\right\| \max\left[\tilde{C}_{\mathbf{h}}, -\frac{1}{2}\tilde{\lambda}_{\max}\rho_{\mathrm{m}}\right]}{\tilde{\lambda}_{\max}} + \rho_{\mathrm{m}}$ (in Theorem 3)	Eqs. (3.4a), (5.9a) Lemma 11
$C'_{\mathbf{m}} = \sqrt{C_{\mathbf{m}}^{2} + (N+1-M)\eta^{2}}$	Eq. (3.4b) Appendix B.3
$C_{\mathbf{rm}} = \|\mathbf{B}\| + \eta\sqrt{M}\,C_{\mathbf{Fmm}}$	Lemma 3 Appendix C.2
$C_{\mathrm{hm}} = \eta\sqrt{M}\left[\widehat{C'_{\mathbf{Fz}}} + 2C'_{\mathbf{F\varepsilon}}\right]$	Lemma 3 Appendix C.2
$C_{\mathbf{r}} = \|\mathbf{P}\|\left\|\mathbf{P}^{-1}\right\|^{2} C_{\mathbf{rm}} = \|\mathbf{P}\|\left\|\mathbf{P}^{-1}\right\|^{2}\left[\|\mathbf{B}\| + \eta\sqrt{M}\,C_{\mathbf{Fmm}}\right]$	Eq. (4.4) Appendix D
$C_{\mathbf{h}} = \|\mathbf{P}\|C_{\mathbf{hm}} = \|\mathbf{P}\|\eta\sqrt{M}\left[\widehat{C'_{\mathbf{Fz}}} + 2C'_{\mathbf{F\varepsilon}}\right]$	Eq. (4.4) Appendix D
$C_{\mathbf{rf}} = \sqrt{M}\,C_{\mathbf{Fmm}}$	Eq. (5.1) Appendix C.1
$C_{\mathbf{hf}} = \sqrt{M}\left[\widehat{C'_{\mathbf{Fz}}} + 2C'_{\mathbf{F\varepsilon}}\right]$	Eq. (5.1) Appendix C.1

Table 2 continued

Constant and formula	Location and proof
$\tilde{C}_{\mathbf{rf}} = C_{\mathbf{rf}}$	Eq. (F.6) AppendixF.1
$\tilde{C}_{\mathbf{hf}} = C_{\mathbf{hf}} + \|\mathbf{B}\| \varrho_{\mathrm{m}} + C_{\mathbf{rf}}[2\eta + \varepsilon \varrho_{\mathrm{m}}] \varrho_{\mathrm{m}}$	Eq. (F.6) Appendix F.1
$\tilde{C}_{\mathbf{rx}} = \|\mathbf{P}\|(1 + \|\mathbf{T}\|) \left\|\mathbf{Q}^{-1}\right\|^2 \left[\|\mathbf{B}\| + \eta \tilde{C}_{\mathbf{rf}}\right]$	Eq. (5.4a) Appendix F.4
$\tilde{C}_{\mathbf{ry}} = \tilde{C}_{\mathbf{rf}} \left\|\mathbf{Q}^{-1}\right\|^2$	Eq. (5.4a) Appendix F.4
$\tilde{C}_{\mathbf{hx}} = \|\mathbf{P}\|(1 + \|\mathbf{T}\|) \eta \tilde{C}_{\mathbf{hf}}$	Eq. (5.4a) Appendix F.4
$\tilde{C}_{\mathbf{hy}} = \tilde{C}_{\mathbf{hf}}$	Eq. (5.4a) Appendix F.4
$\tilde{C}_{\mathbf{r}} = \sqrt{\tilde{C}_{\mathbf{rx}}^2 + d^2 \tilde{C}_{\mathbf{ry}}^2}$	Eq. (5.6c) Appendix F.5
$\tilde{C}_{\mathbf{h}} = \sqrt{\tilde{C}_{\mathbf{hx}}^2 + d^2 \tilde{C}_{\mathbf{hy}}^2}$	Eq. (5.6c) Appendix F.5
$\widehat{C}_{\mathbf{h}} = \left\|\mathbf{P}\mathrm{diag}(\widehat{m})\mathbf{h}_{\mathbf{f}}\right\|_{\mathbf{m}=\widehat{m}}$	Eq. (6.2) Appendix J.2
$C_{\mathbf{H}} = \|\mathbf{P}\| \left\|\mathbf{P}^{-1}\right\| \left[2C_{F\varepsilon} + (N+1)\eta C_{F\varepsilon m}\right]$	Eq. (6.2) Appendix J.2
$C'_{F\mathbf{z}m} = \sqrt{MC_{F\mathbf{z}n}'^2 + (N+1-M)C_{F\mathbf{z}sn}'^2}$	Eq. (8.2b) Appendix K

[a]Tighter estimates that apply under various restrictions are presented in the appendices

[b]Here $\|\mathbf{B}\|_Q := \max\left\{\left\|\mathbf{v}^{\mathbf{T}}\mathbf{B}\mathbf{v}\right\| \, \big| \, |\mathbf{v}| = 1\right\}$

$$\mathbf{b}_i'^{\mathrm{T}} = \left(b_{i1}', \ldots, b_{i\,N+1}' \right) := \left. \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}'} \right|_{\mathbf{m}' = \widehat{m}'} = \left. \left(\frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_1'}, \ldots, \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_{N+1}'} \right) \right|_{\mathbf{m}' = \widehat{m}'},$$

$$(3.3c)$$

and

$$R_i := \frac{1}{2} (\mathbf{m}' - \widehat{m}')^{\mathrm{T}} \left. \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\mathrm{T}}} \right|_{\mathbf{m}' = \widehat{m}'} (\mathbf{m}' - \widehat{m}') + \text{h.o.t.}, \tag{3.3d}$$

with

$$\left. \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\mathrm{T}}} \right|_{\mathbf{m}' = \widehat{m}'} = \begin{pmatrix} \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_1'^2} & \cdots & \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_1' \partial m_{N+1}'} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_1' \partial m_{N+1}'} & \cdots & \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_{N+1}'^2} \end{pmatrix}_{\mathbf{m}' = \widehat{m}'}. \tag{3.3e}$$

Here $\widehat{F}_i = 0$ for $i = 1, \ldots, N$ (from the equilibrium equation of the residents), and

$$\left| \widehat{F}_{N+1} \right| \leq \varepsilon \widehat{C}_{\mathrm{Fz}}' \tag{3.3f}$$

(Appendix B.2). Moreover, by the bounded-world axiom (v) and Taylor's theorem, we have

Lemma 2 *If there exists a constant $C_{\mathbf{m}}$ such that*

$$\left| \mathbf{m} - \widehat{m} \right| \leq \varepsilon C_{\mathbf{m}}, \tag{3.4a}$$

then there exist constants $C_{\mathbf{m}}' = \sqrt{C_{\mathbf{m}}^2 + (N - M + 1)\eta^2}$ and C_{Fmm}' satisfying

$$\left| \mathbf{m}' - \widehat{m}' \right| \leq \varepsilon C_{\mathbf{m}}' \tag{3.4b}$$

and

$$|R_i| \leq \tfrac{1}{2} C_{\mathrm{Fmm}}' \left| \mathbf{m}' - \widehat{m}' \right|^2 \leq \tfrac{1}{2} \varepsilon^2 C_{\mathrm{Fmm}}' C_{\mathbf{m}}'^2, \tag{3.4c}$$

where for vectors $|\cdot|$ denotes the Euclidian norm.

See Appendix B.3 for the proof. Thus, if Eq. (3.4a) is satisfied for a sufficiently small ε, the fitness-generating function is approximated well by a linear function of $\mathbf{m}'$.

3.4 Taylor expansion in the population densities of the original phenotypes

Next, we transform the term linear in $\mathbf{m}'$ in Eq. (3.3a) into one in $\mathbf{n}'$. As $\mathbf{m}'$ is a linear function of $\mathbf{n}'$, $\mathbf{m}'$ can be written as $\mathbf{m}' = \mathbf{W}\mathbf{n}'$, where $\mathbf{W}$ is a $(N + 1)$-by-$(N + 1)$ matrix with components given by Eq. (3.1a). Therefore, substituting this relationship into Eq. (3.3a) gives

$$F(s_i; \mathbf{s}'; \mathbf{n}') = \widehat{F}_i + \mathbf{b}_i'^{\mathrm{T}} \mathbf{W}(\mathbf{n}' - \widehat{n'}) + R_i$$

$$= \widehat{F}_i + \mathbf{a}_i^{\mathrm{T}}(\mathbf{n}' - \widehat{n'}) + R_i, \tag{3.5a}$$

where, since $\breve{F}(s_i; \mathbf{s}'; \mathbf{m}') = F(s_i; \mathbf{s}'; \mathbf{n}')$,

$$\mathbf{a}_i^{\mathrm{T}} = \left(a_{i,1}, \ldots, a_{i,N+1}\right) := \mathbf{b}_i'^{\mathrm{T}} \mathbf{W}$$

$$= \left. \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}'} \right|_{\mathbf{m}'=\widehat{m'}} \mathbf{W}$$

$$= \left. \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial \mathbf{n}'} \right|_{\mathbf{n}'=\widehat{n'}}$$

$$= \left. \left(\frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_1'}, \ldots, \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_{N+1}'} \right) \right|_{\mathbf{n}'=\widehat{n'}}. \tag{3.5b}$$

By combining the equations above with Lemma 2, and by using Eq. (2.3), we get

Theorem 1 *For the population densities $\mathbf{m} = (m_1, \ldots, m_M)^{\mathrm{T}}$ of approximate phenotypes $\mathbf{s}_{\mathrm{a}} = (s_1, \ldots, s_M)^{\mathrm{T}}$ formed by clustering resident and mutant phenotypes $\mathbf{s}' = (s_1, \ldots, s_{N+1})^{\mathrm{T}}$ according to a threshold phenotypic distance $\varepsilon = \varepsilon_\mu \rho_\mu$, if $\mathbf{m}$ satisfies*

$$\left| \mathbf{m} - \widehat{m} \right| \leq \varepsilon C_{\mathbf{m}},$$

during the transient following mutant invasion, then the fitness-generating function can be expanded as

$$F(s_i; \mathbf{s}'; \mathbf{n}') = \widehat{F}_i + \sum_{j=1}^{N+1} a_{ij}(n_j - \widehat{n}_j) + R_i \tag{3.6a}$$

with $|R_i| \leq C'_{\mathbf{Fmm}} \left| \mathbf{m}' - \widehat{m'} \right|^2 \leq \varepsilon^2 C'_{\mathbf{Fmm}} C'_{\mathbf{m}}$, which gives the LV-approximation

$$\frac{dn_i}{dt} = n_i \left[\gamma_i + \sum_{j=1}^{N+1} a_{ij} n_j + R_i \right] \tag{3.6b}$$

with $\gamma_i = \widehat{F}_i - \sum_{j=1}^{N+1} a_{ij} \widehat{n}_j$.

4 Approximability condition when the population densities of the approximate phenotypes are large

In this section, we consider the sufficient condition in Eq. (3.4a) for LV-approximability, $\left| \mathbf{m} - \widehat{m} \right| < \varepsilon C_{\mathbf{m}}$. We refer to this as the approximability condition. If the initial equilibrium population densities of approximate phenotypes are not small, so that $\widehat{m}_i \gg \varepsilon$ is satisfied for all $i = 1, \ldots, M$, their dynamics can be analyzed by a linear stability analysis of the resident equilibrium. On the other hand, if some approximate phenotypes have very small equilibrium population densities, also the corresponding eigenvalues of the associated Jacobian come very close to zero and examining linear terms alone is not sufficient. In this section, we analyze the first, simpler, case to show that the approximability condition (3.4a) can generally be fulfilled. The second, more complicated, case is then analyzed in a similar manner in the next section.

4.1 Dynamics of approximate phenotypes

The dynamics of approximate phenotypes $\mathbf{m} = (m_1, \ldots, m_M)^{\mathrm{T}}$ satisfies, by Eqs. (3.1a) and (2.3),

$$\frac{dm_i}{dt} = m_i f(s_i; \mathbf{s}'; \mathbf{n}') \tag{4.1a}$$

for $i = 1, \ldots, M$, where the growth rate of m_i,

$$f(s_i; \mathbf{s}'; \mathbf{n}') := \sum_{j \in \mathrm{com}(i)} p_j F(s_j; \mathbf{s}'; \mathbf{n}'), \tag{4.1b}$$

is the average growth rate within the ith cluster weighted with the fractions $p_j := n_j / m_{\mathrm{cid}(j)}$ of its component phenotypes. As for the remaining degrees of freedom in $\mathbf{m}'$, i.e., $m_i = \varepsilon n_i$ for $i = M + 1, \ldots, N + 1$, their dynamics are given by

$$\frac{dm_i}{dt} = m_i F(s_i; \mathbf{s}'; \mathbf{n}'). \tag{4.1c}$$

When **m** is kept constant, these remaining degrees of freedom describe the relatively slow dynamics of the cluster compositions, corresponding to the dynamics of the fractions p_j.

4.2 Transformation into perturbed community

For convenience, Eq. (4.1a) is rewritten in vector–matrix form as

$$\frac{d\mathbf{m}}{dt} = \text{diag}(\mathbf{m})\mathbf{f}(\mathbf{s}_a; \mathbf{s}'; \mathbf{n}'), \tag{4.2}$$

where $\text{diag}(\mathbf{m})$ is a diagonal matrix with diagonal entries $m_1, \ldots, m_M$, $\mathbf{s}_a = (s_1, \ldots, s_M)^T$, and $\mathbf{f}(\mathbf{s}_a; \mathbf{s}'; \mathbf{n}') := \left(f(s_1; \mathbf{s}'; \mathbf{n}'), \ldots, f(s_M; \mathbf{s}'; \mathbf{n}') \right)^T$. We decompose the right-hand side of Eq. (4.2) into a component determined by **m** alone and a residual of order ε, which is treated as a perturbation. The former component is further decomposed into linear and higher-order terms. Specifically, we have

Lemma 3 *The dynamics of* **m** *in Eq.* (4.2) *can be transformed into*

$$\frac{d\mathbf{m}}{dt} = \mathbf{J}(\mathbf{m} - \widehat{m}) + \mathbf{r}_m \left| \mathbf{m} - \widehat{m} \right|^2 + \varepsilon \mathbf{h}_m, \tag{4.3a}$$

where

$$\mathbf{J} := \text{diag}(\widehat{m})\mathbf{B} = \begin{pmatrix} \widehat{m}_1 b_{11} & \ldots & \widehat{m}_1 b_{1M} \\ \ldots & \ldots & \ldots \\ \widehat{m}_M b_{M1} & \ldots & \widehat{m}_M b_{MM} \end{pmatrix},$$
$$b_{ij} := \left. \frac{\partial F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial m_j} \right|_{\mathbf{m} = \widehat{m}}, \tag{4.3b}$$

and $\mathbf{r}_m = (r_{m1}, \ldots, r_{mM})^T$ *is a function of* **m** *satisfying* $|\mathbf{r}_m| \leq C_{\mathbf{rm}}$, *while* $\mathbf{h}_m = (h_{m1}, \ldots, h_{mM})^T$ *is a function of* **m'** *and* ε *satisfying* $|\mathbf{h}_m| \leq C_{\mathbf{hm}}$.

See Appendix C for the proof. Notice that **J** and $\mathbf{r}_m$ are both independent of ε.

4.3 Local Lyapunov function

If the perturbation term is neglected in Eq. (4.3a), i.e., $\varepsilon = 0$, we can easily examine the local stability of the fixed point $\widehat{m}$ by checking whether all eigenvalues of **J** have negative real parts. With the perturbation, however, we also have to compare the magnitudes of those eigenvalues with the perturbation. Moreover, as the perturbation causes a deviation of the community from $\widehat{m}$, the effect of the higher-order term $\mathbf{r}_m$ $\left| \mathbf{m} - \widehat{m} \right|^2$ has to be examined as well.

To simplify the analysis, we introduce a new vector $\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{T}} = \mathbf{P}(\mathbf{m} - \widehat{m})$ with a real matrix $\mathbf{P}$ and write Eq. (4.3a) as

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{r}|\mathbf{x}|^2 + \varepsilon\mathbf{h}, \tag{4.4}$$

with $\mathbf{A} := \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$, $\mathbf{r} := \mathbf{P}\mathbf{r}_m|\mathbf{P}^{-1}\mathbf{x}|^2/|\mathbf{x}|^2$, $\mathbf{h} := \mathbf{P}\mathbf{h}_m$, $|\mathbf{r}| \leq C_\mathbf{r}$, and $|\mathbf{h}| \leq C_\mathbf{h}$; see Appendix D. As proved in Appendix E, we have

Lemma 4 *A real matrix* $\mathbf{P}$ *can be chosen so that* $\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$ *satisfies*

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \leq \lambda_{\max}|\mathbf{x}|^2 \tag{4.5a}$$

for $\lambda_{\max} < 0$ *with*

$$\lambda_{\max} := \max\{\mathrm{Re}(\lambda_1), \ldots, \mathrm{Re}(\lambda_M)\} \tag{4.5b}$$

(when the eigenvalues $\lambda_1, \ldots, \lambda_M$ *of* $\mathbf{J}$ *in Eq. (4.3b) are all distinct) or*

$$\lambda_{\max} := \max\{\mathrm{Re}(\lambda_1), \ldots, \mathrm{Re}(\lambda_D), \tfrac{1}{2}\mathrm{Re}(\lambda_{D+1}), \ldots, \tfrac{1}{2}\mathrm{Re}(\lambda_M)\} \tag{4.5c}$$

(when some eigenvalues are repeated, with distinct eigenvalues $\lambda_1, \ldots, \lambda_D$ *and repeated eigenvalues* $\lambda_{D+1}, \ldots, \lambda_M$*).*

When the second and third terms in Eq. (4.4) are both neglected, the time derivative of $|\mathbf{x}|^2$ is a monotonically decreasing function if $\lambda_{\max} < 0$, which gives

Lemma 5 *For Eq. (4.4) with* $\varepsilon = 0$*, if*

$$\lambda_{\max} < 0, \tag{4.6a}$$

then

$$V := \sum_{i=1}^{M} x_i^2 = |\mathbf{x}|^2 = \mathbf{x}^{\mathrm{T}}\mathbf{x} \tag{4.6b}$$

is a local Lyapunov function, i.e., $V = 0$ *for* $\mathbf{x} = \mathbf{0}$ *and* $dV/dt < 0$ *for* $0 < |\mathbf{x}| < \phi$ *with a sufficiently small* ϕ*.*

Proof By Eqs. (4.4) and (4.6a), the time derivative of V equals

$$\begin{aligned}
\frac{dV}{dt} &= 2\mathbf{x}^{\mathrm{T}}\frac{d\mathbf{x}}{dt} \\
&= 2\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + 2\mathbf{x}^{\mathrm{T}}\mathbf{r}|\mathbf{x}|^2 + 2\varepsilon\mathbf{x}^{\mathrm{T}}\mathbf{h} \\
&\leq 2\lambda_{\max}|\mathbf{x}|^2 + 2\mathbf{x}^{\mathrm{T}}\mathbf{r}|\mathbf{x}|^2 + 2\varepsilon\mathbf{x}^{\mathrm{T}}\mathbf{h} \\
&\leq 2\lambda_{\max}|\mathbf{x}|^2 + 2C_\mathbf{r}|\mathbf{x}|^3 + 2\varepsilon C_\mathbf{h}|\mathbf{x}|
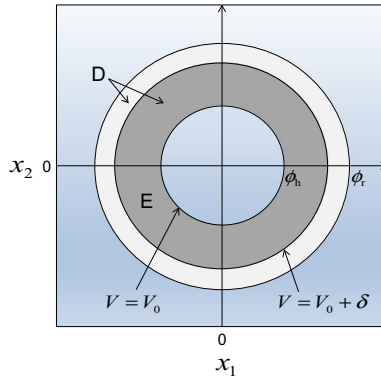\end{aligned}$$

Fig. 2 Stability against perturbation when the initial equilibrium population densities of approximate phenotypes are not small. As a simple example, a community composed of two approximate phenotypes $\mathbf{s}_a = (s_1, s_2)^T$ is considered. Their population densities $\mathbf{m} = (m_1, m_2)^T$ are transformed into $\mathbf{x} = (x_1, x_2)^T$ so that $\mathbf{x} = \mathbf{0}$ corresponds to the initial equilibrium before mutant invasion. A local Lyapunov function $V = |\mathbf{x}|^2$ of the community dynamics monotonically decreases with time within the light-gray and dark-gray regions marked by D. The dark-gray region marked by E is associated with a repeller that prevents the community dynamics from passing its inner boundary $|\mathbf{x}| = \phi_h$ from its inner side $|\mathbf{x}| < \phi_h$

$$= 2C_\mathbf{r}|\mathbf{x}|^2(|\mathbf{x}| - \phi_r) + \lambda_{\max}|\mathbf{x}|(|\mathbf{x}| - \phi_h), \tag{4.7a}$$

where

$$\phi_r := \frac{|\lambda_{\max}|}{2C_\mathbf{r}} > 0,$$
$$\phi_h := \frac{2\varepsilon C_\mathbf{h}}{|\lambda_{\max}|} > 0. \tag{4.7b}$$

Thus, if $\varepsilon = 0$ (i.e., $\phi_h = 0$), then $V = 0$ for $\mathbf{x} = \mathbf{0}$ and $\mathrm{d}V/\mathrm{d}t < 0$ for $0 < |\mathbf{x}| < \phi_r$. Therefore, V is a local Lyapunov function for $\mathbf{x} = \mathbf{0}$. □

4.4 Stability condition under perturbation

For $\lambda_{\max} < 0$ and $\varepsilon < \lambda_{\max}^2/(4C_\mathbf{h}C_\mathbf{r})$, there exists a contour $V = V_0$ with $\phi_h^2 < V_0 < \phi_r^2$ on which $\mathrm{d}V/\mathrm{d}t < 0$ (Fig. 2). Hence, all solutions of Eq. (4.4) that start within this contour stay inside of it. As the initial state $\widehat{x}$ satisfies $\widehat{x} = \mathbf{P}(\widehat{m} - \widehat{m}) = \mathbf{0}$, we have

Lemma 6 *If*

$$\sqrt{\varepsilon} < \frac{-\lambda_{\max}}{2\sqrt{C_\mathbf{h}C_\mathbf{r}}}, \tag{4.8a}$$

then

$$|\mathbf{x}| < 2\varepsilon \frac{C_\mathbf{h}}{|\lambda_{\max}|} \tag{4.8b}$$

during the transient following mutant invasion.

Finally, by translating Eq. (4.8b) back to $\mathbf{m} - \widehat{m} = \mathbf{P}^{-1}\mathbf{x}$, and by substituting it into Eq. (3.4a), we have

Theorem 2 *For the population densities* $\mathbf{m} = (m_1, \ldots, m_M)^{\mathrm{T}}$ *of approximate phenotypes* $\mathbf{s}_{\mathrm{a}} = (s_1, \ldots, s_M)^{\mathrm{T}}$ *formed by clustering resident and mutant phenotypes* $\mathbf{s}' = (s_1, \ldots, s_{N+1})^{\mathrm{T}}$ *according to a threshold phenotypic distance* $\varepsilon = \rho_\mu \varepsilon_\mu$, *if* $\lambda_{\max}$ *defined by Eq. (4.5) satisfies the approximability condition, Eq. (4.8a),*

$$\sqrt{\varepsilon} < -\frac{\lambda_{\max}}{2\sqrt{C_{\mathrm{h}} C_{\mathrm{r}}}},$$

then

$$\left| \mathbf{m} - \widehat{m} \right| \leq C_{\mathbf{m}} \varepsilon \tag{4.9a}$$

during the transient following mutant invasion, where

$$C_{\mathbf{m}} = \frac{2 \left\| \mathbf{P}^{-1} \right\| C_{\mathbf{h}}}{|\lambda_{\max}|} \tag{4.9b}$$

with $\mathbf{P}$ *defined by Eq. (E.24) of Appendix E and with* $\| \cdot \|$ *denoting the induced norm for the matrix·, i.e., the maximum absolute value among its eigenvalues.*

In evolutionary dynamics determined by trait-substitution sequences, i.e., induced by repeated mutant invasions, when the fitness gradients for all residents are sufficiently strong, so that the coexistence of a mutant and its parental resident is impossible for any resident (as explained in Sect. 8), each of the resident phenotypes is not similar to any other. Then, ε may be chosen at $\varepsilon = \varepsilon_\mu$ (i.e., $\rho_\mu = 1$), so that only the mutant and its parental resident are clustered. This parental resident can be chosen as the approximate phenotype of that cluster, in which case all approximate phenotypes are identical to the resident phenotypes before the mutant invasion. Thus, as long as the initial equilibrium before the invasion is linearly stable, $\lambda_{\max}$ is negative, because of Eq. (4.1a) in conjunction with Eq. (2.3). Hence, for sufficiently small ε_μ, Eq. (4.8a) is always satisfied, in accordance with the proof by Dercole and Rinaldi (2008).

On the other hand, when the fitness gradients for some residents become small as a consequence of their directional coevolution toward higher fitnesses, effects of the higher-order properties of the fitness function may induce evolutionary branching. During the early stage of evolutionary branching, phenotypic distances among residents branched from the ancestral resident have magnitudes that are comparable with ε_μ. In this case, clustering only the mutant and its parental resident with $\varepsilon = \varepsilon_\mu$ may provide too small a value of $|\lambda_{\max}|$ to satisfy Eq. (4.8a), while including similar residents in the cluster for an appropriate ε larger than ε_μ may provide a sufficiently large $|\lambda_{\max}|$ to satisfy Eq. (4.8a).

In Dercole and Rinaldi (2008), only the mutant and its parental resident are clustered together, and the other residents are not clustered. Thus, when the phenotypic distance

among some residents is small, say, equal to $\varepsilon_{\text{resident}}$, the leading eigenvalue of the community's Jacobian matrix inevitably is close to zero as well. This problem is avoided in their proof by assuming sufficiently small mutational step sizes compared to $\varepsilon_{\text{resident}}$. However, when we consider a trait substitution sequence under a given magnitude of mutational step sizes, early stages of evolutionary branching inevitably lead to $\varepsilon_{\text{resident}}$ of the same order of magnitude as the mutational step sizes, no matter what is assumed for the latter. From this perspective, the proof by Dercole and Rinaldi (2008) requires all the residents to be dissimilar.

According to Eq. (4.3b), the leading eigenvalue also becomes close to zero when initial equilibrium densities of some residents are small. Analogously to the above case of similar residents, while this problem is seemingly avoided in Dercole and Rinaldi (2008) by assuming sufficiently small mutational step sizes, it inevitably occurs in trait-substitution sequences in which residents gradually go extinct. Hence, their proof fails to cover all the cases that one may wish to consider (and are actually considered in their book).

5 Approximability condition when the population densities of some approximate phenotypes are small

If an approximate phenotype s_1 has a small population density $\widehat{m}_1 = \text{O}(\varepsilon)$ at the initial equilibrium, then the corresponding eigenvalue of $\mathbf{J} = \text{diag}(\widehat{m})\mathbf{B}$ will be close to zero, making it difficult to satisfy the approximability condition in Eq. (4.8a). Even in this case, however, $\left| \mathbf{m} - \widehat{m} \right| = \text{O}(\varepsilon)$ may hold during the transient following mutant invasion. To cover this situation by developing a refined approximability condition, we examine in this section not only linear terms, but also quadratic terms, of the Taylor expansions investigated in the preceding section.

5.1 Transformation into perturbed community

First, we decompose the function $\mathbf{f}$ in Eq. (4.2),

$$\frac{d\mathbf{m}}{dt} = \text{diag}(\mathbf{m})\mathbf{f}(\mathbf{s}_a; \mathbf{s}'; \mathbf{n}'),$$

into the terms that are linear in $\mathbf{m}$, the terms that are of higher order in $\mathbf{m}$, and the perturbation terms, in a manner similar to Lemma 3 in the previous section. Specifically, we have

Lemma 7 *Equation* (4.2) *can be transformed into*

$$\frac{d\mathbf{m}}{dt} = \text{diag}(\mathbf{m})\left[\mathbf{B}(\mathbf{m} - \widehat{m}) + \varepsilon \mathbf{h}_f + \mathbf{r}_f \left| \mathbf{m} - \widehat{m} \right|^2 \right], \tag{5.1}$$

where $\mathbf{r}_f = (r_{f1}, \ldots, r_{fM})^T$ *is a function of* $\mathbf{m}$ *satisfying* $|\mathbf{r}_f| \leq C_{\mathbf{r}f}$ *and* $\mathbf{h}_f = (h_{f1}, \ldots, h_{fM})^T$ *is a function of* $\mathbf{m}'$ *and* ε *satisfying* $|\mathbf{h}_f| \leq C_{\mathbf{h}f}$.

See Appendix C.1 for the proof.

We now consider situations in which L population densities, i.e., $\widehat{m}_i$ for $i = 1, \ldots, L$, are large, while the remaining $K = M - L$ population densities, i.e., $\widehat{m}_i$ for $i = L + 1, \ldots, M$, are small, such that $\left|(\widehat{m}_{L+1}, \ldots, \widehat{m}_M)\right| = \rho_m \varepsilon$ for a positive constant ρ_m. To treat the small population densities differently from the larger ones, we decompose $\mathbf{m}$ into the larger population densities $\mathbf{m}_x = (m_1, \ldots, m_L)^T$ and the small population densities $\mathbf{m}_y = (m_{L+1}, \ldots, m_M)^T$, $\mathbf{m} = (\mathbf{m}_x, \mathbf{m}_y)^T$. Then, Eq. (5.1) is split into

$$\frac{d\mathbf{m}_x}{dt} = \mathrm{diag}(\mathbf{m}_x)\left[\mathbf{B}_{xx}(\mathbf{m}_x - \widehat{m}_x) + \mathbf{B}_{xy}(\mathbf{m}_y - \widehat{m}_y) + \varepsilon\mathbf{h}_{fx} + \mathbf{r}_{fx}\left|\mathbf{m} - \widehat{m}\right|^2\right] \quad (5.2a)$$

and

$$\frac{d\mathbf{m}_y}{dt} = \mathrm{diag}(\mathbf{m}_y)\left[\mathbf{B}_{yx}(\mathbf{m}_x - \widehat{m}_x) + \mathbf{B}_{yy}(\mathbf{m}_y - \widehat{m}_y) + \varepsilon\mathbf{h}_{fy} + \mathbf{r}_{fy}\left|\mathbf{m} - \widehat{m}\right|^2\right], \quad (5.2b)$$

with

$$\begin{pmatrix} \mathbf{B}_{xx} & \mathbf{B}_{xy} \\ \mathbf{B}_{yx} & \mathbf{B}_{yy} \end{pmatrix} = \mathbf{B}, \quad \begin{pmatrix} \mathbf{h}_{fx} \\ \mathbf{h}_{fy} \end{pmatrix} = \mathbf{h}_f, \quad \begin{pmatrix} \mathbf{r}_{fx} \\ \mathbf{r}_{fy} \end{pmatrix} = \mathbf{r}_f. \quad (5.2c)$$

Around the initial equilibrium $\widehat{m}$, the dynamics of $\mathbf{m}_y$ are much slower than those of $\mathbf{m}_x$. When ε goes to zero, the equilibrium population densities $\left|(\widehat{m}_{L+1}, \ldots, \widehat{m}_M)\right| \leq \rho_m \varepsilon$ do so as well, i.e.,

$$\widehat{m} \to (\widehat{m}_1, \ldots, \widehat{m}_L, 0, \ldots, 0)^T = \begin{pmatrix} \widehat{m}_x \\ \mathbf{0} \end{pmatrix} =: \widehat{\widetilde{m}}, \quad (5.3a)$$

and the whole dynamics get confined to a center manifold given by $d\mathbf{m}_x/dt = 0$. The slow dynamics close to the center manifold are governed by $d\mathbf{m}_y/dt$. An approximation for this center manifold can be derived by setting $d\mathbf{m}_x/dt = 0$ with $\varepsilon\mathbf{h}_{fx} + \mathbf{r}_{fx}\left|\mathbf{m} - \widehat{m}\right|^2 = 0$, yielding

$$\mathbf{m}_x = \widehat{m}_x - \mathbf{B}_{xx}^{-1}\mathbf{B}_{xy}\mathbf{m}_y =: \tilde{\mathbf{m}}_x(\mathbf{m}_y), \quad (5.3b)$$

which passes through the fixed point $\tilde{\mathbf{m}}_x(\mathbf{0}) =: \widehat{\widetilde{m}}$.

Although for $\varepsilon > 0$, $\widehat{m} = (\widehat{m}_x, \widehat{m}_y)^T$ will deviate from $\widehat{\widetilde{m}} = (\widehat{m}_x, \mathbf{0})^T$, it is expected that for small ε the dynamics can still be effectively characterized by their projection onto the center manifold $\mathbf{m}_x = \tilde{\mathbf{m}}_x(\mathbf{m}_y)$. Thus, we transform Eq. (5.2a) into

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}_x\mathbf{x} + \varepsilon\tilde{\mathbf{h}}_x + \tilde{\mathbf{r}}_x|\mathbf{w}|^2, \quad (5.4a)$$

$$\frac{d\mathbf{y}}{dt} = \mathrm{diag}(\mathbf{y})\left[\mathbf{J}_y\mathbf{y} + \mathbf{U}\mathbf{x} + \varepsilon\tilde{\mathbf{h}}_y + \tilde{\mathbf{r}}_y|\mathbf{w}|^2\right], \quad (5.4b)$$

with $\left|\tilde{\mathbf{h}}_{\mathbf{x}}\right| \leq \tilde{C}_{\mathbf{hx}}, \left|\tilde{\mathbf{h}}_{\mathbf{y}}\right| \leq \tilde{C}_{\mathbf{hy}}, |\tilde{\mathbf{r}}_{\mathbf{x}}| \leq \tilde{C}_{\mathbf{rx}}$, and $\left|\tilde{\mathbf{r}}_{\mathbf{y}}\right| \leq \tilde{C}_{\mathbf{ry}}$, where $\mathbf{x} := \mathbf{P}(\mathbf{m}_{\mathbf{x}} - \tilde{\mathbf{m}}_{\mathbf{x}}(\mathbf{m}_{\mathbf{y}}))$ describes the convergence to, or deviation from, the center manifold $\mathbf{m}_{\mathbf{x}} = \tilde{\mathbf{m}}_{\mathbf{x}}(\mathbf{m}_{\mathbf{y}})$, and $\mathbf{y} := \mathbf{m}_{\mathbf{y}}$ describes the slow dynamics along the manifold. The other variables and parameters newly introduced in Eqs. (5.4a) and (5.4b) are given by

$$\mathbf{w} := \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix},$$

$$\mathbf{A}_{\mathbf{x}} := \mathbf{P}\mathbf{J}_{\mathbf{x}}\mathbf{P}^{-1} = \mathbf{P}\mathrm{diag}(\widehat{m}_{\mathbf{x}})\mathbf{B}_{\mathbf{xx}}\mathbf{P}^{-1},$$

$$\mathbf{U} := \mathbf{B}_{\mathbf{yx}}\mathbf{P}^{-1},$$

$$\mathbf{J}_{\mathbf{y}} := \mathbf{B}_{\mathbf{yy}} - \mathbf{B}_{\mathbf{yx}}\mathbf{B}_{\mathbf{xx}}^{-1}\mathbf{B}_{\mathbf{xy}}, \tag{5.4c}$$

(Appendix F.1–F.4), where $\mathbf{P}$ is chosen so that Eq. (4.5a) is satisfied for $\mathbf{A} = \mathbf{A}_{\mathbf{x}}$ (Appendix E). Notice that $\mathbf{y}$ must be non-negative while $\mathbf{x}$ is indeterminate. Here, the effect of $\widehat{m} - \widetilde{m}$ is subsumed into the perturbation terms $\tilde{\mathbf{h}}_{\mathbf{x}}$ and $\tilde{\mathbf{h}}_{\mathbf{y}}$. Neglecting them gives a fixed point $\mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{0}$ that corresponds to $\widetilde{m} = (\widehat{m}_{\mathbf{x}}, \mathbf{0})^{\mathrm{T}}$, which is slightly different from the initial equilibrium $\widehat{m} = (\widehat{m}_{\mathbf{x}}, \widehat{m}_{\mathbf{y}})^{\mathrm{T}}$ when $\varepsilon > 0$. In the next subsections, we analyze the magnitude of $\left|\mathbf{m} - \widetilde{m}\right|$ in Eq. (5.4a) during the transient following mutant invasion, to obtain the magnitude of $\left|\mathbf{m} - \widehat{m}\right| = \left|\mathbf{m} - \widetilde{m}\right| + O(\varepsilon)$.

5.2 Local Lyapunov function

Following Mazenc (2001), we construct a local Lyapunov function to examine the magnitude of $|\mathbf{w}|$ during the transient following mutant invasion. We have

Lemma 8 *In Eq. (5.4a) with $\varepsilon = 0$, if the eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_M$ of the symmetric part of*

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_{\mathbf{x}} & 0 \\ d\mathbf{U} & d\mathbf{J}_{\mathbf{y}} \end{pmatrix}, \tag{5.5a}$$

with d being a positive constant, satisfy

$$\tilde{\lambda}_{\max} = \max\left\{\tilde{\lambda}_1, \ldots, \tilde{\lambda}_M\right\} < 0 \tag{5.5b}$$

and $|\mathbf{w}| < \phi$, with ϕ being a sufficiently small constant, then

$$V = \sum_{i=1}^{L} x_i^2 + 2d \sum_{i=L+1}^{M} y_i$$
$$= \mathbf{x}^{\mathrm{T}}\mathbf{x} + 2d\mathbf{c}^{\mathrm{T}}\mathbf{y} \tag{5.5c}$$

with $\mathbf{c} = (1, \ldots, 1)^{\mathrm{T}}$ *is a local Lyapunov function.*

Proof We assume that $\tilde{\lambda}_{\max} < 0$. The time derivative of V is

$$
\begin{aligned}
\frac{\mathrm{d}V}{\mathrm{d}t} &= 2\mathbf{x}^{\mathrm{T}}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} + 2d\mathbf{c}^{\mathrm{T}}\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} \\
&= 2\mathbf{x}^{\mathrm{T}}\Big[\mathbf{A}_x\mathbf{x} + \varepsilon\tilde{\mathbf{h}}_x + \tilde{\mathbf{r}}_x|\mathbf{w}|^2\Big] + 2d\mathbf{c}^{\mathrm{T}}\mathrm{diag}(\mathbf{y})\Big[\mathbf{J}_y\mathbf{y} + \mathbf{U}\mathbf{x} + \varepsilon\tilde{\mathbf{h}}_y + \tilde{\mathbf{r}}_y|\mathbf{w}|^2\Big] \\
&= 2\mathbf{x}^{\mathrm{T}}\Big[\mathbf{A}_x\mathbf{x} + \varepsilon\tilde{\mathbf{h}}_x + \tilde{\mathbf{r}}_x|\mathbf{w}|^2\Big] + 2d\mathbf{y}^{\mathrm{T}}\Big[\mathbf{J}_y\mathbf{y} + \mathbf{U}\mathbf{x} + \varepsilon\tilde{\mathbf{h}}_y + \tilde{\mathbf{r}}_y|\mathbf{w}|^2\Big] \\
&= 2\begin{pmatrix}\mathbf{x}\\\mathbf{y}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\mathbf{A}_x & 0\\ d\mathbf{U} & d\mathbf{J}_y\end{pmatrix}\begin{pmatrix}\mathbf{x}\\\mathbf{y}\end{pmatrix} + 2\begin{pmatrix}\mathbf{x}\\\mathbf{y}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\tilde{\mathbf{r}}_x\\ d\tilde{\mathbf{r}}_y\end{pmatrix}|\mathbf{w}|^2 + 2\varepsilon\begin{pmatrix}\mathbf{x}\\\mathbf{y}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\tilde{\mathbf{h}}_x\\ d\tilde{\mathbf{h}}_y\end{pmatrix} \\
&= 2\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{A}}\mathbf{w} + 2\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{r}}|\mathbf{w}|^2 + 2\varepsilon\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{h}}, \tag{5.6a}
\end{aligned}
$$

with

$$
\tilde{\mathbf{r}} = \begin{pmatrix}\mathbf{r}_x\\ d\mathbf{r}_y\end{pmatrix}, \quad \tilde{\mathbf{h}} = \begin{pmatrix}\mathbf{h}_x\\ d\mathbf{h}_y\end{pmatrix}. \tag{5.6b}
$$

Notice that the last line of Eq. (5.6a) has a form identical to the second line of Eq. (4.7a). Although in this case the initial state $\widehat{w} := (0, \ldots, 0, \widehat{m}_{L+1}, \ldots, \widehat{m}_M)^{\mathrm{T}}$ is not zero, $\big|\widehat{w}\big| = \big|(0, \ldots, 0, \widehat{m}_{L+1}, \ldots, \widehat{m}_M)\big| = \varepsilon\rho_{\mathrm{m}}$ holds. Therefore, Eq. (5.6a) can analogously be transformed further,

$$
\begin{aligned}
\frac{\mathrm{d}V}{\mathrm{d}t} &\leq 2\tilde{\lambda}_{\max}|\mathbf{w}|^2 + 2\tilde{C}_{\mathbf{r}}|\mathbf{w}|^3 + 2\varepsilon\tilde{C}_{\mathbf{h}}|\mathbf{w}| \\
&= 2\tilde{C}_{\mathbf{r}}|\mathbf{w}|^2\left(|\mathbf{w}| - \frac{\big|\tilde{\lambda}_{\max}\big|}{2\tilde{C}_{\mathbf{r}}}\right) + \tilde{\lambda}_{\max}|\mathbf{w}|\left(|\mathbf{w}| - \frac{2\varepsilon\tilde{C}_{\mathbf{h}}}{\big|\tilde{\lambda}_{\max}\big|}\right) \\
&\leq 2\tilde{C}_{\mathbf{r}}|\mathbf{w}|^2\left(|\mathbf{w}| - \frac{\big|\tilde{\lambda}_{\max}\big|}{2\tilde{C}_{\mathbf{r}}}\right) + \tilde{\lambda}_{\max}|\mathbf{w}|\left(|\mathbf{w}| - \max\left\{\frac{2\varepsilon\tilde{C}_{\mathbf{h}}}{\big|\tilde{\lambda}_{\max}\big|}, \varepsilon\rho_{\mathrm{m}}\right\}\right) \\
&= 2\tilde{C}_{\mathbf{r}}|\mathbf{w}|^2\Big(|\mathbf{w}| - \tilde{\phi}_{\mathrm{r}}\Big) + \tilde{\lambda}_{\max}|\mathbf{w}|\Big(|\mathbf{w}| - \tilde{\phi}_{\mathrm{h}}\Big), \tag{5.6c}
\end{aligned}
$$

where $\tilde{C}_{\mathbf{h}} \geq \big|\tilde{\mathbf{h}}\big|$, $\tilde{C}_{\mathbf{r}} \geq |\tilde{\mathbf{r}}|$ (Appendix F.5) and

$$
\tilde{\phi}_{\mathrm{h}} = \max\left\{\frac{2\varepsilon\tilde{C}_{\mathbf{h}}}{\big|\tilde{\lambda}_{\max}\big|}, \varepsilon\rho_{\mathrm{m}}\right\},
$$

$$
\tilde{\phi}_{\mathrm{r}} = \frac{\big|\tilde{\lambda}_{\max}\big|}{2\tilde{C}_{\mathbf{r}}}. \tag{5.6d}
$$

In Eq. (5.6c), the transformation from the second to the third row allows the initial state $\widehat{w}$ to satisfy $\left|\widehat{w}\right| = \varepsilon\rho_m \leq \tilde{\phi}_h$, which is used for the case of $\varepsilon > 0$ in Lemma 10. Therefore, for $\varepsilon = 0$ (i.e., $\tilde{\phi}_h = 0$), V satisfies $V = 0$ for $\mathbf{w} = \mathbf{0}$ and $dV/dt < 0$ for $|\mathbf{w}| < \tilde{\phi}_r$.   □

In addition, the following lemma is proved in Appendix G.

Lemma 9 *Equation* (5.5b), *i.e.,*

$$\tilde{\lambda}_{\max} < 0,$$

holds, if the real parts of the eigenvalues of $\mathbf{J}_x = \mathrm{diag}(\widehat{m}_x)\mathbf{B}_{xx}$ *and the real eigenvalues of* $\frac{1}{2}\left(\mathbf{J}_y + \mathbf{J}_y^{\mathrm{T}}\right)$ *with* $\mathbf{J}_y = \mathbf{B}_{yy} - \mathbf{B}_{yx}\mathbf{B}_{xx}^{-1}\mathbf{B}_{xy}$ *are all negative, and if d is sufficiently small.*

5.3 Stability under perturbation

Next, we take the perturbation into account, i.e., we consider $\varepsilon > 0$. In the previous section, the contour curves of the local Lyapunov function have the same shapes as the boundaries of the region ensuring that $dV/dt < 0$, i.e., as the two circles $|\mathbf{w}| = \phi_h$ and $|\mathbf{w}| = \phi_r$. In this section, although the contours have shapes different from the circles $|\mathbf{w}| = \tilde{\phi}_h$ and $|\mathbf{w}| = \tilde{\phi}_r$, the manner of analysis is the same. First, as the initial state $\widehat{w}$ satisfies $\left|\widehat{w}\right| = \varepsilon\rho_m \leq \tilde{\phi}_h$ according to Eq. (5.6d), we trivially have

Lemma 10 *We assume that ε is sufficiently small so that* $\tilde{\phi}_h < \tilde{\phi}_r$. *For a region* $D = \left\{\mathbf{w}|\ \tilde{\phi}_h < |\mathbf{w}| < \tilde{\phi}_r\right\}$, *within which* $dV/dt < 0$ *with V defined in Eq.* (5.5c), *consider a contour curve* $V = V_0$ *such that its inscribed circle is given by* $|\mathbf{w}| = \tilde{\phi}_h$ *and its circumscribed circle is given by* $|\mathbf{w}| = \alpha\tilde{\phi}_h$ *with* $\alpha > 1$ *(Fig. 3). If*

$$\alpha\tilde{\phi}_h < \tilde{\phi}_r, \tag{5.7a}$$

then there exists a set $E = \{\mathbf{w}|\ V_0 < V < V_0 + \delta\}$ *with a sufficiently small δ where* $dV/dt < 0$ *and which therefore ensures that*

$$|\mathbf{w}| \leq \alpha\tilde{\phi}_h \tag{5.7b}$$

holds during the transient following mutant invasion.

This lemma is an extension of Lemma 5 relaxing the requirement that the contours be circular: Lemma 10 with $\alpha = 1$ corresponds to Lemma 5.

Then, by substituting Eqs. (5.6d) into (5.7a), we have

Lemma 11 *For some d, if*

$$\sqrt{\varepsilon} < \frac{-\tilde{\lambda}_{\max}}{2\sqrt{\alpha\tilde{C}_r\max\left\{\tilde{C}_h, \frac{1}{2}\left|\tilde{\lambda}_{\max}\right|\rho_m\right\}}}, \tag{5.8a}$$
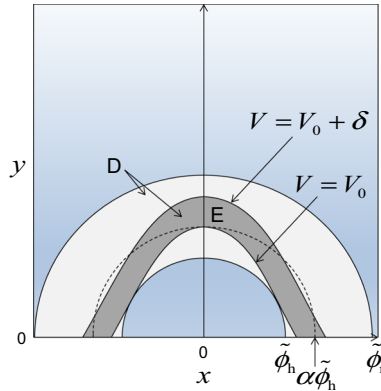
Fig. 3 Stability against perturbation when the initial equilibrium population densities of some approximate phenotypes are small. As a simple example, a community composed of two approximate phenotypes $\mathbf{s_a} = (s_1, s_2)^T$ is considered. Their population densities $\mathbf{m} = (m_1, m_2)^T$ are transformed into $\mathbf{w} = (x, y)^T$ so that $\mathbf{w} = \mathbf{0}$ corresponds to the initial equilibrium before mutant invasion, where y corresponds to the population density of the phenotype with small population density. A local Lyapunov function $V = x^2 + dy$ (with $d > 0$) of the community dynamics monotonically decreases with time within the light-gray and dark-gray regions marked by D. The dark-gray region marked by E is associated with a repeller that prevents the community dynamics from passing a boundary $V = V_0$ from its inner side $V < V_0$, thus keeping $|\mathbf{w}| \le \alpha\tilde{\phi}_h$

then

$$|\mathbf{w}| \le -\frac{2\varepsilon\alpha}{\left|\tilde{\lambda}_{\max}\right|}\max\left\{\tilde{C}_\mathbf{h}, \ \tfrac{1}{2}\left|\tilde{\lambda}_{\max}\right|\rho_\mathrm{m}\right\} \tag{5.8b}$$

holds during the transient following mutant invasion, where

$$\alpha = \begin{cases} 1 & \text{for } K = 0 \\[2mm] \max\left\{\dfrac{\sqrt{\tilde{\phi}_h^2 + Kd^2}}{\tilde{\phi}_h}, \ \dfrac{\tilde{\phi}_h^2 + Kd^2}{2d\tilde{\phi}_h}\right\} & \text{for } 0 < K \le \dfrac{\tilde{\phi}_h^2}{d^2} \\[3mm] \max\left\{\dfrac{\sqrt{2dK^{1/2}\tilde{\phi}_h}}{\tilde{\phi}_h}, \ \sqrt{K}\right\} & \text{for } K > \dfrac{\tilde{\phi}_h^2}{d^2} \end{cases}, \tag{5.8c}$$

and $K = M - L$ is the number of approximate phenotypes with small equilibrium population densities.

See Appendix H for the derivation of the expressions for α. When $K = 0$, this lemma is independent of d and becomes identical to Lemma 6 in the previous section, i.e., $\tilde{\lambda}_{\max} = \lambda_{\max}$, $\tilde{C}_\mathbf{r} = C_\mathbf{r}$, $\tilde{C}_\mathbf{h} = C_\mathbf{h}$, and $\alpha = 1$. Thus, this lemma includes Lemma 6 as a special case. When $K > 0$, on the other hand, $\tilde{\lambda}_{\max}$, α, $\tilde{C}_\mathbf{h}$, and $\tilde{C}_\mathbf{r}$ in Eq. (5.8a) all depend on d. A choice for d may be one that maximizes the right-hand side of Eq. (5.8a).

By translating Lemma 11 into a corresponding statement for $\mathbf{m}$, we get

Theorem 3 *For the population densities* $\mathbf{m} = (m_1, \ldots, m_M)^T$ *of approximate phenotypes* $\mathbf{s}_a = (s_1, \ldots, s_M)^T$ *formed by clustering resident phenotypes* $\mathbf{s}' = (s_1, \ldots, s_{N+1})^T$ *according to a threshold phenotypic distance* $\varepsilon = \rho_\mu \varepsilon_\mu$, *if the approximability condition, Eq.* (5.8a),

$$\sqrt{\varepsilon} < \frac{-\tilde{\lambda}_{max}}{2\sqrt{\alpha \tilde{C}_r \max\left\{\tilde{C}_\mathbf{h}, \ \frac{1}{2}\left|\tilde{\lambda}_{max}\right|\rho_m\right\}}}$$

is satisfied for some d, then

$$\left|\mathbf{m} - \widehat{m}\right| = \left|\mathbf{Q}^{-1}\mathbf{w} + \widehat{m} - \widehat{\tilde{m}}\right| \le C_\mathbf{m}\varepsilon,$$

$$C_\mathbf{m} = \frac{2\alpha\left\|\mathbf{Q}^{-1}\right\|\max\left\{\tilde{C}_\mathbf{h}, \ -\frac{1}{2}\tilde{\lambda}_{max}\rho_m\right\}}{\left|\tilde{\lambda}_{max}\right|} + \rho_m, \tag{5.9a}$$

holds during the transient following mutant invasion, where $\mathbf{m}$ *is split into an L-dimensional vector* $\mathbf{m}_x = (m_1, \ldots, m_L)^T$ *of not-small initial population densities, and a* $K(= M - L)$-*dimensional vector* $\mathbf{m}_y = (m_{L+1}, \ldots, m_M)^T$ *of small initial population densities* $\widehat{m}_{L+1}, \ldots, \widehat{m}_M$ *such that* $\left|\left(\widehat{m}_{L+1}, \ldots, \widehat{m}_M\right)\right| = \varepsilon\rho_m$ ($\rho_m = 0$ *when all initial population densities are not small, i.e.,* $K = 0$), *and* $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_M$ *are the eigenvalues of* $\frac{1}{2}\left(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^T\right)$ *with*

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_x & 0 \\ d\mathbf{B}_{yx}\mathbf{P}^{-1} & d\mathbf{J}_y \end{pmatrix},$$

$$\mathbf{A}_x = \mathbf{P}^{-1}\mathrm{diag}(\widehat{m}_x)\mathbf{B}_{xx}\mathbf{P},$$

$$\mathbf{J}_y = \mathbf{B}_{yy} - \mathbf{B}_{yx}\mathbf{B}_{xx}^{-1}\mathbf{B}_{xy},$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{P} & \mathbf{P}\mathbf{B}_{xx}^{-1}\mathbf{B}_{xy} \\ 0 & \mathbf{I}_y \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{B}_{xx} & \mathbf{B}_{xy} \\ \mathbf{B}_{yx} & \mathbf{B}_{yy} \end{pmatrix} = \mathbf{B}, \tag{5.9b}$$

where $\mathbf{I}_y$ *denotes the* $K \times K$ *identity matrix,* $\mathbf{B}$ *is defined in Eq.* (4.3b), *and* $\mathbf{P}$ *is defined in Eq.* (E.24) *in Appendix* E.

Note that it is arbitrary which phenotypes we choose as having small initial population densities. Thus, whether each approximate phenotype's initial population density is small or not can be decided in such a manner that satisfying the approximability condition becomes easiest. If none of the approximate population densities is treated as small, i.e., $K = 0$, Theorem 3 becomes identical to Theorem 2. Thus, Theorem 3 includes Theorem 2 as a special case. The threshold phenotypic distance ε and the way of clustering can also be chosen arbitrarily, so that satisfying the approximability

condition becomes easiest, as long as $\varepsilon > \varepsilon_\mu$. Notice that $\tilde{C}_{\mathbf{h}}$ depends on ε, although it is bounded when $\varepsilon \to 0$. In addition, all of the other mathematical objects in Eq. (5.8a) depend indirectly on ε, because ε affects how to cluster the existing phenotypes. A procedure for the evaluation of Eq. (5.8a) would be as follows: first choose ε, and then choose approximate phenotypes, choose K, choose d when $K > 0$ (so that the right-hand side of Eq. (5.8a) is maximized), and examine whether the inequality holds good.

Moreover, the initial state $\widehat{n'}$ need not be exactly at an equilibrium of the resident phenotypes $s_1, \ldots, s_N$ if the value of $\widehat{C'_{\text{Fz}}}$ is adjusted such that $F(s_i; \mathbf{s}'; \widehat{n'}) \le \varepsilon \widehat{C'_{\text{Fz}}}$ is satisfied for all $i = 1, \ldots, N+1$. Therefore, this theorem can be applied also to the case of higher mutation rates, in which frequent mutant invasions prevent the community from reaching the next population-dynamical equilibrium.

Note also that the smallness of changes of the population densities of the approximate phenotypes ensures not only the smallness of fitness changes of existing phenotypes $F(s_i; \mathbf{s}'; \mathbf{n}')$ for $i = 1, \ldots, N+1$, i.e., LV-approximability, but also the smallness of fitness changes of any non-existing phenotype z, i.e., of the fitness landscape $F(z; \mathbf{s}'; \mathbf{n}')$. Specifically, from Theorem 3 we immediately see

Corollary 1 *If the approximability condition, Eq. (5.8a), in Theorem 3 is satisfied, then the change of the fitness landscape $F(z; \mathbf{s}'; \mathbf{n}')$ is slight during the transient following mutant invasion, because by using Taylor's theorem we see that*

$$
\left| F(z; \mathbf{s}'; \mathbf{n}') - F(z; \mathbf{s}'; \widehat{n'}) \right|
$$

$$
= \left| \breve{F}(z; \mathbf{s}'; \mathbf{m}') - \breve{F}(z; \mathbf{s}'; \widehat{m'}) \right|
$$

$$
= \left| \left[\breve{F}(z; \mathbf{s}'; \widehat{m'}) + \left. \frac{\partial \breve{F}(z; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}'} \right|_{\mathbf{m}'=\mathbf{m}'_{\mathrm{T}}} (\mathbf{m}' - \widehat{m'}) \right] - \breve{F}(z; \mathbf{s}'; \widehat{m'}) \right|
$$

$$
= \left| \left. \frac{\partial \breve{F}(z; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}'} \right|_{\mathbf{m}'=\mathbf{m}'_{\mathrm{T}}} (\mathbf{m}' - \widehat{m'}) \right|
$$

$$
\le \max_{\left| \mathbf{m}'-\widehat{m'} \right| < C'_{\mathbf{m}}\varepsilon} \left\| \frac{\partial \breve{F}(z; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}'} \right\| C'_{\mathbf{m}}\varepsilon, \tag{5.10a}
$$

where $\mathbf{m}_{\mathrm{T}} := \theta_{\mathrm{T}}(\mathbf{m} - \widehat{m}) + \widehat{m}$ with an appropriately chosen $\theta_{\mathrm{T}} \in [0, 1]$, and by using $C_{\mathbf{m}}$ from Eq. (5.9a) in $C'_{\mathbf{m}} = \sqrt{C_{\mathbf{m}}^2 + (N + 1 - M)\eta^2}$ from Lemma 2, we see that

$$
C'_{\mathbf{m}} = \sqrt{ \left(\frac{2\alpha \|\mathbf{Q}^{-1}\| \max\left\{ \tilde{C}_{\mathbf{h}}, -\frac{1}{2}\tilde{\lambda}_{\max}\rho_{\mathrm{m}} \right\}}{\left| \tilde{\lambda}_{\max} \right|} + \rho_{\mathrm{m}} \right)^2 + (N + 1 - M)\eta^2}. \tag{5.10b}
$$

5.4 Generalization to higher-dimensional trait spaces

Theorems 1–3 and Corollary 1 apply to one-dimensional trait spaces. These results readily generalized to trait spaces of arbitrary dimensions by a slight modification of the analyses so that the derivative of the fitness function with respect to a phenotype in the one-dimensional trait space is replaced with the corresponding directional derivative in the higher-dimensional trait space, as shown in Appendix I.

6 Tighter estimates

For deriving the approximability conditions in Eqs. (4.7a) and (5.8a), we have used the maximum possible values of the perturbation terms ($\mathbf{h}$ and $\tilde{\mathbf{h}}$) and nonlinear terms ($\mathbf{r}$ and $\tilde{\mathbf{r}}$) attainable for $\mathbf{n}' \in [0, \eta]^{N+1}$. These provide the simplest, but rather conservative, approximability conditions. By approximating those terms as linear or higher-order functions of population densities (corresponding to $\mathbf{x}$ and $\mathbf{w}$), we may improve the estimates underlying the approximability conditions. In Theorem 2, for example, the perturbation term $\mathbf{h}$ can be expanded in $\mathbf{w}$ around $\mathbf{w} = \mathbf{0}$ (i.e., $\mathbf{m} = \widehat{m}$) up to the first-order remainder terms,

$$
\begin{aligned}
\mathbf{h} &= \widehat{h} + \mathbf{Hx}, \\
\widehat{h} &= \mathbf{h}|_{\mathbf{m}=\widehat{m}'}, \\
\mathbf{H} &= \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{m}=\mathbf{m}'_\mathrm{T}}
\end{aligned}
\tag{6.1}
$$

with some appropriately chosen $\mathbf{m}_\mathrm{T} \in [0, \eta]^M$. Then, applying Eqs. (6.1) to Lemma 6 gives the condition

$$
\sqrt{\varepsilon} < \frac{-\lambda_{\max}}{\sqrt{4\widehat{C}_\mathbf{h} C_\mathbf{r} + 2C_\mathbf{H}|\lambda_{\max}|}},
\tag{6.2}
$$

where $\widehat{C}_\mathbf{h} \geq \left|\widehat{h}\right|$ and $C_\mathbf{H} \geq \|\mathbf{H}\|$; see Appendix J for the derivation. As the magnitude of the zeroth-order term for $\mathbf{h}$ is estimated more tightly by $\widehat{C}_\mathbf{h}$ at $\mathbf{m} = \widehat{m}$, compared to $C_\mathbf{h}$ used in the original approximability condition in Eq. (4.8a), this condition can work better than the original approximability condition, but is less simple.

7 Example: Approximability condition for a resource-competition model

In this section, we give a simple example of how to examine the approximability condition in a specific ecological model.

7.1 Model description

We consider a resource-competition model based on the Beddington–DeAngelis-type functional response (Beddington 1975; DeAngelis et al. 1975), known to describe both saturation of consumption and interference competition among consumers. Under N coexisting consumer phenotypes $\mathbf{s} = (s_1, \ldots, s_N)^{\mathrm{T}}$ with their densities $\mathbf{n} = (n_1, \ldots, n_N)^{\mathrm{T}}$, we describe the ith phenotype's per-capita growth rate as

$$\frac{1}{n_i} \frac{\mathrm{d}n_i}{\mathrm{d}t} = \beta g(s_i; \mathbf{s}; \mathbf{n}) - \psi, \tag{7.1}$$

$$g(s_i; \mathbf{s}; \mathbf{n}) = \frac{\theta(s_i)}{\zeta_1 + \zeta_2 \theta(s_i) + \zeta_3 \sum_{j=1}^{N} n_j \alpha(s_j, s_i)}, \tag{7.2}$$

In Eq. (7.1), $g(s_i; \mathbf{s}; \mathbf{n})$ is the resource gain of phenotype s_i, β is a constant assimilation efficiency, and ψ is a constant natural death rate. In Eq. (7.2), $\theta(s_i)$ is the density of potential resources for s_i, and $\alpha(s_j, s_i)$ describes the niche overlap between phenotypes s_i and s_j. ζ_1, ζ_2, and ζ_3 are constant parameters related to the encounter rate of resources, handling time of resources, and intensity of interference competition, respectively. Notice that $\zeta_3 = 0$ gives the Holling type-II functional response (Holling 1959). Equation (7.2) can be derived from a generalized Beddington–deAngelis functional response with explicit description of a resource distribution and phenotypes' niches expressed along a resource-quality axis (Appendix L.1).

We assume $\psi = 1$ and $\zeta_3 = 1$ without loss of generality. For simplicity, we assume that $\theta(s_i)$ depends only on s_i (i.e., constant inflows of resources into the system) and that

$$\alpha(s_j, s_i) = \exp\left(-\tfrac{1}{2}(s_j - s_i)^2\right). \tag{7.3}$$

7.2 Approximability condition

As a simplest example for the approximability condition in this model, we consider invasion by a mutant phenotype s_2 into a single resident phenotype s_1 at its population-dynamical equilibrium $\widehat{n}_1 = [\beta - \zeta_2]\theta(s_1) - \zeta_1$. From Eqs. (7.1) and (7.2), their population dynamics are given by

$$\frac{1}{n_1} \frac{\mathrm{d}n_1}{\mathrm{d}t} = F(s_1; \mathbf{s}'; \mathbf{n}') = \frac{\beta \theta(s_1)}{\zeta_1 + \zeta_2 \theta(s_1) + n_1 + n_2 \alpha(s_2, s_1)} - 1, \tag{7.4a}$$

$$\frac{1}{n_2} \frac{\mathrm{d}n_2}{\mathrm{d}t} = F(s_2; \mathbf{s}'; \mathbf{n}') = \frac{\beta \theta(s_2)}{\zeta_1 + \zeta_2 \theta(s_2) + n_1 \alpha(s_2, s_1) + n_2} - 1, \tag{7.4b}$$

with $\mathbf{s}' = (s_1, s_2)^{\mathrm{T}}$ and $\mathbf{n}' = (n_1, n_2)^{\mathrm{T}}$.

We choose s_1 as the approximate phenotype, i.e., $s_a = s_1$ and $\mathbf{m}' = (m_1, m_2)^T = (n_1 + n_2, \varepsilon n_2)^T$ with $\varepsilon = s_2 - s_1$, $0 < \varepsilon \ll 1$. Then, from Eqs. (4.3b) and (4.5b), we see

$$\lambda_{\max} = \widehat{m}_1 b_{11} = \widehat{n}_1 \left[\frac{dF(s_2; s_1; m_1)}{dm_1} \right]_{m_1 = \widehat{n}_1}$$

$$= -\frac{\widehat{n}_1}{\beta\theta(s_1)}$$

$$= -\frac{[\beta - \zeta_2]\theta(s_1) - \zeta_1}{\beta\theta(s_1)}. \tag{7.5}$$

Notice that $\lambda_{\max}$ is always negative because $\widehat{n}_1$ must be positive. As for $C_\mathbf{h}$ and $C_\mathbf{r}$ in the approximability condition $\sqrt{\varepsilon} < -\lambda_{\max}/[2\sqrt{C_\mathbf{h} C_\mathbf{r}}]$ in Theorem 2, we find (as derived in Appendix L.2) that

$$C_\mathbf{h} \leq \breve{C}_\mathbf{h},$$
$$\breve{C}_\mathbf{h} = \frac{\eta\beta[5C_{\partial\theta} + 3\varepsilon C_\theta]}{\zeta_1 + \zeta_2 C_{\theta\min}},$$
$$C_\mathbf{r} = \frac{1}{\beta\theta(s_1)} + \eta\frac{2\beta\theta(s_1)}{[\zeta_1 + \zeta_2\theta(s_1)]^3}, \tag{7.6}$$

with

$$\eta = [\beta - \zeta_2]C_\theta - \zeta_1,$$
$$C_\theta = \max\{\theta(s)|s \in [s_1, s_2]\},$$
$$C_{\theta\min} = \min\{\theta(s)|s \in [s_1, s_2]\},$$
$$C_{\partial\theta} = \max\left\{\frac{d\theta(s)}{ds}|s \in [s_1, s_2]\right\}. \tag{7.7}$$

Therefore, a sufficient condition for the approximability condition is given by

$$\sqrt{\varepsilon} < -\frac{\lambda_{\max}}{2\sqrt{\breve{C}_\mathbf{h} C_\mathbf{r}}}$$

$$= \frac{\widehat{n}_1}{2\sqrt{\frac{\eta\beta[5C_{\partial\theta} + 3\varepsilon C_\theta]}{\zeta_1 + \zeta_2 C_{\theta\min}}\left(\beta\theta(s_1) + \eta\frac{2\beta^3\theta(s_1)^3}{[\zeta_1 + \zeta_2\theta(s_1)]^3}\right)}} \tag{7.8}$$

with $\widehat{n}_1 = [\beta - \zeta_2]\theta(s_1) - \zeta_1$ and Eqs. (7.7). Notice that the right-hand side of Eq. (7.8) includes εC_θ, which is negligible when $\varepsilon C_\theta \ll C_{\partial\theta}$.

8 Application: Extending the invasion–implies–substitution theorem

The derived stability conditions and resultant Lotka–Volterra approximation can be used to analyze the community dynamics triggered by a mutant invasion. In this section, we apply them to extend the invasion–implies–substitution theorem (Dercole and Rinaldi 2008, Appendix B) to an arbitrary set of resident phenotypes that form well-recognizable and -separated clusters in a one-dimensional trait space; see Appendix K for details.

We assume an arbitrary set of resident phenotypes $s_1, \ldots, s_N$ together with a mutant $s' = s_{N+1}$, with the resident and mutant phenotypes clustered into approximate phenotypes $\mathbf{s}_a = (s_1, \ldots, s_M)^\mathrm{T}$ that satisfy the approximability condition of Theorem 3. Then, from Lemma 2, $\left|\Delta \mathbf{m}'\right| \leq \varepsilon C'_\mathrm{m} = \varepsilon \sqrt{C^2_\mathrm{m} + (N + 1 - M)\eta^2}$ is conserved during the transient following mutant invasion. We denote the identity of the cluster containing the mutant by i, i.e., $\mathrm{cid}(N + 1) = i$. Using Eq. (4.1a), the dynamics of the mutant fraction $p_{N+1} = n_{N+1}/m_i$ within this cluster can be expressed as

$$
\begin{aligned}
\frac{\mathrm{d}p_{N+1}}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t} \frac{n_{N+1}}{m_i} \\
&= \frac{n_{N+1}}{m_i} F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - \frac{n_{N+1}}{m_i} f(s_i; \mathbf{s}'; \mathbf{n}') \\
&= p_{N+1} F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - p_{N+1} \sum_{j \in \mathrm{com}(i)} p_j F(s_j; \mathbf{s}'; \mathbf{n}').
\end{aligned}
\tag{8.1}
$$

For convenience, we assume that the representative phenotype s_i of this cluster is chosen as the phenotype most similar to the mutant, i.e., $|s_{N+1} - s_i| = \min_{j \in \mathrm{com}(i)} |s_{N+1} - s_j|$. Then, by Taylor's theorem, we transform Eq. (8.1) into

$$
\frac{\mathrm{d}p_{N+1}}{\mathrm{d}t} \geq p_{N+1}(1 - p_{N+1}) \left[\widehat{F}_z(s_i)(s_{N+1} - s_i) - \varepsilon C'_{\mathrm{Fzm}} \left|\Delta \mathbf{m}'\right| - \varepsilon^2 C_{\mathrm{Fzz}} \right]. \tag{8.2a}
$$

Here, $\widehat{F}_z(s_i) = \partial F(z; \mathbf{s}'; \widehat{n}')/\partial z \big|_{z=s_i}$ is the fitness gradient at s_i, and the constants C'_{Fzm} and C'_{Fzz} bound the remainder terms through

$$
\left| \frac{\partial \breve{F}(z; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial z} \right|_{z=z_{j\mathrm{T}}} \leq C'_{\mathrm{Fzm}}, \quad \left| \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z^2} \right|_{z=z_{j\mathrm{T}}} \leq C'_{\mathrm{Fzz}}, \tag{8.2b}
$$

for $z_{j\mathrm{T}} \in [s_j, s_{\mathrm{cid}(j)}]$ for all $j = 1, \ldots, N + 1$ during the transient following mutant invasion. Then, by substituting our results $\left|\Delta \mathbf{m}'\right| \leq \varepsilon C'_\mathbf{m} = \varepsilon \sqrt{C^2_\mathrm{m} + (N + 1 - M)\eta^2}$ and Eq. (5.9a) into Eq. (8.2a), a sufficient condition for $\mathrm{d}p_{N+1}/\mathrm{d}t$ to be always positive is given by

$$
\widehat{F}_z(s_i)(s_{N+1} - s_i) > \varepsilon^2 [C_{\mathrm{Fzz}} + C'_{\mathrm{Fzm}} C'_\mathbf{m}] \tag{8.3}
$$

with $C'_\mathbf{m}$ in Eq. (5.10b).

If the fitness gradient $\widehat{F}_z(s_i)$ is sufficiently strong, so that it satisfies Eq. (8.3), then p_{N+1} monotonically increases until it reaches 1, i.e., until all other phenotypes within the cluster containing the mutant are excluded. Equation (8.3) means that the fitness advantage of s_{N+1} against s_i due to the fitness gradient must exceed the effects of the curvature of the fitness landscape ($\varepsilon^2 C_{Fzz}$) and of the perturbation due to the population dynamics ($\varepsilon^2 C'_{Fzm} C'_m$). As long as Eq. (8.3) holds for any resident phenotype and its mutants, repeated mutant invasions always result in monomorphic phenotype clusters, i.e., resident phenotypes are kept dissimilar, corresponding to the situation considered by Dercole and Rinaldi (2008). Notice that when $\left|\tilde{\lambda}_{\max}\right|$ becomes close to zero, e.g., when the community is close to a bifurcation point of its population dynamics, C'_m becomes large and Eq. (8.3) thus becomes difficult to satisfy.

9 Discussion

As explained in the beginning of this paper, ecological interactions engender various evolutionary dynamics, including cyclic coevolution, adaptive radiation, adaptive speciation, taxon cycles, and community formation. To analyze how ecological interactions induce selection pressures that drive such dynamics, the following two assumptions are often made (Metz et al. 1992, 1996; Dieckmann and Law 1996). First, mutation rates are sufficiently small relative to the timescale of the population dynamics, so that the evolutionary dynamics are reduced to trait-substitution sequences resulting from repeated mutant invasions. Second, mutational step sizes are sufficiently small, so that a mutant invasion typically results in an equilibrium phenotype distribution similar to that before the invasion. The latter is called attractor inheritance (Geritz et al. 2002). In such cases, each mutant invasion modifies the fitness landscape only slightly. The fitness landscape can then be treated as a smooth function of resident phenotypes at equilibrium population densities, enabling effective analyses of directional coevolution (Dieckmann and Law 1996) and diversification through evolutionary branching (Metz et al. 1992, 1996; Geritz et al. 1997, 1998). Using the concept of approximate phenotypes introduced in the present paper, attractor inheritance can be translated into the smallness of changes of the population densities of approximate phenotypes during the transient population dynamics following mutant invasion, toward the next population-dynamical equilibrium.

9.1 Conditions for attractor inheritance

Prior to our analyses in the present paper, qualitative conditions for attractor inheritance have been proved for sufficiently small mutational step sizes in the following two cases: (1) all residents and the mutant are similar to each other (Geritz et al. 2002; Meszéna et al. 2005; Durinx et al. 2008), or (2) no two residents are similar to each other and their initial equilibrium population densities are not small (Dercole and Rinaldi 2008, Appendix B). In this paper, we have derived quantitative conditions for attractor inheritance for a set of residents and a mutant, by clustering them according to a threshold phenotypic distance into approximate phenotypes. The condi-

tions ensuring attractor inheritance, i.e., the approximability conditions in Theorems 2 and 3, establish relationships among the magnitudes of the mutational step size, the return rate to an equilibrium of the population dynamics of approximate phenotypes, the nonlinearity of the population dynamics, and the perturbation due to within-cluster population dynamics. These conditions are especially important when finite, rather than infinitesimally small, mutational step sizes are required for analyzing the considered evolutionary dynamics, such as when investigating evolutionary suicide (Gyllenberg and Parvinen 2001) and evolutionary branching of directionally evolving populations (Ito and Dieckmann 2012, 2014). A next step would be to analyze whether it is really possible to satisfy the approximability condition, or rather, whether the condition can be satisfied with not too large error bounds in all but a set of theoretically possible but practically irrelevant cases. Although we here have considered only deterministic population dynamics, the impact of demographic stochasticity on trait-substitution sequences (Geritz et al. 2002) can be considered using the same framework we have introduced here, by subsuming its effect in the perturbation terms.

9.2 Assumption of well-recognizable and -separated phenotypic clusters

Our analysis assumes that the number N of existing phenotypes is finite, and that phenotypic clusters are well-recognizable and well-separated from each other so that the largest of within-cluster distances, ε, is much smaller than the smallest of between-cluster distances. We discuss the validity of our two assumptions below.

In principle, ODE population models should be seen as large-system-size limits of stochastic individual-based models. Generally, the larger the number of coexisting phenotypes, the slower is the convergence to the ODE limit. Thus, for all practical purposes, ODE models with very large numbers of phenotypes can be left out of the picture. If we do so, the finiteness of the number of existing phenotypes, N, ensures the existence of the smallest between-cluster distance and the largest within-cluster distance.

However, if a system has long chains of phenotypes in which the distances between any two consecutive members of the chain are small but the distance between the ends of the chain is large, we have no way to cluster them so that ε becomes much smaller than the smallest between-cluster distance. In this case, the error estimate for perturbation terms in Theorems 2 or 3 ($C_{\mathbf{h}}$ or $\tilde{C}_{\mathbf{h}}$), in comparison with the leading eigenvalue of the community Jacobian matrix ($\lambda_{\max}$ or $\tilde{\lambda}_{\max}$), can be too large for the approximability condition to be satisfied.

Fortunately, there is effectively no chance of such configurations occurring in ongoing evolutionary dynamics with sufficiently small mutational step sizes as in such dynamics closely similar phenotypes only occur in the early stages of evolutionary branching. (The local coexistence regions that can occur in higher-dimensional trait spaces around the zero fitness contour for particular residents are that narrow that the chance of a mutant landing in them is practically negligible. The more so since the far more common mutants landing outside these coexistence regions will oust all those inside the regions, so there is no chance of the number of coexisting similar phenotypes ever becoming large (Durinx et al. 2008).) Finally, of phenotypes that evolve

towards each other, only one will survive due to competitive exclusion. Therefore, the assumption of well-recognizable and -separated phenotypic clusters is warranted except for a fraction of cases that will be encountered only very exceptionally, as well as transiently, in the scenarios that have our interest.

9.3 LV-approximation for analyzing evolutionary branching in multidimensional trait spaces

As shown in Sect. 3, attractor inheritance in approximate phenotypes directly enables LV-approximations of the population dynamics of the original phenotypes before clustering, similar to the previous studies (Meszéna et al. 2005; Dercole and Rinaldi 2008; Durinx et al. 2008). The derived LV-approximations may be especially useful for extending conditions for evolutionary branching from one-dimensional trait spaces to higher-dimensional trait spaces. In two-dimensional trait spaces, various numerical analyses have shown that phenotypes that are strongly convergence stable, but not evolutionarily stable, also known as strongly attracting invadable ESSes, induce evolutionary branching (e.g., Vukics et al. 2003; Ackermann and Doebeli 2004; Egas et al. 2005; Ravigné et al. 2009; Ito and Dieckmann 2012). Those phenotypes are fixed-point attractors that can be attained by directional evolution causing the convergence of a monomorphic population (Leimar 2009) to them, with sufficient proximity of a set in $\mathcal{S}^2$ enabling the emergence of dimorphisms followed by directional evolution causing the divergence of the two morphs. However, whether an emergent polymorphism can evolutionarily diversify further into visually distinct morphs without collapse has not been proved until recently for higher-dimensional trait spaces. Based on the rational form of invasion-fitness functions in terms of existing phenotypes, which has been derived by LV-approximation (Durinx et al. 2008), Geritz et al. (2016) derived a set of conditions that ensure that such diversifying evolution does not collapse in trait spaces of arbitrary dimension, by describing the initial diversifying evolution with coupled Lande equations (Lande 1979). While those conditions are satisfied by strongly attracting invadable ESSes in two-dimensional trait spaces, the higher-dimensional cases remain to be analyzed further (Geritz et al. 2016).

9.4 Axioms for fitness functions

The analyses in this paper are based on a set of axioms for the fitness-generating functions characterizing ecologically plausible differential equations describing trait-mediated community dynamics. Our set of axioms are similar to the set of properties assumed in Dercole (2016), which are used by him to derive a general procedure for formulating population-dynamical models resulting from individual pairwise interaction. Properties 1, 2, and 3 in Dercole (2016) are identical to our axiom (iii), (iv), and (ii), respectively, while property 4 in Dercole (2016) corresponds to our axiom (i). Dercole's property 4, however, delimits a smaller class of models than ours.

The symmetry axiom (ii) and the reducibility axiom (iii) are no more than consistency conditions, as is the exchangeability axiom (iv). The latter axiom, however, is together with the remaining smoothness axiom (i) and bounded-world axiom (v) the

root cause of the Lotka–Volterra approximabiliy. Indeed, Lemma 1, which is central for deriving the condition for attractor inheritance and LV-approximation, is proved by applying those three axioms (Appendix A). While the bounded-world axiom (v) seems to be well grounded in reality, the smoothness axiom (i) may not hold in an exact sense, because it assumes that the population-dynamical behavior of individuals depends smoothly on their traits and that all ecological interactions are instantaneous. This instantaneousness can arise when the timescale of the life-history dynamics among individuals is much faster than that of their population dynamics. Durinx et al. (2008) have proved attractor inheritance and LV-approximation in physiologically structured models with multiple birth states, in which the timescales of life-history dynamics and population dynamics are not separated. This instills us with cautious optimism that the assumption of instantaneousness we have used in the present paper might be relaxed as well.

Appendix A: Proof of Lemma 1

Here we derive C'_{Fm} in Eq. (3.2) in the main text. We start from the observation that by the smoothness and exchangeability of F and the compactness of the set of community states, i.e., axioms (i), (iv), and (v), there exists a constant C'_{Fn} such that

$$\left| \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} \right| \leq C'_{\mathrm{Fn}} \tag{A.1}$$

for all $i, j = 1, \ldots, N+1$, with

$$C'_{\mathrm{Fn}} \leq \max \left\{ \left| \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} \right| \; \middle| \; i, j = 1, \ldots, N+1, \; \mathbf{n}' \in [0, \eta]^{N+1} \right\}. \tag{A.2}$$

The derivative of $\breve{F}$ with respect to m_j is calculated from

$$\frac{\partial \breve{F}(s_i;\mathbf{s}';\mathbf{m}')}{\partial m_j} = \sum_{k=1}^{N+1} \frac{\partial n_k}{\partial m_j} \frac{\partial \breve{F}(s_i;\mathbf{s}';\mathbf{m}')}{\partial n_k}. \tag{A.3}$$

To calculate the derivative $\frac{\partial n_k}{\partial m_j}$, we substitute Eq. (3.1b),

$$m_k = \varepsilon n_k, \tag{A.4}$$

for $i = M+1, \ldots, N+1$ into Eq. (3.1a) for $i = 1, \ldots, M$,

$$m_k = \sum_{l \in \mathrm{com}(k)} n_l$$

$$= n_k + \sum_{l \in \mathrm{com}(k), l \neq k} n_l$$

$$= n_k + \sum_{l \in \mathrm{com}(k), l \neq k} \frac{1}{\varepsilon} m_l, \tag{A.5}$$

which gives

$$n_k = \begin{cases} m_k - \frac{1}{\varepsilon} \sum\limits_{l \in \mathrm{com}(k), l \neq k} m_l & \text{for } k = 1, \ldots, M \\ \frac{1}{\varepsilon} m_k & \text{for } k = M+1, \ldots, N+1. \end{cases} \tag{A.6}$$

Since $l \in \mathrm{com}(k)$ and $l \neq k$ require $l \in \{M+1, \ldots, N+1\}$, we see for $j = 1, \ldots, M$

$$\frac{\partial n_k}{\partial m_j} = \begin{cases} 1 & \text{for } k = j \\ 0 & \text{otherwise}, \end{cases} \tag{A.7}$$

and for $j = M+1, \ldots, N+1$

$$\frac{\partial n_k}{\partial m_j} = \begin{cases} -\frac{1}{\varepsilon} & \text{for } k = \mathrm{cid}(j) \text{ (i.e., } j \in \mathrm{com}(k)) \\ \frac{1}{\varepsilon} & \text{for } k = j \\ 0 & \text{otherwise}. \end{cases} \tag{A.8}$$

By substituting Eq. (A.7) into Eq. (A.3), we find for $j = 1, \ldots, M$,

$$\frac{\partial \breve{F}(s_i;\mathbf{s}';\mathbf{m}')}{\partial m_j} = \frac{\partial F(s_i;\mathbf{s}';\mathbf{n}')}{\partial n_j}$$

$$\left| \frac{\partial \breve{F}(s_i;\mathbf{s}';\mathbf{m}')}{\partial m_j} \right| = \left| \frac{\partial F(s_i;\mathbf{s}';\mathbf{n}')}{\partial n_j} \right| \leq C'_{\mathrm{Fn}}. \tag{A.9}$$

Similarly, by substituting Eq. (A.8) into Eq. (A.3), we find for $j = M+1, \ldots, N+1$

$$
\begin{aligned}
\frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} &= \frac{\partial n_j}{\partial m_j} \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} + \frac{\partial n_{\text{cid}(j)}}{\partial m_j} \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_{\text{cid}(j)}} \\
&= \frac{1}{\varepsilon} \left[\frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} - \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_{\text{cid}(j)}} \right].
\end{aligned} \tag{A.10}
$$

We use Taylor's theorem and the exchangeability axiom (iv) to derive an estimate for the term in square brackets,

$$
\begin{aligned}
\frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} - \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_{\text{cid}(j)}} &= \left[\frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} - \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_{\text{cid}(j)}} \right]_{s_j = s_{\text{cid}(j)}} \\
&\quad + \left[\frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j} - \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)}} \right]_{s_j = s_{jT}} (s_j - s_{\text{cid}(j)}) \\
&= \left[\frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j} - \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)}} \right]_{s_j = s_{jT}} (s_j - s_{\text{cid}(j)}) \\
&= \left[\frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j} - \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)}} \right]_{s_j = s_{jT}} \rho_j \varepsilon
\end{aligned} \tag{A.11}
$$

for some appropriately chosen $s_{jT} \in [s_j, s_{\text{cid}(j)}]$, where $\rho_j = (s_j - s_{\text{cid}(j)})/\varepsilon$ with $|\rho_j| \leq 1$ (because by assumption within-cluster phenotypic differences do not exceed ε). By the smoothness of F, there exists a constant C'_{Fsn} such that

$$
\begin{aligned}
&\left| \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j} - \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)}} \right|_{s_j = s_{jT}} \\
&\leq \max \left\{ \left| \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j} - \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)}} \right|_{s_j = s_{jT}} \middle| \begin{array}{l} i = 1, \ldots, N+1, \ j = M+1, \ldots, N+1 \\ \mathbf{n}' \in [0, \eta]^{N+1}, \ s_{jT} \in [s_j, s_{\text{cid}(j)}] \end{array} \right\} \\
&=: C'_{\text{Fsn}},
\end{aligned} \tag{A.12}
$$

Thus, substituting Eqs. (A.11) and (A.12) into Eq. (A.10) yields for $j = M+1, \ldots, N+1$

$$
\begin{aligned}
\frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} &= \left[\frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j} - \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)}} \right]_{s_j = s_{jT}} \rho_j, \\
\left| \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} \right| &\leq C'_{\text{Fsn}}.
\end{aligned} \tag{A.13}
$$

Finally, from Eqs. (A.9) and (A.13),

$$\left| \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} \right| \leq \max\{C'_{\text{Fn}}, C'_{\text{Fsn}}\} =: C'_{\text{Fm}}. \tag{A.14}$$

$\square$

Appendix B: Derivation of Eq. (3.3a) and proof of Lemma 2

B.1 Some preliminary estimates

To get an estimate for the remainder R in

$$G(\mathbf{x} + \Delta\mathbf{x}) = G(\mathbf{x}) + DG(\mathbf{x})\Delta\mathbf{x} + R, \tag{B.1}$$

$G : \mathbb{R}^N \to \mathbb{R}$, consider the function $g : \mathbb{R} \to \mathbb{R} : g(z) = G(\mathbf{x} + z\Delta\mathbf{x})$. Then by Taylor's theorem,

$$g(z) = g(0) + Dg(0)z + \tfrac{1}{2}D^2g(\theta_{\text{T}}z)z^2, \tag{B.2}$$

with θ_{T} some appropriately chosen number between 0 and 1. By the chain rule,

$$g(0) = G(\mathbf{x}), \, Dg(0) = DG(\mathbf{x})\Delta\mathbf{x}, \, D^2g(\theta_{\text{T}}z) = \Delta\mathbf{x}^{\text{T}}D^2G(\mathbf{x} + \theta_{\text{T}}z\Delta\mathbf{x})\Delta\mathbf{x}. \tag{B.3}$$

Now we use that

$$\Delta\mathbf{x}^T D^2 G(\mathbf{x} + \theta_{\text{T}}z\Delta\mathbf{x})\Delta\mathbf{x} \leq \left\| D^2G(\mathbf{x} + \theta_{\text{T}}z\Delta\mathbf{x}) \right\|_{\text{Q}} |\Delta\mathbf{x}|^2 \tag{B.4}$$

with $\|\cdot\|_{\text{Q}}$ defined for an arbitrary $K \times K$ symmetric matrix $\mathbf{B}$ by

$$\|\mathbf{B}\|_{\text{Q}} := \max\left\{ \left| \mathbf{u}^{\text{T}}\mathbf{B}\mathbf{u} \right| \, \Big| \, |\mathbf{u}| = 1 \right\}, \tag{B.5}$$

i.e., the absolute value of the dominant eigenvalue. By using Eq. (B.5), we see

$$|R| \leq \frac{1}{2}\max\left\{ \left\| D^2G(\mathbf{x} + \theta_{\text{T}}\Delta\mathbf{x}) \right\|_{\text{Q}} \, \Big| \, 0 \leq \theta_{\text{T}} \leq 1 \right\}|\Delta\mathbf{x}|^2. \tag{B.6}$$

When it is given that $|\Delta\mathbf{x}| < \delta$ this translates into the uniform estimates

$$|R| < \frac{1}{2}\max\left\{ \left\| D^2G(\mathbf{x} + \mathbf{x}') \right\|_{\text{Q}} \, \Big| \, |\mathbf{x}'| \leq \delta \right\}\delta^2. \tag{B.7}$$

Note that $\|\mathbf{B}\|_{\text{Q}}$ satisfies

$$\|\mathbf{B}\|_{\text{Q}} \leq K B_{\text{max}} \tag{B.8}$$

with $B_{\max} = \max_{i,\,j \in \{1,\ldots,K\}} \left(\left| B_{ij} \right| \right)$, because

$$
\left| \mathbf{u}^{\mathsf{T}} \mathbf{B} \mathbf{u} \right| = \left| \sum_{i=1}^{K} \sum_{j=1}^{K} B_{ij} u_i u_j \right|
$$

$$
\leq B_{\max} \sum_{i=1}^{K} \sum_{j=1}^{K} |u_i||u_j| = B_{\max} \left[\sum_{i=1}^{K} |u_i|^2 + 2 \sum_{i=1}^{K-1} \sum_{j>i}^{K} |u_i||u_j| \right]
$$

$$
= B_{\max} \left[\sum_{i=1}^{K} |u_i|^2 + \sum_{i=1}^{K-1} \sum_{j>i}^{K} |u_i|^2 + \sum_{i=1}^{K-1} \sum_{j>i}^{K} |u_j|^2 - \sum_{i=1}^{K-1} \sum_{j>i}^{K} \left(|u_i| - |u_j| \right)^2 \right]
$$

$$
= B_{\max} \left[\begin{array}{l} \sum\limits_{i=1}^{K} |u_i|^2 + [(K-1)|u_1|^2 + \cdots + |u_{K-1}|^2] \\ \\ + [(|u_2|^2 + \cdots + |u_K|^2) + \cdots + |u_K|^2] - \sum\limits_{i=1}^{K-1} \sum\limits_{j>i}^{K} \left(|u_i| - |u_j| \right)^2 \end{array} \right]
$$

$$
= B_{\max} \left[\sum_{i=1}^{K} |u_i|^2 + (K-1) \sum_{i=1}^{K} |u_i|^2 - \sum_{i=1}^{K-1} \sum_{j>i}^{K} \left(|u_i| - |u_j| \right)^2 \right]
$$

$$
\leq B_{\max} \left[K \sum_{i=1}^{K} |u_i|^2 - \sum_{i=1}^{K-1} \sum_{j>i}^{K} \left(|u_i| - |u_j| \right)^2 \right]
$$

$$
\leq K B_{\max} |\mathbf{u}|^2. \tag{B.9}
$$

B.2 Proof of Eq. (3.3f)

By Taylor's theorem, the $\widehat{F}_j$ for $j = M+1, \ldots, N+1$ can be written as

$$
\widehat{F}_j = F(z_j; \mathbf{s}'; \widehat{n'})
$$

$$
= F(z_j; \mathbf{s}'; \widehat{n'}) \Big|_{z_j = s_{\text{cid}(j)}} + \frac{\partial F(z_j; \mathbf{s}'; \widehat{n'})}{\partial z_j} \Big|_{z_j = s_{jT}} (s_j - s_{\text{cid}(j)})
$$

$$
= \frac{\partial F(z_j; \mathbf{s}'; \widehat{n'})}{\partial z_j} \Big|_{z_j = s_{jT}} \rho_j \varepsilon, \tag{B.10}
$$

with some appropriately chosen parameter $s_{jT} \in [s_j, s_{\text{cid}(j)}]$. Hence,

$$
\frac{1}{\varepsilon} \left| \widehat{F}_j \right| = \frac{1}{\varepsilon} \left| \frac{\partial F(z_j; \mathbf{s}'; \widehat{n'})}{\partial z_j} \Big|_{z_j = s_{jT}} \rho_j \varepsilon \right|
$$

$$\leq \rho_j \max\left\{ \left. \left| \frac{\partial F(z_j; \mathbf{s}'; \widehat{n}')}{\partial z_j} \right| \right|_{z_j = s_{jT}} \middle| s_{jT} \in [s_j, s_{\mathrm{cid}(j)}] \right\} =: \widehat{C'_{\mathrm{Fz},j}}, \qquad (\mathrm{B}.11)$$

and

$$\widehat{C'_{\mathrm{Fz}}} := \max_{j \in \{M+1,\ldots,N+1\}} \widehat{C'_{\mathrm{Fz},j}}$$

$$= \max\left\{ \left. \left| \frac{\partial F(z_j; \mathbf{s}'; \widehat{n}')}{\partial z_j} \right| \right|_{z_j = s_{jT}} \middle| j = M+1,\ldots,N+1, \ s_{jT} \in [s_j, s_{\mathrm{cid}(j)}] \right\}$$

$$(\mathrm{B}.12)$$

because $\rho_j = 0$ for $j = 1, \ldots, M$. $\qquad\square$

B.3 Proof of Lemma 2

Using Taylor's theorem, Eq. (3.3d) in the main text can be transformed to

$$R_i = \frac{1}{2}(\mathbf{m}' - \widehat{m}')^{\mathrm{T}} \left. \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\mathrm{T}}} \right|_{\mathbf{m}' = \mathbf{m}'_{\mathrm{T}}} (\mathbf{m}' - \widehat{m}'), \qquad (\mathrm{B}.13)$$

where $\mathbf{m}'_{\mathrm{T}} = \theta_{\mathrm{T}}\left(\mathbf{m}' - \widehat{m}'\right) + \widehat{m}'$ with some appropriately chosen $\theta_{\mathrm{T}} \in [0, 1]$. By Eq. (3.4a) in the main text,

$$\left| \mathbf{m}' - \widehat{m}' \right| = \sqrt{\left| \mathbf{m} - \widehat{m} \right|^2 + \left| (m_{M+1}, \ldots, m_{N+1})^{\mathrm{T}} - (\widehat{m}_{M+1}, \ldots, \widehat{m}_{N+1})^{\mathrm{T}} \right|^2}$$

$$= \sqrt{\left| \mathbf{m} - \widehat{m} \right|^2 + \left| (\varepsilon n_{M+1}, \ldots, \varepsilon n_{N+1})^{\mathrm{T}} - (\varepsilon \widehat{n}_{M+1}, \ldots, \varepsilon \widehat{n}_{N+1})^{\mathrm{T}} \right|^2}$$

$$\leq \sqrt{\left| \mathbf{m} - \widehat{m} \right|^2 + \varepsilon^2 \max_{i \in \{M+1,\ldots,N+1\}} \left(\left| n_i - \widehat{n}_i \right|^2 \right)(N - M + 1)}$$

$$\leq \sqrt{\varepsilon^2 C_{\mathbf{m}}^2 + \varepsilon^2 \eta^2 (N - M + 1)}$$

$$\leq \varepsilon \sqrt{C_{\mathbf{m}}^2 + (N - M + 1)\eta^2} =: \varepsilon C'_{\mathbf{m}}. \qquad (\mathrm{B}.14)$$

Thus, the result from Appendix B.1 translates as

$$|R_i| = \frac{1}{2} \left| (\mathbf{m}' - \widehat{m}')^{\mathrm{T}} \left. \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\mathrm{T}}} \right|_{\mathbf{m}' = \mathbf{m}'^{\mathrm{T}}} (\mathbf{m}' - \widehat{m}') \right|$$

$$\leq \frac{1}{2} \left\| \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{n}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\mathrm{T}}} \right|_{\mathbf{m}'=\mathbf{m}'_{\mathrm{T}}} \right\|_{\mathrm{Q}} \left| \mathbf{m}' - \widehat{m}' \right|^2$$

$$\leq \frac{1}{2} C'_{\mathrm{Fmm}} \left| \mathbf{m}' - \widehat{m}' \right|^2$$

$$\leq \frac{1}{2} \varepsilon^2 C'_{\mathrm{Fmm}} C'^2_{\mathbf{m}}, \tag{B.15}$$

where, as proved in the next subsection,

$$C'_{\mathrm{Fmm}} := \max \left\{ \left\| \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\mathrm{T}}} \right\|_{\mathrm{Q}} \middle| \begin{array}{l} i = 1, \ldots, N+1, \\ m_1, \ldots, m_M \in [0, \eta], \\ m_{M+1}, \ldots, m_{N+1} \in [0, \varepsilon\eta] \end{array} \right\}$$

$$\leq (N+1)\max \left\{ C'_{\mathrm{Fnn}}, 2C'_{\mathrm{Fsnn}}, C'_{\mathrm{Fssnn}} \right\}. \tag{B.16}$$

$\square$

B.4 Finding C'_{Fmm}

Here we prove Eq. (B.16). Analogously to Appendix A, we start from the observation that by the smoothness axiom (i) and bounded-world axiom (v) there exists a constant C'_{Fnn} such that

$$\left| \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_k \partial n_j} \right| \leq C'_{\mathrm{Fnn}} \tag{B.17}$$

for all $i, j, k = 1, \ldots, N+1$, with

$$C'_{\mathrm{Fnn}} := \max \left\{ \left| \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_k \partial n_j} \right| \middle| i, j, k = 1, \ldots, N+1, \ \mathbf{n}' \in [0, \eta]^{N+1} \right\}. \tag{B.18}$$

From Eqs. (A.9) and (A.10), the second derivative of the fitness-generating function with respect to m_j and m_k is expressed as

$$\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} = \begin{cases} \frac{\partial}{\partial n_k} \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} & \text{for } k = 1, \ldots, M \\ \frac{1}{\varepsilon} \left\{ \frac{\partial}{\partial n_k} \left\{ \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} \right\} - \frac{\partial}{\partial n_{\mathrm{cid}(k)}} \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} \right\} & \text{for } k = M+1, \ldots, N+1, \end{cases} \tag{B.19}$$

while Eqs. (A.9) and (A.13) are combined into

$$\frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_j} = \begin{cases} \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_j} & \text{for } j = 1, \ldots, M \\ \left[\frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j} - \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\mathrm{cid}(j)}} \right]_{s_j = s_{j\mathrm{T}}} \rho_j & \text{for } j = M+1, \ldots, N+1 \end{cases} \tag{B.20}$$

with some appropriately chosen $s_{jT} \in [s_j, s_{\mathrm{cid}(j)}]$. By substituting Eqs. (B.20) into (B.19), we see for $j, k = 1, \ldots, M$

$$\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} = \frac{\partial F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_k \partial n_j}, \tag{B.21}$$

and for $j = M + 1, \ldots, N + 1$ and $k = 1, \ldots, M$ (and equivalently for $k = M + 1, \ldots, N + 1$ and $j = 1, \ldots, M$)

$$\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} = \left[\frac{\partial^3 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_k \partial s_j \partial n_j} - \frac{\partial^3 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial n_k \partial s_j \partial n_{\mathrm{cid}(j)}} \right]_{s_j = s_{jT}} \rho_j. \tag{B.22}$$

For $j, k = M + 1, \ldots, N + 1$, we first transform Eq. (B.19) and then substitute Eq. (B.20) into the equation,

$$
\begin{aligned}
\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} &= \frac{1}{\varepsilon} \left[\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial n_k \partial m_j} - \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial n_{\mathrm{cid}(k)} \partial m_j} \right]_{s_k = s_{\mathrm{cid}(k)}} \\
&\quad + \frac{1}{\varepsilon} \left\{ \frac{\partial}{\partial s_k} \left[\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial n_k \partial m_j} - \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial n_{\mathrm{cid}(k)} \partial m_j} \right] \right\}_{s_k = s_{kT}} (s_k - s_{\mathrm{cid}(k)}) \\
&= \rho_k \left\{ \frac{\partial}{\partial s_k} \left[\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial n_k \partial m_j} - \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial n_{\mathrm{cid}(k)} \partial m_j} \right] \right\}_{s_k = s_{kT}} \\
&= \rho_k \rho_j \left[\frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial n_k \partial s_j \partial n_j} - \frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial n_k \partial s_j \partial n_{\mathrm{cid}(j)}} \right. \\
&\quad \left. - \frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial n_{\mathrm{cid}(k)} \partial s_j \partial n_j} + \frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial n_{\mathrm{cid}(k)} \partial s_j \partial n_{\mathrm{cid}(j)}} \right]_{s_k = s_{kT}, s_j = s_{jT}},
\end{aligned}
\tag{B.23}
$$

with some appropriately chosen $s_{kT} \in [s_k, s_{\mathrm{cid}(k)}]$. Hence, from Eqs. (B.21) to (B.23), we find for $j, k = 1, \ldots, M$

$$\left| \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} \right| \leq C'_{\mathrm{Fnn}}, \tag{B.24}$$

for $j = M+1, \ldots, N+1$ and $k = 1, \ldots, M$ (and equivalently for $k = M+1, \ldots, N+1$ and $j = 1, \ldots, M$)

$$\left| \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} \right| \leq C'_{\mathrm{Fsnn}} \tag{B.25}$$

with

$$
C'_{\text{Fsnn}} := \max \left\{ \left| \frac{\partial^3 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_k \partial n_j} - \frac{\partial^3 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_k \partial n_{\text{cid}(j)}} \right|_{s_j = s_{jT}} \left| \begin{array}{l} i = 1, \ldots, N+1, \\ j = M+1, \ldots, N+1, \\ k = 1, \ldots, M, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \ s_{jT} \in [s_j, s_{\text{cid}(j)}] \end{array} \right. \right\},
$$

(B.26)

and for $j, k = M+1, \ldots, N+1$

$$
\left| \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} \right| \leq C'_{\text{Fssnn}}
$$

(B.27)

with

$$
C'_{\text{Fssnn}} := \max \left\{ \left| \begin{array}{l} \frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial s_j \partial n_k \partial n_j} - \frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial s_j \partial n_k \partial n_{\text{cid}(j)}} \\ - \frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial s_j \partial n_{\text{cid}(k)} \partial n_j} + \frac{\partial^4 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial s_k \partial s_j \partial n_{\text{cid}(k)} \partial n_{\text{cid}(j)}} \end{array} \right|_{s_k = s_{kT}, s_j = s_{jT}} \left| \begin{array}{l} i = 1, \ldots, N+1, \\ j, k = M+1, \ldots, N+1, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \\ s_{jT} \in [s_j, s_{\text{cid}(j)}], \\ s_{kT} \in [s_k, s_{\text{cid}(k)}] \end{array} \right. \right\}.
$$

(B.28)

Collecting Eqs. (B.24), (B.25), and (B.27), we find

$$
\left| \frac{\partial \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_k \partial m_j} \right| \leq \max \left\{ C'_{\text{Fnn}}, C'_{\text{Fsnn}}, C'_{\text{Fssnn}} \right\} =: C'_{\text{Fmm}}.
$$

(B.29)

Finally, for

$$
\frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\text{T}}} = \begin{pmatrix} \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_1 \partial m_1} & \cdots & \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_1 \partial m_{N+1}} \\ \cdots & \ddots & \cdots \\ \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_{N+1} \partial m_1} & \cdots & \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial m_{N+1} \partial m_{N+1}} \end{pmatrix},
$$

(B.30)

we find its upper bound by using Eq. (B.8), as

$$
\left\| \frac{\partial^2 \breve{F}(s_i; \mathbf{s}'; \mathbf{m}')}{\partial \mathbf{m}' \partial \mathbf{m}'^{\text{T}}} \right\|_{\text{Q}} \leq (N+1) C'_{\text{Fmm}} = (N+1) \max \left\{ C'_{\text{Fnn}}, C'_{\text{Fsnn}}, C'_{\text{Fssnn}} \right\} =: C'_{\text{Fmm}}.
$$

(B.31)

$\square$

Appendix C: Proof of Lemmas 3 and 7

C.1 Expansion of *f*

We assume that non-representative phenotypes, s_j for $j = M + 1, \ldots, N + 1$, are functions of ε, $s_j = s_{\text{cid}(j)} + \varepsilon\rho_j$, so that when $\varepsilon \to 0$ they converge to their representative phenotypes, $s_{\text{cid}(j)}$. Then $\mathbf{s}'$ is expressed as $\mathbf{s}' = \mathbf{s}'_a + \varepsilon\boldsymbol{\rho}'$ with $\mathbf{s}'_a := (s_1, \ldots, s_M, s_{\text{cid}(M+1)}, \ldots, s_{\text{cid}(N+1)})^{\text{T}}$ and $\boldsymbol{\rho}' := (0, \ldots, 0, \rho_{M+1}, \ldots, \rho_{N+1})^{\text{T}}$. Notice that $|\rho_j| \leq 1$ because within-cluster phenotypic differences are assumed not to exceed ε, i.e., $|s_j - s_{\text{cid}(j)}| \leq \varepsilon$ for all $j = 1, \ldots, N + 1$. By Taylor's theorem and the exchangeability axiom (vi), the fitness of s_j can be expanded in ε around $\varepsilon = 0$, as

$$
\begin{aligned}
F(s_j; \mathbf{s}'; \mathbf{n}') &= \left. F(s_{\text{cid}(j)} + \varepsilon\rho_j; \mathbf{s}'_a + \varepsilon\boldsymbol{\rho}'; \mathbf{n}') \right|_{\varepsilon=0} + \varepsilon \left. \frac{\partial F(s_{\text{cid}(j)} + \varepsilon\rho_j; \mathbf{s}'_a + \varepsilon\boldsymbol{\rho}'; \mathbf{n}')}{\partial\varepsilon} \right|_{\varepsilon=\varepsilon_{\text{T}j}} \\
&= F(s_{\text{cid}(j)}; \mathbf{s}_a; \mathbf{m}) + \varepsilon F_{\varepsilon j}, \quad\quad\quad\quad\quad\quad\quad\quad (\text{C.1})
\end{aligned}
$$

with some appropriately chosen $\varepsilon_{\text{T}j} \in [0, \varepsilon]$ and

$$
F_{\varepsilon j} := \left. \frac{\partial F(s_{\text{cid}(j)} + \varepsilon\rho_j; \mathbf{s}'_a + \varepsilon\boldsymbol{\rho}'; \mathbf{n}')}{\partial\varepsilon} \right|_{\varepsilon=\varepsilon_{\text{T}j}}. \quad\quad\quad\quad (\text{C.2})
$$

Here, $|F_{\varepsilon j}|$ satisfies

$$
|F_{\varepsilon j}| \leq \max\left\{ \left. \left| \frac{\partial F(s_{\text{cid}(j)} + \varepsilon\rho_j; \mathbf{s}'_a + \varepsilon\boldsymbol{\rho}'; \mathbf{n}')}{\partial\varepsilon} \right| \right|_{\varepsilon=\varepsilon_{\text{T}j}} \left| \begin{matrix} j = 1, \ldots, N + 1, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \ \varepsilon_{\text{T}j} \in [0, \varepsilon] \end{matrix} \right. \right\} =: C'_{\text{F}\varepsilon}. \quad (\text{C.3})
$$

By using Eq. (C.1), $f(s_i; \mathbf{s}'; \mathbf{n}')$ from Eq. (4.1b) can be expressed as

$$
\begin{aligned}
f(s_i; \mathbf{s}'; \mathbf{n}') &= \sum_{j \in \text{com}(i)} p_j F(s_j; \mathbf{s}'; \mathbf{n}') \\
&= F(s_i; \mathbf{s}_a; \mathbf{m}) + \varepsilon \sum_{j \in \text{com}(i)} p_j F_{\varepsilon j}. \quad\quad\quad\quad (\text{C.4})
\end{aligned}
$$

We further expand $F(s_i; \mathbf{s}_a; \mathbf{m})$ around $\mathbf{m} = \widehat{m} = (\widehat{m}_1, \ldots, \widehat{m}_M)^{\text{T}}$ as

$$
F(s_i; \mathbf{s}_a; \mathbf{m}) = F(s_i; \mathbf{s}_a; \widehat{m}) + \mathbf{b}_i^{\text{T}}(\mathbf{m} - \widehat{m}) + (\mathbf{m} - \widehat{m})^{\text{T}}\mathbf{F}_{\text{mm}i}(\mathbf{m} - \widehat{m}), \quad (\text{C.5})
$$

where

$$
\mathbf{b}_i^{\text{T}} := (b_{i1}, \ldots, b_{iM}) = \left. \left(\frac{\partial F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial m_1}, \ldots, \frac{\partial F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial m_M} \right) \right|_{\mathbf{m}=\widehat{m}}, \quad (\text{C.6})
$$

and

$$\mathbf{F}_{\mathrm{mm}i} := \left. \frac{\partial^2 F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial \mathbf{m} \partial \mathbf{m}^{\mathrm{T}}} \right|_{\mathbf{m}=\mathbf{m}_{\mathrm{T}}} = \begin{pmatrix} \frac{\partial^2 F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial m_1^2} & \cdots & \frac{\partial^2 F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial m_1 \partial m_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial m_M \partial m_1} & \cdots & \frac{\partial^2 F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial m_M^2} \end{pmatrix}_{\mathbf{m}=\mathbf{m}_{\mathrm{T}}}, \quad (\text{C.7})$$

where $\mathbf{m}_{\mathrm{T}} := \theta_{\mathrm{T}}(\mathbf{m} - \widehat{m}) + \widehat{m}$ with some appropriately chosen $\theta_{\mathrm{T}} \in [0, 1]$. Notice that by Eq. (C.1) $F(s_i; \mathbf{s}_a; \widehat{m})$ satisfies

$$F(s_i; \mathbf{s}'; \widehat{n'}) = F(s_i; \mathbf{s}_a; \widehat{m}) + \varepsilon \left. \frac{\partial F(s_i + \varepsilon \rho_i; \mathbf{s}'_a + \varepsilon \boldsymbol{\rho}'; \widehat{n'})}{\partial \varepsilon} \right|_{\varepsilon = \varepsilon_{\mathrm{T}i}}$$

$$= F(s_i; \mathbf{s}_a; \widehat{m}) + \varepsilon \widehat{F_{\varepsilon i}}, \quad (\text{C.8})$$

which gives

$$F(s_i; \mathbf{s}_a; \widehat{m}) = F(s_i; \mathbf{s}'; \widehat{n'}) - \varepsilon \widehat{F_{\varepsilon i}}. \quad (\text{C.9})$$

Thus, combining Eqs. (C.4), (C.5), and (C.9), we obtain

$$f(s_i; \mathbf{s}'; \mathbf{n}') = \mathbf{b}_i^{\mathrm{T}}(\mathbf{m} - \widehat{m}) + (\mathbf{m} - \widehat{m})^{\mathrm{T}} \mathbf{F}_{\mathrm{mm}i}(\mathbf{m} - \widehat{m})$$

$$+ \varepsilon \left[\frac{1}{\varepsilon} F(s_i; \mathbf{s}'; \widehat{n'}) - \widehat{F_{\varepsilon i}} + \sum_{j \in \mathrm{com}(i)} p_j F_{\varepsilon j} \right]$$

$$= \mathbf{b}_i^{\mathrm{T}}(\mathbf{m} - \widehat{m}) + r_{\mathrm{f}i} \left| \mathbf{m} - \widehat{m} \right|^2 + \varepsilon h_{\mathrm{f}i}, \quad (\text{C.10})$$

where

$$r_{\mathrm{f}i} := \frac{(\mathbf{m} - \widehat{m})^{\mathrm{T}} \mathbf{F}_{\mathrm{mm}i}(\mathbf{m} - \widehat{m})}{\left| \mathbf{m} - \widehat{m} \right|^2},$$

$$h_{\mathrm{f}i} := \frac{1}{\varepsilon} F(s_i; \mathbf{s}'; \widehat{n'}) - \widehat{F_{\varepsilon i}} + \sum_{j \in \mathrm{com}(i)} p_j F_{\varepsilon j}. \quad (\text{C.11})$$

Here

$$|r_{\mathrm{f}i}| = \left| \frac{(\mathbf{m} - \widehat{m})^{\mathrm{T}} \mathbf{F}_{\mathrm{mm}i}(\mathbf{m} - \widehat{m})}{\left| \mathbf{m} - \widehat{m} \right|^2} \right|$$

$$\leq \max\left\{\left.\left\|\frac{\partial^2 F(s_i; \mathbf{s}_a; \mathbf{m})}{\partial \mathbf{m} \partial \mathbf{m}^{\mathrm{T}}}\right\|_Q\right| i = 1, \ldots, M, \mathbf{m} \in [0, \eta]^M\right\} =: C_{\mathbf{Fmm}} \quad (C.12)$$

and by Eqs. (C.3) and (3.3f) in the main text, i.e., $\left|F(s_i; \mathbf{s}'; \widehat{n'})\right| = \left|\widehat{F_i}\right| = 0$ for

$i = 1, \ldots, N$ and $\left|F(s_{N+1}; \mathbf{s}'; \widehat{n'})\right| = \left|\widehat{F_{N+1}}\right| \leq C'_{\mathrm{Fz}} \varepsilon,$

$$|h_{\mathrm{f}i}| = \left|\frac{1}{\varepsilon} F(s_i; \mathbf{s}'; \widehat{n'}) - \widehat{F_{\varepsilon i}} + \sum_{j \in \mathrm{com}(i)} p_j F_{\varepsilon j}\right|$$

$$\leq \frac{1}{\varepsilon}\left|F(s_i; \mathbf{s}'; \widehat{n'})\right| + \left|\widehat{F_{\varepsilon i}}\right| + \left|\sum_{j \in \mathrm{com}(i)} p_j F_{\varepsilon j}\right|$$

$$\leq \widehat{C'_{\mathrm{Fz}}} + 2C'_{\mathrm{F}\varepsilon}. \quad (C.13)$$

Substituting Eq. (C.10) into Eq. (4.2) in the main text yields

$$\frac{d\mathbf{m}}{dt} = \mathrm{diag}(\mathbf{m})\left[\mathbf{B}(\mathbf{m} - \widehat{m}) + \mathbf{r}_{\mathrm{f}}\left|\mathbf{m} - \widehat{m}\right|^2 + \varepsilon \mathbf{h}_{\mathrm{f}}\right], \quad (C.14)$$

where

$$\mathbf{B} := \begin{pmatrix} \mathbf{b}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{b}_M^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} b_{11} & \ldots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{M1} & \ldots & b_{MM} \end{pmatrix} \quad (C.15)$$

with $\mathbf{b}_i^{\mathrm{T}}$ defined by Eq. (C.6), $\mathbf{h}_{\mathrm{f}} = (h_{\mathrm{f}1}, \ldots, h_{\mathrm{f}M})^{\mathrm{T}}$, and $\mathbf{r}_{\mathrm{f}} = (r_{\mathrm{f}1}, \ldots, r_{\mathrm{f}M})^{\mathrm{T}}$, with

$$|\mathbf{r}_{\mathrm{f}}| = \sqrt{\sum_{i=1}^M |\mathbf{r}_{\mathrm{f}i}|^2} \leq \sqrt{M} C_{\mathbf{Fmm}} =: C_{\mathbf{rf}},$$

$$|\mathbf{h}_{\mathrm{f}}| = \sqrt{\sum_{i=1}^M |\mathbf{h}_{\mathrm{f}i}|^2} \leq \sqrt{M} \max_i(|\mathbf{h}_{\mathrm{f}i}|) = \sqrt{M}\left[\widehat{C'_{\mathrm{Fz}}} + 2C'_{\mathrm{F}\varepsilon}\right] =: C_{\mathbf{hf}}. \quad (C.16)$$

C.2 Expansion of dm/dt

Eq. (C.14) is further transformed into

$$\frac{d\mathbf{m}}{dt} = \mathrm{diag}(\mathbf{m})\mathbf{B}(\mathbf{m} - \widehat{m}) + \mathrm{diag}(\mathbf{m})\mathbf{r}_{\mathrm{f}}\left|\mathbf{m} - \widehat{m}\right|^2 + \varepsilon \mathrm{diag}(\mathbf{m})\mathbf{h}_{\mathrm{f}}$$

$$= \mathrm{diag}\left(\left[\mathbf{m} - \widehat{m}\right] + \widehat{m}\right)\mathbf{B}(\mathbf{m} - \widehat{m}) + \mathrm{diag}(\mathbf{m})\mathbf{r}_{\mathrm{f}}\left|\mathbf{m} - \widehat{m}\right|^2 + \varepsilon \mathrm{diag}(\mathbf{m})\mathbf{h}_{\mathrm{f}}$$

$$= \text{diag}(\widehat{m})\mathbf{B}(\mathbf{m} - \widehat{m})$$

$$+ \left[\text{diag}(\mathbf{m} - \widehat{m})\mathbf{B}(\mathbf{m} - \widehat{m}) + \text{diag}(\mathbf{m})\mathbf{r}_f \left| \mathbf{m} - \widehat{m} \right|^2 \right] + \varepsilon \text{diag}(\mathbf{m})\mathbf{h}_f$$

$$= \mathbf{J}(\mathbf{m} - \widehat{m}) + \mathbf{r}_m \left| \mathbf{m} - \widehat{m} \right|^2 + \varepsilon \mathbf{h}_m \tag{C.17}$$

where

$$\mathbf{J} := \text{diag}(\widehat{m})\mathbf{B} = \begin{pmatrix} \widehat{m}_1 b_{11} & \dots & \widehat{m}_1 b_{1M} \\ \dots & \ddots & \dots \\ \widehat{m}_M b_{M1} & \dots & \widehat{m}_M b_{MM} \end{pmatrix} \tag{C.18}$$

is the Jacobian matrix at $\widehat{m}$, and $\mathbf{r_m} = (r_{m1}, \dots, r_{mM})^{\mathrm{T}}$ and $\mathbf{h_m} = (h_{m1}, \dots, h_{mM})^{\mathrm{T}}$ are given by

$$\mathbf{r}_m := \frac{\text{diag}(\mathbf{m} - \widehat{m})\mathbf{B}(\mathbf{m} - \widehat{m})}{\left| \mathbf{m} - \widehat{m} \right|^2} + \text{diag}(\mathbf{m})\mathbf{r}_f,$$

$$\mathbf{h}_m := \text{diag}(\mathbf{m})\mathbf{h}_f, \tag{C.19}$$

where from Eq. (C.16) we find

$$|\mathbf{r}_m| \leq \frac{\left| \text{diag}(\mathbf{m} - \widehat{m})\mathbf{B}(\mathbf{m} - \widehat{m}) \right|}{\left| \mathbf{m} - \widehat{m} \right|^2} + |\text{diag}(\mathbf{m})\mathbf{r}_f|$$

$$\leq \frac{\left\| \text{diag}(\mathbf{m} - \widehat{m}) \right\| \|\mathbf{B}\|}{\left| \mathbf{m} - \widehat{m} \right|} + \|\text{diag}(\mathbf{m})\| |\mathbf{r}_f|$$

$$= \frac{\max\limits_{i=1,\dots M}\left(\left| m_i - \widehat{m}_i \right| \right) \|\mathbf{B}\|}{\left| \mathbf{m} - \widehat{m} \right|} + \max\limits_{i=1,\dots M}(|m_i|)|\mathbf{r}_f|$$

$$\leq \sqrt{\frac{\max\limits_{i=1,\dots M}\left(\left| m_i - \widehat{m}_i \right| \right)^2}{\sum_{i=1}^{M} \left| m_i - \widehat{m}_i \right|^2}} \|\mathbf{B}\| + \eta |\mathbf{r}_f| \leq \|\mathbf{B}\| + \eta C_{rf} = \|\mathbf{B}\| + \eta \sqrt{MC}_{\mathbf{Fmm}} =: C_{rm},$$

$$|\mathbf{h}_m| \leq \|\text{diag}(\mathbf{m})\| |\mathbf{h}_f| \leq \max\limits_{i=1,\dots M}(|m_i|)|\mathbf{h}_f| \leq \eta C_{hf} = \eta \sqrt{M} \left[\widehat{C}'_{Fz} + 2C'_{F\varepsilon} \right] =: C_{hm} \tag{C.20}$$

$\square$

Appendix D: Finding $C_{\mathbf{r}}$ and $C_{\mathbf{h}}$

Substituting $\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{T}} = \mathbf{P}(\mathbf{m} - \widehat{m})$ into Eq. (4.3a) in the main text gives Eq. (4.4), $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{A}\mathbf{x} + \mathbf{r}|\mathbf{x}|^2 + \varepsilon\mathbf{h}$, where $\mathbf{A} = \mathbf{PJP}^{-1}$ and

$$\mathbf{r} := (r_1, \ldots, r_M)^{\mathrm{T}} = \frac{\mathbf{Pr_m}\left|\mathbf{P}^{-1}\mathbf{x}\right|^2}{|\mathbf{x}|^2},$$

$$\mathbf{h} := (h_1, \ldots, h_M)^{\mathrm{T}} = \mathbf{Ph_m}. \tag{D.2}$$

There exist constants $C_{\mathbf{r}}$ and $C_{\mathbf{h}}$ such that $|\mathbf{r}| \leq C_{\mathbf{r}}$ and $|\mathbf{h}| \leq C_{\mathbf{h}}$, given by, e.g.,

$$|\mathbf{r}| \leq \frac{\|\mathbf{P}\| |\mathbf{r_m}| \left\|\mathbf{P}^{-1}\right\|^2 |\mathbf{x}|^2}{|\mathbf{x}|^2}$$

$$= \|\mathbf{P}\| \left\|\mathbf{P}^{-1}\right\|^2 |\mathbf{r_m}|$$

$$\leq \|\mathbf{P}\| \left\|\mathbf{P}^{-1}\right\|^2 C_{\mathbf{rm}} =: C_{\mathbf{r}}, \tag{D.3}$$

and

$$|\mathbf{h}| \leq \|\mathbf{P}\| |\mathbf{h_m}| \leq \|\mathbf{P}\| C_{\mathbf{hm}} =: C_{\mathbf{h}}. \tag{D.4}$$

By substituting Eqs. (C.20) into Eqs. (D.3) and (D.4), we obtain

$$C_{\mathbf{r}} = \|\mathbf{P}\| \left\|\mathbf{P}^{-1}\right\|^2 \left[\|\mathbf{B}\| + \eta\sqrt{M}C_{\mathbf{Fmm}}\right],$$

$$C_{\mathbf{h}} = \|\mathbf{P}\| \eta\sqrt{M}\left[\widehat{C'_{\mathbf{Fz}}} + 2C'_{\mathbf{F}\varepsilon}\right]. \tag{D.5}$$

Appendix E: Proof of Lemma 4

We first prove three simple cases: all eigenvalues are (1) distinct, (2) the same real number, or (3) the same complex number, and then combine them for proving the general case.

E.1 All eigenvalues are distinct

We assume that all eigenvalues of the $M \times M$ matrix $\mathbf{J}$ are distinct, comprising G real eigenvalues $\lambda_1, \ldots, \lambda_G$ with corresponding real eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_G$ and $2H$ complex eigenvalues $\beta_1 \pm i\omega_1, \ldots, \beta_H \pm i\omega_H$ with corresponding complex eigen-

vectors $\mathbf{u}_1 \pm i\mathbf{w}_1, \ldots, \mathbf{u}_H \pm i\mathbf{w}_H$, where $G + 2H = M$. Then $\mathbf{J}$ can be decomposed as $\mathbf{J} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$, where

$$\mathbf{P} := (\mathbf{v}_1, \ldots, \mathbf{v}_G, \mathbf{u}_1, \mathbf{w}_1, \mathbf{u}_2, \mathbf{w}_2, \ldots, \mathbf{u}_H, \mathbf{w}_H),$$

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & & \cdots\cdots\cdots & & 0 \\ 0 & \ddots & 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & 0 & \lambda_G & 0 & & 0 & \ddots & \vdots \\ \vdots & & \ddots & 0 & \beta_1 & \omega_1 & 0 & \vdots \\ \vdots & & \ddots & 0 & -\omega_1 & \beta_1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & 0 & 0 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \ddots & 0 & \beta_H & \omega_H \\ 0 & & \cdots\cdots\cdots & & & \cdots & 0 & -\omega_H & \beta_H \end{pmatrix}. \tag{E.1}$$

Thus, for $\lambda_{\max} = \max\{\mathrm{Re}(\lambda_1), \ldots, \mathrm{Re}(\beta_H \pm i\omega_H)\} = \max \{\lambda_1, \ldots, \lambda_G, \beta_1, \ldots, \beta_H\}$ defined by Eq. (4.5b),

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\frac{1}{2}\left(\mathbf{A} + \mathbf{A}^{\mathrm{T}}\right)\mathbf{x} \le \lambda_{\max}|\mathbf{x}|^2 \tag{E.2}$$

$\square$

E.2 All eigenvalues are the same real number

We assume that $\mathbf{m} = (m_1, m_2, m_3)^{\mathrm{T}}$ for which all three eigenvalues of $\mathbf{J}$ are the same real number, i.e., $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$. Then there exists a regular matrix $\mathbf{P}_{\mathrm{J}}$ that transforms $\mathbf{J}$ into the Jordan normal form $\mathbf{\Lambda} := \mathbf{P}_{\mathrm{J}}\mathbf{J}\mathbf{P}_{\mathrm{J}}^{-1}$ with

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}. \tag{E.3}$$

Here we modify $\mathbf{P}_{\mathrm{J}}$ to $\mathbf{P}$ by multiplying it with $\mathbf{\Theta}$,

$$\mathbf{P} := \mathbf{P}_{\mathrm{J}}\mathbf{\Theta} \tag{E.4}$$

with

$$\mathbf{\Theta} := \begin{pmatrix} a^2 & a & 1 \\ 0 & a & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{\Theta}^{-1} = \begin{pmatrix} a^{-2} & -a^{-2} & 0 \\ 0 & a^{-1} & -a^{-1} \\ 0 & 0 & 1 \end{pmatrix}. \tag{E.5}$$

Then

$$\mathbf{A} = \mathbf{PJP}^{-1}$$

$$= \mathbf{P_J J P_J^{-1} \Theta^{-1}} = \begin{pmatrix} a^2 & a & 1 \\ 0 & a & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} a^{-2} & -a^{-2} & 0 \\ 0 & a^{-1} & -a^{-1} \\ 0 & 0 & 1 \end{pmatrix}.$$

$$= \begin{pmatrix} a^2\lambda & a^2 + a\lambda & a + \lambda \\ 0 & a\lambda & a + \lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} a^{-2} & -a^{-2} & 0 \\ 0 & a^{-1} & -a^{-1} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda & a & 0 \\ 0 & \lambda & a \\ 0 & 0 & \lambda \end{pmatrix}.$$

$$\text{(E.6)}$$

Thus, for $\lambda < 0$, $a = \lambda/2$ and $\mathbf{x} = (x_1, x_2, x_3)^{\mathrm{T}} = \mathbf{P}(\mathbf{m} - \widehat{m})$, we see

$$\mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} = \mathbf{x}^{\mathrm{T}} \begin{pmatrix} \lambda & \lambda/2 & 0 \\ 0 & \lambda & \lambda/2 \\ 0 & 0 & \lambda \end{pmatrix} \mathbf{x} = \frac{\lambda}{4} \left[4x_1^2 + 4x_2^2 + 4x_3^2 + 2x_1 x_2 + 2x_2 x_3 \right]$$

$$= \frac{\lambda}{4} \left[3x_1^2 + 2x_2^2 + 3x_3^2 + (x_1 + x_2)^2 + (x_2 + x_3)^2 \right] \qquad \text{(E.7)}$$

$$\leq \frac{\lambda}{4} \left[3x_1^2 + 2x_2^2 + 3x_3^2 \right] \leq \frac{\lambda}{2} |\mathbf{x}|^2.$$

Similarly, for a $W \times W$ matrix $\mathbf{J}$ with W repeated real eigenvalues λ, its Jordan normal form $\mathbf{\Lambda} = \mathbf{P_J J P_J^{-1}}$ is expressed as

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}, \qquad \text{(E.8)}$$

where, here and below, all blank components are zero. We define

$$\mathbf{\Theta} := \begin{pmatrix} a^{W-1} & a^{W-2} & a^{W-3} & \cdots & a & 1 \\ & a^{W-2} & a^{W-3} & \cdots & a & 1 \\ & & \ddots & \ddots & \vdots & \vdots \\ & & & \ddots & \ddots & \vdots \\ & & & & a & 1 \\ & & & & & 1 \end{pmatrix} \qquad \text{(E.9)}$$

with its inverse

$$\Theta^{-1} = \begin{pmatrix} a^{-W+1} & -a^{-W+1} & & & & \\ & a^{-W+2} & -a^{-W+2} & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & a^{-1} & -a^{-1} \\ & & & & & 1 \end{pmatrix}, \tag{E.10}$$

which gives

$$\mathbf{A} = \mathbf{P} \mathbf{J} \mathbf{P}^{-1} = \mathbf{P}_{\mathrm{J}} \mathbf{J} \mathbf{P}_{\mathrm{J}}^{-1} \Theta^{-1} = \begin{pmatrix} \lambda & a & & & \\ & \lambda & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda & a \\ & & & & \lambda \end{pmatrix}. \tag{E.11}$$

Thus, for $\lambda < 0$, $a = \lambda/2$, and $\mathbf{x} = (x_1, \ldots, x_W)^{\mathrm{T}} = \mathbf{P}(\mathbf{m} - \widehat{m})$, we see

$$\begin{aligned} \mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} &= \frac{\lambda}{4} \Big[4x_1^2 + \cdots + 4x_W^2 + 2x_1 x_2 + 2x_2 x_3 + \cdots + 2x_{W-1} x_W \Big] \\ &= \frac{\lambda}{4} \Big[3x_1^2 + 2x_2^2 + \cdots + 2x_{W-1}^2 + 3x_W^2 + (x_1 + x_2)^2 + \cdots + (x_{W-1} + x_W)^2 \Big]. \\ &\leq \frac{\lambda}{4} \Big[3x_1^2 + 2x_2^2 + \cdots + 2x_{W-1}^2 + 3x_W^2 \Big] \leq \frac{\lambda}{2} |\mathbf{x}|^2. \end{aligned} \tag{E.12}$$

$\square$

E.3 All eigenvalues are the same complex number

We assume that $\mathbf{m} = (m_1, m_2, m_3, m_4, m_5, m_6)^{\mathrm{T}}$ for which the six-by-six matrix $\mathbf{J}$ has three sets of repeated complex eigenvalues, i.e., $\beta_1 \pm i_1 \omega$, $\beta_2 \pm i_2 \omega$ and $\beta_3 \pm i_3 \omega$ with $\beta_1 = \beta_2 = \beta_3 = \beta$ and $\omega_1 = \omega_2 = \omega_3 = \omega$. Then there exists a regular matrix $\mathbf{P}_{\mathrm{J}}$ that standardizes $\mathbf{J}$ to $\mathbf{\Lambda} = \mathbf{P}_{\mathrm{J}} \mathbf{J} \mathbf{P}_{\mathrm{J}}^{-1}$ in the following normal form

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{C} & \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} \beta & \omega \\ -\omega & \beta \end{pmatrix}, \quad \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{E.13}$$

(Morris et al. 2003). Here we define a regular matrix $\mathbf{P}$ as

$$\mathbf{P} := \mathbf{P_J}\Theta \tag{E.14}$$

with

$$\Theta := \begin{pmatrix} a^2\mathbf{E} & a\mathbf{E} & \mathbf{E} \\ \mathbf{0} & a\mathbf{E} & \mathbf{E} \\ \mathbf{0} & \mathbf{0} & \mathbf{E} \end{pmatrix}, \quad \Theta^{-1} = \begin{pmatrix} a^{-2}\mathbf{E}^{-1} & -a^{-1}\mathbf{E}^{-1} & \mathbf{0} \\ \mathbf{0} & a^{-1}\mathbf{E}^{-1} & -a^{-1}\mathbf{E}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}^{-1} \end{pmatrix},$$

$$\mathbf{E} := \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{E}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \tag{E.15}$$

Then we see

$$\mathbf{A} = \mathbf{PJP}^{-1}$$

$$= \Theta\mathbf{P_J}\mathbf{J}\mathbf{P_J}^{-1}\Theta^{-1} = \begin{pmatrix} a^2\mathbf{E} & a\mathbf{E} & \mathbf{E} \\ \mathbf{0} & a\mathbf{E} & \mathbf{E} \\ \mathbf{0} & \mathbf{0} & \mathbf{E} \end{pmatrix}\begin{pmatrix} \mathbf{C} & \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix}\begin{pmatrix} a^{-2}\mathbf{E}^{-1} & -a^{-2}\mathbf{E}^{-1} & \mathbf{0} \\ \mathbf{0} & a^{-1}\mathbf{E}^{-1} & -a^{-1}\mathbf{E}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} a^2\mathbf{EC} & a^2\mathbf{E}+a\mathbf{EC} & a\mathbf{E}+\mathbf{EC} \\ \mathbf{0} & a\mathbf{EC} & a\mathbf{E}+\mathbf{EC} \\ \mathbf{0} & \mathbf{0} & \mathbf{EC} \end{pmatrix}\begin{pmatrix} a^{-2}\mathbf{E}^{-1} & -a^{-2}\mathbf{E}^{-1} & \mathbf{0} \\ \mathbf{0} & a^{-1}\mathbf{E}^{-1} & -a^{-1}\mathbf{E}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{ECE}^{-1} & a\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{ECE}^{-1} & a\mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{ECE}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{C} & a\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & a\mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix}. \tag{E.16}$$

Thus, for $\lambda < 0$, $a = \lambda/2$, and $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)^{\mathrm{T}} = \mathbf{P}(\mathbf{m} - \widehat{m})$, we see

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\begin{pmatrix} \mathbf{C} & a\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & a\mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\begin{pmatrix} \beta & \omega & \beta/2 & 0 & 0 & 0 \\ -\omega & \beta & 0 & \beta/2 & 0 & 0 \\ 0 & 0 & \beta & \omega & \beta/2 & 0 \\ 0 & 0 & -\omega & \beta & 0 & \beta/2 \\ 0 & 0 & 0 & 0 & \beta & \omega \\ 0 & 0 & 0 & 0 & -\omega & \beta \end{pmatrix}\mathbf{x}$$

$$= \frac{\beta}{4}\left[4x_1^2 + 4x_2^2 + 4x_3^2 + 4x_4^2 + 4x_5^2 + 4x_6^2 + 2x_1x_3 + 2x_2x_4 + 2x_3x_5 + 2x_4x_6\right]$$

$$= \frac{\beta}{4}\left[3x_1^2 + 3x_2^2 + 2x_3^2 + 2x_4^2 + 3x_5^2 + 3x_6^2 + (x_1+x_3)^2 + (x_2+x_4)^2 + (x_3+x_5)^2 + (x_4+x_6)^2\right]$$

$$\leq \frac{\beta}{4}\left[3x_1^2 + 3x_2^2 + 2x_3^2 + 2x_4^2 + 3x_5^2 + 3x_6^2\right] \leq \frac{\beta}{2}|\mathbf{x}|^2. \tag{E.17}$$

Similarly, for a $2W \times 2W$ matrix $\mathbf{J}$ with W sets of repeated complex eigenvalues, i.e., $\beta_1 \pm i\omega_1, \ldots, \beta_W \pm i\omega_W$ with $\beta_1 = \cdots = \beta_W = \beta$ and $\omega_1 = \cdots = \omega_W = \omega$, $\mathbf{\Lambda} = \mathbf{P_J J P_J^{-1}}$ is expressed as

$$
\mathbf{\Lambda} = \begin{pmatrix} \mathbf{C\ I_2} & & & & \\ & \mathbf{C} & \ddots & & \\ & & \ddots & \ddots & \\ & & & \mathbf{C\ I_2} \\ & & & & \mathbf{C} \end{pmatrix}, \mathbf{C} = \begin{pmatrix} \beta & \omega \\ -\omega & \beta \end{pmatrix}, \tag{E.18}
$$

where, here and below, all blank components are zero. We define

$$
\mathbf{\Theta} := \begin{pmatrix} a^{W-1}\mathbf{E} & a^{W-2}\mathbf{E} & a^{W-3}\mathbf{E} & \cdots & a\mathbf{E} & \mathbf{E} \\ & a^{W-2}\mathbf{E} & a^{W-3}\mathbf{E} & \cdots & a\mathbf{E} & \mathbf{E} \\ & & \ddots & \ddots & \vdots & \vdots \\ & & & \ddots & \ddots & \vdots \\ & & & & a\mathbf{E} & \mathbf{E} \\ & & & & & \mathbf{E} \end{pmatrix} \tag{E.19}
$$

with

$$
\mathbf{\Theta}^{-1} = \begin{pmatrix} a^{-W+1}\mathbf{E}^{-1} & -a^{-W+1}\mathbf{E}^{-1} & & & \\ & a^{-W+2}\mathbf{E}^{-1} & -a^{-W+2}\mathbf{E}^{-1} & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots & \\ & & & & a^{-1}\mathbf{E}^{-1} & -a^{-1}\mathbf{E}^{-1} \\ & & & & & \mathbf{E}^{-1} \end{pmatrix}, \tag{E.20}
$$

which gives

$$
\mathbf{A} = \mathbf{PJP}^{-1} = \mathbf{P_J J P_J^{-1} \Theta^{-1}} = \begin{pmatrix} \mathbf{C}\ a\mathbf{I_2} & & & & \\ & \mathbf{C} & \ddots & & \\ & & \ddots & \ddots & \\ & & & \mathbf{C}\ a\mathbf{I_2} \\ & & & & \mathbf{C} \end{pmatrix}. \tag{E.21}
$$

Thus, for $\lambda < 0$, $a = \lambda/2$, and $\mathbf{x} = (x_1, \ldots, x_{2W})^{\mathrm{T}} = \mathbf{P}(\mathbf{m} - \widehat{m})$, we see

$$
\mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} = \frac{\beta}{4} \left[4x_1^2 + \cdots + 4x_{2W}^2 + 2x_1 x_3 + \cdots + 2x_{2W-2} x_{2W} \right]
$$

$$
= \frac{\beta}{4} \left[3x_1^2 + 3x_2^2 + 2x_3^2 + \cdots + 2x_{2W-2}^2 + 3x_{2W-1}^2 \right]
$$

$$+3x_{2W}^2 + (x_1 + x_3)^2 + \cdots + (x_{2W-2} + x_{2W})^2\Big]$$

$$\leq \frac{\beta}{4}\Big[3x_1^2 + 3x_2^2 + 2x_3^2 + \cdots + 2x_{2W-2}^2 + 3x_{2W-1}^2 + 3x_{2W}^2\Big] \leq \frac{\beta}{2}|\mathbf{x}|^2. \quad \text{(E.22)}$$

$\square$

E.4 General case

By combining the aforementioned three cases, here we prove Lemma 4. For an arbitrary $M \times M$ matrix $\mathbf{J}$, there exists a regular matrix $\mathbf{P_J}$ that transforms $\mathbf{J}$ into $\mathbf{\Lambda} = \mathbf{P_J}\mathbf{J}\mathbf{P_J}^{-1}$ in the following form

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}^{\mathrm{R}} & & & & & & & \\ & \mathbf{\Lambda}^{\mathrm{C}} & & & & & & \\ & & \mathbf{\Lambda}_1^{\mathrm{repR}} & & & & & \\ & & & \ddots & & & & \\ & & & & \mathbf{\Lambda}_K^{\mathrm{repR}} & & & \\ & & & & & \mathbf{\Lambda}_1^{\mathrm{repC}} & & \\ & & & & & & \ddots & \\ & & & & & & & \mathbf{\Lambda}_L^{\mathrm{repC}} \end{pmatrix}, \quad \text{(E.23)}$$

where, here and below, all blank components are zero, $\mathbf{\Lambda}^{\mathrm{R}}$ is a $G \times G$ diagonal matrix with its entries given by the distinct real eigenvalues $\lambda_1^{\mathrm{R}}, \ldots, \lambda_G^{\mathrm{R}}$, and $\mathbf{\Lambda}^{\mathrm{C}}$ is a $2H \times 2H$ block diagonal matrix with its entries determined by the distinct complex eigenvalues $\beta_1^{\mathrm{C}} \pm i\omega_1^{\mathrm{C}}, \ldots, \beta_H^{\mathrm{C}} \pm i\omega_H^{\mathrm{C}}$, i.e.,

$$\mathbf{\Lambda}^{\mathrm{C}} = \begin{pmatrix} \mathbf{C}_1^{\mathrm{C}} & & \\ & \ddots & \\ & & \mathbf{C}_H^{\mathrm{C}} \end{pmatrix}, \quad \mathbf{C}_h^{\mathrm{C}} = \begin{pmatrix} \beta_h^{\mathrm{C}} & \omega_h^{\mathrm{C}} \\ -\omega_h^{\mathrm{C}} & \beta_h^{\mathrm{C}} \end{pmatrix}, \quad \text{(E.24)}$$

for $h = 1, \ldots, H$ (Morris et al. 2003). The subsequent matrices in Eq. (E.23), $\mathbf{\Lambda}_k^{\mathrm{repR}}$ for $k = 1, \ldots, K$ are for repeated real eigenvalues, with their repetition numbers denoted by $\mathrm{repr}(k)$. $\mathbf{\Lambda}_k^{\mathrm{repR}}$ is a $\mathrm{repr}(k) \times \mathrm{repr}(k)$ matrix expressed in the form of Eq. (E.8) with its diagonal entries $\lambda = \lambda_k^{\mathrm{repR}}$. Finally, $\mathbf{\Lambda}_l^{\mathrm{repC}}$ for $l = 1, \ldots, L$ are for repeated complex eigenvalues, with their repetition numbers denoted by $\mathrm{repc}(l)$. $\mathbf{\Lambda}_l^{\mathrm{repC}}$ is a $2\mathrm{repc}(l) \times 2\mathrm{repc}(l)$ matrix expressed in the form of Eq. (E.18) with

$$\mathbf{C} = \mathbf{C}_l^{\mathrm{repC}} = \begin{pmatrix} \beta_l^{\mathrm{repC}} & \omega_l^{\mathrm{repC}} \\ -\omega_l^{\mathrm{repC}} & \beta_l^{\mathrm{repC}} \end{pmatrix}. \quad \text{(E.25)}$$

Since $\mathbf{J}$ is an $M \times M$ matrix, $G + 2H + \mathrm{repr}(1) + \cdots + \mathrm{repr}(K) + 2[\mathrm{repc}(1) + \cdots + \mathrm{rep}(L)] = M$ is fulfilled. Here we define

$$\mathbf{P} := \mathbf{P_J}\boldsymbol{\Theta} \qquad (E.26)$$

with

$$\boldsymbol{\Theta} := \begin{pmatrix} \mathbf{I}^R & & & & & & & & \\ & \mathbf{I}^C & & & & & & & \\ & & \boldsymbol{\Theta}_1^{\mathrm{repR}} & & & & & & \\ & & & \ddots & & & & & \\ & & & & \boldsymbol{\Theta}_K^{\mathrm{repR}} & & & & \\ & & & & & \boldsymbol{\Theta}_1^{\mathrm{repC}} & & & \\ & & & & & & \ddots & & \\ & & & & & & & \boldsymbol{\Theta}_L^{\mathrm{repC}} \end{pmatrix},$$

$$(E.27)$$

where $\mathbf{I}^R$ and $\mathbf{I}^C$ are, respectively, $G \times G$ and $2H \times 2H$ identity matrices, $\boldsymbol{\Theta}_k^{\mathrm{repR}}$ is given by Eq. (E.9) with $W = \mathrm{repr(k)}$, and $\boldsymbol{\Theta}_l^{\mathrm{repC}}$ is given by Eq. (E.18) with $W = \mathrm{repc}(l)$. Then we see

$$\mathbf{A} = \boldsymbol{\Theta}\mathbf{P_J}\mathbf{J}\mathbf{P_J}^{-1}\boldsymbol{\Theta}^{-1} = \boldsymbol{\Theta}\boldsymbol{\Lambda}\boldsymbol{\Theta}^{-1}$$

$$= \begin{pmatrix} \boldsymbol{\Lambda}^R & & & & & & & & \\ & \boldsymbol{\Lambda}^C & & & & & & & \\ & & \breve{\boldsymbol{\Lambda}}_1^{\mathrm{repR}} & & & & & & \\ & & & \ddots & & & & & \\ & & & & \breve{\boldsymbol{\Lambda}}_K^{\mathrm{repR}} & & & & \\ & & & & & \breve{\boldsymbol{\Lambda}}_1^{\mathrm{repC}} & & & \\ & & & & & & \ddots & & \\ & & & & & & & \breve{\boldsymbol{\Lambda}}_L^{\mathrm{repC}} \end{pmatrix}, \qquad (E.28)$$

where $\breve{\boldsymbol{\Lambda}}_k^{\mathrm{repR}}$ is the same as $\boldsymbol{\Lambda}_k^{\mathrm{repR}}$ except that all 1 s in $\boldsymbol{\Lambda}_k^{\mathrm{repR}}$ are replaced with $\lambda_k^{\mathrm{repR}}/2$ for $k = 1, \ldots, K$, and in the same manner $\breve{\boldsymbol{\Lambda}}_l^{\mathrm{repC}}$ is the same as $\boldsymbol{\Lambda}_l^{\mathrm{repC}}$ except that all $\mathbf{I}_2$ in $\boldsymbol{\Lambda}_l^{\mathrm{repC}}$ are replaced with $\beta_l^{\mathrm{repC}}\mathbf{I}_2/2$ for $l = 1, \ldots, L$. Therefore, we get

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \begin{pmatrix} \mathbf{x}^{\mathrm{R}} \\ \mathbf{x}^{\mathrm{C}} \\ \mathbf{x}_1^{\mathrm{repR}} \\ \vdots \\ \mathbf{x}_K^{\mathrm{repR}} \\ \mathbf{x}_1^{\mathrm{repC}} \\ \vdots \\ \mathbf{x}_L^{\mathrm{repC}} \end{pmatrix}^{\mathrm{T}} \mathbf{A} \begin{pmatrix} \mathbf{x}^{\mathrm{R}} \\ \mathbf{x}^{\mathrm{C}} \\ \mathbf{x}_1^{\mathrm{repR}} \\ \vdots \\ \mathbf{x}_K^{\mathrm{repR}} \\ \mathbf{x}_1^{\mathrm{repC}} \\ \vdots \\ \mathbf{x}_L^{\mathrm{repC}} \end{pmatrix}.$$

(E.29)

$$= [\mathbf{x}^{\mathrm{R}}]^{\mathrm{T}}\boldsymbol{\Lambda}^{\mathrm{R}}\mathbf{x}^{\mathrm{R}} + [\mathbf{x}^{\mathrm{C}}]^{\mathrm{T}}\boldsymbol{\Lambda}^{\mathrm{C}}\mathbf{x}^{\mathrm{C}}$$
$$+ [\mathbf{x}_1^{\mathrm{repR}}]^{\mathrm{T}}\breve{\boldsymbol{\Lambda}}_1^{\mathrm{repR}}\mathbf{x}_1^{\mathrm{repR}} + \cdots + [\mathbf{x}_K^{\mathrm{repR}}]^{\mathrm{T}}\breve{\boldsymbol{\Lambda}}_K^{\mathrm{repR}}\mathbf{x}_K^{\mathrm{repR}}$$
$$+ [\mathbf{x}_1^{\mathrm{repC}}]^{\mathrm{T}}\breve{\boldsymbol{\Lambda}}_1^{\mathrm{repC}}\mathbf{x}_1^{\mathrm{repC}} + \cdots + [\mathbf{x}_L^{\mathrm{repC}}]^{\mathrm{T}}\breve{\boldsymbol{\Lambda}}_L^{\mathrm{repC}}\mathbf{x}_L^{\mathrm{repC}},$$

where

$$[\mathbf{x}^{\mathrm{R}}]^{\mathrm{T}}\mathbf{3}^{\mathrm{R}}\mathbf{x}^{\mathrm{R}} \le \max\{\lambda_1^{\mathrm{R}}, \ldots, \lambda_G^{\mathrm{R}}\}|\mathbf{x}^{\mathrm{R}}|^2,$$
$$[\mathbf{x}^{\mathrm{C}}]^{\mathrm{T}}\mathbf{3}^{\mathrm{C}}\mathbf{x}^{\mathrm{C}} \le \max\{\beta_1^{\mathrm{C}}, \ldots, \beta_H^{\mathrm{C}}\}|\mathbf{x}^{\mathrm{C}}|^2$$

(E.30)

are satisfied. In addition, from Eqs. (E.12) and (E.22) we see

$$[\mathbf{x}_k^{\mathrm{repR}}]^{\mathrm{T}}\breve{\Lambda}_k^{\mathrm{repR}}\mathbf{x}_k^{\mathrm{repR}} \le \frac{\lambda_k^{\mathrm{repR}}}{2}|\mathbf{x}_k^{\mathrm{repR}}|^2,$$
$$[\mathbf{x}_l^{\mathrm{repC}}]^{\mathrm{T}}\breve{\Lambda}_l^{\mathrm{repC}}\mathbf{x}_l^{\mathrm{repC}} \le \frac{\beta_l^{\mathrm{repC}}}{2}|\mathbf{x}_l^{\mathrm{repC}}|^2$$

(E.31)

for $k = 1, \ldots, K$ and $l = 1, \ldots, L$. Thus, substituting Eqs. (E.30) and (E.31) into Eq. (E.29) gives

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \le \max\{\lambda_1^{\mathrm{R}}, \ldots, \lambda_G^{\mathrm{R}}\}\left|\mathbf{x}^{\mathrm{R}}\right|^2 + \max\{\beta_1^{\mathrm{C}}, \ldots, \beta_H^{\mathrm{C}}\}\left|\mathbf{x}^{\mathrm{C}}\right|^2$$
$$+ \frac{\lambda_1^{\mathrm{repR}}}{2}\left|\mathbf{x}_1^{\mathrm{repR}}\right|^2 + \cdots + \frac{\lambda_K^{\mathrm{repR}}}{2}\left|\mathbf{x}_K^{\mathrm{repR}}\right|^2$$
$$+ \frac{\beta_1^{\mathrm{repC}}}{2}\left|\mathbf{x}_1^{\mathrm{repC}}\right|^2 + \cdots + \frac{\beta_L^{\mathrm{repC}}}{2}\left|\mathbf{x}_L^{\mathrm{repC}}\right|^2,$$
$$\le \breve{\lambda}_{\max}\left[\left|\mathbf{x}^{\mathrm{R}}\right|^2 + \left|\mathbf{x}^{\mathrm{C}}\right|^2 + \left|\mathbf{x}_1^{\mathrm{repR}}\right|^2 + \cdots + \left|\mathbf{x}_K^{\mathrm{repR}}\right|^2 + \left|\mathbf{x}_1^{\mathrm{repC}}\right|^2 + \cdots + \left|\mathbf{x}_L^{\mathrm{repC}}\right|^2\right]$$
$$= \breve{\lambda}_{\max}|\mathbf{x}|^2$$

(E.32)

with

$$
\breve{\lambda}_{\max} := \max\left\{ \lambda_1^{\mathrm{R}}, \ldots, \lambda_G^{\mathrm{R}}, \beta_1^{\mathrm{C}}, \ldots, \beta_H^{\mathrm{C}}, \frac{\lambda_1^{\mathrm{repR}}}{2}, \ldots, \frac{\lambda_K^{\mathrm{repR}}}{2}, \frac{\beta_1^{\mathrm{repC}}}{2}, \ldots, \frac{\beta_L^{\mathrm{repC}}}{2} \right\},
$$
(E.33)

which is identical to $\lambda_{\max}$ in Lemma 4. □

Appendix F: Derivation of Eq. (5.4a)

F.1 Adjustment of equilibrium point

As introduced in the main text just before Eq. (5.2a), $\mathbf{m}$ is expressed as $\mathbf{m} = \begin{pmatrix} \mathbf{m}_{\mathrm{x}} \\ \mathbf{m}_{\mathrm{y}} \end{pmatrix}$ with $\mathbf{m}_{\mathrm{x}} = (m_1, \ldots, m_L)^{\mathrm{T}}$ and $\mathbf{m}_{\mathrm{y}} = (m_{L+1}, \ldots, m_M)^{\mathrm{T}}$ with $|\widehat{m}_{\mathrm{y}}| = |(\widehat{m}_{L+1}, \ldots, \widehat{m}_M)| \leq \rho_m \varepsilon$ for all $i = L+1, \ldots, M$. First, we transform

$$
\mathbf{f}(\mathbf{s}_{\mathrm{a}}; \mathbf{s}_{\mathrm{a}}; \mathbf{m}) = \mathbf{B}(\mathbf{m} - \widehat{m}) + \mathbf{r}_{\mathrm{f}}\left| \mathbf{m} - \widehat{m} \right|^2 + \varepsilon \mathbf{h}_{\mathrm{f}}
$$
(F.1)

by adjusting the perturbation term to get an equilibrium not at $\mathbf{m} = \widehat{m} = \begin{pmatrix} \widehat{m}_{\mathrm{x}} \\ \widehat{m}_{\mathrm{y}} \end{pmatrix}$ but at $\mathbf{m} = \widehat{\widetilde{m}} = \begin{pmatrix} \widehat{m}_{\mathrm{x}} \\ \mathbf{0} \end{pmatrix}$ when $\varepsilon = 0$. This is done by the transformations

$$
\mathbf{B}(\mathbf{m} - \widehat{m}) = \mathbf{B}\begin{pmatrix} \mathbf{m}_{\mathrm{x}} - \widehat{m}_{\mathrm{x}} \\ \mathbf{m}_{\mathrm{y}} - \widehat{m}_{\mathrm{y}} \end{pmatrix} = \mathbf{B}\begin{pmatrix} \mathbf{m}_{\mathrm{x}} - \widehat{m}_{\mathrm{x}} \\ \mathbf{m}_{\mathrm{y}} \end{pmatrix} - \mathbf{B}\begin{pmatrix} 0 \\ \widehat{m}_{\mathrm{y}} \end{pmatrix}
$$
$$
= \mathbf{B}\left(\mathbf{m} - \widehat{\widetilde{m}}\right) - \varepsilon \left[\mathbf{B}\begin{pmatrix} 0 \\ \frac{\widehat{m}_{\mathrm{y}}}{\varepsilon} \end{pmatrix} \right]
$$
(F.2)

and

$$
\mathbf{r}_{\mathrm{f}}\left| \mathbf{m} - \widehat{m} \right|^2 = \mathbf{r}_{\mathrm{f}}\left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \mathbf{r}_{\mathrm{f}}\left[\left| \mathbf{m} - \widehat{m} \right|^2 - \left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 \right]
$$
$$
= \mathbf{r}_{\mathrm{f}}\left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \mathbf{r}_{\mathrm{f}}\left[\left| \mathbf{m} - \widehat{m} \right|^2 - \left| (\mathbf{m} - \widehat{m}) + (\widehat{m} - \widehat{\widetilde{m}}) \right|^2 \right]
$$
$$
= \mathbf{r}_{\mathrm{f}}\left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \mathbf{r}_{\mathrm{f}}\left[\left(\mathbf{m} - \widehat{m} \right)^{\mathrm{T}}\left(\mathbf{m} - \widehat{m} \right) \right.
$$
$$
\left. -[(\mathbf{m} - \widehat{m}) + (\widehat{m} - \widehat{\widetilde{m}})]^{\mathrm{T}}[(\mathbf{m} - \widehat{m}) + (\widehat{m} - \widehat{\widetilde{m}})] \right]
$$

$$= \mathbf{r}_f \left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \mathbf{r}_f \left[-2 \left(\mathbf{m} - \widehat{m} \right)^T (\widehat{m} - \widehat{\widetilde{m}}) - (\widehat{m} - \widehat{\widetilde{m}})^T (\widehat{m} - \widehat{\widetilde{m}}) \right]$$

$$= \mathbf{r}_f \left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \mathbf{r}_f \left[-2 \left(\mathbf{m} - \widehat{m} \right) - (\widehat{m} - \widehat{\widetilde{m}}) \right]^T (\widehat{m} - \widehat{\widetilde{m}})$$

$$= \mathbf{r}_f \left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \mathbf{r}_f \left[-2 \left(\mathbf{m} - \widehat{m} \right) - \begin{pmatrix} \mathbf{0} \\ \widehat{m}_y \end{pmatrix} \right]^T \begin{pmatrix} \mathbf{0} \\ \widehat{m}_y \end{pmatrix}, \tag{F.3}$$

which upon substitution into Eq. (F.1) gives

$$\mathbf{f}(\mathbf{s}_a; \mathbf{s}_a; \mathbf{m}) = \mathbf{B} \left(\mathbf{m} - \widehat{\widetilde{m}} \right) + \tilde{\mathbf{r}}_f \left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \varepsilon \tilde{\mathbf{h}}_f, \tag{F.4}$$

where

$$\tilde{\mathbf{r}}_f = \mathbf{r}_f,$$

$$\tilde{\mathbf{h}}_f = \mathbf{h}_f - \mathbf{B} \begin{pmatrix} 0 \\ \frac{\widehat{m}_y}{\varepsilon} \end{pmatrix} + \frac{1}{\varepsilon} \mathbf{r}_f \left[-2 \left(\mathbf{m} - \widehat{m} \right) - \begin{pmatrix} \mathbf{0} \\ \widehat{m}_y \end{pmatrix} \right]^T \begin{pmatrix} \mathbf{0} \\ \widehat{m}_y \end{pmatrix}, \tag{F.5}$$

and

$$|\tilde{\mathbf{r}}_f| = |\mathbf{r}_f| \le C_{\mathbf{rf}} =: \tilde{C}_{\mathbf{rf}},$$

$$\left| \tilde{\mathbf{h}}_f \right| \le |\mathbf{h}_f| + \|\mathbf{B}\| \left| \frac{\widehat{m}_y}{\varepsilon} \right| + \frac{1}{\varepsilon} |\mathbf{r}_f| \left[2 \left| \mathbf{m}_y - \widehat{m}_y \right| + \left| \widehat{m}_y \right| \right] \left| \widehat{m}_y \right|$$

$$\le |\mathbf{h}_f| + \|\mathbf{B}\| \rho_m + \frac{1}{\varepsilon} |\mathbf{r}_f| [2\eta + \varepsilon\rho_m] \varepsilon\rho_m$$

$$\le C_{\mathbf{hf}} + \|\mathbf{B}\| \rho_m + C_{\mathbf{rf}} [2\eta + \varepsilon\rho_m] \rho_m =: \tilde{C}_{\mathbf{hf}}. \tag{F.6}$$

Substituting Eq. (E.4) into Eq. (4.2) in the main text gives

$$\frac{d\mathbf{m}}{dt} = \text{diag}(\mathbf{m}) \left[\mathbf{B}(\mathbf{m} - \widehat{\widetilde{m}}) + \tilde{\mathbf{r}}_f \left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \varepsilon \tilde{\mathbf{h}}_f \right]. \tag{F.7}$$

F.2 Decomposition into not-small and small population densities

We decompose Eq. (F.7) into equations for the L phenotypes with larger population densities $\mathbf{m}_x = (m_1, \ldots, m_L)^T$ and the other $K = M - L$ phenotypes with small population densities $\mathbf{m}_y = (m_{L+1}, \ldots, m_M)^T$,

$$\frac{d\mathbf{m}_x}{dt} = \text{diag}(\mathbf{m}_x) \left[\mathbf{B}_{xx}(\mathbf{m}_x - \widehat{m}_x) + \mathbf{B}_{xy}\mathbf{m}_y + \tilde{\mathbf{r}}_{fx} \left| \mathbf{m} - \widehat{\widetilde{m}} \right|^2 + \varepsilon \tilde{\mathbf{h}}_{fx} \right], \tag{F.8}$$

$$\frac{d\mathbf{m_y}}{dt} = \mathrm{diag}(\mathbf{m_y})\left[\mathbf{B_{yx}}(\mathbf{m_x} - \widehat{m}_x) + \mathbf{B_{yy}}\mathbf{m_y} + \tilde{\mathbf{r}}_{fy}\left|\mathbf{m} - \widehat{m}\right|^2 + \varepsilon\tilde{\mathbf{h}}_{fy}\right], \qquad (F.9)$$

where $\mathbf{m} = \begin{pmatrix} \mathbf{m_x} \\ \mathbf{m_y} \end{pmatrix}$, $\tilde{\mathbf{r}}_x = (\tilde{r}_1, \ldots, \tilde{r}_L)^T$, $\tilde{\mathbf{r}}_y = (\tilde{r}_{L+1}, \ldots, \tilde{r}_M)^T$, $\tilde{\mathbf{h}}_x = (\tilde{h}_1, \ldots, \tilde{h}_L)^T$, $\tilde{\mathbf{h}}_y = (\tilde{h}_{L+1}, \ldots, \tilde{h}_M)^T$, and

$$\mathbf{B_{xx}} = \begin{pmatrix} b_{1,1} & \ldots & b_{1,L} \\ \ldots & \ldots & \ldots \\ b_{L,1} & \ldots & b_{L,L} \end{pmatrix}, \quad \mathbf{B_{xy}} = \begin{pmatrix} b_{1,L+1} & \ldots & b_{1,M} \\ \ldots & \ldots & \ldots \\ b_{L,L+1} & \ldots & b_{L,M} \end{pmatrix}. \qquad (F.10)$$

$$\mathbf{B_{yx}} = \begin{pmatrix} b_{L+1,1} & \ldots & b_{L+1,L} \\ \ldots & \ldots & \ldots \\ b_{M,1} & \ldots & b_{M,L} \end{pmatrix}, \quad \mathbf{B_{yy}} = \begin{pmatrix} b_{L+1,L+1} & \ldots & b_{L+1,M} \\ \ldots & \ldots & \ldots \\ b_{M,L+1} & \ldots & b_{M,M} \end{pmatrix}. \qquad (F.11)$$

F.3 Variable transformation

If $|\mathbf{m} - \tilde{\mathbf{m}}|$ is sufficiently small, so that the nonlinear term in Eq. (F.8) can be neglected, and $\varepsilon = 0$, then $\frac{d\mathbf{m_x}}{dt} = 0$ equals

$$\mathbf{B_{xx}}(\mathbf{m_x} - \widehat{m}_x) + \mathbf{B_{xy}}\mathbf{m_y} = 0, \qquad (F.12)$$

which gives an approximate slow manifold defined by Eq. (5.3b) in the main text,

$$\tilde{\mathbf{m}}_x(\mathbf{m_y}) = \widehat{m}_x - \mathbf{T}\mathbf{m_y} \qquad (F.13)$$

with $\mathbf{T} := \mathbf{B_{xx}^{-1}}\mathbf{B_{xy}}$. Even when $\varepsilon > 0$, the dynamics is expected to be effectively characterized by its projection on this manifold. Thus, we introduce $\mathbf{x} = \mathbf{P}(\mathbf{m_x} - \tilde{\mathbf{m}}_x(\mathbf{m_y}))$ and $\mathbf{y} = \mathbf{m_y}$, to capture the fast convergence to the manifold and slow dynamics along it, respectively. The vector $\mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$ is written as

$$\mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{P}(\mathbf{m_x} - \widehat{m}_x + \mathbf{T}\mathbf{m_y}) \\ \mathbf{m_y} \end{pmatrix} = \mathbf{Q}\begin{pmatrix} \mathbf{m_x} - \widehat{m}_x \\ \mathbf{m_y} \end{pmatrix} = \mathbf{Q}(\mathbf{m} - \tilde{\mathbf{m}}), \qquad (F.14)$$

with

$$\mathbf{Q} := \begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & \mathbf{I_y} \end{pmatrix},$$

$$\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{P}^{-1} & -\mathbf{T} \\ 0 & \mathbf{I_y} \end{pmatrix}, \qquad (F.15)$$

where $\mathbf{I}_y$ is the $K \times K$ identity matrix. Then, by Eq. (F.7),

$$
\begin{aligned}
\frac{d\mathbf{w}}{dt} &= \mathbf{Q}\frac{d\mathbf{m}}{dt} \\
&= \mathbf{Q}\mathrm{diag}(\mathbf{m})\left[\mathbf{B}(\mathbf{m} - \widehat{\tilde{m}}) + \tilde{\mathbf{r}}_f \left|\mathbf{m} - \widehat{\tilde{m}}\right|^2 + \varepsilon\tilde{\mathbf{h}}_f \right] \\
&= \mathbf{Q}\mathrm{diag}(\mathbf{m})\mathbf{B}(\mathbf{m} - \widehat{\tilde{m}}) + \mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f \left|\mathbf{m} - \widehat{\tilde{m}}\right|^2 + \varepsilon\mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f. \quad \text{(F.16)}
\end{aligned}
$$

$\mathbf{Q}\mathrm{diag}(\mathbf{m})$ is further transformed as

$$
\begin{aligned}
\mathbf{Q}\mathrm{diag}(\mathbf{m}) &= \begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & \mathbf{I}_y \end{pmatrix} \begin{pmatrix} \mathrm{diag}(\widehat{m}_x) + \mathrm{diag}(\mathbf{m}_x - \widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{m}_y) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{P} & 0 \\ 0 & \mathbf{I}_y \end{pmatrix} \begin{pmatrix} \mathrm{diag}(\widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{m}_y) \end{pmatrix} + \begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathrm{diag}(\mathbf{m}_x - \widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{m}_y) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{P} & 0 \\ 0 & \mathbf{I}_y \end{pmatrix} \begin{pmatrix} \mathrm{diag}(\widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{m}_y) \end{pmatrix} + \begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & 0 \end{pmatrix} \mathrm{diag}(\mathbf{m} - \tilde{\mathbf{m}}). \quad \text{(F.17)}
\end{aligned}
$$

Substituting Eq. (F.17) into Eq. (F.16) gives

$$
\begin{aligned}
\frac{d\mathbf{w}}{dt} &= \left[\begin{pmatrix} \mathbf{P} & 0 \\ 0 & \mathbf{I}_y \end{pmatrix} \begin{pmatrix} \mathrm{diag}(\widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{m}_y) \end{pmatrix} + \begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & 0 \end{pmatrix} \mathrm{diag}(\mathbf{m} - \widehat{\tilde{m}}) \right] \mathbf{B}(\mathbf{m} - \widehat{\tilde{m}}) \\
&\quad + \mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f \left|\mathbf{m} - \widehat{\tilde{m}}\right|^2 + \varepsilon\mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f \\
&= \begin{pmatrix} \mathbf{P}\mathrm{diag}(\widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{m}_y) \end{pmatrix} \mathbf{B}\mathbf{Q}^{-1}\mathbf{w} \\
&\quad + \left[\begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & 0 \end{pmatrix} \mathrm{diag}(\mathbf{m} - \widehat{\tilde{m}})\mathbf{B}(\mathbf{m} - \widehat{\tilde{m}}) + \mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f \left|\mathbf{m} - \widehat{\tilde{m}}\right|^2 \right] + \varepsilon\mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f \\
&= \begin{pmatrix} \mathbf{P}\mathrm{diag}(\widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{m}_y) \end{pmatrix} \mathbf{B}\mathbf{Q}^{-1}\mathbf{w} + \mathbf{r}^*|\mathbf{w}|^2 + \varepsilon\mathbf{h}^*, \quad \text{(F.18)}
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{r}^* &:= \frac{1}{|\mathbf{w}|^2}\left[\begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & 0 \end{pmatrix} \mathrm{diag}(\mathbf{m} - \widehat{\tilde{m}})\mathbf{B}(\mathbf{m} - \widehat{\tilde{m}}) + \mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f \left|\mathbf{m} - \widehat{\tilde{m}}\right|^2 \right], \\
\mathbf{h}^* &:= \mathbf{Q}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f, \quad \text{(F.19)}
\end{aligned}
$$

where $\mathbf{r}^*$ is worked out below after the derivation of the equation for $d\mathbf{w}/dt$, and estimated in Appendix F.4. $\mathbf{B}\mathbf{Q}^{-1}$ is transformed by using $\mathbf{T} = \mathbf{B}_{xx}^{-1}\mathbf{B}_{xy}$, as

$$
\mathbf{B}\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{B}_{xx} & \mathbf{B}_{xy} \\ \mathbf{B}_{yx} & \mathbf{B}_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{P}^{-1} & -\mathbf{T} \\ 0 & \mathbf{I}_y \end{pmatrix}
$$

$$= \begin{pmatrix} \mathbf{B}_{xx}\mathbf{P}^{-1} & -\mathbf{B}_{xx}\mathbf{T} + \mathbf{B}_{xy} \\ \mathbf{B}_{yx}\mathbf{P}^{-1} & -\mathbf{B}_{yx}\mathbf{T} + \mathbf{B}_{yy} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{B}_{xx}\mathbf{P}^{-1} & 0 \\ \mathbf{B}_{yx}\mathbf{P}^{-1} & -\mathbf{B}_{yx}\mathbf{B}_{xx}^{-1}\mathbf{B}_{yy} + \mathbf{B}_{yy} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_{xx}\mathbf{P}^{-1} & 0 \\ \mathbf{U} & \mathbf{J}_y \end{pmatrix}, \tag{F.20}$$

with $\mathbf{U} = \mathbf{B}_{yx}\mathbf{P}^{-1}$ and $\mathbf{J}_y = -\mathbf{B}_{yx}\mathbf{B}_{xx}^{-1}\mathbf{B}_{yy} + \mathbf{B}_{yy}$. Eq. (F.20) transforms the first term of the last line of Eq. (F.18) into

$$\begin{pmatrix} \mathbf{P}\mathrm{diag}(\widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{y}) \end{pmatrix}\mathbf{B}\mathbf{Q}^{-1}\mathbf{w}$$

$$= \begin{pmatrix} \mathbf{P}\mathrm{diag}(\widehat{m}_x) & 0 \\ 0 & \mathrm{diag}(\mathbf{y}) \end{pmatrix}\begin{pmatrix} \mathbf{B}_{xx}\mathbf{P}^{-1} & 0 \\ \mathbf{U} & \mathbf{J} \end{pmatrix}\mathbf{w}$$

$$= \begin{pmatrix} \mathbf{P}\mathrm{diag}(\widehat{m}_x)\mathbf{B}_{xx}\mathbf{P}^{-1} & 0 \\ \mathrm{diag}(\mathbf{y})\mathbf{U} & \mathrm{diag}(\mathbf{y})\mathbf{J} \end{pmatrix}\mathbf{w}$$

$$= \begin{pmatrix} \mathbf{A}_x & 0 \\ \mathrm{diag}(\mathbf{y})\mathbf{U} & \mathrm{diag}(\mathbf{y})\mathbf{J} \end{pmatrix}\mathbf{w}$$

$$= \begin{pmatrix} \mathbf{A}_x\mathbf{x} \\ \mathrm{diag}(\mathbf{y})[\mathbf{U}\mathbf{x} + \mathbf{J}_y\mathbf{y}] \end{pmatrix}, \tag{F.21}$$

with $\mathbf{A}_x = \mathbf{P}\mathrm{diag}(\widehat{m}_x)\mathbf{B}_{xx}\mathbf{P}^{-1}$. Therefore, Eq. (F.18) is transformed into

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_x\mathbf{x} \\ \mathrm{diag}(\mathbf{y})[\mathbf{U}\mathbf{x} + \mathbf{J}_y\mathbf{y}] \end{pmatrix} + \mathbf{r}^*|\mathbf{w}|^2 + \varepsilon\mathbf{h}^*. \tag{F.22}$$

In addition, $\mathbf{r}^*$ and $\mathbf{h}^*$ given in Eq. (F.19) are further transformed into

$$\mathbf{r}^* = \frac{1}{|\mathbf{w}|^2}\left[\begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & 0 \end{pmatrix}\mathrm{diag}(\mathbf{m} - \widehat{m})\mathbf{B}(\mathbf{m} - \widehat{m}) + \begin{pmatrix} \mathbf{P} & \mathbf{PT} \\ 0 & \mathbf{I}_y \end{pmatrix}\mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f\left|\mathbf{m} - \widehat{m}\right|^2\right]$$

$$= \frac{1}{|\mathbf{w}|^2}\begin{pmatrix} (\mathbf{P} \ \mathbf{PT})\mathrm{diag}(\mathbf{m} - \widehat{m})\mathbf{B}(\mathbf{m} - \widehat{m}) + (\mathbf{P} \ \mathbf{PT})\mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f\left|\mathbf{m} - \widehat{m}\right|^2 \\ (0 \ \mathbf{I}_y)\mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f\left|\mathbf{m} - \widehat{m}\right|^2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{|\mathbf{w}|^2}(\mathbf{P} \ \mathbf{PT})\left[\mathrm{diag}(\mathbf{m} - \widehat{m})\mathbf{B}(\mathbf{m} - \widehat{m}) + \mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f\left|\mathbf{m} - \widehat{m}\right|^2\right] \\ \mathrm{diag}(\mathbf{y})\frac{1}{|\mathbf{w}|^2}\tilde{\mathbf{r}}_{fy}\left|\mathbf{m} - \widehat{m}\right|^2 \end{pmatrix}$$

$$=: \begin{pmatrix} \tilde{\mathbf{r}}_x \\ \mathrm{diag}(\mathbf{y})\tilde{\mathbf{r}}_y \end{pmatrix}, \tag{F.23}$$

$$\mathbf{h}^* = \begin{pmatrix} \mathbf{P} \ \mathbf{PT} \\ \mathbf{0} \ \mathbf{I}_y \end{pmatrix} \mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f = \begin{pmatrix} (\mathbf{P} \ \mathbf{PT})\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f \\ (\mathbf{0} \ \mathbf{I}_y)\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f \end{pmatrix} = \begin{pmatrix} (\mathbf{P} \ \mathbf{PT})\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f \\ \mathrm{diag}(\mathbf{y})\tilde{\mathbf{h}}_{fy} \end{pmatrix}$$

$$=: \begin{pmatrix} \tilde{\mathbf{h}}_x \\ \mathrm{diag}(\mathbf{y})\tilde{\mathbf{h}}_y \end{pmatrix}, \tag{F.24}$$

with

$$\begin{aligned}
\tilde{\mathbf{r}}_x &= \frac{1}{|\mathbf{w}|^2}(\mathbf{P} \ \mathbf{PT})\left[\mathrm{diag}(\mathbf{m}-\widehat{\tilde{m}})\mathbf{B}(\mathbf{m}-\widehat{\tilde{m}}) + \mathrm{diag}(\mathbf{m})\tilde{\mathbf{r}}_f \left|\mathbf{m}-\widehat{\tilde{m}}\right|^2 \right], \\
\tilde{\mathbf{r}}_y &= \frac{1}{|\mathbf{w}|^2}\tilde{\mathbf{r}}_{fy}\left|\mathbf{m}-\widehat{\tilde{m}}\right|^2, \\
\tilde{\mathbf{h}}_x &= (\mathbf{P} \ \mathbf{PT})\mathrm{diag}(\mathbf{m})\tilde{\mathbf{h}}_f, \\
\tilde{\mathbf{h}}_y &= \tilde{\mathbf{h}}_{fy}.
\end{aligned} \tag{F.25}$$

Substituting Eqs. (F.23) and (F.24) into Eq. (F.22) gives Eq. (5.4a) in the main text,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} &= \begin{pmatrix} \mathbf{A}_x\mathbf{x} \\ \mathrm{diag}(\mathbf{y})[\mathbf{U}\mathbf{x}+\mathbf{J}_y\mathbf{y}] \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{r}}_x \\ \mathrm{diag}(\mathbf{y})\tilde{\mathbf{r}}_y \end{pmatrix}|\mathbf{w}|^2 + \varepsilon\begin{pmatrix} \tilde{\mathbf{h}}_x \\ \mathrm{diag}(\mathbf{y})\tilde{\mathbf{h}}_y \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}_x\mathbf{x} + \varepsilon\tilde{\mathbf{h}}_x + \tilde{\mathbf{r}}_x|\mathbf{w}|^2 \\ \mathrm{diag}(\mathbf{y})\left[\mathbf{J}_y\mathbf{y}+\mathbf{U}\mathbf{x}+\varepsilon\tilde{\mathbf{h}}_y+\tilde{\mathbf{r}}_y|\mathbf{w}|^2\right] \end{pmatrix}.
\end{aligned} \tag{F.26}$$

F.4 Finding $\tilde{C}_{rx}$, $\tilde{C}_{ry}$, $\tilde{C}_{hx}$, and $\tilde{C}_{hy}$

By using the following relationships,

$$\begin{aligned}
\frac{\left|\mathbf{m}-\widehat{\tilde{m}}\right|}{|\mathbf{w}|} &\leq \frac{|\mathbf{Q}^{-1}\mathbf{w}|}{|\mathbf{w}|} = \left\|\mathbf{Q}^{-1}\right\|, \\
\frac{\left\|\mathrm{diag}(\mathbf{m}-\widehat{\tilde{m}})\right\|}{|\mathbf{w}|} &\leq \frac{\max_i\left(\left|m_i-\widehat{\tilde{m}}_i\right|\right)}{|\mathbf{w}|} = \frac{\left|\mathbf{m}-\widehat{\tilde{m}}\right|}{|\mathbf{w}|}\frac{\max_i\left(\left|m_i-\widehat{\tilde{m}}_i\right|\right)}{\left|\mathbf{m}-\widehat{\tilde{m}}\right|} \\
&\leq \left\|\mathbf{Q}^{-1}\right\|\sqrt{\frac{\max_i\left(\left|m_i-\widehat{\tilde{m}}_i\right|^2\right)}{\sum_{i=1}^M\left|m_i-\widehat{\tilde{m}}_i\right|^2}} \leq \left\|\mathbf{Q}^{-1}\right\|, \tag{F.27}
\end{aligned}$$

the following formulas for $\tilde{C}_{\mathbf{rx}}$, $\tilde{C}_{\mathbf{ry}}$, $\tilde{C}_{\mathbf{hx}}$, and $\tilde{C}_{\mathbf{hy}}$ are found,

$$
|\tilde{\mathbf{r}}_{\mathbf{x}}| = \left| (\mathbf{P}\ \mathbf{PT}) \left[\frac{\mathrm{diag}(\mathbf{m} - \widehat{\tilde{m}})}{|\mathbf{w}|} \mathbf{B} \frac{(\mathbf{m} - \widehat{\tilde{m}})}{|\mathbf{w}|} + \mathrm{diag}(\mathbf{m}) \tilde{\mathbf{r}}_{\mathrm{f}} \frac{\left| \mathbf{m} - \widehat{\tilde{m}} \right|^2}{|\mathbf{w}|^2} \right] \right|
$$

$$
\leq (\|\mathbf{P}\| + \|\mathbf{PT}\|) \left[\left\| \mathbf{Q}^{-1} \right\| \|\mathbf{B}\| \left\| \mathbf{Q}^{-1} \right\| + \max_i (m_i) \tilde{C}_{\mathrm{rf}} \left\| \mathbf{Q}^{-1} \right\|^2 \right]
$$

$$
\leq \|\mathbf{P}\|(1 + \|\mathbf{T}\|) \left\| \mathbf{Q}^{-1} \right\|^2 \left[\|\mathbf{B}\| + \eta \tilde{C}_{\mathrm{rf}} \right] =: \tilde{C}_{\mathbf{rx}}, \tag{F.28}
$$

$$
\left| \tilde{\mathbf{r}}_{\mathbf{y}} \right| \leq \left| \tilde{\mathbf{r}}_{\mathrm{fy}} \right| \frac{\left| \mathbf{m} - \widehat{\tilde{m}} \right|^2}{|\mathbf{w}|^2} \leq \tilde{C}_{\mathrm{rf}} \left\| \mathbf{Q}^{-1} \right\|^2 =: \tilde{C}_{\mathbf{ry}}, \tag{F.29}
$$

$$
\left| \tilde{\mathbf{h}}_{\mathbf{x}} \right| \leq \|\mathbf{P}\|(1 + \|\mathbf{T}\|) \eta \tilde{C}_{\mathrm{hf}} =: \tilde{C}_{\mathbf{hx}}, \tag{F.30}
$$

$$
\left| \tilde{\mathbf{h}}_{\mathbf{y}} \right| = \left| \tilde{\mathbf{h}}_{\mathrm{fy}} \right| \leq \tilde{C}_{\mathrm{hf}} =: \tilde{C}_{\mathbf{hy}}. \tag{F.31}
$$

F.5 Finding $\tilde{C}_{\mathbf{r}}$ and $\tilde{C}_{\mathbf{h}}$

From Eq. (5.6b) in the main text and Eqs. (F.28) to (F.31), we see

$$
|\tilde{\mathbf{r}}| = \left| \begin{pmatrix} \tilde{\mathbf{r}}_{\mathbf{x}} \\ d\tilde{\mathbf{r}}_{\mathbf{y}} \end{pmatrix} \right| = \sqrt{|\tilde{\mathbf{r}}_{\mathbf{x}}|^2 + d^2 |\tilde{\mathbf{r}}_{\mathbf{y}}|^2} \leq \sqrt{\tilde{C}_{\mathbf{rx}}^2 + d^2 \tilde{C}_{\mathbf{ry}}^2} =: \tilde{C}_{\mathbf{r}},
$$

$$
\left| \tilde{\mathbf{h}} \right| = \left| \begin{pmatrix} \tilde{\mathbf{h}}_{\mathbf{x}} \\ d\tilde{\mathbf{h}}_{\mathbf{y}} \end{pmatrix} \right| = \sqrt{\left| \tilde{\mathbf{h}}_{\mathbf{x}} \right|^2 + d^2 \left| \tilde{\mathbf{h}}_{\mathbf{y}} \right|^2} \leq \sqrt{\tilde{C}_{\mathbf{hx}}^2 + d^2 \tilde{C}_{\mathbf{hy}}^2} =: \tilde{C}_{\mathbf{h}}. \tag{F.32}
$$

Appendix G: Proof of Lemma 9

We denote the eigenvalues of $\mathbf{J}_{\mathbf{x}}$ and $\frac{1}{2} \left(\mathbf{J}_{\mathbf{y}} + \mathbf{J}_{\mathbf{y}}^{\mathrm{T}} \right)$ by $\lambda_1, \ldots, \lambda_L$ and $\lambda_{L+1}, \ldots, \lambda_M$, respectively. We prove Lemma 9 by showing that for $\lambda_{\mathbf{x}\,\mathrm{max}} := \max_{i=1,\ldots,L} (\mathrm{Re}(\lambda_i)) < 0$, $\lambda_{\mathbf{y}\,\mathrm{max}} := \max_{i=L+1,\ldots,M} (\mathrm{Re}(\lambda_i)) < 0$, and a sufficiently small d, $\frac{1}{2} \left[\tilde{\mathbf{A}}^{\mathrm{T}} + \tilde{\mathbf{A}} \right]$ is negative definite,

$$
\mathbf{w}^{\mathrm{T}} \tilde{\mathbf{A}} \mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathbf{A}_{\mathbf{x}} & 0 \\ d\mathbf{U} & d\mathbf{J}_{\mathbf{y}} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}
$$

$$
= \mathbf{x}^{\mathrm{T}} \mathbf{A}_{\mathbf{x}} \mathbf{x} + d\mathbf{y}^{\mathrm{T}} \mathbf{U} \mathbf{x} + d\mathbf{y}^{\mathrm{T}} \mathbf{J}_{\mathbf{y}} \mathbf{y}
$$

$$
= \mathbf{x}^{\mathrm{T}} \mathbf{A}_{\mathbf{x}} \mathbf{x} + d\mathbf{y}^{\mathrm{T}} \mathbf{U} \mathbf{x} + d\mathbf{y}^{\mathrm{T}} \frac{1}{2} \left(\mathbf{J}_{\mathbf{y}} + \mathbf{J}_{\mathbf{y}}^{\mathrm{T}} \right) \mathbf{y}
$$

$$
\leq \lambda_{\mathbf{x}\,\mathrm{max}} |\mathbf{x}|^2 + d\|\mathbf{U}\| |\mathbf{x}| |\mathbf{y}| + d\lambda_{\mathbf{y}\,\mathrm{max}} |\mathbf{y}|^2
$$

$$= \lambda_{x\,\text{max}} \left[|\mathbf{x}|^2 + 2 \frac{d\|\mathbf{U}\| |\mathbf{x}| |\mathbf{y}|}{2\lambda_{x\,\text{max}}} + \frac{d^2\|\mathbf{U}\|^2 |\mathbf{y}|^2}{4\lambda_{x\,\text{max}}^2} \right] - \frac{d^2\|\mathbf{U}\|^2 |\mathbf{y}|^2}{4\lambda_{x\,\text{max}}} + d\lambda_{y\,\text{max}} |\mathbf{y}|^2$$

$$= \lambda_{x\,\text{max}} \left[|\mathbf{x}| + \frac{d\|\mathbf{U}\| |\mathbf{y}|}{2\lambda_{x\,\text{max}}} \right]^2 - \frac{d^2\|\mathbf{U}\|^2 |\mathbf{y}|^2}{4\lambda_{x\,\text{max}}} + d\lambda_{y\,\text{max}} |\mathbf{y}|^2$$

$$= \lambda_{x\,\text{max}} \left[|\mathbf{x}| + \frac{d\|\mathbf{U}\| |\mathbf{y}|}{2\lambda_{x\,\text{max}}} \right]^2 + \frac{d|\mathbf{y}|^2 \|\mathbf{U}\|^2}{4\lambda_{x\,\text{max}}} \left[\frac{4\lambda_{x\,\text{max}} \lambda_{y\,\text{max}}}{\|\mathbf{U}\|^2} - d \right]. \tag{G.1}$$

Thus, $\mathbf{w}^\mathrm{T} \tilde{\mathbf{A}} \mathbf{w}$ is always negative for any $\mathbf{w}$, under $\lambda_x\text{max} < 0$, $\lambda_y\text{max} < 0$, and a sufficiently small d so that

$$d < \frac{4\lambda_x\text{max}\lambda_y\text{max}}{\|\mathbf{U}\|^2}. \tag{G.2}$$

In this case, all eigenvalues of $\frac{1}{2}\left[\tilde{\mathbf{A}}^\mathrm{T} + \tilde{\mathbf{A}} \right]$ are negative. $\qquad\square$

Appendix H: Derivation of Eq. (5.8c)

Here we derive Eq. (5.8c) in the main text. As explained in the main text, the contour curve $V = V_0$ has an inscribed circle $|\mathbf{w}| = \tilde{\phi}_\mathrm{h}$, and a circumscribed circle $|\mathbf{w}| = \alpha\tilde{\phi}_\mathrm{h}$. Specifically, the contour curve is defined by

$$E_\mathrm{V} = \left\{ \mathbf{w} = (x_1, \ldots, x_L, y_1, \ldots, y_K)^\mathrm{T} \, \middle| \, \sum_{i=1}^{L} x_i^2 + 2d \sum_{j=1}^{K} y_j = V_0 \right\}, \tag{H.1}$$

where V_0 is determined so that $|\mathbf{w}| = \tilde{\phi}_\mathrm{h}$ is an inscribed circle of $V = V_0$, i.e.,

$$\tilde{\phi}_\mathrm{h}^2 = \min_{\mathbf{w} \in E_\mathrm{V}} \left(|\mathbf{w}|^2 \right). \tag{H.2}$$

For convenience, we write $\mathbf{y}$ for $\mathbf{y} = (y_1, \ldots, y_K)^\mathrm{T}$, which is identical to $\mathbf{y} = (y_{L+1}, \ldots, y_M)^\mathrm{T}$ in the other appendices and in the main text. On the other hand, the radius of the circumscribed circle, $\alpha\tilde{\phi}_\mathrm{h}$, satisfies

$$\alpha^2 \tilde{\phi}_\mathrm{h}^2 = \max_{\mathbf{w} \in E_\mathrm{V}} \left(|\mathbf{w}|^2 \right). \tag{H.3}$$

Thus, α is given by

$$\alpha = \sqrt{\frac{\alpha^2 \tilde{\phi}_\mathrm{h}^2}{\tilde{\phi}_\mathrm{h}^2}} = \sqrt{\frac{\max_{\mathbf{w} \in E_\mathrm{V}} \left(|\mathbf{w}|^2 \right)}{\min_{\mathbf{w} \in E_\mathrm{V}} \left(|\mathbf{w}|^2 \right)}}. \tag{H.4}$$

To calculate the maximum and minimum of $|\mathbf{w}|^2$ for $\mathbf{w} \in E_V$ we proceed as follows. To make calculation easier, we decompose the expression of E_V as $E_V = \{\mathbf{w}| \ \mathbf{x} \in E_{Vx}, \ \mathbf{y} \in E_{Vy}, \ q \in [0, 1]\}$, with

$$E_{Vx} = \left\{ \mathbf{x}| \sum_{i=1}^{L} x_i^2 = (1-q)V_0 \right\}, \tag{H.5}$$

$$E_{Vy} = \left\{ \mathbf{y}|2d \sum_{j=1}^{K} y_j = q V_0, \ y_j \geq 0 \ \text{for all} \ j = 1, \dots, K \right\}. \tag{H.6}$$

Then, $\min_{\mathbf{w} \in E_V}(|\mathbf{w}|^2)$ and $\max_{\mathbf{w} \in E_V}(|\mathbf{w}|^2)$ are expressed as

$$\min_{\mathbf{w} \in E_V}(|\mathbf{w}|^2) = \min_{q \in [0,1]} \left(\min_{\mathbf{x} \in E_{Vx}}(|\mathbf{x}|^2) + \min_{\mathbf{y} \in E_{Vy}}(|\mathbf{y}|^2) \right), \tag{H.7}$$

$$\max_{\mathbf{w} \in E_V}(|\mathbf{w}|^2) = \max_{q \in [0,1]} \left(\max_{\mathbf{x} \in E_{Vx}}(|\mathbf{x}|^2) + \max_{\mathbf{y} \in E_{Vy}}(|\mathbf{y}|^2) \right). \tag{H.8}$$

Before performing the extremisations, we first work out the various expressions for a given value of q. We start with the $\mathbf{x}$-part of $\mathbf{w}$. Clearly,

$$\min_{\mathbf{x} \in E_{Vx}}(|\mathbf{x}|^2) = \max_{\mathbf{x} \in E_{Vx}}(|\mathbf{x}|^2) = (1-q)V_0. \tag{H.9}$$

For the calculation of $\max_{\mathbf{y} \in E_{Vy}}(|\mathbf{y}|^2)$, the $|\mathbf{y}|^2$ can be transformed by $2d \sum_{j=1}^{K} y_j = q V_0$ into

$$|\mathbf{y}|^2 = \sum_{i=1}^{K} y_i^2 = \left[\sum_{i=1}^{K} y_i \right]^2 - 2 \sum_{j=1}^{K} \sum_{k=j+1}^{K} y_j y_k = \frac{q^2 V_0^2}{4d^2} - 2 \sum_{j=1}^{K} \sum_{k=j+1}^{K} y_j y_k. \tag{H.10}$$

Since $y_i \geq 0$ for $i = 1, \dots, K$, $\max_{\mathbf{y} \in E_{Vy}}(|\mathbf{y}|^2)$ is given by

$$\max_{\mathbf{y} \in E_{Vy}}(|\mathbf{y}|^2) = \frac{q^2 V_0^2}{4d^2}, \tag{H.11}$$

which is attained when $y_{j_0} = \frac{q V_0}{2d}$ and $y_{j \neq j_0} = 0$ for some $j_0 = 1, \dots, K$.

To calculate $\min_{\mathbf{y} \in E_{Vy}}(|\mathbf{y}|^2)$, Eq. (H.10) is further transformed into

$$\sum_{i=1}^{K} y_i^2 = \left[\sum_{i=1}^{K} y_i \right]^2 - 2 \sum_{j=1}^{K} \sum_{k=j+1}^{K} y_j y_k$$
$$= \frac{q^2 V_0^2}{4d^2} + \sum_{j=1}^{K} \sum_{k=j+1}^{K} (y_j - y_k)^2 - \sum_{j=1}^{K} \sum_{k=j+1}^{K} (y_j^2 + y_k^2)$$

$$= \frac{q^2 V_0^2}{4d^2} + \sum_{j=1}^{K} \sum_{k=j+1}^{K} (y_j - y_k)^2$$

$$- \sum_{j=1}^{K} \sum_{k=j+1}^{K} y_j^2 - \sum_{j=1}^{K} \sum_{k=j+1}^{K} y_k^2$$

$$= \frac{q^2 V_0^2}{4d^2} + \sum_{j=1}^{K} \sum_{k=j+1}^{K} (y_j - y_k)^2$$

$$- \left[(K-1)y_1^2 + \cdots + y_{K-1}^2 \right] - \left[(y_2^2 + \cdots + y_K^2) + \cdots + (y_K^2) \right]$$

$$= \frac{q^2 V_0^2}{4d^2} + \sum_{j=1}^{K} \sum_{k=j+1}^{K} (y_j - y_k)^2 - (K-1) \sum_{j=1}^{K} y_j^2. \tag{H.12}$$

Solving Eq. (H.12) for $\sum_{i=1}^{K} y_i^2 = |\mathbf{y}|^2$ gives

$$|\mathbf{y}|^2 = \sum_{i=1}^{K} y_i^2 = \frac{q^2 V_0^2}{4d^2 K} + \frac{1}{K} \sum_{j=1}^{K} \sum_{k=j+1}^{K} (y_j - y_k)^2. \tag{H.13}$$

Therefore,

$$\min_{\mathbf{y} \in E_{V_y}} \left(|\mathbf{y}|^2 \right) = \frac{q^2 V^2}{4d^2 K}, \tag{H.14}$$

which is attained when $y_i = \frac{q V_0}{2dK}$ for all $i = 1, \ldots K$. Substituting Eqs. (H.9), (H.11), and (H.14) into Eqs. (H.7) and (H.8) gives

$$\min_{\mathbf{w} \in E_V} \left(|\mathbf{w}|^2 \right) = \min_{q \in [0,1]} \left((1-q)V_0 + \frac{q^2 V_0^2}{4d^2 K} \right)$$

$$= \min_{q \in [0,1]} \left(\frac{V_0^2}{4d^2 K} q^2 - V_0 q + V_0 \right)$$

$$= \min_{q \in [0,1]} \left(\frac{V_0^2}{4d^2 K} \left[q - \frac{2d^2 K}{V_0} \right]^2 + V_0 - Kd^2 \right)$$

$$= \begin{cases} V_0 - Kd^2 & \text{for } \frac{2d^2 K}{V_0} \leq 1 \quad \left(\text{attained when } q = \frac{2d^2 K}{V_0} \right) \\ \frac{V_0^2}{4d^2 K} & \text{for } \frac{2d^2 K}{V_0} > 1 \quad \text{(attained when } q = 1) \end{cases} \tag{H.15}$$

and

$$
\begin{aligned}
\max_{\mathbf{w} \in E_V} \left(|\mathbf{w}|^2 \right) &= \max_{q \in [0,1]} \left(\frac{q^2 V_0^2}{4d^2} + (1-q) V_0 \right) \\
&= \max \left\{ \frac{V_0^2}{4d^2}, V_0 \right\} \quad \text{(attained when } q = 0 \text{ or } 1\text{).}
\end{aligned}
\tag{H.16}
$$

Thus, from Eqs. (H.2) and (H.15), V_0 is obtained as

$$
V_0 = \begin{cases} \tilde{\phi}_h^2 + K d^2 & \text{for } K \leq \frac{\tilde{\phi}_h^2}{d^2} \\ 2d\sqrt{K} \tilde{\phi}_h & \text{for } K > \frac{\tilde{\phi}_h^2}{d^2}. \end{cases}
\tag{H.17}
$$

Finally, substituting Eqs. (H.16) and (H.17) into Eq. (H.4) gives

$$
\alpha = \sqrt{\frac{\max_{\mathbf{w} \in E_V} \left(|\mathbf{w}|^2 \right)}{\tilde{\phi}_h^2}} = \begin{cases} 1 & \text{for } K = 0 \\ \max \left\{ \frac{\tilde{\phi}_h^2 + K d^2}{2d\tilde{\phi}_h}, \frac{\sqrt{\tilde{\phi}_h^2 + K d^2}}{\tilde{\phi}_h} \right\} & \text{for } 0 < K \leq \frac{\tilde{\phi}_h^2}{d^2} \\ \max \left\{ \sqrt{K}, \frac{\sqrt{2dK^{1/2}\tilde{\phi}_h}}{\tilde{\phi}_h} \right\} & \text{for } \frac{\tilde{\phi}_h^2}{d^2} < K, \end{cases}
\tag{H.18}
$$

where $\max_{\mathbf{w} \in E_V} \left(|\mathbf{w}|^2 \right) = \max_{\mathbf{w} \in E_V} \left(|\mathbf{x}|^2 \right) = \tilde{\phi}_h^2$ for $K = 0$ is used.

Appendix I: Generalization to higher-dimensional trait spaces

Throughout the manuscript, the dimensionality of the considered trait space matters only when the fitness function (or its derivatives with respect to population densities or phenotypes) is Taylor expanded, as in Eqs. (A.11), (B.10), (B.23), (C.1), and (C.8), for a non-representative phenotype s_j for $j = M+1, \ldots, N+1$, around its representative phenotype $s_{\mathrm{cid}(j)}$. Those equations are readily extended to higher-dimensional trait spaces, by replacing those derivatives with the corresponding directional derivatives, as explained below.

We consider a trait space of arbitrary dimension Z, having N resident phenotypes $\mathbf{v}_1, \ldots, \mathbf{v}_N$ and a mutant $\mathbf{v}' = \mathbf{v}_{N+1}$, where the jth phenotype is denoted by $\mathbf{v}_j = (v_{j,1}, \ldots, v_{j,Z})^{\mathrm{T}} \in \mathbb{R}^Z$. The fitness function, denoted by $F(\mathbf{v}_j; \mathbf{V}'; \mathbf{n}')$ with $\mathbf{V}' = (\mathbf{v}_1, \ldots, \mathbf{v}_{N+1})$ and $\mathbf{n}' = (n_1, \ldots, n_{N+1})^{\mathrm{T}}$, is assumed to satisfy all axioms (i–v) in Sect. 2. Analogously to the one-dimensional case, we permute and cluster those $N+1$ phenotypes into M groups so that their representatives, i.e., approximate phenotypes, $\mathbf{V}_a = (\mathbf{v}_1, \ldots, \mathbf{v}_M)^{\mathrm{T}}$ satisfy $\left| \mathbf{v}_j - \mathbf{v}_{\mathrm{cid}(j)} \right| < \varepsilon$ for all $j = M+1, \ldots, N+1$ and $\mathrm{cid}(j) \in \{1, \ldots, M\}$. To expand $F(\mathbf{v}_j; \mathbf{V}'; \mathbf{n}')$ using directional derivatives, we introduce $\mathbf{u}_j(s_j)$ for $j = M+1, \ldots, N+1$ as

$$
\mathbf{u}_j(s_j) := \mathbf{v}_{\mathrm{cid}(j)} + (s_j - s_{\mathrm{cid}(j)}) \mathbf{e}_{j,\mathrm{cid}(j)},
\tag{I.1}
$$

with a scalar parameter s_j,

$$\mathbf{e}_{j,\text{cid}(j)} := \frac{\mathbf{v}_j - \mathbf{v}_{\text{cid}(j)}}{\left|\mathbf{v}_j - \mathbf{v}_{\text{cid}(j)}\right|}, \tag{I.2}$$

and $s_{\text{cid}(j)} \in \{s_1, \ldots, s_M\}$, where $s_1, \ldots, s_M$ can be chosen arbitrarily as long as $\mathbf{u}_j(s_j) = \mathbf{v}_j$ holds for $j = M+1, \ldots, N+1$ (because all the expansions of the fitness function in this paper are in non-representative phenotypes that correspond to $s_{M+1}, \ldots, s_{N+1}$). Notice that $\mathbf{u}_j(s_{\text{cid}(j)}) = \mathbf{v}_{\text{cid}(j)}$. For notational convenience, we also introduce for $j = 1, \ldots, M$

$$\mathbf{u}_j(s_j) := \mathbf{v}_j. \tag{I.3}$$

Then we define for $j = 1, \ldots, N+1$

$$F_j(s_j; \mathbf{s}'; \mathbf{n}') := F(\mathbf{u}_j(s_j); \mathbf{U}'(\mathbf{s}'); \mathbf{n}') \tag{I.4}$$

with $\mathbf{s}' := (s_1, \ldots, s_{N+1})^{\text{T}}$, $\mathbf{u}_j(s_j) = \mathbf{v}_j$, and $\mathbf{U}'(\mathbf{s}') = (\mathbf{u}_1(s_1), \ldots, \mathbf{u}_{N+1}(s_{N+1})) = \mathbf{V}'$. Note that the $F_j(s_j; \mathbf{s}'; \mathbf{n}')$ satisfy the smoothness axiom (i), the reducibility axiom (ii), and the bounded-world axiom (iii). The exchangeability axiom (iv) is also satisfied between s_j and its representative phenotype $s_{\text{cid}(j)}$, for $j = M+1, \ldots, N+1$ (i.e., $F_j(s_j; \mathbf{s}'; \mathbf{n}') = F_{\text{cid}(j)}(s_{\text{cid}(j)}; \mathbf{s}'; \mathbf{n}')$ for $s_j = s_{\text{cid}(j)}$). Thus, by Taylor's theorem, $F_j(s_j; \mathbf{s}'; \mathbf{n}')$ for $j = M+1, \ldots, N+1$ can be expanded in s_j around $s_{\text{cid}(j)}$ as

$$\begin{aligned}
F_j(s_j; \mathbf{s}'; \mathbf{n}') &= F_j(s_j; \mathbf{s}'; \mathbf{n}')\big|_{s_j = s_{\text{cid}(j)}} + \left.\frac{\partial F_j(s_j; \mathbf{s}'; \mathbf{n}')}{\partial s_j}\right|_{s_j = s_{j\text{T}}} (s_j - s_{\text{cid}(j)}) \\
&= F_{\text{cid}(j)}(s_j; \mathbf{s}'; \mathbf{n}')\big|_{s_j = s_{\text{cid}(j)}} + \left.\frac{\partial F_j(s_j; \mathbf{s}'; \mathbf{n}')}{\partial s_j}\right|_{s_j = s_{j\text{T}}} (s_j - s_{\text{cid}(j)})
\end{aligned} \tag{I.5}$$

with some appropriately chosen $s_{j\text{T}} \in [s_j, s_{\text{cid}(j)}]$. The derivatives of $F_j(s_j; \mathbf{s}'; \mathbf{n}')$ with respect to population densities or phenotypes can be expanded in the same manner. Therefore, for all $\cdot = i, j, k$ replacing $F(s.; \mathbf{s}'; \mathbf{n}')$ with $F.(s.; \mathbf{s}'; \mathbf{n}')$, $F(z.; \mathbf{s}'; \mathbf{n}')$ with $F.(z.; \mathbf{s}'; \mathbf{n}')$, $F(s.; \mathbf{s}'; \widehat{n}')$ with $F.(s.; \mathbf{s}'; \widehat{n}')$, $\breve{F}(s.; \mathbf{s}'; \mathbf{m}')$ with $\breve{F}.(s.; \mathbf{s}'; \mathbf{m}')$, $F(s.; \mathbf{s}_{\text{a}}; \mathbf{m})$ with $F.(s.; \mathbf{s}_{\text{a}}; \mathbf{m})$ and $F(s_{\text{cid}(\cdot)} + \varepsilon\rho.; \mathbf{s}'; \mathbf{n}')$ with $F.(s_{\text{cid}(\cdot)} + \varepsilon\rho.; \mathbf{s}'; \mathbf{n}')$ throughout this paper (except in this section, Sect. 7, and Appendix K) gives the complete proofs for Theorems 1–3 for the fitness function $F(\mathbf{v}_j; \mathbf{V}'; \mathbf{n}')$.

Appendix J: Tighter estimates

J.1 Main derivation

Here we derive an approximability condition based on the first-order approximation of the perturbation term, corresponding to the zeroth-order approximability condition in Sect. 4, Eq. (4.8a),

$$\sqrt{\varepsilon} < \frac{-\lambda_{\max}}{2\sqrt{C_{\mathbf{h}}C_{\mathbf{r}}}}, \tag{J.1}$$

which is applied when the initial equilibrium population densities of approximate phenotypes are not small. The standardized dynamics of the approximate phenotypes is given by Eq. (4.4) in the main text,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{r}|\mathbf{x}|^2 + \varepsilon\mathbf{h}. \tag{J.2}$$

By Taylor's theorem, the perturbation term expands as

$$\mathbf{h} = \mathbf{h}|_{\mathbf{x}=\mathbf{0}} + \left.\frac{\partial\mathbf{h}}{\partial\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_{\mathrm{T}}} \mathbf{x}$$

$$= \widehat{h} + \mathbf{H}\mathbf{x}, \tag{J.3}$$

where $\mathbf{H}$ is an $M \times M$ matrix given by

$$\mathbf{H} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_M} \\ \cdots & \cdots & \cdots \\ \frac{\partial h_M}{\partial x_1} & \cdots & \frac{\partial h_M}{\partial x_M} \end{pmatrix}_{\mathbf{x}=\mathbf{x}_{\mathrm{T}}}, \tag{J.4}$$

where $\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{T}}$ and $\mathbf{x}_{\mathrm{T}} = \theta_{\mathrm{T}}\mathbf{x}$ with some appropriately chosen $\theta_{\mathrm{T}} \in [0, 1]$. There exist constants $\widehat{C}_h$ and $C_{\mathbf{H}}$ such that $\left|\widehat{h}\right| \leq \widehat{C}_h$ and $\|\mathbf{H}\| \leq C_{\mathbf{H}}$; see next subsection. We introduce

$$c_{\mathbf{h}}(\phi_{\mathbf{x}}) := \widehat{C}_h + C_{\mathbf{H}}\phi_{\mathbf{x}}, \tag{J.5}$$

which clearly satisfies

$$\max_{|\mathbf{x}|\leq\phi_{\mathbf{x}}} (|\mathbf{h}|) \leq c_{\mathbf{h}}(\phi_{\mathbf{x}}). \tag{J.6}$$

This means that $|\mathbf{h}|$ does not exceed $c_{\mathbf{h}}(\phi_x)$ when $\mathbf{x}$ is within a circle of radius ϕ_x. Thus, replacing C_h with $c_{\mathbf{h}}(\phi_x)$ in Eqs. (4.8a) and (4.8b) in Lemma 5 in Sect. 4, we have

$$\varepsilon < \frac{\lambda_{\max}^2}{4c_{\mathbf{h}}(\phi_x)C_{\mathbf{r}}} = \frac{\lambda_{\max}^2}{4(\widehat{C}_h + C_{\mathbf{H}}\phi_x)C_{\mathbf{r}}} \tag{J.7}$$

and

$$|\mathbf{x}| \le 2\varepsilon\frac{c_{\mathbf{h}}(\phi_x)}{|\lambda_{\max}|} = 2\varepsilon\frac{\widehat{C}_h + C_{\mathbf{H}}\phi_x}{|\lambda_{\max}|}. \tag{J.8}$$

Equation (J.8) is ensured by Eq. (J.7), as long as ϕ_x is appropriately chosen so that the right-hand side of Eq. (J.8) does not exceed ϕ_x, i.e.,

$$2\varepsilon\frac{\widehat{C}_h + C_{\mathbf{H}}\phi_x}{|\lambda_{\max}|} \le \phi_x. \tag{J.9}$$

Clearly, the smallest ϕ_x that satisfies both Eq. (J.7) and Eq. (J.9) makes the approximability condition Eq. (J.7) the easiest to satisfy for a given $\lambda_{\max}$, and simultaneously makes the right-hand side of Eq. (J.8) the smallest. Such a ϕ_x is given by solving Eq. (J.9) assuming equality, i.e.,

$$\phi_x = \frac{2\varepsilon\widehat{C}_h}{|\lambda_{\max}| - 2\varepsilon C_{\mathbf{H}}}. \tag{J.10}$$

Substituting Eq. (J.10) into Eq. (J.7) gives the improved stability condition, Eq. (6.2) in the main text,

$$\sqrt{\varepsilon} < \frac{-\lambda_{\max}}{\sqrt{4\widehat{C}_h C_{\mathbf{r}} + 2C_{\mathbf{H}}|\lambda_{\max}|}}. \tag{J.11}$$

This approximability condition can be further improved by the higher-order approximation of the nonlinear term and/or perturbation term, although the resultant conditions will be less simple than Eq. (J.11).

J.2 Finding $C_{\mathbf{h}}$ and $C_{\mathbf{H}}$

From Eqs. (C.11), (C.19), and (D.2), $\mathbf{h}$ is given by

$$\mathbf{h} = \mathbf{P}\mathbf{h}_{\mathrm{m}} = \mathbf{P}\mathrm{diag}(\mathbf{m})\mathbf{h}_{\mathrm{f}}. \tag{J.12}$$

with $\mathbf{h}_{\mathrm{f}} = (h_{\mathrm{f}1}, \ldots, h_{\mathrm{f}M})^{\mathrm{T}}$ and

$$h_{\mathrm{f}i} = \frac{1}{\varepsilon} F(s_i; \mathbf{s}'; \widehat{n'}) - \widehat{F}_{\varepsilon i} + \sum_{j \in \mathrm{com}(i)} p_j F_{\varepsilon j}. \tag{J.13}$$

As for $\widehat{C}_h$,

$$\widehat{h} = \mathbf{h}|_{\mathbf{m}=\widehat{m}} = \left[\mathbf{P}\mathrm{diag}(\widehat{m})\mathbf{h}_{\mathrm{f}} \right]_{\mathbf{m}=\widehat{m}} \tag{J.14}$$

gives

$$\left| \widehat{h} \right| \leq \left\| \mathbf{P}\mathrm{diag}(\widehat{m})\mathbf{h}_{\mathrm{f}} \right\|_{\mathbf{m}=\widehat{m}} =: \widehat{C}_h. \tag{J.15}$$

As for $C_{\mathbf{H}}$, we first transform $\mathbf{H}$ by using $\mathbf{x} = \mathbf{P}(\mathbf{m} - \widehat{m})$ and Eq. (J.12) as

$$\mathbf{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \frac{\partial \mathbf{h}}{\partial \mathbf{m}} \frac{\partial \mathbf{m}}{\partial \mathbf{x}} = \mathbf{P}\frac{\partial [\mathrm{diag}(\mathbf{m})\mathbf{h}_{\mathrm{f}}]}{\partial \mathbf{m}}\mathbf{P}^{-1}$$

$$= \mathbf{P}\begin{pmatrix} \frac{\partial m_1 h_{\mathrm{f}1}}{\partial m_1} & \cdots & \frac{\partial m_1 h_{\mathrm{f}1}}{\partial m_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial m_M h_{\mathrm{f}M}}{\partial m_1} & \cdots & \frac{\partial m_M h_{\mathrm{f}M}}{\partial m_M} \end{pmatrix}\mathbf{P}^{-1}. \tag{J.16}$$

To calculate $\frac{\partial m_i h_{\mathrm{f}i}}{\partial m_k}$ for $i, k = 1, \ldots, M$, we transform $m_i h_{\mathrm{f}i}$ by using Eq. (J.13), Eqs. (A.6), and $F(s_i; \mathbf{s}'; \widehat{n'}) = 0$ for $i = 1, \ldots, M$ as

$$m_i h_{\mathrm{f}i} = -m\widehat{F}_{\varepsilon i} + \sum_{j \in \mathrm{com}(i)} n_j F_{\varepsilon j}$$

$$= -m_i \widehat{F}_{\varepsilon i} + n_i F_{\varepsilon i} + \sum_{j \in \mathrm{com}(i), j \neq i} n_j F_{\varepsilon j}$$

$$= -m\widehat{F}_{\varepsilon i} + \left(m_i - \frac{1}{\varepsilon} \sum_{j \in \mathrm{com}(i), j \neq i} m_j \right) F_{\varepsilon i} + \frac{1}{\varepsilon} \sum_{j \in \mathrm{com}(i), j \neq i} m_j F_{\varepsilon j}$$

$$= m_i(F_{\varepsilon i} - \widehat{F}_{\varepsilon i}) - \frac{1}{\varepsilon} \sum_{j \in \mathrm{com}(i), j \neq i} m_j F_{\varepsilon i} + \frac{1}{\varepsilon} \sum_{j \in \mathrm{com}(i), j \neq i} m_j F_{\varepsilon j}. \tag{J.17}$$

Since $j \in \mathrm{com}(i)$ and $j \neq i$ require $j \in \{M+1, \ldots, N+1\}$, we see for $i, k = 1, \ldots, M$

$$\frac{\partial m_i h_{\mathrm{f}i}}{\partial m_k} = \frac{\partial m_i}{\partial m_k}(F_{\varepsilon i} - \widehat{F}_{\varepsilon i}) + m_i \frac{\partial F_{\varepsilon i}}{\partial m_k} - \frac{1}{\varepsilon} \sum_{j \in \mathrm{com}(i), j \neq i} m_j \frac{\partial F_{\varepsilon i}}{\partial m_k} + \frac{1}{\varepsilon} \sum_{j \in \mathrm{com}(i), j \neq i} m_j \frac{\partial F_{\varepsilon j}}{\partial m_k}$$

$$= \delta(i-k)(F_{\varepsilon i} - \widehat{F}_{\varepsilon i}) + \sum_{j \in \mathrm{com}(i)} n_j \frac{\partial F_{\varepsilon i}}{\partial m_k} - \sum_{j \in \mathrm{com}(i), j \neq i} n_j \frac{\partial F_{\varepsilon i}}{\partial m_k} + \sum_{j \in \mathrm{com}(i), j \neq i} n_j \frac{\partial F_{\varepsilon j}}{\partial m_k}$$

$$= \delta(i-k)(F_{\varepsilon i} - \widehat{F}_{\varepsilon i}) + \sum_{j \in \mathrm{com}(i)} n_j \frac{\partial F_{\varepsilon j}}{\partial m_k}$$

$$= \delta(i-k)(F_{\varepsilon i} - \widehat{F}_{\varepsilon i}) + \mathbf{c}_i^{\mathrm{T}} \mathrm{diag}(\mathbf{n}') \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial m_k}, \tag{J.18}$$

where $\delta(i-k) = 1$ for $i = k$, $\delta(i-k) = 0$ otherwise, $\mathbf{F}'_{\varepsilon} := (F_{\varepsilon 1}, \ldots, F_{\varepsilon, N+1})^{\mathrm{T}}$, and $\mathbf{c}_i$ is a vector of length $N + 1$ with entries 0 or 1 such that $\mathbf{c}_i^{\mathrm{T}} \mathbf{n}' = \sum_{j \in \mathrm{com}(i)} n_j = m_i$. We transform Eq. (J.18) into vector–matrix form,

$$\frac{\partial [\mathrm{diag}(\mathbf{m})\mathbf{h}_{\mathrm{f}}]}{\partial \mathbf{m}} = \mathrm{diag}(\mathbf{F}_{\varepsilon} - \widehat{F}_{\varepsilon}) + \begin{pmatrix} \mathbf{c}_1^{\mathrm{T}} \mathrm{diag}(\mathbf{n}') \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial m_1} & \cdots & \mathbf{c}_1^{\mathrm{T}} \mathrm{diag}(\mathbf{n}') \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial m_M} \\ \vdots & \ddots & \vdots \\ \mathbf{c}_M^{\mathrm{T}} \mathrm{diag}(\mathbf{n}') \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial m_1} & \cdots & \mathbf{c}_M^{\mathrm{T}} \mathrm{diag}(\mathbf{n}') \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial m_M} \end{pmatrix}$$

$$= \mathrm{diag}(\mathbf{F}_{\varepsilon} - \widehat{F}_{\varepsilon}) + \begin{pmatrix} \mathbf{c}_1^{\mathrm{T}} \mathrm{diag}(\mathbf{n}') \\ \vdots \\ \mathbf{c}_M^{\mathrm{T}} \mathrm{diag}(\mathbf{n}') \end{pmatrix} \begin{pmatrix} \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial m_1} & \cdots & \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial m_M} \end{pmatrix}$$

$$= \mathrm{diag}(\mathbf{F}_{\varepsilon} - \widehat{F}_{\varepsilon}) + \begin{pmatrix} \mathbf{c}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{c}_M^{\mathrm{T}} \end{pmatrix} \mathrm{diag}(\mathbf{n}') \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial \mathbf{m}}. \tag{J.19}$$

Substituting Eq. (J.19) into Eq. (J.16), we find

$$\|\mathbf{H}\| \leq \|\mathbf{P}\| \left\| \frac{\partial [\mathrm{diag}(\mathbf{m})\mathbf{h}_{\mathrm{f}}]}{\partial \mathbf{m}} \right\| \|\mathbf{P}^{-1}\|$$

$$\leq \|\mathbf{P}\| \left[\left\| \mathrm{diag}(\mathbf{F}_{\varepsilon} - \widehat{F}_{\varepsilon}) \right\| + \left\| \begin{pmatrix} \mathbf{c}_1^{\mathrm{T}} \\ \mathbf{c}_M^{\mathrm{T}} \end{pmatrix} \right\| \|\mathrm{diag}(\mathbf{n}')\| \left\| \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial \mathbf{m}} \right\| \right] \|\mathbf{P}^{-1}\|$$

$$\leq \|\mathbf{P}\| \left[2 \max_{i \in \{1, \ldots, M\}} (|F_{\varepsilon i}|) + (N+1) \max_{i \in \{1, \ldots, N+1\}} (|n_i|) \left\| \frac{\partial \mathbf{F}'_{\varepsilon}}{\partial \mathbf{m}} \right\| \right] \|\mathbf{P}^{-1}\|$$

$$\leq \|\mathbf{P}\| \|\mathbf{P}^{-1}\| [2 C_{\mathrm{F}\varepsilon} + (N+1)\eta C_{\mathbf{F}\varepsilon\mathbf{m}}] =: C_{\mathbf{H}}, \tag{J.20}$$

with

$$C_{\mathrm{F}\varepsilon} := \max \left\{ \left. \left| \frac{\partial F(s_{\mathrm{cid}(i)} + \varepsilon \rho_i; \mathbf{s}'_a + \varepsilon \boldsymbol{\rho}'; \mathbf{n}')}{\partial \varepsilon} \right| \right|_{\varepsilon = \varepsilon_{\mathrm{T}i}} \left| \begin{array}{l} i = 1, \ldots, M, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \ \varepsilon_{\mathrm{T}i} \in [0, \varepsilon] \end{array} \right\} \right. \tag{J.21}$$

and

$$
C_{\mathbf{F}\varepsilon\mathbf{m}} := \max\left\{ \left\| \frac{\partial \mathbf{F}'_\varepsilon}{\partial \mathbf{m}} \right\| \; \left| \; \begin{array}{c} \varepsilon_{\mathrm{T}i} \in [0, \varepsilon], \\ m_1, \ldots, m_M \in [0, \eta], \\ m_{M+1}, \ldots, m_{N+1} \in [0, \varepsilon\eta] \end{array} \right. \right\},
$$

$$
\mathbf{F}'_\varepsilon = \begin{pmatrix} F_{\varepsilon 1} \\ \vdots \\ F_{\varepsilon,N+1} \end{pmatrix} = \begin{pmatrix} \frac{\partial \breve{F}(s_{\mathrm{cid}(1)} + \varepsilon\rho_1; \mathbf{s}'_{\mathrm{a}} + \varepsilon\boldsymbol{\rho}'; \mathbf{m}')}{\partial \varepsilon} \\ \vdots \\ \frac{\partial \breve{F}(s_{\mathrm{cid}(N+1)} + \varepsilon\rho_{N+1}; \mathbf{s}'_{\mathrm{a}} + \varepsilon\boldsymbol{\rho}'; \mathbf{m}')}{\partial \varepsilon} \end{pmatrix}, \tag{J.22}
$$

where $F_{\varepsilon i}$ is defined by Eq. (C.2) and $\breve{F}(s_i; \mathbf{s}'; \mathbf{m}') = F(s_i; \mathbf{s}'; \mathbf{n}')$ as in Lemma 1.

Appendix K: Proof of Eq. (8.2)

K.1 Main proof

First, we derive Eqs. (8.1) from Eq. (4.1a) in the main text,

$$
\begin{aligned}
\frac{\mathrm{d}p_{N+1}}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{n_{N+1}}{m_i}\right] = \frac{1}{m_i}\frac{\mathrm{d}n_{N+1}}{\mathrm{d}t} - \frac{n_{N+1}}{m_i^2}\frac{\mathrm{d}m_i}{\mathrm{d}t} \\
&= \frac{n_{N+1}}{m_i}F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - \frac{n_{N+1}}{m_i}f(s_i; \mathbf{s}'; \mathbf{n}') \\
&= p_{N+1}F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - p_{N+1}\sum_{j\in\mathrm{com}(i)} p_j F(s_j; \mathbf{s}'; \mathbf{n}') \\
&= p_{N+1}\left[F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - \sum_{j\in\mathrm{com}(i)} p_j F(s_j; \mathbf{s}'; \mathbf{n}')\right]. \tag{K.1}
\end{aligned}
$$

This is further transformed as

$$
\begin{aligned}
\frac{\mathrm{d}p_{N+1}}{\mathrm{d}t} &= p_{N+1}\left[(1 - p_{N+1})F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - \sum_{j\in\mathrm{com}(i), j\neq N+1} p_j F(s_j; \mathbf{s}'; \mathbf{n}')\right] \\
&= p_{N+1}(1 - p_{N+1})\left[F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - \frac{1}{1 - p_{N+1}}\sum_{j\in\mathrm{com}(i), j\neq N+1} p_j F(s_j; \mathbf{s}'; \mathbf{n}')\right] \\
&= p_{N+1}(1 - p_{N+1})\left[F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - \sum_{j\in\mathrm{com}(i), j\neq N+1} \tilde{p}_j F(s_j; \mathbf{s}'; \mathbf{n}')\right], \tag{K.2}
\end{aligned}
$$

where $\tilde{p}_j = p_j/(1 - p_{N+1}) = n_j/(m_i - n_{N+1})$ is the proportion of phenotype s_j in the ith cluster when the mutant s_{N+1} is removed. Thus, the sign of $\frac{\mathrm{d}p_{N+1}}{\mathrm{d}t}$ is determined

by the terms inside the square bracket. The $F(s_j; \mathbf{s}'; \mathbf{n}')$ is expanded in s_j around $s_i = s_{\text{cid}(j)}$ as

$$
F(s_j; \mathbf{s}'; \mathbf{n}') = F(s_i; \mathbf{s}'; \mathbf{n}') + \left. \frac{\partial F(z_j; \mathbf{s}'; \mathbf{n}')}{\partial z_j} \right|_{z_j = s_i} (s_j - s_i) + \frac{1}{2} \left. \frac{\partial^2 F(z_j; \mathbf{s}'; \mathbf{n}')}{\partial z_j^2} \right|_{z_j = s_{jT}} (s_j - s_i)^2
$$

$$
= F(s_i; \mathbf{s}'; \mathbf{n}') + \left. \frac{\partial F(z_j; \mathbf{s}'; \mathbf{n}')}{\partial z_j} \right|_{z_j = s_i} \rho_j \varepsilon + \frac{1}{2} \left. \frac{\partial^2 F(z_j; \mathbf{s}'; \mathbf{n}')}{\partial z_j^2} \right|_{z_j = s_{jT}} \rho_j^2 \varepsilon^2
$$

$$
= F(s_i; \mathbf{s}'; \mathbf{n}') + F_z(s_i)\rho_j \varepsilon + \frac{1}{2} F_{zz}(s_{jT})\rho_j^2 \varepsilon^2 \tag{K.3}
$$

with some appropriately chosen $s_{jT} \in [s_i, s_j]$, where $|\rho_j| \leq 1$ because the within-cluster phenotypic differences do not exceed ε. Substituting Eq. (K.3) into Eq. (K.2) transforms the term inside of the square bracket as

$$
F(s_{N+1}; \mathbf{s}'; \mathbf{n}') - \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j F(s_j; \mathbf{s}'; \mathbf{n}')
$$

$$
= F(s_i; \mathbf{s}'; \mathbf{n}') + F_z(s_i)\rho_{N+1}\varepsilon + \frac{1}{2}F_{zz}(s_{N+1,\text{T}})\rho_{N+1}^2 \varepsilon^2
$$

$$
- \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j \left[F(s_i; \mathbf{s}'; \mathbf{n}') + F_z(s_i)\rho_j \varepsilon + \frac{1}{2}F_{zz}(s_{jT})\rho_j^2 \varepsilon^2 \right]
$$

$$
= \varepsilon F_z(s_i) \left[\rho_{N+1} - \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j \rho_j \right]
$$

$$
+ \frac{\varepsilon^2}{2} \left[F_{zz}(s_{N+1,\text{T}})\rho_{N+1}^2 - \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j F_{zz}(s_{jT})\rho_j^2 \right]. \tag{K.4}
$$

The second term satisfies

$$
\left| F_{zz}(s_{N+1,\text{T}})\rho_{N+1}^2 - \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j F_{zz}(s_{jT})\rho_j^2 \right|
$$

$$
\leq \left| F_{zz}(s_{N+1,\text{T}}) \right| \rho_{N+1}^2 + \max_{j \in \text{com}(i), j \neq N+1} \left| F_{zz}(s_{jT}) \right| \rho_j^2
$$

$$
\leq 2 \max_{j \in \text{com}(i), j \neq i} \left(\left| F_{zz}(s_{jT}) \right| \right)
$$

$$
\leq 2 C'_{\text{Fzz}}, \tag{K.5}
$$

where the constant C'_{Fzz} is defined by

$$
C'_{\text{Fzz}} := \max \left\{ \left. \left| \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z^2} \right| \right|_{z = s_{jT}} \right| \left| \Delta \mathbf{m}' \right| \in [0, \varepsilon C'_{\mathbf{m}}],
$$

$$
j = M+1, \ldots, N+1, \ s_{jT} \in [s_j, s_{\text{cid}(j)}] \right\}, \tag{K.6}
$$

while the first term is further transformed by expanding $F_z(s_i)$ in $\mathbf{m}'$ as

$$
\varepsilon F_z(s_i) \left[\rho_{N+1} - \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j(s_i)\rho_j \right]
$$

$$
= F_z(s_i)\varepsilon[\rho_{N+1} - \bar{\rho}_i]
$$

$$
= \left[\left. \frac{\partial F(z; \mathbf{s}'; \widehat{n'})}{\partial z} \right|_{z=s_i} + \left. \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z \partial \mathbf{m}'} \right|_{z=s_i, \Delta\mathbf{m}'=\Delta\mathbf{m}'_{\mathrm{T}}} \Delta\mathbf{m}' \right] \varepsilon[\rho_{N+1} - \bar{\rho}_i]
$$

$$
\geq \left[\widehat{F}_z(s_i) - C'_{\mathrm{Fzm}} \left| \Delta\mathbf{m}' \right| \right] \varepsilon[\rho_{N+1} - \bar{\rho}_i]
$$

$$
\geq \widehat{F}_z(s_i)\varepsilon(\rho_{N+1} - \bar{\rho}_i) - \varepsilon C'_{\mathrm{Fzm}} \left| \Delta\mathbf{m}' \right|, \tag{K.7}
$$

where $\bar{\rho} := \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j(s_i)\rho_j$ is the average value for ρ_j of this cluster when the mutant s_{N+1} is removed, satisfying $|\rho_{N+1} - \bar{\rho}_i| \leq 1$, and C'_{Fzm} is calculated in the next subsection. Substituting Eqs. (K.4), (K.5), and (K.7) into Eq. (K.2) gives

$$
\frac{\mathrm{d}p_{N+1}}{\mathrm{d}t} \geq p_{N+1}(1 - p_{N+1}) \left[\widehat{F}_z(s_i)\varepsilon(\rho_{N+1} - \bar{\rho}_i) - \varepsilon C'_{\mathrm{Fzm}} \left| \Delta\mathbf{m}' \right| - \varepsilon^2 C_{\mathrm{Fzz}} \right]
$$

$$
= p_{N+1}(1 - p_{N+1}) \left[\widehat{F}_z(s_i)(s_{N+1} - \bar{s}_i) - \varepsilon C'_{\mathrm{Fzm}} \left| \Delta\mathbf{m}' \right| - \varepsilon^2 C_{\mathrm{Fzz}} \right], \tag{K.8}
$$

where $\bar{s}_i = \varepsilon\bar{\rho}_i + s_i = \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j(\varepsilon\rho_j + s_i) = \sum_{j \in \text{com}(i), j \neq N+1} \tilde{p}_j s_j$ is the average trait value of this cluster when the mutant s_{N+1} is removed. For convenience, we assume that the representative phenotype is the most similar phenotype to the mutant, i.e., $|s_{N+1} - s_i| = \min_{k \in \text{com}(i)}(|s_{N+1} - s_k|)$. Notice that $|s_{N+1} - s_i| \leq |s_{N+1} - \bar{s}_i|$. Then, by exploiting $\left| \Delta\mathbf{m}' \right| \leq \varepsilon C'_{\mathbf{m}}$, a sufficient condition for Eq. (K.8) being always positive is given by

$$
\widehat{F}_z(s_i)(s_{N+1} - s_i) > \varepsilon^2 \left[C'_{\mathrm{Fzm}} C'_{\mathbf{m}} + C_{\mathrm{Fzz}} \right]. \tag{K.9}
$$

K.2 Finding C'_{Fzm}

From Eq. (A.7), we find for $j = 1, \ldots, M$

$$
\left| \left. \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial m_j \partial z} \right|_{z=s_{j\mathrm{T}}} \right| = \left| \left[\sum_{k=1}^{N+1} \frac{\partial n_k}{\partial m_j} \left[\frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial n_k \partial z} \right] \right]_{z=s_{j\mathrm{T}}} \right| = \left| \left. \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial n_j \partial z} \right|_{z=s_{j\mathrm{T}}} \right| \leq C'_{\mathrm{Fzn}}, \tag{K.10}
$$

for any $s_{jT} \in [s_j, s_{\text{cid}(j)}]$ with

$$C'_{\text{Fzn}} := \max\left\{ \left| \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z \partial n_j} \right|_{z=s_{jT}} \right| \; j = 1, \ldots, N+1, \; \mathbf{n}' \in [0, \eta]^{N+1}, \; s_{jT} \in [s_{\text{cid}(j)}, s_j] \right\}. \tag{K.11}$$

From Eq. (A.8) and Taylor's theorem, we find for $j = M+1, \ldots, N+1$ and for any $s_{jT} \in [s_j, s_{\text{cid}(j)}]$

$$\left| \left[\frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial m_j \partial z} \right]_{z=s_{jT}} \right| = \frac{1}{\varepsilon} \left| \left[\frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial n_j \partial z} - \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial n_{\text{cid}(j)} \partial z} \right]_{z=s_{jT}} \right|$$

$$= \frac{1}{\varepsilon} \left| \left[\frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial n_j \partial z} - \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial n_{\text{cid}(j)} \partial z} \right]_{\substack{z=s_{jT}, \\ s_j=s_{\text{cid}(j)}}} \right.$$

$$\left. + \left[\frac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j \partial z} - \frac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)} \partial z} \right]_{\substack{z=s_{jT}, \\ s_j=s_{jTT}}} (s_j - s_{\text{cid}(j)}) \right|$$

$$= \left| \rho_j \left[\frac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_j \partial z} - \frac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial s_j \partial n_{\text{cid}(j)} \partial z} \right]_{\substack{z=s_{jT}, \\ s_j=s_{jTT}}} \right| \leq C'_{\text{Fzsn}} \tag{K.12}$$

with some appropriately chosen $s_{jTT} \in [s_j, s_{\text{cid}(j)}]$ and

$$C'_{\text{Fzsn}} = \max\left\{ \left| \frac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z \partial s_j \partial n_j} - \frac{\partial^3 F(z; \mathbf{s}'; \mathbf{n}')}{\partial z \partial s_j \partial n_{\text{cid}(j)}} \right|_{z=s_{jT}, s_j=s_{jTT}} \right| \begin{array}{l} j = 1, \ldots, N+1, \; \mathbf{n}' \in [0, \eta]^{N+1}, \\ s_{jT} \in [s_{\text{cid}(j)}, s_j], s_{jTT} \in [s_{\text{cid}(j)}, s_j] \end{array} \right\}. \tag{K.13}$$

From Eqs. (K.10) and (K.12), we find

$$\left| \frac{\partial^2 F(z; \mathbf{s}'; \mathbf{n}')}{\partial \mathbf{m}' \partial z} \right|_{z=s_{jT}} = \left| \begin{pmatrix} \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial m_1 \partial z} \\ \cdots \\ \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial m_M \partial z} \\ \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial m_{M+1} \partial z} \\ \cdots \\ \frac{\partial^2 F(s_i; \mathbf{s}'; \mathbf{n}')}{\partial m_{N+1} \partial z} \end{pmatrix}_{z=s_{jT}} \right|$$

$$\leq \left| \begin{pmatrix} C'_{\text{Fzn}} \\ \cdots \\ C'_{\text{Fzn}} \\ C'_{\text{Fzsn}} \\ \cdots \\ C'_{\text{Fzsn}} \end{pmatrix} \right| = \sqrt{M C'^2_{\text{Fzn}} + (N+1-M) C'^2_{\text{Fzsn}}} := C'_{\text{Fzm}}. \tag{K.14}$$

Appendix L: Derivation of functional response and error estimates in Sect. 7

L.1 Functional response

For an arbitrary resource distribution expressed along a resource-quality axis w (size, hardness, toxicity, nutrient composition, etc.), incorporation of interference competition into Holling's disc equation (Holling 1959) gives a generalized Beddington–deAngelis-type functional response,

$$g(s_i; \mathbf{s}; \mathbf{n}) = \frac{\int R(w)c(w, s_i)\mathrm{d}w}{\zeta_1 + \zeta_2 \int R(w)c(w, s_i)\mathrm{d}w + \zeta_3 \int C(w)c(w, s_i)\mathrm{d}w}, \quad \text{(L.1)}$$

(Ito et al. 2009; Ito and Dieckmann 2012), where $R(w)$ is the resource distribution and $c(w, s_i)$ describes the niche of a phenotype s_i, in the form of the individual consumption effort as a function of the resource quality w. $C(w)$ is the consumption-effort distribution along the resource-quality axis w invested by all existing phenotypes,

$$C(w) = \sum_{i=1}^{N} n_i c(w, s_i). \quad \text{(L.2)}$$

Notice that the derivation of Eq. (L.1) in Ito et al. (2009) implicitly assumes that an individual searching resources engages in interference competition not only with the other individuals searching resources but also with individuals handling resources. Even when interference competition occurs only among individuals searching resources, a functional response can be derived by applying the "Holling square argument" (Heesterbeek and Metz 1993), but in a different form from Eq. (L.1). Eq. (L.1) can be transformed into

$$g(s_i; \mathbf{s}; \mathbf{n}) = \frac{\theta(s_i)}{\zeta_1 + \zeta_2\theta(s_i) + \zeta_3 \sum_{j=1}^{N} n_j \alpha(s_j, s_i)},$$

$$\alpha(s_j, s_i) = \int c(w, s_j)c(w, s_i)\mathrm{d}w,$$

$$\theta(s_i) = \int R(w)c(w, s_i)\mathrm{d}w. \quad \text{(L.3)}$$

L.2 Derivation of η, C_h, and C_r

As for η, we see from Eqs. (7.1) and (7.2) in the main text that for any i

$$\frac{1}{n_i}\frac{\mathrm{d}n_i}{\mathrm{d}t} = \frac{\beta\theta(s_i)}{\zeta_1 + \zeta_2\theta(s_i) + \sum_{j=1}^{2} n_j\alpha(s_j, s_i)} - 1$$

$$\leq \frac{\beta\theta(s_i)}{\zeta_1 + \zeta_2\theta(s_i) + n_i} - 1, \quad \text{(L.4)}$$

which is negative whenever n_i satisfies

$$n_i > [\beta - \zeta_2]C_\theta - \zeta_1, \tag{L.5}$$

with $C_\theta = \max\{\theta(s)|s \in [s_1, s_2]\}$. Thus, we find

$$\eta = [\beta - \zeta_2]C_\theta - \zeta_1. \tag{L.6}$$

As for $C_{\mathbf{r}}$, we see from Table 2 that

$$C_{\mathbf{r}} = \|\mathbf{P}\| \left\|\mathbf{P}^{-1}\right\|^2 \left[\|\mathbf{B}\| + \eta\sqrt{M}C_{\mathrm{Fmm}}\right]$$
$$= \frac{1}{\beta\theta(s_1)} + \eta C_{\mathrm{Fmm}}, \tag{L.7}$$

since $\mathbf{P} = 1$, $M = 1$, and $\mathbf{B} = b_{11} = -1/[\beta\theta(s_1)]$. In addition, we see from Table 2 that

$$C_{\mathrm{Fmm}} = \max\left\{ \left.\left\|\frac{\partial^2 F(s_i; \mathbf{s}_{\mathrm{a}}; \mathbf{m})}{\partial\mathbf{m}\partial\mathbf{m}^{\mathrm{T}}}\right\|_Q \right| i = 1, \ldots, M, \ \mathbf{m} \in [0, \eta]^M \right\}$$
$$= \max\left\{ \left.\left\|\frac{\partial^2 F(s_1; s_1; n_1)}{\partial n_1^2}\right\| \right| n_1 \in [0, \eta] \right\}$$
$$= \max\left\{ \left.\left\|\frac{2\beta\theta(s_1)}{[\zeta_1 + \zeta_2\theta(s_1) + n_1]^3}\right\| \right| n_1 \in [0, \eta] \right\}$$
$$= \frac{2\beta\theta(s_1)}{[\zeta_1 + \zeta_2\theta(s_1)]^3}, \tag{L.8}$$

which upon substitution into Eq. (L.7) gives

$$C_{\mathbf{r}} = \frac{1}{\beta\theta(s_1)} + \eta\frac{2\beta\theta(s_1)}{[\zeta_1 + \zeta_2\theta(s_1)]^3}. \tag{L.9}$$

As for $C_{\mathbf{h}}$, we see from Table 2 that

$$C_{\mathbf{h}} = \|\mathbf{P}\|C_{\mathbf{hm}} = \|\mathbf{P}\|\eta\sqrt{M}\left[\widehat{C'_{\mathrm{Fz}}} + 2C'_{\mathrm{F}\varepsilon}\right]$$
$$= \eta\left[\widehat{C'_{\mathrm{Fz}}} + 2C'_{\mathrm{F}\varepsilon}\right]. \tag{L.10}$$

For $\widehat{C'_{\mathrm{Fz}}}$, we see from Table 2 that

$$
\widehat{C'_{\mathrm{Fz}}} = \max \left\{ \left. \left| \frac{\partial F(z_j; \mathbf{s}'; \widehat{n'})}{\partial z_j} \right| \right|_{z_j = s_{j\mathrm{T}}} \; \middle| \; j = M+1, \ldots, N+1, \; s_{j\mathrm{T}} \in [s_j, s_{\mathrm{cid}(j)}] \right\}
$$

$$
= \max \left\{ \left. \left| \frac{\partial F(z_2; \mathbf{s}'; \widehat{n'})}{\partial z_2} \right| \right|_{z_2 \in [s_1, s_2]} \right\}
$$

$$
= \max \left\{ \left. \left| \frac{\partial}{\partial z_2} \left[\frac{\beta \theta(z_2)}{\zeta_1 + \zeta_2 \theta(s_1) + \widehat{n}_1 \alpha(z_2, s_1)} - 1 \right] \right| \right|_{z_2 \in [s_1, s_2]} \right\}
$$

$$
= \max \left\{ \left. \left| \frac{\beta \frac{\mathrm{d}\theta(z_2)}{\mathrm{d}z_2}}{\zeta_1 + \zeta_2 \theta(s_1) + \widehat{n}_1 \alpha(z_2, s_1)} + \frac{\beta \theta(z_2)[-\widehat{n}_1(z_2 - s_1)\alpha(z_2, s_1)]}{[\zeta_1 + \zeta_2 \theta(s_1) + \widehat{n}_1 \alpha(z_2, s_1)]^2} \right| \right|_{z_2 \in [s_1, s_2]} \right\}
$$

$$
\leq \max \left\{ \left. \left| \frac{\beta \frac{\mathrm{d}\theta(z_2)}{\mathrm{d}z_2}}{\zeta_1 + \zeta_2 \theta(s_1)} \right| \right|_{z_2 \in [s_1, s_2]} + \left. \left| \frac{\beta \theta(z_2)\varepsilon}{\zeta_1 + \zeta_2 \theta(s_1)} \right| \right|_{z_2 \in [s_1, s_2]} \right\}
$$

$$
\leq \frac{\beta C_{\partial\theta}}{\zeta_1 + \zeta_2 \theta(s_1)} + \frac{\beta C_\theta \varepsilon}{\zeta_1 + \zeta_2 \theta(s_1)}, \tag{L.11}
$$

with $C_\theta = \max\{\theta(s) | s \in [s_1, s_2]\}$ and $C_{\partial\theta} = \max\{\mathrm{d}\theta(s)/\mathrm{d}s | s \in [s_1, s_2]\}$. For $C'_{\mathrm{F}\varepsilon}$, we see from Table 2 that

$$
C'_{\mathrm{F}\varepsilon} = \max \left\{ \left. \left| \frac{\partial F(s_{\mathrm{cid}(j)} + \varepsilon\rho_j; \mathbf{s}'_\mathrm{a} + \varepsilon\boldsymbol{\rho}'; \mathbf{n}')}{\partial \varepsilon} \right| \right|_{\varepsilon = \varepsilon_{\mathrm{T}j}} \; \middle| \; \begin{array}{l} j = 1, \ldots, N+1, \\ \mathbf{n}' \in [0, \eta]^{N+1}, \\ \varepsilon_{\mathrm{T}j} \in [0, \varepsilon] \end{array} \right\}
$$

$$
= \max \left\{ \left. \left| \frac{\partial F(s_{\mathrm{cid}(j)} + \varepsilon\rho_j; (s_1, s_1 + \varepsilon)^{\mathrm{T}}; (n_1, n_2)^{\mathrm{T}})}{\partial \varepsilon} \right| \right|_{\varepsilon = \varepsilon_{\mathrm{T}}} \; \middle| \; \begin{array}{l} j = 1, 2, \\ 0 \leq n_1 \leq \eta, \; 0 \leq n_2 \leq \eta, \\ \varepsilon_{\mathrm{T}} \in [0, \varepsilon] \end{array} \right\},
$$

$$\tag{L.12}$$

where for $j = 1$ we see from Eqs. (7.4a) in the main text that

$$
\left| \frac{\partial F(s_1; (s_1, s_1 + \varepsilon)^{\mathrm{T}}; (n_1, n_2)^{\mathrm{T}})}{\partial \varepsilon} \right|
$$

$$
= \left| \frac{\partial}{\partial \varepsilon} \left[\frac{\beta \theta(s_1)}{\zeta_1 + \zeta_2 \theta(s_1) + n_1 + n_2 \alpha(s_1 + \varepsilon, s_1)} \right] \right|
$$

$$
= \frac{\beta \theta(s_1)\varepsilon n_2 \alpha(s_1 + \varepsilon, s_1)}{[\zeta_1 + \zeta_2 \theta(s_1) + n_1 + n_2 \alpha(s_1 + \varepsilon, s_1)]^2}
$$

$$
\leq \frac{\beta \theta(s_1)\varepsilon}{\zeta_1 + \zeta_2 \theta(s_1) + n_1 + n_2 \alpha(s_1 + \varepsilon, s_1)}
$$

$$
\leq \frac{\varepsilon \beta C_\theta}{\zeta_1 + \zeta_2 C_{\theta \min}} \tag{L.13}
$$

with $C_{\theta\min} = \min\{\theta(s)|s \in [s_1, s_2]\}$, and for $j = 2$ we see analogously

$$
\begin{aligned}
&\left| \frac{\partial F(s_1 + \varepsilon; (s_1, s_1 + \varepsilon)^{\mathrm{T}}; (n_1, n_2)^{\mathrm{T}})}{\partial \varepsilon} \right| \\
&= \left| \frac{\partial}{\partial \varepsilon} \left[\frac{\beta}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon) + n_1 \alpha(s_1 + \varepsilon, s_1) + n_2} \right] \right| \\
&\le \left[\frac{\beta}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon) + n_1 \alpha(s_1 + \varepsilon, s_1) + n_2} + \frac{\beta\theta(s_1 + \varepsilon)\zeta_2}{[\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon) + n_1 \alpha(s_1 + \varepsilon, s_1) + n_2]^2} \right] \left| \frac{\partial\theta(s_1 + \varepsilon)}{\partial \varepsilon} \right| \\
&\quad + \frac{\beta\theta(s_1 + \varepsilon)n_1\varepsilon\alpha(s_1 + \varepsilon, s_1)}{[\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon) + n_1 \alpha(s_1 + \varepsilon, s_1) + n_2]^2} \\
&\le \left[\frac{\beta}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon) + n_1 \alpha(s_1 + \varepsilon, s_1) + n_2} + \frac{\beta}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon) + n_1 \alpha(s_1 + \varepsilon, s_1) + n_2} \right] C_{\partial\theta} \\
&\quad + \frac{\beta\theta(s_1 + \varepsilon)\varepsilon}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon) + n_1 \alpha(s_1 + \varepsilon, s_1) + n_2} \\
&\le \frac{\beta}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon)} C_{\partial\theta} + \frac{\beta}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon)} C_{\partial\theta} \\
&\quad + \frac{\beta\theta(s_1 + \varepsilon)\varepsilon}{\zeta_1 + \zeta_2 \theta(s_1 + \varepsilon)} \le \frac{\beta[2C_{\partial\theta} + \varepsilon C_\theta]}{\zeta_1 + \zeta_2 C_{\theta\min}}
\end{aligned}
\tag{L.14}
$$

Combining Eqs. (L.6), (L.11), (L.12), (L.13), and (L.14), we find

$$
\begin{aligned}
C_{\mathbf{h}} &= \eta \left[\widehat{C'_{\mathrm{Fz}}} + 2C'_{\mathrm{F}\varepsilon} \right] \\
&\le \eta \left[\frac{\beta C_{\partial\theta}}{\zeta_1 + \zeta_2 \theta(s_1)} + \frac{\beta C_\theta \varepsilon}{\zeta_1 + \zeta_2 \theta(s_1)} + 2\max\left\{ \frac{\varepsilon\beta C_\theta}{\zeta_1 + \zeta_2 C_{\theta\min}}, \frac{\beta[2C_{\partial\theta} + \varepsilon C_\theta]}{\zeta_1 + \zeta_2 C_{\theta\min}}, \right\} \right] \\
&\le \eta \left[\frac{\beta C_{\partial\theta} + \beta C_\theta \varepsilon}{\zeta_1 + \zeta_2 C_{\theta\min}} + \frac{2\beta[2C_{\partial\theta} + \varepsilon C_\theta]}{\zeta_1 + \zeta_2 C_{\theta\min}} \right] \\
&\le \frac{\eta\beta[5C_{\partial\theta} + 3\varepsilon C_\theta]}{\zeta_1 + \zeta_2 C_{\theta\min}}
\end{aligned}
\tag{L.15}
$$

with $C_\theta = \max\{\theta(s)|s \in [s_1, s_2]\}$, $C_{\theta\min} = \min\{\theta(s)|s \in [s_1, s_2]\}$, and $C_{\partial\theta} = \max\{d\theta(s)/ds|s \in [s_1, s_2]\}$.

References

Ackermann M, Doebeli M (2004) Evolution of niche width and adaptive diversification. Evolution 58:2599–2612

Beddington JR (1975) Mutual interference between parasites or predators and its effect on searching efficiency. J Anim Ecol 44:331–340

Brown JS, Vincent TL (1987) Coevolution as an evolutionary game. Evolution 41:66–79

Cohen Y, Vincent TL, Brown JS (1999) A G-function approach to fitness minima, fitness maxima, evolutionarily stable strategies and adaptive landscapes. Evol Ecol Res 1:923–942

DeAngelis DL, Goldstein RA, O'Neill RV (1975) A model for trophic interaction. Ecology 56:881–892

Dercole F (2016) The ecology of asexual pairwise interactions: the generalized law of mass action. Theor Ecol 9:299–321

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

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. Nature 400:354–357

Dieckmann U, Law R (1996) The dynamical theory of coevolution: a derivation from stochastic ecological processes. J Math Biol 34:579–612

Dieckmann U, Marrow P, Law R (1995) Evolutionary cycling in predator-prey interactions: population dynamics and the Red Queen. J Theor Biol 176:91–102

Dieckmann U, Metz JAJ, Doebeli M, Tautz D (2004) Adaptive speciation. Cambridge University Press, Cambridge

Dieckmann U, Brännström Å, HilleRisLambers R, Ito HC (2007) The adaptive dynamics of community structure. In: Takeuchi Y, Sato K, Iwasa Y (eds) Mathematics for ecology and environmental sciences. Springer, Berlin, pp 145–177

Doebeli M, Dieckmann U (2000) Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am Nat 156:S77–S101

Durinx M, Metz JAJ, Meszéna G (2008) Adaptive dynamics for physiologically structured population models. J Math Biol 56:673–742

Egas M, Sabelis MW, Dieckmann U (2005) Evolution of specialization and ecological character displacement of herbivores along a gradient of plant quality. Evolution 59:507–520

Geritz SAH (2005) Resident-invader dynamics and the coexistence of similar strategies. J Math Biol 50:67–82

Geritz SAH, Metz JAJ, Kisdi É, Meszéna G (1997) Dynamics of adaptation and evolutionary branching. Phys Rev Lett 78:2024–2027

Geritz SAH, Kisdi É, Meszéna G, Metz JAJ (1998) Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol Ecol 12:35–57

Geritz SAH, Gyllenberg M, Jacobs FJA, Parvinen K (2002) Invasion dynamics and attractor inheritance. J Math Biol 44:548–560

Geritz SAH, Metz JAJ, Rueffler C (2016) Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. J Math Biol 72:1081–1099

Gyllenberg M, Parvinen K (2001) Necessary and sufficient conditions for evolutionary suicide. Bull Math Biol 63:981–993

Heesterbeek JAP, Metz JAJ (1993) The saturating contact rate in marriage-and epidemic models. J Math Biol 31:529–539

Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University Press, Cambridge

Holling CS (1959) Some characteristics of simple types of predation and parasitism. Can Entomol 91:385–398

Ito HC, Dieckmann U (2007) A new mechanism for recurrent adaptive radiations. Am Nat 170:E96–E111

Ito HC, Dieckmann U (2012) Evolutionary branching lines and areas in bivariate trait spaces. Evol Ecol Res 14:555–582

Ito HC, Dieckmann U (2014) Evolutionary branching under slow directional evolution. J Theor Biol 360:290–314

Ito HC, Shimada M, Ikegami T (2009) Coevolutionary dynamics of adaptive radiation for food-web development. Popul Ecol 51:65–81

Kisdi É, Jacobs FJA, Geritz SAH (2001) Red Queen evolution by cycles of evolutionary branching and extinction. Selection 2:161–176

Lande R (1979) Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 33:402–416

Leimar O (2009) Multidimensional convergence stability. Evol Ecol Res 11:191–208

Loeuille N, Loreau M (2005) Evolutionary emergence of size-structured food webs. PNAS 102:5761–5766

Mazenc F (2001) Differentiable Lyapunov function and center manifold theory. In: Isidori A (ed) Nonlinear control in the year 2000, volume 2, Lecture notes in control and information sciences, vol 259. Springer, London, pp 143–148

Meszéna G, Gyllenberg M, Jacobs FJ, Metz JAJ (2005) Link between population dynamics and dynamics of Darwinian evolution. Phys Rev Lett 95:078105

Metz JAJ, Nisbet RM, Geritz SAH (1992) How should we define 'fitness' for general ecological scenarios? Trends Ecol Evol 7:198–202

Metz JAJ, Geritz SAH, Meszéna G, Jacobs FJA, van Heerwaarden JS (1996) Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. In: van Strien SJ, Verduyn-Lunel SM (eds) Stochastic and spatial structures of dynamical systems. North Holland, Amsterdam, pp 83–231

Morris HW, Smale S, Devaney LR (2003) Differential equations, dynamical systems, and an introduction to chaos, 2nd edn. Academic Press, Waltham

Ravigné V, Dieckmann U, Olivieri I (2009) Live where you thrive: joint evolution of habitat choice and local adaptation facilitates specialization and promotes diversity. Am Nat 174:E141–E169

Rundle HD, Nosil R (2004) Ecological speciation. Ecol Lett 8:336–352

Takahashi D, Brännström Å, Mazzucco R, Yamauchi A, Dieckmann U (2013) Abrupt community transitions and cyclic evolutionary dynamics in complex food webs. J Theor Biol 337:181–189

Vukics A, Asboth J, Meszéna G (2003) Speciation in multidimensional evolutionary space. Phys Rev E 68:041903

Zeeman ML (1993) Hopf bifurcations in competitive three-dimensional Lotka–Volterra systems. Dyn Stab Syst 8:189–217