The components of directional and disruptive selection in heterogeneous group-structured populations

Hisashi Ohtsuki[1,*], Claus Rueffler[2], Joe Yuichiro Wakano[3,4,5], Kalle Parvinen[6,7] & Laurent Lehmann[8]

[1]Department of Evolutionary Studies of Biosystems, School of Advanced Sciences, SOKENDAI (The Graduate University for Advanced Studies), Shonan Village, Hayama, Kanagawa 240-0193, Japan

[2]Animal Ecology, Department of Ecology and Genetics, University of Uppsala, Sweden

[3]School of Interdisciplinary Mathematical Sciences, Meiji University, Tokyo 164-8525, Japan

[4]Meiji Institute for Advanced Study of Mathematical Sciences, Tokyo 164-8525, Japan

[5]Department of Mathematics, The University of British Columbia, 1984 Mathematics Road, Vancouver, B.C. Canada, V6T 1Z2.

[6]Department of Mathematics and Statistics, FI-20014 University of Turku, Finland

[7]Evolution and Ecology Program, International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria

[8]Department of Ecology and Evolution, University of Lausanne, Switzerland.

*Corresponding Author (email: ohtsuki_hisashi@soken.ac.jp)

keywords: stabilizing selection, disruptive selection, kin selection, evolutionary dynamics

Abstract

We derive how directional and disruptive selection operate on scalar traits in a heterogeneous group-structured population for a general class of models. In particular, we assume that each group in the population can be in one of a finite number of states, where states can affect group size and/or other environmental variables, at a given time. Using up to second-order perturbation expansions of the invasion fitness of a mutant allele, we derive expressions for the directional and disruptive selection coefficients, which are sufficient to classify the singular strategies of adaptive dynamics. These expressions include first- and second-order perturbations of individual fitness (expected number of settled offspring produced by an individual, possibly including self through survival); the first-order perturbation of the stationary distribution of mutants (derived here explicitly for the first time); the first-order perturbation of pairwise relatedness; and reproductive values, pairwise and three-way relatedness, and stationary distribution of mutants, each evaluated under neutrality. We introduce the concept of individual k-fitness (defined as the expected number of settled offspring of an individual for which $k-1$ randomly chosen neighbors are lineage members) and show its usefulness for calculating relatedness and its perturbation. We then demonstrate that the directional and disruptive selection coefficients can be expressed in terms individual k-fitnesses with $k = 1, 2, 3$ only. This representation has two important benefits. First, it allows for a significant reduction in the dimensions of the system of equations describing the mutant dynamics that needs to be solved to evaluate explicitly the two selection coefficients. Second, it leads

to a biologically meaningful interpretation of their components. As an application of our methodology, we analyze directional and disruptive selection in a lottery model with either hard or soft selection and show that many previous results about selection in group-structured populations can be reproduced as special cases of our model.

1 Introduction

Many natural populations are both group-structured – with the number of individuals interacting at the local scale being finite – and heterogeneous – with different groups being subject to different demographic and environmental conditions (*e.g.*, varying group size and temperature, respectively). Understanding how evolution, and in particular natural selection, moulds phenotypic traits in such systems is complicated as both local heterogeneity and demographic stochasticity need to be taken into account. In order to predict the outcome of evolution in heterogeneous populations, evolutionists are generally left with the necessity to approximate the evolutionary dynamics, as a full understanding of this process is yet out of reach.

A standard approximation to predict evolutionary outcomes is to assume that traits are quantitative, that the details of inheritance do not matter ("phenotypic gambit", Grafen, 1991), and that mutations have weak (small) phenotypic effects (*e.g.* Grafen, 1985; Taylor, 1989; Parker and Maynard Smith, 1990; Rousset, 2004). Under these assumptions, directional trait evolution can be quantified by a phenotypic selection gradient that captures first-order effects of selection. Thus, phenotypic change occurs in an uphill direction on the fitness landscape. This directional selection either causes the trait value to change endlessly (for instance, due to macro environmental changes or cycles in the evolutionary dynamics), or the trait value eventually approaches a local equilibrium point, a so-called *singular strategy*, where directional selection vanishes. Such a singular strategy may be locally uninvadable ("evolutionary stable") and thus a local end-point of the evolutionary dynamics. However, when the fitness landscape is dynamic due to selection being frequency-dependent, then it is also possible that, as the population evolves uphill on the fitness landscape, this landscape changes such that the population eventually finds itself at a singular strategy that is located in a fitness valley. In this case, directional selection turns into disruptive selection, which means that a singular strategy that is an attractor of the evolutionary dynamics (and thus convergence stable) is invadable by nearby mutants and thus an *evolutionary branching point* (Metz et al., 1996; Geritz et al., 1998). Further evolutionary dynamics can then result in genetic polymorphism in the population, thus possibly favoring the maintenance of adaptive diversity in the long term (see Rueffler et al., 2006, for a review). Disruptive selection at a singular point is quantified by the disruptive selection coefficient (called quadratic selection gradient in the older literature: Lande and Arnold, 1983; Phillips and Arnold, 1989), which involves second-order effects of selection.

A central question concerns the nature and interpretation of the components of the selection gradient and the disruptive selection coefficient on a quantitative trait in heterogeneous populations. For the selection gradient, this question has been studied for a long time and a general answer has been given under the assumption that individuals can be in a finite number of states (summarized in Rousset, 2004). Then, regardless of the complexity of the spatial, demographic, environmental, or physiological states

individuals can be in or experience (in the kin-selection literature commonly referred to as class-structure, *e.g.* Taylor, 1990; Frank, 1998; Rousset, 2004), the selection gradient on a quantitative trait depends on three key components (Taylor, 1990; Frank, 1998; Rousset, 2004). The first component are individual fitness differentials, which capture the marginal gains and losses of producing offspring in particular states to parents in particular states. The second component are (neutral) reproductive values weighting these fitness differentials. These capture the fact that offspring settling in different states contribute differently to the future gene pool. The third component are (neutral) relatedness coefficients. These also weight the fitness differentials, and capture the fact that some pairs of individuals are more likely to carry the same phenotype (inherited from a common ancestor) than randomly sampled individuals. This results in correlations between the trait values of interacting individuals. Such correlations matter for selection ("kin selection", *e.g.* Michod, 1982) and occur in populations subject to limited genetic mixing and small local interaction groups. At the risk of oversimplifying, reproductive values can be thought of as capturing the effect of population heterogeneity on directional selection, while relatedness captures the effect of demographic stochasticity under limited genetic mixing.

The situation is different with respect to the coefficient of disruptive selection, *i.e.*, the second-order effects of selection. The components of the disruptive selection coefficient have not been worked out in general and are studied only under the assumptions of well-mixed or spatially structured populations, but with otherwise homogeneous individuals. For the spatially structured case the effects of selection on relatedness has been shown to matter, as selection changes the number of individuals expressing similar trait values in a certain group (Ajar, 2003; Wakano and Lehmann, 2014; Mullon et al., 2016), resulting in a reduced strength of disruptive selection under limited dispersal. For the general case that individuals can be in different states one expects intuitively that disruptive selection also depends on how selection affects the distribution of individuals over the different states. But this has not been analyzed so far even though it is captured implicitly when second-order derivatives of invasion fitness are computed as has been done in several previous works investigating evolutionary branching in some specific models of class-structured populations (*e.g.* Massol et al., 2011; Rueffler et al., 2013; Massol and Débarre, 2015; Kisdi, 2016; Parvinen et al., 2018, 2020).

In the present paper, we develop an evolutionary model for a heterogeneous group-structured population that covers a large class of biological scenarios. For this model, we show that the disruptive selection coefficient can be expressed in terms of individual fitness differentials weighted by the neutral quantities appearing in the selection gradient. This both significantly facilitates concrete calculations under complex scenarios and allows for a biological interpretation of selection. Our results contain several previous models as special cases.

The remainder of this paper is organized as follows. (1) We start by describing a demographic model for a heterogeneous group-structured population and present some background material underlying the characterization of uninvadable ("evolutionary stable") strategies by way of invasion fitness for this model. We here also introduce a novel individual fitness concept – individual k-fitness – defined as the expected number of settled offspring of an individual for which $k-1$ randomly chosen neighbors are relatives (*i.e.*, members of the same lineage). This fitness concept plays a central role in our analysis. (2)

Assuming quantitative scalar traits, we present first- and second-order perturbations of invasion fitness (*i.e.*, the selection gradient and disruptive selection coefficient, respectively), discuss their components and the interpretations thereof, and finally express all quantities in terms of individual k-fitness with $k = 1, 2, 3$. (3) We present a generic lottery model under spatial heterogeneity for both soft and hard selection regimes and show that the selection gradient and the disruptive selection coefficient can be computed explicitly under any scenario falling into this class of models. We then apply these results to a concrete local adaptation scenario where we derive conditions for evolutionary stability and convergence stability of singular trait values, and show their dependence on migration rate and group size. In doing so, we recover and extend previous results from the literature and show how our model connects seemingly different approaches.

2 Model

2.1 Biological assumptions

We consider a population of haploid individuals that is subdivided into infinitely many groups that are connected to each other by dispersal (*i.e.*, the infinite island model). Dispersal between groups may occur by individuals alone or by groups of individuals as in propagule dispersal, but is always random with respect to the destination group . We consider a discrete-time reproductive process and thus discrete census steps. At each census, each group is in a state $s \in \mathcal{S}$ with $\mathcal{S} = \{s_1, s_2, ..., s_N\}$ where N denotes the number of possible states. The state s determines the number of individuals in a group and/or any environmental factor determining the survival, reproduction, and dispersal of all individuals within a group. For the sake of simplicity, we will consider only a finite number of discrete states in this paper. The state s does not need to be a fixed property of a group but can change in time and be affected by individual trait values and thus be determined endogenously. However, we assume that such state changes are governed by a time-homogeneous Markov chain, meaning that there are no temporal trends in dynamics of the group states. We denote by n_s the finite number of adult individuals in a group in state s, which can thus change over time if the group state changes. We assume that group size is bounded as a result of density dependence acting at the local scale (hence there is an upper bound on group size). The described set-up includes a variety of classical models.

1. Purely spatially structured populations: The state s is identical for all groups ($N = 1$) and so there is only one group size. This is essentially the island model as developed by Wright (1943), which has been a long-term work horse for understanding the effect of spatial structure on evolutionary dynamics (*e.g.* Eshel, 1972; Bulmer, 1986; Rousset, 2004).

2. Stochastic population dynamics at the group level: The state s determines the number of individuals in a group, which can potentially vary in time (*e.g.* Metz and Gyllenberg, 2001; Rousset and Ronce, 2004). This case covers the situation in which each group is embedded in a community consisting of several interacting species and where the state s determines the number of individuals for each of the other species (*e.g.* Chesson, 1981).

3. Environmental heterogeneity: The state s determines an aspect of the within-group environment, which affects the survival and/or reproduction of its group members. An example is heterogeneity in patch quality or size (*e.g.* Wild et al., 2009; Massol et al., 2011; Rodrigues and Gardner, 2012). We note that in the limit of infinite group size this coincides with models of temporal and spatial heterogeneity as reviewed in Svardal et al. (2015).

4. Group splitting: This is a special case in which migration between groups is in fact absent but groups can be connected to each other if they originate through splitting of a parental group. The state s again determines the number of adults in a group. This model is inspired by compartmentalized replication in prebiotic evolution (stochastic corrector model, Szathmary and Demeter, 1987; Grey et al., 1995).

5. Purely physiologically structured population: In the special case with only a single individual in a group, the state s can be taken to represent the physiological state of an individual such as age or size or combinations thereof (*e.g.* Ronce and Promislow, 2010). In the special case of complete and independent offspring dispersal (*i.e.*, no group dispersal) but arbitrary group size, the state s can be taken to represent the combination of individual physiological states of all members in a group so that the model covers within group heterogeneity.

Since we are mainly interested in natural selection driven by recurrent invasions by possibly different mutants, we can focus on the initial invasion of a mutant allele into a monomorphic resident population. Hence, we assume that at any time at most two alleles segregate in the population, a mutant allele whose carriers express the trait value x and a resident allele whose carriers express the trait value y. We furthermore assume that traits are one-dimensional and real-valued ($x, y \in \mathbb{R}$). Suppose that initially the population is monomorphic (*i.e.*, fixed) for the resident allele y and a single individual mutates to trait value x. How do we ascertain the extinction or spread of the mutant?

2.2 Multitype branching process and invasion fitness

Since any mutant is initially rare, we can focus on the initial invasion of the mutant into the total population and approximate its dynamics as a discrete-time multitype branching process (Harris, 1963; Karlin and Taylor, 1975; Wild, 2011). In doing so, we largely follow the model construction and notation used in Lehmann et al. (2016) (see section A in the Supplementary Material for a mathematical description of the stochastic process underlying our model). In particular, in order to ascertain uninvadability of mutants into a population of residents it is sufficient to focus on the transition matrix $\boldsymbol{A} = \{a(s', i'|s, i)\}$ whose entry in position $(s', i'; s, i)$, denoted by $a(s', i'|s, i)$, is the expected number of groups in state s' with $i' \geq 1$ mutant individuals that descend from a group in state s with $i \geq 1$ mutant individuals over one time step in a population that is otherwise monomorphic for y. In the following, we refer to a group in state s with i mutants and $n_s - i$ residents as an (s, i)-group for short. The transition matrix $\boldsymbol{A}$ is a square matrix that is assumed to be primitive (we note that primitivity will obtain under all models listed in section 2.1 but may be induced for different reasons). Thus, a positive integer ℓ (possibly depending on x and y) exists such that every entry of $\boldsymbol{A}^\ell$ (ℓth power of $\boldsymbol{A}$) is positive. The

entries $a(s', i'|s, i)$ of the matrix $\boldsymbol{A}$ generally depend on both x and y, but for ease of exposition we do not write these arguments explicitly unless necessary. The same convention applies to all other variables that can in principle depend on x and y.

From standard results on multitype branching processes (Harris, 1963; Karlin and Taylor, 1975) it follows that a mutant x arising as a single copy in an arbitrary group of the population, *i.e.*, in any $(s, 1)$-group, goes extinct with probability one if and only if the largest eigenvalue of $\boldsymbol{A}$, denoted by ρ, is less than or equal to one,

$$\rho \leq 1, \tag{1}$$

where ρ satisfies

$$\boldsymbol{A}\boldsymbol{u} = \rho \boldsymbol{u} \tag{2}$$

and where $\boldsymbol{u}$ is the leading right eigenvector of $\boldsymbol{A}$. We refer to ρ as the *invasion fitness* of the mutant. If eq.(1) holds, then we say that y is *uninvadable* by x. To better understand what determines invasion fitness, we introduce the concept of the *mutant lineage*, which we define as the collection of descendants of the initial mutant: its direct descendants (possibly including self through survival), the descendants of its immediate descendants, and so on. Invasion fitness then gives the expected number of mutant copies produced over one time step by a randomly sampled mutant from its lineage in an otherwise monomorphic resident population that has reached demographic stationarity (Mullon et al., 2016; Lehmann et al., 2016). The mutant stationary distribution is given by the vector $\boldsymbol{u}$ with entries $u(s, i)$ describing, after normalization, the asymptotic probability that a randomly sampled group containing at least one mutant is in state s and contains $i \geq 1$ mutants. In other words, invasion fitness is the expected number of mutant copies produced by a lineage member randomly sampled from the distribution $\boldsymbol{u}$ (see eq.(C8) in the Supplementary Material and the explanation thereafter).

2.3 Statistical description of the mutant lineage

We use the matrix $\boldsymbol{A} = \{a(s', i'|s, i)\}$ and its leading right eigenvector $\boldsymbol{u}$ to derive several quantities allowing us to obtain an explicit representation of invasion fitness, which will be the core of our sensitivity analysis.

2.3.1 Asymptotic probabilities and relatedness of k-individuals

We start by noting that the asymptotic probability for a mutant to find itself in an (s, i)-group is given by

$$q(s, i) \equiv \frac{iu(s, i)}{\sum_{s' \in \mathcal{S}} \sum_{i'=1}^{n_{s'}} i' u(s', i')}. \tag{3}$$

From this, we can compute two state probabilities. First, the asymptotic probability that a randomly sampled mutant finds itself in a group in state s is given by

$$q(s) \equiv \sum_{i=1}^{n_s} q(s, i). \tag{4}$$

Second, the asymptotic probability that, conditional on being sampled in a group in state s, a randomly sampled mutant finds itself in a group with i mutants is given by

$$q(i|s) \equiv \frac{q(s,i)}{q(s)}. \tag{5}$$

Let us further define

$$\phi_k(s,i) \equiv \begin{cases} 1 & (k=1) \\ \displaystyle\prod_{j=1}^{k-1} \frac{i-j}{n_s-j} & (2 \leq k \leq i) \\ 0 & (i+1 \leq k \leq n_s), \end{cases} \tag{6}$$

which, for $k > 1$, can be interpreted as the probability that, given a mutant is sampled from an (s,i)-group, $k-1$ randomly sampled group neighbors without replacement are all mutants. This allows us to define the relatedness between k individuals in a group in state s as

$$r_k(s) \equiv \sum_{i=1}^{n_s} \phi_k(s,i) q(i|s). \tag{7}$$

This is the probability that $k-1$ randomly sampled neighbors without replacement of a randomly sampled mutant in state s are also mutants (*i.e.*, they all descend from the lineage founder). For example,

$$r_2(s) = \sum_{i=1}^{n_s} \frac{i-1}{n_s-1} q(i|s) \tag{8}$$

is the asymptotic probability of sampling a mutant among the neighbors of a random mutant individual from a group in state s and thus provides a measure of pairwise relatedness among group members. Likewise,

$$r_3(s) = \sum_{i=1}^{n_s} \frac{(i-1)(i-2)}{(n_s-1)(n_s-2)} q(i|s) \tag{9}$$

is the asymptotic probability that, conditional on being sampled in a group in state s, two random neighbors of a random mutant individual are also mutants.

2.3.2 Individual fitness and individual k-fitness

Consider a mutant in an (s,i)-group and define

$$w(s'|s,i) \equiv \frac{1}{i} \sum_{i'=1}^{n_{s'}} i' a(s',i'|s,i). \tag{10}$$

The sum on the right-hand side of eq.(10) counts the expected total number of mutants in groups in state s' produced by an (s,i)-group, and the share from a single mutant in this (s,i)-group is calculated by dividing this lineage productivity by i. Hence, $w(s'|s,i)$ is the expected number of offspring of a mutant individual (possibly including self through survival), which settle in a group in state s', given that the mutant resided in a group in state (s,i) in the previous time period. Thus $w(s'|s,i)$ is an individual fitness[1].

[1] It is important to note that the conditioning in $w(s'|s,i)$ is only on the state of the parental generation (as emphasized by the notation) and that $w(s'|s,i)$ depends on group transition probabilities in models in which the state s of a group can change in each generation. See eqs.(E.1–E.2) in Lehmann et al. (2016) as well as section G.2 in the Supplementary Material for more details.

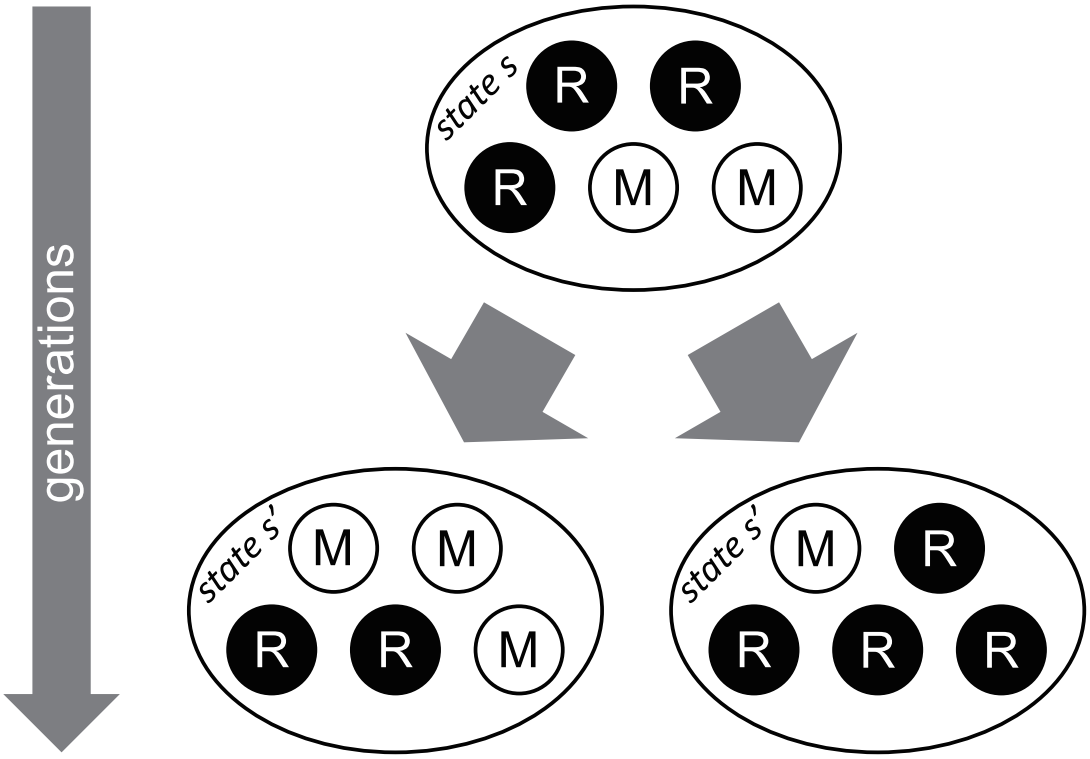

Figure 1: A schematic example for the calculation of individual k-fitness. Symbols M and R represent mutants and resident individuals, respectively. In this example, an $(s, 2)$-group "produced" one $(s', 3)$-group and one $(s', 1)$-group. Individual 1-fitness of each mutant in the parental generation is the total number of mutants in the following generation $(3 + 1 = 4)$ divided by the number of mutants in the $(s, 2)$-group $(= 2)$. Thus $w_1(s'|s, 2) = 4/2 = 2$. For individual 2-fitness we calculate the weighted number of mutants in the following generation, where the weights are the probabilities that a random neighbor of a mutant is also a mutant, and then divide it by the number of mutants in the $(s, 2)$-group $(= 2)$. These probabilities are $2/4$ for the $(s', 3)$-group and $0/4$ for the $(s', 1)$-group. Thus, the weighted number of mutants is $3 \cdot (2/4) + 1 \cdot (0/4) = 3/2$, and the individual 2-fitness is $w_2(s'|s, 2) = (3/2)/2 = 3/4$. Similarly, $w_3(s'|s, 2) = \{3 \cdot (1/6) + 1 \cdot (0/6)\}/2 = 1/4$ and $w_4(s'|s, 2) = w_5(s'|s, 2) = 0$.

We now extend the concept of individual fitness to consider a collection of offspring descending from a mutant individual. More formally, for any integer k $(1 \leq k \leq n_{s'})$ we let

$$w_k(s'|s, i) \equiv \frac{1}{i} \sum_{i'=1}^{n_{s'}} \phi_k(s', i') i' a(s', i'|s, i) \tag{11}$$

be the expected number of offspring produced by a single mutant individual in an (s, i)-group (possibly including self through survival) that settle in a group in state s' and have $k - 1$ randomly sampled group neighbors (without replacement) that are also mutants. We refer to $w_k(s'|s, i)$ as "individual k-fitness" regardless of the states s' and (s, i) (see Figure 1 for an illustrative example).

Note that individual 1-fitness equals $w(s'|s, i)$ as defined in eq.(10). Hence, individual k-fitness $w_k(s'|s, i)$ is a generalization of this fitness concept. The difference between eq.(10) and eq.(11) is the term $\phi_k(s', i')$, which shows that k-fitness counts an individual's number of offspring (possibly including self through survival) that experience a certain identity-by-descent genetic state in their group. Under our assumption of infinitely many groups, more than one dispersing offspring can settle in the same group only with propagule dispersal. Thus, without propagule dispersal dispersing offspring do not contribute

to k-fitness for $k > 1$.

2.3.3 Notation for perturbation analysis

Since our goal is to perform a sensitivity analysis of ρ to evaluate the selection gradient and disruptive selection coefficient, we assume that the mutant and resident trait values are close to each other and write

$$x = y + \delta \tag{12}$$

with δ sufficiently small (*i.e.*, $|\delta| \ll 1$). Thus, ρ can be Taylor-expanded with respect to δ.

For invasion fitness ρ, or more generally, for any smooth function F that depends on δ, we will use the following notation throughout this paper. The Taylor-expansion of F with respect to δ is written as

$$F(\delta) = F^{(0)} + \delta F^{(1)} + \delta^2 F^{(2)} + \cdots , \tag{13a}$$

where $F^{(\ell)}$ is given by

$$F^{(\ell)} = \frac{1}{\ell!} \frac{\mathrm{d}^\ell F(\delta)}{\mathrm{d}\delta^\ell}\bigg|_{\delta=0} . \tag{13b}$$

2.3.4 Properties of the monomorphic resident population

The zeroth-order coefficient in eq.(13) corresponds to the situation where the function F is evaluated under the supposition that individuals labelled as "mutant" and "resident" are the same. In that case, individuals in groups with the same state are assumed to be exchangeable in the sense that they have the same reproductive characteristics (the same distribution of fitnesses, *i.e.*, the same mean fitness, the same variance in fitness, and so on). This results in a neutral evolutionary process, *i.e.*, a monomorphic population.

We now characterize the mutant lineage dynamics under a neutral process as this plays a crucial role in our analysis. From eq.(10), the individual 1-fitness in an (s, i)-group, written under neutrality, equals

$$w_1^{(0)}(s'|s, i) = \frac{1}{i} \sum_{i'=1}^{n_{s'}} i' a^{(0)}(s', i'|s, i), \tag{14}$$

where each $a^{(0)}(s', i'|s, i)$ is an entry of the matrix $\boldsymbol{A}$ under neutrality. By our exchangeability assumption, eq.(14) does not depend on i, the number of the individuals labeled as "mutants" in this group (see section A.2 (iv) in the Supplementary Material). If this would not be the case, mutants in a group (s, i_1) and in a group (s, i_2) with $i_1 \neq i_2$ would have different reproductive outputs and mutants and residents would not be exchangeable. Therefore, from now on we write $w_1^{(0)}(s'|s, i)$ simply as $w_1^{(0)}(s'|s)$. We collect these neutral fitnesses in the $N \times N$ matrix $\boldsymbol{W}^{(0)} = \{w_1^{(0)}(s'|s)\}$. Its entry (s', s) gives the expected number of descendants (possibly including self through survival) settling in groups of state s' that descend from an individual residing in an s-group (mutant or resident since they are phenotypically indistinguishable).

The assumptions that each group is density regulated (see Section 2.1) and that the resident population has reached stationarity guarantee that the largest eigenvalue of $\boldsymbol{W}^{(0)}$ equals 1 (see section A.2 (v) in the Supplementary Material). This is the unique largest eigenvalue because $\boldsymbol{W}^{(0)}$ is primitive due to

the assumption that $\boldsymbol{A}$ is primitive. Thus, there is no demographic change in populations in which all individuals carry the same trait y and that have reached stationarity.

The fact that under neutrality $w_1^{(0)}(s'|s,i)$ is independent of i and $\boldsymbol{W}^{(0)}$ has the unique largest eigenvalue of 1 imposes constraints on the matrix $\boldsymbol{A}^{(0)} = \{a^{(0)}(s',i'|s,i)\}$ that describes the growth of a mutant lineage under neutrality. Let us denote the left eigenvector of $\boldsymbol{W}^{(0)}$ corresponding to the eigenvalue 1 by $\boldsymbol{v}^{(0)} = \{v^{(0)}(s)\}$, which is a strictly positive row vector of length N. Each entry $v^{(0)}(s)$ gives the reproductive value of an individual in state s, which is the asymptotic contribution of that individual to the gene pool. Note that $v^{(0)}(s)$ does not depend on δ because it is defined from $\boldsymbol{W}^{(0)}$, which is independent of δ. We now construct a row vector $\hat{\boldsymbol{v}}^{(0)} = \{\hat{v}^{(0)}(s,i)\}$ of length $n \equiv \sum_{s\in\mathcal{S}} n_s$ by setting $\hat{v}^{(0)}(s,i) = v^{(0)}(s)i$. It has been shown that $\hat{\boldsymbol{v}}^{(0)}$ is a positive left eigenvector of the matrix $\boldsymbol{A}^{(0)} = \{a^{(0)}(s',i'|s,i)\}$ corresponding to the eigenvalue 1, and therefore – since $\hat{\boldsymbol{v}}^{(0)}$ is strictly positive – the Perron-Frobenius theorem implies that the largest eigenvalue of $\boldsymbol{A}^{(0)}$ is $\rho^{(0)} = 1$ (see Appendix A in Lehmann et al., 2016, for a proof and more details). We also show that the column vector $\{q^{(0)}(s)\}$ of length N, denoting the stable asymptotic distribution given by eq.(4) under neutrality, is the right eigenvector of the matrix $\boldsymbol{W}^{(0)}$ corresponding to the eigenvalue of 1 (see section C.2.1 in the Supplementary Material). There is freedom of choice for how to normalize the left eigenvector $\boldsymbol{v}^{(0)}$ and here we employ the convention that $\sum_{s\in\mathcal{S}} v^{(0)}(s)q^{(0)}(s) = 1$. This means that the reproductive value of a randomly sampled mutant individual from its lineage is unity.

To summarize, under neutrality, the stable asymptotic distribution of mutants and the reproductive value of individuals satisfy

$$q^{(0)}(s') = \sum_{s\in\mathcal{S}} w_1^{(0)}(s'|s)q^{(0)}(s) \qquad (\boldsymbol{q}^{(0)} = \boldsymbol{W}^{(0)}\boldsymbol{q}^{(0)}), \tag{15a}$$

$$v^{(0)}(s) = \sum_{s'\in\mathcal{S}} v^{(0)}(s')w_1^{(0)}(s'|s) \qquad (\boldsymbol{v}^{(0)} = \boldsymbol{v}^{(0)}\boldsymbol{W}^{(0)}), \tag{15b}$$

$$1 = \sum_{s\in\mathcal{S}} v^{(0)}(s)q^{(0)}(s) \qquad (1 = \boldsymbol{v}^{(0)}\boldsymbol{q}^{(0)}), \tag{15c}$$

where $\boldsymbol{v}^{(0)}$ is a row-vector with entries $v^{(0)}(s)$ and $\boldsymbol{q}^{(0)}$ is a column-vector with entries $q^{(0)}(s)$.

2.4 Invasion fitness as reproductive-value-weighted fitness

Equation (2) for the leading eigenvalue and eigenvector of the matrix $\boldsymbol{A}$ can be left-multiplied on both sides by any non-zero vector of weights. This allows to express ρ in terms of this vector of weights and $\boldsymbol{A}$ and $\boldsymbol{u}$. If one chooses for the vector of weights the vector of neutral reproductive values $\hat{\boldsymbol{v}}^{(0)}$ discussed above, then invasion fitness can be expressed as

$$\rho = \frac{1}{V}\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s')w_1(s'|s,i)q(i|s)q(s), \tag{16a}$$

where

$$V \equiv \sum_{s\in\mathcal{S}} v^{(0)}(s)q(s) \tag{16b}$$

(see Lehmann et al., 2016, Appendix C, eq.(C.5), for the proof). This representation of ρ is useful to do concrete calculations. The intuition behind it is as follows. The inner sum, taken over i, represents

the reproductive-value-weighted average number of offspring in states s' given a parental mutant resides in an s-group, where the average is taken over all possible mutant numbers experienced by the parental mutant in an s-group. The middle sum takes the average over all states s in which mutants can reside in the parental generation, and the outer sum takes the average over all possible states s' in which mutant offspring can reside (possibly including parents through survival).

Hence, the numerator in eq.(16a) is the reproductive-value-weighted average individual 1-fitness of a mutant individual randomly sampled from the mutant lineage, while the denominator V can be interpreted (in force of eq.(15b)) as the reproductive-value-weighted average of the neutral 1-fitness of an individual sampled from the asymptotic state distribution of the mutant lineage. Hence, ρ is the ratio of the reproductive-value-weighted average fitness of a mutant individual and that of a mutant individual under neutrality where both individuals are sampled from the same distribution. Note that in eq.(16a) the quantities $w_1(s'|s,i)$, $q(s)$ and $q(i|s)$ depend on δ while $v^{(0)}(s')$ does not.

Our goal is to compute from eq.(16a) the selection gradient and disruptive selection coefficients,

$$\rho^{(1)}(y) \equiv \left.\frac{\partial \rho}{\partial \delta}\right|_{\delta=0} \quad \text{and} \quad \rho^{(2)}(y) \equiv \frac{1}{2}\left.\frac{\partial^2 \rho}{\partial \delta^2}\right|_{\delta=0}. \tag{17}$$

These coefficients are all we need to classify singular strategies (Metz et al., 1996; Geritz et al., 1998). Indeed, a singular strategy y^* satisfies

$$\rho^{(1)}(y^*) = 0. \tag{18}$$

This strategy is locally convergence stable (*i.e.*, a local attractor point of the evolutionary dynamics) when

$$c(y^*) \equiv \left.\frac{\mathrm{d}\rho^{(1)}(y)}{\mathrm{d}y}\right|_{y=y^*} < 0. \tag{19}$$

Note that convergence stability hinges on mutants with small phenotypic deviation δ invading and substituting residents ("invasion implies substitution"), which holds true when $|\delta| \ll 1$ under the demographic assumptions of our model (Rousset, 2004, pp. 196 and 206). Furthermore, the singular point is locally uninvadable if

$$\rho^{(2)}(y^*) < 0. \tag{20}$$

A singular strategy can then be classified by determining the combination of signs of the disruptive selection coefficient $\rho^{(2)}(y^*)$ and the convergence stability coefficient $c(y^*)$ at y^* (Metz et al., 1996; Geritz et al., 1998).

3 Sensitivity analysis

3.1 Eigenvalue perturbations

Using eq.(16a), as well as the normalization of reproductive values given in eq.(15c), we show in section B in the Supplementary Material that the first-order perturbation of ρ with respect to δ is given by

$$\rho^{(1)} = \sum_{s'\in\mathcal{S}} \sum_{s\in\mathcal{S}} \sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s). \tag{21}$$

Thus, $\rho^{(1)}$ is simply a weighted perturbation of individual 1-fitnesses w_1. For the second-order perturbation of ρ with respect to δ, given that $\rho^{(1)} = 0$, we find that

$$\rho^{(2)} = \rho^{(2\mathrm{w})} + \rho^{(2\mathrm{q})} + \rho^{(2\mathrm{r})} \tag{22a}$$

where

$$\rho^{(2\mathrm{w})} = \sum_{s'\in\mathcal{S}} \sum_{s\in\mathcal{S}} \sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(2)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s) \tag{22b}$$

$$\rho^{(2\mathrm{q})} = \sum_{s'\in\mathcal{S}} \sum_{s\in\mathcal{S}} \sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(1)}(s) \tag{22c}$$

$$\rho^{(2\mathrm{r})} = \sum_{s'\in\mathcal{S}} \sum_{s\in\mathcal{S}} \sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(1)}(s'|s,i) q^{(1)}(i|s) q^{(0)}(s) \tag{22d}$$

(section B in the Supplementary Material). The first term, labelled $\rho^{(2\mathrm{w})}$, comes from the second-order perturbation of individual 1-fitnesses. The second term, labelled $\rho^{(2\mathrm{q})}$, comes from the first-order perturbation of the stationary distribution of mutants in the different states, and the third term, labelled $\rho^{(2\mathrm{r})}$, comes from the first-order perturbation of the stationary distribution of the number of mutants in any given state.

While eqs.(21) and (22) give some insights into how selection acts on mutants, in particular, they emphasize the role of selection on the distributions $q(s)$ and $q(i|s)$, these expressions remain complicated as they involve weighted averages of fitness derivatives $w_k^{(\ell)}(s'|s,i)$ $(\ell = 1,2)$ over the neutral and perturbed mutant distributions $q^{(1)}(i|s)$ and $q^{(1)}(s)$. To obtain more insightful expressions for these sensitivities, we express in the next section $w_k(s'|s,i)$ for $k = 1,2,3$ in terms of trait values. This will allow us to carry out rearrangements and simplifications of $\rho^{(1)}$ and $\rho^{(2)}$.

3.2 Individual fitness functions

3.2.1 Individual 1-fitness

Consider a focal individual in a focal group in state s and denote by z_1 the trait value of that individual. Suppose that the other $n_s - 1$ neighbors adopt the trait values $z_2, \cdots, z_{n_s}$ and almost all individuals outside this focal group adopt the trait value z. Let then

$$w_{1,s'|s}(z_1, z_2, \cdots, z_{n_s}, z) \qquad (s', s \in \mathcal{S},\ z_1, \cdots, z_{n_s}, z \in \mathbb{R}) \tag{23a}$$

be the expected number of offspring in state s' that descend from a focal in state s. Equation (23a) expresses individual 1-fitness in terms of the phenotypes of all interacting individuals and will be referred to as an individual fitness function. It is a common building block of phenotypic models (see Frank, 1998; Rousset, 2004, for textbook treatments) and is the fitness that has to be considered if an exact description of a population is required, for instance, in an individual-based stochastic model, where each individual may have a different phenotype.

Because the only heterogeneity we consider are the different group states (we have no heterogeneity in individual states within groups), the individual 1-fitness function $w_{1,s'|s}$ is invariant under permutations of $z_2, \cdots, z_{n_s}$. With this, we can rewrite eq.(23a) as

$$w_{1,s'|s}(z_1, \boldsymbol{z}_{\{2,\cdots,n_s\}}, z) \qquad \text{or} \qquad w_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z), \tag{23b}$$

where the set-subscripted vector $\boldsymbol{z}_{\{2,\cdots,n_s\}}$ represents a vector of length n_s-1 in which each of $z_2,\cdots,z_{n_s}$ appears in an arbitrary order but exactly once. The subscript $-\{1\}$ is used as a shorthand notation of the set difference $\{1,2,\cdots,n_s\}\setminus\{1\}=\{2,\cdots,n_s\}$ and used when the baseline set $\{1,2,\cdots,n_s\}$ is clear from the context. Therefore, $\boldsymbol{z}_{-\{1\}}$ is the same as $\boldsymbol{z}_{\{2,\cdots,n_s\}}$. Similarly, in the following the subscript $-\{1,2\}$ represents the set difference $\{1,2,\cdots,n_s\}\setminus\{1,2\}=\{3,\cdots,n_s\}$, and so forth. For example, $\boldsymbol{z}_{-\{1,2\}}=\boldsymbol{z}_{\{3,\cdots,n_s\}}$ represents a vector of length n_s-2 in which each of $z_3,\cdots,z_{n_s}$ appears in an arbitrary order but exactly once.

For our two allele model $z_i, z\in\{x,y\}$, we can write a mutant's individual 1-fitness as

$$w_1(s'|s,i)=w_{1,s'|s}(x,\underbrace{x,\cdots,x}_{i-1},\underbrace{y,\cdots,y}_{n_s-i},y). \tag{24}$$

By using the chain rule and permutation invariance, the zeroth, first, and second order perturbations of $w_1(s'|s,i)$ with respect to δ are

$$w_1^{(0)}(s'|s,i)=w_{1,s'|s}, \tag{25a}$$

$$w_1^{(1)}(s'|s,i)=\frac{\partial w_{1,s'|s}}{\partial z_1}+(i-1)\frac{\partial w_{1,s'|s}}{\partial z_2}, \tag{25b}$$

$$w_1^{(2)}(s'|s,i)=\frac{1}{2}\frac{\partial^2 w_{1,s'|s}}{\partial z_1^2}+\frac{i-1}{2}\frac{\partial^2 w_{1,s'|s}}{\partial z_2^2}+(i-1)\frac{\partial^2 w_{1,s'|s}}{\partial z_1\partial z_2}+\frac{(i-1)(i-2)}{2}\frac{\partial^2 w_{1,s'|s}}{\partial z_2\partial z_3}. \tag{25c}$$

Here, all functions and derivatives that appear without arguments are evaluated at the resident population, $(y,\cdots,y)$, a convention we adopt throughout. Note that some derivatives appearing in eqs.(25) are ill-defined for $n_s=1$ and $n_s=2$, but they are always nullified by the factors $(i-1)$ and $(i-1)(i-2)$. Thus, by simply neglecting these ill-defined terms, eq.(25) is valid for any $1\le i\le n_s$.

3.2.2 Individual 2- and 3-fitness

Consider again a focal individual with trait value z_1 in a group in state s in which the n_s-1 group neighbors have the trait values $\boldsymbol{z}_{-\{1\}}=\boldsymbol{z}_{\{2,\cdots,n_s\}}$ in a population that is otherwise monomorphic for z.

For this setting, we define two types of individual 2-fitness functions. First, let

$$w^{\mathrm{I}}_{2,s'|s}(z_1,\boldsymbol{z}_{-\{1\}},z)\qquad (s',s\in\mathcal{S},\ z_1,\cdots,z_{n_s},z\in\mathbb{R}) \tag{26}$$

be the expected number of offspring in state s' that descend from the focal individual and that have a random neighbor that also descends from the focal individual (see Figure 2). Intuitively speaking, $w^{\mathrm{I}}_{2,s'|s}$ measures the number of sibling pairs produced by a focal individual. Hence, when one considers the reproductive process backward in time, $w^{\mathrm{I}}_{2,s'|s}>0$ means that coalescence events do occur. We call $w^{\mathrm{I}}_{2,s'|s}$ the "same-parent individual 2-fitness", because the offspring involved in it descend from the same individual.

Second, for $n_s\ge 2$ consider a neighbor of the focal individual with trait value z_2, called the *target* individual, in a group in which the remaining n_s-2 neighbors have the trait profile $\boldsymbol{z}_{-\{1,2\}}=\boldsymbol{z}_{\{3,\cdots,n_s\}}$. Let

$$w^{\mathrm{II}}_{2,s'|s}(z_1,z_2,\boldsymbol{z}_{-\{1,2\}},z)\qquad (s',s\in\mathcal{S},\ z_1,\cdots,z_{n_s},z\in\mathbb{R}) \tag{27}$$

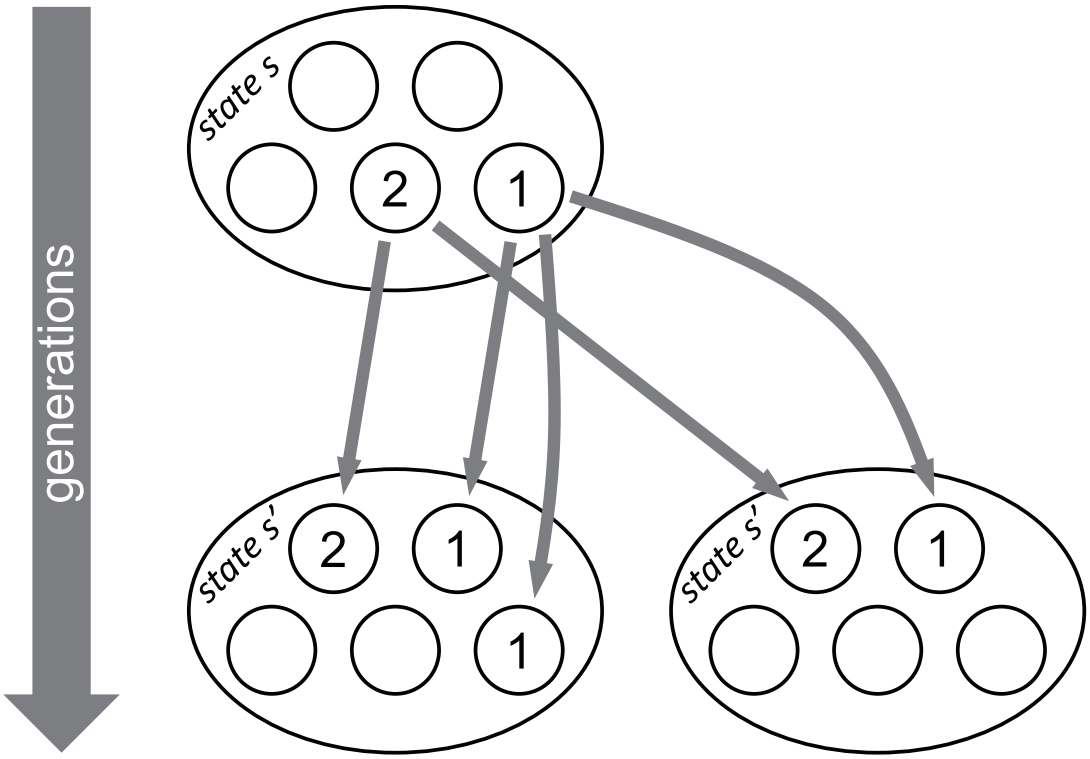

Figure 2: A schematic example of how we calculate the individual 2-fitnesses w_2^{I} and w_2^{II}. Gray arrows represent reproduction (or survival). We label by "1" the focal individual with trait value z_1 in the parental generation and its offspring (possibly including self through survival) in the following generation. Similarly, we label by "2" the target individual with trait value z_2 in the parental generation and its offspring (possibly including self through survival) in the offspring generation. Because each of the two descendants of the focal individual in the bottom-left group (those with label "1") finds with probability 1/4 a random neighbor whose label is "1", whereas the one descendant of the focal individual in the bottom-right group in the offspring generation finds no neighbors whose label is "1", the same-parent individual 2-fitness of the focal is calculated as $w^{\mathrm{I}}_{2,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) = 2 \cdot (1/4) + 1 \cdot (0/4) = 1/2$. Similarly, because each of the two descendants of the focal individual in the bottom-left group finds a random neighbor whose label is "2" with probability 1/4, and because the one descendant of the focal in the bottom-right group finds a random neighbor whose label is "2" with probability 1/4, the different-parent individual 2-fitness of the focal is $w^{\mathrm{II}}_{2,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z) = 2 \cdot (1/4) + 1 \cdot (1/4) = 3/4$.

be the expected number of offspring in state s' that descend from the focal individual with trait value z_1 and that have a random neighbor that descends from the *target* individual with trait value z_2 (see Figure 2). We call $w^{\mathrm{II}}_{2,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z)$ the "different-parent individual 2-fitness", because the offspring involved in it descend from two different pre-selected individuals, which can thus collectively be thought of as the focal set of individuals under consideration. We note that this fitness function is invariant under the permutation of the trait values z_1 and z_2 of individuals from the focal set[2] and it is also invariant under the permutation of the trait values in $\boldsymbol{z}_{-\{1,2\}}$. But since $w^{\mathrm{II}}_{2,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z)$ counts the offspring number (of a certain type) per individual with trait z_1, $w^{\mathrm{II}}_{2,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z)$ is a type of individual fitness.

Using the notation of mutant and resident phenotypes we have for $2 \leq i \leq n_s$ that

$$w_2(s'|s,i) = w^{\mathrm{I}}_{2,s'|s}(x, \underbrace{x, \cdots, x}_{i-1}, \underbrace{y, \cdots, y}_{n_s-i}, y) + (i-1) w^{\mathrm{II}}_{2,s'|s}(x, x, \underbrace{x, \cdots, x}_{i-2}, \underbrace{y, \cdots, y}_{n_s-i}, y), \tag{28}$$

because a mutant neighbor of an offspring of a focal mutant either descends from the focal itself or is an offspring of one of the $i-1$ mutant neighbors of the focal. The zeroth and first order perturbations of $w_2(s'|s,i)$ with respect to δ are given by

$$w_2^{(0)}(s'|s,i) = w^{\mathrm{I}}_{2,s'|s} + (i-1) w^{\mathrm{II}}_{2,s'|s}, \tag{29a}$$

$$w_2^{(1)}(s'|s,i) = \frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_1} + (i-1)\frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_2} + 2(i-1)\frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_1} + (i-1)(i-2)\frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_3}, \tag{29b}$$

where the derivatives $\partial w^{\mathrm{II}}_{2,s'|s}/\partial z_1$ and $\partial w^{\mathrm{II}}_{2,s'|s}/\partial z_2$ (the latter is equal to $\partial w^{\mathrm{II}}_{2,s'|s}/\partial z_1$ due to the permutation invariance, and hence the coefficient "2" appears in eq.(29b)) involve the trait values of the individuals of the focal set and $\partial w^{\mathrm{II}}_{2,s'|s}/\partial z_3$ involves the trait values of a third individual. Note that some derivatives in eqs.(29) are ill-defined for $n_s = 1, 2$ but they are always nullified by the factor $(i-1)$ or $(i-1)(i-2)$. Thus, by simply neglecting these ill-defined terms eq.(29) is valid for any $1 \leq i \leq n_s$.

Following the same line of reasoning as for individual 1- and 2-fitness, we similarly define three different types of individual 3-fitness. See section D in the Supplementary Material for more detailed explanations. Specifically, $w^{\mathrm{I}}_{3,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)$ is defined as the expected number of offspring in state s' that descend from a focal individual in state s with trait value z_1 and that have two random neighbors sampled without replacement both descending from the focal individual. Furthermore, $w^{\mathrm{II}}_{3,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z)$ is defined as the expected number of offspring in state s' that descend from the focal individual in state s with trait value z_1 and with two random neighbors sampled without replacement both descending from a target individual with trait value z_2. Finally, $w^{\mathrm{III}}_{3,s'|s}(z_1, z_2, z_3, \boldsymbol{z}_{-\{1,2,3\}}, z)$ is defined as the expected number of offspring in state s' that descend from the focal individual in state s with trait value z_1 with two random neighbors sampled without replacement, one of which descends from a first target individual

[2]This can be seen by noting that when the focal and target individual from the focal set leave a realized number of A_1 and A_2 offspring, respectively, in the same group of size n_s, then this group contributes to the focal's 2-fitness w_2^{II} with A_1 (the number of focal's offspring) times $A_2/(n_s-1)$ (the probability that a random neighbor of focal's offspring is the target's offspring), which equals to $A_1 A_2/(n_s-1)$. Since $A_1 A_2/(n_s-1)$ is symmetric with respect to A_1 and A_2, changing the roles of the focal and target individual does not alter the realized fitness count. The same logic applies when the focal and target individuals leave offspring to different groups, because in this case the counts per group are simply summed over all groups. A single individual's w_2^{II} is the expectation of such counts over all realizations of offspring number in the same and different groups (where the expectation is taken over all single generation stochastic events affecting reproduction and survival), and the invariance holds because it holds for each realization.

with trait value z_2 and the other descends from a second target individual with trait value z_3. With these definitions, we show in section D in the Supplementary Material that the zeroth-order perturbation of 3-fitness $w_3(s'|s,i)$ with respect to δ is given by

$$w_3^{(0)}(s'|s,i) = w^{\mathrm{I}}_{3,s'|s} + 3(i-1)w^{\mathrm{II}}_{3,s'|s} + \frac{(i-1)(i-2)}{2}w^{\mathrm{III}}_{3,s'|s}, \tag{30}$$

where $w^{\mathrm{I}}_{3,s'|s}, w^{\mathrm{II}}_{3,s'|s}, w^{\mathrm{III}}_{3,s'|s}$ are those three different individual 3-fitness functions evaluated in a resident monomorphic population, $(y,\cdots,y)$.

3.3 Sensitivity results

We now write $\rho^{(1)}$ and $\rho^{(2)}$ from Section 3.1 in terms of the just defined derivatives of the individual fitness functions.

3.3.1 Selection gradient

By substituting eq.(25b) into eq.(21) we obtain

$$\rho^{(1)} = \sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}} v^{(0)}(s')\left[\frac{\partial w_{1,s'|s}}{\partial z_1} + \sum_{i=1}^{n_s}(i-1)\frac{\partial w_{1,s'|s}}{\partial z_2}q^{(0)}(i|s)\right]q^{(0)}(s), \tag{31}$$

and by applying eq.(8) to the second term in square brackets we obtain

$$\rho^{(1)} = \sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}} v^{(0)}(s')\left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (n_s-1)\frac{\partial w_{1,s'|s}}{\partial z_2}r_2^{(0)}(s)\right]q^{(0)}(s). \tag{32}$$

Thus, in order to be able to evaluate $\rho^{(1)}$ it is sufficient to compute the neutral pairwise relatedness $r_2^{(0)}(s)$ while the explicit evaluation of the $q^{(0)}(i|s)$ distribution is not needed. It is indeed a well-known result that the selection gradient $\rho^{(1)}$ can be expressed in terms of reproductive values and relatedness-weighted fitness derivatives (see Frank, 1998; Rousset, 2004, for textbook treatments) and where $q^{(0)}(s)$ and $v^{(0)}(s)$ are given by eq.(15) with $w_{1,s'|s} = w_1^{(0)}(s'|s)$.

Equation (32) can be interpreted as the expected first-order effect of all members of a lineage changing to expressing the mutant allele on the fitness of a focal individual that is a random member of this lineage. The recipient is sampled from state s with probability $q^{(0)}(s)$ and the derivative in the first term in the square brackets of $\rho^{(1)}$ is the effect of the focal changing its own trait value on its individual fitness. The derivative in the second term in the square brackets describes the effect of the group neighbors of the focal changing their trait value on the focal's individual fitness. This term is weighted by pairwise neutral relatedness since this is the likelihood that any such neighbor carries the same allele as the focal in the neutral process. Equation (32) is the inclusive fitness effect of mutating from the resident to the mutant allele for a demographically and/or environmentally structured population and the term in brackets can be thought of as the state-s-specific inclusive fitness effect on offspring in state s'. Equation (32) has previously been derived by Lehmann et al. (2016, Box 2) and is in agreement with eqs.(26) and (27) of Rousset and Ronce (2004), who derived the first-order perturbation $\rho^{(1)}$ in terms of other quantities under the assumptions of fluctuating group size.

We show in section E in the Supplementary Material that by substituting eq.(29a) into eq.(C15), pairwise relatedness (eq.(8)) under neutrality satisfies the recursion

$$r_2^{(0)}(s') = \frac{1}{q^{(0)}(s')} \sum_{s\in\mathcal{S}} \left[w^{\mathrm{I}}_{2,s'|s} + (n_s - 1) w^{\mathrm{II}}_{2,s'|s} r_2^{(0)}(s) \right] q^{(0)}(s). \tag{33}$$

This expression for $r_2^{(0)}(s)$, formulated in terms of individual 2-fitnesses, is novel but is in full agreement with previous results. In particular, eq.(29) of Rousset and Ronce (2004) can be shown to reduce to eq.(33) (see section G.2 in the Supplementary Material for a proof of this connection).

In summary, consistent with well established results, we present a biologically meaningful representation of $\rho^{(1)}$. The ingredients in this representation can be obtained from the three systems of linear equations defined by eqs.(15a), (15b) and (33). This system of equations is fully determined once the individual k-fitnesses functions for $k = 1, 2$, namely, $w_{1,s'|s}$, $w^{\mathrm{I}}_{2,s'|s}$, and $w^{\mathrm{II}}_{2,s'|s}$ are specified for a resident population, and the k-fitness functions can usually be evaluated once a life-cycle has been specified. The dimension of this combined equation system has maximally three times the number of states N. This is significantly lower than the dimension of the matrix $\boldsymbol{A}$ we began with, especially, if group size > 10. In the next section, we extend these results to the disruptive selection coefficient.

3.3.2 Disruptive selection coefficient

Assuming that $\rho^{(1)} = 0$ and substituting eq.(25) into eq.(22), rearrangements given in section E in the Supplementary Material show that

$$\rho^{(2\mathrm{w})} = \frac{1}{2} \sum_{s'\in\mathcal{S}} \sum_{s\in\mathcal{S}} v^{(0)}(s') \left[\frac{\partial^2 w_{1,s'|s}}{\partial z_1^2} + (n_s - 1) \frac{\partial^2 w_{1,s'|s}}{\partial z_2^2} r_2^{(0)}(s) \right.$$
$$\left. + 2(n_s - 1) \frac{\partial^2 w_{1,s'|s}}{\partial z_1 \partial z_2} r_2^{(0)}(s) + (n_s - 1)(n_s - 2) \frac{\partial^2 w_{1,s'|s}}{\partial z_2 \partial z_3} r_3^{(0)}(s) \right] q^{(0)}(s) \tag{34a}$$

$$\rho^{(2\mathrm{q})} = \sum_{s'\in\mathcal{S}} \sum_{s\in\mathcal{S}} v^{(0)}(s') \left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (n_s - 1) \frac{\partial w_{1,s'|s}}{\partial z_2} r_2^{(0)}(s) \right] q^{(1)}(s) \tag{34b}$$

$$\rho^{(2\mathrm{r})} = \sum_{s'\in\mathcal{S}} \sum_{s\in\mathcal{S}} v^{(0)}(s') \left[(n_s - 1) \frac{\partial w_{1,s'|s}}{\partial z_2} r_2^{(1)}(s) \right] q^{(0)}(s). \tag{34c}$$

Equation (34a) depends on four different types of qualitative effects on the fitness of a focal individual: (i) The second-order effect on own fitness of the focal changing its trait value, which is positive, and then contributes to disruptive selection, if fitness is convex in own phenotype. (ii) The second-order effect resulting from the neighbors of the focal changing their trait values, which is positive if the focal's fitness is convex in phenotype of group neighbors. This contributes to disruptive selection proportionally to pairwise relatedness $r_2^{(0)}(s)$, since this is the likelihood that a random neighbor carries the same allele as the focal individual. (iii) The joint effect of the focal individual and any of its neighbors changing their trait value, which is positive if the effect of increased trait values of own and others complement each other. This again contributes to disruptive selection in proportion to the likelihood that any neighbor is a mutant. (iv) The joint effect of pairs of neighbors of the focal changing their trait values, which is positive if the effect of increased trait values in neighbors complement each other. This contributes to disruptive selection with the probability $r_3^{(0)}(s)$ that a pair of neighbors carry the the same allele as the focal individual.

Equation (34b) depends, for each state, on the product of the state specific inclusive fitness effect (recall the term in brackets in eq.(32)) multiplied with the perturbation $q^{(1)}(s)$ of the group state probability. A contribution to disruptive selection occurs if the mutant allele increases its probability to be in a given state while simultaneously increasing the individual fitness of its carriers in that state. Similarly, eq.(34c) depends, for each state, on the product of the state specific indirect effect of others on own fitness (recall the second term in brackets in eq.(32)) and the relatedness perturbation $r_2^{(1)}(s)$. This contributes to disruptive selection if the mutant allele increases the probability that a focal has mutant neighbors while simultaneously increasing the individual fitness of those neighbors. Finally, we note that in the presence of a single state (*i.e.*, no state heterogeneity among groups) $\rho^{(2\mathrm{q})} = 0$. This is the case in all previously published expressions for the disruptive selection coefficient (Day, 2001; Ajar, 2003; Wakano and Lehmann, 2014; Mullon et al., 2016), which therefore reduce to $\rho^{(2\mathrm{w})} + \rho^{(2\mathrm{r})}$ as defined by eqs.(34a) and (34c).

In order to compute $\rho^{(2)}$ we need, in addition to eqs.(15a), (15b) and (33), expressions for $q^{(1)}(s)$, $r_3^{(0)}(s)$, and $r_2^{(1)}(s)$. In section E in the Supplementary Material, we derive the corresponding recursions for $\rho^{(1)} = 0$. In particular, we show that $q^{(1)}(s)$ satisfies

$$q^{(1)}(s') = \sum_{s\in\mathcal{S}} \left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (n_s - 1)\frac{\partial w_{1,s'|s}}{\partial z_2} r_2^{(0)}(s) \right] q^{(0)}(s) + \sum_{s\in\mathcal{S}} w_{1,s'|s} q^{(1)}(s) \tag{35}$$

and that $r_3^{(0)}(s)$ satisfies

$$r_3^{(0)}(s') = \frac{1}{q^{(0)}(s')} \sum_{s\in\mathcal{S}} \left[w^{\mathrm{I}}_{3,s'|s} + 3(n_s - 1) w^{\mathrm{II}}_{3,s'|s} r_2^{(0)}(s) + \frac{(n_s-1)(n_s-2)}{2} w^{\mathrm{III}}_{3,s'|s} r_3^{(0)}(s) \right] q^{(0)}(s). \tag{36}$$

Finally, we show that $r_2^{(1)}(s)$ satisfies the recursion

$$\begin{aligned} r_2^{(1)}(s') = \frac{1}{q^{(0)}(s')} \sum_{s\in\mathcal{S}} \Bigg[& \frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_1} + (n_s - 1)\frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_2} r_2^{(0)}(s) \\ & + 2(n_s - 1)\frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_1} r_2^{(0)}(s) + (n_s-1)(n_s-2)\frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_3} r_3^{(0)}(s) \Bigg] q^{(0)}(s) \\ & + \frac{1}{q^{(0)}(s')} \sum_{s\in\mathcal{S}} \left[(n_s - 1) w^{\mathrm{II}}_{2,s'|s} r_2^{(1)}(s) \right] q^{(0)}(s) \\ & + \frac{1}{q^{(0)}(s')} \sum_{s\in\mathcal{S}} \left[w^{\mathrm{I}}_{2,s'|s} + (n_s - 1) w^{\mathrm{II}}_{2,s'|s} r_2^{(0)}(s) \right] q^{(1)}(s) \\ & - r_2^{(0)}(s') \frac{q^{(1)}(s')}{q^{(0)}(s')}. \end{aligned} \tag{37}$$

Equation (35) shows that $q^{(1)}(s)$ depends on the state-specific inclusive fitness effect (compare the first summand in eq.(35) to the term in brackets in eq.(32)). Thus, the probability that a mutant is in a certain state s increases with its state-specific inclusive fitness effect. Equation (36) for the three-way relatedness coefficient depends on $w^{\mathrm{I}}_{3,s'|s}, w^{\mathrm{II}}_{3,s'|s}$ and $w^{\mathrm{III}}_{3,s'|s}$ and it is a generalization of the pairwise relatedness coefficient given by eq.(33). Finally, eq.(37) shows that $r_2^{(1)}(s)$ depends on direct and indirect effects on $w^{\mathrm{I}}_{2,s'|s}$ and $w^{\mathrm{II}}_{2,s'|s}$. Note, that eq.(37) together with eqs.(15a), (15b), (33), (35), and (36) form a linear system of equations with a dimension equal to six times the number of states N. Its solution allows us to determine the disruptive selection coefficient $\rho^{(2)}$. This system of equations in turn is fully

determined once the the k-fitnesses for $k = 1, 2, 3$ are specified for a resident population, namely, $w_{1,s'|s}$, $w^{\mathrm{I}}_{2,s'|s}$, $w^{\mathrm{II}}_{2,s'|s}$, $w^{\mathrm{I}}_{3,s'|s}$, $w^{\mathrm{II}}_{3,s'|s}$, and $w^{\mathrm{III}}_{3,s'|s}$.

In general, if the state space $\mathcal{S}$ is large, solving this system of equations (and those needed for $\rho^{(1)}$) may be complicated. Similarly, the 2- and 3-fitnesses may be complicated. We here give two directions for approximating $\rho^{(1)}$ and $\rho^{(2)}$. First, individual fitness generally depends on vital rates like fecundity and survival (see eqs.(45)–(46) for a concrete example) and variation of these vital rates may have small effects on fitness, which induces weak selection regardless of the magnitude of the phenotypic deviation δ (called "ω-weak selection" by Wild and Traulsen, 2007, and "weak payoff" by Van Cleve, 2015). For "weak payoffs" (or ω-weak selection), $\rho^{(2)} \approx \rho^{(2\mathrm{w})}$ because one can neglect $\rho^{(2\mathrm{q})}$ and $\rho^{(2\mathrm{r})}$. Indeed, both these terms involve products of marginal changes in fitness, which implies that these products are of second-order effect under weak payoffs and first-order effects will thus dominate. Since $\rho^{(2\mathrm{w})}$ only involves first-order effects it dominates the disruptive selection coefficient. See Van Cleve (2015) for an applications of this approximation to $\rho^{(1)}$ and Wakano and Lehmann (2014) and Mullon et al. (2016) to $\rho^{(2)}$. Second, variation of vital rates and fitness across states may be small under certain biological scenarios in which case one may apply a so-called small noise approximation (e.g., Tuljapurkar, 1990; Caswell, 2001) to $\rho^{(1)}$ and $\rho^{(2)}$, whereby the magnitude of variation are taken to be small. This simplification has been used to approximate $\rho^{(1)}$ in a multi-species meta-population model that is covered by our general model (Mullon and Lehmann, 2018), but has not yet been applied to $\rho^{(2)}$, which would be interesting in future work.

Finally, for some specific life-cycles the 2- and 3-fitness functions can be expressed in terms of components of the 1-fitness functions. This greatly simplifies the calculations because all recursions can then be solved explicitly. We will now provide an application of our model along this latter line, which still covers a large class of models.

4 Application to a lottery model with spatial heterogeneity

We now study a lottery model with overlapping generations and spatial heterogeneity. Such a model can be formulated for a variety of life-cycles and we here take a hierarchical approach in which we make increasingly more specific assumptions. Accordingly, this section is divided in three parts. Section 4.1 provides general results about the components of the selection coefficients based on the assumption of fixed group states s. In Section 4.2 we introduce two forms of population regulation resulting in hard and soft selection, respectively. Finally, in Section 4.3 we specify an explicit fitness function which allows us to present a fully worked example for the effect of group size and spatial heterogeneity on disruptive selection.

4.1 Spatial lottery model

4.1.1 Decomposition into philopatric and dispersal components

We start by making the following three assumptions. (i) Group states s describe environmental variables that do not change in time. Thus, group states are fixed and we here refer to them as habitats. By π_s we denote the relative proportion of groups in habitat s, hence $\sum_{s\in\mathcal{S}} \pi_s = 1$. (ii) Individuals survive

independently of each other with probability $\gamma_s < 1$ to the next time step in a group in habitat s. Note that $\gamma_s = 0$ corresponds to the Wright-Fisher update where all adults die simultaneously, and that $\gamma_s \sim 1$ corresponds to the Moran update where at most one individual dies in a group. (iii) Dispersal occurs individually and independently to a random destination (no propagule dispersal). (iv) The evolving trait does not affect survival. With these assumptions we can decompose the 1-fitness of a focal individual into a philopatric and dispersal component as

$$w_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) = \begin{cases} \underbrace{w^{\mathrm{p}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)}_{\text{philopatric}} + \underbrace{w^{\mathrm{d}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)}_{\text{dispersal}} & (s' = s) \\ \underbrace{w^{\mathrm{d}}_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)}_{\text{dispersal}} & (s' \neq s). \end{cases} \tag{38a}$$

Offspring that have left from their natal group and successfully settled elsewhere are counted in the dispersal component $w^{\mathrm{d}}_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)$. The philopatric component $w^{\mathrm{p}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)$ counts the number of non-dispersing offspring, possibly including self trough survival. Thus, we further decompose the philopatric part into a survival part and a reproduction part as

$$w^{\mathrm{p}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) = \underbrace{\gamma_s}_{\text{philopatric survival}} + \underbrace{(1-\gamma_s) w^{\mathrm{pr}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)}_{\text{philopatric reproduction}}. \tag{38b}$$

Similarly, for the dispersal part we write

$$w^{\mathrm{d}}_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) = \underbrace{(1-\gamma_{s'}) w^{\mathrm{dr}}_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)}_{\text{dispersal reproduction}}. \tag{38c}$$

4.1.2 General results for spatial lottery model

For this model, we explicitly compute the components of the selection gradient and disruptive selection coefficient in sections F.1 and F.2 of the Supplementary Material. In particular, we show that the probability that a random lineage member is sampled from a group in state s under neutrality equals the weighted frequency

$$q^{(0)}(s) = \frac{\pi_s n_s}{\sum_{s' \in \mathcal{S}} \pi_{s'} n_{s'}}, \tag{39}$$

where the weights are the number of individuals in the group state.

For the reproductive value, it is instructive to provide a formula for $v^{(0)}(s')q^{(0)}(s)$, because the reproductive value always appears as a product with $q^{(0)}(s)$ in $\rho^{(1)}$ (eq.(32)) and $\rho^{(2)}$ (eq.(34)) (the only exception is eq.(34b), but see the discussion below eq.(43)). This product is given by

$$v^{(0)}(s')q^{(0)}(s) = \frac{w^{\mathrm{dr}}_{1,s|s'}}{(1-\gamma_{s'})(1-w^{\mathrm{pr}}_{1,s'|s'})(1-w^{\mathrm{pr}}_{1,s|s})} \Bigg/ \left(\sum_{s'' \in \mathcal{S}} \frac{w^{\mathrm{dr}}_{1,s''|s''}}{(1-\gamma_{s''})(1-w^{\mathrm{pr}}_{1,s''|s''})^2} \right) \tag{40}$$

(section F.1 in the Supplementary Material). Furthermore, the neutral pairwise relatedness coefficient equals

$$r^{(0)}_2(s) = \frac{2\gamma_s w^{\mathrm{pr}}_{1,s|s} + (1-\gamma_s)\left(w^{\mathrm{pr}}_{1,s|s}\right)^2}{n_s(1+\gamma_s) - 2(n_s-1)\gamma_s w^{\mathrm{pr}}_{1,s|s} - (n_s-1)(1-\gamma_s)\left(w^{\mathrm{pr}}_{1,s|s}\right)^2} \tag{41}$$

(section F.2 in the Supplementary Material). The general solution for $r_3^{(0)}(s)$ remains complicated (see eq.(F32) for the full expression), but for special cases it is

$$r_3^{(0)}(s) = \begin{cases} \dfrac{\left(w_{1,s|s}^{\mathrm{pr}}\right)^3 \left[n_s + 2(n_s-1)\left(w_{1,s|s}^{\mathrm{pr}}\right)^2\right]}{\left[n_s - (n_s-1)\left(w_{1,s|s}^{\mathrm{pr}}\right)^2\right]\left[n_s^2 - (n_s-1)(n_s-2)\left(w_{1,s|s}^{\mathrm{pr}}\right)^3\right]} & \text{(Wright-Fisher process, } \gamma_s = 0\text{)}, \\ \dfrac{2\left(w_{1,s|s}^{\mathrm{pr}}\right)^2}{\left[n_s - (n_s-1)w_{1,s|s}^{\mathrm{pr}}\right]\left[n_s - (n_s-2)w_{1,s|s}^{\mathrm{pr}}\right]} & \text{(Moran process, } \gamma_s \sim 1\text{)}. \end{cases} \tag{42}$$

If the resident trait value is equal to the singular strategy where $\rho^{(1)} = 0$, then the first-order perturbation of the stationary mutant distribution is

$$\begin{aligned} q^{(1)}(s) = \Bigg\{ & \frac{1}{1-w_{1,s|s}^{\mathrm{pr}}}\left[\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_1} + (n_s-1)\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2} r_2^{(0)}(s)\right] \\ & - \sum_{s'\in\mathcal{S}} \frac{1}{1-w_{1,s'|s'}^{\mathrm{pr}}}\left[\frac{\partial w_{1,s'|s'}^{\mathrm{pr}}}{\partial z_1} + (n_{s'}-1)\frac{\partial w_{1,s'|s'}^{\mathrm{pr}}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s') \Bigg\} q^{(0)}(s) \end{aligned} \tag{43}$$

(section F.1 in the Supplementary Material). Note that we can obtain the fraction $q^{(1)}(s)/q^{(0)}(s)$ by dividing both sides of eq.(43) by $q^{(0)}(s)$, which, when combined with eq.(40), allows to directly obtain the product $v^{(0)}(s')q^{(1)}(s)$. This quantity is required to compute eq.(34b). Finally, for $\rho^{(1)} = 0$ we have

$$\begin{aligned} r_2^{(1)}(s) = 2r_2^{(0)}(s) & \frac{\gamma_s + (1-\gamma_s)w_{1,s|s}^{\mathrm{pr}}}{2\gamma_s w_{1,s|s}^{\mathrm{pr}} + (1-\gamma_s)\left(w_{1,s|s}^{\mathrm{pr}}\right)^2} \\ & \times \left\{ [1+(n_s-1)r_2^{(0)}(s)]\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_1} + (n_s-1)[2r_2^{(0)}(s) + (n_s-2)r_3^{(0)}(s)]\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2} \right\} \end{aligned} \tag{44}$$

(see Section F.2 in the Supplementary Material where we also make the connection to previous work).

With eqs.(40) and (41) we can compute the first-order perturbation of invasion fitness, eq.(32), explicitly given specific life-cycle assumptions (since all recursions have been solved). Similarly, under the assumption that $\rho^{(1)} = 0$, and with eqs.(39)–(44) in hand, we can explicitly compute the second-order perturbation of invasion fitness, eq.(34).

4.2 Fecundity selection under two different forms of density regulation

We further refine our assumptions in order to arrive at two life-cycles with concrete expressions for $w_{1,s|s}^{\mathrm{pr}}$ and $w_{1,s'|s}^{\mathrm{dr}}$. The first one is as follows. (1) Each adult individual in a group in habitat s produces on average a very large number f_s of offspring, and then either survives with probability γ_s or dies with the complementary probability. (2) Offspring disperse independently of each other to a uniformly randomly chosen non-natal group with the non-zero probability m_s. An offspring survives dispersal with probability p_s when dispersing from a group in habitat s. (3) All offspring aspiring to settle in a group in habitat s compete for the average number $(1-\gamma_s)n_s$ of breeding sites vacated by the death of adults and are recruited until all n_s breeding sites are occupied. (4) The evolving trait does not affect dispersal.

In this life cycle, density-dependent population regulation occurs after dispersal when offspring aspire to settle and we refer to this regime as *hard selection*. We also consider a *soft-selection* variant in which

density regulation occurs in two steps (as in Fig. 1 of Svardal et al., 2015). First, a local trait-dependent stage of density-dependent regulation occurs immediately after reproduction (after stage (1) in the above life cycle) in which the offspring pool in each group is brought back to a size proportional to the local group size n_s, say size Kn_s, where K is a large number. From here on dispersal and recruitment (second regulation step) proceed as in the hard-selection life cycle.

For these two life cycles, the philopatric and dispersal fitness components can be written as

$$w^{\mathrm{pr}}_{1,s|s}(z_1,\boldsymbol{z}_{-\{1\}},z)=\begin{cases} n_s\dfrac{(1-m_s)f_s(z_1,\boldsymbol{z}_{-\{1\}},z)}{(1-m_s)\sum_{i=1}^{n_s}f_s(z_i,\boldsymbol{z}_{-\{i\}},z)+I_{\mathrm{hard}}(z)} & \text{(hard selection)} \quad (45a)\\ \underbrace{n_s\dfrac{(1-m_s)}{(1-m_s)n_s+I_{\mathrm{soft}}}}_{\text{trait independent recruitment}}\times\underbrace{n_s\dfrac{f_s(z_1,\boldsymbol{z}_{-\{1\}},z)}{\sum_{i=1}^{n_s}f_s(z_i,\boldsymbol{z}_{-\{i\}},z)}}_{\text{trait dependent regulation}} & \text{(soft selection)}, \quad (45b)\end{cases}$$

and

$$w^{\mathrm{dr}}_{1,s'|s}(z_1,\boldsymbol{z}_{-\{1\}},z)=\begin{cases} \pi_{s'}n_{s'}\dfrac{p_sm_sf_s(z_1,\boldsymbol{z}_{-\{1\}},z)}{(1-m_{s'})\sum_{i=1}^{n_{s'}}f_{s'}(z,\boldsymbol{z},z)+I_{\mathrm{hard}}(z)} & \text{(hard selection)} \quad (46a)\\ \underbrace{\pi_{s'}n_{s'}\dfrac{p_sm_s}{(1-m_{s'})n_{s'}+I_{\mathrm{soft}}}}_{\text{trait independent recruitment}}\times\underbrace{n_s\dfrac{f_s(z_1,\boldsymbol{z}_{-\{1\}},z)}{\sum_{i=1}^{n_s}f_s(z_i,\boldsymbol{z}_{-\{i\}},z)}}_{\text{trait dependent regulation}} & \text{(soft selection)}, \quad (46b)\end{cases}$$

respectively, where $f_s(z_i,\boldsymbol{z}_{-\{i\}},z)$ is the fecundity of individual i in a group in habitat s and

$$I_{\mathrm{hard}}(z)\equiv\sum_{s\in\mathcal{S}}\pi_sn_sp_sm_sf_s(z,\boldsymbol{z},z) \quad (47a)$$

$$I_{\mathrm{soft}}\equiv\sum_{s\in\mathcal{S}}\pi_sn_sp_sm_s \quad (47b)$$

are the trait-dependent immigration terms for the hard-selection model and trait-independent immigration term for the soft selection model, respectively.

Equations (45b) and (46b) can be understood as follows. During the stage of trait-dependent regulation the local offspring pool in a group in habitat s is brought back to a size proportional to n_s, namely Kn_s, whereby the proportion of individuals among the surviving offspring descending from a focal individual is $f_s(z_1,\boldsymbol{z}_{-\{1\}},z)/\sum_{i=1}^{n_s}f_s(z_i,\boldsymbol{z}_{-\{i\}},z)$. Each of these offspring either disperses or stays local and then competes to be recruited. With probability $1-m_s$ an offspring is philopatric, and this philopatric offspring gets recruited with probability $1/\left[K((1-m_s)n_s+I_{\mathrm{soft}})\right]$ per open spot. Here $K((1-m_s)n_s+I_{\mathrm{soft}})$ is the expected number of local competitors, where the number of migrant offspring competing in a given group for recruitment and coming from a group in habitat s is proportional to $\pi_sn_sp_sm_s$. Offspring dispersing to a group in habitat s' experience on average $K\left((1-m_{s'})n_{s'}+I_{\mathrm{soft}}\right)$ competitors and the probability to compete in such a group is $\pi_{s'}$. The likelihood to be recruited (either after dispersing or without dispersing) is then multiplied by the expected number of open breeding sites, which equals $n_s(1-\gamma_s)$ in the natal group and $n_{s'}(1-\gamma_{s'})$ in non-natal groups in habitat s', but the factors $(1-\gamma_s)$ and $(1-\gamma_{s'})$ are already accounted for in eqs.(38b) and (38c). Note that the constant K does not appear in eqs.(45b) and (46b) because it appears both in the numerator and denominator of these equations and thus cancels out.

Using eq.(45) and eq.(46) along with eqs.(39)–(44) allows to compute $\rho^{(1)}$ and $\rho^{(2)}$ for a large class of models. In sections G, H.3 and I.3 in the Supplementary Material, we show that we recover a number

of previously published results belonging to this class of models, some of which were derived with quite different calculations (Pen, 2000; Ohtsuki, 2010; Lehmann and Rousset, 2010; Rodrigues and Gardner, 2012; Wakano and Lehmann, 2014; Svardal et al., 2015; Mullon et al., 2016; Parvinen et al., 2018). This indirectly confirms the validity of our calculations. For simplicity of notation we assumed that the evolving trait does neither affect survival nor dispersal (it only affects fecundity), extensions to include effects on survival and dispersal are in principle straightforward.

4.3 Selection analysis

In this section, we finally present explicit expressions for the selection gradient $\rho^{(1)}$ and the coefficient of disruptive selection $\rho^{(2)}$ for both the model of hard and soft selection. We then introduce an explicit fecundity function, which, under some additional symmetry assumptions, allows us to have a completely worked example.

4.3.1 Hard selection

Inserting eqs.(45a) and (46a) into eqs.(38b) and (38c), respectively, we show in section H in the Supplementary Material that the selection gradient for the hard selection lottery model is

$$\rho^{(1)} \propto \sum_{s\in\mathcal{S}} \frac{\pi_s n_s p_s m_s f_s}{d_{s,\text{hard}}} \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + r_2^{(0)}(s)(n_s-1)\frac{\frac{\partial f_s}{\partial z_2}}{f_s} - (1-d_{s,\text{hard}})^2 r_{2,\text{R}}^{(0)}(s)\left(\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1)\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\}, \tag{48}$$

where the proportionality constant is positive (and given by the inverse of eq.(H4)) and $d_{s,\text{hard}}$ is the backward migration rate from groups in state s under neutrality defined as

$$d_{s,\text{hard}} \equiv \frac{I_{\text{hard}}}{(1-m_s)n_s f_s + I_{\text{hard}}}. \tag{49}$$

This rate depends on y because I_{hard} and f_s are evaluated at $(y,\cdots,y)$. Equation (48) further depends on

$$r_{2,\text{R}}^{(0)}(s) \equiv \frac{1}{n_s} + \frac{n_s-1}{n_s} r_2^{(0)}(s), \tag{50}$$

which is the relatedness between two individuals sampled with replacement in a group in state s and where

$$r_2^{(0)}(s) = \frac{2\gamma_s(1-d_{s,\text{hard}}) + (1-\gamma_s)(1-d_{s,\text{hard}})^2}{n_s(1+\gamma_s) - 2(n_s-1)\gamma_s(1-d_{s,\text{hard}}) - (n_s-1)(1-\gamma_s)(1-d_{s,\text{hard}})^2}. \tag{51}$$

Equation (48) can be understood as follows. The first term in the curly brackets is the marginal fecundity effect by a focal individual on itself, while the second term is the marginal fecundity effect conferred by all group members to the focal individual weighted by the coefficient of pairwise relatedness. Finally, the third term reflects competition for the finite number of breeding spots in a group. A change in the trait value of a focal individual that increases its fecundity or that of its neighbors increases the strength of local competition. This reduces the fitness of the focal individual if the additional offspring remain philopatric and compete with own offspring. Equation (48) is a generalization of previous results obtained for the island model (see section H in the Supplementary Material for the detail of these connections).

Similarly, inserting eqs.(45a) and (46a) into eqs.(38b) and (38c), respectively, and using these in eq.(34), we obtain a general expression for the disruptive selection coefficient $\rho^{(2)}$ under hard selection. The resulting expression, while useful for numerical calculations, is too lengthy to be presented here and we refer to section H in the Supplementary Material for details. Therein, we show that under a Wright-Fisher process ($\gamma_s = 0$) the results of Parvinen et al. (2018) are recovered, who obtained an expression of $\rho^{(2)}$ expressed in terms of first- and second-order derivatives of f_s.

To complement these results and to approach a fully worked example, we assume a Moran process (*i.e.*, $\gamma_s \sim 1$) and that fecundity of an adult individual depends only on its own phenotype (*i.e.*, $f_s(z_1, \boldsymbol{z}_{-\{1\}}, z) = f_s(z_1)$). Under these assumptions, we show in section J.1 in the Supplementary Material that the selection gradient is a weighted sum of $\mathrm{d}f_s/\mathrm{d}z_1$ over different states s (see eq.(J1)), and that the disruptive selection coefficient is

$$\rho^{(2)} \propto \sum_{s\in\mathcal{S}} \frac{\pi_s n_s p_s m_s f_s}{d_{s,\mathrm{hard}}} \left\{ X_{1,s,\mathrm{hard}} \frac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} + X_{2,s,\mathrm{hard}} \left(\frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s} \right)^2 \right\}, \tag{52a}$$

where the positive proportionality constant is the same as in eq.(48), and

$$X_{1,s,\mathrm{hard}} = \frac{1}{2} \frac{d_{s,\mathrm{hard}}(1 - d_{s,\mathrm{hard}} + n_s)}{1 + d_{s,\mathrm{hard}}(n_s - 1)} \quad (\geq 0) \tag{52b}$$

$$X_{2,s,\mathrm{hard}} = \frac{d_{s,\mathrm{hard}}(1 - d_{s,\mathrm{hard}})(1 - d_{s,\mathrm{hard}} + n_s)n_s}{\{2 + d_{s,\mathrm{hard}}(n_s - 2)\}\{1 + d_{s,\mathrm{hard}}(n_s - 1)\}} \quad (\geq 0) \tag{52c}$$

For complete dispersal (*i.e.*, $d_{s,\mathrm{hard}} = 1$)[3] we obtain that $X_{1,s,\mathrm{hard}} = 1/2$ and $X_{2,s,\mathrm{hard}} = 0$. As the dispersal rate $d_{s,\mathrm{hard}}$ decreases, the ratio $X_{2,s,\mathrm{hard}}/X_{1,s,\mathrm{hard}}$ increases monotonically. Hence, as dispersal becomes more limited, relatively more weight is put on the squared first-order derivative $(\mathrm{d}f_s/\mathrm{d}z_1)^2$ compared to the second-order derivative $\mathrm{d}^2 f_s/\mathrm{d}z_1^2$, indicating that limited dispersal facilitates disruptive selection (and, if the singular strategy y^* is convergence stable and remains so when varying dispersal, then evolutionary branching is facilitated). On the other hand, for a fixed $d_{s,\mathrm{hard}} < 1$, the ratio $X_{2,s,\mathrm{hard}}/X_{1,s,\mathrm{hard}}$ monotonically decreases as group size decreases. Hence, with decreasing group size less weight is put on the squared first-order derivative $(\mathrm{d}f_s/\mathrm{d}z_1)^2$, which acts to limit disruptive selection. We finally note that the functional form of eq.(52a) holds beyond the Moran process, provided all other assumptions are the same. While the weights will depend on the specifics of the reproductive process, we conjecture that the weights will feature the same qualitative dependence on dispersal and group size.

We now make two further assumptions. First, we follow Svardal et al. (2015) and assume that fecundity is under Gaussian stabilising selection with habitat specific optimum $y_{\mathrm{op},s}$. Thus,

$$f_s(z_1) = f_{\max} \exp\left[-\frac{(z_1 - y_{\mathrm{op},s})^2}{2\sigma_{\mathrm{st}}^2} \right], \tag{53}$$

where $f_{\max}$ is the maximal fecundity of an individual and σ_{st}^2 is inversely proportional to the strength of stabilising selection. Second, we assume that group size, migration and juvenile survival are identical for all habitats, *i.e.*, $n_s = n$, $m_s = m$, and $p_s = p$ for all s. Hence, habitats only differ in the trait value $y_{\mathrm{op},s}$ that maximizes fecundity.

[3]For a homogeneous population with a single habitat s, a singular point is characterized by $\mathrm{d}f_s/\mathrm{d}z_1 = 0$, and therefore eq.(52) predicts that the sign of the disruptive selection coefficient is solely determined by the sign of $\mathrm{d}^2 f_s/\mathrm{d}z_1^2$ no matter whether dispersal is complete or locally limited. A similar result has been shown in Parvinen et al. (2017) by assuming a Wright-Fisher process.

Under these assumptions, the singular strategy y^* is implicitly given by

$$y^* = \sum_{s \in \mathcal{S}} \psi_s(y^*) y_{\mathrm{op},s}, \tag{54}$$

which is a weighted average of the habitat specific trait optima with the weights ψ_s being complicated functions of the model parameters (see section J.1 in the Supplementary Material). The condition for the disruptive selection coefficient at the singular point y^* (eq.(52a)) being positive can be expressed as

$$\sum_{s \in \mathcal{S}} \Psi_s(y^*)(y_{\mathrm{op},s} - y^*)^2 > \sigma^2_{\mathrm{st}}, \tag{55}$$

where the Ψ_s's are again complicated weights (section J.1 in the Supplementary Material).

These expressions greatly simplify when we consider only two habitats with equal proportions, *i.e.* $\mathcal{S} = \{1, 2\}$ with $\pi_1 = \pi_2 = 1/2$, no mortality in dispersal, $p = 1$, and symmetric optima in the sense that $y_{\mathrm{op},2} = -y_{\mathrm{op},1}$. Due to this symmetry, $y^* = 0$ is a solution of eq.(53) and therefore a singular strategy. Furthermore, in section J.1 in the Supplementary Material, we find that under the aforementioned assumptions

$$\Psi_s(y^*) = \frac{1}{2}\left(\frac{2-m}{m} - \frac{4(1-m)^2}{m(2+m(n-2))}\right). \tag{56}$$

Then, by using the variance of the habitat optima defined by

$$\sigma^2_{\mathrm{op}} = \sum_{s \in \mathcal{S}} \pi_s \left(y_{\mathrm{op},s} - y^*\right)^2 \tag{57}$$

(in the current case, with $\pi_1 = \pi_2 = 1/2$), condition (55) can be written as

$$\left(\frac{2-m}{m} - \underbrace{\frac{4(1-m)^2}{m(2+m(n-2))}}_{\to 0 \text{ when } n \to \infty}\right) \sigma^2_{\mathrm{op}} > \sigma^2_{\mathrm{st}}. \tag{58}$$

The first term in the parenthesis is the effect of limited dispersal on disruptive selection in the absence of kin selection (that is, under infinite group size). This term increases with decreasing dispersal, which facilitates disruptive selection. Indeed, low dispersal increases the probability that lineage members experience the same group-specific state favoring local adaptation. The second term in the parenthesis captures the effect of kin selection. The absolute value of this negative term increases with both decreasing dispersal and decreasing group size, which inhibits disruptive selection. This effect can be understood as follows. All philopatric offspring within a group compete with each other for the limited number of spots to settle within a group. Relatedness among individuals within a group increases with decreasing group size. Thus, in smaller groups competing individuals are more likely to be related with each other and this diminishes the benefit of mutations increasing adaptation to the group-specific state. This effect becomes more pronounced with decreasing dispersal since this increases relatedness within groups even more. We therefore expect that the singular point y^* is more likely to be uninvadable for small groups and this is indeed what we observe in Figure 3, especially evident in panel (f). It can be shown that the effect of decreasing dispersal on the first term on the left-hand side of (58) dominates the effect on the second term. Thus, decreasing m indeed facilitates disruptive selection as illustrated in Figure 3(b-f).

In the limit of $m = 0$ and $m = 1$ the condition for the disruptive selection coefficient being positive (58) becomes

$$\begin{cases} (1+n)\sigma_{\text{op}}^2 > \sigma_{\text{st}}^2 & \text{when } m \to 0 \\ \sigma_{\text{op}}^2 > \sigma_{\text{st}}^2 & \text{when } m = 1. \end{cases} \tag{59}$$

Thus, at very low dispersal the singular point changes from being uninvadable to invadable when group size exceeds $n = (\sigma_{\text{st}}^2 - \sigma_{\text{op}}^2)/\sigma_{\text{op}}^2$ (as can be seen in Figure 3(f) where the boundary between CSS and branching point for very low m occurs at $n = 4$). At complete dispersal, the singular point is uninvadable for $\sigma_{\text{op}}^2 < \sigma_{\text{st}}^2$ and invadable otherwise. Finally, the singular strategy is more likely to be under stabilizing selection the larger the ratio $\sigma_{\text{st}}^2/\sigma_{\text{op}}^2$, as is clearly illustrated in Figure 3(a-f).

A singular point at which selection is disruptive is an evolutionary branching point if it is also convergence stable. Substituting eq.(48) under all mentioned assumptions into eq.(19) we obtain after rearrangements that $y^* = 0$ is convergence stable if

$$\left(2 - m - \underbrace{\frac{(1-m)^2(1-m+(1+m)n)}{(1+m(n-1))(1-m+n)}}_{\to 0 \text{ when } n \to \infty}\right)\sigma_{\text{op}}^2 < \sigma_{\text{st}}^2 \tag{60}$$

and repelling otherwise. From inspecting the left-hand side of this condition, the coefficient of σ_{op}^2 is a unimodal function of m and takes the minimum value 1 at $m = 0, 1$ and the maximum at

$$m = \frac{\sqrt{1+n}}{n + \sqrt{1+n}} \tag{61}$$

for any fixed n. Therefore, it is clear that $\sigma_{\text{op}}^2 < \sigma_{\text{st}}^2$ is a necessary but not sufficient condition for convergence stability. More generally, increasing σ_{st}^2 relative to σ_{op}^2 increases the space in the (m, n)-plane for which the singular point is convergence stable (*cf.* Figure 3(a-f)). In section J.1 in the Supplementary Material we show that $2\sigma_{\text{op}}^2 < \sigma_{\text{st}}^2$ is a sufficient condition for convergence stability (*cf.* Figure 3(e-f)). Interestingly, from the unimodality above, the singular point can be repelling for intermediate values of m as can be seen in Figure 3(b-d). For large group size, condition (60) becomes $(2 - m)\sigma_{\text{op}}^2 < \sigma_{\text{st}}^2$ and therefore convergence stability changes at $m = 2 - (\sigma_{\text{st}}^2/\sigma_{\text{op}}^2)$, which coincides very well with where the singular point turns from convergence stable to repelling at group size $n = 100$ in Figure 3(b-d). For the effect of group size n on convergence stability, the coefficient of σ_{op}^2 in condition (60) is, for any fixed $0 < m < 1$, an increasing function of n. Thus, smaller group sizes are more favorable for convergence stability of the singular point $y^* = 0$.

An immediate conclusion from these observations is that for $m = 1$ evolutionary branching does not occur under hard selection (with fecundity given by eq.(53)). This is so because for $m = 1$ competition is global and does not occur between individuals within a group. This removes any frequency-dependent selection effect. Indeed, under our assumptions setting $m = 1$ (and $p = 1$) in eq.(45a) and eq.(46a) results in $w_{1,s|s}^{\text{pr}}(z_1, \boldsymbol{z}_{-\{1\}}, z) = 0$ and $w_{1,s'|s}^{\text{dr}}(z_1, \boldsymbol{z}_{-\{1\}}, z) = \pi_{s'} n_{s'} f_s(z_1)/I_{\text{hard}}(z)$ for all s' and s. Thus, there is no longer any state specific frequency-dependence, since $I_{\text{hard}}(z)$ is common to all fitness functions. In this case, the singular point is both convergence stable and uninvadable if $\sigma_{\text{op}}^2 < \sigma_{\text{st}}^2$ and both repelling and invadable if $\sigma_{\text{op}}^2 > \sigma_{\text{st}}^2$. This is in agreement with the well-known finding that under hard selection and complete dispersal selection is frequency-independent and adaptive polymorphism cannot

be maintained by spatial heterogeneity alone (Dempster, 1955; Ravigné, 2004; Ravigné et al., 2009; Débarre and Gandon, 2011).

4.3.2 Soft selection

Inserting eqs.(45b) and (46b) into eqs.(38b) and (38c), respectively, we show in section I in the Supplementary Material that the selection gradient for the soft selection lottery model is

$$\rho^{(1)} \propto \sum_{s\in\mathcal{S}} \frac{\pi_s n_s p_s m_s}{d_{s,\text{soft}}} \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + r_2^{(0)}(s)(n_s-1)\frac{\frac{\partial f_s}{\partial z_2}}{f_s} - r_{2,\text{R}}^{(0)}(s)\left(\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1)\frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}, \tag{62}$$

where the positive proportionality constant is positive (and given by the inverse of eq.(I4)) and

$$d_{s,\text{soft}} \equiv \frac{I_{\text{soft}}}{(1-m_s)n_s + I_{\text{soft}}} \tag{63}$$

is the backward migration rate from groups in habitat s under neutrality. In contrast to the case of hard selection, eq.(63) is independent of y. Pairwise relatedness under neutrality $r_2^{(0)}(s)$ takes the same form as in eq.(51) where all $d_{s,\text{hard}}$ have to be replaced with $d_{s,\text{soft}}$. The key difference between eq.(48) and eq.(62) is that under soft selection the competition term is larger than under hard selection because the weighting by the backward dispersal probability has disappeared in the latter case. This reflects the fact that under soft selection density regulation occurs before dispersal. Again, eq.(62) is a generalization of previous results as detailed in section I in the Supplementary Material.

Similarly, inserting eqs.(45b) and (46b) into eqs.(38b) and (38c), respectively, and using these in eq.(34), we obtain a general expression for the disruptive selection coefficient $\rho^{(2)}$ under soft selection. As was the case for hard selection, the resulting expression can be useful for numerical calculations, but is too lengthy to be presented here and we refer to section I in the Supplementary Material for details.

Paralleling the analysis under hard selection, we assume a Moran process (*i.e.*, $\gamma_s \sim 1$) and that the fecundity of adult individuals depends only on their own phenotype ($f_s(z_1, \boldsymbol{z}_{-\{1\}}, z) = f_s(z_1)$). Under these assumptions we show in section J.2 in the Supplementary Material that

$$\rho^{(2)} \propto \sum_{s\in\mathcal{S}} \frac{\pi_s n_s p_s m_s}{d_{s,\text{soft}}} \left\{ X_{1,s,\text{soft}} \frac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} + X_{2,s,\text{soft}} \left(\frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s} \right)^2 \right\}, \tag{64a}$$

where the positive proportionality constant is the same as in eq.(62), and

$$X_{1,s,\text{soft}} = \frac{1}{2}\frac{d_{s,\text{soft}}(n_s-1)}{1+d_{s,\text{soft}}(n_s-1)} \quad (\geq 0) \tag{64b}$$

$$X_{2,s,\text{soft}} = \frac{d_{s,\text{soft}}(n_s-1)\{d_{s,\text{soft}}(1-d_{s,\text{soft}})(n_s-1)(n_s-2) - 2d_{s,\text{soft}}(n_s-1) + (n_s-2)\}}{\{2+d_{s,\text{soft}}(n_s-2)\}\{1+d_{s,\text{soft}}(n_s-1)\}^2}. \tag{64c}$$

The ratio of these weights, $X_{2,s,\text{soft}}/X_{1,s,\text{soft}}$, shows qualitatively the same behavior as the corresponding expressions under hard selection (eqs.(52b) and (52c)) with respect to changes in $d_{s,\text{soft}}$ and n_s. However, a notable difference from the hard selection case is that $X_{2,s,\text{soft}}$ (and hence the ratio, $X_{2,s,\text{soft}}/X_{1,s,\text{soft}}$) can be negative for small n_s and large d_s. We finally note that, as was the case for eq.(52a), the functional form of eq.(64a) holds beyond the Moran process, provided all other assumptions are the same.

Under the assumption of Gaussian fecundity selection (eq.(53)) and $n_s = n$, $m_s = m$, $p_s = p = 1$ for all states s, which entails $d_{\text{soft}} = m$, we again obtain a fully worked example. The value y^* for the

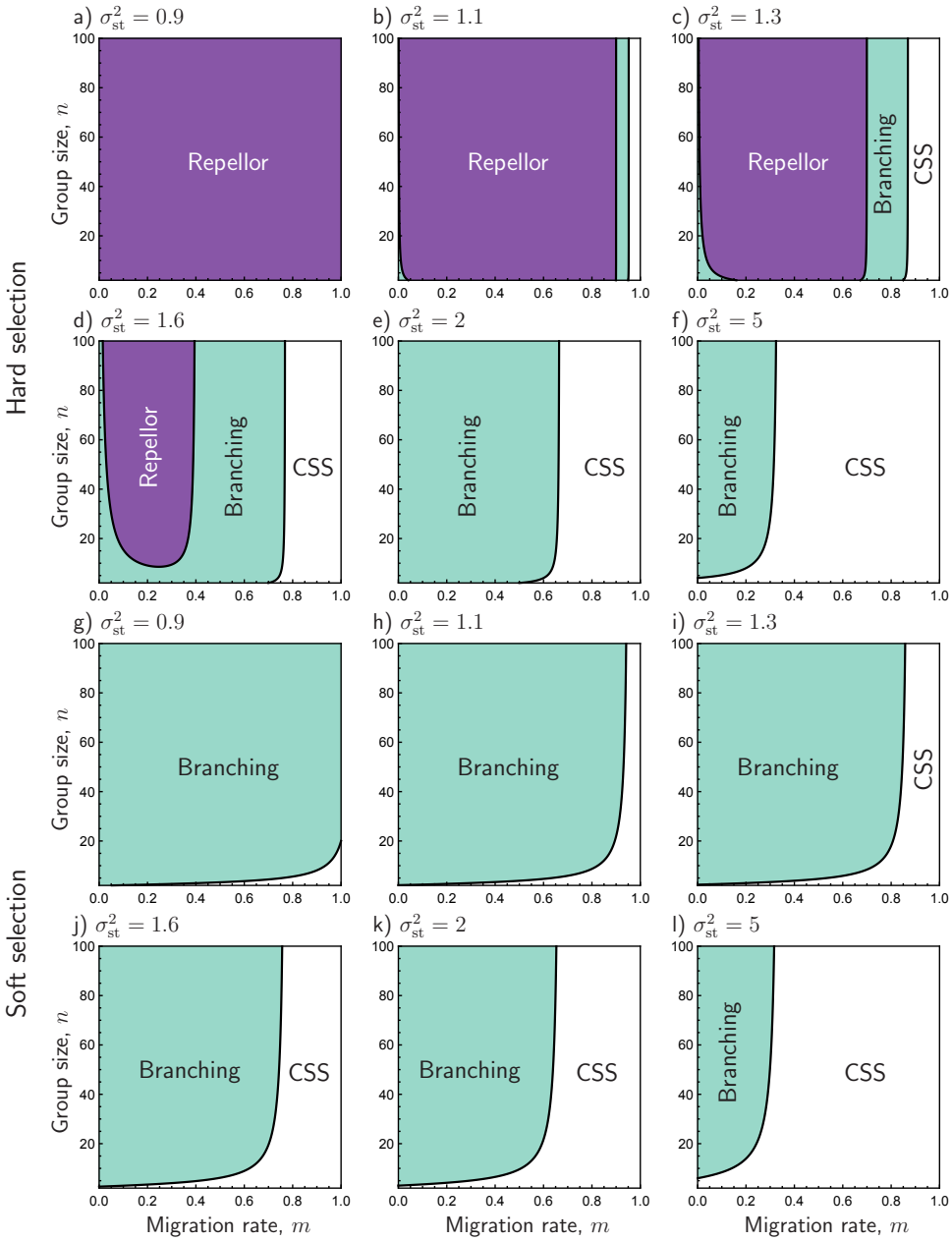

Figure 3: Bifurcation diagrams for the singular point $y^* = 0$ as a function of the migration rate m (x-axis) and group size n (y-axis) for six different values of the within group selection parameter $\sigma^2_{\rm st}$ (see eq.(53)). (a-f) Hard selection, (g-l) soft selection. Purple: evolutionary repellor, blue: evolutionary branching point, white: uninvadable and convergence stable singular point, *i.e.*, continuously stable strategy (CSS). Other parameter values: $y_{\rm op,1} = 1 = -y_{\rm op,2}$ (implying $\sigma^2_{\rm op} = 1$).

singular strategy is given by the average habitat optimum,

$$y^* = \sum_{s \in \mathcal{S}} \pi_s y_{\mathrm{op},s} \tag{65}$$

(section J.2 in the Supplementary Material). Furthermore, the coefficient of disruptive selection is positive if and only if

$$\left(\frac{2-m}{m} - \underbrace{\frac{4 + 2m\,(2-m)\,(n-2)}{m(2+m(n-2))(1+m(n-1))}}_{\to 0 \text{ when } n \to \infty} \right) \sigma^2_{\mathrm{op}} > \sigma^2_{\mathrm{st}}, \tag{66}$$

where σ^2_{op} is the variance in the habitat optima defined by eq.(57). Note that condition (66) is valid only for $n \geq 2$ (because otherwise eqs.(64b) and (64c) evaluate to zero). The two terms in parenthesis on the left-hand side of condition (66) have the same interpretation as the corresponding terms in condition (58) for the case of hard selection and they respond in the same direction with respect to changes in dispersal m and group size n. In the limit of infinitely large group size ($n \to \infty$) the second term vanishes and we recover eq.(C.15) of Svardal et al. (2015).

In section J.2 in the Supplementary Material, we show that y^* as given by eq.(65) is convergence stable for any value of σ^2_{st} and σ^2_{op} and independent of group size n and dispersal probability m. Thus, the singular point is an evolutionary branching point when it is invadable and an endpoint of the evolutionary dynamics (continuously stable strategy, CSS) when uninvadable. For the special case of only two habitats with $y_{\mathrm{op},1} = 1 = -y_{\mathrm{op},2}$, Figure 3 shows how n, m and σ^2_{st} determine whether $y^* = 0$ is a branching point or a CSS. In summary, stronger selection (smaller values of σ^2_{st}), lower migration and larger groups favor adaptive diversification at an evolutionary branching point.

5 Discussion

The main result of this paper is an expression for the disruptive selection coefficient $\rho^{(2)}$ in heterogeneous group-structured populations (eq.(34)). We show that $\rho^{(2)}$ depends on three types of differentials: (a) the first- and second-order perturbations of the expected number of offspring in different states produced by an individual in a given state, (b) the first-order perturbation of the probability that an individual is in the different states, and (c) the first-order perturbation of the probability that a randomly sampled neighbor of an individual carries alleles identical by descent (perturbation of relatedness). These differentials depend on and are weighted by three quantities evaluated under neutrality: (i) the reproductive values $v^{(0)}(s)$ of individuals in state s, (ii) the pairwise and three-way relatedness coefficients $r_2^{(0)}(s)$ and $r_3^{(0)}(s)$ in state s, and (iii) the probability $q^{(0)}(s)$ that a randomly sampled individual resides in a group in state s.

At a conceptual level, our results about the components of $\rho^{(2)}$ can be thought of as a direct extension of the result that the three types of neutral weights – reproductive values, relatednesses, and probabilities of occurrence in state s – are needed to evaluate the selection gradient $\rho^{(1)}$ for quantitative traits in group-structured populations (Taylor and Frank, 1996; Frank, 1998; Rousset, 2004). All the above mentioned differentials and their weights can be obtained by solving systems of linear equations that are at most

of dimension N, *i.e.*, the number of states groups can be in. This represents a significant reduction compared to the dimension of the state space of the original evolutionary process, which is equal to the dimension of the mutant transition matrix $\boldsymbol{A}$.

A distinctive and novel feature of our analysis is the introduction of the concept of individual k-fitness, $w_k(s'|s,i)$, which describes the expected number of descendants of a mutant in an (s,i) group (possibly including self through survival) that settle in state-s' groups and have $k-1$ randomly sampled neighbors that are also mutants (*i.e.*, that descend from the same common ancestor). In the context of our perturbation analysis, we show that $w_k(s'|s,i)$ can be themselves expressed in terms of individual k-fitness functions for $k=1,2,3$ where individuals are labelled as focal, group neighbor and population member, and which are sufficient to evaluate all aforementioned quantities and thus $\rho^{(1)}$ and $\rho^{(2)}$ (see sections 3.2.1-3.3). These latter individual k-fitness functions do not depend on the mutant type and provide for $k=2,3$ the generalizations of the fitness functions for $k=1$ already in use in the direct fitness method (Taylor and Frank, 1996; Frank, 1998; Rousset, 2004). They are thus sufficient biological ingredients to determine whether or not disruptive selection occurs. In a well-mixed populations in which individuals do not interact with relatives only individual 1-fitness functions are required to evaluate $\rho^{(1)}$ and $\rho^{(2)}$. Individual 2- and 3-fitnesses describe the possibility that under limited dispersal the offspring of a given parent can have neighbors (here one or two) that belong to the same lineage and are thus more likely to have the same trait value than randomly sampled individuals from the population. This causes non-random mutant-mutant interactions, which is well known to critically affect the nature of selection on traits affecting own and others' reproduction and survival (Hamilton, 1964; Michod, 1982; Frank, 1998; Rousset, 2004). Because the individual k-fitnesses describe group configurations in which offspring have neighbors that belong to the same lineage, the ancestral lineages of the k interacting individuals must coalesce in a common ancestor, and this can occur only if there is a non-zero probability that at least two individuals descend from the same parent over a generation (see section G.2 in the Supplementary Material for the connection to coalescence theory). Neutral relatedness in evolutionary models is indeed usually computed by using coalescence arguments and thus use a "backward" perspective on allele transmission (*e.g.* Taylor and Frank, 1996; Frank, 1998; Rousset, 2004). This may somewhat disconnect relatedness from the "forward" perspective of allele transmission induced by reproduction. Using individual 2-fitnesses to evaluate relatedness (see eq.(33)) brings upfront the connection between relatedness and reproduction (note that the "backward" approach may nevertheless be more useful for concrete calculations of relatedness).

As an application of our results, we analyze a lottery model with overlapping generations in heterogeneous habitats that allows for both hard and soft selection regimes. For this scenario, we show that $\rho^{(1)}$ and $\rho^{(2)}$ can in principle be solved explicitly (all systems of equation can be solved explicitly) but that generic expressions remain complicated functions, since they apply to any kind of social interactions (*i.e.*, any "game") and different ecologies. In doing these calculations, we recover a number of previous results concerning relatedness, selection gradients and disruptive selection coefficients for lottery models (in particular those of Pen, 2000; Rousset and Ronce, 2004; Ohtsuki, 2010; Lehmann and Rousset, 2010; Rodrigues and Gardner, 2012; Wakano and Lehmann, 2014; Svardal et al., 2015; Mullon et al., 2016;

Parvinen et al., 2018, see sections G, H.3 and I.3 in the Supplementary Material for details), which confirms the validity of our approach. Finally, as a fully worked example, we investigate the evolution of adaptive polymorphism due to local adaption by extending the soft selection model of Svardal et al. (2015) to finite group size and hard selection. We confirm that adaptive polymorphism is generally favored by limited migration under soft selection and that small group size does not change this result qualitatively but tends to inhibit disruptive selection. For hard selection, however, the situation is more complicated as limited dispersal and finite group size favors not only disruptive selection but also repelling generalist strategies so that it becomes less likely that polymorphism can emerge from gradual evolution (Figure 3). With respect to limited migration this finding is also described by Débarre and Gandon (2011).

While our model allows for many different types of interactions between individuals within groups, it also has several limitations. At the individual level, we consider only scalar traits, but multidimensional (or functional-valued) traits can be taken into account by replacing derivatives by directional derivatives, which will not change the structure of our perturbation analysis. At the group level, we do not consider heterogeneity within groups, but in natural populations individuals within groups are likely to differ in their physiological state such as age, size and sex. To incorporate physiological heterogeneity requires an extension of the state space $\mathcal{S}$ and to take into account the distribution of mutants within sub-groups of individuals belonging to the same physiological state in a group. The structure of our perturbation analysis, however, will remain unchanged by adding within-group heterogeneity, and only additional reproductive values and relatednesses will be needed. Likewise, in order to take isolation-by-distance into account, one again needs to extend the state space $\mathcal{S}$, while to include diploidy one needs to extend the number of genetic states and this should only impact the relatedness coefficients. While such extensions remain to be done (and have all been done for the selection gradient $\rho^{(1)}$ (*e.g.* Rousset, 2004)), they are unlikely to change the required components of the disruptive selection coefficient $\rho^{(2)}$ and how they are connected algebraically. We thus conjecture that the representation of $\rho^{(2)}$ holds generally.

In conclusion, for a large class of models we describe the consequences of limited dispersal and finite group size on evolutionary stability and diversification in heterogeneous populations, which we hope will help to formulate and analyze concrete biological models.

Acknowledgments

HO and KP received support from the SOKENDAI Advanced Sciences Synergy Program (SASSP). HO and JYW received support from JSPS KAKENHI (No.16K07524). HO received support from JSPS KAKENHI (No.20K06812). JYW received support from JSPS KAKENHI (No.16K05283 and 16H06412). We are grateful to Dr. Yu Uchiumi for discussions.

References

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

Bulmer, M. G. 1986. Sex ratio theory in geographically structured populations. Heredity 56:69–73.

Caswell, H. 2001. Matrix Population Models. 2nd ed. Sinauer.

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

Day, T. 2001. Population structure inhibits evolutionary diversification under competition for resources. Genetica 112-113:71–86.

Débarre, F., and S. Gandon. 2011. Evolution in heterogeneous environments: Between soft and hard selection. The American Naturalist 177:E84–E97.

Dempster, E. 1955. Maintenance of genetic heterogeneity. Cold Spring Harbor Symposia on Quantitative Biology 20:25–32.

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

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

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

Grafen, A. 1985. A geometric view of relatedness. Oxford surveys in evolutionary biology 2:28–89.

———. 1991. Modeling in behavioural ecology. Pages 5–31 *in* J. R. Krebs and N. Davies, eds. Behavioural Ecology. Blackwell Scientific Publications, Oxford.

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

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

Harris, T. E. 1963. The Theory of Branching Processes. Springer-Verlag Berlin Heidelberg.

Karlin, S., and H. M. Taylor. 1975. A First Course in Stochastic Processes. Academic Press, San Diego.

Kisdi, E. 2016. Dispersal polymorphism in stable habitats. Journal of Theoretical Biology 392:69–82.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characers. Evolution 37:1210–1226.

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

Lehmann, L., and F. Rousset. 2010. How life-history and demography promote or inhibit the evolution of helping behaviors. Philosophical Transactions of the Royal Society B 365:2599–2617.

Massol, F., and F. Débarre. 2015. Evolution of dispersal in spatially and temporally variable environments: The importance of life cycles. Evolultion 69:1925–1937.

Massol, F., A. Duputiè, P. David, and P. Jarne. 2011. Asymmetric patch size distribution leads to disruptive selection on dispersal. Evolution 65:490–500.

Metz, J. A. J., S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs, and J. S. Van Heerwaarden. 1996. Adaptive dynamics: A geometrical study of the consequences of nearly faithful reproduction. Pages 183–231 *in* S. van Strien and S. Verduyn Lunel, eds. Stochastic and spatial structures of dynamical systems, Proceedings of the Royal Dutch Academy of Science. North Holland, Dordrecht, Netherlands; available at http://pure.iiasa.ac.at/id/eprint/4497/.

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

Michod, R. E. 1982. The theory of kin selection. Annual Review of Ecology and Systematics 13:23–55.

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

Mullon, C., and L. Lehmann. 2018. Eco-evolutionary dynamics in metacommunities: ecological inheritance, helping within species, and harming between species. American Naturalist 192:664–685.

Ohtsuki, H. 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. Evolution 64:3344–53.

Parker, G. A., and J. Maynard Smith. 1990. Optimality theory in evolutionary biology. Nature 348:27–33.

Parvinen, K., H. Ohtsuki, and J. Y. Wakano. 2017. The effect of fecundity derivatives on the condition of evolutionary branching in spatial models. Journal of Theoretical Biology 416:129–143.

———. 2018. Spatial heterogeneity and evolution of fecundity-affecting traits. Journal of Theoretical Biology 454:190–204.

———. 2020. Evolution of dispersal in a spatially heterogeneous population with finite patch sizes. Proceedings of the National Academy of Sciences USA 117:7290–7295.

Pen, I. 2000. Reproductive effort in viscous populations. Evolution 54:293–297.

Phillips, P. A., and S. J. Arnold. 1989. Visualizing multivariate selection. Evolution 43:1209–1222.

Ravigné. 2004. Implications of habitat choice for protected polymorphisms. Evolutionary Ecology Research 6:125–145.

Ravigné, V., U. Dieckmann, and I. Olivieri. 2009. Live where you thrive: Joint evolution of habitat choice and local adaptation facilitates specialization and promotes diversity. The American Naturalist 174:E141–E169.

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

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

Rousset, F. 2004. Genetic Structure and Selection in Subdivided Populations, vol. 40 of *Monographs in Population Biology*. Princeton University Press, Princeton, N.J.

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

Rueffler, C., J. A. J. Metz, and T. J. M. Van Dooren. 2013. What life cycle graphs can tell about the evolution of life histories. Journal of Mathematical Biology 66:225–279.

Rueffler, C., T. J. M. Van Dooren, O. Leimar, and P. A. Abrams. 2006. Disruptive selection and then what? Trends in Ecology and Evolution 21:238–245.

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

Szathmary, E., and L. Demeter. 1987. Group selection of early replicators and the origin of life. Journal of Theoretical Biology 128:463–486.

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

———. 1990. Allele-frequency change in a class structured population. The American Naturalist 135:95–106.

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

Tuljapurkar, S. 1990. Population Dynamics in variable environments, vol. 85 of *Lecture Notes in Biomathematics*. Springer Verlag, Berlin, Germany.

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

Wakano, J. Y., and L. Lehmann. 2014. Evolutionary branching in deme-structured populations. Journal of Theoretical Biology 351:83–95.

Wild, G. 2011. Inclusive fitness from multitype branching processes. Bulletin of Mathematical Biology 73:1028–1051.

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

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

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

Supplementary Material of
"The components of directional and disruptive selection in heterogeneous group-structured populations"

Hisashi Ohtsuki, Claus Rueffler, Joe Yuichiro Wakano, Kalle Parvinen & Laurent Lehmann

Table of Contents

A Mathematical properties of the baseline model **37**
A.1 Multitype branching process . . . 37
A.2 List of assumptions and their implications . . . 38

B Derivation of perturbations of invasion fitness **39**

C Derivation of recursions **41**
C.1 Recursions of $q(s)$ and $r_k(s)$ for arbitrary δ . . . 42
C.1.1 Recursion for $q(s)$ . . . 42
C.1.2 Recursion for $r_k(s)$ . . . 43
C.2 Perturbations of $q(s)$ and $r_k(s)$ . . . 43
C.2.1 Perturbation of $q(s)$ . . . 44
C.2.2 Perturbation of $r_k(s)$ . . . 44
C.3 Closedness of the recursions in $q(s)$ and $r_k(s)$ . . . 45

D Individual 3-fitness **47**

E Derivation of the quantities with individual fitness functions **48**

F Derivation of the results for the lottery model **51**
F.1 Calculations for $q(s)$ and $v^{(0)}(s)$ . . . 51
F.2 Individual 2-fitness and 3-fitness for the lottery model and calculations for relatedness 54
F.2.1 Pairwise relatedness . . . 55
F.2.2 Three-way relatedness . . . 57
F.2.3 First-order perturbation of pairwise relatedness . . . 60

G Consistency with previous results about relatedenss **61**
G.1 Neutral relatedness in the lottery model . . . 61
G.2 Neutral relatedness under fluctuating group size . . . 61
G.3 Perturbation of relatedness . . . 63

H Perturbations for the hard selection lottery model **64**
H.1 First-order perturbation of invasion fitness . . . 64
H.2 Second-order perturbation of invasion fitness . . . 66
H.3 Consistency with previous results about perturbations . . . 72
H.3.1 A model with overlapping generations by Lehmann and Rousset (2010) . . . 72
H.3.2 A spatially heterogeneous model by Rodrigues and Gardner (2012) . . . 72
H.3.3 A cancellation result by Mullon et al. (2016) . . . 73
H.3.4 Second-order results by Parvinen et al. (2018) . . . 73

I Perturbations for the soft selection lottery model **74**
I.1 First-order perturbation of invasion fitness . . . 74
I.2 Second-order perturbation of invasion fitness . . . 76

I.3 Consistency with previous results 80
I.3.1 A model with "Regulation before dispersal" by Lehmann and Rousset (2010) . 80
I.3.2 A model with local adaptation by Svardal et al. (2015) 81

J Evolutionary analysis of a lottery model with local adaptation 81
J.1 Hard selection 81
J.2 Soft selection 84

A Mathematical properties of the baseline model

In this section, we provide a mathematical description of the stochastic process underlying the mutant dynamics that we consider in our paper.

A.1 Multitype branching process

We study the process of invasion of a mutant arising as a single copy (or a finite number of copies) in a monomorphic resident population and consider mutant dynamics as long as the mutant remains rare. Specifically, we pay attention to the number of groups including at least one adult mutant. Groups can differ in their state $s \in \mathcal{S}$ and in the number $i \in \{1, \cdots, n_s\}$ of mutants. We therefore count the number of each "type" of group, where a type, denoted by τ here and thereafter, is specified by the vector $\tau = (s, i)$. The set of all possible types is $\mathcal{T} \equiv \cup_{s\in\mathcal{S}}\{(s, i) \mid i \in \{1, \cdots, n_s\}\}$ and there are $n \equiv \sum_{s\in\mathcal{S}} n_s$ different group types.

A "population state" describes how the mutant is distributed among group types as long as it is rare in a population otherwise monomorphic for the resident type. It is specified by a vector $\boldsymbol{M} = \{M_\tau\} \in \mathbb{N}^n$, where $\mathbb{N} = \{0, 1, 2, \cdots\}$ and where each M_τ represents the number of type-τ groups at a given time. We consider that the change in state is given by a discrete-time and time-homogeneous Markov chain, denoted by $\mathcal{M}$, defined on $\mathbb{N}^n$, where the transition probability from population state $\boldsymbol{M}$ to $\boldsymbol{M}'$ is given by $P(\boldsymbol{M} \to \boldsymbol{M}')$. Here, P implicitly depends both on the mutant and resident trait values, x and y, respectively, and allows to define the generating function

$$G_{\boldsymbol{M}}(\boldsymbol{\xi}) \equiv \sum_{\boldsymbol{M}'\in\mathbb{N}^n} \left(P(\boldsymbol{M} \to \boldsymbol{M}') \prod_{\tau\in\mathcal{T}} \xi_\tau^{M'_\tau} \right) \tag{A1}$$

induced by the Markov chain, where $\boldsymbol{\xi} \equiv \{\xi_\tau\}_{\tau\in\mathcal{T}}$ is a vector of dummy variables. This Markov chain has one absorbing state, which is the extinction of the mutant, and otherwise only transient states with the possibility that the absorbing state is never reached and so the number of mutants grows without bound.

We assume that the Markov chain $\mathcal{M}$ is a multitype branching process, meaning that each group "behaves" independently of the other groups. Mathematically, this assumption is embodied by the generating function of the Markov chain (eq.(A1)) being given by

$$G_{\boldsymbol{M}}(\boldsymbol{\xi}) = \prod_{\tau\in\mathcal{T}} \{G_{\boldsymbol{E}_\tau}(\boldsymbol{\xi})\}^{M_\tau}, \tag{A2}$$

where $\boldsymbol{E}_\tau$ is a vector of length n whose τ-th component is 1 and all the others are zero. Intuitively speaking, eq.(A2) shows that $P(\boldsymbol{M} \to \boldsymbol{M}')$ is uniquely determined by the "fundamental" transitions

probabilities, $P(\boldsymbol{E}_\tau \to \boldsymbol{M}')$ ($\tau \in \mathcal{T}$), each representing how many groups of the same and different types are "produced" by a single type-τ group in the previous time step (by "produced" we mean that the survival and reproduction of individuals in a single type-τ group affect the composition of that group in the descendant generation, as well as the composition of other groups by emigration of offspring, *e.g.*, eq.(1) of Lehmann et al., 2016).

For a given initial population state $\boldsymbol{M}_0 \in \mathbb{N}^n$, the ultimate extinction probability of mutants is defined as

$$\pi(\boldsymbol{M}_0) = \text{Prob}(\boldsymbol{M}_{t=\infty} = \boldsymbol{0} \,|\, \boldsymbol{M}_0), \tag{A3}$$

where $\boldsymbol{M}_{t=\infty}$ is the population state vector when the number of time steps $t \to \infty$, and $\boldsymbol{0}$ represents a vector of zeroes of length n. Now we define

$$a(\tau'|\tau) = \sum_{\boldsymbol{M}' \in \mathbb{N}^n} M'_{\tau'} P(\boldsymbol{E}_\tau \to \boldsymbol{M}'), \tag{A4}$$

which is the expected number of type-τ' groups that are "produced" by a single type-τ group[1]. We collect the expectations $a(\tau'|\tau)$ for all τ, τ' to construct matrix $\boldsymbol{A} = \{a(\tau'|\tau)\}$. We assume that (i) matrix $\boldsymbol{A}$ is primitive, as specified in section 2.2 in the main text, and that (ii) $\mathcal{M}$ is not "singular" (where $\mathcal{M}$ being "singular" means that all of $G_{\boldsymbol{E}_\tau}(\boldsymbol{\xi})$ ($\tau \in \mathcal{T}$) are linear functions without a constant term; see Harris, 1963). Let ρ be the largest eigenvalue of $\boldsymbol{A}$ (since $\boldsymbol{A}$ is primitive it follows from the Perron-Frobenius theorem that such a unique positive ρ exists). Then, a standard result in multitype branching process theory (Harris, 1963; Karlin and Taylor, 1975) guarantees that the following relations hold between ρ and the $\pi(\boldsymbol{E}_\tau)$:

$$\begin{aligned} \rho \leq 1 \quad &\Longrightarrow \quad \pi(\boldsymbol{E}_\tau) = 1 \quad (\text{for all } \tau \in \mathcal{T}), \\ \rho > 1 \quad &\Longrightarrow \quad \pi(\boldsymbol{E}_\tau) < 1 \quad (\text{for all } \tau \in \mathcal{T}). \end{aligned} \tag{A5}$$

It is precisely matrix $\boldsymbol{A} = \{a(\tau'|\tau)\} = \{a(s', i'|s, i)\}$ that we study in the main text. Because our focus is only on the uninvadability of the resident with respect to invasion of single mutants, which translates to whether or not $\pi(\boldsymbol{E}_{(s,1)}) = 1$, we do not have to distinguish two different multitype branching processes that yield the same $\boldsymbol{A}$-matrix. Therefore, an evolutionary invasion analysis can be started from $\boldsymbol{A}$, and does not need to detail the transition probabilities of the underlying multitype branching process.

A.2 List of assumptions and their implications

We here summarize the basic mathematical assumptions for our model and their biological implications.

(i) A time-homogeneous Markov chain $\mathcal{M}$, which depends on x, y and whose state space is $\{(M_\tau)_{\tau \in \mathcal{T}} \in \mathbb{N}^n\}$, is a multitype branching process. This implies that each group that includes at least one mutant behaves independently of all other groups with mutants.

(ii) Matrix $\boldsymbol{A} = \{a(\tau'|\tau)\}$, calculated from $\mathcal{M}$ as in eq.(A4), is primitive. This implies that each type of group has a positive contribution to the production of any other type of group after some finite number of time steps.

[1] We assume that this expectation exists for all τ and τ'. Using generating the function $G_{\boldsymbol{E}_\tau}(\boldsymbol{\xi})$, eq.(A4) is calculated as $\frac{\partial}{\partial \xi_{\tau'}} G_{\boldsymbol{E}_\tau}(\boldsymbol{\xi})\Big|_{\boldsymbol{\xi}=\boldsymbol{1}}$, where $\boldsymbol{1}$ is a vector of ones.

(iii) $\mathcal{M}$ is not singular. Hence, we do not consider a degenerate multitype branching process in which each group always produces exactly one group at the next time step.

(iv) Individual fitness $w_1^{(0)}(s'|s,i)$ as defined in eq.(14) does not depend on i (and therefore can be written as $w_1^{(0)}(s'|s)$). This implies that resident and mutant individuals are indistinguishable (and exchangeable) under neutrality.

(v) Matrix $\boldsymbol{W}^{(0)}$, whose entries are determined by $w_1^{(0)}(s'|s)$, has 1 as its largest eigenvalue. This implies that a monomorphic population of resident individuals stays at the same average group size due to density-dependent regulation.

B Derivation of perturbations of invasion fitness

We here prove eq.(21) and eq.(22). Before doing so, we list some frequently used relations:

$$w_1^{(0)}(s'|s,i) = w_1^{(0)}(s'|s) \qquad \text{(see section 2.3.4)} \tag{B1a}$$

$$\sum_{s\in\mathcal{S}} q(s) = 1 \qquad \text{(see eqs.(3) and (4))} \tag{B1b}$$

$$\sum_{i=1}^{n_s} q(i|s) = 1 \qquad \text{(see eqs.(4) and (5)).} \tag{B1c}$$

We decompose $w_1(s'|s,i)$ in eq.(16a) into a neutral part and a non-neutral part,

$$\begin{aligned} w_1(s'|s,i) &= \underbrace{w_1^{(0)}(s'|s,i)}_{\text{neutral}} + \underbrace{w_{1,\text{sel}}(s'|s,i)}_{\text{non-neutral}} \\ &\underset{\text{eq.(B1a)}}{=} w_1^{(0)}(s'|s) + w_{1,\text{sel}}(s'|s,i). \end{aligned} \tag{B2}$$

Note that, by definition, the ℓ-th order perturbation of $w_{1,\text{sel}}(s'|s,i)$ with respect to δ equals

$$w_{1,\text{sel}}^{(\ell)}(s'|s,i) = \begin{cases} 0 & (\ell = 0) \\ w_1^{(\ell)}(s'|s,i) & (\ell \geq 1). \end{cases} \tag{B3}$$

From eq.(16a), we then have

$$\begin{aligned} \rho &= \frac{1}{V}\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}} v^{(0)}(s') w_1^{(0)}(s'|s) \underbrace{\sum_{i=1}^{n_s} q(i|s)}_{\substack{=1 \\ \text{(from eq.(B1c))}}} q(s)\right] + \frac{1}{V}\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right] \\ &= \frac{1}{V}\left[\sum_{s\in\mathcal{S}} \underbrace{\sum_{s'\in\mathcal{S}} v^{(0)}(s') w_1^{(0)}(s'|s)}_{=v^{(0)}(s)\text{ (from eq.(15b))}} q(s)\right] + \frac{1}{V}\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right] \\ &= \frac{1}{V}\underbrace{\sum_{s\in\mathcal{S}} v^{(0)}(s) q(s)}_{=V\text{ (from eq.(16b))}} + \frac{1}{V}\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right] \\ &= 1 + \frac{1}{V}\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right]. \end{aligned} \tag{B4}$$

As a check, the zeroth order perturbation is

$$\begin{aligned}
\rho^{(0)} &= 1 + \left(V^{-1}\right)^{(0)} \left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right]^{(0)} \\
&= 1 + \left(V^{-1}\right)^{(0)} \left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') \underbrace{w^{(0)}_{1,\text{sel}}(s'|s,i)}_{=0\,(\text{from eq.(B3)})} q^{(0)}(i|s) q^{(0)}(s)\right] \\
&= 1,
\end{aligned} \tag{B5}$$

as expected. The first-order perturbation of eq.(B4) is

$$\begin{aligned}
\rho^{(1)} &= \left(V^{-1}\right)^{(1)} \underbrace{\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right]^{(0)}}_{=0\,(\text{shown in eq.(B5)})} \\
&\quad + \underbrace{\left(V^{-1}\right)^{(0)}}_{=(V^{(0)})^{-1}} \left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right]^{(1)} \\
&= \left(\underbrace{\sum_{s\in\mathcal{S}} v^{(0)}(s) q^{(0)}(s)}_{=1\,(\text{from eq.(15c)})}\right)^{-1} \left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\text{sel}}(s'|s,i) q(i|s) q(s)\right]^{(1)} \\
&= \sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') \underbrace{w^{(1)}_{1,\text{sel}}(s'|s,i)}_{\substack{=w_1^{(1)}(s'|s,i) \\ (\text{from eq.(B3)})}} q^{(0)}(i|s) q^{(0)}(s) \\
&= \sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s),
\end{aligned} \tag{B6}$$

which reproduces eq.(E.14) in Lehmann et al. (2016). Note that the first-order perturbation of the term in square brackets in the third line of eq.(B6) can potentially produce more terms in the fourth line, but they are null because $w^{(0)}_{1,\text{sel}}(s'|s,i) = 0$ (see eq.(B3)). Equation (B6) proves eq.(21) in the main text.

Next, we study $\rho^{(2)}$ under the condition $\rho^{(1)} = 0$. The second-order perturbation of eq.(B4) is

$$
\begin{aligned}
\rho^{(2)} = \left(V^{-1}\right)^{(2)} & \underbrace{\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\mathrm{sel}}(s'|s,i) q(i|s) q(s)\right]^{(0)}}_{=0\,(\text{shown in eq.(B5)})} \\
& + \left(V^{-1}\right)^{(1)} \underbrace{\left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\mathrm{sel}}(s'|s,i) q(i|s) q(s)\right]^{(1)}}_{=\rho^{(1)}=0,\,(\text{shown in (B6) \& assumption})} \\
& + \underbrace{\left(V^{-1}\right)^{(0)}}_{=1\,(\text{shown in eq.(B6)})} \left[\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_{1,\mathrm{sel}}(s'|s,i) q(i|s) q(s)\right]^{(2)} \\
= & \sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') \underbrace{w^{(2)}_{1,\mathrm{sel}}(s'|s,i)}_{\substack{=w_1^{(2)}(s'|s,i)\\ (\text{from eq.(B3)})}} q^{(0)}(i|s) q^{(0)}(s) \\
& + \sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') \underbrace{w^{(1)}_{1,\mathrm{sel}}(s'|s,i)}_{\substack{=w_1^{(1)}(s'|s,i)\\ (\text{from eq.(B3)})}} q^{(0)}(i|s) q^{(1)}(s) \\
& + \sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') \underbrace{w^{(1)}_{1,\mathrm{sel}}(s'|s,i)}_{\substack{=w_1^{(1)}(s'|s,i)\\ (\text{from eq.(B3)})}} q^{(1)}(i|s) q^{(0)}(s) \\
= & \underbrace{\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(2)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s)}_{=\rho^{(2\mathrm{w})}} + \underbrace{\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(1)}(s)}_{=\rho^{(2\mathrm{q})}} \\
& + \underbrace{\sum_{s'\in\mathcal{S}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n_s} v^{(0)}(s') w_1^{(1)}(s'|s,i) q^{(1)}(i|s) q^{(0)}(s)}_{=\rho^{(2\mathrm{r})}} .
\end{aligned}
\tag{B7}
$$

Note that the second-order perturbation of the term in square brackets in the third line of eq.(B7) can potentially produce more terms in the following lines, but they are null because $w^{(0)}_{1,\mathrm{sel}}(s'|s,i) = 0$ (see eq.(B3)). Equation (B7) proves eq.(22) in the main text.

C Derivation of recursions

In order to practically use the general formulae for the first and second order derivatives of invasion fitness, eq.(21) and eq.(22), or to use the corresponding formulae derived for the individual fitness functions, eq.(32) and eq.(34), we need to know $q(s)$ (the asymptotic distribution that a randomly sampled mutant finds itself in a group of state s) and $q(i|s)$ (the asymptotic distribution that a randomly sampled mutant, given that it is sampled from a group in state s, finds itself in a group with i mutants) under neutrality as well as the first-order perturbation of these quantities with respect to δ. Due to eq.(7), knowing $q(i|s)\,(i = 1, \cdots, n_s)$ is equivalent to knowing relatedness, $r_k(s)\,(k = 1, \cdots, n_s)$. The purpose of this section is to derive recursions that $q(s)$ and $r_k(s)$ satisfy. Specifically, in section C.1 we derive recursions that are valid for any δ, and in section C.2 we describe their perturbations to the zeroth- (hence under

neutrality) and first-order of δ.

C.1 Recursions of $q(s)$ and $r_k(s)$ for arbitrary δ

For simplicity, we will from here on omit the lower and upper bound of the summation when obvious from the context.

C.1.1 Recursion for $q(s)$

Writing $\rho \boldsymbol{u} = \boldsymbol{A}\boldsymbol{u}$ component-wise gives

$$\rho u(s', i') = \sum_s \sum_i a(s', i'|s, i) u(s, i). \tag{C1}$$

By multiplying both sides with i' we obtain

$$\rho i' u(s', i') = \sum_s \sum_i i' a(s', i'|s, i) u(s, i). \tag{C2}$$

Summing eq.(C2) over s' and i' gives

$$\rho \sum_{s'} \sum_{i'} i' u(s', i') = \sum_{s'} \sum_{i'} \sum_s \sum_i i' a(s', i'|s, i) u(s, i). \tag{C3}$$

Dividing eq.(C2) by eq.(C3) results in

$$\frac{i' u(s', i')}{\sum_{s''} \sum_{i''} i'' u(s'', i'')} = \frac{\sum_s \sum_i i' a(s', i'|s, i) u(s, i)}{\sum_{s''} \sum_{i''} \sum_s \sum_i i'' a(s'', i''|s, i) u(s, i)}. \tag{C4}$$

Using eq.(3) we note that the left-hand side equals $q(s', i')$. Thus, eq.(C4) can be rewritten as

$$q(s', i') = \frac{\sum_s \sum_i \{i' a(s', i'|s, i)/i\}\{i u(s, i)\}}{\sum_{s''} \sum_{i''} \sum_s \sum_i \{i'' a(s'', i''|s, i)/i\}\{i u(s, i)\}}. \tag{C5}$$

We divide both the numerator and the denominator of the right-hand side of eq.(C5) by the constant, $\sum_{s'} \sum_{i'} i' u(s', i')$. By using eq.(3) once again we obtain

$$q(s', i') = \frac{\sum_s \sum_i \{i' a(s', i'|s, i)/i\} q(s, i)}{\sum_{s''} \sum_{i''} \sum_s \sum_i \{i'' a(s'', i''|s, i)/i\} q(s, i)}, \tag{C6}$$

which is the recursion that $q(s, i)$ obeys.

To obtain the recursion that $q(s)$ obeys, we sum eq.(C6) over i' and obtain

$$\begin{aligned} q(s') = \sum_{i'} q(s', i') &= \frac{\sum_s \sum_i \sum_{i'} \{i' a(s', i'|s, i)/i\} q(s, i)}{\sum_{s''} \sum_s \sum_i \sum_{i''} \{i'' a(s'', i''|s, i)/i\} q(s, i)} \\ &\underset{\text{eq.(10)}}{=} \frac{\sum_s \sum_i w_1(s'|s, i) q(s, i)}{\sum_{s''} \sum_s \sum_i w_1(s''|s, i) q(s, i)} \\ &\underset{\text{eq.(5)}}{=} \frac{\sum_s \sum_i w_1(s'|s, i) q(i|s) q(s)}{\sum_{s''} \sum_s \sum_i w_1(s''|s, i) q(i|s) q(s)}. \end{aligned} \tag{C7}$$

Interpretation of ρ: Dividing eq.(C3) by the constant $\sum_s \sum_i iu(s,i)$ gives

$$\begin{aligned}\rho &= \frac{\sum_{s'} \sum_{i'} \sum_s \sum_i i'a(s',i'|s,i)u(s,i)}{\sum_{s''} \sum_{i''} i''u(s'',i'')} \\ &= \sum_{s'} \sum_s \sum_i \underbrace{\sum_{i'} \{i'a(s',i'|s,i)/i\}}_{=w_1(s'|s,i)\text{ (from eq.(10))}} \frac{iu(s,i)}{\sum_{s''} \sum_{i''} i''u(s'',i'')} \\ &= \sum_{s'} \left[\sum_s \sum_i w_1(s'|s,i) \underbrace{\frac{iu(s,i)}{\sum_{s''} \sum_{i''} i''u(s'',i'')}}_{\substack{=q(s,i) \\ =q(i|s)q(s)}} \right], \end{aligned} \tag{C8}$$

which reproduces eq.(5) of Lehmann et al. (2016). The term inside the square bracket of eq.(C8) can be interpreted as the state-s' component of the expected individual fitness of a mutant randomly sampled from the asymptotic distribution $u(s,i)$ (*i.e.*, the probability for a randomly sampled mutant to find itself in an (s,i)-group is proportional to $iu(s,i)$). Thus, invasion fitness ρ can be interpreted as the expected number of mutant copies produced by a lineage member randomly sampled from the distribution $q(s,i)$.

Combining eqs.(C7) and (C8) gives us a useful relationship

$$q(s') = \frac{1}{\rho} \left\{ \sum_s \sum_i w_1(s'|s,i)q(i|s)q(s) \right\}. \tag{C9}$$

C.1.2 Recursion for $r_k(s)$

From the definition of r_1 in eq.(7) we have $r_1(s) = 1$. Thus, we are interested in the recursions for $r_k(s)$ for $k \geq 2$. Using the definition for $q(s,i)$ in eq.(5) and the expression eq.(C6), we have

$$q(i'|s') = \frac{\sum_s \sum_i \{i'a(s',i'|s,i)/i\}q(s,i)}{\sum_{i''} \sum_s \sum_i \{i''a(s',i''|s,i)/i\}q(s,i)}. \tag{C10}$$

Multiplying both sides of the last equation by $\phi_k(s',i')$ and summing over i' gives

$$\begin{aligned} r_k(s') = \sum_{i'} \phi_k(s',i')q(i'|s') &= \frac{\sum_s \sum_i \sum_{i'} \phi_k(s',i')\{i'a(s',i'|s,i)/i\}q(s,i)}{\sum_s \sum_i \sum_{i''} \{i''a(s',i''|s,i)/i\}q(s,i)} \\ &\underset{\text{eqs.(10, 11)}}{=} \frac{\sum_s \sum_i w_k(s'|s,i)q(s,i)}{\sum_s \sum_i w_1(s'|s,i)q(s,i)} \\ &\underset{\text{eq.(5)}}{=} \frac{\sum_s \sum_i w_k(s'|s,i)q(i|s)q(s)}{\sum_s \sum_i w_1(s'|s,i)q(i|s)q(s)}, \end{aligned} \tag{C11}$$

where w_k denotes k-fitness as defined in eq.(11) in the main text.

C.2 Perturbations of $q(s)$ and $r_k(s)$

Here, we derive recursions satisfied by the zeroth- and first-order perturbations of $q(s)$ and $r_k(s)$. These will be of practical use when computing the selection gradient and disruptive selection coefficient, *e.g.*, eq.(21) and eq.(22) or eq.(32) and eq.(34). We also show that the recursions for the perturbation of $r_k(s)$ can be greatly simplified if we make some additional assumptions on the fitness functions.

C.2.1 Perturbation of $q(s)$

The following observation is useful in later calculations. For the perturbations of eqs.(B1b) and (B1c) we obtain

$$\sum_s q^{(0)}(s) = 1, \quad \sum_s q^{(1)}(s) = 0, \quad \sum_i q^{(0)}(i|s) = 1, \quad \sum_i q^{(1)}(i|s) = 0. \tag{C12}$$

Zeroth-order perturbation of $q(s)$: The zeroth-order perturbation of eq.(C9) with respect to δ is given by

$$\begin{aligned} q^{(0)}(s') &= \underbrace{\frac{1}{\rho^{(0)}}}_{=1} \left\{ \sum_s \sum_i w_1(s'|s,i) q(i|s) q(s) \right\}^{(0)} \\ &= \sum_s \sum_i \underbrace{w_1^{(0)}(s'|s,i)}_{\substack{=w_1^{(0)}(s'|s) \\ \text{(from eq.(B1a))}}} q^{(0)}(i|s) q^{(0)}(s) \\ &= \sum_s w_1^{(0)}(s'|s) \underbrace{\sum_i q^{(0)}(i|s)}_{=1\,\text{(from eq.(C12))}} q^{(0)}(s) \\ &= \sum_s w_1^{(0)}(s'|s) q^{(0)}(s). \end{aligned} \tag{C13}$$

This proves eq.(15a) in the main text.

First-order perturbation of $q(s)$: Assuming $\rho^{(1)} = 0$ and using the quotient rule the first-order perturbation of eq.(C9) with respect to δ is given by

$$\begin{aligned} q^{(1)}(s') &= \underbrace{\frac{1}{\rho^{(0)}}}_{=1} \left\{ \sum_s \sum_i w_1(s'|s,i) q(i|s) q(s) \right\}^{(1)} - \frac{1}{\left(\rho^{(0)}\right)^2} \left\{ \sum_s \sum_i w_1(s'|s,i) q(i|s) q(s) \right\}^{(0)} \underbrace{\rho^{(1)}}_{=0} \\ &= \sum_s \sum_i w_1^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s) \\ &\quad + \sum_s \sum_i \underbrace{w_1^{(0)}(s'|s,i)}_{\substack{=w_1^{(0)}(s'|s) \\ \text{(from eq.(B1a))}}} q^{(1)}(i|s) q^{(0)}(s) + \sum_s \sum_i \underbrace{w_1^{(0)}(s'|s,i)}_{\substack{=w_1^{(0)}(s'|s) \\ \text{(from eq.(B1a))}}} q^{(0)}(i|s) q^{(1)}(s) \\ &= \sum_s \sum_i w_1^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s) \\ &\quad + \sum_s w_1^{(0)}(s'|s) \underbrace{\sum_i q^{(1)}(i|s)}_{\substack{=0 \\ \text{(from eq.(C12))}}} q^{(0)}(s) + \sum_s w_1^{(0)}(s'|s) \underbrace{\sum_i q^{(0)}(i|s)}_{\substack{=1 \\ \text{(from eq.(C12))}}} q^{(1)}(s) \\ &= \sum_s \sum_i w_1^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s) + \sum_s w_1^{(0)}(s'|s) q^{(1)}(s). \end{aligned} \tag{C14}$$

C.2.2 Perturbation of $r_k(s)$

Zeroth-order perturbation of $r_k(s)$: With respect to the zeroth-order perturbation of eq.(C11) with respect to δ we obtain

$$r_k^{(0)}(s') = \frac{\{\sum_s \sum_i w_k(s'|s,i) q(i|s) q(s)\}^{(0)}}{\{\sum_s \sum_i w_1(s'|s,i) q(i|s) q(s)\}^{(0)}} \underset{\text{eq.(C13)}}{=} \frac{1}{q^{(0)}(s')} \sum_s \sum_i w_k^{(0)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s). \tag{C15}$$

Of practical importance for $k \geq 2$ is the case that $w_k^{(0)}(s'|s,i) = 0$ holds for all $s' \neq s$ and all $i = 1, \cdots, n_s$. This applies, for example, when the state of a given group does not change and mutants settle in new groups only as single individuals (no propagule dispersal). Then eq.(C15) simplifies to

$$r_k^{(0)}(s) = \sum_i w_k^{(0)}(s|s,i) q^{(0)}(i|s), \tag{C16}$$

and $r_k^{(0)}(s)$ can be calculated independently of $\{q^{(0)}(s)\}_{s\in\mathcal{S}}$.

First-order perturbation of $r_k(s)$: Assuming $\rho^{(1)} = 0$ and using the quotient rule the first-order perturbation of eq.(C11) with respect to δ equals

$$\begin{aligned} r_k^{(1)}(s') = & \frac{\{\sum_s \sum_i w_k(s'|s,i)q(i|s)q(s)\}^{(1)}}{\{\sum_s \sum_i w_1(s'|s,i)q(i|s)q(s)\}^{(0)}} \\ & - \underbrace{\frac{\{\sum_s \sum_i w_k(s'|s,i)q(i|s)q(s)\}^{(0)}}{\{\sum_s \sum_i w_1(s'|s,i)q(i|s)q(s)\}^{(0)}}}_{=r_k^{(0)}(s') \text{ (from eq.(C15))}} \cdot \frac{\{\sum_s \sum_i w_1(s'|s,i)q(i|s)q(s)\}^{(1)}}{\{\sum_s \sum_i w_1(s'|s,i)q(i|s)q(s)\}^{(0)}}. \end{aligned} \tag{C17}$$

For $\rho^{(1)} = 0$ we observed in eq.(C14) that $\{\sum_s \sum_i w_1(s'|s,i)q(i|s)q(s)\}^{(1)} = q^{(1)}(s')$ holds. Applying this relation and eq.(C13) we obtain

$$r_k^{(1)}(s') = \frac{1}{q^{(0)}(s')} \left\{ \sum_s \sum_i w_k(s'|s,i)q(i|s)q(s) \right\}^{(1)} - r_k^{(0)}(s') \frac{q^{(1)}(s')}{q^{(0)}(s')}, \tag{C18}$$

which, upon expanding the first-order perturbation on the right hand side more explicitly, becomes

$$\begin{aligned} r_k^{(1)}(s') = \frac{1}{q^{(0)}(s')} \Bigg[& \left\{ \sum_s \sum_i w_k^{(1)}(s'|s,i) q^{(0)}(i|s) q^{(0)}(s) \right\} \\ & + \left\{ \sum_s \sum_i w_k^{(0)}(s'|s,i) q^{(1)}(i|s) q^{(0)}(s) \right\} \\ & + \left\{ \sum_s \sum_i w_k^{(0)}(s'|s,i) q^{(0)}(i|s) q^{(1)}(s) \right\} \Bigg] - r_k^{(0)}(s') \frac{q^{(1)}(s')}{q^{(0)}(s')}. \end{aligned} \tag{C19}$$

When considering the case that $w_k^{(0)}(s'|s,i) = 0$ holds for all $s' \neq s$ and all $i = 1, \cdots, n_s$, just as we did for the case of the zeroth-order perturbation, eq.(C19) simplifies to

$$\begin{aligned} r_k^{(1)}(s) = & \sum_i w_k^{(1)}(s|s,i) q^{(0)}(i|s) + \sum_i w_k^{(0)}(s|s,i) q^{(1)}(i|s) \\ & + \underbrace{\left\{ \sum_i w_k^{(0)}(s|s,i) q^{(0)}(i|s) \right\}}_{=r_k^{(0)}(s) \text{ (from eq.(C16))}} \frac{q^{(1)}(s)}{q^{(0)}(s)} - r_k^{(0)}(s) \frac{q^{(1)}(s)}{q^{(0)}(s)} \\ = & \sum_i w_k^{(1)}(s|s,i) q^{(0)}(i|s) + \sum_i w_k^{(0)}(s|s,i) q^{(1)}(i|s), \end{aligned} \tag{C20}$$

and $r_k^{(1)}(s)$ can be calculated independently of $\{q^{(0)}(s)\}_{s\in\mathcal{S}}$ and $\{q^{(1)}(s)\}_{s\in\mathcal{S}}$.

C.3 Closedness of the recursions in $q(s)$ and $r_k(s)$

In the previous sections, we obtained recursions for $q(s)$, $r_k(s)$ and the perturbations thereof to the first order. However, it is not clear whether these actually form a closed system of equations in terms of

the variables. The purpose of this section is to show that eqs.(C7) and (C11) (and the perturbation thereof) indeed constitute such a closed system of equations. To do this, we pay attention to the sum $\sum_i w_k(s'|s,i)q(i|s)$, which frequently appears in eqs.(C7) and (C11). We prove that this sum can be written as a linear combination of the n_s relatedness coefficients $r_1(s), \cdots, r_{n_s}(s)$. With this, it follows that the desired result indeed holds.

Proof: For fixed k, s' and s (and for fixed x and y), consider n_s distinct points on a two-dimensional plane, $(i, w_k(s'|s,i)) \in \mathbb{R}^2$ $(i = 1, \ldots, n_s)$. Then, from a standard result of polynomial interpolation, there exists a unique polynomial in ξ, denoted by $L_{k,s',s}(\xi)$ (called a Lagrange polynomial), whose graph $\{(\xi, L_{k,s',s}(\xi)) \mid \xi \in \mathbb{R}\}$ passes through the above n_s points and whose order as a polynomial in ξ is equal to or less than $n_s - 1$. Take such a polynomial $L_{k,s',s}$. By definition

$$w_k(s'|s,i) = L_{k,s',s}(i) \quad (i = 1, \ldots, n_s) \tag{C21}$$

holds. Now we define a set of polynomials in ξ, $\{\Phi_{1,s}(\xi), \cdots, \Phi_{n_s,s}(\xi)\}$, as

$$\Phi_{\ell,s}(\xi) = \prod_{j=1}^{\ell-1} \frac{\xi - j}{n_s - j} \qquad (1 \le \ell \le n_s,\ s \in \mathcal{S}), \tag{C22}$$

where we define $\Phi_{1,s}(\xi) = 1$. This set of n_s polynomials of order 0 to $n_s - 1$ can be written more explicitly as

$$\left\{1, \frac{\xi - 1}{n_s - 1}, \frac{(\xi-1)(\xi-2)}{(n_s-1)(n_s-2)}, \cdots, \frac{(\xi-1)(\xi-2)\cdots(\xi-n_s+1)}{(n_s-1)(n_s-2)\cdots 1}\right\}. \tag{C23}$$

It is thus a basis of the vector space composed of all polynomials in ξ of order equal to or less than $n_s - 1$. Because $L_{k,s',s}(\xi)$ is one such polynomial it can be written as

$$L_{k,s',s}(\xi) = \sum_{\ell=1}^{n_s} a_{\ell,k,s',s} \Phi_{\ell,s}(\xi) \tag{C24}$$

for some $a_{\ell,k,s',s} \in \mathbb{R}$ $(\ell = 1, \cdots, n_s)$. By construction

$$\Phi_{\ell,s}(i) = \phi_\ell(s,i) \qquad (i = 1, \cdots, n_s) \tag{C25}$$

holds (compare eq.(C22) with eq.(6)).

We now consider $\sum_i w_k(s'|s,i)q(i|s)$. Using $L_{k,s',s}$ and eqs.(C21),(C24) and (C25) we obtain

$$\begin{aligned}
\sum_{i=1}^{n_s} w_k(s'|s,i)q(i|s) &\underset{\text{eq.(C21)}}{=} \sum_{i=1}^{n_s} L_{k,s',s}(i)q(i|s) \\
&\underset{\text{eq.(C24)}}{=} \sum_{i=1}^{n_s}\sum_{\ell=1}^{n_s} a_{\ell,k,s',s}\Phi_{\ell,s}(i)q(i|s) \\
&\underset{\text{eq.(C25)}}{=} \sum_{i=1}^{n_s}\sum_{\ell=1}^{n_s} a_{\ell,k,s',s}\phi_\ell(s,i)q(i|s) \\
&= \sum_{\ell=1}^{n_s} a_{\ell,k,s',s} \sum_{i=1}^{n_s} \phi_\ell(s,i)q(i|s) \\
&\underset{\text{eq.(7)}}{=} \sum_{\ell=1}^{n_s} a_{\ell,k,s',s} r_\ell(s).
\end{aligned} \tag{C26}$$

Given $r_1(s) = 1$ for all $s \in \mathcal{S}$, eqs.(C7) and (C11) form a large but closed system of equations with $q(s)$ and $r_k(s)$ for all $k = 2, \cdots, n_s$ and all $s \in \mathcal{S}$. Its size is $N + \sum_{s\in\mathcal{S}}(n_s - 1) = N + (n - N) = n$, which makes sense since n is the dimension of matrix $\boldsymbol{A}$. □

D Individual 3-fitness

We define the following four different individual 3-fitness functions. For that purpose, consider a focal individual in a group in state s who adopts z_1, and its $n_s - 1$ neighbors who adopt $\boldsymbol{z}_{-\{1\}}$ in an otherwise monomorphic population for z.

Individual 3-fitness of type I: Define

$$w^{\mathrm{I}}_{3,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) \tag{D1}$$

as the expected number of offspring in state s' that descend from the focal individual adopting z_1 in state s and that have two random neighbors (sampled without replacement) that both descend from the focal individual.

Individual 3-fitness of type II: Consider one of the focal's neighbor, called the target individual, who adopts z_2. We define

$$w^{\mathrm{II}}_{3,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z) \tag{D2}$$

as the expected number of offspring in state s' that descend from the focal individual adopting z_1 in state s and whose two random neighbors (sampled without replacement) both descend from the target individual (adopting z_2).

Individual 3-fitness of type II': Consider one of the focal's neighbor, called the target individual, who adopts z_2. We define

$$w^{\mathrm{II'}}_{3,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z) \tag{D3}$$

as the expected number of offspring in state s' that descend from the focal individual adopting z_1 in state s, and where one of its two random neighbors (sampled without replacement) descends from the focal individual while the other descends from the target individual (adopting trait value z_2). The following useful symmetry between $w^{\mathrm{II'}}_{3,s'|s}$ and $w^{\mathrm{II}}_{3,s'|s}$ holds:

$$w^{\mathrm{II'}}_{3,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z) = 2w^{\mathrm{II}}_{3,s'|s}(z_2, z_1, \boldsymbol{z}_{-\{1,2\}}, z). \tag{D4}$$

The reason for this symmetry is as follows. Consider a focal individual adopting z_1 and a target individual adopting z_2 in the same state-s group. Suppose that a group in state s' at the next time step comprises A_1 individuals that descend from the focal individual and A_2 individuals that descend from the target individual. Then, by definition such a group contributes to the 3-fitness of type II' of the focal individual $w^{\mathrm{II'}}_{3,s'|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z)$ by

$$A_1 \cdot \frac{2(A_1 - 1)A_2}{(n_{s'} - 1)(n_{s'} - 2)}. \tag{D5}$$

The same group contributes to the 3-fitness of type II of the target individual, $w^{\mathrm{II}}_{3,s'|s}(z_2, z_1, \boldsymbol{z}_{-\{1,2\}}, z)$, by

$$A_2 \cdot \frac{A_1(A_1 - 1)}{(n_{s'} - 1)(n_{s'} - 2)}. \tag{D6}$$

Equation (D5) is twice as large as eq.(D6), and therefore the symmetry eq.(D4) holds.

Individual 3-fitness of type III: Consider two neighbors of the focal individual, the one who adopts z_2 (called the first target individual) and the one who adopts z_3 (called the second target individual). Then we define

$$w^{\mathrm{III}}_{3,s'|s}(z_1, z_2, z_3, \boldsymbol{z}_{-\{1,2,3\}}, z) \tag{D7}$$

as the expected number of offspring in state s' that descend from the focal individual adopting z_1 in state s, and where one of its two random neighbors (sampled without replacement) descends from the first target individual (adopting z_2) and the other descends from the second target individual (adopting z_3). The following symmetry exists for a similar reason as above. For any permutation of σ of the set $\{1, 2, 3\}$ holds

$$w^{\mathrm{III}}_{3,s'|s}(z_{\sigma(1)}, z_{\sigma(2)}, z_{\sigma(3)}, \boldsymbol{z}_{-\{1,2,3\}}, z) = w^{\mathrm{III}}_{3,s'|s}(z_1, z_2, z_3, \boldsymbol{z}_{-\{1,2,3\}}, z). \tag{D8}$$

Calculation of w_3: With these four individual 3-fitness functions the individual 3-fitness of a mutant in an (s, i)-group, for $3 \le i \le n_s$, can be written as

$$\begin{aligned}
&w_3(s'|s,i) \\
&= w^{\mathrm{I}}_{3,s'|s}(x, \underbrace{x, \cdots, x}_{i-1}, \underbrace{y, \cdots, y}_{n_s-i}, y) + (i-1) w^{\mathrm{II}}_{3,s'|s}(x, x, \underbrace{x, \cdots, x}_{i-2}, \underbrace{y, \cdots, y}_{n_s-i}, y) \\
&\quad + (i-1) w^{\mathrm{II'}}_{3,s'|s}(x, x, \underbrace{x, \cdots, x}_{i-2}, \underbrace{y, \cdots, y}_{n_s-i}, y) + \frac{(i-1)(i-2)}{2} w^{\mathrm{III}}_{3,s'|s}(x, x, x, \underbrace{x, \cdots, x}_{i-3}, \underbrace{y, \cdots, y}_{n_s-i}, y) \\
&\underset{\text{eq.(D4)}}{=} w^{\mathrm{I}}_{3,s'|s}(x, \underbrace{x, \cdots, x}_{i-1}, \underbrace{y, \cdots, y}_{n_s-i}, y) \\
&\quad + 3(i-1) w^{\mathrm{II}}_{3,s'|s}(x, x, \underbrace{x, \cdots, x}_{i-2}, \underbrace{y, \cdots, y}_{n_s-i}, y) + \frac{(i-1)(i-2)}{2} w^{\mathrm{III}}_{3,s'|s}(x, x, x, \underbrace{x, \cdots, x}_{i-3}, \underbrace{y, \cdots, y}_{n_s-i}, y).
\end{aligned} \tag{D9}$$

Its zeroth-order perturbation with respect to δ is

$$w_3^{(0)}(s'|s,i) = w^{\mathrm{I}}_{3,s'|s} + 3(i-1) w^{\mathrm{II}}_{3,s'|s} + \frac{(i-1)(i-2)}{2} w^{\mathrm{III}}_{3,s'|s}, \tag{D10}$$

where all 3-fitness functions w_3 are evaluated at $(y, \cdots, y)$. Equation (D10) is the expected number of offspring in state s' under neutrality that descend from a focal mutant individual in state s and have two random neighbors (sampled without replacement) that are both mutants.

Note that the derivation above assumed $3 \le i \le n_s$, but we can separately confirm that eq.(D10) is valid for any $1 \le i \le n_s$, because some ill-defined terms for $n_s = 1$ and 2 become nullified by the factors $i - 1$ and $i - 2$.

E Derivation of the quantities with individual fitness functions

Here, we derive all the key results presented in section 3.3 in the main text.

Equation (33)**:** By setting $k = 2$ in eq.(C15) and substituting $w_2^{(0)}$ with eq.(29a), we obtain

$$\begin{aligned}
r_2^{(0)}(s') &= \frac{1}{q^{(0)}(s')} \sum_s \sum_i \left[w^{\mathrm{I}}_{2,s'|s} + (i-1) w^{\mathrm{II}}_{2,s'|s} \right] q^{(0)}(i|s) q^{(0)}(s) \\
&\underset{\text{eqs.(8,C12)}}{=} \frac{1}{q^{(0)}(s')} \sum_s \left[w^{\mathrm{I}}_{2,s'|s} + (n_s - 1) w^{\mathrm{II}}_{2,s'|s} r_2^{(0)}(s) \right] q^{(0)}(s).
\end{aligned} \tag{E1}$$

This proves eq.(33).

If $w^{\mathrm{I}}_{2,s'|s} = w^{\mathrm{II}}_{2,s'|s} = 0$ for $s' \neq s$ (this is the case, for example, when propagule dispersal is not allowed), we can use eq.(C16) instead of eq.(C15). Substituting eq.(29a) in eq.(C16) for $k = 2$ gives

$$\begin{aligned} r^{(0)}_2(s) &= \sum_i \left[w^{\mathrm{I}}_{2,s|s} + (i-1) w^{\mathrm{II}}_{2,s|s} \right] q^{(0)}(i|s) \\ &\underset{\text{eqs.(8,C12)}}{=} w^{\mathrm{I}}_{2,s|s} + (n_s - 1) w^{\mathrm{II}}_{2,s|s} r^{(0)}_2(s). \end{aligned} \tag{E2}$$

Equation (34)**:** First, substituting $w^{(2)}_1$ in eq.(22b) with eq.(25c) gives

$$\begin{aligned} \rho^{(2\mathrm{w})} &= \frac{1}{2} \sum_{s'} \sum_s \sum_i v^{(0)}(s') \Bigg[\frac{\partial^2 w_{1,s'|s}}{\partial z_1^2} + (i-1) \frac{\partial^2 w_{1,s'|s}}{\partial z_2^2} \\ &\qquad + 2(i-1) \frac{\partial^2 w_{1,s'|s}}{\partial z_1 \partial z_2} + (i-1)(i-2) \frac{\partial^2 w_{1,s'|s}}{\partial z_2 \partial z_3} \Bigg] q^{(0)}(i|s) q^{(0)}(s) \\ &\underset{\text{eqs.(8,9,C12)}}{=} \frac{1}{2} \sum_{s'} \sum_s v^{(0)}(s') \Bigg[\frac{\partial^2 w_{1,s'|s}}{\partial z_1^2} + (n_s - 1) \frac{\partial^2 w_{1,s'|s}}{\partial z_2^2} r^{(0)}_2(s) \\ &\qquad + 2(n_s - 1) \frac{\partial^2 w_{1,s'|s}}{\partial z_1 \partial z_2} r^{(0)}_2(s) + (n_s - 1)(n_s - 2) \frac{\partial^2 w_{1,s'|s}}{\partial z_2 \partial z_3} r^{(0)}_3(s) \Bigg] q^{(0)}(s). \end{aligned} \tag{E3a}$$

Second, substituting $w^{(1)}_1$ in eq.(22c) with eq.(25b) gives

$$\begin{aligned} \rho^{(2\mathrm{q})} &= \sum_{s'} \sum_s \sum_i v^{(0)}(s') \left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (i-1) \frac{\partial w_{1,s'|s}}{\partial z_2} \right] q^{(0)}(i|s) q^{(1)}(s) \\ &\underset{\text{eqs.(8,C12)}}{=} \sum_{s'} \sum_s v^{(0)}(s') \left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (n_s - 1) \frac{\partial w_{1,s'|s}}{\partial z_2} r^{(0)}_2(s) \right] q^{(1)}(s). \end{aligned} \tag{E3b}$$

Third, substituting $w^{(1)}_1$ in eq.(22d) with eq.(25b) gives

$$\begin{aligned} \rho^{(2\mathrm{r})} &= \sum_{s'} \sum_s \sum_i v^{(0)}(s') \left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (i-1) \frac{\partial w_{1,s'|s}}{\partial z_2} \right] q^{(1)}(i|s) q^{(0)}(s) \\ &\underset{\text{eqs.(8,C12)}}{=} \sum_{s'} \sum_s v^{(0)}(s') \left[(n_s - 1) \frac{\partial w_{1,s'|s}}{\partial z_2} r^{(1)}_2(s) \right] q^{(0)}(s). \end{aligned} \tag{E3c}$$

Equation (35)**:** Substituting $w^{(0)}_1$ and $w^{(1)}_1$ in eq.(C14) with eq.(25a) and eq.(25b), respectively, gives

$$\begin{aligned} q^{(1)}(s') &= \sum_s \sum_i \left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (i-1) \frac{\partial w_{1,s'|s}}{\partial z_2} \right] q^{(0)}(i|s) q^{(0)}(s) + \sum_s w_{1,s'|s} q^{(1)}(s) \\ &\underset{\text{eqs.(8,C12)}}{=} \sum_s \left[\frac{\partial w_{1,s'|s}}{\partial z_1} + (n_s - 1) \frac{\partial w_{1,s'|s}}{\partial z_2} r^{(0)}_2(s) \right] q^{(0)}(s) + \sum_s w_{1,s'|s} q^{(1)}(s). \end{aligned} \tag{E4}$$

Equation (36)**:** By setting $k = 3$ in eq.(C15) and substituting $w^{(0)}_3$ with eq.(30) we obtain

$$\begin{aligned} r^{(0)}_3(s') &= \frac{1}{q^{(0)}(s')} \sum_s \sum_i \left[w^{\mathrm{I}}_{3,s'|s} + 3(i-1) w^{\mathrm{II}}_{3,s'|s} + \frac{(i-1)(i-2)}{2} w^{\mathrm{III}}_{3,s'|s} \right] q^{(0)}(i|s) q^{(0)}(s) \\ &\underset{\text{eqs.(8,9,C12)}}{=} \frac{1}{q^{(0)}(s')} \sum_s \left[w^{\mathrm{I}}_{3,s'|s} + 3(n_s - 1) w^{\mathrm{II}}_{3,s'|s} r^{(0)}_2(s) + \frac{(n_s-1)(n_s-2)}{2} w^{\mathrm{III}}_{3,s'|s} r^{(0)}_3(s) \right] q^{(0)}(s). \end{aligned} \tag{E5}$$

This proves eq.(36) in the main text.

If $w^{\mathrm{I}}_{3,s'|s} = w^{\mathrm{II}}_{3,s'|s} = w^{\mathrm{III}}_{3,s'|s} = 0$ for $s' \neq s$ (this is the case, for example, when propagule dispersal is not allowed), we can use eq.(C16) instead of eq.(C15). Substituting eq.(30) in eq.(C16) for $k = 3$ then gives us

$$\begin{aligned} r^{(0)}_3(s) &= \sum_i \left[w^{\mathrm{I}}_{3,s|s} + 3(i-1) w^{\mathrm{II}}_{3,s|s} + \frac{(i-1)(i-2)}{2} w^{\mathrm{III}}_{3,s|s} \right] q^{(0)}(i|s) \\ &\underset{\text{eqs.(8,9,C12)}}{=} w^{\mathrm{I}}_{3,s|s} + 3(n_s-1) w^{\mathrm{II}}_{3,s|s} r^{(0)}_2(s) + \frac{(n_s-1)(n_s-2)}{2} w^{\mathrm{III}}_{3,s|s} r^{(0)}_3(s). \end{aligned} \tag{E6}$$

Equation (37)**:** By setting $k = 2$ in eq.(C19), substituting $w^{(0)}_2$ and $w^{(1)}_2$ with eq.(29a) and (29b), respectively, we obtain

$$\begin{aligned} r^{(1)}_2(s') &= \frac{1}{q^{(0)}(s')} \sum_s \sum_i \left[\frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_1} + (i-1) \frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_2} \right. \\ &\qquad \left. + 2(i-1) \frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_1} + (i-1)(i-2) \frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_3} \right] q^{(0)}(i|s) q^{(0)}(s) \\ &\quad + \frac{1}{q^{(0)}(s')} \sum_s \sum_i \left[w^{\mathrm{I}}_{2,s'|s} + (i-1) w^{\mathrm{II}}_{2,s'|s} \right] q^{(1)}(i|s) q^{(0)}(s) \\ &\quad + \frac{1}{q^{(0)}(s')} \sum_s \sum_i \left[w^{\mathrm{I}}_{2,s'|s} + (i-1) w^{\mathrm{II}}_{2,s'|s} \right] q^{(0)}(i|s) q^{(1)}(s) \\ &\quad - r^{(0)}_2(s') \frac{q^{(1)}(s')}{q^{(0)}(s')} \\ &\underset{\text{eqs.(8,9,C12)}}{=} \frac{1}{q^{(0)}(s')} \sum_s \left[\frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_1} + (n_s-1) \frac{\partial w^{\mathrm{I}}_{2,s'|s}}{\partial z_2} r^{(0)}_2(s) \right. \\ &\qquad \left. + 2(n_s-1) \frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_1} r^{(0)}_2(s) + (n_s-1)(n_s-2) \frac{\partial w^{\mathrm{II}}_{2,s'|s}}{\partial z_3} r^{(0)}_3(s) \right] q^{(0)}(s) \\ &\quad + \frac{1}{q^{(0)}(s')} \sum_s \left[(n_s-1) w^{\mathrm{II}}_{2,s'|s} r^{(1)}_2(s) \right] q^{(0)}(s) \\ &\quad + \frac{1}{q^{(0)}(s')} \sum_s \left[w^{\mathrm{I}}_{2,s'|s} + (n_s-1) w^{\mathrm{II}}_{2,s'|s} r^{(0)}_2(s) \right] q^{(1)}(s) \\ &\quad - r^{(0)}_2(s') \frac{q^{(1)}(s')}{q^{(0)}(s')}. \end{aligned} \tag{E7}$$

This proves eq.(37) in the main text.

If $w^{\mathrm{I}}_{2,s'|s} = w^{\mathrm{II}}_{2,s'|s} = 0$ for $s' \neq s$ (this is the case, for example, when propagule dispersal is not allowed), we can use eq.(C20) instead of eq.(C19). Substituting eq.(29) in eq.(C20) for $k = 2$ gives us

$$\begin{aligned} r^{(1)}_2(s) &= \sum_i \left[\frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_1} + (i-1) \frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_2} + 2(i-1) \frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_1} + (i-1)(i-2) \frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_3} \right] q^{(0)}(i|s) \\ &\quad + \sum_i \left[w^{\mathrm{I}}_{2,s|s} + (i-1) w^{\mathrm{II}}_{2,s|s} \right] q^{(1)}(i|s) \\ &= \frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_1} + (n_s-1) \frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_2} r^{(0)}_2(s) + 2(n_s-1) \frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_1} r^{(0)}_2(s) + (n_s-1)(n_s-2) \frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_3} r^{(0)}_3(s) \\ &\quad + (n_s-1) w^{\mathrm{II}}_{2,s|s} r^{(1)}_2(s). \end{aligned} \tag{E8}$$

Therefore, $r_2^{(1)}(s)$ is solved as

$$r_2^{(1)}(s) = \left\{1-(n_s-1)w^{\mathrm{II}}_{2,s|s}\right\}^{-1} \times \left[\frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_1} + (n_s-1)\frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_2} r_2^{(0)}(s) + 2(n_s-1)\frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_1} r_2^{(0)}(s) + (n_s-1)(n_s-2)\frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_3} r_3^{(0)}(s)\right]. \tag{E9}$$

From eq.(E2) we see that $\{1-(n_s-1)w^{\mathrm{II}}_{2,s|s}\}^{-1} = r_2^{(0)}(s)/w^{\mathrm{I}}_{2,s|s}$ holds, so eq.(E9) can also be written as

$$r_2^{(1)}(s) = \frac{r_2^{(0)}(s)}{w^{\mathrm{I}}_{2,s|s}} \times \left[\frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_1} + (n_s-1)\frac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_2} r_2^{(0)}(s) + 2(n_s-1)\frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_1} r_2^{(0)}(s) + (n_s-1)(n_s-2)\frac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_3} r_3^{(0)}(s)\right]. \tag{E10}$$

F Derivation of the results for the lottery model

Here, we derive all the results presented in section 4.1 of the main text.

F.1 Calculations for $q(s)$ and $v^{(0)}(s)$

Under the assumptions (i)-(iii) described in section 4.1 of the main text, eq.(38c) therein can be further decomposed as

$$\begin{aligned} w^{\mathrm{d}}_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) &= \underbrace{(1-\gamma_{s'})w^{\mathrm{dr}}_{1,s'|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)}_{\text{allopatric, reproduction}} \\ &= \underbrace{(1-\gamma_{s'})w^{\mathrm{col}}_{s'}(z)}_{\text{colonization success}} \cdot \underbrace{w^{\mathrm{em}}_{s}(z_1, \boldsymbol{z}_{-\{1\}}, z)}_{\text{emigrant production}}, \end{aligned} \tag{F1}$$

where $w^{\mathrm{col}}_{s'}$ represents a component of colonization success of dispersing offspring arriving in groups in state s', w^{em}_{s} represents the emigrant production in groups in state s, and $w^{\mathrm{dr}}_{1,s'|s}$ is given as a product of these two terms. This multiplicative decomposition is a key property that greatly simplifies the following analysis. It follows from the fact that when dispersal occurs individually and independently to a random destination, the production of emigrants in a group in habitat s does not depend on the habitat s' of the destination group and the colonization success depends only on the habitat of the destination group and the resident trait value.

Before proceeding to the derivation, we show that

$$w_{1,s|s} < 1, \quad w^{\mathrm{p}}_{1,s|s} < 1, \quad w^{\mathrm{pr}}_{1,s|s} < 1 \tag{F2}$$

hold for all $s \in \mathcal{S}$, because we will frequently use these facts without a particular notice below. The proof starts from the observation that eq.(15a) is rewritten as

$$(1-w_{1,s'|s'})q^{(0)}(s') = \sum_{s\in\mathcal{S}, s\neq s'} w_{1,s'|s} q^{(0)}(s). \tag{F3}$$

Remember that $q^{(0)}(s) > 0$ holds for all $s \in \mathcal{S}$, and therefore the right hand side of eq.(F3) is non-negative. Suppose the right hand side of eq.(F3) is zero. Then we have $w_{1,s'|s} = 0$ for all $s \neq s'$, which

contradicts that matrix $\boldsymbol{W}^{(0)}$ is primitive (see section 2.3.4). Therefore, the right hand side of eq.(F3) must be strictly positive. Because $q^{(0)}(s') > 0$, it follows that $w_{1,s'|s'} < 1$ holds, and this argument is valid for all $s' \in \mathcal{S}$. Second, because $w_{1,s|s} = w^{\mathrm{p}}_{1,s|s} + w^{\mathrm{d}}_{1,s|s} < 1$, we have $w^{\mathrm{p}}_{1,s|s} < 1$. Third, from eq.(38b), one can see that the relation

$$1 - w^{\mathrm{p}}_{1,s|s} = (1-\gamma_s)(1 - w^{\mathrm{pr}}_{1,s|s}) \tag{F4}$$

holds. Because we have just proven that $w^{\mathrm{p}}_{1,s|s} < 1$ and because $\gamma_s < 1$, one concludes that $w^{\mathrm{pr}}_{1,s|s} < 1$ holds. End of the proof.

We now first calculate explicitly $q^{(0)}(s)$ and $v^{(0)}(s)$. For that purpose, the vector-matrix notation in eqs.(15a) and (15b) is helpful. In fact, from eqs.(38) and (F1), the fitness under neutrality is written as

$$\boldsymbol{W}^{(0)} = \underbrace{\begin{pmatrix} w^{\mathrm{p}}_{1,s_1|s_1} & & 0 \\ & \ddots & \\ 0 & & w^{\mathrm{p}}_{1,s_N|s_N} \end{pmatrix}}_{\text{diagonal matrix}\,(N\times N)} + \underbrace{\begin{pmatrix} (1-\gamma_{s_1})w^{\mathrm{col}}_{s_1} \\ \vdots \\ (1-\gamma_{s_N})w^{\mathrm{col}}_{s_N} \end{pmatrix}}_{N\times 1} \underbrace{\begin{pmatrix} w^{\mathrm{em}}_{s_1} & \cdots & w^{\mathrm{em}}_{s_N} \end{pmatrix}}_{1\times N}. \tag{F5}$$

Note that w^{col}_s's and $w^{\mathrm{em}}_{s_1}$'s without variables are evaluated respectively at y and $(y, \cdots, y)$ here and thereafter. To solve eq.(15a), we right-multiply eq.(F5) by the column vector $\boldsymbol{q}^{(0)}$ and obtain

$$\begin{aligned}
\boldsymbol{q}^{(0)} &= \begin{pmatrix} w^{\mathrm{p}}_{1,s_1|s_1} & & 0 \\ & \ddots & \\ 0 & & w^{\mathrm{p}}_{1,s_N|s_N} \end{pmatrix} \boldsymbol{q}^{(0)} + \begin{pmatrix} (1-\gamma_{s_1})w^{\mathrm{col}}_{s_1} \\ \vdots \\ (1-\gamma_{s_N})w^{\mathrm{col}}_{s_N} \end{pmatrix} \underbrace{\begin{pmatrix} w^{\mathrm{em}}_{s_1} & \cdots & w^{\mathrm{em}}_{s_N} \end{pmatrix} \boldsymbol{q}^{(0)}}_{=\text{some constant}} \\
\Longrightarrow \quad & \begin{pmatrix} 1 - w^{\mathrm{p}}_{1,s_1|s_1} & & 0 \\ & \ddots & \\ 0 & & 1 - w^{\mathrm{p}}_{1,s_N|s_N} \end{pmatrix} \boldsymbol{q}^{(0)} \propto \begin{pmatrix} (1-\gamma_{s_1})w^{\mathrm{col}}_{s_1} \\ \vdots \\ (1-\gamma_{s_N})w^{\mathrm{col}}_{s_N} \end{pmatrix} \\
\Longrightarrow \quad & q^{(0)}(s) \propto \frac{(1-\gamma_s)w^{\mathrm{col}}_s}{1 - w^{\mathrm{p}}_{1,s|s}} \underset{\text{eq.(F4)}}{=} \frac{(1-\gamma_s)w^{\mathrm{col}}_s}{(1-\gamma_s)(1 - w^{\mathrm{pr}}_{1,s|s})} = \frac{w^{\mathrm{col}}_s}{1 - w^{\mathrm{pr}}_{1,s|s}}.
\end{aligned} \tag{F6}$$

To solve eq.(15b) we left-multiply eq.(F5) by the row vector $\boldsymbol{v}^{(0)}$ and obtain

$$\begin{aligned}
\boldsymbol{v}^{(0)} &= \boldsymbol{v}^{(0)} \begin{pmatrix} w^{\mathrm{p}}_{1,s_1|s_1} & & 0 \\ & \ddots & \\ 0 & & w^{\mathrm{p}}_{1,s_N|s_N} \end{pmatrix} + \underbrace{\boldsymbol{v}^{(0)} \begin{pmatrix} (1-\gamma_{s_1})w^{\mathrm{col}}_{s_1} \\ \vdots \\ (1-\gamma_{s_N})w^{\mathrm{col}}_{s_N} \end{pmatrix}}_{=\text{some constant}} \begin{pmatrix} w^{\mathrm{em}}_{s_1} & \cdots & w^{\mathrm{em}}_{s_N} \end{pmatrix} \\
\Longrightarrow \quad & \boldsymbol{v}^{(0)} \begin{pmatrix} 1 - w^{\mathrm{p}}_{1,s_1|s_1} & & 0 \\ & \ddots & \\ 0 & & 1 - w^{\mathrm{p}}_{1,s_N|s_N} \end{pmatrix} \propto \begin{pmatrix} w^{\mathrm{em}}_{s_1} & \cdots & w^{\mathrm{em}}_{s_N} \end{pmatrix} \\
\Longrightarrow \quad & v^{(0)}(s) \propto \frac{w^{\mathrm{em}}_s}{1 - w^{\mathrm{p}}_{1,s|s}} \underset{\text{eq.(F4)}}{=} \frac{w^{\mathrm{em}}_s}{(1-\gamma_s)(1 - w^{\mathrm{pr}}_{1,s|s})}.
\end{aligned} \tag{F7}$$

We normalize $q^{(0)}(s)$ and $v^{(0)}(s)$ to satisfy eq.(15c) and eq.(B1b) and obtain the following result:

$$q^{(0)}(s) = \frac{w^{\mathrm{col}}_s}{1 - w^{\mathrm{pr}}_{1,s|s}} \Bigg/ \left(\sum_{s'} \frac{w^{\mathrm{col}}_{s'}}{1 - w^{\mathrm{pr}}_{1,s'|s'}} \right) \tag{F8}$$

$$v^{(0)}(s) = \frac{w^{\mathrm{em}}_s}{(1-\gamma_s)(1 - w^{\mathrm{pr}}_{1,s|s})} \left(\sum_{s'} \frac{w^{\mathrm{col}}_{s'}}{1 - w^{\mathrm{pr}}_{1,s'|s'}} \right) \Bigg/ \left(\sum_{s'} \frac{w^{\mathrm{dr}}_{1,s'|s'}}{(1-\gamma_{s'})(1 - w^{\mathrm{pr}}_{1,s'|s'})^2} \right). \tag{F9}$$

Combining eqs.(F8) and (F9) gives eq.(40) in the main text. In particular, their product

$$v^{(0)}(s)q^{(0)}(s) = \frac{w^{\mathrm{dr}}_{1,s|s}}{(1-\gamma_s)(1 - w^{\mathrm{pr}}_{1,s|s})(1 - w^{\mathrm{pr}}_{1,s|s})} \Bigg/ \left(\sum_{s'} \frac{w^{\mathrm{dr}}_{1,s'|s'}}{(1-\gamma_{s'})(1 - w^{\mathrm{pr}}_{1,s'|s'})^2} \right), \tag{F10}$$

can be recognized as the the class reproductive value of a group in habitat s (Rousset, 2004).

We note that $q^{(0)}(s)$ can also be derived via a different pathway. To ease our understanding, let us temporarily consider a finite population that consists of N_{G} groups (later we will take the limit, $N_{\mathrm{G}} \to \infty$, because we actually consider an infinitely large population in this paper) and consider the neutral case. The total number of individuals in groups in habitat s' in a next time step is the sum of the number of individuals in habitat s' produced by individuals in habitat s over all $s \in \mathcal{S}$, which is

$$\sum_{s} w_{1,s'|s} N_{\mathrm{G}} \pi_s n_s. \tag{F11}$$

Here, $N_{\mathrm{G}}\pi_s$ is the total number of groups in habitat s, and $N_{\mathrm{G}}\pi_s n_s$ is the total number of individuals in groups in habitat s. However, eq.(F11) can also be written as $N_{\mathrm{G}}\pi_{s'} n_{s'}$, so equating these two quantities cancels out N_{G} and gives us

$$\pi_{s'} n_{s'} = \sum_{s} w_{1,s'|s} \pi_s n_s, \tag{F12}$$

and therefore eq.(F12) is valid for $N_{\mathrm{G}} \to \infty$ as well (see also eq.(E.21) of Lehmann et al., 2016). Equation (F12) suggests that the column vector $\{\pi_s n_s\}_{s\in\mathcal{S}}$ is an eigenvector of matrix $\boldsymbol{W}^{(0)}$ corresponding to the leading eigenvalue of 1. From the Perron-Frobenius theorem the eigenspace of the matrix $\boldsymbol{W}^{(0)}$ corresponding to the leading eigenvalue 1 is one-dimensional. Thus, from the uniqueness of the normalized eigenvector we obtain

$$q^{(0)}(s) = \frac{\pi_s n_s}{\sum_{s'} \pi_{s'} n_{s'}}. \tag{F13}$$

Note that eq.(F8) and eq.(F13) are both correct. Their equivalence suggests the existence of the following constraint in the choice of w_s^{col} and $1 - w_{1,s|s}^{\mathrm{pr}}$ in our spatial lottery model, namely, that there exists a positive constant C which is independent of s and

$$\frac{w_s^{\mathrm{col}}}{1 - w_{1,s|s}^{\mathrm{pr}}} = C \pi_s n_s \tag{F14}$$

is satisfied for all $s \in \mathcal{S}$.

Finally, we calculate $q^{(1)}(s)$ under the assumption that $\rho^{(1)} = 0$. From eq.(35), we then obtain

$$\begin{aligned}
q^{(1)}(s) &\underset{\text{eq.(38a)}}{=} \left[\frac{\partial w_{1,s|s}^{\mathrm{p}}}{\partial z_1} + (n_s - 1)\frac{\partial w_{1,s|s}^{\mathrm{p}}}{\partial z_2} r_2^{(0)}(s)\right] q^{(0)}(s) + w_{1,s|s}^{\mathrm{p}} q^{(1)}(s) \\
&\quad + \sum_{s'} \left[\frac{\partial w_{1,s|s'}^{\mathrm{d}}}{\partial z_1} + (n_{s'} - 1)\frac{\partial w_{1,s|s'}^{\mathrm{d}}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s') + \sum_{s'} w_{1,s|s'}^{\mathrm{d}} q^{(1)}(s') \\
&\underset{\text{eqs.(38b, F1)}}{=} (1-\gamma_s)\left[\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_1} + (n_s - 1)\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2} r_2^{(0)}(s)\right] q^{(0)}(s) + \{\gamma_s + (1-\gamma_s) w_{1,s|s}^{\mathrm{pr}}\} q^{(1)}(s) \\
&\quad + (1-\gamma_s) w_s^{\mathrm{col}} \left\{\sum_{s'} \left[\frac{\partial w_{s'}^{\mathrm{em}}}{\partial z_1} + (n_{s'} - 1)\frac{\partial w_{s'}^{\mathrm{em}}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s') + \sum_{s'} w_{s'}^{\mathrm{em}} q^{(1)}(s')\right\},
\end{aligned} \tag{F15}$$

which is implicitly solved as

$$\begin{aligned}
q^{(1)}(s) &= \frac{1}{1 - w_{1,s|s}^{\mathrm{pr}}} \left[\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_1} + (n_s - 1)\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2} r_2^{(0)}(s)\right] q^{(0)}(s) \\
&\quad + \frac{w_s^{\mathrm{col}}}{1 - w_{1,s|s}^{\mathrm{pr}}} \left\{\sum_{s'} \left[\frac{\partial w_{s'}^{\mathrm{em}}}{\partial z_1} + (n_{s'} - 1)\frac{\partial w_{s'}^{\mathrm{em}}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s') + \sum_{s'} w_{s'}^{\mathrm{em}} q^{(1)}(s')\right\}.
\end{aligned} \tag{F16}$$

Using $\sum_s q^{(1)}(s) = 0$ (see eq.(C12)), we have

$$\sum_s q^{(1)}(s) = \sum_s \frac{1}{1-w^{\rm pr}_{1,s|s}}\left[\frac{\partial w^{\rm pr}_{1,s|s}}{\partial z_1} + (n_s-1)\frac{\partial w^{\rm pr}_{1,s|s}}{\partial z_2} r_2^{(0)}(s)\right] q^{(0)}(s) + \left(\sum_s \frac{w^{\rm col}_s}{1-w^{\rm pr}_{1,s|s}}\right)\left\{\sum_{s'}\left[\frac{\partial w^{\rm em}_{s'}}{\partial z_1} + (n_{s'}-1)\frac{\partial w^{\rm em}_{s'}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s') + \sum_{s'} w^{\rm em}_{s'} q^{(1)}(s')\right\} = 0, \tag{F17}$$

which shows that

$$\sum_{s'}\left[\frac{\partial w^{\rm em}_{s'}}{\partial z_1} + (n_{s'}-1)\frac{\partial w^{\rm em}_{s'}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s') + \sum_{s'} w^{\rm em}_{s'} q^{(1)}(s') = -\left(\sum_{s'} \frac{w^{\rm col}_{s'}}{1-w^{\rm pr}_{1,s'|s'}}\right)^{-1}\left(\sum_{s'} \frac{1}{1-w^{\rm pr}_{1,s'|s'}}\left[\frac{\partial w^{\rm pr}_{1,s'|s'}}{\partial z_1} + (n_{s'}-1)\frac{\partial w^{\rm pr}_{1,s'|s'}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s')\right). \tag{F18}$$

holds. Putting eq.(F18) back into eq.(F16) and using eq.(F8) gives us

$$q^{(1)}(s) = \left\{\frac{1}{1-w^{\rm pr}_{1,s|s}}\left[\frac{\partial w^{\rm pr}_{1,s|s}}{\partial z_1} + (n_s-1)\frac{\partial w^{\rm pr}_{1,s|s}}{\partial z_2} r_2^{(0)}(s)\right] - \sum_{s'} \frac{1}{1-w^{\rm pr}_{1,s'|s'}}\left[\frac{\partial w^{\rm pr}_{1,s'|s'}}{\partial z_1} + (n_{s'}-1)\frac{\partial w^{\rm pr}_{1,s'|s'}}{\partial z_2} r_2^{(0)}(s')\right] q^{(0)}(s')\right\} q^{(0)}(s), \tag{F19}$$

which proves eq.(43) in the main text.

F.2 Individual 2-fitness and 3-fitness for the lottery model and calculations for relatedness

The purpose of this subsection is to derive pairwise relatedness and three-way relatedness under neutrality, as well as the first-order perturbation of pairwise relatedness under the assumption of the lottery model in section 4.1. For that purpose we will show that for this lottery model individual 2-fitness and individual 3-fitness are written in terms of the philopatric component of individual 1-fitness as in eq.(F23) and eq.(F31).

Let us label individuals in a focal group in habitat s from 1 to n_s, and let their trait values be z_1 to z_{n_s}, respectively. Let us also operationally define "positions" in this group, from "position 1" to "position n_s". If adult k survives (with occurs with probability γ_s) we operationally assume that he/she occupies "position k" in the next generation. If adult k dies (which occurs with probability $1-\gamma_s$) we assume that "position k" in the next generation becomes open to competition. In the latter case, adult i bears a descendant in this "position k" with probability $w^{\rm pr}_{1,s|s}(z_i, \boldsymbol{z}_{-\{i\}}, z)/n_s$, which we will write as ω_i in what follows as a short-hand notation.

With these, the probability $\zeta_{i,k}$ that an individual in position k in the next generation in the focal group "descends" from adult i in that group in the previous generation (thus including self through survival) is, for the lottery model, given by

$$\zeta_{i,k} \equiv \gamma_s \delta_{i,k} + (1-\gamma_s)\omega_i, \tag{F20}$$

where $\delta_{i,k}$ equals one if $i=k$ and otherwise zero. In eq.(F20), the first term represents the probability that adult i survives and occupies position k (thus equal to γ_s if $i=k$, zero otherwise), and the second term represents the probability that adult k dies, position k becomes open to competition, and a descendant of adult i occupies this position.

F.2.1 Pairwise relatedness

Then, same-parent individual 2-fitness of adult 1, which we take as a focal individual, denoted by $w^{\mathrm{I}}_{2,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)$ (recall eq.(26)), is written as

$$
\begin{aligned}
& w^{\mathrm{I}}_{2,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) \\
&= \sum_{k_1=1}^{n_s} \zeta_{1,k_1} \left(\frac{1}{n_s - 1} \sum_{\substack{k_2=1 \\ k_2 \neq k_1}}^{n_s} \zeta_{1,k_2} \right) \\
&= \frac{1}{n_s - 1} \sum_{\substack{1 \leq k_1, k_2 \leq n_s \\ k_1 \neq k_2}} \left[\gamma_s \delta_{1,k_1} + (1 - \gamma_s)\omega_1 \right] \left[\gamma_s \delta_{1,k_2} + (1 - \gamma_s)\omega_1 \right] \\
&= \frac{1}{n_s - 1} \Bigg[\gamma_s^2 \underbrace{\left(\sum_{\substack{1 \leq k_1, k_2 \leq n_s \\ k_1 \neq k_2}} \delta_{1,k_1} \delta_{1,k_2} \right)}_{=0} + \gamma_s (1 - \gamma_s) \underbrace{\left(\sum_{\substack{1 \leq k_1, k_2 \leq n_s \\ k_1 \neq k_2}} (\delta_{1,k_1} + \delta_{1,k_2}) \right)}_{=2(n_s - 1)} \omega_1 \\
&\qquad + (1 - \gamma_s)^2 \omega_1^2 \underbrace{\left(\sum_{\substack{1 \leq k_1, k_2 \leq n_s \\ k_1 \neq k_2}} 1 \right)}_{=n_s(n_s - 1)} \Bigg] \\
&= 2\gamma_s (1 - \gamma_s)\omega_1 + n_s (1 - \gamma_s)^2 \omega_1^2 \\
&= \frac{1}{n_s} \left\{ 2\gamma_s (1 - \gamma_s) \left[w^{\mathrm{pr}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) \right] + (1 - \gamma_s)^2 \left[w^{\mathrm{pr}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) \right]^2 \right\}.
\end{aligned}
\tag{F21}
$$

In a similar vein, different-parent individual 2-fitness is

$$
\begin{aligned}
&w^{\mathrm{II}}_{2,s|s}(z_1,z_2,\boldsymbol{z}_{-\{1,2\}},z)\\
&=\sum_{k_1=1}^{n_s}\zeta_{1,k_1}\left(\frac{1}{n_s-1}\sum_{\substack{k_2=1\\k_2\neq k_1}}^{n_s}\zeta_{2,k_2}\right)\\
&=\frac{1}{n_s-1}\sum_{\substack{1\le k_1,k_2\le n_s\\k_1\neq k_2}}\left[\gamma_s\delta_{1,k_1}+(1-\gamma_s)\omega_1\right]\left[\gamma_s\delta_{2,k_2}+(1-\gamma_s)\omega_2\right]\\
&=\frac{1}{n_s-1}\Bigg[\gamma_s^2\underbrace{\left(\sum_{\substack{1\le k_1,k_2\le n_s\\k_1\neq k_2}}\delta_{1,k_1}\delta_{2,k_2}\right)}_{=1}+\gamma_s(1-\gamma_s)\left(\underbrace{\left(\sum_{\substack{1\le k_1,k_2\le n_s\\k_1\neq k_2}}\delta_{1,k_1}\right)}_{=n_s-1}\omega_2+\underbrace{\left(\sum_{\substack{1\le k_1,k_2\le n_s\\k_1\neq k_2}}\delta_{2,k_2}\right)}_{=n_s-1}\omega_1\right)\\
&\qquad+(1-\gamma_s)^2\omega_1\omega_2\underbrace{\left(\sum_{\substack{1\le k_1,k_2\le n_s\\k_1\neq k_2}}1\right)}_{=n_s(n_s-1)}\Bigg]\\
&=\frac{\gamma_s^2}{n_s-1}+\gamma_s(1-\gamma_s)(\omega_1+\omega_2)+n_s(1-\gamma_s)^2\omega_1\omega_2\\
&=\frac{\gamma_s^2}{n_s-1}+\frac{1}{n_s}\Bigg\{\gamma_s(1-\gamma_s)\left(\left[w^{\mathrm{pr}}_{1,s|s}(z_1,\boldsymbol{z}_{-\{1\}},z)\right]+\left[w^{\mathrm{pr}}_{1,s|s}(z_2,\boldsymbol{z}_{-\{2\}},z)\right]\right)\\
&\qquad+(1-\gamma_s)^2\left[w^{\mathrm{pr}}_{1,s|s}(z_1,\boldsymbol{z}_{-\{1\}},z)\right]\left[w^{\mathrm{pr}}_{1,s|s}(z_2,\boldsymbol{z}_{-\{2\}},z)\right]\Bigg\}.
\end{aligned}
\tag{F22}
$$

In contrast, when $s'\neq s$, we have $w^{\mathrm{I}}_{2,s'|s}=0$ and $w^{\mathrm{II}}_{2,s'|s}=0$ from the assumption of independent dispersal. With these expressions, we obtain

$$
\begin{cases}
w^{\mathrm{I}}_{2,s|s}=\dfrac{1}{n_s}(1-\gamma_s)\left\{2\gamma_s w^{\mathrm{pr}}_{1,s|s}+(1-\gamma_s)\left(w^{\mathrm{pr}}_{1,s|s}\right)^2\right\}\\
w^{\mathrm{II}}_{2,s|s}=\dfrac{\gamma_s^2}{n_s-1}+\dfrac{1}{n_s}(1-\gamma_s)\left\{2\gamma_s w^{\mathrm{pr}}_{1,s|s}+(1-\gamma_s)\left(w^{\mathrm{pr}}_{1,s|s}\right)^2\right\}\\
\dfrac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_1}=\dfrac{2}{n_s}(1-\gamma_s)\left[\gamma_s+(1-\gamma_s)w^{\mathrm{pr}}_{1,s|s}\right]\dfrac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_1}\\
\dfrac{\partial w^{\mathrm{I}}_{2,s|s}}{\partial z_2}=\dfrac{2}{n_s}(1-\gamma_s)\left[\gamma_s+(1-\gamma_s)w^{\mathrm{pr}}_{1,s|s}\right]\dfrac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_2}\\
\dfrac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_1}=\dfrac{1}{n_s}(1-\gamma_s)\left[\gamma_s+(1-\gamma_s)w^{\mathrm{pr}}_{1,s|s}\right]\left[\dfrac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_1}+\dfrac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_2}\right]\\
\dfrac{\partial w^{\mathrm{II}}_{2,s|s}}{\partial z_3}=\dfrac{2}{n_s}(1-\gamma_s)\left[\gamma_s+(1-\gamma_s)w^{\mathrm{pr}}_{1,s|s}\right]\dfrac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_2}.
\end{cases}
\tag{F23}
$$

By substituting eq.(F23) in eq.(E2), we obtain a recursion on $r_2^{(0)}(s)$ as

$$
\begin{aligned}
r_2^{(0)}(s)&=\frac{1}{n_s}(1-\gamma_s)\left\{2\gamma_s w^{\mathrm{pr}}_{1,s|s}+(1-\gamma_s)\left(w^{\mathrm{pr}}_{1,s|s}\right)^2\right\}\\
&\quad+\left[\gamma_s^2+\frac{n_s-1}{n_s}(1-\gamma_s)\left\{2\gamma_s w^{\mathrm{pr}}_{1,s|s}+(1-\gamma_s)\left(w^{\mathrm{pr}}_{1,s|s}\right)^2\right\}\right]r_2^{(0)}(s),
\end{aligned}
\tag{F24}
$$

which is explicitly solved as eq.(41) in the main text.

F.2.2 Three-way relatedness

Next we calculate individual 3-fitness of three different types (from type-I to III) in order to calculate three-way relatedness. For type-I, we have

$$
\begin{aligned}
&w^{\mathrm{I}}_{3,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z) \\
&= \sum_{k_1=1}^{n_s} \zeta_{1,k_1} \left(\frac{1}{(n_s-1)(n_s-2)} \sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \zeta_{1,k_2}\zeta_{1,k_3} \right) \\
&= \frac{1}{(n_s-1)(n_s-2)} \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \left[\gamma_s\delta_{1,k_1} + (1-\gamma_s)\omega_1\right]\left[\gamma_s\delta_{1,k_2} + (1-\gamma_s)\omega_1\right]\left[\gamma_s\delta_{1,k_3} + (1-\gamma_s)\omega_1\right] \\
&= \frac{1}{(n_s-1)(n_s-2)} \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \left[\gamma_s(1-\gamma_s)^2(\delta_{1,k_1}+\delta_{1,k_2}+\delta_{1,k_3})\omega_1^2 + (1-\gamma_s)^3\omega_1^3\right] \\
&= 3\gamma_s(1-\gamma_s)^2\omega_1^2 + n_s(1-\gamma_s)^3\omega_1^3 \\
&= \frac{1}{n_s^2}\left\{3\gamma_s(1-\gamma_s)^2\left[w^{\mathrm{pr}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)\right]^2 + (1-\gamma_s)^3\left[w^{\mathrm{pr}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)\right]^3\right\},
\end{aligned}
\tag{F25}
$$

where we have used

$$
\begin{cases}
\displaystyle\sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1}\delta_{1,k_2}\delta_{1,k_3} = 0, \\
\displaystyle\sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1}\delta_{1,k_2} = \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1}\delta_{1,k_3} = \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_2}\delta_{1,k_3} = 0, \\
\displaystyle\sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1} = \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_2} = \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_3} = (n_s-1)(n_s-2), \\
\displaystyle\left(\sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} 1\right) = n_s(n_s-1)(n_s-2).
\end{cases}
\tag{F26}
$$

For type-II, we obtain

$$
\begin{aligned}
&w^{\mathrm{II}}_{3,s|s}(z_1, z_2, \boldsymbol{z}_{-\{1,2\}}, z) \\
&= \sum_{k_1=1}^{n_s} \zeta_{1,k_1} \left(\frac{1}{(n_s-1)(n_s-2)} \sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \zeta_{2,k_2}\zeta_{2,k_3} \right) \\
&= \frac{1}{(n_s-1)(n_s-2)} \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \left[\gamma_s\delta_{1,k_1} + (1-\gamma_s)\omega_1\right]\left[\gamma_s\delta_{2,k_2} + (1-\gamma_s)\omega_2\right]\left[\gamma_s\delta_{2,k_3} + (1-\gamma_s)\omega_2\right] \\
&= \frac{1}{(n_s-1)(n_s-2)} \sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \Big[\gamma_s^2(1-\gamma_s)(\delta_{1,k_1}\delta_{2,k_2} + \delta_{1,k_1}\delta_{2,k_3})\omega_2 \\
&\qquad\qquad + \gamma_s(1-\gamma_s)^2(\delta_{1,k_1}\omega_2^2 + \delta_{2,k_2}\omega_1\omega_2 + \delta_{2,k_3}\omega_1\omega_2) \\
&\qquad\qquad + (1-\gamma_s)^3\omega_1\omega_2^2\Big] \\
&= \frac{2}{n_s-1}\gamma_s^2(1-\gamma_s)\omega_2 + \gamma_s(1-\gamma_s)^2(2\omega_1\omega_2 + \omega_2^2) + n_s(1-\gamma_s)^3\omega_1\omega_2^2 \\
&= \frac{2}{n_s(n_s-1)}\gamma_s^2(1-\gamma_s)\left[w^{\mathrm{pr}}_{1,s|s}(z_2, \boldsymbol{z}_{-\{2\}}, z)\right] \\
&\quad + \frac{1}{n_s^2}\Big\{\gamma_s(1-\gamma_s)^2\left(2\left[w^{\mathrm{pr}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)\right]\left[w^{\mathrm{pr}}_{1,s|s}(z_2, \boldsymbol{z}_{-\{2\}}, z)\right] + \left[w^{\mathrm{pr}}_{1,s|s}(z_2, \boldsymbol{z}_{-\{2\}}, z)\right]^2\right) \\
&\qquad + (1-\gamma_s)^3\left[w^{\mathrm{pr}}_{1,s|s}(z_1, \boldsymbol{z}_{-\{1\}}, z)\right]\left[w^{\mathrm{pr}}_{1,s|s}(z_2, \boldsymbol{z}_{-\{2\}}, z)\right]^2\Big\},
\end{aligned}
\tag{F27}
$$

where we have used

$$
\begin{cases}
\displaystyle\sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1}\delta_{2,k_2}\delta_{2,k_3} = 0, \\
\displaystyle\sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{2,k_2}\delta_{2,k_3} = 0, \\
\displaystyle\sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1}\delta_{2,k_2} = \sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1}\delta_{2,k_3} = n_s - 2, \\
\displaystyle\sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{1,k_1} = \sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{2,k_2} = \sum_{\substack{1\le k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} \delta_{2,k_3} = (n_s-1)(n_s-2), \\
\displaystyle\left(\sum_{\substack{1\le k_1,k_2,k_3\le n_s \\ k_1\neq k_2\neq k_3\neq k_1}} 1\right) = n_s(n_s-1)(n_s-2).
\end{cases}
\tag{F28}
$$

Finally, for type-III we obtain

$$
\begin{aligned}
&w^{\mathrm{III}}_{3,s|s}(z_1,z_2,z_3,\boldsymbol{z}_{-\{1,2,3\}},z)\\
&=\sum_{k_1=1}^{n_s}\zeta_{1,k_1}\left(\frac{1}{(n_s-1)(n_s-2)}\sum_{\substack{1\le k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}(\zeta_{2,k_2}\zeta_{3,k_3}+\zeta_{3,k_2}\zeta_{2,k_3})\right)\\
&=\frac{2}{(n_s-1)(n_s-2)}\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\zeta_{1,k_1}\zeta_{2,k_2}\zeta_{3,k_3}\\
&=\frac{2}{(n_s-1)(n_s-2)}\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}[\gamma_s\delta_{1,k_1}+(1-\gamma_s)\omega_1]\,[\gamma_s\delta_{2,k_2}+(1-\gamma_s)\omega_2]\,[\gamma_s\delta_{3,k_3}+(1-\gamma_s)\omega_3]\\
&=\frac{2}{(n_s-1)(n_s-2)}\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\big[\gamma_s^3\delta_{1,k_1}\delta_{2,k_2}\delta_{3,k_3}\\
&\qquad\qquad+\gamma_s^2(1-\gamma_s)(\delta_{1,k_1}\delta_{2,k_2}\omega_3+\delta_{1,k_1}\delta_{3,k_3}\omega_2+\delta_{2,k_2}\delta_{3,k_3}\omega_1)\\
&\qquad\qquad+\gamma_s(1-\gamma_s)^2(\delta_{1,k_1}\omega_2\omega_3+\delta_{2,k_2}\omega_1\omega_3+\delta_{3,k_3}\omega_1\omega_2)+(1-\gamma_s)^3\omega_1\omega_2\omega_3\big]\\
&=\frac{2}{(n_s-1)(n_s-2)}\gamma_s^3+\frac{2}{n_s-1}\gamma_s^2(1-\gamma_s)(\omega_1+\omega_2+\omega_3)\\
&\quad+2\gamma_s(1-\gamma_s)^2(\omega_1\omega_2+\omega_1\omega_3+\omega_2\omega_3)+2n_s(1-\gamma_s)^3\omega_1\omega_2\omega_3\\
&=\frac{2}{(n_s-1)(n_s-2)}\gamma_s^3\\
&\quad+\frac{2}{n_s(n_s-1)}\gamma_s^2(1-\gamma_s)\Big(\Big[w^{\mathrm{pr}}_{1,s|s}(z_1,\boldsymbol{z}_{-\{1\}},z)\Big]+\Big[w^{\mathrm{pr}}_{1,s|s}(z_2,\boldsymbol{z}_{-\{2\}},z)\Big]+\Big[w^{\mathrm{pr}}_{1,s|s}(z_3,\boldsymbol{z}_{-\{3\}},z)\Big]\Big)\\
&\quad+\frac{2}{n_s^2}\Big\{\gamma_s(1-\gamma_s)^2\Big(\Big[w^{\mathrm{pr}}_{1,s|s}(z_1,\boldsymbol{z}_{-\{1\}},z)\Big]\Big[w^{\mathrm{pr}}_{1,s|s}(z_2,\boldsymbol{z}_{-\{2\}},z)\Big]\\
&\qquad\qquad+\Big[w^{\mathrm{pr}}_{1,s|s}(z_1,\boldsymbol{z}_{-\{1\}},z)\Big]\Big[w^{\mathrm{pr}}_{1,s|s}(z_3,\boldsymbol{z}_{-\{3\}},z)\Big]\\
&\qquad\qquad+\Big[w^{\mathrm{pr}}_{1,s|s}(z_2,\boldsymbol{z}_{-\{2\}},z)\Big]\Big[w^{\mathrm{pr}}_{1,s|s}(z_3,\boldsymbol{z}_{-\{3\}},z)\Big]\Big)\\
&\qquad+(1-\gamma_s)^3\Big[w^{\mathrm{pr}}_{1,s|s}(z_1,\boldsymbol{z}_{-\{1\}},z)\Big]\Big[w^{\mathrm{pr}}_{1,s|s}(z_2,\boldsymbol{z}_{-\{2\}},z)\Big]\Big[w^{\mathrm{pr}}_{1,s|s}(z_3,\boldsymbol{z}_{-\{3\}},z)\Big]\Big\}.
\end{aligned}
\tag{F29}
$$

Here we have used

$$
\begin{cases}
\displaystyle\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\delta_{1,k_1}\delta_{2,k_2}\delta_{3,k_3}=1,\\
\displaystyle\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\delta_{1,k_1}\delta_{2,k_2}=\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\delta_{1,k_1}\delta_{3,k_3}=\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\delta_{2,k_2}\delta_{3,k_3}=n_s-2,\\
\displaystyle\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\delta_{1,k_1}=\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\delta_{2,k_2}=\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}\delta_{3,k_3}=(n_s-1)(n_s-2),\\
\displaystyle\left(\sum_{\substack{1\le k_1,k_2,k_3\le n_s\\ k_1\neq k_2\neq k_3\neq k_1}}1\right)=n_s(n_s-1)(n_s-2).
\end{cases}
\tag{F30}
$$

In contrast, when $s'\neq s$, from the assumption of random dispersal we have $w^{\mathrm{I}}_{3,s'|s}=w^{\mathrm{II}}_{3,s'|s}=$

$w^{\mathrm{III}}_{3,s'|s} = 0$. From these calculations we arrive at

$$\begin{cases} w^{\mathrm{I}}_{3,s|s} = \dfrac{1}{n_s^2}\left\{3\gamma_s(1-\gamma_s)^2\left(w^{\mathrm{pr}}_{1,s|s}\right)^2 + (1-\gamma_s)^3\left(w^{\mathrm{pr}}_{1,s|s}\right)^3\right\} \\ w^{\mathrm{II}}_{3,s|s} = \dfrac{2}{n_s(n_s-1)}\gamma_s^2(1-\gamma_s)w^{\mathrm{pr}}_{1,s|s} + \dfrac{1}{n_s^2}\left\{3\gamma_s(1-\gamma_s)^2\left(w^{\mathrm{pr}}_{1,s|s}\right)^2 + (1-\gamma_s)^3\left(w^{\mathrm{pr}}_{1,s|s}\right)^3\right\} \\ w^{\mathrm{III}}_{3,s|s} = \dfrac{2}{(n_s-1)(n_s-2)}\gamma_s^3 + \dfrac{6}{n_s(n_s-1)}\gamma_s^2(1-\gamma_s)w^{\mathrm{pr}}_{1,s|s} \\ \qquad + \dfrac{2}{n_s^2}\left\{3\gamma_s(1-\gamma_s)^2\left(w^{\mathrm{pr}}_{1,s|s}\right)^2 + (1-\gamma_s)^3\left(w^{\mathrm{pr}}_{1,s|s}\right)^3\right\}. \end{cases} \tag{F31}$$

Substituting eq.(F31) in eq.(E6) gives us a recursion on $r_3^{(0)}$ that includes $r_2^{(0)}$. Solving it with the help of eq.(41) gives the following result:

$$r_3^{(0)}(s) = \frac{(w^{\mathrm{pr}}_{1,s|s})^2\{a_0 + a_1(w^{\mathrm{pr}}_{1,s|s}) + a_2(w^{\mathrm{pr}}_{1,s|s})^2 + a_3(w^{\mathrm{pr}}_{1,s|s})^2\}}{\{b_0 + b_1(w^{\mathrm{pr}}_{1,s|s}) + b_2(w^{\mathrm{pr}}_{1,s|s})^2\}\{c_0 + c_1(w^{\mathrm{pr}}_{1,s|s}) + c_2(w^{\mathrm{pr}}_{1,s|s})^2 + c_3(w^{\mathrm{pr}}_{1,s|s})^3\}}, \tag{F32a}$$

where

$$\begin{cases} a_0 = 3n_s\gamma_s(1+3\gamma_s^2) \\ a_1 = \{n_s + (17n_s - 12)\gamma_s^2\}(1-\gamma_s) \\ a_2 = 10(n_s-1)\gamma_s(1-\gamma_s)^2 \\ a_3 = 2(n_s-1)(1-\gamma_s)^3 \\ b_0 = n_s(1+\gamma_s) \\ b_1 = -2(n_s-1)\gamma_s \\ b_2 = -(n_s-1)(1-\gamma_s) \\ c_0 = n_s^2(1+\gamma_s+\gamma_s^2) \\ c_1 = -3n_s(n_s-2)\gamma_s^2 \\ c_2 = -3(n_s-1)(n_s-2)\gamma_s(1-\gamma_s) \\ c_3 = -(n_s-1)(n_s-2)(1-\gamma_s)^2. \end{cases} \tag{F32b}$$

Setting $\gamma_s = 0$ or $\gamma_s \sim 1$ (note that setting $\gamma_s = 1$ means individuals never die and hence evolution does not occur, so we need γ_s to be "close" to 1) in eq.(F32) gives eq.(42) in the main result.

F.2.3 First-order perturbation of pairwise relatedness

Here we assume $\rho^{(1)} = 0$ and calculate the first-order perturbation of pairwise relatedness, $r_2^{(1)}$.

Substituting eq.(F23) in eq.(E10) yields

$$
\begin{aligned}
r_2^{(1)}(s) = & \frac{r_2^{(0)}(s)}{\frac{1}{n_s}(1-\gamma_s)\left\{2\gamma_s w_{1,s|s}^{\mathrm{pr}} + (1-\gamma_s)\left(w_{1,s|s}^{\mathrm{pr}}\right)^2\right\}} \cdot \frac{2}{n_s}(1-\gamma_s)\left[\gamma_s + (1-\gamma_s)w_{1,s|s}^{\mathrm{pr}}\right] \times \\
& \left[\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_1} + (n_s-1)\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2} r_2^{(0)}(s)\right. \\
& \left. + (n_s-1)\left\{\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_1} + \frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2}\right\} r_2^{(0)}(s) + (n_s-1)(n_s-2)\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2} r_3^{(0)}(s)\right] \\
= & \, 2r_2^{(0)}(s)\frac{\gamma_s + (1-\gamma_s)w_{1,s|s}^{\mathrm{pr}}}{2\gamma_s w_{1,s|s}^{\mathrm{pr}} + (1-\gamma_s)\left(w_{1,s|s}^{\mathrm{pr}}\right)^2} \\
& \times \left\{[1+(n_s-1)r_2^{(0)}(s)]\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_1} + (n_s-1)[2r_2^{(0)}(s) + (n_s-2)r_3^{(0)}(s)]\frac{\partial w_{1,s|s}^{\mathrm{pr}}}{\partial z_2}\right\},
\end{aligned}
\tag{F33}
$$

which is eq.(44) in the main text.

G Consistency with previous results about relatedenss

We here show that we recover several previous results concerning relatedness from our model.

G.1 Neutral relatedness in the lottery model

Our result for the neutral pairwise relatedness given in eq.(41) agrees with R_d in Pen (2000) (the solution to his eq.(A2)). To see this one has to set $S = \gamma_s$, $h = w_{1,s|s}^{\mathrm{pr}}$ and $N = n_s$ in Pen (2000). Furthermore, for the special case that $\gamma_s = 0$ in eq.(41) we obtain

$$
r_2^{(0)}(s) = \frac{\left(w_{1,s|s}^{\mathrm{pr}}\right)^2}{n_s - (n_s-1)\left(w_{1,s|s}^{\mathrm{pr}}\right)^2}, \tag{G1}
$$

which agrees with $\dot{Q}$ in eq.(2.9) in Rousset (2004), a standard results for the island model, by using $\gamma = 1$, $1 - m = w_{1,s|s}^{\mathrm{pr}}$ and $N = n_s$ there.

On the other hand, taking the limit $\gamma_s \to 1$ in eq.(41) gives

$$
r_2^{(0)}(s) = \frac{w_{1,s|s}^{\mathrm{pr}}}{n_s - (n_s-1)w_{1,s|s}^{\mathrm{pr}}}. \tag{G2}
$$

This result agrees with $r_2(\boldsymbol{z},\boldsymbol{z}) = (1-m(\boldsymbol{z}))/(1+m(\boldsymbol{z})(N-1))$ in Table 1 in Mullon et al. (2016) by setting $1 - m(\boldsymbol{z}) = w_{1,s|s}^{\mathrm{pr}}$ and $N = n_s$ in their formula. Note, that there was a typo in their original expression, which we corrected here.

Finally, as for $r_3^{(0)}(s)$, the first line of eq.(42) agrees with R_3 in eq.(12b) in Ohtsuki (2010) by setting $1 - m = w_{1,s|s}^{\mathrm{pr}}$ and $N = n_s$ there. The second line agrees with $r_3(\boldsymbol{z},\boldsymbol{z})$ in Table 1 in Mullon et al. (2016) by setting $1 - m(\boldsymbol{z}) = w_{1,s|s}^{\mathrm{pr}}$ and $N = n_s$ there.

G.2 Neutral relatedness under fluctuating group size

We here prove the connection between our eq.(33) and eq.(29) of Rousset and Ronce (2004). In this latter model, individuals migrate independently from each other (no propagule migration) and states determine

group size, which fluctuates stochasticaly between generations according to an ergodic Markov chain. The probability $p^{(0)}(s)$ that a group is in state s in the neutral process at stationarity then satisfies

$$p^{(0)}(s') = \sum_{s\in\mathcal{S}} p^{(0)}(s'|s)p^{(0)}(s), \tag{G3}$$

where $p^{(0)}(s'|s) = p^{(0)}(s'|s;\boldsymbol{p}^{(0)})$ is the forward transition probability that a group in state s in the parental generation is in state s' in the offspring generation and this generally depends on whole population state $\boldsymbol{p}^{(0)} = (p^{(0)}(s_1), \cdots, p^{(0)}(s_N))$ (since groups are connected to each other by dispersal, see Metz and Gyllenberg (2001); Lehmann et al. (2006); Alizon and Taylor (2008) for concrete examples of such transition probabilities) and on the resident trait value.

In terms of these quantities, first note

$$\frac{q^{(0)}(s)}{q^{(0)}(s')} = \frac{n_s p^{(0)}(s)}{n_{s'} p^{(0)}(s')}, \tag{G4}$$

where the equality follows from eq.(E.21) of Lehmann et al. (2016). Intuitively speaking, eq.(G4) tells that the stationary fraction of individuals in group s under neutrality, $n_s p^{(0)}(s)$, is proportional to the stationary distribution of mutants under neutrality, $q^{(0)}(s)$. Second, because the model of Rousset and Ronce (2004) did not allow any propagule dispersal, the only way that a focal individual in a group in state s earns individual 2-fitness is that the state of the group changes from state s in the parental generation to state s' in the offspring generation and the focal individual produces offspring in the focal group. Thus, we obtain

$$\begin{aligned} w^{\mathrm{I}}_{2,s'|s} &= w^{\mathrm{I}}_{2,s'|(s'\leftarrow s)} p^{(0)}(s'|s) \\ w^{\mathrm{II}}_{2,s'|s} &= w^{\mathrm{II}}_{2,s'|(s'\leftarrow s)} p^{(0)}(s'|s), \end{aligned} \tag{G5}$$

where $w^{\mathrm{I}}_{2,s'|(s'\leftarrow s)}$ and $w^{\mathrm{II}}_{2,s'|(s'\leftarrow s)}$ are conditional 2-fitness components of the focal individual (they follow the same definition as $w^{\mathrm{I}}_{2,s'|s}$ and $w^{\mathrm{II}}_{2,s'|s}$ but are conditional on the parental generation being in state s and the offspring generation being in state s') evaluated under neutrality.

Substituting eqs.(G4)–(G5) into eq.(33) and using the backward transition probability

$$p^{(0)}_{\mathrm{back}}(s|s') \equiv \frac{p^{(0)}(s'|s)p^{(0)}(s)}{p^{(0)}(s')} \tag{G6}$$

that a group in state s' in the offspring generation was in state s in the previous generation, we obtain that

$$r^{(0)}_2(s') = \sum_{s\in\mathcal{S}} \left[F^{\mathrm{I}}_{2,s'|(s'\leftarrow s)} + F^{\mathrm{II}}_{2,s'|(s'\leftarrow s)} r^{(0)}_2(s) \right] p^{(0)}_{\mathrm{back}}(s \mid s'), \tag{G7}$$

where

$$\begin{aligned} F^{\mathrm{I}}_{2,s'|(s'\leftarrow s)} &\equiv \frac{n_s}{n_{s'}} w^{\mathrm{I}}_{2,s'|(s'\leftarrow s)} \\ F^{\mathrm{II}}_{2,s'|(s'\leftarrow s)} &\equiv \frac{(n_s-1)n_s}{n_{s'}} w^{\mathrm{II}}_{2,s'|(s'\leftarrow s)}. \end{aligned} \tag{G8}$$

We now claim that $F^{\mathrm{I}}_{2,s'|(s'\leftarrow s)}$ is the probability that two randomly sampled offspring in state s' both descend from the same parent in state s (*i.e.*, the coalescence probability) under neutrality and that $F^{\mathrm{II}}_{2,s'|(s'\leftarrow s)}$ is the probability that two randomly sampled offspring in state s' both descend from

two distinct parents in state s under neutrality. If this interpretation holds, then our eq.(G7) reduces to eq.(29) of Rousset and Ronce (2004).

We now proceed to prove this claim. Given that the state of the group is s in the parental generation and s' in the offspring generation, from the definition of (conditional) same-parent 2-fitness, we can write

$$w^{\mathrm{I}}_{2,s'|(s'\leftarrow s)} = \sum_{k_1=1}^{n_{s'}} \zeta_{1,k_1} \left(\frac{1}{n_{s'}-1} \sum_{\substack{k_2=1 \\ k_2\neq k_1}}^{n_{s'}} \zeta_{1,k_2} \right), \tag{G9}$$

where, we use the ζ-notation once used in section F.2; this time $\zeta_{i,k}$ represents the probability that an individual in position $k \in \{1,\cdots,n_{s'}\}$ in the next generation in the focal group descends from adult $i \in \{1,\cdots,n_s\}$ in that group in the previous generation under neutrality conditioned on that the group state has changed (forwardly in time) from s to s'.

$$F^{\mathrm{I}}_{2,s'|(s'\leftarrow s)} = n_s \times \sum_{k_1=1}^{n_{s'}} \sum_{\substack{k_2=1 \\ k_2\neq k_1}}^{n_{s'}} \frac{\zeta_{1,k_1}\zeta_{1,k_2}}{n_{s'}(n_{s'}-1)}, \tag{G10}$$

which, owing to neutrality (and thus exchangeability of individuals), can be rewritten as

$$F^{\mathrm{I}}_{2,s'|(s'\leftarrow s)} = \sum_{i=1}^{n_s} \sum_{k_1=1}^{n_{s'}} \sum_{\substack{k_2=1 \\ k_2\neq k_1}}^{n_{s'}} \frac{\zeta_{i,k_1}\zeta_{i,k_2}}{n_{s'}(n_{s'}-1)}, \tag{G11}$$

and therefore can be read as the ratio of the total number of ways of sampling two offspring from the same parent in a group of size n_s to the total number of ways of sampling two offspring in a group of size $n_{s'}$ (*i.e.*, the coalescence probability). Likewise, we have

$$w^{\mathrm{II}}_{2,s'|(s'\leftarrow s)} = \sum_{k_1=1}^{n_{s'}} \zeta_{1,k_1} \left(\frac{1}{n_{s'}-1} \sum_{\substack{k_2=1 \\ k_2\neq k_1}}^{n_{s'}} \zeta_{2,k_2} \right), \tag{G12}$$

whereby

$$F^{\mathrm{II}}_{2,s'|(s'\leftarrow s)} = (n_s-1)n_s \times \sum_{k_1=1}^{n_{s'}} \sum_{\substack{k_2=1 \\ k_2\neq k_1}}^{n_{s'}} \frac{\zeta_{1,k_1}\zeta_{2,k_2}}{n_{s'}(n_{s'}-1)}, \tag{G13}$$

which, owing to neutrality and exchangeability of individuals, can be rewritten as

$$F^{\mathrm{II}}_{2,s'|(s'\leftarrow s)} = \sum_{i_1=1}^{n_s} \sum_{\substack{i_2=1 \\ i_2\neq i_1}}^{n_s} \sum_{k_1=1}^{n_{s'}} \sum_{\substack{k_2=1 \\ k_2\neq k_1}}^{n_{s'}} \frac{\zeta_{i_1,k_1}\zeta_{i_2,k_2}}{n_{s'}(n_{s'}-1)}, \tag{G14}$$

and thus can be read as the ratio of the total number of ways of sampling two offspring from the distinct parents in a group of size n_s to the total number of ways of sampling two offspring in a group of size $n_{s'}$ (*i.e.*, the proability that offspring in state s' descend from two distinct parents in state s). This ends the proof of our aforementioned claim.

G.3 Perturbation of relatedness

As for $r_2^{(1)}(s)$, eq.(44) evaluated at $\gamma_s = 0$ reads

$$r_2^{(1)}(s) = 2r_2^{(0)}(s)\frac{1}{w^{\mathrm{pr}}_{1,s|s}} \left\{ \left[1+(n_s-1)r_2^{(0)}(s)\right] \frac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_1} + (n_s-1)\left[2r_2^{(0)}(s)+(n_s-2)r_3^{(0)}(s)\right] \frac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_2} \right\}, \tag{G15}$$

which corresponds to the expression that appear in Δr in Wakano and Lehmann (2014) (see the bottom of their Appendix B, after their eq.(B.46)), which essentially shows that the perturbation of pair-relatedness is

$$2R_2\frac{(1+(N-1)R_2)w_{\mathrm{S}}^{\mathrm{P}}+(2R_2+(N-2)R_3)w_{\mathrm{D}}^{\mathrm{P}}}{1-m}. \tag{G16}$$

The correspondence between eq.(G15) and eq.(G16) is as follows. Set $R_2=r_2^{(0)}(s), R_3=r_3^{(0)}(s), N=n_s, 1-m=w_{1,s|s}^{\mathrm{pr}}, w_{\mathrm{S}}^{\mathrm{P}}=\partial w_{1,s|s}^{\mathrm{pr}}/\partial z_1$, and $w_{\mathrm{D}}^{\mathrm{P}}=(n_s-1)(\partial w_{1,s|s}^{\mathrm{pr}}/\partial z_2)$ in eq.(G16). Note that the original expression at the bottom of Appendix B in Wakano and Lehmann (2014) contains the factor 4, not 2. This two-fold difference comes from the fact that Wakano and Lehmann (2014) considered variance dynamics of a trait distribution, whereas we here consider directly the perturbation of $r_2(s)$.

H Perturbations for the hard selection lottery model

Here, we derive the expressions for $\rho^{(1)}$ and $\rho^{(2)}$ for the lottery model under hard selection. The resulting expression are complicated and do not appear in the main text, but they can be useful for numerical calculations as they apply to any fecundity function given the model's other assumptions.

We recall that for hard selection, from eqs.(38b) and (38c) with eqs.(45a) and (46a), we have

$$\begin{cases} w_{1,s|s}^{\mathrm{p}}(z_1,\boldsymbol{z}_{-\{1\}},z)=\gamma_s+(1-\gamma_s)\underbrace{\dfrac{(1-m_s)n_sf_s(z_1,\boldsymbol{z}_{-\{1\}},z)}{(1-m_s)\sum_{i=1}^{n_s}f_s(z_i,\boldsymbol{z}_{-\{i\}},z)+I_{\mathrm{hard}}(z)}}_{=w_{1,s|s}^{\mathrm{pr}}(z_1,\boldsymbol{z}_{-\{1\}},z)} \\ w_{1,s'|s}^{\mathrm{d}}(z_1,\boldsymbol{z}_{-\{1\}},z)=\underbrace{(1-\gamma_{s'})\underbrace{\dfrac{\pi_{s'}n_{s'}}{(1-m_{s'})\sum_{i=1}^{n_{s'}}f_{s'}(z,\boldsymbol{z},z)+I_{\mathrm{hard}}(z)}}_{=w_{s'}^{\mathrm{col}}(z)}\cdot\underbrace{p_sm_sf_s(z_1,\boldsymbol{z}_{-\{1\}},z)}_{=w_s^{\mathrm{em}}(z_1,\boldsymbol{z}_{-\{1\}},z)}}_{=w_{1,s'|s}^{\mathrm{dr}}(z_1,\boldsymbol{z}_{-\{1\}},z)}. \end{cases} \tag{H1}$$

H.1 First-order perturbation of invasion fitness

Our goal here is to calculate the first-order perturbation of invasion fitness, given by eq.(32). For that purpose, we will calculate the components that appear in eq.(32), as below.

Firstly, we will calculate $v^{(0)}(s')q^{(0)}(s)$. With the definition of backward migration probability, eq.(49), we see that

$$\begin{cases} w_{1,s|s}^{\mathrm{pr}}=1-d_{s,\mathrm{hard}} \\ w_{1,s'|s}^{\mathrm{dr}}=d_{s',\mathrm{hard}}\dfrac{\pi_{s'}n_{s'}p_sm_sf_s}{I_{\mathrm{hard}}} \end{cases} \tag{H2}$$

hold. Substituting them into eq.(40) gives

$$\begin{aligned} v^{(0)}(s')q^{(0)}(s)&=\frac{\pi_sn_sp_{s'}m_{s'}f_{s'}}{(1-\gamma_{s'})d_{s',\mathrm{hard}}I_{\mathrm{hard}}}\Bigg/\left(\sum_{s''}\frac{\pi_{s''}n_{s''}p_{s''}m_{s''}f_{s''}}{(1-\gamma_{s''})d_{s'',\mathrm{hard}}I_{\mathrm{hard}}}\right) \\ &=\frac{\pi_sn_sp_{s'}m_{s'}f_{s'}}{(1-\gamma_{s'})d_{s',\mathrm{hard}}}\Bigg/\left\langle\frac{npmf}{(1-\gamma)d_{\mathrm{hard}}}\right\rangle, \end{aligned} \tag{H3}$$

where we set

$$\left\langle\frac{npmf}{(1-\gamma)d_{\mathrm{hard}}}\right\rangle\equiv\sum_{s\in\mathcal{S}}\pi_s\frac{n_sp_sm_sf_s}{(1-\gamma_s)d_{s,\mathrm{hard}}}. \tag{H4}$$

Secondly, we will calculate two different derivatives of individual 1-fitness that appear in eq.(32). A direct calculation of eq.(H1) shows that

$$\begin{cases} \dfrac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_1} = \underbrace{(1-d_{s,\mathrm{hard}})}_{=w^{\mathrm{pr}}_{1,s|s}} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\ \dfrac{\partial w^{\mathrm{pr}}_{1,s|s}}{\partial z_2} = \underbrace{(1-d_{s,\mathrm{hard}})}_{=w^{\mathrm{pr}}_{1,s|s}} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\ \dfrac{\partial w^{\mathrm{dr}}_{1,s'|s}}{\partial z_1} = \underbrace{d_{s',\mathrm{hard}} \dfrac{\pi_{s'} n_{s'} p_s m_s f_s}{I_{\mathrm{hard}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \cdot \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} \\ \dfrac{\partial w^{\mathrm{dr}}_{1,s'|s}}{\partial z_2} = \underbrace{d_{s',\mathrm{hard}} \dfrac{\pi_{s'} n_{s'} p_s m_s f_s}{I_{\mathrm{hard}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \cdot \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}. \end{cases} \tag{H5}$$

Remember that from eq.(38)

$$w_{1,s'|s} = \delta_{s',s}[\gamma_s + (1-\gamma_s) w^{\mathrm{pr}}_{1,s|s}] + (1-\gamma_{s'}) w^{\mathrm{dr}}_{1,s'|s} \tag{H6}$$

(where $\delta_{s',s}$ is Kronecker's delta; it is 1 if $s' = s$ and is 0 if $s' \neq s$) holds. Taking the first-order derivative of eq.(H6), substituting eq.(H5) therein, and using eq.(H6) again for rewriting give us

$$\begin{cases} \dfrac{\partial w_{1,s'|s}}{\partial z_1} = (w_{1,s'|s} - \delta_{s',s}\gamma_s) \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} - \delta_{s',s}(1-\gamma_s)(1-d_{s,\mathrm{hard}})^2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\ \dfrac{\partial w_{1,s'|s}}{\partial z_2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s) \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} - \delta_{s',s}(1-\gamma_s)(1-d_{s,\mathrm{hard}})^2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right). \end{cases} \tag{H7}$$

Now we are ready to calculate the first-order perturbation of invasion fitness, eq.(32). We substitute eq.(H7) into eq.(32). We show this calculation piece by piece. Firstly,

$$\begin{aligned} &\sum_{s'} \sum_{s} v^{(0)}(s') \frac{\partial w_{1,s'|s}}{\partial z_1} q^{(0)}(s) \\ &= \sum_{s} \underbrace{\sum_{s'} v^{(0)}(s') w_{1,s'|s}}_{\substack{=v^{(0)}(s) \\ \text{(from eq.(15b))}}} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} q^{(0)}(s) - \sum_{s} v^{(0)}(s) \gamma_s \frac{\frac{\partial f_s}{\partial z_1}}{f_s} q^{(0)}(s) \\ &\quad - \sum_{s} v^{(0)}(s)(1-\gamma_s)(1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) q^{(0)}(s) \\ &= \sum_{s} (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}. \end{aligned} \tag{H8}$$

Secondly, a quite similar calculation to above leads to

$$\begin{aligned} &\sum_{s'} \sum_{s} v^{(0)}(s')(n_s-1) \frac{\partial w_{1,s'|s}}{\partial z_2} r^{(0)}_2(s) q^{(0)}(s) \\ &= \sum_{s} (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) \cdot (n_s-1) r^{(0)}_2(s) \left\{ \frac{\frac{\partial f_s}{\partial z_2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}. \end{aligned} \tag{H9}$$

By combining eqs.(H8) and (H9) in eq.(32), we obtain the first-order perturbation as

$$\rho^{(1)} = \sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s - 1) r_2^{(0)}(s) \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right.$$
$$\left. - [1 + (n_s-1) r_2^{(0)}(s)](1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}. \tag{H10}$$

Rewriting it by using eqs.(50) and (H3) gives eq.(48) in the main text. Here, $r_2^{(0)}(s)$ is calculated by substituting eq.(H2) into eq.(41). The result is shown in eq.(51) in the main text.

H.2 Second-order perturbation of invasion fitness

We now assume that $\rho^{(1)} = 0$. Our goal here is to calculate the second-order perturbation of invasion fitness, given by eq.(34). We have already calculated first-order derivatives that appear there, as given in eq.(H7). We will then calculate various second-order derivatives that appear in eq.(34). A direct calculation of eq.(H1) shows that relevant terms are

$$\begin{cases}
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_1^2} = \underbrace{(1-d_{s,\mathrm{hard}})}_{=w^{\mathrm{pr}}_{1,s|s}} \frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \right) \\
\qquad - 2(1-d_{s,\mathrm{hard}})^2 \frac{\frac{\partial f_s}{\partial z_1}}{f_s} \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\
\qquad + 2(1-d_{s,\mathrm{hard}})^3 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \\
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_2^2} = \underbrace{(1-d_{s,\mathrm{hard}})}_{=w^{\mathrm{pr}}_{1,s|s}} \frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \right) \\
\qquad - 2(1-d_{s,\mathrm{hard}})^2 \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\
\qquad + 2(1-d_{s,\mathrm{hard}})^3 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \\
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_1 z_2} = \underbrace{(1-d_{s,\mathrm{hard}})}_{=w^{\mathrm{pr}}_{1,s|s}} \frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{2}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \frac{n_s-2}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \\
\qquad - (1-d_{s,\mathrm{hard}})^2 \left(\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\
\qquad + 2(1-d_{s,\mathrm{hard}})^3 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \\
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_2 z_3} = \underbrace{(1-d_{s,\mathrm{hard}})}_{=w^{\mathrm{pr}}_{1,s|s}} \frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{2}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \frac{n_s-2}{n_s} \frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \\
\qquad - 2(1-d_{s,\mathrm{hard}})^2 \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\
\qquad + 2(1-d_{s,\mathrm{hard}})^3 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2
\end{cases} \tag{H11a}$$

and

$$\begin{cases} \dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_1^2} = \underbrace{d_{s',\mathrm{hard}} \dfrac{\pi_{s'} n_{s'} p_s m_s f_s}{I_{\mathrm{hard}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \cdot \dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} \\ \dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_2^2} = \underbrace{d_{s',\mathrm{hard}} \dfrac{\pi_{s'} n_{s'} p_s m_s f_s}{I_{\mathrm{hard}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \cdot \dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \\ \dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_1 z_2} = \underbrace{d_{s',\mathrm{hard}} \dfrac{\pi_{s'} n_{s'} p_s m_s f_s}{I_{\mathrm{hard}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \cdot \dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} \\ \dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_2 z_3} = \underbrace{d_{s',\mathrm{hard}} \dfrac{\pi_{s'} n_{s'} p_s m_s f_s}{I_{\mathrm{hard}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \cdot \dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s}. \end{cases} \tag{H11b}$$

Taking the second-order derivative of eq.(H6), substituting eq.(H11) therein, and using eq.(H6) again for rewriting yields

$$
\begin{cases}
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_1^2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} - \delta_{s',s}(1-\gamma_s)\times \\
\qquad \left\{(1-d_{s,\text{hard}})^2\left(\dfrac{1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right)\right. \\
\qquad\qquad + 2(1-d_{s,\text{hard}})^2\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s}\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) \\
\qquad\qquad\qquad \left. - 2(1-d_{s,\text{hard}})^3\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_2^2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} - \delta_{s',s}(1-\gamma_s)\times \\
\qquad \left\{(1-d_{s,\text{hard}})^2\left(\dfrac{1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right)\right. \\
\qquad\qquad + 2(1-d_{s,\text{hard}})^2\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) \\
\qquad\qquad\qquad \left. - 2(1-d_{s,\text{hard}})^3\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_1 z_2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} - \delta_{s',s}(1-\gamma_s)\times \\
\qquad \left\{(1-d_{s,\text{hard}})^2\left(\dfrac{2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s-2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s}\right)\right. \\
\qquad\qquad + (1-d_{s,\text{hard}})^2\left(\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) \\
\qquad\qquad\qquad \left. - 2(1-d_{s,\text{hard}})^3\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_2 z_3} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} - \delta_{s',s}(1-\gamma_s)\times \\
\qquad \left\{(1-d_{s,\text{hard}})^2\left(\dfrac{2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s-2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s}\right)\right. \\
\qquad\qquad + 2(1-d_{s,\text{hard}})^2\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) \\
\qquad\qquad\qquad \left. - 2(1-d_{s,\text{hard}})^3\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\}.
\end{cases}
\tag{H12}
$$

Now we are ready to calculate eq.(34). We start from $\rho^{(2\text{w})}$, which is given by eq.(34a). We substitute

eq.(H12) in eq.(34a). The four terms in eq.(34a) are then calculated as follows. For example,

$$
\begin{aligned}
Z_{1,\mathrm{hard}} &\equiv \frac{1}{2}\sum_{s'}\sum_{s} v^{(0)}(s')\frac{\partial^2 w_{1,s'|s}}{\partial z_1^2} q^{(0)}(s) \\
&= \frac{1}{2}\sum_{s}\underbrace{\sum_{s'} v^{(0)}(s') w_{1,s'|s}}_{\substack{=v^{(0)}(s) \\ \text{(from eq.(15b))}}} \frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} q^{(0)}(s) - \frac{1}{2}\sum_{s} v^{(0)}(s)\gamma_s \frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} q^{(0)}(s) \\
&\quad - \frac{1}{2}\sum_{s} v^{(0)}(s)(1-\gamma_s)\Bigg\{(1-d_{s,\mathrm{hard}})^2\left(\frac{1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right) \\
&\qquad + 2(1-d_{s,\mathrm{hard}})^2 \frac{\frac{\partial f_s}{\partial z_1}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) \\
&\qquad - 2(1-d_{s,\mathrm{hard}})^3\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\Bigg\} q^{(0)}(s) \qquad \text{(H13a)} \\
&= \frac{1}{2}\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\times \\
&\quad \Bigg\{\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} - (1-d_{s,\mathrm{hard}})^2\left(\frac{1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right) \\
&\qquad - 2(1-d_{s,\mathrm{hard}})^2 \frac{\frac{\partial f_s}{\partial z_1}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) \\
&\qquad\quad + 2(1-d_{s,\mathrm{hard}})^3\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\Bigg\}.
\end{aligned}
$$

The other terms that appear in eq.(34a) are calculated in essentially the same way as

$$
\begin{cases}
Z_{2,\mathrm{hard}} \equiv \dfrac{1}{2}\displaystyle\sum_{s'}\sum_{s} v^{(0)}(s')(n_s-1)\dfrac{\partial^2 w_{1,s'|s}}{\partial z_2^2} r_2^{(0)}(s) q^{(0)}(s) \\
\qquad = \dfrac{1}{2}\displaystyle\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)r_2^{(0)}(s)\times \\
\qquad \left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\dfrac{1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \right) \right. \\
\qquad\qquad - 2(1-d_{s,\mathrm{hard}})^2 \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\
\qquad\qquad\qquad \left. + 2(1-d_{s,\mathrm{hard}})^3 \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
Z_{3,\mathrm{hard}} \equiv \dfrac{1}{2}\displaystyle\sum_{s'}\sum_{s} v^{(0)}(s')2(n_s-1)\dfrac{\partial^2 w_{1,s'|s}}{\partial z_1 z_2} r_2^{(0)}(s) q^{(0)}(s) \\
\qquad = \dfrac{1}{2}\displaystyle\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\cdot 2(n_s-1)r_2^{(0)}(s)\times \\
\qquad \left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\dfrac{2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s-2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \right. \\
\qquad\qquad - (1-d_{s,\mathrm{hard}})^2 \left(\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\
\qquad\qquad\qquad \left. + 2(1-d_{s,\mathrm{hard}})^3 \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
Z_{4,\mathrm{hard}} \equiv \dfrac{1}{2}\displaystyle\sum_{s'}\sum_{s} v^{(0)}(s')(n_s-1)(n_s-2)\dfrac{\partial^2 w_{1,s'|s}}{\partial z_2 z_3} r_3^{(0)}(s) q^{(0)}(s) \\
\qquad = \dfrac{1}{2}\displaystyle\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)(n_s-2)r_3^{(0)}(s)\times \\
\qquad \left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\dfrac{2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s-2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \right. \\
\qquad\qquad - 2(1-d_{s,\mathrm{hard}})^2 \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \\
\qquad\qquad\qquad \left. + 2(1-d_{s,\mathrm{hard}})^3 \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\}.
\end{cases}
\tag{H13b}
$$

Collecting the Z-terms in eqs.(H13a) and (H13b) gives us

$$
\rho^{(2\mathrm{w})} = Z_{1,\mathrm{hard}} + Z_{2,\mathrm{hard}} + Z_{3,\mathrm{hard}} + Z_{4,\mathrm{hard}}. \tag{H13c}
$$

Here, $r_3^{(0)}(s)$ is calculated by substituting eq.(H2) in eq.(F32).

Next, we calculate $\rho^{(2\mathrm{q})}$, which is given by eq.(34b). We repeat the same calculations as eqs.(H8) and

(H9), but this time $q^{(0)}$ is replaced with $q^{(1)}$ there. From eq.(H10), the result is

$$\begin{aligned}\rho^{(2q)} &= \sum_s (1-\gamma_s) v^{(0)}(s) q^{(1)}(s) \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1) r_2^{(0)}(s) \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right. \\ &\qquad \left. - [1+(n_s-1) r_2^{(0)}(s)](1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} \\ &= \sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) \frac{q^{(1)}(s)}{q^{(0)}(s)} \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1) r_2^{(0)}(s) \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right. \\ &\qquad \left. - [1+(n_s-1) r_2^{(0)}(s)](1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}.\end{aligned} \tag{H14}$$

By substituting eqs.(H2) and (H5) into eq.(43) we obtain

$$\begin{aligned}\frac{q^{(1)}(s)}{q^{(0)}(s)} &= \left\{ \frac{1-d_{s,\mathrm{hard}}}{d_{s,\mathrm{hard}}} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1) r_2^{(0)}(s) \frac{1-d_{s,\mathrm{hard}}}{d_{s,\mathrm{hard}}} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right. \\ &\qquad \left. - [1+(n_s-1) r_2^{(0)}(s)] \frac{(1-d_{s,\mathrm{hard}})^2}{d_{s,\mathrm{hard}}} \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} \\ &\quad - \sum_{s'} \left\{ \frac{1-d_{s',\mathrm{hard}}}{d_{s',\mathrm{hard}}} \frac{\frac{\partial f_{s'}}{\partial z_1}}{f_{s'}} + (n_{s'}-1) r_2^{(0)}(s') \frac{1-d_{s',\mathrm{hard}}}{d_{s',\mathrm{hard}}} \frac{\frac{\partial f_{s'}}{\partial z_2}}{f_{s'}} \right. \\ &\qquad \left. - [1+(n_{s'}-1) r_2^{(0)}(s')] \frac{(1-d_{s',\mathrm{hard}})^2}{d_{s',\mathrm{hard}}} \left(\frac{1}{n_{s'}} \frac{\frac{\partial f_{s'}}{\partial z_1}}{f_{s'}} + \frac{n_{s'}-1}{n_{s'}} \frac{\frac{\partial f_{s'}}{\partial z_2}}{f_{s'}} \right) \right\} q^{(0)}(s').\end{aligned} \tag{H15}$$

Now we substitute eq.(H15) into eq.(H14). Before doing so, we observe that the second term in eq.(H15), which is made of the sum over s', is just a constant, and that putting a constant in the place of $q^{(1)}(s)/q^{(0)}(s)$ in eq.(H14) gives $\rho^{(1)}$ times this constant (see eq.(H10)), which is zero by the assumption in this subsection. Thus we can substitute only the first term of eq.(H15) into eq.(H14) to obtain $\rho^{(2q)}$, which leads to

$$\begin{aligned}\rho^{(2q)} &= \sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) \times \\ &\left\{ \frac{1-d_{s,\mathrm{hard}}}{d_{s,\mathrm{hard}}} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1) r_2^{(0)}(s) \frac{1-d_{s,\mathrm{hard}}}{d_{s,\mathrm{hard}}} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right. \\ &\qquad \left. - [1+(n_s-1) r_2^{(0)}(s)] \frac{(1-d_{s,\mathrm{hard}})^2}{d_{s,\mathrm{hard}}} \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} \times \\ &\left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1) r_2^{(0)}(s) \frac{\frac{\partial f_s}{\partial z_2}}{f_s} - [1+(n_s-1) r_2^{(0)}(s)](1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}.\end{aligned} \tag{H16}$$

Thirdly, we calculate $\rho^{(2r)}$, which is given by eq.(34c). For this, we repeat the calculation in eq.(H9), but this time $r_2^{(0)}$ is replaced with $r_2^{(1)}$ there. The result is

$$\rho^{(2r)} = \sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) \cdot (n_s-1) r_2^{(1)}(s) \left\{ \frac{\frac{\partial f_s}{\partial z_2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}. \tag{H17}$$

Here, we obtain $r_2^{(1)}(s)$ by substituting eqs.(H2) and (H5) in eq.(44), as

$$
\begin{aligned}
&r_2^{(1)}(s) \\
&= 2r_2^{(0)}(s)\frac{\gamma_s + (1-\gamma_s)(1-d_{s,\mathrm{hard}})}{2\gamma_s(1-d_{s,\mathrm{hard}}) + (1-\gamma_s)(1-d_{s,\mathrm{hard}})^2} \\
&\times \left\{ [1+(n_s-1)r_2^{(0)}(s)] \left[(1-d_{s,\mathrm{hard}})\frac{\frac{\partial f_s}{\partial z_1}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right] \right. \\
&\quad \left. + (n_s-1)[2r_2^{(0)}(s) + (n_s-2)r_3^{(0)}(s)] \left[(1-d_{s,\mathrm{hard}})\frac{\frac{\partial f_s}{\partial z_2}}{f_s} - (1-d_{s,\mathrm{hard}})^2 \left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right] \right\}.
\end{aligned}
\tag{H18}
$$

H.3 Consistency with previous results about perturbations

We here show that we recover several previous results from our model.

H.3.1 A model with overlapping generations by Lehmann and Rousset (2010)

Suppose there is a single habitat ($N = 1$) and no mortality in migration ($p_s = 1$). Then eq.(48) becomes

$$
\rho^{(1)} = (1-\gamma)\left\{ \frac{\frac{\partial f}{\partial z_1}}{f} + (n-1)r_2^{(0)}\frac{\frac{\partial f}{\partial z_2}}{f} - (1-m)^2\frac{1+(n-1)r_2^{(0)}}{n}\left(\frac{\frac{\partial f}{\partial z_1}}{f} + (n-1)\frac{\frac{\partial f}{\partial z_2}}{f} \right) \right\}, \tag{H19}
$$

which precisely recovers the inclusive fitness effect S_{IF} shown in eq.(A.21) in Lehmann and Rousset (2010) with the following correspondence; $\gamma \to s$, $(\partial f/\partial z_1)/f \to -C$, $r_2^{(0)} \to R$, $(n-1)(\partial f/\partial z_2)/f \to B$, and $\{1+(n-1)r_2^{(0)}\}/n \to R^{\mathrm{R}}$.

H.3.2 A spatially heterogeneous model by Rodrigues and Gardner (2012)

Rodrigues and Gardner (2012) studied effects of spatial and temporal heterogeneity of patch quality on the evolution of helping/harming. Their results on spatial heterogeneity (Results 1 and 2 therein) readily follow from our eq.(48). To recover them, set $N = 2$ (high/low quality patches) and $\mathcal{S} = \{1, 2\}$. Also, set $\gamma_1 = \gamma_2 = 0$, $n_1 = n_2 = n(\geq 2)$, $p_1 = p_2 = 1$, and $m_1 = m_2 = m(> 0)$, which leads to

$$
\begin{aligned}
d_1 &= \frac{m(\pi_1 f_1 + \pi_2 f_2)}{(1-m)f_1 + m(\pi_1 f_1 + \pi_2 f_2)} \\
d_2 &= \frac{m(\pi_1 f_1 + \pi_2 f_2)}{(1-m)f_2 + m(\pi_1 f_1 + \pi_2 f_2)}.
\end{aligned}
\tag{H20}
$$

Therefore, eq.(48) becomes

$$
\begin{aligned}
\rho^{(1)} = \left\langle \frac{f}{d} \right\rangle^{-1} \sum_{s=1}^{2} \frac{\pi_s f_s}{d_s} \Bigg\{ & \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1)r_2^{(0)}(s)\frac{\frac{\partial f_s}{\partial z_2}}{f_s} \\
& - \frac{1+(n_s-1)r_2^{(0)}(s)}{n_s}(1-d_s)^2 \left(\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1)\frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \Bigg\},
\end{aligned}
\tag{H21a}
$$

where

$$
\left\langle \frac{f}{d} \right\rangle \equiv \sum_{s=1}^{2} \pi_s \frac{f_s}{d_s}. \tag{H21b}
$$

For the Wright-Fisher update ($\gamma_1 = \gamma_2 = 0$), via a direct calculation of eq.(51) we can confirm

$$
\frac{1+(n_s-1)r_2^{(0)}(s)}{n_s}(1-d_s)^2 = r_2^{(0)}(s) \tag{H22}
$$

holds for $n \geq 2$. Simplifying eq.(H21) by using eq.(H22) leads to

$$\rho^{(1)} = \left\langle \frac{f}{d} \right\rangle^{-1} \sum_{s=1}^{2} \frac{\pi_s f_s}{d_s} \underbrace{\left\{1 - r_2^{(0)}(s)\right\}}_{(>0)} \left(\frac{\frac{\partial f_s}{\partial z_1}}{f_s} \right). \tag{H23}$$

Notice that effect on others' fecundity, $\partial f_s / \partial z_2$, is completely absent above. Thus we conclude that, as long as there is cost in helping/harming in both group states (*i.e.* $\partial f_s / \partial z_1 < 0$ for $s = 1, 2$), neither obligate nor facultative helping/harming evolves, which essentially echoes Results 1 and 2 in Rodrigues and Gardner (2012).

H.3.3 A cancellation result by Mullon et al. (2016)

Consider $N = 1$ (only one state) and the limit of $\gamma_s \to 1$ (Moran process). Also, suppose $\rho^{(1)} = 0$ (first-order perturbation of invasion fitness is null). This means that the expression inside the curly brackets of eq.(H10) is null. By applying eq.(51), this condition can be rewritten as

$$\frac{d_{s,\mathrm{hard}}(1 - d_{s,\mathrm{hard}} + n_s)}{1 + d_{s,\mathrm{hard}}(n_s - 1)} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{d_{s,\mathrm{hard}}(1 - d_{s,\mathrm{hard}})(n_s - 1)}{1 + d_{s,\mathrm{hard}}(n_s - 1)} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} = 0. \tag{H24}$$

Meanwhile, consider $r_2^{(1)}(s)$. It is proportional to the expression inside the curly brackets of eq.(H18), that is

$$\frac{n_s(1 - d_{s,\mathrm{hard}})}{2 + d_{s,\mathrm{hard}}(n_s - 2)} \times \left[\frac{d_{s,\mathrm{hard}}(1 - d_{s,\mathrm{hard}} + n_s)}{1 + d_{s,\mathrm{hard}}(n_s - 1)} \frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{d_{s,\mathrm{hard}}(1 - d_{s,\mathrm{hard}})(n_s - 1)}{1 + d_{s,\mathrm{hard}}(n_s - 1)} \frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right], \tag{H25}$$

which is zero. Hence $r_2^{(1)}(s) = 0$. This suggests that the first-order perturbation of relatedness in the Moran process is null, and therefore that the component of second-order perturbation of invasion fitness due to perturbation of relatedness is

$$\rho^{(2\mathrm{r})} = 0 \tag{H26}$$

in the Moran process.

Mullon et al. (2016) found essentially the same result (see their eq.(16)) for their Moran model, although their model was slightly different from ours here; Mullon et al. (2016) assumed that in each (non-extinct) patch exactly one adult individual always dies at each update step, whereas our model assumes that death occurs randomly to each individual and that it occurs rarely ($\gamma_s \to 1$).

H.3.4 Second-order results by Parvinen et al. (2018)

Parvinen et al. (2018) calculated the metapopulation fitness of mutants, R_m, in a subdivided population by assuming non-overlapping generations, $\gamma_s = 0$ (Wright-Fisher process), uniform migration rate ($m_s = m$), and uniform death rate during dispersal ($p_s = p$). It is known that $R_\mathrm{m} - 1$ has the same sign as $\rho - 1$ (Lehmann et al. (2016)), and therefore that metapopulation fitness can be used as a proxy to determine evolutionary success of mutants.

Parvinen et al. (2018) calculated a Taylor expansion of R_m with respect to mutational deviation, δ:

$$R_\mathrm{m} = 1 + \delta R_\mathrm{m}^{(1)} + \delta^2 R_\mathrm{m}^{(2)} + \cdots, \tag{H27}$$

where $R_{\mathrm{m}}^{(1)}$ corresponds to their $D_1(s_{\mathrm{res}})$ (see their eq.(3.5)), and $R_{\mathrm{m}}^{(2)}$ corresponds to their $D_2(s_{\mathrm{res}})/2$ (see their eq.(3.10)).

By using our eq.(H10) we can confirm (calculations are not shown here because they are too long) that

$$R_{\mathrm{m}}^{(1)} = \frac{\left\langle \frac{nf}{d_{\mathrm{hard}}} \right\rangle}{\langle nf \rangle} \rho^{(1)} \tag{H28}$$

holds, where

$$\left\langle \frac{nf}{d_{\mathrm{hard}}} \right\rangle \equiv \sum_s \pi_s \frac{n_s f_s(y, \boldsymbol{y}, y)}{d_{s,\mathrm{hard}}}, \qquad \langle nf \rangle \equiv \sum_s \pi_s n_s f_s(y, \boldsymbol{y}, y). \tag{H29}$$

This result was firstly shown by Parvinen et al. (2018) (see their eq.(B.26)). Similarly, when $\rho^{(1)} = 0$, by using eqs.(H13), (H16) and (H17) we can confirm (calculations are not shown here because they are too lengthy) that

$$R_{\mathrm{m}}^{(2)} = \frac{\left\langle \frac{nf}{d_{\mathrm{hard}}} \right\rangle}{\langle nf \rangle} \rho^{(2)} \tag{H30}$$

holds, as expected. These work as indirect confirmations that our results, eq.(H10) (for first-order) and eqs.(H13), (H16) and (H17) (for second-order), are correct.

I Perturbations for the soft selection lottery model

In this section, we derive the expressions for $\rho^{(1)}$ and $\rho^{(2)}$ for the lottery model under soft selection. Recall that for this model; namely, eqs.(38b) and (38c) with eqs.(45b) and (46b), we have

$$\begin{cases} w_{1,s|s}^{\mathrm{p}}(z_1, \boldsymbol{z}_{-\{1\}}, z) = \gamma_s + (1-\gamma_s) \underbrace{\dfrac{(1-m_s)n_s}{(1-m_s)n_s + I_{\mathrm{soft}}} \times \dfrac{f_s(z_1, \boldsymbol{z}_{-\{1\}}, z)}{\sum_{i=1}^{n_s} f_s(z_i, \boldsymbol{z}_{-\{i\}}, z)/n_s}}_{=w_{1,s|s}^{\mathrm{pr}}(z_1, \boldsymbol{z}_{-\{1\}}, z)} \\ w_{1,s'|s}^{\mathrm{d}}(z_1, \boldsymbol{z}_{-\{1\}}, z) = (1-\gamma_{s'}) \underbrace{\underbrace{\dfrac{\pi_{s'} n_{s'}}{(1-m_{s'})n_{s'} + I_{\mathrm{soft}}}}_{=w_{s'}^{\mathrm{col}}(z)} \cdot \underbrace{p_s m_s \dfrac{f_s(z_1, \boldsymbol{z}_{-\{1\}}, z)}{\sum_{i=1}^{n_s} f_s(z_i, \boldsymbol{z}_{-\{i\}}, z)/n_s}}_{=w_s^{\mathrm{em}}(z_1, \boldsymbol{z}_{-\{1\}}, z)}}_{=w_{1,s'|s}^{\mathrm{dr}}(z_1, \boldsymbol{z}_{-\{1\}}, z)} . \end{cases} \tag{I1}$$

I.1 First-order perturbation of invasion fitness

The goal here is to calculate the first-order perturbation of invasion fitness, given by eq.(32). For this purpose we will calculate its components one by one. First, we derive $v^{(0)}(s')q^{(0)}(s)$. With the definition of the backward migration probability (eq.(63)) we obtain

$$\begin{cases} w_{1,s|s}^{\mathrm{pr}} = 1 - d_{s,\mathrm{soft}} \\ w_{1,s'|s}^{\mathrm{dr}} = d_{s',\mathrm{soft}} \dfrac{\pi_{s'} n_{s'} p_s m_s}{I_{\mathrm{soft}}}. \end{cases} \tag{I2}$$

Substituting them into eq.(40) yields

$$\begin{aligned} v^{(0)}(s')q^{(0)}(s) &= \frac{\pi_s n_s p_{s'} m_{s'}}{(1-\gamma_{s'}) d_{s',\mathrm{soft}} I_{\mathrm{soft}}} \Bigg/ \left(\sum_{s''} \frac{\pi_{s''} n_{s''} p_{s''} m_{s''}}{(1-\gamma_{s''}) d_{s'',\mathrm{soft}} I_{\mathrm{soft}}} \right) \\ &= \frac{\pi_s n_s p_{s'} m_{s'}}{(1-\gamma_{s'}) d_{s',\mathrm{soft}}} \Bigg/ \left\langle \frac{npm}{(1-\gamma) d_{\mathrm{soft}}} \right\rangle, \end{aligned} \tag{I3}$$

where

$$\left\langle \frac{npm}{(1-\gamma)d_{\text{soft}}} \right\rangle \equiv \sum_s \pi_s \frac{n_s p_s m_s}{(1-\gamma_s)d_{s,\text{soft}}}. \tag{I4}$$

Second, we calculate first-order derivatives of 1-fitness that appear in eq.(32). A direct calculation of eq.(I1) shows that relevant first-order derivatives are

$$\begin{cases} \dfrac{\partial w^{\text{pr}}_{1,s|s}}{\partial z_1} = \underbrace{(1-d_{s,\text{soft}})}_{=w^{\text{pr}}_{1,s|s}} \left\{ \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} \\ \dfrac{\partial w^{\text{pr}}_{1,s|s}}{\partial z_2} = \underbrace{(1-d_{s,\text{soft}})}_{=w^{\text{pr}}_{1,s|s}} \left\{ \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} \\ \dfrac{\partial w^{\text{dr}}_{1,s'|s}}{\partial z_1} = \underbrace{d_{s',\text{soft}} \dfrac{\pi_{s'} n_{s'} p_s m_s}{I_{\text{soft}}}}_{=w^{\text{dr}}_{1,s'|s}} \left\{ \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} \\ \dfrac{\partial w^{\text{dr}}_{1,s'|s}}{\partial z_2} = \underbrace{d_{s',\text{soft}} \dfrac{\pi_{s'} n_{s'} p_s m_s}{I_{\text{soft}}}}_{=w^{\text{dr}}_{1,s'|s}} \left\{ \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}. \end{cases} \tag{I5}$$

Taking the first-order derivative of eq.(H6), substituting eq.(I5) therein, and using eq.(H6) again for rewriting give us

$$\begin{cases} \dfrac{\partial w_{1,s'|s}}{\partial z_1} = (w_{1,s'|s} - \delta_{s',s}\gamma_s) \left\{ \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} \\ \dfrac{\partial w_{1,s'|s}}{\partial z_2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s) \left\{ \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}. \end{cases} \tag{I6}$$

Finally, with eqs.(I3) and (I6) we can calculate $\rho^{(1)}$. We observe that each equation in eq.(I6) has the factor $(w_{1,s'|s} - \delta_{s',s}\gamma_s)$ in common. Thus the following relation is useful in the following calculation; for any function $F(s)$ we have

$$\begin{aligned} &\sum_{s'}\sum_{s} v^{(0)}(s') \left[(w_{1,s'|s} - \delta_{s',s}\gamma_s) F(s) \right] q^{(0)}(s) \\ =&\sum_{s} \underbrace{\sum_{s'} v^{(0)}(s') w_{1,s'|s}}_{\substack{=v^{(0)}(s) \\ \text{(from eq.(15b))}}} F(s) q^{(0)}(s) - \sum_s v^{(0)}(s)\gamma_s F(s) q^{(0)}(s) \\ =&\sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) F(s). \end{aligned} \tag{I7}$$

The first term in eq.(32) is then

$$\begin{aligned} &\sum_{s'}\sum_{s} v^{(0)}(s') \frac{\partial w_{1,s'|s}}{\partial z_1} q^{(0)}(s) \\ =&\sum_{s'}\sum_{s} v^{(0)}(s') (w_{1,s'|s} - \delta_{s',s}\gamma_s) \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} - \left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\} q^{(0)}(s) \\ \underset{\text{eq.(I7)}}{=}&\sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s) \left\{ \frac{\frac{\partial f_s}{\partial z_1}}{f_s} - \left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \right\}. \end{aligned} \tag{I8}$$

The second term in eq.(32) is similarly calculated as

$$\sum_{s'}\sum_{s} v^{(0)}(s')(n_s-1)\frac{\partial w_{1,s'|s}}{\partial z_2} r_2^{(0)}(s) q^{(0)}(s)$$
$$= \sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s)\cdot (n_s-1) r_2^{(0)}(s)\left\{\frac{\frac{\partial f_s}{\partial z_2}}{f_s} - \left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\}. \tag{I9}$$

By combining eqs.(I8) and (I9) we obtain

$$\rho^{(1)} = \sum_s (1-\gamma_s) v^{(0)}(s) q^{(0)}(s)\left\{\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1) r_2^{(0)}(s)\frac{\frac{\partial f_s}{\partial z_2}}{f_s} - [1+(n_s-1)r_2^{(0)}(s)]\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\}. \tag{I10}$$

Rewriting eq.(I10) with eqs.(50) and (I3) gives eq.(62) in the main text. Here, $r_2^{(0)}(s)$ is obtained by substituting eq.(I2) into eq.(41). We find that $r_2^{(0)}(s)$ takes exactly the same form as eq.(51) except that all $d_{s,\mathrm{hard}}$ there should be replaced by $d_{s,\mathrm{soft}}$.

I.2 Second-order perturbation of invasion fitness

Below we assume $\rho^{(1)} = 0$. The goal here is to calculate the second-order perturbation of invasion fitness, given by eq.(34). As in the hard selection case, we have already calculated first-order derivatives that appear there, as given in eq.(I6). We will then calculate various second-order derivatives that appear in eq.(34). A direct calculation of eq.(I1) shows that relevant terms are

$$\begin{cases}
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_1^2} = \underbrace{(1-d_{s,\mathrm{soft}})}_{=w^{\mathrm{pr}}_{1,s|s}}\left\{\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} - \left(\frac{1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right) - 2\frac{\frac{\partial f_s}{\partial z_1}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_2^2} = \underbrace{(1-d_{s,\mathrm{soft}})}_{=w^{\mathrm{pr}}_{1,s|s}}\left\{\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} - \left(\frac{1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right) - 2\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_1 z_2} = \underbrace{(1-d_{s,\mathrm{soft}})}_{=w^{\mathrm{pr}}_{1,s|s}}\left\{\frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} - \left(\frac{2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \frac{n_s-2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s}\right) - \left(\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
\frac{\partial^2 w^{\mathrm{pr}}_{1,s|s}}{\partial z_2 z_3} = \underbrace{(1-d_{s,\mathrm{soft}})}_{=w^{\mathrm{pr}}_{1,s|s}}\left\{\frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} - \left(\frac{2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \frac{n_s-2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s}\right) - 2\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}.\right)^2\right\}.
\end{cases} \tag{I11a}$$

and

$$
\begin{cases}
\dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_1^2} = \underbrace{d_{s',\mathrm{soft}} \dfrac{\pi_{s'} n_{s'} p_s m_s}{I_{\mathrm{soft}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} - \left(\dfrac{1}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \right) \right. \\
\qquad\qquad \left. - 2 \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
\dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_2^2} = \underbrace{d_{s',\mathrm{soft}} \dfrac{\pi_{s'} n_{s'} p_s m_s}{I_{\mathrm{soft}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} - \left(\dfrac{1}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \right) \right. \\
\qquad\qquad \left. - 2 \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
\dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_1 z_2} = \underbrace{d_{s',\mathrm{soft}} \dfrac{\pi_{s'} n_{s'} p_s m_s}{I_{\mathrm{soft}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} - \left(\dfrac{2}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s - 2}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \right. \\
\qquad\qquad \left. - \left(\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
\dfrac{\partial^2 w^{\mathrm{dr}}_{1,s'|s}}{\partial z_2 z_3} = \underbrace{d_{s',\mathrm{soft}} \dfrac{\pi_{s'} n_{s'} p_s m_s}{I_{\mathrm{soft}}}}_{=w^{\mathrm{dr}}_{1,s'|s}} \left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} - \left(\dfrac{2}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s - 2}{n_s} \dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \right. \\
\qquad\qquad \left. - 2 \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2 \left(\dfrac{1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s - 1}{n_s} \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\}.
\end{cases}
\tag{I11b}
$$

Taking the second-order derivative of eq.(H6), substituting eq.(I11) therein, and using eq.(H6) again for rewriting yields

$$
\begin{cases}
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_1^2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \right) \right. \\
\qquad\qquad \left. - 2\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s}\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_2^2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} - \left(\dfrac{1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} \right) \right. \\
\qquad\qquad \left. - 2\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_1 z_2} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} - \left(\dfrac{2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s-2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \right. \\
\qquad\qquad \left. - \left(\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\} \\
\dfrac{\partial^2 w_{1,s'|s}}{\partial z_2 z_3} = (w_{1,s'|s} - \delta_{s',s}\gamma_s)\left\{ \dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} - \left(\dfrac{2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \dfrac{n_s-2}{n_s}\dfrac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} \right) \right. \\
\qquad\qquad \left. - 2\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right) + 2\left(\dfrac{1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_1}}{f_s} + \dfrac{n_s-1}{n_s}\dfrac{\frac{\partial f_s}{\partial z_2}}{f_s} \right)^2 \right\}.
\end{cases}
\tag{I12}
$$

Now we are ready for calculating the second-order derivative of invasion fitness, given by eq.(34). We start from from $\rho^{(2\mathrm{w})}$, given by eq.(34a). Substituting eq.(I12) in eq.(34) and using eq.(I7) produces the

following four terms:

$$
\begin{cases}
Z_{1,\mathrm{soft}} \equiv \dfrac{1}{2}\displaystyle\sum_{s'}\sum_{s} v^{(0)}(s')\frac{\partial^2 w_{1,s'|s}}{\partial z_1^2} q^{(0)}(s) \\
\quad = \dfrac{1}{2}\displaystyle\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\left\{\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} - \left(\frac{1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right)\right. \\
\qquad \left. - 2\frac{\frac{\partial f_s}{\partial z_1}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
Z_{2,\mathrm{soft}} \equiv \dfrac{1}{2}\displaystyle\sum_{s'}\sum_{s} v^{(0)}(s')(n_s-1)\frac{\partial^2 w_{1,s'|s}}{\partial z_2^2} r_2^{(0)}(s)q^{(0)}(s) \\
\quad = \dfrac{1}{2}\displaystyle\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)r_2^{(0)}(s)\left\{\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s} - \left(\frac{1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1^2}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2^2}}{f_s}\right)\right. \\
\qquad \left. - 2\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
Z_{3,\mathrm{soft}} \equiv \dfrac{1}{2}\displaystyle\sum_{s'}\sum_{s} v^{(0)}(s')2(n_s-1)\frac{\partial^2 w_{1,s'|s}}{\partial z_1 z_2} r_2^{(0)}(s)q^{(0)}(s) \\
\quad = \dfrac{1}{2}\displaystyle\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\cdot 2(n_s-1)r_2^{(0)}(s)\left\{\frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} - \left(\frac{2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \frac{n_s-2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s}\right)\right. \\
\qquad \left. - \left(\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\} \\
Z_{4,\mathrm{soft}} \equiv \dfrac{1}{2}\displaystyle\sum_{s'}\sum_{s} v^{(0)}(s')(n_s-1)(n_s-2)\frac{\partial^2 w_{1,s'|s}}{\partial z_2 z_3} r_3^{(0)}(s)q^{(0)}(s) \\
\quad = \dfrac{1}{2}\displaystyle\sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)(n_s-2)r_3^{(0)}(s)\left\{\frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s} - \left(\frac{2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_1 z_2}}{f_s} + \frac{n_s-2}{n_s}\frac{\frac{\partial^2 f_s}{\partial z_2 z_3}}{f_s}\right)\right. \\
\qquad \left. - 2\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right) + 2\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)^2\right\},
\end{cases}
\tag{I13a}
$$

and we have

$$
\rho^{(2\mathrm{w})} = Z_{1,\mathrm{soft}} + Z_{2,\mathrm{soft}} + Z_{3,\mathrm{soft}} + Z_{4,\mathrm{soft}}. \tag{I13b}
$$

Here, $r_3^{(0)}(s)$ is calculated by substituting eq.(I2) in eq.(F32).

Next, we calculate $\rho^{(2\mathrm{q})}$, given by eq.(34b). We repeat the same calculations as eqs.(I8) and (I9), but $q^{(0)}$ there should be replaced with $q^{(1)}$. From eq.(I10) we have

$$
\begin{aligned}
\rho^{(2\mathrm{q})} &= \sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(1)}(s)\left\{\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1)r_2^{(0)}(s)\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right. \\
&\qquad \left. - [1+(n_s-1)r_2^{(0)}(s)]\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\} \\
&= \sum_{s}(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\frac{q^{(1)}(s)}{q^{(0)}(s)}\left\{\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + (n_s-1)r_2^{(0)}(s)\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right. \\
&\qquad \left. - [1+(n_s-1)r_2^{(0)}(s)]\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s} + \frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\}.
\end{aligned}
\tag{I14}
$$

By substituting eqs.(I2) and (I5) in eq.(43) we obtain $q^{(1)}(s)/q^{(0)}(s)$ as

$$\begin{aligned}\frac{q^{(1)}(s)}{q^{(0)}(s)} &= \frac{1-d_{s,\mathrm{soft}}}{d_{s,\mathrm{soft}}}\left\{\frac{\frac{\partial f_s}{\partial z_1}}{f_s}+(n_s-1)r_2^{(0)}(s)\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right.\\ &\qquad\left.-[1+(n_s-1)r_2^{(0)}(s)]\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s}+\frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\}\\ &\quad-\sum_{s'}\frac{1-d_{s',\mathrm{soft}}}{d_{s',\mathrm{soft}}}\left\{\frac{\frac{\partial f_{s'}}{\partial z_1}}{f_{s'}}+(n_{s'}-1)r_2^{(0)}(s')\frac{\frac{\partial f_{s'}}{\partial z_2}}{f_{s'}}\right.\\ &\qquad\left.-[1+(n_{s'}-1)r_2^{(0)}(s')]\left(\frac{1}{n_{s'}}\frac{\frac{\partial f_{s'}}{\partial z_1}}{f_{s'}}+\frac{n_{s'}-1}{n_{s'}}\frac{\frac{\partial f_{s'}}{\partial z_2}}{f_{s'}}\right)\right\}q^{(0)}(s').\end{aligned} \tag{I15}$$

Now we substitute eq.(I15) into eq.(I14). Note that the second term in eq.(I15), which consists of the sum over s', is just a constant, and that putting a constant in the place of $q^{(1)}(s)/q^{(0)}(s)$ in eq.(I14) gives $\rho^{(1)}$ times this constant (see eq.(I10)), which is zero by the assumption in this subsection. Thus we can substitute only the first term of eq.(I15) into eq.(I14) to obtain $\rho^{(2\mathrm{q})}$, by which we obtain

$$\begin{aligned}\rho^{(2\mathrm{q})} &= \sum_s(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\frac{1-d_{s,\mathrm{soft}}}{d_{s,\mathrm{soft}}}\left\{\frac{\frac{\partial f_s}{\partial z_1}}{f_s}+(n_s-1)r_2^{(0)}(s)\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right.\\ &\qquad\left.-[1+(n_s-1)r_2^{(0)}(s)]\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s}+\frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\}^2.\end{aligned} \tag{I16}$$

Quite notably, $\rho^{(2\mathrm{q})}$ in eq.(I16) is always non-negative, which is apparently seen from its expression.

Thirdly, we calculate $\rho^{(2\mathrm{r})}$, given by eq.(34c). For this, we repeat the calculation in eq.(I9), but $r_2^{(0)}$ there should be replaced by $r_2^{(1)}$. The result is

$$\rho^{(2\mathrm{r})} = \sum_s(1-\gamma_s)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)r_2^{(1)}(s)\left\{\frac{\frac{\partial f_s}{\partial z_2}}{f_s}-\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s}+\frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right\}. \tag{I17}$$

Here, we obtain $r_2^{(1)}(s)$ by substituting eqs.(I2) and (I5) in eq.(44), as

$$\begin{aligned}r_2^{(1)}(s) &= 2r_2^{(0)}(s)\frac{\gamma_s+(1-\gamma_s)(1-d_{s,\mathrm{soft}})}{2\gamma_s(1-d_{s,\mathrm{soft}})+(1-\gamma_s)(1-d_{s,\mathrm{soft}})^2}\\ &\quad\times\left\{[1+(n_s-1)r_2^{(0)}(s)](1-d_{s,\mathrm{soft}})\left[\frac{\frac{\partial f_s}{\partial z_1}}{f_s}-\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s}+\frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right]\right.\\ &\qquad\left.+(n_s-1)[2r_2^{(0)}(s)+(n_s-2)r_3^{(0)}(s)](1-d_{s,\mathrm{soft}})\left[\frac{\frac{\partial f_s}{\partial z_2}}{f_s}-\left(\frac{1}{n_s}\frac{\frac{\partial f_s}{\partial z_1}}{f_s}+\frac{n_s-1}{n_s}\frac{\frac{\partial f_s}{\partial z_2}}{f_s}\right)\right]\right\}.\end{aligned} \tag{I18}$$

I.3 Consistency with previous results

Here, we again show that we recover previous results.

I.3.1 A model with "Regulation before dispersal" by Lehmann and Rousset (2010)

When there is a single habitat type ($N=1$), no mortality in migration ($p_s=1$), and no overlap of generations ($\gamma_s=0$), eq.(62) reduces to

$$\rho^{(1)} = \left(1-r_2^{(0)}\right)\left\{\frac{\frac{\partial f}{\partial z_1}}{f}-\frac{1}{n}\left(\frac{\frac{\partial f}{\partial z_1}}{f}+(n-1)\frac{\frac{\partial f}{\partial z_2}}{f}\right)\right\}, \tag{I19}$$

which reproduces eq.(A.7) of Lehmann and Rousset (2010) for their "regulation before dispersal" model with the following correspondence; $r_2^{(0)}\to R$, $(\partial f/\partial z_1)/f\to -C$, $n\to N$, and $(n-1)(\partial f/\partial z_2)/f\to B$.

I.3.2 A model with local adaptation by Svardal et al. (2015)

Svardal et al. (2015) studied a soft selection model with spatial and temporal environmental heterogeneity and effectively infinitely large group size. In the absence of temporal heterogeneity, their model fits our soft selection framework by setting $\pi_s = c_s$, $n_s = n_0(\to \infty)$, $p_s = 1$, $m_s = m$, $\gamma_s = \gamma$ and $f_s(z_1, \boldsymbol{z}_{-\{1\}}, z) = f_s(z_1)$. Then our eq.(62) predicts

$$\left.\frac{\mathrm{d}\rho}{\mathrm{d}x}\right|_{x=y} = \rho^{(1)} = (1-\gamma)\sum_s c_s \frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s} = (1-\gamma)\sum_s c_s \frac{\partial}{\partial x}\left(\frac{f_s(x)}{f_s(y)}\right)\Bigg|_{x=y}. \tag{I20}$$

In the notation of Svardal et al. (2015) this can be written as $(1-\gamma)\mathrm{E_s}[\partial\rho]$, which equals their first-order derivative of invasion fitness (see their Appendix B.1). Similarly, when the first-order derivative is null, eqs.(I13), (I16), and (I17) predict

$$\left.\frac{\mathrm{d}^2\rho}{\mathrm{d}x^2}\right|_{x=y} = 2\rho^{(2)} = (1-\gamma)\sum_s c_s \left\{\frac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} + 2\frac{1-m}{m}\left(\frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2\right\}. \tag{I21}$$

In the notation of Svardal et al. (2015) this can be written as $(1-\gamma)\{\mathrm{E_S}[\partial^2\rho] + \{2(1-m)/m\}\mathrm{E_S}[(\partial\rho)^2]\}$, which agrees with their second-order derivative of invasion fitness, because in the absence of temporal heterogeneity their eq.(B.27) becomes

$$\begin{aligned}
&(1-\gamma)\left[\mathrm{E_S}[\partial^2\rho - (\partial\rho)^2] + \mathrm{Var_S}[\partial\rho] + 2\frac{1-m}{m}\mathrm{Var_S}[\partial\rho]\right] \\
&= (1-\gamma)\left[\mathrm{E_S}[\partial^2\rho - (\partial\rho)^2] + \mathrm{E_S}[(\partial\rho)^2] - \underbrace{\mathrm{E_S}[\partial\rho]^2}_{=0} + 2\frac{1-m}{m}\left\{\mathrm{E_S}[(\partial\rho)^2] - \underbrace{\mathrm{E_S}[\partial\rho]^2}_{=0}\right\}\right] \\
&= (1-\gamma)\left[\mathrm{E_S}[\partial^2\rho] + 2\frac{1-m}{m}\mathrm{E_S}[(\partial\rho)^2]\right].
\end{aligned}$$

J Evolutionary analysis of a lottery model with local adaptation

In this section, we derive the results for the special case presented in section 4.3 of the main text based on the assumptions that (i) the demography follows the assumptions in section 4.2 (either hard or soft selection), (ii) the Moran process limit is considered ($\gamma_s = \gamma \sim 1$ for all s), and (iii) fecundity of an adult only depends on its trait value, as $f_s(z_1, \boldsymbol{z}_{-\{1\}}, z) = f_s(z_1)$.

J.1 Hard selection

Analysis under assumptions (i) to (iii): Under these assumptions, eq.(48) simplifies to

$$\rho^{(1)} \propto \sum_s \frac{\pi_s n_s p_s m_s f_s}{d_{s,\mathrm{hard}}} \frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\left\{1 - r_{2,\mathrm{R}}^{(0)}(s)(1-d_{s,\mathrm{hard}})^2\right\}, \tag{J1a}$$

where the proportionality constant is

$$\left(\sum_{s\in\mathcal{S}} \pi_s \frac{n_s p_s m_s f_s}{(1-\gamma)d_{s,\mathrm{hard}}}\right)^{-1}. \tag{J1b}$$

Remember that in eq.(J1), $d_{s,\mathrm{hard}}$, f_s and its derivative, and $r_{2,\mathrm{R}}^{(0)}(s)$ all depend on y. A singular strategy y^* is the one at which $\rho^{(1)}$ vanishes.

Next we study the second order effect of selection at the singular point y^* by using eqs.(H13), (H16) and (H17). Note that our f_s depends only on its bearer's trait value, so most of the derivatives of f_s that appear there are null. We start from eq.(H13):

$$\begin{cases} \rho^{(2\mathrm{w})} = Z_{1,\mathrm{hard}} + Z_{2,\mathrm{hard}} + Z_{3,\mathrm{hard}} + Z_{4,\mathrm{hard}}, \qquad \text{where} \\ Z_{1,\mathrm{hard}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma)v^{(0)}(s)q^{(0)}(s)\times \\ \qquad \left\{ \dfrac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} - \dfrac{(1-d_{s,\mathrm{hard}})^2}{n_s}\dfrac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} - 2\dfrac{(1-d_{s,\mathrm{hard}})^2}{n_s}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 + 2\dfrac{(1-d_{s,\mathrm{hard}})^3}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\} \\ Z_{2,\mathrm{hard}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)r_2^{(0)}(s)\times \\ \qquad \left\{ -\dfrac{(1-d_{s,\mathrm{hard}})^2}{n_s}\dfrac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} + 2\dfrac{(1-d_{s,\mathrm{hard}})^3}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\} \\ Z_{3,\mathrm{hard}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma)v^{(0)}(s)q^{(0)}(s)\cdot 2(n_s-1)r_2^{(0)}(s)\times \\ \qquad \left\{ -\dfrac{(1-d_{s,\mathrm{hard}})^2}{n_s}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 + 2\dfrac{(1-d_{s,\mathrm{hard}})^3}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\} \\ Z_{4,\mathrm{hard}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)(n_s-2)r_3^{(0)}(s)\times \left\{ 2\dfrac{(1-d_{s,\mathrm{hard}})^3}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\}. \end{cases} \tag{J2}$$

Next we calculate eq.(H16):

$$\rho^{(2\mathrm{q})} = \sum_s (1-\gamma)v^{(0)}(s)q^{(0)}(s)\times \left\{ \frac{1-d_{s,\mathrm{hard}}}{d_{s,\mathrm{hard}}} - [1+(n_s-1)r_2^{(0)}(s)]\frac{(1-d_{s,\mathrm{hard}})^2}{d_{s,\mathrm{hard}}n_s} \right\} \times \left\{ 1-[1+(n_s-1)r_2^{(0)}(s)]\frac{(1-d_{s,\mathrm{hard}})^2}{n_s} \right\} \left(\frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2. \tag{J3}$$

Then, we calculate eq.(H17):

$$\rho^{(2\mathrm{r})} = \sum_s (1-\gamma)v^{(0)}(s)q^{(0)}(s)\cdot(n_s-1)r_2^{(1)}(s)\times \left\{ -\frac{(1-d_{s,\mathrm{hard}})^2}{n_s}\frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s} \right\}. \tag{J4}$$

Relatedness values are calculated from eqs.(51), (42) (with $w^{\mathrm{pr}}_{1,s|s} = 1-d_{s,\mathrm{hard}}$; see eq.(H2)), and (H18) respectively, as

$$\begin{cases} r_2^{(0)}(s) = \dfrac{1-d_{s,\mathrm{hard}}}{1+d_{s,\mathrm{hard}}(n_s-1)} \\ r_3^{(0)}(s) = \dfrac{2(1-d_{s,\mathrm{hard}})^2}{\{2+d_{s,\mathrm{hard}}(n_s-2)\}\{1+d_{s,\mathrm{hard}}(n_s-1)\}} \\ r_2^{(1)}(s) = \dfrac{d_{s,\mathrm{hard}}(1-d_{s,\mathrm{hard}})(1-d_{s,\mathrm{hard}}+n_s)n_s}{\{2+d_{s,\mathrm{hard}}(n_s-2)\}\{1+d_{s,\mathrm{hard}}(n_s-1)\}^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right). \end{cases} \tag{J5}$$

By using eqs.(J2)–(J5) and eq.(H3), and after doing some algebra, we arrive at the expression eq.(52) in the main text.

With two additional assumptions: Here, we add two more assumptions to make our model fully tractable: (iv) group size n_s, the probability to migrate m_s, and survival of migrating individuals p_s are identical across habitats ($n_s = n$, $m_s = m$, and $p_s = p$ for all s), and (v) fecundity in groups in habitat s is given by eq.(53) in the main text.

For the selection gradient, by using eqs.(50), (53) and (J5), we calculate eq.(J1) as

$$\rho^{(1)} \propto \sum_s \pi_s f_s(y) \frac{y_{\mathrm{op},s} - y}{\sigma_{\mathrm{st}}^2} \frac{1 - d_{s,\mathrm{hard}}(y) + n}{1 + d_{s,\mathrm{hard}}(y) \cdot (n-1)}, \tag{J6a}$$

where the proportionality constant is

$$\left(\sum_{s \in \mathcal{S}} \pi_s \frac{f_s(y)}{(1-\gamma) d_{s,\mathrm{hard}}(y)} \right)^{-1}, \tag{J6b}$$

and the dependence of f_s and $d_{s,\mathrm{hard}}$ on y is made explicit. The singular strategy y^* is only implicitly solved as

$$y^* = \sum_s \psi_s(y^*) y_{\mathrm{op},s}, \tag{J7}$$

which is a weighted average of $y_{\mathrm{op},s}$, the optimal trait values in groups in state s, with the weights ψ_s given by

$$\psi_s(y^*) \equiv \pi_s f_s(y^*) \frac{1 - d_{s,\mathrm{hard}}(y^*) + n}{1 + d_{s,\mathrm{hard}}(y^*) \cdot (n-1)} \Big/ \sum_{s'} \pi_{s'} f_{s'}(y^*) \frac{1 - d_{s',\mathrm{hard}}(y^*) + n}{1 + d_{s',\mathrm{hard}}(y^*) \cdot (n-1)}. \tag{J8}$$

For the disruptive selection coefficient at the singular strategy, y^*, a direct calculation of derivatives of eq.(53) in eq.(52a) leads to

$$\begin{aligned} \rho^{(2)} &\propto \sum_s \frac{\pi_s f_s(y^*)}{d_{s,\mathrm{hard}}(y^*)} \left\{ X_{1,s,\mathrm{hard}}(y^*) \frac{(y_{\mathrm{op},s} - y^*)^2 - \sigma_{\mathrm{st}}^2}{\sigma_{\mathrm{st}}^4} + X_{2,s,\mathrm{hard}}(y^*) \frac{(y_{\mathrm{op},s} - y^*)^2}{\sigma_{\mathrm{st}}^4} \right\} \\ &\propto \sum_s \frac{\pi_s f_s(y^*)}{d_{s,\mathrm{hard}}(y^*)} \left\{ (X_{1,s,\mathrm{hard}}(y^*) + X_{2,s,\mathrm{hard}}(y^*))(y_{\mathrm{op},s} - y^*)^2 - X_{1,s,\mathrm{hard}}(y^*) \sigma_{\mathrm{st}}^2 \right\}, \end{aligned} \tag{J9}$$

so the condition for $\rho^{(2)} > 0$ can be expressed as

$$\sum_s \Psi_s(y^*)(y_{\mathrm{op},s} - y^*)^2 > \sigma_{\mathrm{st}}^2, \tag{J10}$$

where

$$\begin{aligned} \Psi_s(y^*) &\equiv \frac{\pi_s f_s(y^*)}{d_{s,\mathrm{hard}}(y^*)} (X_{1,s,\mathrm{hard}}(y^*) + X_{2,s,\mathrm{hard}}(y^*)) \Big/ \sum_{s'} \frac{\pi_{s'} f_{s'}(y^*)}{d_{s',\mathrm{hard}}(y^*)} X_{1,s',\mathrm{hard}}(y^*) \\ &= \pi_s f_s(y^*) \frac{(1 - d_{s,\mathrm{hard}}(y^*) + n)\{2(n+1) - d_{s,\mathrm{hard}}(y^*) \cdot (n+2)\}}{\{2 + d_{s,\mathrm{hard}}(y^*) \cdot (n-2)\}\{1 + d_{s,\mathrm{hard}}(y^*) \cdot (n-1)\}} \\ &\qquad \Big/ \sum_{s'} \pi_{s'} f_{s'}(y^*) \frac{1 - d_{s',\mathrm{hard}}(y^*) + n}{1 + d_{s',\mathrm{hard}}(y^*) \cdot (n-1)}. \end{aligned} \tag{J11}$$

Further assumptions: Here we make three more assumptions: (vi) no mortality during dispersal, $p = 1$, (vii) two habitats with equal proportions, $\mathcal{S} = \{1, 2\}$ and $\pi_1 = \pi_2 = 1/2$, and (viii) optima are symmetric in the sense of $y_{\mathrm{op},2} = -y_{\mathrm{op},1}$.

In this case $y^* = 0$ is a singular strategy due to symmetry. To see this, from the symmetry of the fecundity functions $f_1(y)$ and $f_2(y)$, we have $f_1(0) = f_2(0)$ (see eq.(53)). Then, from the definition of $d_{s,\mathrm{hard}}$ (eq.(49); see also eq.(47)), we have

$$d_{s,\mathrm{hard}}(0) = \frac{nm \frac{f_1(0) + f_2(0)}{2}}{(1-m) n f_s(0) + nm \frac{f_1(0) + f_2(0)}{2}} = m \tag{J12}$$

for $s = 1, 2$. Substituting these into eq.(J8) gives us $\psi_1(0) = \psi_2(0) = 1/2$, and hence eq.(J7) is satisfied.

We are interested under what conditions $\rho^{(2)} > 0$ holds. By using $\pi_1 = \pi_2 = 1/2$, $f_1(0) = f_2(0)$ and eq.(J12), the weight given by eq.(J11) can be written as

$$\Psi_s(0) = \frac{1}{2}\left(\frac{2-m}{m} - \frac{4(1-m)^2}{m(2+m(n-2))}\right). \tag{J13}$$

Therefore, when we define the variance of the habitat optima σ^2_{op} by eq.(57), the condition (J10) can be re-written as eq.(58) in the main text (note that we use $\pi_1 = \pi_2 = 1/2$ there).

Finally, we determine convergence stability of $y^* = 0$. Calculating eq.(19) for the selection gradient eq.(J6) gives us, after some algebra,

$$c(0) = \left.\frac{\mathrm{d}\rho^{(1)}(y)}{\mathrm{d}y}\right|_{y=0} \propto (1+m(n-1))(1-m+n)(\sigma^2_{\text{op}} - \sigma^2_{\text{st}}) + n^2 m(1-m)\sigma^2_{\text{op}} \tag{J14}$$

with a proportionality constant that is positive for $m > 0$. Rearranging the convergence stability condition, $c(0) < 0$, gives condition (60) in the main text. To obtain a sufficient condition for convergence stability, we substitute eq.(61) into the coefficient of σ^2_{op} in (60) and obtain

$$\left.\left(2 - m - \underbrace{\frac{(1-m)^2(1-m+(1+m)n)}{(1+m(n-1))(1-m+n)}}_{\to 0 \text{ when } n\to\infty}\right)\right|_{m=\frac{\sqrt{1+n}}{n+\sqrt{1+n}}} = \frac{2\sqrt{1+n}}{1+\sqrt{1+n}}, \tag{J15}$$

and therefore

$$\frac{2\sqrt{1+n}}{1+\sqrt{1+n}}\sigma^2_{\text{op}} < \sigma^2_{\text{st}} \tag{J16}$$

is a sufficient condition for convergence stability for fixed n. In addition, since eq.(J15) is upper-bounded by two, $2\sigma^2_{\text{op}} < \sigma^2_{\text{st}}$ is a sufficient condition for convergence stability for any n.

J.2 Soft selection

Analysis under assumptions (i) to (iii): Under these assumptions, eq.(62) simplifies to

$$\rho^{(1)} \propto \sum_s \frac{\pi_s n_s p_s m_s}{d_{s,\text{soft}}} \frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\left\{1 - r^{(0)}_{2,\text{R}}(s)\right\}, \tag{J17a}$$

where the proportionality constant is

$$\left(\sum_{s\in\mathcal{S}} \pi_s \frac{n_s p_s m_s}{(1-\gamma)d_{s,\text{soft}}}\right)^{-1}. \tag{J17b}$$

Remember that in eq.(J17), f_s, its derivative, and $r^{(0)}_{2,\text{R}}(s)$ depend on the resident strategy y, whereas $d_{s,\text{soft}}$ is independent of y (see eqs.(47) and (63)). A singular strategy y^* is the one at which $\rho^{(1)}$ vanishes.

The disruptive selection coefficient at the singular point y^* is calculated through eqs.(I13), (I16) and (I17) under the assumption that f_s depends only on its bearer's trait value, so most of the derivatives

of f_s that appear in $\rho^{(2)}$ are null. In particular, from eq.(I13), we have

$$\begin{cases} \rho^{(2\mathrm{w})} = Z_{1,\mathrm{soft}} + Z_{2,\mathrm{soft}} + Z_{3,\mathrm{soft}} + Z_{4,\mathrm{soft}}, \qquad \text{where} \\ Z_{1,\mathrm{soft}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma) v^{(0)}(s) q^{(0)}(s) \times \\ \qquad \left\{ \dfrac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} - \dfrac{1}{n_s}\dfrac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} - \dfrac{2}{n_s}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 + \dfrac{2}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\} \\ Z_{2,\mathrm{soft}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma) v^{(0)}(s) q^{(0)}(s) \cdot (n_s - 1) r_2^{(0)}(s) \times \\ \qquad \left\{ -\dfrac{1}{n_s}\dfrac{\frac{\mathrm{d}^2 f_s}{\mathrm{d}z_1^2}}{f_s} + \dfrac{2}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\} \\ Z_{3,\mathrm{soft}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma) v^{(0)}(s) q^{(0)}(s) \cdot 2(n_s - 1) r_2^{(0)}(s) \times \\ \qquad \left\{ -\dfrac{1}{n_s}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 + \dfrac{2}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\} \\ Z_{4,\mathrm{soft}} = \dfrac{1}{2}\displaystyle\sum_s (1-\gamma) v^{(0)}(s) q^{(0)}(s) \cdot (n_s - 1)(n_s - 2) r_3^{(0)}(s) \times \left\{ \dfrac{2}{n_s^2}\left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s}\right)^2 \right\}. \end{cases} \tag{J18}$$

Second, eq.(I16) becomes

$$\rho^{(2\mathrm{q})} = \sum_s (1-\gamma) v^{(0)}(s) q^{(0)}(s) \frac{1 - d_{s,\mathrm{soft}}}{d_{s,\mathrm{soft}}} \left\{ 1 - \frac{1 + (n_s - 1) r_2^{(0)}(s)}{n_s} \right\}^2 \left(\frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s} \right)^2. \tag{J19}$$

Third, eq.(I17) becomes

$$\rho^{(2\mathrm{r})} = \sum_s (1-\gamma) v^{(0)}(s) q^{(0)}(s) \cdot (n_s - 1) r_2^{(1)}(s) \times \left\{ -\frac{1}{n_s} \left(\frac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s} \right) \right\}. \tag{J20}$$

Relatedness values are calculated in the following way: $r_2^{(0)}(s)$ is from eq.(51) (where all $d_{s,\mathrm{hard}}$ are replaced by $d_{s,\mathrm{soft}}$; see section 4.3.2 in the main text), $r_3^{(0)}(s)$ is from eq.(42) (with $w_{1,s|s}^{\mathrm{pr}} = 1 - d_{s,\mathrm{soft}}$; see eq.(I2)), and $r_2^{(1)}(s)$ is from eq.(I18) combined with $r_2^{(0)}(s)$ and $r_3^{(0)}(s)$ as just derived. We then obtain

$$\begin{cases} r_2^{(0)}(s) = \dfrac{1 - d_{s,\mathrm{soft}}}{1 + d_{s,\mathrm{soft}}(n_s - 1)}, \\ r_3^{(0)}(s) = \dfrac{2(1 - d_{s,\mathrm{soft}})^2}{\{2 + d_{s,\mathrm{soft}}(n_s - 2)\}\{1 + d_{s,\mathrm{soft}}(n_s - 1)\}}, \\ r_2^{(1)}(s) = \dfrac{d_{s,\mathrm{soft}}(1 - d_{s,\mathrm{soft}}) n_s (n_s - 1)}{\{2 + d_{s,\mathrm{soft}}(n_s - 2)\}\{1 + d_{s,\mathrm{soft}}(n_s - 1)\}^2} \left(\dfrac{\frac{\mathrm{d}f_s}{\mathrm{d}z_1}}{f_s} \right). \end{cases} \tag{J21}$$

By using eqs.(J18)–(J21) and eq.(I3), and after some algebra, we arrive at the expression eq.(64) in the main text.

With two additional assumptions: Similarly to the hard selection case, we here add two more assumptions: (iv) group size n_s, the probability to migrate m_s, and survival of migrating individuals p_s are identical across habitats ($n_s = n$, $m_s = m$, and $p_s = p$ for all s), and (v) fecundity in groups in habitat s is explicitly given by eq.(53) in the main text.

With these assumptions, from eqs.(47) and (63) we see that

$$d_{s,\mathrm{soft}} = \frac{pm}{(1 - m) + pm} \tag{J22}$$

does not depend on s, and hence we write it as d_{soft} from now on. An immediate consequence is that $X_{1,s,\text{soft}}$ and $X_{2,s,\text{soft}}$ in eqs.(64b) and (64c) are independent of s, and we write them as

$$X_{1,\text{soft}} = \frac{1}{2}\frac{d_{\text{soft}}(n-1)}{1+d_{\text{soft}}(n-1)} \tag{J23a}$$

$$X_{2,\text{soft}} = \frac{d_{\text{soft}}(n-1)\{d_{\text{soft}}(1-d_{\text{soft}})(n-1)(n-2)-2d_{\text{soft}}(n-1)+(n-2)\}}{\{2+d_{\text{soft}}(n-2)\}\{1+d_{\text{soft}}(n-1)\}^2}. \tag{J23b}$$

For the selection gradient, by using eqs.(50), (53) and (J21), we calculate eq.(J17) as

$$\rho^{(1)} = (1-\gamma)\frac{d_{\text{soft}}(n-1)}{1+d_{\text{soft}}(n-1)}\sum_s \pi_s \frac{y_{\text{op},s}-y}{\sigma_{\text{st}}^2}, \tag{J24}$$

and therefore, the singular strategy is simply given by

$$y^* = \sum_s \pi_s y_{\text{op},s}. \tag{J25}$$

For convergence stability, we take the derivative of eq.(J24) with respect to y and obtain

$$c(y^*) = \left.\frac{\mathrm{d}\rho^{(1)}(y)}{\mathrm{d}y}\right|_{y=y^*} = (1-\gamma)\frac{d_{\text{soft}}(n-1)}{1+d_{\text{soft}}(n-1)}\left(-\frac{1}{\sigma_{\text{st}}^2}\right) < 0, \tag{J26}$$

which shows that a singular strategy is always convergence stable. This parallels the results of Svardal et al. (2015).

For disruptive selection coefficient at the singular strategy y^* a direct calculation of eq.(53) in eq.(64a) leads to

$$\begin{aligned}\rho^{(2)} &\propto \sum_s \pi_s \left\{X_{1,\text{soft}}\frac{(y_{\text{op},s}-y^*)^2-\sigma_{\text{st}}^2}{\sigma_{\text{st}}^4} + X_{2,\text{soft}}\frac{(y_{\text{op},s}-y^*)^2}{\sigma_{\text{st}}^4}\right\} \\ &\propto \sum_s \pi_s \left\{(X_{1,\text{soft}}+X_{2,\text{soft}})(y_{\text{op},s}-y^*)^2 - X_{1,\text{soft}}\sigma_{\text{st}}^2\right\},\end{aligned} \tag{J27}$$

and the condition for $\rho^{(2)} > 0$ becomes

$$\frac{X_{1,\text{soft}}+X_{2,\text{soft}}}{X_{1,\text{soft}}}\underbrace{\sum_s \pi_s (y_{\text{op},s}-y^*)^2}_{=\sigma_{\text{op}}^2 \text{ (from eq.(57))}} > \sigma_{\text{st}}^2. \tag{J28}$$

No mortality in dispersal: If we further assume that (vi) no mortality occurs during dispersal, $p = 1$, then substituting $p = 1$ into eq.(J22) gives $d_{\text{soft}} = m$. Inserting this into eq.(J23) and calculating condition (J28) with the expressions obtained for $X_{1,\text{soft}}$ and $X_{2,\text{soft}}$ results, after rearranging terms, in condition (66) in the main text.

References

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

Harris, T. E. 1963. The Theory of Branching Processes. Springer-Verlag Berlin Heidelberg.

Karlin, S., and H. M. Taylor. 1975. A First Course in Stochastic Processes. Academic Press, San Diego.

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

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

Lehmann, L., and F. Rousset. 2010. How life-history and demography promote or inhibit the evolution of helping behaviors. Philosophical Transactions of the Royal Society B 365:2599–2617.

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

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

Ohtsuki, H. 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. Evolution 64:3344–53.

Parvinen, K., H. Ohtsuki, and J. Y. Wakano. 2018. Spatial heterogeneity and evolution of fecundity-affecting traits. Journal of Theoretical Biology 454:190–204.

Pen, I. 2000. Reproductive effort in viscous populations. Evolution 54:293–297.

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

Rousset, F. 2004. Genetic Structure and Selection in Subdivided Populations, vol. 40 of *Monographs in Population Biology*. Princeton University Press, Princeton, N.J.

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

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

Wakano, J. Y., and L. Lehmann. 2014. Evolutionary branching in deme-structured populations. Journal of Theoretical Biology 351:83–95.