bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

An evolutionary quantitative genetics model for phenotypic (co)variance under limited dispersal, with an application to socially synergistic traits

Charles Mullon*[1] and Laurent Lehmann[1]

[1] *Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland*

Abstract

Darwinian evolution consists of the gradual transformation of heritable traits due to natural selection and the input of random variation by mutation. Here, we use a quantitative genetics approach to investigate the coevolution of multiple quantitative traits under selection, mutation, and limited dispersal. We track the dynamics of trait means and variance-covariances between traits that experience frequency-dependent selection. Assuming a multivariate-normal trait distribution, we recover classical dynamics of quantitative genetics, as well as stability and evolutionary branching conditions of invasion analyses, except that due to limited dispersal, selection depends on indirect fitness effects and relatedness. In particular, correlational selection that associates different traits *within*-individuals depends on the fitness effects of such associations *between*-individuals. We find that these kin selection effects can be as relevant as pleiotropy for the evolution of correlation between traits. We illustrate this with an example of the coevolution of two social traits whose association within-individuals is costly but synergistically beneficial between-individuals. As dispersal becomes limited and relatedness increases, associations between-traits between-individuals become increasingly targeted by correlational selection. Consequently, the trait distribution goes from being bimodal with a negative correlation under panmixia to unimodal with a positive correlation under limited dispersal.

Keywords. Island model, G-matrix evolution, evolutionary branching, division of labour, social evolution

1 Introduction

Understanding how heritable quantitative traits are molded by natural selection and mutation has been a long-standing goal of evolutionary biology. This research endeavour has led to an abundant theoretical literature that seeks to understand the roles of ecology and genetics in the gradual transformation of quantitative phenotypes. Notwithstanding this abundance, models of gradual evolution usually follow one of two approaches, depending on whether the focus is put on ecological or genetic processes.

One approach consists in investigating the invasion success of a rare phenotypic mutant (i.e., an evolutionary invasion analysis, e.g. Michod, 1979, Eshel and Feldman, 1984, Parker and Smith, 1990, Eshel et al., 1997; also referred to as "Adaptive Dynamics", e.g., Dercole and Rinaldi, 2008, for a textbook treatment) and places emphasis on ecology (or on how organisms interact with one another via effects on resources and the environment). In most practical applications, this emphasis comes at the expense of genetics realism. In particular, trait dynamics inferred from invasion analyses most often assume that mutations have weak quantitative effects and are so rare (relative to the strength of selection) that at most two alleles can segregate in the population. In this case, a sensitivity analysis of the invasion fitness of a rare mutant in a resident monomorphic

*Electronic address: charles.mullon@unil.ch

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

population that is at its ecological equilibrium (e.g., Michod, 1979, Eshel and Motro, 1981, Eshel and Feldman, 1984, Taylor, 1989, Parker and Smith, 1990, Charlesworth, 1994) can be used to understand gradual trait evolution and the ecological transformations due to this evolution (Metz et al., 1996, Geritz et al., 1998, Rousset, 2004, Dercole and Rinaldi, 2008, Metz, 2011). Evolutionary invasion analysis is therefore particularly well-suited to investigate the evolution of traits under ecological feedbacks and the frequency-dependent selection that emerges due to such feedbacks (e.g., Kisdi and Geritz, 2009, Lion, 2018, and references therein). This approach has revealed that in the presence of trade-offs, gradual evolution under ecological feedbacks often leads to the emergence of polymorphism. Here, the population evolves under directional selection to express a trait value such that any rare mutant has an advantage over the common resident (Eshel and Motro, 1981, Eshel, 1983, Taylor, 1989, Christiansen, 1991, Abrams et al., 1993b). As a result, the population subsequently splits into two lineages of distinct phenotypes, or morphs, in a process referred to as evolutionary branching (Geritz et al., 1998; see Rueffler et al., 2006, Kisdi and Geritz, 2009, for reviews).

By contrast to invasion analysis, evolutionary quantitative genetics models of gradual evolution tend to be more preoccupied with the genetic basis of traits (Roff, 1997, Lynch and Walsh, 1998). Importantly, quantitative genetics models envisage that substantial heritable phenotypic variation segregates in the population. The continuum-of-alleles model, in particular, posits that quantitative traits are determined by a continuum of possible alleles produced by mutation (e.g., Kimura, 1965b, Latter, 1970, Fleming, 1979, Bürger, 1986). A quantitative genetics approach aims to investigate the roles of selection and mutation in the gradual evolution of a phenotypic distribution of arbitrary complexity. Due to the complication of dealing with multiple phenotypic variants, however, analytical explorations of quantitative genetics models usually come at the expense of generality. Notably, the vast majority of quantitative genetics models of traits under frequency-dependent selection, which either is implicit or emerges from ecological interactions, focuses on the evolution of mean phenotypic values in the population, assuming that heritable phenotypic variation is constant (i.e., additive genetic variances and covariances are fixed, e.g., Lande, 1976, 1981, Iwasa et al., 1991, Gomulkiewicz and Kirkpatrick, 1992, Abrams et al., 1993a, Iwasa and Pomiankowski, 1995, Day and Taylor, 1996, Tazzyman and Iwasa, 2009, Nuismer et al., 2010, Connallon, 2015).

But phenotypic variance should be especially sensitive to frequency-dependent selection. This is because such selection either favors or disfavors rare variants that differ from the most common, and thus either increases or decreases trait variance (Slatkin, 1980, Taper and Chase, 1985, Taylor and Day, 1997, Day and Proulx, 2004, Sasaki and Dieckmann, 2011, Wakano and Iwasa, 2013, Wakano and Lehmann, 2014, Débarre et al., 2014, Débarre and Otto, 2016). In fact, recent quantitative genetics models investigating populations of individuals experiencing frequency-dependent interactions have revealed links between the dynamics of phenotypic variance and evolutionary branching (Sasaki and Dieckmann, 2011, Wakano and Iwasa, 2013, Wakano and Lehmann, 2014, Débarre et al., 2014, Débarre and Otto, 2016), thereby extending the links between the dynamics of the phenotypic mean in quantitative genetics models and directional selection in invasion analysis models (Charlesworth, 1990, Iwasa et al., 1991, Taper and Case, 1992, Abrams et al., 1993a; for reviews: Abrams, 2001, Lion, 2018). Specifically, evolutionary branching occurs in a quantitative genetics model when the phenotypic variance is predicted to grow without bound while the phenotypic mean is held constant, under the assumption that the phenotypic distribution is normal (this assumption allows to only have to consider the dynamics of the mean and variance of the phenotypic distribution, Wakano and Iwasa, 2013, Wakano and Lehmann, 2014, Débarre et al., 2014, Débarre and Otto, 2016). As evolutionary branching occurs, the variance may in fact converge to a bounded value (see Fig. 2e-f of Débarre and Otto, 2016), but these dynamics cannot be captured by models that assume that the phenotypic distribution is normal and thus unimodal (instead of a bi- or multi-modal distribution; see Sasaki and Dieckmann, 2011 and Appendix D of Débarre and Otto, 2016 for a relaxation of the unimodal assumption). In spite of this limitation, quantitative genetics approaches have been useful to investigate relevant factors for frequency-dependent selection and evolutionary branching, such as genetic drift (with fixed, Wakano and Iwasa, 2013, or fluctuating, Débarre and Otto, 2016, population size) or the interaction between multiple traits (Débarre et al., 2014).

One factor that is particularly relevant for frequency-dependent interactions is limited dispersal. This is because limited dispersal creates genetic structure, whereby individuals that interact and compete with one another are more likely to share identical alleles at loci determining social or competitive traits than individuals

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

randomly sampled from the population, resulting in kin selection on traits (Hamilton, 1964, Michod, 1982, Frank, 1998, Rousset, 2004). Using an invasion analysis, a number of models have investigated the conditions that lead to disruptive selection (usually followed by evolutionary branching) due to frequency-dependent interactions among individuals under limited dispersal (Day, 2001, Ajar, 2003, Rousset, 2004, Mullon et al., 2016, Parvinen et al., 2018; see also Svardal et al., 2015 for evolutionary branching due to spatial and temporal heterogeneities in selection but without kin selection). Using a quantitative genetics approach, Wakano and Lehmann (2014) (WL14 hereafter) found branching conditions equivalent to those obtained from invasion analysis by studying the dynamics of the variance of a trait under limited dispersal. The analysis of frequency-dependent and disruptive selection under limited dispersal has helped reveal further connections between invasion analysis and fundamental branches of evolutionary theory. In particular, Ajar (2003), WL14 and Mullon et al. (2016) expressed disruptive selection coefficients in terms of relatedness coefficients, which are quantities central to population genetics, kin selection and social evolution theory (i.e., the evolution of traits that influence the fitness of their actor and recipient, such as helping or harming, e.g., Hamilton, 1964, Frank, 1998, Rousset, 2004, Wenseleers, 2010; see also Kisdi, 2016 for a kin selection perspective on evolutionary branching of dispersal).

In this paper, we incorporate two additional factors that have previously been omitted in the gradual evolution of quantitative traits when selection is frequency-dependent and dispersal is limited. First, we consider the joint evolution of multiple traits (whereas WL14 focuses on a single trait). This enables us to investigate how phenotypic covariances among traits within individuals are molded by frequency-dependent selection and pleiotropic mutations (i.e., when traits share a common genetic basis so that mutations have correlated effects across traits). Second, we model the coupled dynamics of the phenotypic means and (co)variances (whereas WL14 looks at the dynamics of the variance only once selection on means is negligible). This allows for a more complete picture of the dynamics of the phenotypic distribution. By expressing these dynamics in terms of relatedness coefficients, we further connect kin selection theory with the evolutionary quantitative genetics of multiple traits (Lande, 1979, Lande and Arnold, 1983, Phillips and Arnold, 1989, Brodie et al., 1995; in particular with the evolution of the **G**-matrix of additive genetic variance-covariance, Steppan et al., 2002, Arnold et al., 2008)

The rest of this paper is organized as follows. We describe the life-cycle and population structure under consideration in section 2. Our first result, presented in section 3.1, is an equation for the one-generational change of a multi-variate phenotypic distribution under limited dispersal, mutation, and selection. Next, in section 3.2, we present a closed dynamical system for the mean vector and variance-covariance matrix of the phenotypic distribution, under the assumption that the distribution in the whole population is normal. Further, we express this dynamical system in terms of effects on individual fitness and relatedness in section 3.3, and highlight some equilibrium properties of our dynamical system in section 3.4. In section 3.5, we apply our framework to study the coevolution of two traits that have socially synergistic effects between individuals. Finally, we discuss the implications of our results for understanding patterns of intra-specific variation, with special reference to social and competitive traits.

2 Model

We consider a population of haploid individuals, divided among an infinite number of groups, each of fixed size N (the total population size is therefore constant). Each individual bears a multidimensional phenotype that consists of n genetically determined quantitative traits. The discrete-time life cycle of this population is as follows. (1) Groups may go extinct (in which case all N adult individuals in a group die before reproduction) and do so independently of one another. (2) Adults reproduce clonally (producing offspring in sufficient number for the size of each group in the population to be N by the end of the life cycle) then either survive or die, which frees up breeding spots. (3) The phenotype of each individual independently mutates with probability ν, causing a random quantitative deviation in trait values. (4) Each offspring either remains in its natal group, or disperses to another randomly chosen group (i.e., we consider the island model of dispersal, Wright, 1931,

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Rousset, 2004). (5) Offspring compete locally in each group to fill open breeding spots, if any.

This life-cycle allows for one, several, or all adults to die per life-cycle iteration (including through whole group extinction before reproduction). Generations can thus overlap but the expression of traits is assumed to be independent of age (e.g., the fertility or mortality of an individual is independent from its age and that of any other individual it interacts with). Dispersal can occur before or after density-dependent competition (as long as the number of adults in each groups remains constant), and in groups, so that more than one offspring from the same natal patch can establish in a non-natal patch. This life cycle is equivalent to that considered in Mullon et al. 2016, except that here, we allow for the constant input of mutations in the population (step 3 of the life cycle).

3 Results

3.1 Dynamics of the phenotypic distribution

In order to track phenotypic evolution in the population, we denote by $p_t(\mathbf{z})$ the phenotypic density distribution in the population at a demographic time point t, where $\mathbf{z} = (z_1, z_2, \ldots, z_n) \in \mathbb{R}^n$ is a vector collecting the variable z_a for each trait $a = 1, \ldots, n$. To capture group structure, we introduce the density distribution ϕ_t of phenotypic groups states in the population at time t (i.e., $\phi_t(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$ is the probability density distribution of groups in which individuals arbitrarily labelled 1 to N have phenotypes $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$, respectively, see Appendix eq. A-3 for more details).

In Appendix A, we show that the recurrence equation for the phenotypic distribution in the population from demographic time step t to $t+1$ (one iteration of the life cycle) can be expressed as

$$p_{t+1}(\mathbf{z}) = (1-\nu) W(\mathbf{z}, \phi_t) p_t(\mathbf{z}) + \nu \int_{\mathbb{R}^n} \upsilon(\mathbf{z}', \mathbf{z}) W(\mathbf{z}', \phi_t) p_t(\mathbf{z}') \mathrm{d}\mathbf{z}'. \tag{1}$$

The first summand represents changes in the distribution due to reproduction and survival of individuals that have not mutated (with probability $1-\nu$), and the second summand, changes due to those that have (with probability ν; and where $\upsilon(\mathbf{z}', \mathbf{z})$ denotes the probability density function for the event that an individual mutates from $\mathbf{z}'$ to $\mathbf{z}$ given that a mutation has occurred). The quantity $W(\mathbf{z}, \phi_t)$ in eq. (1) is a measure of fitness of phenotype $\mathbf{z}$, which depends on the way phenotypes are distributed across groups (i.e., on ϕ_t). When $W(\mathbf{z}, \phi_t) > 1$, the frequency of $\mathbf{z}$ in the population increases due to selection, and conversely decreases when $W(\mathbf{z}, \phi_t) < 1$.

To gain insights into the fitness measure $W(\mathbf{z}, \phi_t)$, note that recurrence eq. (1) has the same form as the classical recurrence of the phenotypic distribution in well-mixed populations under the continuum-of-alleles model (e.g., Kimura, 1965b, eqs. 1-2; Fleming, 1979, eq. 2.4; Bürger, 1986, eq. 1; Taylor and Day, 1997, eq. 1; Champagnat et al., 2006, eq. 4.1). In a well-mixed population of constant size, the fitness of phenotype $\mathbf{z}$ is equal to the *individual* fitness of a focal individual with phenotype $\mathbf{z}$; namely, its expected number of successful offspring produced over one iteration of the life-cycle (including self through survival). Because individuals interact at random in a well-mixed population, such individual fitness function only depends on the population wide phenotypic distribution, $p_t(\mathbf{z})$ (i.e., the phenotype of any group neighbor to a focal individual, captured by ϕ_t, is in fact independently and identically distributed according to $p_t(\mathbf{z})$). The fitness of an individual with phenotype $\mathbf{z}$ in a population with trait distribution $p_t(\mathbf{z})$ can thus be written as $w(\mathbf{z}, p_t(\mathbf{z}))$, and $W(\mathbf{z}, \phi_t) = w(\mathbf{z}, p_t(\mathbf{z}))$ in eq. (1) (to distinguish between fitness at the phenotype and individual level, we generically denote the former by an upper case W and the latter by a lower case w).

Defining individual fitness in terms of expected number of successful offspring is standard in social evolution theory (e.g., Hamilton, 1964, Rousset, 2004), and takes its roots in population dynamics: when $w(\mathbf{z}, p_t(\mathbf{z})) > 1$, the number of individuals with phenotype $\mathbf{z}$ increases and conversely decreases when $w(\mathbf{z}, p_t(\mathbf{z})) < 1$ (e.g., eq. 2.2 of Nagylaki 1992). As such, it is sometimes referred to as "absolute" fitness. Many quantitative ge-

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

netics models, by contrast, employ the notion of "relative" fitness to track changes in phenotypic frequencies. This can stem from two non-mutually exclusive modelling choices: (1) one in fact considers the effect of the phenotype on a vital rate, $f(\mathbf{z}, p_t(\mathbf{z}))$ (such as fecundity or offspring survival), that influences the number of offspring that enter competition before regulation, which requires normalisation by mean vital rate, $W(\mathbf{z}, \phi_t) = w(\mathbf{z}, p_t(\mathbf{z})) = f(\mathbf{z}, p_t(\mathbf{z}))/[\int f(\mathbf{z}, p_t(\mathbf{z}))p_t(\mathbf{z})\mathrm{d}\mathbf{z}]$; (2) the population size fluctuates, in which case it is necessary to normalise by mean fitness, $W(\mathbf{z}, \phi_t) = w(\mathbf{z}, p_t(\mathbf{z}))/[\int w(\mathbf{z}, p_t(\mathbf{z}))p_t(\mathbf{z})\mathrm{d}\mathbf{z}]$. In our model, because group size and therefore population size is constant, $W(\mathbf{z}, \phi_t)$ in eq. (1) can be viewed as an absolute measure of fitness.

In contrast to a well-mixed population, the fitness of an individual w in a dispersal-limited group-structured population depends on the way phenotypes are distributed across groups (so on ϕ_t), and specifically on the collection of phenotypes carried by the individuals that belong to its own group. The fitness of an individual with phenotype $\mathbf{z}$ in a population with group distribution given by ϕ_t can thus be written as $w_\mu(\mathbf{z}, \phi_t)$, where μ is the collection of phenotypes carried by individuals in the focal group (formally, μ is a counting measure in our analysis – see Appendix A.1.2 – but for the purpose of the main text, it can simply be thought of as the phenotypic composition of the focal group). In terms of this individual fitness function, we find that the fitness at the level of the phenotype that is relevant for phenotypic dynamics, $W(\mathbf{z}, \phi_t)$ in eq. (1), is

$$W(\mathbf{z}, \phi_t) = \int w_\mu(\mathbf{z}, \phi_t) q(\mu|\mathbf{z}, \phi_t)\mathrm{d}\mu, \tag{2}$$

where the integral runs through every possible group state, μ, and $q(\mu|\mathbf{z}, \phi_t)$ is the probability density function for the event that an individual randomly picked from the collection of all carriers of the $\mathbf{z}$ phenotype in the population at time t resides in a group in state μ (see eq. A-17 in Appendix A for derivation). According to eq. (2), $W(\mathbf{z}, \phi_t)$ is the average expected number of successful offspring of an individual with phenotype $\mathbf{z}$, where the average is taken over all group states μ in which an individual with phenotype $\mathbf{z}$ can reside at time t.

An alternative interpretation for $W(\mathbf{z}, \phi_t)$ can be reached by noting that because there is an infinite number of possible alleles, all individuals with the same phenotype $\mathbf{z}$ belong to the same genetic lineage (as the same allele cannot appear twice via mutation). The function $q(\mu|\mathbf{z}, \phi_t)$ in eq. (2) then corresponds to the probability that an individual sampled from this lineage at time t resides in a group in state μ. As such, $W(\mathbf{z}, \phi_t)$ can be interpreted as the average direct fitness of an individual randomly sampled from the lineage of individuals carrying phenotype $\mathbf{z}$ at time t. If on average individuals from the $\mathbf{z}$-lineage produce more than one successful offspring at time t, this lineage will be larger at time $t+1$ and in a population of constant size, the frequency of individuals with phenotype $\mathbf{z}$ will increase. The fitness measure $W(\mathbf{z}, \phi_t)$ can thus be seen as the multi-allelic version of the concept of a mutant's *lineage fitness* used previously in invasion analyses (which turns out to be equal to the mutant's growth rate when the mutant is rare in an otherwise monomorphic population, Mullon et al., 2016, Lehmann et al., 2016; see also Wild, 2011 for similar branching processes approach to social evolution in group-structured populations). We will therefore refer to $W(\mathbf{z}, \phi_t)$ as the lineage fitness (or average direct fitness) of phenotype $\mathbf{z}$, keeping in mind that unlike in invasion analyses, $W(\mathbf{z}, \phi_t)$ here applies for any frequency of $\mathbf{z}$ (rare or common) and for any population composition (monomorphic or polymorphic).

3.2 Tracking the dynamics of the phenotypic distribution

The dynamical equation for the phenotypic distribution eq. (1) has no straightforward solution, even when the population is well-mixed (Kimura, 1965b, Fleming, 1979, Lande, 1979, Bürger, 1986). Under limited dispersal, this problem is further complicated by the necessity of simultaneously tracking the dynamics of group composition ϕ_t. To proceed further in our analysis and track the dynamics of the phenotypic distribution, we therefore make additional assumptions.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

3.2.1 Weak selection, weak mutation, normal closure and quasi-equilibrium of local genetic associations

We first assume that selection is weak, in the sense that the phenotypic variance in the population is small (allowing for second-order approximation of lineage fitness, see Appendix B.1.1 for details, and Iwasa et al., 1991 for a similar approach for quantitative genetics of traits under frequency-dependent selection in well-mixed populations). This enables us to express lineage fitness in terms of time-dependent local genetic associations among individuals of the same group (i.e., relatedness coefficients), which capture relevant moments of the distribution of group composition ϕ_t, and therefore avoids us having to keep track of the full distribution (see eq. B-1-B-4). Next, we assume that mutations are rare, so that we can ignore the joint effects of selection and mutation on the phenotypic distribution over one time period (see Appendix B.1.2).

Following previous authors (e.g., see Taylor and Day, 1997, Wakano and Iwasa, 2013, Débarre et al., 2014, Wakano and Lehmann, 2014, Débarre and Otto, 2016, for social traits), we further assume that the processes of selection and mutation are such that $p_t(\mathbf{z})$ is approximately multivariate normal (allowing for moment closure, see Appendix B.2.1). The assumption of normality is a strong one but it is noteworthy that it does not require that the realized distribution of phenotypes within a focal group at any given demographic time period is normal. In addition, the assumption of normality has been shown to give accurate predictions for the change of mean and variance, which is our main goal, even when selection generates significant deviations from normality (in well-mixed populations, Turelli and Barton, 1994). Under the assumption of normality, the distribution $p_t(\mathbf{z})$ is characterised by its mean vector $\bar{\mathbf{z}}_t = (\bar{z}_{1,t}, \bar{z}_{2,t}, \ldots, \bar{z}_{n,t})$, whose a-entry is the average value of trait a in the population at time period t, $\bar{z}_{a,t} = \int_{\mathbb{R}^n} z_a p_t(\mathbf{z})\mathrm{d}\mathbf{z}$; and its variance-covariance matrix $\boldsymbol{G}_t$ whose (a,b)-entry is the (co)variance among traits a and b in the population at time period t, $\sigma_{ab,t} = \int_{\mathbb{R}^n} (z_a - \bar{z}_{a,t})(z_b - \bar{z}_{b,t}) p_t(\mathbf{z})\mathrm{d}\mathbf{z}$. The dynamics of $p_t(\mathbf{z})$ can therefore be tracked through the dynamics of its mean vector $\bar{\mathbf{z}}_t$ and variance-covariance matrix $\boldsymbol{G}_t$.

But due to limited dispersal, the dynamics of $\bar{\mathbf{z}}_t$ and $\boldsymbol{G}_t$ still depend on time-dependent local genetic associations among individuals of the same group. To close evolutionary dynamics on $\bar{\mathbf{z}}_t$ and $\boldsymbol{G}_t$ and avoid tracking the dynamics of these genetic associations, we assume that selection is weak relative to dispersal so that genetic associations reach their steady state before significant changes have occurred in the phenotypic distribution, $p_t(\mathbf{z})$ (see section B.2.2 for details). This "quasi-equilibrium" assumption, which is frequently used in population genetic theory (e.g., Kimura, 1965a, Nagylaki, 1993, Kirkpatrick et al., 2002, Roze and Rousset, 2005, 2008), finally allows us to characterize the dynamics of $p_t(\mathbf{z})$ entirely by the coupled dynamics of its mean vector $\bar{\mathbf{z}}_t$ and variance-covariance matrix $\boldsymbol{G}_t$.

3.2.2 Dynamics of phenotypic mean vector and variance-covariance matrix

Under the above assumptions, we show in Appendix B that the coupled changes of the mean trait vector and variance-covariance matrix over one demographic time period are respectively given by

$$\Delta\bar{\mathbf{z}}_t = \boldsymbol{G}_t \boldsymbol{s}(\bar{\mathbf{z}}_t) \tag{3a}$$

$$\Delta\boldsymbol{G}_t = \boldsymbol{M} + \boldsymbol{G}_t \left(\boldsymbol{H}(\bar{\mathbf{z}}_t) - \boldsymbol{s}(\bar{\mathbf{z}}_t)\boldsymbol{s}(\bar{\mathbf{z}}_t)^{\mathrm{T}} \right) \boldsymbol{G}_t, \tag{3b}$$

where $\boldsymbol{s}(\bar{\mathbf{z}}_t) = (s_1(\bar{\mathbf{z}}_t), \ldots, s_n(\bar{\mathbf{z}}_t))^{\mathrm{T}}$ ($.^{\mathrm{T}}$ denotes the transpose of a vector or matrix) is a $n \times 1$ is vector of directional selection coefficients (or selection gradients), i.e., $s_a(\bar{\mathbf{z}}_t)$ is the first-order, marginal, effect of an (infinitesimal) change in trait a away from the population mean $\bar{\mathbf{z}}_t$ on lineage fitness ($s_a(\bar{\mathbf{z}}_t) = \partial W(\mathbf{z}, \phi_t)/\partial z_a$). The $n \times n$ matrix $\boldsymbol{M}$ collects the effects of mutation; its (a,b)-entry,

$$(\boldsymbol{M})_{ab} = \nu \underbrace{\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (z_a - z'_a)(z_b - z'_b) \upsilon(\mathbf{z}', \mathbf{z}) \mathrm{d}\mathbf{z}' \mathrm{d}\mathbf{z}}_{\sigma^{\mathrm{m}}_{ab}}, \tag{4}$$

is the product of the mutation probability, ν, with the (co)variance, σ^{m}_{ab}, in mutational effects on traits a and b conditional on the appearance of a mutation (which captures the pleiotropic effects of mutations on a and b: when $\sigma^{\mathrm{m}}_{ab} > 0$, mutations tend to change a and b in a similar way; and when $\sigma^{\mathrm{m}}_{ab} < 0$, in opposite ways).

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The $n \times n$ Hessian matrix $\boldsymbol{H}(\bar{\mathbf{z}}_t)$ collects the second-order effects of traits on lineage fitness; its (a,b)-entry $\boldsymbol{H}(\bar{\mathbf{z}}_t)_{ab} = h_{ab}(\bar{\mathbf{z}}_t)$ is the marginal effect of joint changes in traits a and b away from the population mean $\bar{\mathbf{z}}_t$ on lineage fitness ($h_{ab}(\bar{\mathbf{z}}_t) = \partial^2 W(\mathbf{z},\phi_t)/(\partial z_a \partial z_b)$). Finally, the notation $\boldsymbol{s}(\bar{\mathbf{z}}_t)\boldsymbol{s}(\bar{\mathbf{z}}_t)^{\mathrm{T}}$ denotes the outer product between two column vectors, so that $\boldsymbol{s}(\bar{\mathbf{z}}_t)\boldsymbol{s}(\bar{\mathbf{z}}_t)^{\mathrm{T}}$ is $n \times n$ matrix with (a,b)-entry $s_a(\bar{\mathbf{z}}_t)s_b(\bar{\mathbf{z}}_t)$.

3.2.3 Directional, disruptive, and correlational selection coefficients

Dynamical equations (3) have the same form as in well-mixed populations (e.g., eqs. 1-2 of Phillips and Arnold, 1989, see also eq. 7 of Lande, 1979 and eqs. 6 and 15 of Lande and Arnold, 1983). In such models, the effects of selection depend on the marginal effects of traits on *individual* rather than *lineage* fitness. Nevertheless, the parallels between eq. (3) and previous works allow us to use the same vocabulary and interpretations on the evolution of phenotypic means and (co)variances (Brodie et al., 1995). First, the evolution of the mean of each trait (eq. 3a) depends on the vector of directional selection (or the selection gradient), $\boldsymbol{s}(\bar{\mathbf{z}}_t)$, which points in the direction favored by selection in multivariate phenotypic space (Lande, 1979). The effect of directional selection on the mean of each trait, however, is constrained by the genetic variation available and these constraints are captured by $\boldsymbol{G}_t$ in eq. (3a) (Lande, 1979).

Second, the evolution of the variance-covariance matrix $\boldsymbol{G}_t$ (eq. 3b) depends on the effects of mutations ($\boldsymbol{M}$), of directional selection ($\boldsymbol{s}(\bar{\mathbf{z}}_t)\boldsymbol{s}(\bar{\mathbf{z}}_t)^{\mathrm{T}}$), and of quadratic selection given by the matrix $\boldsymbol{H}(\bar{\mathbf{z}}_t)$ (Lande, 1979, Lande and Arnold, 1983, Phillips and Arnold, 1989). This matrix $\boldsymbol{H}(\bar{\mathbf{z}}_t)$ captures two relevant features of selection. First, the sign of its diagonal entry (a,a) indicates whether selection favors a decrease (when $h_{aa}(\bar{\mathbf{z}}_t) < 0$) or an increase (when $h_{aa}(\bar{\mathbf{z}}_t) > 0$) in the variance of trait a when this trait evolves in isolation of other traits (Phillips and Arnold, 1989), hence $h_{aa}(\bar{\mathbf{z}}_t)$ is referred to as the coefficient of disruptive selection on trait a. Second, the off-diagonal entry (a,b) tells us whether selection favors a positive (when $h_{ab}(\bar{\mathbf{z}}_t) > 0$) or negative (when $h_{ab}(\bar{\mathbf{z}}_t) < 0$) covariance or correlation among traits a and b. The off-diagonal entry $h_{ab}(\bar{\mathbf{z}}_t)$ is therefore referred to as the coefficient of correlational selection among traits a and b (Lande and Arnold, 1983, Phillips and Arnold, 1989).

3.3 Selection in terms of individual fitness effects and relatedness coefficients

So far, the effects of limited dispersal on evolutionary dynamics (eqs. 1 and 3) have been hidden behind the notion of lineage fitness, $W(\mathbf{z},\phi_t)$. To highlight more tangibly how selection depends on limited dispersal, we express the selection coefficients ($\boldsymbol{s}(\bar{\mathbf{z}}_t)$ and $\boldsymbol{H}(\bar{\mathbf{z}}_t)$) in terms of the effects of traits on individual fitness and relatedness. For this, let us first rewrite the individual fitness of a focal individual, that we label as individual "i", as a function $w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)$ of three arguments: (1) the phenotype of the focal individual, $\mathbf{z}_i = (z_{i,1}, z_{i,2}, \ldots, z_{i,n})$; (2) the collection of phenotypes of its $N-1$ neighbors $\mathbf{z}_{-i} = (\mathbf{z}_1, \ldots \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \ldots, \mathbf{z}_N)$ (where $\mathbf{z}_j = (z_{j,1}, z_{j,2}, \ldots, z_{j,n})$ is the phenotype of a neighbor indexed j); and (3) the average phenotype in the population $\bar{\mathbf{z}}_t$ (see eq. 15 for an example of such a fitness function). This individual fitness function is equal to the fitness function $w_\mu(\mathbf{z},\phi_t)$ that appears in eq. (2),

$$w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t) = w_\mu(\mathbf{z},\phi_t), \tag{5}$$

when focal phenotype is $\mathbf{z}_i = \mathbf{z}$, the state of the focal group is $\mu = \{\mathbf{z}_i\} \cup \mathbf{z}_{-i} = (\mathbf{z}_1, \ldots, \mathbf{z}_N)$, and groups other than the focal one are considered to be monomorphic for the population average $\bar{\mathbf{z}}_t$ (i.e., we consider that all individuals in other groups express $\bar{\mathbf{z}}_t$ so that the distribution ϕ_t is delta peaked on $\bar{\mathbf{z}}_t$; we can do this because the phenotypic distribution is assumed to be centered around $\bar{\mathbf{z}}_t$ with small variance and individuals from different groups interact at random in the island model; see Iwasa et al., 1991 for a similar approach in panmictic populations).

We further introduce two neutral relatedness coefficients that will be relevant for selection: let $r_2^\circ(\bar{\mathbf{z}}_t)$ and $r_3^\circ(\bar{\mathbf{z}}_t)$ respectively be the probabilities that, in the absence of selection and when the population phenotypic average is $\bar{\mathbf{z}}_t$, one and two neighbors of a focal individual carry a phenotype that is identical-by-descent to that of the focal (i.e., the set of individuals under consideration have a common ancestor). Alternatively, $r_2^\circ(\bar{\mathbf{z}}_t)$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and $r_3^\circ(\bar{\mathbf{z}}_t)$ can be interpreted as the probabilities that, in the absence of selection and when the population phenotypic average is $\bar{\mathbf{z}}_t$, two and three individuals sampled in the same group carry identical-by-descent phenotypes. This interpretation is in line with the definition of relatedness in the infinite island model (see e.g., Rousset, 2004, Taylor et al., 2007, for further considerations on relatedness in the finite island model).

3.3.1 Directional selection

We find that the selection gradient on a trait a can be expressed as

$$s_a(\bar{\mathbf{z}}_t) = \left.\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} + (N-1)r_2^\circ(\bar{\mathbf{z}}_t)\left.\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}}, \tag{6}$$

where $\mathbf{z}_{-i}=\bar{\mathbf{z}}_t$ means that the derivative is evaluated when all neighbors express the mean phenotype $\bar{\mathbf{z}}_t$ ($\mathbf{z}_j=\bar{\mathbf{z}}_t$ for all $j\neq i$). The first derivative in eq. (6) captures the direct effect of trait a: the effect of a change in trait a in a focal individual on its own fitness. In a well-mixed population, this is all that matters for directional selection (i.e., $s_a(\bar{\mathbf{z}}_t)=\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)/\partial z_{i,a}$ when the population size is constant, Phillips and Arnold, 1989 [1]). The second derivative, which is weighted by pairwise relatedness $r_2^\circ(\bar{\mathbf{z}}_t)$, is the indirect effect of trait a: the effect focal fitness of a change in trait a in a neighbor of the focal (we arbitrarily chose this neighbor to be individual $j\neq i$). The selection gradient on trait a, eq. (6), is therefore the inclusive fitness effect of trait a (Hamilton, 1964, Rousset, 2004). Hence, in the absence of covariance among traits, the change in mean trait value is proportional to this trait's inclusive fitness effect (substituting eq. 6 into 3a with the off-diagonal elements of $\boldsymbol{G}_t$ all zeros). This finding is in line with much previous modeling work on the quantitative genetics of spatially- or family-structured populations (for e.g., Cheverud, 1985, Queller, 1992a,b, Frank, 1998, McGlothlin et al., 2014, Wakano and Lehmann, 2014).

3.3.2 Correlational and disruptive selection

We find that the correlational selection coefficient on two traits a and b (or the disruptive selection coefficient when $a=b$) can be expressed as the sum of two terms,

$$h_{ab}(\bar{\mathbf{z}}_t) = h_{\mathrm{w},ab}(\bar{\mathbf{z}}_t) + h_{\mathrm{r},ab}(\bar{\mathbf{z}}_t), \tag{7a}$$

where the first term,

$$\begin{aligned} h_{\mathrm{w},ab}(\bar{\mathbf{z}}_t) = &\left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,a}\partial z_{i,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} + (N-1)r_2^\circ(\bar{\mathbf{z}}_t)\left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}\partial z_{j,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} \\ &+ (N-1)r_2^\circ(\bar{\mathbf{z}}_t)\left(\left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,a}\partial z_{j,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} + \left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,b}\partial z_{j,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}}\right) \\ &+ (N-1)(N-2)r_3^\circ(\bar{\mathbf{z}}_t)\left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}\partial z_{k,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}}, \end{aligned} \tag{7b}$$

depends on the effects of joint changes in traits a and b within- (first line of eq. 7b) and between-individuals (second and third line of eq. 7b) on focal fitness. The first derivative on the first line of eq. (7b) is the effect of a joint change in traits a and b in a focal individual on its own fitness, which can be viewed as the *direct* synergistic effects of traits a and b (Figure 1a). In a well-mixed population, there are no other effects participating to correlational selection (i.e., $h_{ab}(\bar{\mathbf{z}}_t)=\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)/(\partial z_{i,a}\partial z_{i,b})$, Phillips and Arnold, 1989).

But when dispersal is limited (so that $r_2^\circ(\bar{\mathbf{z}}_t)>0$ and $r_3^\circ(\bar{\mathbf{z}}_t)>0$), three *indirect* synergistic effects become

[1] When the size of the population fluctuates, $s_a(\bar{\mathbf{z}}_t)=\partial\log w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)/\partial z_{i,a}$, due to normalization of focal fitness with respect to mean fitness (see section 3.1; and eq. A6 of Iwasa et al., 1991 for how this holds when selection is frequency-dependent). If the size of the population fluctuates but selection is frequency-independent, then the selection gradient can be expressed as the derivative of the log of mean fitness in the population with respect to the trait under scrutiny (e.g., eq. 7b of Lande, 1979)

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

relevant for correlational selection. These are the effect of a change in: (i) both traits in one neighbor of the focal (second derivative on the first line weighted by the neutral probability that the focal and this neighbor are identical-by-descent, $r_2^\circ(\bar{\mathbf{z}}_t)$, Figure 1b); (ii) in one trait in the focal and in the other in a neighbor (the two derivatives of the second line weighted by $r_2^\circ(\bar{\mathbf{z}}_t)$, Figure 1c); and (iii) in one trait in a neighbor and in the other in another neighbor indexed as k (last derivative weighted by the neutral probability that the focal and these two neighbors are identical-by-descent, $r_3^\circ(\bar{\mathbf{z}}_t)$, Figure 1d). Collectively, these terms capture the effects of non-random (due to limited dispersal) frequency-dependent interactions among individuals on correlational selection, revealing that under limited dispersal, selection favors the association of traits when these have positive effects between individuals.

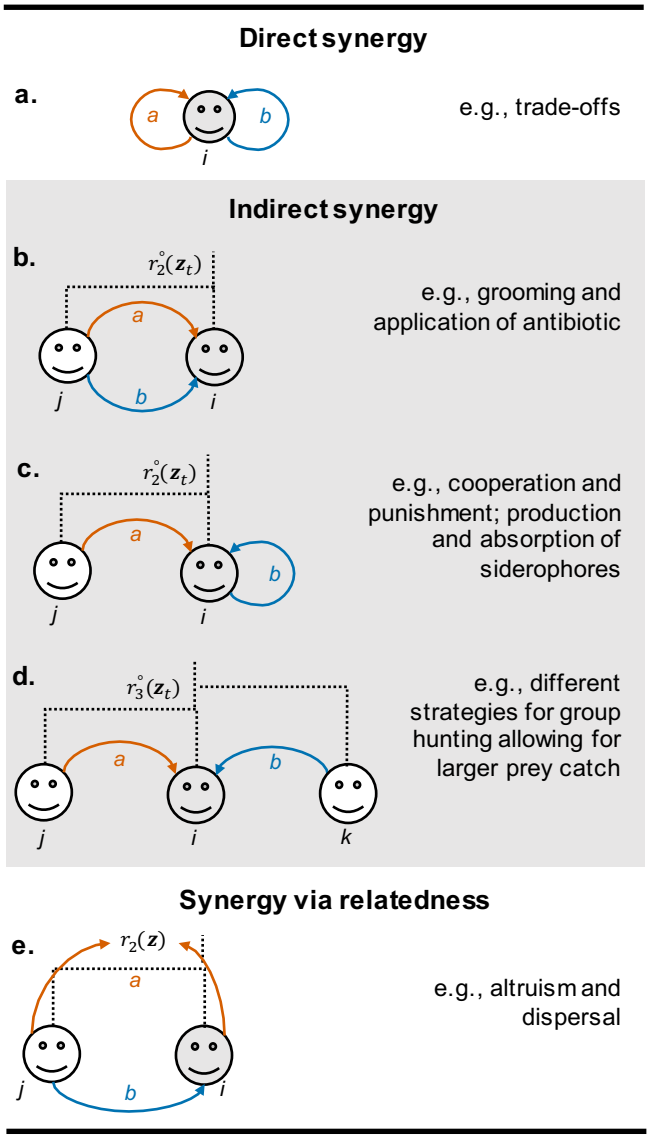

Figure 1: Within- and between-individual fitness effects relevant for correlational selection and examples of traits likely to be influenced by such effects. As revealed by eq. (7), there are five types of fitness effects due to perturbations in two traits a and b that are relevant for correlational selection when dispersal is limited: **a.** effect of a joint changes in a and b within the focal individual (first term of eq. 7b); **b.** effect of joint changes in a and b within neighbors of the focal (second term of eq. 7b, weighted by neutral pairwise relatedness, $r_2^\circ(\bar{\mathbf{z}}_t)$); **c.** effect of joint changes in a and b between the focal (here b) and its neighbors (here, a; second line of eq. 7b, weighted by $r_2^\circ(\bar{\mathbf{z}}_t)$); **d.** effect of joint changes in a and b between neighbors of the focal (third line of eq. 7b, weighted by neutral three-way relatedness, $r_3^\circ(\bar{\mathbf{z}}_t)$); **e.** the effect of the indirect effect of one trait (here b) multiplied to the effect of the other (here a) on pairwise relatedness, which reflects the tendency of relatives to receive the effects of b (eq. 7c).

The second term of eq. (7a), $h_{\mathrm{r},ab}(\bar{\mathbf{z}}_t)$, captures another type of synergistic effect relevant for correlational selection in group-structured populations. This term can be expressed as

$$h_{\mathrm{r},ab}(\bar{\mathbf{z}}_t) = (N-1)\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}}\bigg|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} \times \frac{\partial r_2(\mathbf{z})}{\partial z_b}\bigg|_{\mathbf{z}=\bar{\mathbf{z}}_t} + (N-1)\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,b}}\bigg|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} \times \frac{\partial r_2(\mathbf{z})}{\partial z_a}\bigg|_{\mathbf{z}=\bar{\mathbf{z}}_t}, \tag{7c}$$

where $\partial r_2(\mathbf{z})/\partial z_a$ is the effect of trait a on the probability that a neighbor of a focal individual with phenotype $\mathbf{z}$ carries a phenotype that is identical-by-descent to that of the focal (and $\partial r_2(\mathbf{z})/\partial z_b$ the effect of trait b). We refer to this as the effect of traits on relatedness. So eq. (7c) reveals that correlational selection depends on the product between the indirect effect of one trait ($\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)/\partial z_{j,a}$ and $\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)/\partial z_{j,b}$), with the effect of the other trait on relatedness. Such synergy via relatedness (Figure 1e) reflects that in group structured

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

populations, selection will favor an association among two traits when such an association results in indirect fitness benefits (e.g., trait a is cooperative, $\partial w(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)/\partial z_{j,a} > 0$) being preferentially directed towards relatives (e.g., trait b is the tendency to stay in natal group, $\partial r_2(\mathbf{z})/\partial z_b > 0$).

Group-structure and limited dispersal may thus lead to significant changes to the way selection molds phenotypic correlations, especially when traits have synergistic effects that are either indirect (Figure 1b-d) or via relatedness (Figure 1e). This will be illustrated later when we study the coevolution of two social traits in section 3.5. Before doing so, let us remark that when a single traits evolves ($n = 1$) and the selection gradient on this trait is zero ($s_a(\bar{\mathbf{z}}_t) = 0$), the change in phenotypic variance that we obtain (eq. 7 substituted into eq. 3b) reduces to previously derived expressions from quantitative genetics in the island model (eqs. 26 and 31 of Wakano and Lehmann, 2014). Further, eqs. (6)-(7) are consistent with evolutionary invasion analyses, i.e., with the first- and second-order effects of selection on the growth rate (or invasion fitness) of a rare mutant that arises in a monomorphic group-structured population and that differs from the resident in a single (eqs. 8 & 9 of Ajar, 2003) or multiple (eqs. 12 & 13 of Mullon et al., 2016) traits. We discuss further the correspondence between quantitative genetics, invasion analyses, and adaptive dynamics models in the next section, in which we study the equilibrium properties of the phenotypic distribution.

3.4 Equilibrium properties of the phenotypic distribution

Eq. (3) with eqs. (6)-(7) is a closed dynamical system that allows to track the evolution of the mean trait value and of the (co)variance between traits. In this section, we first investigate key features of the equilibrium of these phenotypic dynamics, and then discuss their connections with notions of evolutionary stability that come from invasion analyses and adaptive dynamics.

3.4.1 Equilibrium mean trait values

We denote the mean trait vector and variance-covariance matrix of the equilibrium phenotypic distribution by $\bar{\mathbf{z}}^*$ and $\boldsymbol{G}^*$, respectively. Such equilibrium simultaneously satisfies $\Delta\bar{\mathbf{z}}_t = \mathbf{0}$ and $\Delta\boldsymbol{G}_t = \mathbf{0}$ (where $\mathbf{0}$ is used to denote a n vector and $n \times n$ matrix whose entries are all zero, respectively). Rather than solving both systems of equations simultaneously, we can use the fact that in eq. (3a), the matrix $\boldsymbol{G}$ is a positive-definite matrix with real-entries (since it is a variance-covariance matrix). From standard linear algebra theory (Hines, 1980, Leimar, 2005, 2009), it then follows that the equilibrium for the phenotypic means must satisfy

$$\boldsymbol{s}(\bar{\mathbf{z}}^*) = \mathbf{0}, \tag{8}$$

i.e., all selection gradients (eq. 6) vanish at $\bar{\mathbf{z}}^*$, independently of the $\boldsymbol{G}$ matrix. An alternative argument to ignore the $\boldsymbol{G}$ matrix when determining the equilibrium trait vector $\bar{\mathbf{z}}^*$ can be made from our assumption that (co)variances are small (weak selection). As a consequence, the dynamics of the $\boldsymbol{G}$ matrix are slower than those of the mean vector $\bar{\mathbf{z}}_t$ (see eq. B-20 in Appendix B.2). Trait means should therefore reach their equilibrium before the variance-covariance $\boldsymbol{G}$ matrix stabilizes.

We can further ask whether a population with a mean vector that is close to an equilibrium $\bar{\mathbf{z}}^*$ will eventually converge to it as a result of selection and mutation. From the fact that $\boldsymbol{G}$ is positive-definite, it can be shown (see Leimar, 2009, for e.g.) that a necessary condition for a population to converge to $\bar{\mathbf{z}}^*$ for all possible $\boldsymbol{G}$ matrices is that the Jacobian matrix $\boldsymbol{J}(\bar{\mathbf{z}}^*)$ of the selection gradients with (a, b) entry

$$\boldsymbol{J}(\bar{\mathbf{z}}^*)_{ab} = \left.\frac{\partial s_a(\bar{\mathbf{z}})}{\partial z_b}\right|_{\bar{\mathbf{z}}=\bar{\mathbf{z}}^*} \tag{9}$$

is negative-definite at $\bar{\mathbf{z}}^*$, which means that the symmetric real part of $\boldsymbol{J}(\bar{\mathbf{z}}^*)$, $(\boldsymbol{J}(\bar{\mathbf{z}}^*) + \boldsymbol{J}(\bar{\mathbf{z}}^*)^{\mathrm{T}})/2$ has only negative eigenvalues. This type of equilibrium is referred to as (strongly) convergence stable (Leimar, 2005, 2009).

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

3.4.2 Equilibrium variance-covariance matrix

The dynamics of the variance-covariance matrix can then be studied at a convergence stable equilibrium $\bar{\mathbf{z}}^*$ for mean trait values (eq. 8). In this case, the equilibrium $\boldsymbol{G}^*$ for the variance-covariance matrix solves

$$\boldsymbol{M} + \boldsymbol{G}^* \boldsymbol{H}(\bar{\mathbf{z}}^*)\boldsymbol{G}^* = \mathbf{0}. \tag{10}$$

Eq. (10) has an admissible solution (i.e., such that $\boldsymbol{G}^*$ is positive-definite) if, and only if, the Hessian matrix, $\boldsymbol{H}(\bar{\mathbf{z}}^*)$, is negative-definite (Bhatia, 2015). This corresponds to the case under which selection is stabilizing at $\bar{\mathbf{z}}^*$. In fact, if $\boldsymbol{H}(\bar{\mathbf{z}}^*)$ is negative-definite, then the population will remain unimodally distributed around the mean vector $\bar{\mathbf{z}}^*$ and exhibit a variance-covariance matrix,

$$\boldsymbol{G}^* = \boldsymbol{M}\left[\boldsymbol{M}^{-1}\left(-\boldsymbol{H}(\bar{\mathbf{z}}^*)\right)^{-1}\right]^{1/2}, \tag{11}$$

where the operation $\mathbf{X}^{1/2}$ denotes the square root of $\mathbf{X}$ such that all the eigenvalues of $\mathbf{X}^{1/2}$ are positive (Bhatia, 2015; see also eq. 21c of Lande, 1980).

3.4.3 Connections with notions of stability from invasion analyses

Using a quantitative genetics approach, we have derived the conditions under which the multivariate phenotypic distribution of a dispersal limited population converges and remains at an equilibrium (eqs. 8-11). Here, we highlight the connections between these conditions and notions of evolutionary stability that have emerged from invasion analyses and adaptive dynamics under limited dispersal.

Singular strategy. First, the selection gradient eq. (6) substituted into condition (8) is equivalent to the definition of evolutionarily singular strategies/phenotypes under limited dispersal (i.e., phenotypes which when expressed by the whole population, the gradient of invasion fitness is zero, e.g., Rousset, 2004; see also Geritz et al., 1998, for general definition).

Convergence stability. Second, the condition for a mean trait vector to be an attractor of directional selection (condition 9 with eq. 6) is equivalent to the condition for a multi-trait phenotype to be convergence stable in invasion analysis (Mullon et al., 2016; see also Brown and Taylor, 2010, for a graphical approach to the coevolution of two traits in a genetically structured population; and Leimar, 2009, Geritz et al., 2016, for general considerations on multi-trait in invasion analysis). It is noteworthy that in spite of this equivalence, the phenotypic dynamics envisaged by a quantitative genetics model (given by eq. 3a, see also eq. 7 of Lande, 1979, or eq. 1 of Phillips and Arnold, 1989) differ from the dynamics inferred from invasion analysis (which are captured by the so-called "canonical equation", eq. 1 of Dieckmann and Law, 1996, or eq. 3 of Leimar, 2009). In a quantitative genetics model, the mean trait vector changes as a result of selection acting on a standing genetic variation, which is large enough to be captured by a statistical distribution ($\boldsymbol{G}_t$ in eq. 3a). Under the "canonical equation", traits evolve under a trait substitution sequence, whereby a selected mutant fixes before another mutant arises, so that the population "jumps" from one monomorphic state to another and in principle cannot sustain polymorphism (see Fig. 1c, upper right panel of Champagnat et al., 2006, for a useful depiction of a trait substitution sequence; see Van Cleve, 2015, for a review of trait substitution sequences with kin selection effects).

Uninvadability. Third, the condition that $\boldsymbol{H}(\bar{\mathbf{z}}^*)$ with eq. (7) is negative-definite for the population to remain unimodally distributed around $\bar{\mathbf{z}}^*$ is consistent with the condition derived from invasion analyses for $\bar{\mathbf{z}}^*$ to be locally uninvadable (i.e., that any rare mutant that arises in a population for monomorphic for $\bar{\mathbf{z}}^*$ and that causes a slight deviation from $\bar{\mathbf{z}}^*$ eventually vanishes, Mullon et al., 2016; see also Ajar, 2003 for a single evolving trait in dispersal limited population; and Leimar, 2009, Geritz et al., 2016, for general considerations on multi-trait analyses).

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Evolutionary branching. Invasion analyses have revealed that a phenotype that is convergence stable is not necessarily uninvadable (Eshel and Motro, 1981, Eshel, 1983, Taylor, 1989, Christiansen, 1991, Abrams et al., 1993b). In fact, when a singular phenotype is convergence stable but invadable, disruptive selection can lead to evolutionary branching, whereby two lineages stably coexist in polymorphism (Metz et al., 1996, Geritz et al., 1998). When multiple traits are evolving, a sufficient condition for the initiation of evolutionary branching is that the Jacobian is negative-definite and the Hessian matrix is positive-definite at the singular phenotype $\bar{\mathbf{z}}^*$ (note that this does not ensure that the resulting polymorphism is stable, Geritz et al., 2016, for further considerations). In the context of quantitative genetics, this means that the mean trait vector is held at $\bar{\mathbf{z}}^*$ (as $\boldsymbol{J}(\bar{\mathbf{z}}^*)$ is negative-definite) while the dynamics of the variance-covariance matrix (eq. 3b) diverges to infinity (as $\boldsymbol{H}(\bar{\mathbf{z}}^*)$ is positive-definite). In other words, at the onset of evolutionary branching, directional selection maintains the population mean vector at $\bar{\mathbf{z}}^*$ all the while disruptive selection favors extreme phenotypes, leading to the explosion of the variance-covariance matrix (in line with previous quantitative genetics approaches to study evolutionary branching, Wakano and Iwasa, 2013, Débarre et al., 2014, Wakano and Lehmann, 2014, Débarre and Otto, 2016).

3.4.4 The molding of phenotypic correlations by selection and mutation

Invasion analyses can be used to infer on the phenotypic correlations or associations generated by disruptive selection (by studying the eigenvector associated with the greatest eigenvalue of $\boldsymbol{H}(\bar{\mathbf{z}}^*)$, which gives the axis in phenotypic space along which selection is disruptive and along which the population becomes dimorphic, Mullon et al., 2016, Geritz et al., 2016). This approach, however, only incorporates the effect of selection and is limited to studying phenotypic correlations at the onset of evolutionary branching (inferring on the long term outcome of evolutionary branching requires studying invasion in dimorphic populations, which is typically much more involved mathematically, e.g., Geritz et al., 1998, Sasaki and Dieckmann, 2011). A quantitative genetics approach such as ours here allows two further considerations on phenotypic correlations (e.g., Lande, 1980, Jones et al., 2007). First, it allows to incorporate the influence of pleiotropy (through the distribution of mutational input, captured by the variance-covariance matrix $\boldsymbol{M}$ in eq. 3b). Second, eq. (11) allows to study equilibrium phenotypic correlations as a balance between mutation and stabilizing selection (and not only disruptive selection). We investigate this balance in more detail in the next section.

3.5 Application to the coevolution of two synergistic social traits

We now apply the quantitative genetics approach elaborated above to study the coevolution of two social traits under limited dispersal. Our main goal is to illustrate the potential significance of indirect synergistic effects for the molding of phenotypic correlations when dispersal is limited (Figure 1b-d).

3.5.1 Two public goods model

We model the coevolution of two nonnegative quantitative traits, labelled 1 and 2, that each capture participation to a different public good. For examples, in group living mammals, one trait could be the time/energy invested into foraging for the group's offspring, and the other, investment into defending the group by standing sentry against predators (e.g., Carter et al., 2014); in microorganisms, each trait could be the production of a specific amino-acid that is released into the external environment from which it can then be absorbed and used by group members (e.g., Özkaya et al., 2017).

Benefits and costs. We assume that both public goods are shared equally among group members, and that individuals receive extra benefits from obtaining both goods together. The benefits, $B(\mathbf{z}_i, \mathbf{z}_{-i})$, received by a focal individual (with traits $\mathbf{z}_i$ in a group composed of $\mathbf{z}_{-i}$) can then be written in terms of the group trait

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

averages ($\tilde{z}_1 = \sum_{j=1}^{N} z_{j,1}/N$ and $\tilde{z}_2 = \sum_{j=1}^{N} z_{j,2}/N$) as

$$B(\mathbf{z}_i, \mathbf{z}_{-i}) = b(\tilde{z}_1 + \tilde{z}_2) + b_{\mathrm{M}} \tilde{z}_1 \tilde{z}_2, \tag{12}$$

where the parameter $b > 0$ tunes the independent benefit of each public good produced (assumed to be the same for both goods for simplicity); and parameter $b_{\mathrm{M}} > 0$, the multiplicative benefits of receiving both goods together. Conversely, participation to both public goods simultaneously is assumed to be extra costly, for instance because the different goods call upon different biological functions that are costly to co-maintain, so that the cost $C(\mathbf{z}_i)$ paid by a focal individual (with traits $\mathbf{z}_i$) can be written as

$$C(\mathbf{z}_i) = \frac{c}{2}(z_{i,1}^2 + z_{i,2}^2) + c_{\mathrm{M}} z_{i,1} z_{i,2}, \tag{13}$$

where the parameter $c > 0$ tunes the independent cost of each trait, and parameter $c_{\mathrm{M}} > 0$, the multiplicative costs of the traits. The fecundity of a focal individual, $f(\mathbf{z}_i, \mathbf{z}_{-i})$, is then the difference between the benefits received and the costs paid,

$$f(\mathbf{z}_i, \mathbf{z}_{-i}) = 1 + B(\mathbf{z}_i, \mathbf{z}_{-i}) - C(\mathbf{z}_i), \tag{14}$$

where 1 is the baseline fecundity when no one in the group participates to either public good ($z_{i,1} = z_{i,2} = 0$ for all i).

These benefits (eq. 12) and costs (eq. 13) entail that it is best for a focal individual to express a negative *within*-individual association between traits (if expressed at all), and simultaneously be in a group in which traits are positively associated *between*-individuals. Such a configuration is possible when the population is well-mixed (so that there are no genetic correlations – or no relatedness – among individuals of the same group), but difficult when individuals of the same group are related due to limited dispersal. As relatedness increases, associations within- and between-individuals become aligned due to the co-inheritance of linked traits (in fact, the covariance between-traits between-individuals is equal to the product of pairwise relatedness with total covariance in the absence of selection; i.e., the between-individuals covariance of traits a and b is equal to $r_2^{\circ}(\bar{\mathbf{z}}_t)\sigma_{ab,t}$, see eq. C-25 in Appendix C). We therefore expect limited dispersal to be relevant to the coevolution of the two traits of our model and to the way selection associates these traits within individuals.

Fitness. Before proceeding to the analysis, let us give the individual fitness function of a focal individual $w(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)$. For this model, we assume that there is no group extinction, that offspring disperse independently from one another before local density regulation, and that all adults die after reproduction (so that the population follows a Wright-Fisher life cycle). In this case, individual fitness is,

$$w(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t) = \underbrace{\frac{(1-m) f(\mathbf{z}_i, \mathbf{z}_{-i})}{(1-m)\sum_{i=1}^{N} f(\mathbf{z}_i, \mathbf{z}_{-i})/N + m f(\bar{\mathbf{z}}_t, \bar{\mathbf{z}}_t)}}_{w^{\mathrm{P}}(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)} + \underbrace{\frac{m f(\mathbf{z}_i, \mathbf{z}_{-i})}{f(\bar{\mathbf{z}}_t, \bar{\mathbf{z}}_t)}}_{w^{\mathrm{D}}(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)}, \tag{15}$$

where $0 < m \leq 1$ is the dispersal probability. Individual fitness is the addition of two terms: (1) the expected number of offspring of the focal that successfully establish in their natal group, $w^{\mathrm{P}}(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)$, which is the ratio of the number of philopatric offspring of the focal to the total number of offspring that enter the competition in the focal group; and (2) the expected number of offspring of the focal that successfully settle in other groups, $w^{\mathrm{D}}(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)$, which is the ratio of offspring the focal sends in a non-focal group to the expected number of offspring in such a group (fitness function of the form eq. 15 is standard under limited dispersal, e.g., Rousset, 2004, Ohtsuki, 2010).

Relatedness. The final pieces of information that are necessary to apply our framework are the neutral relatedness coefficients, $r_2^{\circ}(\bar{\mathbf{z}}_t)$ and $r_3^{\circ}(\bar{\mathbf{z}}_t)$, and the effect of each trait on pairwise relatedness ($\partial r_2(\mathbf{z})/\partial z_a$). These expressions, which have been derived elsewhere for the Wright-Fisher life-cycle considered here (e.g., Rousset, 2004, Ajar, 2003, Ohtsuki, 2010, Wakano and Lehmann, 2014), are given in Appendix B.2.2 (eqs. B-21-B-22).

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

3.5.2 Analysis

We now proceed to analyze the evolution of both social traits using the approach established in section 3.2. We first focus on the equilibrium properties of the phenotypic distribution.

Convergence of mean trait values. Substituting eq. (15) and pairwise relatedness coefficient (eq. B-21) into eq. (6), we obtain that the selection gradient vector is

$$\boldsymbol{s}(\bar{\mathbf{z}}_t) = [1 - r_2^\circ(\bar{\mathbf{z}}_t)] \begin{pmatrix} b/N - c\bar{z}_{1,t} + \bar{z}_{2,t}(-c_\mathrm{M} + b_\mathrm{M}/N) \\ b/N - c\bar{z}_{2,t} + \bar{z}_{1,t}(-c_\mathrm{M} + b_\mathrm{M}/N) \end{pmatrix} + \mathcal{O}(\epsilon^2), \tag{16}$$

where ϵ is a small parameter capturing the magnitude of the effect of the public good on fecundity (i.e., ϵ is the largest of the parameters b, b_M, c, c_M). Solving eq. (16) for zero then yields the unique singular strategy

$$\bar{\mathbf{z}}^* = (\bar{z}_1^*, \bar{z}_2^*) = \left(\frac{b/N}{c + c_\mathrm{M} - b_\mathrm{M}/N}, \frac{b/N}{c + c_\mathrm{M} - b_\mathrm{M}/N}\right), \tag{17}$$

which unsurprisingly decreases with costs c and c_M, and increases with "direct" benefits b/N and b_M/N (as an individual recoups a share $1/N$ of its participation to each public good). Note that this singular strategy does not depend on dispersal (or relatedness). This is due our assumptions that group size is fixed and that generations are non-overlapping (in which case indirect fitness benefits of interacting with relatives are "cancelled" by the fitness costs of kin competition, e.g., Taylor, 1992, Rousset, 2004).

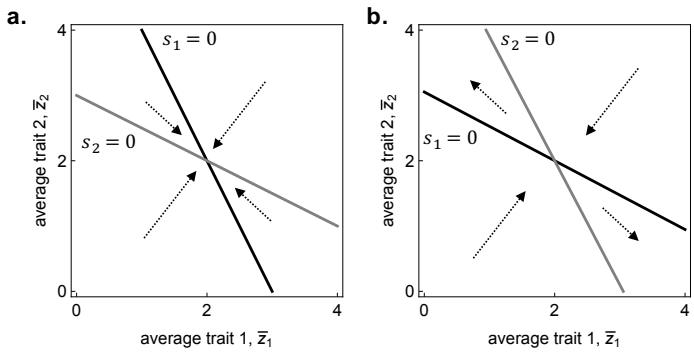

Figure 2: Directional selection on synergistic social traits. Qualitative dynamics of population means due to selection when equilibrium eq. (17) is: **a.** an attractor (with $b/N = 3$, $b_\mathrm{M}/N = 1.5$); **b.** a repeller (with $b/N = 5.8$, $b_\mathrm{M}/N = 0.1$). Solid lines show when the selection gradient eq. (16) on each trait vanishes (black for trait 1, $s_1(\bar{z}_1, \bar{z}_2) = 0$; grey for trait 2, $s_2(\bar{z}_1, \bar{z}_2) = 0$; other parameters: $c = 1$, $c_\mathrm{M} = 2$).

To establish whether the phenotypic distribution will converge to have mean $\bar{\mathbf{z}}^*$, we substitute eq. (16) into the symmetric part of the Jacobian matrix eq. (9), which we evaluate at the equilibrium eq. (17). It is straightforward to show that the two eigenvalues of the resulting matrix are given by

$$[1 - r_2^\circ(\bar{\mathbf{z}}_t)]\{-c - c_\mathrm{M} + b_\mathrm{M}/N, -c + c_\mathrm{M} - b_\mathrm{M}/N\} + \mathcal{O}(\epsilon^2). \tag{18}$$

Both are negative provided

$$-1 < \frac{-c_\mathrm{M} + b_\mathrm{M}/N}{c} < 1, \tag{19}$$

i.e., when the difference between the multiplicative costs, c_M, and direct multiplicative benefits, b_M/N, is small compared to the independent cost, c. In that case, the population will evolve to have mean given by eq. (17) and produce an equal amount of each public good (Figure 2a). Otherwise, the population will evolve to express a single trait and thus produce a single public good (depending on initial conditions, Figure 2b). Eqs. (17) and (19) reveal that limited dispersal does not influence the evolution of the mean of the phenotypic distribution. But what about the shape of the distribution around this mean?

Stabilisation of the distribution around the mean. Assuming eq. (19) holds true, whether or not the population distribution stabilises around the equilibrium trait values (eq. 17) depends on the Hessian matrix, $\boldsymbol{H}(\bar{\mathbf{z}}^*)$.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Let us start with analysing the diagonal elements of $\boldsymbol{H}(\bar{\mathbf{z}}^*)$, which reveal whether selection on each trait is independently stabilizing or disruptive. Substituting eq. (15) and relatedness coefficients (Appendix B.2.2) into eq. (7) for traits 1 and 2 (i.e., $a = b = 1$ and $a = b = 2$), and evaluating it at equilibrium eq. (17), we obtain that the diagonal entries of $\boldsymbol{H}(\bar{\mathbf{z}}^*)$ are

$$h_{11}(\bar{\mathbf{z}}^*) = h_{22}(\bar{\mathbf{z}}^*) = -[1 - r_2^\circ(\bar{\mathbf{z}}_t)]c + \mathcal{O}(\epsilon^2). \tag{20}$$

Since $0 \leq r_2^\circ(\bar{\mathbf{z}}_t) < 1$, the diagonal entries of $\boldsymbol{H}(\bar{\mathbf{z}}^*)$ are always negative, which means that selection on each trait is stabilizing when they evolve independently from one another.

Whether selection is stabilizing when both traits co-evolve also depends on the correlational coefficient of selection, $h_{12}(\bar{\mathbf{z}}^*)$. In particular, stabilizing selection requires that: (1) $h_{11}(\bar{\mathbf{z}}^*) < 0$ and $h_{22}(\bar{\mathbf{z}}^*) < 0$; and (2) $h_{12}(\bar{\mathbf{z}}^*)^2 < h_{11}(\bar{\mathbf{z}}^*)h_{22}(\bar{\mathbf{z}}^*)$, i.e., that the correlational selection coefficient is weak relative to the strength of stabilizing selection on both independent traits; this is because a 2×2 symmetric matrix Hessian matrix is negative-definite if and only if its diagonal entries are both negative and the off-diagonal satisfies condition (2) (e.g., Horn and Johnson, 2012). Condition (2) can equivalently be written as

$$-1 < \rho_s^* = \frac{h_{12}(\bar{\mathbf{z}}^*)}{\sqrt{h_{11}(\bar{\mathbf{z}}^*)h_{22}(\bar{\mathbf{z}}^*)}} < 1, \tag{21}$$

where ρ_s^* is the strength of correlational selection, relative to the strength of stabilizing selection on each independent trait at $\bar{\mathbf{z}}^*$. If eq. (21) does not hold, then selection is disruptive due to correlational selection.

The correlational coefficient of selection is derived by first substituting eq. (15) into eq. (7) with $a = 1$ and $b = 2$, and second, evaluating the result at equilibrium eq. (17). This yields

$$h_{12}(\bar{\mathbf{z}}^*) = [1 - r_2^\circ(\bar{\mathbf{z}}_t)]\Big[-c_M + \big(1/N + \alpha(N-1)/N\big)b_M/N\Big] + \mathcal{O}(\epsilon^2), \tag{22}$$

where

$$\alpha = \frac{(1-m)^2(3N-2-(N-2)m)}{3N-2+(N-1)(N-2)(1-(1-m)^3)} \tag{23}$$

decreases as dispersal and group size increases (i.e., α decreases as relatedness coefficients decrease, see Figure 3a). Eq. (22) reveals that as α (and relatedness) increases, the within-individual association favored by selection goes from negative to positive (Figure 3b-c). This is because as relatedness increases, indirect synergistic effects become increasingly targeted by correlational selection (Figure 1b-d).

Substituting eqs. (20) and (22) into eq. (21), we find that selection is stabilizing around $\bar{\mathbf{z}}^*$ when

$$-1 < \rho_s^* = \frac{-c_M + \big(1/N + \alpha(N-1)/N\big)b_M/N}{c} < 1, \tag{24}$$

which reveals that high relatedness, or large α, favors stabilizing selection (Figure 3b-c, dark grey and black regions), and conversely, low relatedness, or small α, favors disruptive selection and thus polymorphism (when eq. 19 holds but eq. 24 does not, Figure 3b-c, light grey region). This finding is in line with a recent computational eco-evolutionary model which found that when species can evolve cross-feeding interactions, mutualistic coexistence is compromised by spatial structure and limited dispersal (Oliveira et al., 2014). This is also in line with previous results on the evolution of single traits that have found that evolutionary branching is inhibited by limited dispersal (e.g., Day, 2001, Ajar, 2003, Wakano and Lehmann, 2014, Parvinen et al., 2017). In such models and ours, limited dispersal inhibits evolutionary branching because it creates genetic correlations among competing individuals, so that a mutant cannot be as different to common types as in well-mixed population. As a result, frequency-dependent disruptive selection is weaker under limited dispersal.

Effect of selection on phenotypic correlation. Putting our stability analyses together (especially eqs. 17, 19, 22, and 24) and validating them using individual-based simulations (see Appendix D for details), we find that there are three possible outcomes for the phenotypic distribution once it has converged to be unimodal around the equilibrium eq. (17) due to selection: (1) when relatedness is low, correlational selection is negative and strong enough to make selection disruptive, leading to the stable coexistence of individuals specialized in pro-

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

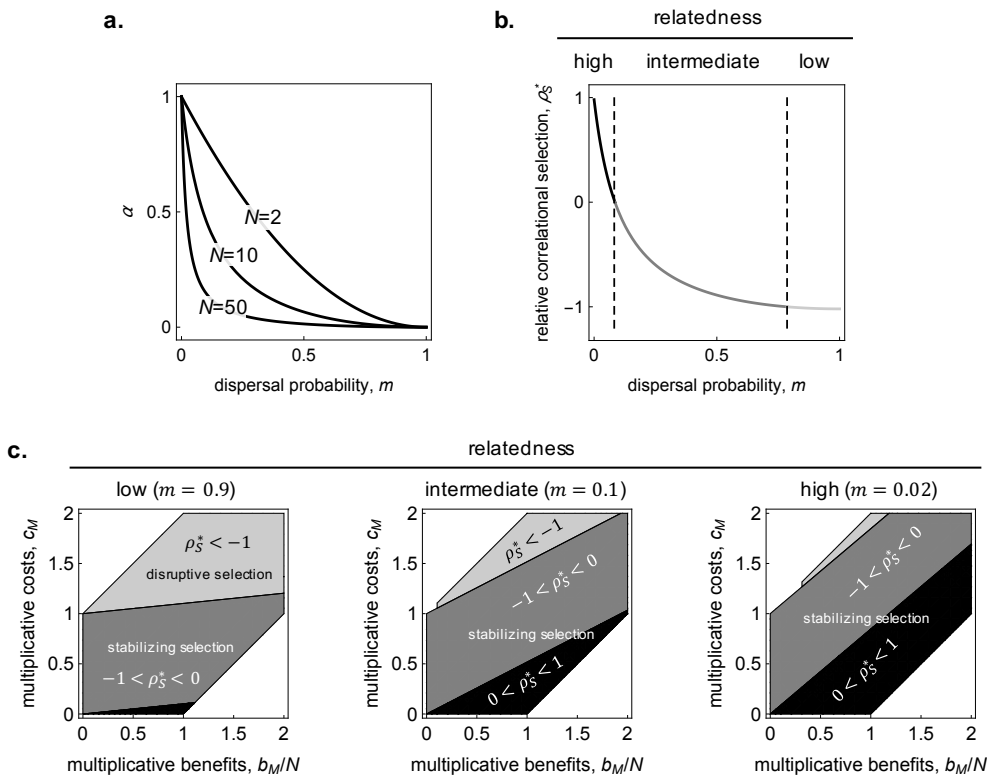

Figure 3: Correlational selection on synergistic social traits. a. Weight α, eq. (23), to multiplicative benefits in the coefficient of correlational selection (see eq. 22). **b.** Relative correlational selection, ρ_s^* (eq. 24), as a function of dispersal m, with critical levels of dispersal for which: $\rho_s^* < -1$ (light grey); $-1 < \rho_s^* < 0$ (dark grey); and $0 < \rho_s^* < 1$ (black, with $N = 10$, $b/N = 0.03$, $b_M/N = 1.8$, $c = 0.8$, $c_M = 1$). **c.** Parameter combinations (with $N = 10$, $c = 1$) for which correlational selection at the equilibrium eq. (17) is: (1) strongly negative (and causes selection to be disruptive, $\rho_s^* < -1$ & eq. 24 does not hold, light grey regions); (2) negative (and selection is stabilizing, $-1 < \rho_s^* < 0$ & eq. 24 holds, dark grey regions); and (3) positive (and selection is stabilizing, $0 < \rho_s^* < 1$ & eq. 24 holds, black regions). White regions correspond to parameter combinations under which the equilibrium is not evolutionary convergent (i.e., eq. 19 does not hold).

ducing a single public good (Figure 4a). In this instance, evolutionary dynamics follow so-called "Black queen" dynamics (Morris et al., 2012, Morris, 2015, with special reference to microorganisms): individuals first evolve to produce the same amount of leaky product that is shared among individuals, but the costly maintenance of both traits leads to specialization in a single product and the evolution of cross-feeding among types (see Rueffler et al., 2012, Vásárhelyi et al., 2015, for similar models on the evolution of specialization in well-mixed populations). (2) Over a critical level of relatedness, selection becomes stabilizing but correlational selection remains negative, which prevents evolutionary branching and thus specialization, but still results in a negative association among traits within individuals (Figure 4b). (3) Over another threshold of relatedness, correlational selection becomes positive, so that the traits become positively associated within individuals (Figure 4c). Hence, even though limited dispersal and relatedness have no bearing on the mean of the phenotypic distribution in our model (eqs. 17 and 19), indirect synergistic effects entail that relatedness has a significant influence on the shape of this distribution (which goes from being bimodal with a negative correlation under panmixia to unimodal with a positive correlation under limited dispersal, Figure 4).

Effect of pleiotropy on phenotypic correlation. So far, our analysis has focused on the effects of selection on the stability of jointly evolving traits (an analysis that could have equally been performed using invasion

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

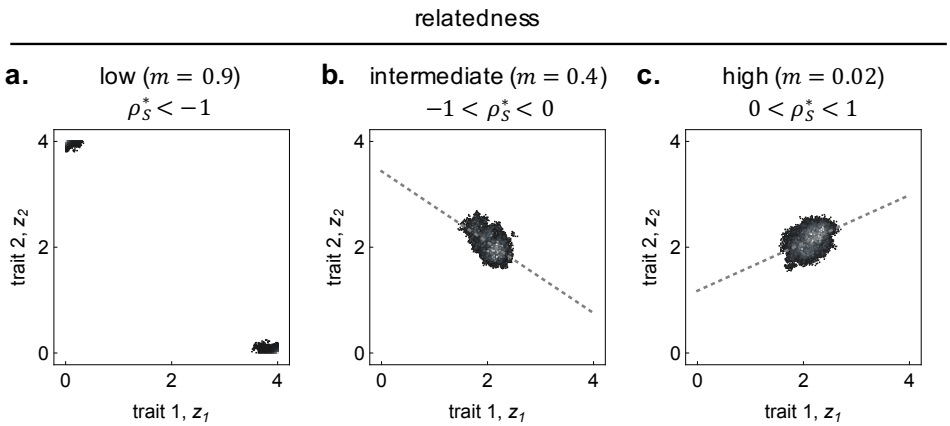

Figure 4: The effect of relatedness and indirect synergy on the phenotypic distribution. Equilibrium phenotypic density distribution, $p_t(\mathbf{z})$, of a simulated population, initially monomorphic for both traits at equilibrium $(2,2)$ (population composed of 1000 groups of size $N = 10$; sampled every 500 generations for 20'000 generations after 30'000 generations of evolution; other parameters: $b_M/N = 1.$, $\nu = 0.01$, $\sigma_{11}^m = \sigma_{22}^m = 0.02$, $\sigma_{12}^m = 0$; see Appendix D for details on simulations). **a.** Strong negative association with social polymorphism (with $b/N = 0.2$, $c = 0.1$, $c_M = 1$). **b.** Negative association (correlation = -0.67, $p < 10^{-10}$; with $b/N = 2.2$, $c = 1$, $c_M = 1.1$). **c.** Positive association (correlation = 0.45, $p < 10^{-10}$; with $b/N = 0.1$, $c = 1$, $c_M = .05$)

analysis; see Mullon et al., 2016, for such an approach to the joint evolution of multiple traits under limited dispersal). But selection is not the only relevant process for the way phenotypic distributions are shaped. As highlighted by the present quantitative genetic approach, the equilibrium variance-covariance matrix of the phenotypic distributions also depends on the patterns of mutation (captured by matrix $\boldsymbol{M}$ in eq. 11). In particular, pleiotropy is expected to influence the correlations among traits within individuals at an evolutionary equilibrium.

In order to investigate the joint effects of pleiotropy and correlational selection, let us assume that the variance of mutational effect on both traits is the same ($\sigma_{11}^m = \sigma_{22}^m = \sigma_m$), in which case the variance-covariance matrix of mutation effects can be written as

$$\boldsymbol{M} = \nu\sigma_m \begin{pmatrix} 1 & \rho_m \\ \rho_m & 1 \end{pmatrix}, \tag{25}$$

where $\rho_m = \sigma_{12}^m/\sigma_m$ is the correlation of the effect of mutations on traits 1 and 2. The parameter $-1 < \rho_m < 1$ thus captures the degree of pleiotropy between both traits (when it is zero, both traits change independently due to mutation, when it is positive, they tend to change in similar ways, and when it is negative, in opposite ways).

Substituting eqs. (20), (22) and (25) into eq. (11), we find that the correlation ρ_{12}^* among traits 1 and 2 at equilibrium is

$$-1 < \rho_{12}^* = \frac{\sigma_{12}^*}{\sqrt{\sigma_{11}^*\sigma_{22}^*}} = \frac{\rho_m + \rho_s^*}{1 + \rho_m\rho_s^* + \sqrt{(1-\rho_m^2)(1-\rho_s^{*2})}} < 1, \tag{26}$$

where ρ_s^* is given in eq. (21). This shows that at equilibrium, the sign of the correlation among between two traits reflects the balance, $\rho_m + \rho_s^*$, between the degree of pleiotropy, ρ_m, and the relative strength of correlational selection ρ_s^* (see Figure 5a; note that eq. 26 can be directly deduced from eq. 11 whenever the variance of mutational effect on both traits is the same, $\sigma_{11}^m = \sigma_{22}^m$, and the coefficients of disruptive selection on independent traits are equal, $h_{11}(\bar{\mathbf{z}}^*) = h_{22}(\bar{\mathbf{z}}^*)$, see eq. 8 of Jones et al., 2007). Since limited dispersal and relatedness has a significant influence on relative correlational selection ρ_s^* (eq. 21), it can affect the correlation ρ_{12}^* among traits in the population as much as pleiotropy, ρ_m.

We additionally checked that our model captured pleiotropy correctly by comparing the phenotypic correla-

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

tion among the two traits at equilibrium predicted by our model (eq. 26) and that observed in simulations for different levels of pleiotropy. We found that model predictions and observations from simulations also matched well in the presence of pleiotropy (Figure 5b).

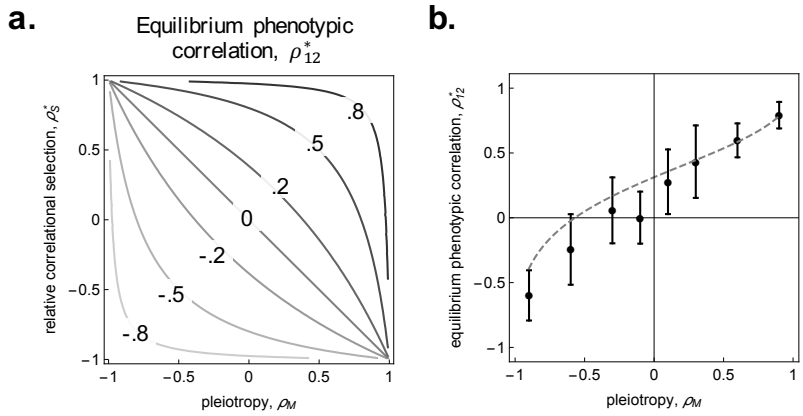

Figure 5: The effect of pleiotropy on phenotypic correlation. **a**. Contours of predicted phenotypic correlation among traits 1 and 2 at mutation-selection balance, ρ^*_{12}, according to pleiotropy, ρ_{m}, and the relative strength of correlational selection, ρ^*_{s} (from eq. 26). **b**. Predicted phenotypic correlation among traits 1 and 2 (dashed grey curve, from eq. 26), and corresponding observations from individual based simulations of a population initially monomorphic for $(2,2)$ divided among 1000 groups of size $N = 10$ (black dots, averaged correlation over 20'000 generations after 30'000 generations of evolution, error bars indicate standard deviation; other parameters: $m = 0.05$, $b/N = 0.2$, $b_{\mathrm{M}}/N = 1$, $c = 1$, $c_{\mathrm{M}} = 0.1$, $\nu = 0.01$, $\sigma^{\mathrm{m}}_{11} = \sigma^{\mathrm{m}}_{22} = 0.02$; see Appendix D for details).

Dynamics of the distribution. We further tested the accuracy of our dynamical model by comparing individual-based simulations with numerical recursions of eqs. (3). We found that simulated populations tend to have lower phenotypic variance than eqs. (3) would predict (Figure 6). This is probably due to global genetic drift, which our model ignores and which depletes phenotypic variance (as in well-mixed populations, e.g., Wakano and Iwasa, 2013, Débarre and Otto, 2016), and/or the presence of phenotypic skew, which is ignored under our assumption that the phenotypic distribution in the population is normal, but which can influence the dynamics of phenotypic variance (Appendix B, eq. B-18). Nonetheless, we observed overall a good qualitative fit between the predicted and observed dynamics of the phenotypic distribution (Figure 6). This suggests that the assumption of normality yields accurate predictions for the change of mean and variance when dispersal is limited (like in well-mixed populations, Turelli and Barton, 1994).

4 Discussion

In this paper, we have modeled the evolution of the distribution of genetically-determined quantitative traits under limited dispersal, frequency-dependent selection and pleiotropic mutation. By doing so, we have generalized two classical quantitative genetics results to include limited dispersal: first for the general recurrence eq. (1) of the phenotypic distribution under the continuum of alleles model (Kimura, 1965b, Fleming, 1979, Lande, 1979, Bürger, 1986); and second for the closed dynamical system eq. (3) of the vector of means and matrix of variance-covariance when the distribution is normal (Lande, 1979, Lande and Arnold, 1983, Phillips and Arnold, 1989). In both cases, genetic structure due to limited dispersal leads to the replacement of individual fitness in classical quantitative genetics equations by lineage fitness. This is the fitness of a typical carrier of a given phenotype (randomly sampled from the lineage of all members carrying that phenotype), i.e., the average direct fitness of a phenotype, which depends on the phenotypes expressed in the whole population and how they are distributed among groups (eq. 2).

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

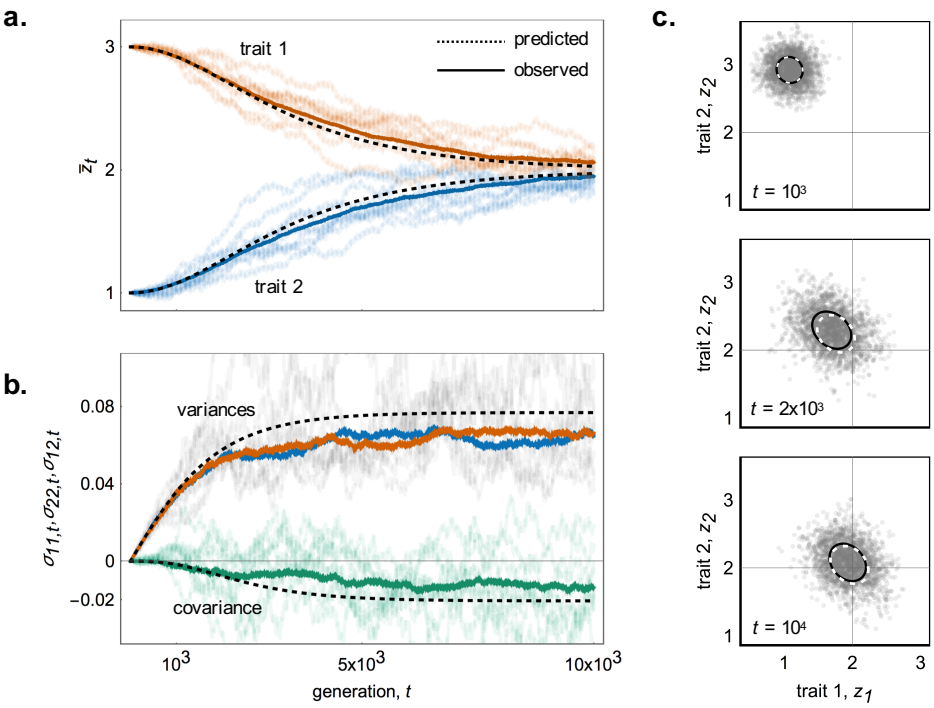

Figure 6: Observed and predicted evolution of the phenotypic distribution, $p_t(\mathbf{z})$. The observed (full lines, from individual based simulations) and predicted (dashed lines, from eq. 3) evolution of the traits' means (**a**. trait 1 in orange and 2 in blue), variances (**b**. trait 1 in orange and 2 in blue) and covariance (**b**. green) for 64 replicates (10 randomly chosen replicates in lighter shade, average over all 64 replicates in darker shade, initial population monomorphic with $z_1 = 3$ and $z_2 = 1$, distributed over 1000 groups of size $N = 10$, other parameters: $m = 0.4$, $b/N = 14.8$, $b_M/N = 0.1$, $c = 5$, $c_M = 2.5$, $\nu = 0.1$, $\sigma^m_{11} = \sigma^m_{22} = 0.02$, $\sigma^m_{12} = 0$; see Appendix D for details). **c**. Snapshot of the population (2'500 individuals randomly sampled across 64 replicates shown by grey points) and variance-covariance ellipses given by the (right) eigenvectors of the **G** matrix (observed across all 64 replicates in full lines and predicted in dashed), at generations: 1'000 (top panel); 2'000 (middle panel); and 10'000 (bottom panel).

From lineage fitness, we were able to reinforce existing links between concepts of evolutionary stability and evolutionary quantitative genetics: (1) the vector of means evolves to convergence stable phenotypic values (eqs. 8-9; see Charlesworth, 1990, Iwasa et al., 1991, Taper and Case, 1992, Abrams et al., 1993a, Abrams, 2001, Lion, 2018, for well-mixed populations, and Cheverud, 1985, Queller, 1992a,b, Frank, 1998, McGlothlin et al., 2014, Lehmann and Rousset, 2014, for limited dispersal); and (2) the distribution remains unimodal around such values when they are locally uninvadable or may become bimodal when they are invadable (eq. 10; see Sasaki and Dieckmann, 2011, Wakano and Iwasa, 2013, Débarre et al., 2014, Débarre and Otto, 2016, for well-mixed populations, and Lehmann and Rousset, 2014, Wakano and Lehmann, 2014, for the dynamics of the variance of a single trait around a singular strategy under limited dispersal). Specifically, we have shown that in a dispersal-limited population with infinitely many types, the selection gradient, which determines the change in mean trait values (eq. 3a), and the Hessian matrix, which shapes the variance-covariance matrix (eq. 3b), are respectively equal to the selection gradient vector and Hessian matrix computed from the invasion fitness of a rare mutant in an otherwise monomorphic population (i.e., eqs. 6-7 are equal to eqs. 12-13 of Mullon et al., 2016. Since the correspondence between the selection gradients of the two approaches is well-established for dispersal-limited populations (Lehmann and Rousset, 2014, for review), it may be felt that the correspondence between the Hessian matrix obtained from the evolutionary invasion analysis and that determining the change in the variance-covariance matrix is intuitive and must hold. Demonstrating this, however, required surprisingly lengthy calculations (see Appendix) showing that it is actually not obvious that when many alleles segregate in a dispersal-limited population and traits are far away from a convergence sta-

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ble point, selection on phenotypic (co)variances only depends on simple pairwise and three-way probabilities of identity-by-descent. Having established these relationships, we expect them to hold under more general demographic settings (e.g., with local demographic fluctuations) and hope that simpler arguments than our present ones can be found to prove it.

The extension of evolutionary invasion analyses to a quantitative genetics model allows to specify the phenotypic distribution at a mutation-selection balance (eq. 11). In particular, it allows to study the effects of selection and mutation on the phenotypic associations that emerge among traits at equilibrium (eq. 26). Our analyses of such associations suggest that kin selection due to limited dispersal can mold phenotypic associations as much as pleiotropic mutations (eq. 26 and Fig. 5). By expressing correlational selection on traits in terms of their direct and indirect fitness effects, we gained insights into the influence of kin selection on phenotypic associations (eq. 7). Motivated by our explicit formula for the variance-covariance matrix (eq. 11) and our example (section 3.5), we complement here the discussion found in Mullon et al. (2016) (based on an invasion analysis) on the implications of kin selection for the evolution of within-individual phenotypic associations. As indicated by the decomposition of correlational selection eq. (7a), there are two ways kin selection influences such associations.

The first is through the fitness effects that traits have when co-expressed among relatives, so when traits have indirect synergistic effects (eq. 7b, Figure 1b-d). Under limited dispersal, selection favors an association among two traits within individuals, when such an association between individuals has indirect fitness benefits. Due to such kin selection effects, different levels of dispersal can lead to significantly different evolutionary outcomes for phenotypic associations, as highlighted by our example on the coevolution of two traits whose association within-individual is costly but beneficial between-individuals due to social synergy. In this example, populations with little genetic structure evolved a division of social labor, with individuals specialized in only one trait coexisting with one another (Figure 4a), but populations with strong genetic structure evolved no such specialization, with traits in fact becoming positively associated within individuals (Figure 4c). In line with our results, populations of *E. coli* that experience frequent mixing (so show little genetic structure) readily evolve cross feeding interactions in poor environments, so that different strains specialize in the production of a specific amino acid (D'Souza and Kost, 2016). By contrast, in meerkat social groups (which typically show high levels of relatedness), individuals tend to participate to all social activities, with participation to different tasks such as babysitting and pup feeding positively associated within individuals (Clutton-Brock et al., 2003). Such patterns can be explained by our results if participation to different tasks in these systems is genetically determined, at least partially.

It is also worthy of note that in our example, relatedness has a substantial influence on the shape of the phenotypic distribution but none on the mean of this distribution (Figure 4, eqs. 17 and 18). Hence, the effects of genetic structure on phenotypic evolution that we report would have gone unnoticed from the study of the dynamics of the mean only (which is the focus of the vast majority of study of quantitative genetics in family-structured populations, e.g., Cheverud, 1985, Queller, 1992a,b, Frank, 1998, McGlothlin et al., 2014), or from the analysis of the selection gradient vector only (as done in the majority of evolutionary analyses to synergistic social traits, e.g., Gandon, 1999, Perrin and Mazalov, 2000, Reuter and Keller, 2001, Lehmann and Perrin, 2002, Rousset and Gandon, 2002, Gardner and West, 2004, Leturque and Rousset, 2004, Hochberg et al., 2008, Brown and Taylor, 2010, Kuijper and Johnstone, 2017). Overall, our example highlights that when traits have indirect synergistic effects between individuals (Figure 1b-d), relatedness is important for the way natural selection molds phenotypic associations within individuals. This consideration should be especially relevant to the evolution of specialization and the emergence of division of labor.

A relevant pair of traits likely to be influenced by such kin selection effects is costly helping and punishment, which have synergistic indirect benefits when expressed by different individuals (e.g., Raihani et al., 2012, and references therein). According to our results, kin selection should favor a positive association among helping and punishment, which interestingly, has been observed in humans (Fehr and Gächter, 2000). Another pair of traits whose evolution is likely to be influenced by their joint expression in different individuals is the production and exploitation of a public good, such as the secretion and use of siderophores by microorganisms (West et al., 2006). Under limited diffusion of siderophores and limited bacterial dispersal (Nadell et al., 2009,

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Kümmerli et al., 2014, Ross-Gillespie et al., 2015), we expect kin selection effects to be ecologically relevant for how secretion and use of siderophores are associated, and more generally for patterns of multi-trait social variation within microbial communities (Cordero and Polz, 2014, van Gestel et al., 2015, Özkaya et al., 2017, Schiessl et al., 2019, Rodríguez Amor and Dal Bello, 2019).

The second way kin selection influences phenotypic associations is via the combination of the indirect effect of one trait with the effect of the other on the tendency to interact with relatives ("synergy via relatedness", eq. 7c, Figure 1e). Specifically, selection favors an association among two traits when it results in fitness benefits being preferentially directed towards relatives or fitness costs towards non-relatives. For example, if trait a has positive indirect fitness effects (e.g., altruistic helping) and trait b decreases the tendency to interact with relatives (e.g., dispersal), then selection favors a negative correlation between traits a and b (e.g., Koella, 2000, Purcell et al., 2012, Mullon et al., 2018). We refer readers interested in this effect to Mullon et al. (2016), in which it is discussed at greater length, in particular in the context of dispersal syndromes (Edelaar and Bolnick, 2012, Ronce and Clobert, 2012).

More generally, our evolutionary perspective on phenotypic associations may be useful to empiricists who investigate correlational selection among traits in experimental or natural populations (e.g., Blows and Brooks, 2003, Blows, 2007, for reviews, and ch. 30 of Walsh and Lynch, 2018). Based on Lande and Arnold (1983) paper, the typical starting point of such studies is to perform a quadratic regression of individual fitness on the multiple traits expressed by this individual (for e.g., eq. 30.11 of Walsh and Lynch, 2018). The linear regression coefficients are collected in a vector usually denoted $\boldsymbol{\beta}$ with entry β_a interpreted as directional selection on trait a, and the quadratic coefficients in a matrix $\boldsymbol{\gamma}$ with entry γ_{ab} interpreted as correlational selection on traits a and b (in our notation, $\beta_a = \partial w(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)/\partial z_{i,a}$ and $\gamma_{ab} = \partial^2 w(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)/(\partial z_{i,a} \partial z_{i,b})$). This correspondence between selection on traits and regression coefficients on individual fitness, however, is only valid in well-mixed populations. Indeed, as our analysis has shown, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are respectively equal to the selection gradient $\boldsymbol{s}(\bar{\mathbf{z}}_t)$ and Hessian matrix $\boldsymbol{H}(\bar{\mathbf{z}}_t)$, only when all relatedness coefficients are zero (eq. 6 and 7).

For populations that are genetically structured, empirical estimates of selection on multiple traits require to: first regress individual fitness on the traits of the focal individual and on those of its social partners; and second, weigh these indirect fitness effects by relatedness coefficients (according to eqs. 6 and 7). Estimates of pairwise relatedness can be obtained from F_{ST} statistics at neutral sites such as microsatellite loci (i.e., the genetic covariance among pairs of interacting individuals relative to the average genetic covariance in the population, Charlesworth and Charlesworth, 2010, chapter 7.1). Similarly, three-way relatedness coefficients can be estimated from comparisons between the genetic skew at neutral sites among triplets of interacting individuals and the average genetic skew in the population. While the importance of indirect fitness effects and relatedness has long been emphasized for the directional selection gradient (so considering only linear regression coefficients, $\boldsymbol{\beta}$, e.g., Cheverud, 1985, Queller, 1992a,b, Frank, 1998, McGlothlin et al., 2014, see also ch. 5 of Walsh and Lynch, 2018), our analysis has further quantified the relationship between correlational selection and quadratic regression coefficients ($\boldsymbol{\gamma}$, eq. 7, Figure 1b-d), which is necessary to understand patterns of phenotypic variation within populations.

In practice, it is likely to be challenging to obtain reliable estimates of all the quadratic regression coefficients necessary to quantify the strength and direction of correlational selection (eq. 7). But our results can nevertheless be of use when designing experimental assays or interpreting collected data. For instance, our results show that for traits that underlie social or competitive behaviors, such as mating, aggression or cooperation, there is little reason to believe that a quadratic regression of an individual's fitness on its own traits provides a full picture of correlational selection. A corollary to this is that when there is mismatch between phenotypic correlations among two traits observed in a population on one hand, and the quadratic regression coefficient on individual fitness from experimental assays on the other (e.g., Bell and Sih, 2007, Adriaenssens and Johnsson, 2012, Han and Brooks, 2013, Akçay et al., 2015), this may indicate the presence of indirect synergistic fitness effects among traits and genetic structure in the population (rather than genetic constraints as typically inferred). One first accessible step towards testing this hypothesis would be to estimate genetic relatedness among interacting individuals. A high relatedness would suggest that phenotypic correlations are influenced by indirect synergistic fitness effects, which could then be estimated though quadratic fitness regressions of

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

fitness on partners' traits.

Our results further provide insight into the effects of limited dispersal on how selection influences the **G** matrix of additive genetic variances-covariances (Steppan et al., 2002, Arnold et al., 2008). Previous theoretical works have studied how linkage disequilibrium, pleiotropy and epistasis influence **G** under selection (Lande, 1980, 1984, Turelli, 1985, Turelli and Barton, 1990, Revell, 2007, Jones et al., 2014), but the effects of limited dispersal on **G** have either been assessed in the absence of selection (Lande, 1992), or when selection is frequency-independent (Jones et al., 2004, Guillaume and Whitlock, 2007, Guillaume, 2011, Björklund and Gustafsson, 2015). Here, we have shown that kin selection effects due to limited dispersal are relevant for the way selection favors phenotypic associations (i.e., for correlational selection, eq. 7), which in the long run can lead to genetic correlations through genetic integration (Sinervo and Svensson, 2002, Roff and Fairbairn, 2012). Of course, our model ignores many relevant features for quantitative genetics: environmental effects, genetic dominance, genetic linkage or sexual reproduction for examples. In particular, by assuming that individuals are haploid and reproduce clonally, our model does not allow to distinguish between different possible genetic architectures such as one pleiotropic locus that determines all evolving traits *versus* one independent locus for each trait. In the latter case, correlations among traits in dispersal-limited sexuals would depend on epistatic effects between loci within individuals and genetic linkage between loci (like in well mixed populations, Lande, 1984), as well as epistatic effects between loci located in different individuals, weighted by the genetic associations between these loci within groups (which would depend on dispersal, inbreeding and linkage; see Roze and Rousset, 2008 for how to compute such associations). Incorporating these features into our framework is likely to make the analysis of selection more complicated, but it would allow to study the genetic basis of variation, such as how genetic architecture and its evolution influence trait associations (e.g., Saltz et al., 2017).

One other significant limitation to our present approach is that it assumes that the phenotypic distribution is normal. This assumption is likely to be violated under frequency-dependent selection, which can lead to skewed and complicated distributions. In particular, the normal assumption precludes investigating what happens to the phenotypic distribution once evolutionary branching has occurred (like adaptive dynamics models based on the invasion analyses of monomorphic populations). To relax this assumption would entail tracking the dynamics of higher moments of the phenotypic distribution. One possible way to retain some mathematical tractability would be to use the oligomorphic approximation proposed by Sasaki and Dieckmann (2011). This approximation decomposes a multimodal trait distribution into a sum of unimodal distributions, each corresponding to a morph. Applying Sasaki and Dieckmann (2011)'s approach, which was developed for a large and well-mixed population, to a dispersal limited one, would be an interesting avenue of future research, as well as including class-structure (e.g., age- or sex-structure).

To conclude, we have derived a quantitative genetics model to study the gradual evolution of multiple traits that experience frequency-dependent selection and pleiotropic mutations when dispersal is limited. This model has revealed that limited dispersal opens previously unattended pathways for correlational selection, through the synergistic effects of traits: (1) between interacting individuals (Figure 1b-d), due to non-random frequency-dependent interactions; and (2) via relatedness (Figure 1e), owing to preferential interactions with relatives. This suggests that limited dispersal can profoundly influence how associations between social traits emerge in response to mutation and selection. Given the ubiquity of genetic structure in natural populations (e.g., Bohonak, 1999, Charlesworth and Charlesworth, 2010, p. 310), our results can help understand a wide range of patterns of intra-specific variation in competitive or social traits (such as behavioral syndromes, Dall et al., 2004, Dingemanse et al., 2012; social niche specialization, Bergmüller and Taborsky, 2007, Montiglio et al., 2013; or social division of labour, Boehm, 2002, Wright et al., 2014), which are increasingly thought to be ecologically significant (Bolnick et al., 2011, Wolf and Weissing, 2012, Sih et al., 2012, Canestrelli et al., 2016, Chaturvedi et al., 2017, Estrela et al., 2019). More broadly, by connecting different branches of theoretical evolutionary biology, from invasion analysis to adaptive dynamics to quantitative genetics, the present framework further bolsters the notion that whatever modeling approach is taken, natural selection cannot be divorced from kin selection when dispersal is limited (Hamilton, 1964, Frank, 1998, Rousset, 2004, van Baalen M, 2013, Lehmann et al., 2016).

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Literature cited

Abrams, P. A. 2001. Modelling the adaptive dynamics of traits involved in inter- and intraspecific interactions: An assessment of three methods. *Ecol. Lett.* 4:166–175.

Abrams, P. A., Y. Harada, and H. Matsuda. 1993a. On the relationship between quantitative genetic and ESS models. *Evolution.* 47:982–985.

Abrams, P. A., H. Matsuda, and Y. Harada. 1993b. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. *Evol. Ecol.* 7:465–487.

Adriaenssens, B. and J. I. Johnsson. 2012. Natural selection, plasticity and the emergence of a behavioural syndrome in the wild. *Ecol. Lett.* 16:47–55.

Ajar, É. 2003. Analysis of disruptive selection in subdivided populations. *BMC Evol. Biol.* 3:22.

Akçay, Ç., S. E. Campbell, and M. D. Beecher. 2015. The fitness consequences of honesty: under-signalers have a survival advantage in song sparrows. *Evolution.* 69:3186–3193.

Arnold, S. J., R. Bürger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the evolution and stability of the G-matrix. *Evolution.* 62:2451–2461.

Bell, A. M. and A. Sih. 2007. Exposure to predation generates personality in threespined sticklebacks (*Gasterosteus aculeatus*). *Ecol. Lett.* 10:828–834.

Bergmüller, R. and M. Taborsky. 2007. Adaptive behavioural syndromes due to strategic niche specialization. *BMC Ecol* 7:12–17.

Bhatia, R. 2015. Positive Definite Matrices. Princeton University Press, Princeton, NJ, USA.

Björklund, M. and L. Gustafsson. 2015. The stability of the G-matrix: The role of spatial heterogeneity. *Evolution.* 69:1953–1958.

Blows, M. W. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. *J. Evol. Biol.* 20:1–8.

Blows, M. W. and R. Brooks. 2003. Measuring nonlinear selection. *Am. Nat.* 162:815–820.

Boehm, C. 2002. Cooperative hunting roles among Tai chimpanzees. *Hum. Nat.* 13:27–46.

Bohonak, A. J. 1999. Dispersal, Gene Flow, and Population Structure. *Q Rev Biol* 74:21–45.

Bolnick, D. I., P. Amarasekare, M. S. Araújo, R. Bürger, J. M. Levine, M. Novak, V. H. W. Rudolf, S. J. Schreiber, M. C. Urban, and D. A. Vasseur. 2011. Why intraspecific trait variation matters in community ecology. *Trends Ecol. Evol.* 26:183–192.

Brodie, E. D., A. J. Moore, and F. J. Janzen. 1995. Visualizing and quantifying natural selection. *Trends Ecol. Evol.* 10:313–318.

Brown, S. P. and P. D. Taylor. 2010. Joint evolution of multiple social traits: a kin selection analysis. *Proc. R. Soc. B.* 277:415–422.

Bürger, R. 1986. On the maintenance of genetic variation: global analysis of kimura's continuum-of-alleles model. *J. Math. Biol.* 24:341–351.

Bürger, R. and I. M. Bomze. 1996. Stationary distributions under mutation-selection balance: structure and properties. *Adv. Appl. Probab.* 28:227–251.

Canestrelli, D., R. Bisconti, and C. Carere. 2016. Bolder takes all? The behavioral dimension of biogeography. *Trends Ecol. Evol.* 31:35–43.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Carter, A. J., S. English, and T. H. Clutton-Brock. 2014. Cooperative personalities and social niche specialization in female meerkats. *J. Evol. Biol.* 27:815–825.

Champagnat, N., R. Ferrière, and S. Méléard. 2006. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. *Theor. Popul. Biol.* 69:297–321.

Champagnat, N. and A. Lambert. 2007. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *Ann. Appl. Probab.* 17:102–155.

Charlesworth, B. 1990. Optimization models, quantitative genetics, and mutation. *Evolution.* 44:520–538.

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

Charlesworth, B. and D. Charlesworth. 2010. Elements of Evolutionary Genetics. Roberts and Company Publishers.

Chaturvedi, A., D. Croll, M. C. Fischer, F. Guillaume, S. Karrenberg, B. Kerr, G. Rolshausen, and J. Stapley. 2017. Can evolution supply what ecology demands? *Trends Ecol. Evol.* 32:187–197.

Chesson, P. L. 1981. Models for spatially distributed populations: The effect of within-patch variability. *Theor. Popul. Biol.* 19:288–325.

Chesson, P. L. 1984. Persistence of a markovian population in a patchy environment. *Z. Wahrscheinlichkeitstheor.* 66:97–107.

Cheverud, J. M. 1985. A quantitative genetic model of altruistic selection. *Behav. Ecol. Sociobiol.* 16:239–243.

Christiansen, F. B. 1991. On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* 138:37–50.

Clutton-Brock, T. H., A. F. Russell, and L. L. Sharpe. 2003. Meerkat helpers do not specialize in particular activities. *Animal Behav.* 66:531–540.

Connallon, T. 2015. The geography of sex-specific selection, local adaptation, and sexual dimorphism. *Evolution.* 69:2333–2344.

Cordero, O. X. and M. F. Polz. 2014. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Micro.* 12:263–273.

Daley, D. J. and D. Vere-Jones. 2003. An Introduction to the Theory of Point Processes, vol. II of *Probability and Its Applications.* Springer New York.

Dall, S. R. X., A. I. Houston, and J. M. McNamara. 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol. Lett.* 7:734–739.

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

Day, T. and S. R. Proulx. 2004. A general theory for the evolutionary dynamics of virulence. *Am. Nat.* 163:E40–E63.

Day, T. and P. D. Taylor. 1996. Evolutionarily stable versus fitness maximizing life histories under frequency-dependent selections. *Proc. R. Soc. Lond., B, Biol. Sci.* 263:333–338.

Débarre, F., S. L. Nuismer, and M. Doebeli. 2014. Multidimensional (co)evolutionary stability. *Am. Nat.* 184:158–171.

Débarre, F. and S. P. Otto. 2016. Evolutionary dynamics of a quantitative trait in a finite asexual population. *Theor. Popul. Biol.* 108:75–88.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Dieckmann, U. and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* 34:579–612.

Dingemanse, N. J., N. A. Dochtermann, and S. Nakagawa. 2012. Defining behavioural syndromes and the role of 'syndrome deviation' in understanding their evolution. *Behav. Ecol. Sociobiol.* 66:1543–1548.

D'Souza, G. D. and C. Kost. 2016. Experimental evolution of metabolic dependency in bacteria. *PLoS Genet.* 12:e1006364.

Edelaar, P. and D. I. Bolnick. 2012. Non-random gene flow: An underappreciated force in evolution and ecology. *Trends Ecol. Evol.* 27:659–665.

Eshel, I. 1983. Evolutionary and continuous stability. *J. Theor. Biol.* 103:99–111.

Eshel, I. and M. W. Feldman. 1984. Initial increase of new mutants and some continuity properties of ESS in two-locus systems. *Am. Nat.* 124:631.

Eshel, I. and U. Motro. 1981. Kin selection and strong evolutionary stability of mutual help. *Theor. Popul. Biol.* 19:420–433.

Eshel, I., U. Motro, and E. Sansone. 1997. Continuous stability and evolutionary convergence. *J. Theor. Biol.* 185:333–343.

Estrela, S., E. Libby, J. Van Cleve, F. Débarre, M. Deforet, W. R. Harcombe, J. Peña, S. P. Brown, and M. E. Hochberg. 2019. Environmentally mediated social dilemmas. *Trends Ecol. Evol.* 34:6–18.

Fehr, E. and S. Gächter. 2000. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* 90:980–994.

Fleming, W. H. 1979. Equilibrium distributions of continuous polygenic. *SIAM J. Appl. Math.* 36:148–168.

Frank, S. A. 1997. The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution.* 51:1712–1729.

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

Gandon, S. 1999. Kin competition, the cost of inbreeding and the evolution of dispersal. *J. Theor. Biol.* 200:345–364.

Gardner, A. and S. A. West. 2004. Cooperation and punishment, especially in humans. *Am. Nat.* 164:753–764.

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. *Evol. Ecol.* 12:35–57.

Geritz, S. A. H., J. A. J. Metz, and C. Rueffler. 2016. Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. *J. Math. Biol.* 72:1081–1099.

Gomulkiewicz, R. and M. Kirkpatrick. 1992. Quantitative genetics and the evolution of reaction norms. *Evolution.* 46:390.

Grafen, A. 1985. A geometric view of relatedness. In Dawkins, R. and M. Ridley (eds.), *Oxford Surv. Evol. Biol.*, pp. 28–90. Oxford University Press, Oxford.

Guillaume, F. 2011. Migration-induced phenotypic divergence: The migration-selection balance of correlated traits. *Evolution.* 65:1723–1738.

Guillaume, F. and M. C. Whitlock. 2007. Effects of migration on the genetic covariance matrix. *Evolution.* 61:2398–2409.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. i. *J. Theor. Biol.* 7:1–16.

Han, C. S. and R. C. Brooks. 2013. Correlational selection does not explain the evolution of a behavioural syndrome. *J. Evol. Biol.* 26:2260–2270.

Harris, T. E. 1963. The Theory of Branching Processes, vol. 6. Courier Corporation, Mineola.

Hines, W. G. S. 1980. Three characterizations of population strategy stability. *J. Appl. Probab.* 17:333–340.

Hochberg, M. E., D. J. Rankin, and M. Taborsky. 2008. The coevolution of cooperation and dispersal in social groups and its implications for the emergence of multicellularity. *BMC Evol. Biol.* 8:238.

Horn, R. A. and C. R. Johnson. 2012. Matrix analysis. Cambridge university press.

Iwasa, Y. and A. Pomiankowski. 1995. Continual change in mate preferences. *Nature.* 377:420–422.

Iwasa, Y., A. Pomiankowski, and S. Nee. 1991. The evolution of costly mate preferences ii. the 'handicap' principle. *Evolution.* 45:1431.

Jones, A. G., S. J. Arnold, and R. Burger. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution.* 58:1639.

Jones, A. G., S. J. Arnold, and R. Bürger. 2007. The mutation matrix and the evolution of evolvability. *Evolution.* 61:727–745.

Jones, A. G., R. Bürger, and S. J. Arnold. 2014. Epistasis and natural selection shape the mutational architecture of complex traits. *Nat. Commun.* 5:3709.

Karlin, S. 1968. Equilibrium behavior of population genetic models with non-random mating: Part ii: Pedigrees, homozygosity and stochastic models. *J. Appl. Probab.* 5:487–566.

Kimura, M. 1965a. Attainment of quasi-linkage equilibrium when gene frequencies are changing by natural selection. *Genetics.* 52:875–890.

Kimura, M. 1965b. A stochastic model concerning the maintenance of genetic variability in quantitative character. *Proc. Natl. Acad. Sci. U.S.A.* 54:731–736.

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

Kisdi, E. 2016. Dispersal polymorphism in stable habitats. *J. Theor. Biol.* 392:69–82.

Kisdi, E. and S. A. H. Geritz. 2009. Adaptive dynamics: a framework to model evolution in the ecological theatre. *J. Math. Biol.* 61:165–169.

Koella, J. C. 2000. The spatial spread of altruism versus the evolutionary response of egoists. *Proc. R. Soc. Lond., B, Biol. Sci.* 267:1979–1985.

Kuijper, B. and R. A. Johnstone. 2017. How sex-biased dispersal affects sexual conflict over care. *Am. Nat.* 189:501–514.

Kümmerli, R., K. T. Schiessl, T. Waldvogel, K. McNeill, and M. Ackermann. 2014. Habitat structure and the evolution of diffusible siderophores in bacteria. *Ecol. Lett.* 17:1536–1544.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution.* 30:314–334.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution.* 33:402–416.

Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. *Genetics.* 94:203–215.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. U.S.A.* 78:3721–3725.

Lande, R. 1984. The genetic correlation between characters maintained by selection, linkage and inbreeding. *Genet. Res.* 44:309–320.

Lande, R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution.* 46:381–389.

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

Latter, B. D. 1970. Selection in finite populations with multiple alleles. ii. centripetal selection, mutation, and isoallelic variation. *Genetics.* 66:165–86.

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

Lehmann, L. and N. Perrin. 2002. Altruism, dispersal and phenotype-matching kin recognition. *Am. Nat.* 159:451–468.

Lehmann, L. and F. Rousset. 2014. The genetical theory of social behaviour. *Philosophical Transactions of the Royal Society B* 369:1–18.

Leimar, O. 2005. The evolution of phenotypic polymorphism: Randomized strategies versus evolutionary branching. *Am. Nat.* 165:669–681.

Leimar, O. 2009. Multidimensional convergence stability. *Evol. Ecol. Res.* 11:191–208.

Leturque, H. and F. Rousset. 2004. Intersexual competition as an explanation for sex-ratio and dispersal biases in polygynous species. *Evolution.* 58:2398–2408.

Lion, S. 2018. Theoretical approaches in evolutionary ecology: Environmental feedback as a unifying perspective. *Am. Nat.* 191:21–44.

Lynch, M. and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer, Massachusetts.

McGlothlin, J. W., J. B. Wolf, E. D. Brodie, and A. J. Moore. 2014. Quantitative genetic versions of Hamilton's rule with empirical applications. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 369:20130358.

Metz, J. A. J. 2011. Thoughts on the geometry of meso-evolution: collecting mathematical elements for a post-modern synthesis. In Chalub, F. A. C. C. and J. Rodrigues (eds.), *Math. Darwin's Leg.*, pp. 193–231. Birkhäuser, Basel.

Metz, J. A. J., S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs, and J. S. van Heerwaarden. 1996. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In van Strien, S. J. and S. M. Verduyn Lunel (eds.), *Stoch. Spat. Struct. Dyn. Syst. Proc. R. Dutch Acad. Sci.*, no. September, pp. 183–231. North-Holland, Amsterdam.

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

Michod, R. E. 1979. Evolution of life histories in response to age-specific mortality factors. *Am. Nat.* 113:531–550.

Michod, R. E. 1982. The theory of kin selection. *Annu. Rev. Ecol. Evol. Syst* 13:23–55.

Montiglio, P. O., C. Ferrari, and D. Reale. 2013. Social niche specialization under constraints: personality, social interactions and environmental heterogeneity. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 368:20120343.

Morale, D., V. Capasso, and K. Oelschläger. 2005. An interacting particle system modelling aggregation behavior: from individuals to populations. *J. Math. Biol.* 50:49–66.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Morris, J. J. 2015. Black queen evolution: the role of leakiness in structuring microbial communities. *Trends Genet.* 31:475–482.

Morris, J. J., R. E. Lenski, and E. R. Zinser. 2012. The black queen hypothesis: evolution of dependencies through adaptive gene loss. *MBio* 3:379.

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

Mullon, C., L. Keller, and L. Lehmann. 2018. Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nat. Ecol. Evol.* 2:132–140.

Nadell, C. D., J. B. Xavier, and K. R. Foster. 2009. The sociobiology of biofilms. *FEMS Microbiol Rev* 33:206–224.

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

Nagylaki, T. 1993. The evolution of multilocus systems under weak selection. *Genetics.* 134:627–647.

Nuismer, S. L., R. Gomulkiewicz, and B. J. Ridenhour. 2010. When is correlation coevolution? *Am. Nat.* 175:525–537.

Oechssler, J. and F. Riedel. 2001. Evolutionary dynamics on infinite strategy spaces. *Econ. Theory* 17:141–162.

Ohtsuki, H. 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. *Evolution.* 64:3344–3353.

Oliveira, N. M., R. Niehus, and K. R. Foster. 2014. Evolutionary limits to cooperation in microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* 111:17941–17946.

Özkaya, Ö., K. B. Xavier, F. Dionisio, and R. Balbontín. 2017. Maintenance of microbial cooperation mediated by public goods in single- and multiple-trait scenarios. *J. Bacteriol.* 199:e00297–17.

Parker, G. and J. 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. *J. Theor. Biol.* 416:129–143.

Parvinen, K., H. Ohtsuki, and J. Y. Wakano. 2018. Spatial heterogeneity and evolution of fecundity-affecting traits. *J. Theor. Biol.* 454:190–204.

Perrin, N. and V. Mazalov. 2000. Local competition, inbreeding, and the evolution of sex-biased dispersal. *Am. Nat.* 155:116–127.

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

Price, G. R. 1970. Selection and covariance. *Nature.* 227:520–521.

Purcell, J., A. Brelsford, and L. Avilés. 2012. Co-evolution between sociality and dispersal: the role of synergistic cooperative benefits. *J. Theor. Biol.* 312:44–54.

Queller, D. C. 1992a. A general-model for kin selection. *Evolution.* 46:376–380.

Queller, D. C. 1992b. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* 139:540.

Raihani, N. J., A. Thornton, and R. Bshary. 2012. Punishment and cooperation in nature. *Trends Ecol. Evol.* 27:288–295.

Reuter, M. and L. Keller. 2001. Sex ratio conflict and worker production in eusocial Hymenoptera. *Am. Nat.* 158:166–177.

Revell, L. J. 2007. The g matrix under fluctuating correlational mutation and selection. *Evolution.* 61:1857–1872.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Rodríguez Amor, D. and M. Dal Bello. 2019. Bottom-up approaches to synthetic cooperation in microbial communities. *Life* 9:22–17.

Roff, D. 1997. Evolutionary Quantitative Genetics. Evolutionary biology: Genetics. Springer US.

Roff, D. A. and D. J. Fairbairn. 2012. A test of the hypothesis that correlational selection generates genetic correlations. *Evolution*. 66:2953–2960.

Ronce, O. and J. Clobert. 2012. Dispersal syndromes. In *Dispersal Ecol. Evol.*, pp. 119–138. Oxford University Press.

Ross-Gillespie, A., Z. Dumas, and R. Kümmerli. 2015. Evolutionary dynamics of interlinked public goods traits: an experimental study of siderophore production in *Pseudomonas aeruginosa*. *J. Evol. Biol.* 28:29–39.

Rousset, F. 2002. Inbreeding and relatedness coefficients: what do they measure? *Heredity*. 88:371–380.

Rousset, F. 2004. Genetic structure and selection in subdivided populations, vol. 40. Princeton University Press, Princeton, NJ.

Rousset, F. and S. Gandon. 2002. Evolution of the distribution of dispersal distance under distance-dependent cost of dispersal. *J. Evol. Biol.* 15:515–523.

Roze, D. and F. Rousset. 2005. Inbreeding depression and the evolution of dispersal rates: a multilocus model. *Am. Nat.* 166:708–721.

Roze, D. and F. Rousset. 2008. Multilocus models in the infinite island model of population structure. *Theor. Popul. Biol.* 73:529–542.

Rueffler, C., J. Hermisson, and G. P. Wagner. 2012. Evolution of functional specialization and division of labor. *Proc Natl Acad Sci USA* 109:E326–E335.

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

Saltz, J. B., F. C. Hessel, and M. W. Kelly. 2017. Trait correlations in the genomics era. *Trends Ecol. Evol.* 32:1–12.

Sasaki, A. and U. Dieckmann. 2011. Oligomorphic dynamics for analyzing the quantitative genetics of adaptive speciation. *J. Math. Biol.* 63:601–635.

Schiessl, K. T., A. Ross-Gillespie, D. M. Cornforth, M. Weigert, C. Bigosch, S. P. Brown, M. Ackermann, and R. Kümmerli. 2019. Individual- versus group-optimality in the production of secreted bacterial compounds. *Evolution*. p. doi:10.1111/evo.13701.

Sih, A., J. Cote, M. Evans, S. Fogarty, and J. Pruitt. 2012. Ecological implications of behavioural syndromes. *Ecol. Lett.* 15:278–289.

Simon, B. 2008. A stochastic model of evolutionary dynamics with deterministic large-population asymptotics. *J. Theor. Biol.* 254:719–730.

Sinervo, B. and E. Svensson. 2002. Correlational selection and the evolution of genomic architecture. *Heredity*. 89:329–338.

Slatkin, M. 1980. Ecological character displacement. *Ecology* 61:163–177.

Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: Evolution of the G matrix. *Trends Ecol. Evol.* 17:320–327.

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

Taper, M. L. and T. J. Case. 1992. Models of character displacement adn the theoretical robustness of taxon cycles. *Evolution*. 46:317–333.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Taper, M. L. and T. J. Chase. 1985. Quantitative genetic models for the coevolution of character displacement. *Ecology* 66:355–371.

Taylor, P. and T. Day. 1997. Evolutionary stability under the replicator and the gradient dynamics. *Evol. Ecol.* 11:579–590.

Taylor, P. D. 1989. Evolutionary stability in one-parameter models under weak selection. *Theor. Popul. Biol.* 36:125–143.

Taylor, P. D. 1992. Altruism in viscous populations - an inclusive fitness model. *Evol. Ecol.* 6:352–356.

Taylor, P. D., T. Day, and G. Wild. 2007. From inclusive fitness to fixation probability in homogeneous structured populations. *J. Theor. Biol.* 249:101–110.

Tazzyman, S. J. and Y. Iwasa. 2009. Sexual selection can increase the effect of random genetic drift-a quantitative genetic model of polymorphism in *Oophaga Pumilio*, the strawberry poison-dart frog. *Evolution.* 64:1719–1728.

Turelli, M. 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics.* 111:165–195.

Turelli, M. and N. H. Barton. 1990. Dynamics of polygenic characters under selection. *Theor. Popul. Biol.* 38:1–57.

Turelli, M. and N. H. Barton. 1994. Genetic and statistical-analyses of strong selection on polygenic traits - what, me normal. *Genetics.* 138:913–941.

van Baalen M. 2013. The unit of adaptation, the emergence of individuality, and the loss of evolutionary sovereignty. In Huneman, P. and F. Bouchard (eds.), *From Groups to Individ. Evol. Emerg. Individ.*, pp. 117–140. MIT Press.

Van Cleve, J. 2015. Social evolution and genetic interactions in the short and long term. *Theor. Popul. Biol.* 103:2–26.

van Gestel, J., H. Vlamakis, and R. Kolter. 2015. Division of labor in biofilms: the ecology of cell differentiation. *Microbiol. Spectr.* 3:1–24.

Vásárhelyi, Z., G. Meszena, and I. Scheuring. 2015. Evolution of heritable behavioural differences in a model of social division of labour. *PeerJ* 3:e977–22.

Wakano, J. Y. and Y. Iwasa. 2013. Evolutionary branching in a finite population: deterministic branching vs. stochastic branching. *Genetics.* 193:229–241.

Wakano, J. Y. and L. Lehmann. 2014. Evolutionary branching in deme-structured populations. *J. Theor. Biol.* 351:83–95.

Walsh, B. and M. Lynch. 2018. Evolution and Selection of Quantitative Traits. Oxford University Press.

Wenseleers, T. 2010. Social evolution theory: a review of methods and approaches. In Szekely, T., A. Moore, and J. Komdeur (eds.), *Soc. Behav. Genes, Ecol. Evol.*, pp. 132–158. Cambridge University Press, Cambridge.

West, S. A., A. S. Griffin, A. Gardner, and S. P. Diggle. 2006. Social evolution theory for microorganisms. *Nat. Rev. Micro.* 4:597–607.

Wild, G. 2011. Inclusive fitness from multitype branching processes. *Bull. Math. Biol.* 73:1028–1051.

Wolf, M. and F. J. Weissing. 2012. Animal personalities: consequences for ecology and evolution. *Trends Ecol. Evol.* 27:452–461.

Wolfram Research, I. 2016. Mathematica. Wolfram Research, Inc., Champaign, Illinois.

Wright, C. M., C. T. Holbrook, and J. N. Pruitt. 2014. Animal personality aligns task specialization and task proficiency in a spider society. *Proc. Natl. Acad. Sci. U.S.A.* 111:9533–9537.

Wright, S. 1931. Evolution in mendelian populations. *Genetics.* 16:290.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Appendix

Contents of appendix

A Phenotypic distribution dynamics **2**

A.1 Process construction . . . 2

A.1.1 Markov chain . . . 2

A.1.2 Markov chain in terms of counting measures . . . 2

A.1.3 State dynamics . . . 3

A.2 Recurrence for the phenotypic distribution . . . 3

A.2.1 Individual fitness . . . 4

A.2.2 Lineage fitness . . . 5

B The dynamics of trait means and variance-covariance **5**

B.1 Weak selection and mutation . . . 5

B.1.1 Weak selection . . . 5

B.1.2 Weak mutation . . . 6

B.1.3 Dynamics of the mean trait values . . . 7

B.1.4 Dynamics of the phenotypic variance-covariance . . . 7

B.2 Closure assumptions . . . 8

B.2.1 Normal closure . . . 9

B.2.2 Quasi-equilibrium . . . 9

C Second-order approximation of lineage fitness **10**

C.1 Marginal effects . . . 11

C.1.1 Average marginal effects . . . 12

C.1.2 Pairwise relatedness . . . 12

C.2 Interaction effects . . . 14

C.2.1 Interaction effects within individuals . . . 14

C.2.2 Interaction effects between neighbors . . . 15

C.2.3 Interaction effects between individuals between groups . . . 17

C.3 Putting it all together . . . 17

D Individual-based simulations **18**

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A Phenotypic distribution dynamics

In this appendix, we derive eq. (1) of the main text.

A.1 Process construction

We first lay the foundations of our analysis by describing how phenotypic evolution in our model population (see section 2) is represented mathematically.

A.1.1 Markov chain

The phenotypic state, or state for short, of a group at given time point is given by the set of phenotypic values of all individuals residing in that group, $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ (where $\mathbf{z}_i = \{z_{i,1}, ..., z_{i,n}\} \in \mathbb{R}^n$ is the phenotype of individual indexed $i \in \{1, \ldots, N\}$). The state of each group in the population changes stochastically from one time period to the next (i.e., after one iteration of the life cycle) due to selection, mutation and dispersal. We assume that these changes can be modeled as a discrete time Markov chain on a continuous state space (as traits are continuous; see Meyn and Tweedie, 2009, for general state spaces). Because groups affect one another through dispersal, the transition kernel of a group depends on the state of all the other groups. But since there is an infinite number of groups and there is no isolation-by-distance (i.e., all groups are equally connected to one another through dispersal), the infinite set of interacting Markov chains (one for each group) can be described as a single Markov chain (for a focal group), whose kernel is a function of the expected value of the process (see Chesson, 1981, 1984, for ecological models). In other words, we can focus on the stochastic dynamics of a focal group and ignore the stochasticity stemming from groups other than the focal one.

A.1.2 Markov chain in terms of counting measures

Note that to describe the state of a focal group, the order of elements in $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ does not matter (because there is no class structure in our population, we do not care which specific individual carries a given phenotype within a group). What matters is how many individuals carry which phenotypes. We can thus represent the state of the focal group by a function, a *counting measure*, that counts the number of individuals with phenotypes that belong to an arbitrary set. Specifically, the counting measure, denoted μ, takes a subset $E \subseteq \mathbb{R}^n$ and sends it to a non-negative integer by counting the number of individuals within the focal group with phenotypes that belong to E according to the following definition

$$\mu(E) = \sum_{i=1}^{N} \delta_{\mathbf{z}_i}(E), \tag{A-1}$$

where δ is the dirac measure,

$$\delta_{\mathbf{z}_i}(E) = \begin{cases} 1, & \text{if } \mathbf{z}_i \in E; \\ 0, & \text{otherwise,} \end{cases} \tag{A-2}$$

(p. 51 of Harris, 1963, p. 3 of Daley and Vere-Jones, 2003; see also pp. 228-229 of Bürger and Bomze, 1996 for further formal considerations on using counting measures to describe a population under the continuum-of-alleles model). Applied to a single phenotypic value $\mathbf{z} = \{z_1, ..., z_n\} \in \mathbb{R}^n$, where z_a is the value of trait a, $\mu(\mathbf{z})$ returns the number of individuals with phenotype $\mathbf{z}$, and applied to the whole set of possible phenotypic values, it returns group size, $\mu(\mathbb{R}^n) = N$.

Under definition eq. (A-1), each possible state that a group can assume is uniquely determined by a specific counting measure, i.e., for each unordered set of N vectors in $\mathbb{R}^n$, there exists a unique counting measure (p. 52 of Harris, 1963 and p. 7 of Daley and Vere-Jones, 2003). We can therefore study the dynamics of the state of the

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

focal group by studying the dynamics of its equivalent counting measure (Daley and Vere-Jones, 2003, p. 13-15). So, if S_t denotes the (random) counting measure of a focal group at time period t, we can study the Markov chain $\{S_t\}$ on the space of all finite counting measures, which we write as $\mathcal{S}$. This type of construction has so far primarily been used to study phenotypic evolution in populations that are well-mixed and when time is continuous (e.g., Bürger and Bomze, 1996, Oechssler and Riedel, 2001, Champagnat et al., 2006, Champagnat and Lambert, 2007; but see Morale et al., 2005, Simon, 2008, for populations in explicit space).

A.1.3 State dynamics

To describe the stochastic dynamics of the state of a focal group, we let

$$\phi_t(T) = \Pr[S_t \in T], \tag{A-3}$$

denote the probability that a focal group is in a state that belongs to a subset $T \subseteq \mathcal{S}$ at time period t (equivalent to eq. 3.3 of Harris, 1963, p. 55). Since there is an infinite number of groups, $\phi_t(T)$ also gives the distribution of group states in the whole population. The dynamics $\phi_t(T)$ are governed by the Markov kernel transition function,

$$P(T|\mu,\phi_t) = \Pr[S_{t+1} \in T|S_t = \mu,\phi_t], \tag{A-4}$$

which is the probability that a group will be in a state that belongs to a subset $T \subseteq \mathcal{S}$ at time period $t+1$, given that it was in state μ at time period t and that the population distribution of states is ϕ_t (this is a non-homogeneous Markov chain, eq. 6.1 of Harris, 1963, p. 60).

State dynamics, or the probability that the focal group is in a state that belongs to $T \subseteq \mathcal{S}$ at time period $t+1$, can then be written as

$$\phi_{t+1}(T) = \int_{\mathcal{S}} \int_{T} P(\mu'|\mu,\phi_t)\phi_t(\mu)\mathrm{d}\mu'\mathrm{d}\mu, \tag{A-5}$$

i.e., the sum of weighted probabilities of going from all states $\mu \in \mathcal{S}$ to states $\mu' \in T$. Because one iteration of the life cycle (from t to $t+1$) encompasses many events, like selection, mutation, and dispersal, the transition kernel for our model is difficult to characterize (studies like Champagnat et al., 2006, are capable of constructing explicit transition kernels by considering time steps small enough so that only one event can occur per step). To model the evolutionary process in a more practical way, we will focus on the dynamics of the distribution of phenotypes across the entire population rather than on the dynamics of the distribution of group states $\phi_t(T)$.

A.2 Recurrence for the phenotypic distribution

The distribution of phenotypes across the entire population at time t is given by the density function

$$p_t(\mathbf{z}) = \int_{\mathcal{S}} \frac{\mu(\mathbf{z})}{N}\phi_t(\mu)\mathrm{d}\mu, \tag{A-6}$$

where $\mu(\mathbf{z})/N$ is the frequency of individuals with phenotype $\mathbf{z}$ within a group in state μ (recall that all groups have the same size N). Using eq. (A-5), the phenotypic distribution at time period $t+1$ can be written as

$$p_{t+1}(\mathbf{z}) = \int_{\mathcal{S}} \frac{\mu(\mathbf{z})}{N}\phi_{t+1}(\mu)\mathrm{d}\mu = \frac{1}{N}\int_{\mathcal{S}}\int_{\mathcal{S}} \mu'(\mathbf{z})P(\mu'|\mu,\phi_t)\phi_t(\mu)\mathrm{d}\mu'\mathrm{d}\mu = \frac{1}{N}\int_{\mathcal{S}} \lambda_\mu(\mathbf{z},\phi_t)\phi_t(\mu)\mathrm{d}\mu, \tag{A-7}$$

where

$$\lambda_\mu(\mathbf{z},\phi_t) = \int_{\mathcal{S}} \mu'(\mathbf{z})P(\mu'|\mu,\phi_t)\mathrm{d}\mu' \tag{A-8}$$

is the expected number of individuals with phenotype $\mathbf{z}$ residing in a focal group at time $t+1$, given that this focal group was in state μ at time t (and the population state distribution was ϕ_t). We can decompose this expected number as

$$\lambda_\mu(\mathbf{z},\phi_t) = \lambda^{\mathrm{P}}_\mu(\mathbf{z},\phi_t) + \lambda^{\mathrm{I}}_\mu(\mathbf{z},\phi_t), \tag{A-9}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

where $\lambda^{\mathrm{P}}_{\mu}(\mathbf{z},\phi_t)$ is the expected number of philopatric individuals (i.e., surviving adults or offspring that have remained in their natal group) and $\lambda^{\mathrm{I}}_{\mu}(\mathbf{z},\phi_t)$ is the expected number of immigrant offspring (i.e., coming from other groups) with phenotype $\mathbf{z}$. We aim to express these expected numbers in terms of the fitness of individuals at time t.

A.2.1 Individual fitness

The number, $\lambda^{\mathrm{P}}_{\mu}(\mathbf{z},\phi_t)$, of philopatric individuals with phenotype $\mathbf{z}$ can be expressed in terms of fitness components of individuals at time t as

$$\lambda^{\mathrm{P}}_{\mu}(\mathbf{z},\phi_t) = \int_{\mathbb{R}^n} \mu(\mathbf{z}')\, w^{\mathrm{P}}_{\mu}(\mathbf{z}',\phi_t)\, u(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}', \tag{A-10}$$

where $w^{\mathrm{P}}_{\mu}(\mathbf{z}',\phi_t)$ is philopatric fitness: it is the expected number of offspring produced by a *single* individual (including itself if it survives) bearing $\mathbf{z}'$ a time t (given that it resides in a state μ group); and $u(\mathbf{z}',\mathbf{z})$ is the p.d.f. for the event that the offspring produced by an individual with phenotype $\mathbf{z}'$ has phenotype $\mathbf{z}$. Note that we assume that surviving adults and offspring mutate alike. While this is relevant to unicellular organisms, an application specific to multicellular organisms would require distinguishing between two components of philopatric fitness: adult survival and offspring production. This would only complicate eq. (1) but would not affect our other results presented in the main text (eq. 3 onwards) as we later assume that mutations are rare, so that the chances of mutating during an individual's lifetime would be unlikely (Appendix B.1.2).

Likewise, we can write the expected number of immigrant offspring as

$$\lambda^{\mathrm{I}}_{\mu}(\mathbf{z},\phi_t) = \int_{\mathbb{R}^n}\int_{\mathcal{S}} \mu'(\mathbf{z}')\, w^{\mathrm{D}}_{\mu,\mu'}(\mathbf{z}',\phi_t)\, u(\mathbf{z}',\mathbf{z})\phi_t(\mu')\mathrm{d}\mu'\mathrm{d}\mathbf{z}', \tag{A-11}$$

where $w^{\mathrm{D}}_{\mu,\mu'}(\mathbf{z}',\phi_t)$ is the expected number of successful emigrant offspring of a *single* individual with phenotype $\mathbf{z}'$, given that it resides in a group in state $\mu' \in \mathcal{S}$, and that the colonized group (i.e., the group the offspring lands in) was in state μ at time t.

Substituting eqs. (A-10) and (A-11) into eq. (A-9), which is in turn substituted into eq. (A-7), the phenotypic distribution at $t+1$ reads as

$$p_{t+1}(\mathbf{z}) = \frac{1}{N}\int_{\mathcal{S}}\int_{\mathbb{R}^n}\left(\mu(\mathbf{z}')\, w^{\mathrm{P}}_{\mu}(\mathbf{z}',\phi_t) + \int_{\mathcal{S}} \mu'(\mathbf{z}')\, w^{\mathrm{D}}_{\mu,\mu'}(\mathbf{z}',\phi_t)\phi_t(\mu')\mathrm{d}\mu'\right) u(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\phi_t(\mu)\mathrm{d}\mu. \tag{A-12}$$

By exchanging integral variables μ and μ' in the second summand within brackets, we obtain

$$p_{t+1}(\mathbf{z}) = \frac{1}{N}\int_{\mathcal{S}}\int_{\mathbb{R}^n} \mu(\mathbf{z}')\left(w^{\mathrm{P}}_{\mu}(\mathbf{z}',\phi_t) + w^{\mathrm{D}}_{\mu}(\mathbf{z}',\phi_t)\right) u(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\phi_t(\mu)\mathrm{d}\mu, \tag{A-13}$$

where

$$w^{\mathrm{D}}_{\mu}(\mathbf{z}',\phi_t) = \int_{\mathcal{S}} w^{\mathrm{D}}_{\mu',\mu}(\mathbf{z}',\phi_t)\phi_t(\mu')\mathrm{d}\mu' \tag{A-14}$$

is the expected number of successful dispersing offspring produced by an individual with phenotype $\mathbf{z}'$, given that this individual resides in a group in state μ at time t.

Individual fitness is then defined as

$$w_{\mu}(\mathbf{z}',\phi_t) = w^{\mathrm{P}}_{\mu}(\mathbf{z}',\phi_t) + w^{\mathrm{D}}_{\mu}(\mathbf{z}',\phi_t), \tag{A-15}$$

which gives the expected number of successful offspring produced by an individual with phenotype $\mathbf{z}'$, given that this individual resides in a group in state μ at time t (and the population state distribution was ϕ_t). In terms of this individual fitness function, the phenotypic distribution at time $t+1$ (eq. A-13) reads as

$$p_{t+1}(\mathbf{z}) = \frac{1}{N}\int_{\mathcal{S}}\int_{\mathbb{R}^n} \mu(\mathbf{z}')\, w_{\mu}(\mathbf{z}',\phi_t)\, u(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\phi_t(\mu)\mathrm{d}\mu. \tag{A-16}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A.2.2 Lineage fitness

To go from eq. (A-16) to eq. (1) of the main text, let us define

$$W(\mathbf{z}',\phi_t) = \int_{\mathcal{S}} w_\mu(\mathbf{z}',\phi_t) q(\mu|\mathbf{z}',\phi_t) \mathrm{d}\mu, \tag{A-17}$$

where

$$q(\mu|\mathbf{z}',\phi_t) = \frac{\mu(\mathbf{z}')}{N} \frac{\phi_t(\mu)}{p_t(\mathbf{z}')} \tag{A-18}$$

is the p.d.f. for the event that an individual resides in a group in state μ at time t given that this individual bears phenotype $\mathbf{z}'$. In other words, $q(\mu|\mathbf{z}',\phi_t)$ gives the probability that an individual, randomly sampled at time t from the collection of individuals with phenotype $\mathbf{z}'$ in the population (the "$\mathbf{z}'$-lineage"), resides in a group in a state μ. As such, $W(\mathbf{z}',\phi_t)$ (eq. A-17) is the expected fitness of a member of the $\mathbf{z}'$-lineage at time t (where expectation is taken over all possible groups this member can belong to) and a multi-allelic version of *lineage fitness* (Mullon et al., 2016, Lehmann et al., 2016).

Substituting eq. (A-17) into eq. (A-16), we obtain that the individual phenotypic density distribution is

$$p_{t+1}(\mathbf{z}) = \int_{\mathbb{R}^n} W(\mathbf{z}',\phi_t) u(\mathbf{z}',\mathbf{z}) p_t(\mathbf{z}') \mathrm{d}\mathbf{z}', \tag{A-19}$$

which combines the forces of mutation and selection on phenotypic change. To start disentangling these, note that when the probability of a mutation is independent from parental phenotype, the p.d.f. for the event that the offspring of an individual with phenotype $\mathbf{z}'$ has phenotype $\mathbf{z}$ can be expressed as

$$u(\mathbf{z}',\mathbf{z}) = (1-\nu)\delta(\mathbf{z}'-\mathbf{z}) + \nu v(\mathbf{z}',\mathbf{z}), \tag{A-20}$$

where ν is the probability that an offspring has a mutant phenotype (i.e., $1-u(\mathbf{z},\mathbf{z}) = \nu$ for all $\mathbf{z}$), $\delta(\mathbf{z}'-\mathbf{z})$ is the Dirac delta function, and $v(\mathbf{z}',\mathbf{z})$ is the conditional probability of mutating from $\mathbf{z}'$ to $\mathbf{z}$ given that a mutation has occurred. So the first term of eq. (A-20) captures the event of no mutation, in which case the offspring has the same phenotype than its parent, and the second term captures the event of a mutation. Substituting eq. (A-20) into eq. (A-19), we finally obtain eq. (1) in the main text, as required.

B The dynamics of trait means and variance-covariance

Here, we derive eqs. (3)-(7) of the main text, which govern the closed dynamics of trait means and variance-covariance. As mentioned in the main text, this derivation hinges upon several assumptions that we detail below.

B.1 Weak selection and mutation

B.1.1 Weak selection

We first assume that the phenotypic distribution, $p_t(\mathbf{z})$, is peaked around the population mean $\bar{\mathbf{z}}_t = \int_{\mathbb{R}^n} \mathbf{z} p_t(\mathbf{z}) \mathrm{d}\mathbf{z}$ (i.e., the phenotypic variance is small). We can thus approximate lineage fitness, $W(\mathbf{z},\phi_t)$, as a second-order Taylor expansion around $\bar{\mathbf{z}}_t$. We do so in Appendix C, in which we show that lineage fitness can be written as

$$W(\mathbf{z},\phi_t) = W(\bar{\mathbf{z}}_t,\phi_t) + \sum_{a=1}^{n} \xi_t(z_a) s_{a,t}(\bar{\mathbf{z}}_t) + \frac{1}{2}\sum_{a=1}^{n}\sum_{b=1}^{n} \xi_t(z_a)\xi_t(z_b) h_{ab,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t^3), \tag{B-1}$$

where

$$W(\bar{\mathbf{z}}_t,\phi_t) = 1 - \frac{1}{2}\sum_{a=1}^{n}\sum_{b=1}^{n} \sigma_{ab,t} h_{ab,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t^3) \tag{B-2}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

is the lineage fitness of the average phenotype $\bar{\mathbf{z}}_t$; $\xi_t(z_a) = z_a - \bar{z}_{a,t}$ denotes the difference between a value z_a and the average trait value a; $\sigma_{ab,t} = \int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_b)p_t(\mathbf{z})\mathrm{d}\mathbf{z}$ is the (co)variance among traits a and b in the population; $s_{a,t}(\bar{\mathbf{z}}_t)$ is the first-order effect of change in trait a away from $\bar{\mathbf{z}}_t$ on lineage fitness (i.e., $s_{a,t}(\bar{\mathbf{z}}_t) = \partial W(\mathbf{z},\phi_t)/\partial z_a|_{\mathbf{z}=\bar{\mathbf{z}}_t}$); $h_{ab,t}(\bar{\mathbf{z}}_t)$ is the second-order effect of a joint change in traits a and b away from $\bar{\mathbf{z}}_t$ on lineage fitness (i.e., $h_{ab,t}(\bar{\mathbf{z}}_t) = \partial^2 W(\mathbf{z},\phi_t)/\partial z_a \partial z_b|_{\mathbf{z}=\bar{\mathbf{z}}_t}$); and ξ_t is the maximum deviation between individual trait value in the population and the population mean trait value at time t. We detail the first- and second-order effects below.

The first-order effect is given by

$$s_{a,t}(\bar{\mathbf{z}}_t) = \left.\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} + (N-1)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\left.\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}}, \tag{B-3}$$

where individual fitness, $w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)$, is written as in the main text eq. (5) and $r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)$ is a neutral time-dependent coefficient of pairwise relatedness (i.e., the probability that two individuals sampled at random within a group at time t are identical-by-descent in the absence of selection, see section C.1.2 in Appendix C for more details).

The second-order effect is given by

$$h_{ab,t}(\bar{\mathbf{z}}_t) = h_{\mathrm{w},ab,t}(\bar{\mathbf{z}}_t) + h_{\mathrm{r},ab,t}(\bar{\mathbf{z}}_t), \tag{B-4a}$$

with

$$\begin{aligned} h_{\mathrm{w},ab,t}(\bar{\mathbf{z}}_t) = {} & \left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,a}\partial z_{i,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} + r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)(N-1)\left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}\partial z_{j,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} \\ & + r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)(N-1)\left(\left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,a}\partial z_{j,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} + \left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,b}\partial z_{j,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}}\right) \\ & + r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)(N-1)(N-2)\left.\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}\partial z_{k,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}}, \end{aligned} \tag{B-4b}$$

and

$$h_{\mathrm{r},ab,t}(\bar{\mathbf{z}}_t) = (N-1)\left.\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,b}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} \times \left.\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}\right|_{\mathbf{z}=\bar{\mathbf{z}}_t} + (N-1)\left.\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} \times \left.\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_b}\right|_{\mathbf{z}=\bar{\mathbf{z}}_t}, \tag{B-4c}$$

where $r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)$ is the neutral time-dependent three-way relatedness (i.e., the probability that three individuals sampled at random within a group at time t are identical-by-descent in the absence of selection, see eq. C-44 for details); and $\partial r_{2,t}(\mathbf{z})/\partial z_a$ is the marginal effect of a change in trait a on time-dependent pairwise relatedness (i.e., the effect of trait a on the probability that a neighbor of a focal individual with phenotype $\mathbf{z}$ carries a phenotype that is identical-by-descent to that of the focal at time t, see section C.1.2 in Appendix C for more details).

The first (eq. B-3) and second (eq. B-4) order effects are the same as the selection gradient (eq. 6) and correlational selection (eq. 7) of the main text, respectively, with the exception that relatedness coefficients ($r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)$, $r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)$, $\partial r_{2,t}(\mathbf{z})/\partial z_a$) are time-dependent in eqs. (B-3) and (B-4) and independent in eqs. (6) and (7). We will specify in section B.2.2 below how we can get rid of this time dependence, but first, we need to make a further assumption.

B.1.2 Weak mutation

Our next assumption is that mutations are rare, with the probability of mutating, ν, of the order $\mathcal{O}(\xi_t^2)$. Under this assumption, note that $\nu W(\mathbf{z},\phi_t) = \nu + \mathcal{O}(\xi_t^3)$ (from eqs. B-1-B-2). We can therefore rewrite eq. (1) of the

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

main text as

$$p_{t+1}(\mathbf{z}) = W(\mathbf{z},\phi_t)p_t(\mathbf{z}) + \nu\left(\int_{\mathbb{R}^n} p_t(\mathbf{z}')\upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}' - p_t(\mathbf{z})\right) + \mathcal{O}(\xi_t^3), \tag{B-5}$$

where the first term captures the effects of selection only, and the next term, the effects of mutation only. Eq. (B-5) takes the same functional form as classical recurrence for the phenotypic distribution in well-mixed populations when selection and mutation are weak (under the continuum-of-alleles model, e.g., eq. 1 of Bürger, 1986; for fluctuating population size, see eq. 4.9 of Champagnat et al., 2006), but with lineage, $W(\mathbf{z},\phi_t)$, instead of individual fitness. Next, we use eq. (B-5) to derive recurrence equations for the changes in mean trait values and the phenotypic variance-covariance matrix over one time period.

B.1.3 Dynamics of the mean trait values

By definition, the change in the mean of trait a over one time period is

$$\Delta\bar{z}_{a,t} = \bar{z}_{a,t+1} - \bar{z}_{a,t} = \int_{\mathbb{R}^n} \xi_t(z_a)p_{t+1}(\mathbf{z})\mathrm{d}\mathbf{z}. \tag{B-6}$$

Substituting eq. (B-5) into eq. (B-6), we obtain

$$\Delta\bar{z}_{a,t} = \int_{\mathbb{R}^n} \xi_t(z_a)W(\mathbf{z},\phi_t)p_t(\mathbf{z})\mathrm{d}\mathbf{z} + \nu\left(\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \xi_t(z_a)p_t(\mathbf{z}')\upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}\right) + \mathcal{O}(\xi_t^4), \tag{B-7}$$

But since the effects of mutation are assumed to be unbiased, we have

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \xi_t(z_a)p_t(\mathbf{z}')\upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} = 0. \tag{B-8}$$

Eq. (B-7) then reduces to

$$\Delta\bar{z}_{a,t} = \bar{z}_{a,t+1} - \bar{z}_{a,t} = \int_{\mathbb{R}^n} \xi_t(z_a)W(\mathbf{z},\phi_t)p_t(\mathbf{z})\mathrm{d}\mathbf{z} + \mathcal{O}(\xi_t^4), \tag{B-9}$$

which corresponds to the first term of the Price equation: the change in average trait value in a population is equal to the covariance between trait and fitness (Price, 1970; see eq. 3 of Frank, 1997).

Substituting eq. (B-1) into eq. (B-9), we obtain that the change in the mean of trait a is,

$$\Delta\bar{z}_{a,t} = \sum_{b=1}^{n}\sigma_{ab,t}s_{b,t}(\bar{\mathbf{z}}_t) + \frac{1}{2}\sum_{b=1}^{n}\sum_{c=1}^{n}\kappa_{abc,t}h_{bc,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t^4), \tag{B-10}$$

which depends on the skew,

$$\kappa_{abc,t} = \int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_b)\xi_t(z_c)p_t(\mathbf{z})\mathrm{d}\mathbf{z}, \tag{B-11}$$

in the population at time period t (in line with e.g., eq. 8a of Wakano and Iwasa, 2013 and eq. A20 b of Débarre and Otto, 2016 in well-mixed populations; eq. 17 of Wakano and Lehmann, 2014 for the island model).

B.1.4 Dynamics of the phenotypic variance-covariance

By definition, the change in the (co)variance (within individuals) between two traits a and b over one time period is

$$\begin{aligned}\Delta\sigma_{ab,t} = \sigma_{ab,t+1} - \sigma_{ab,t} &= \int_{\mathbb{R}^n} (z_a - \bar{z}_{a,t+1})(z_b - \bar{z}_{b,t+1})p_{t+1}(\mathbf{z})\mathrm{d}\mathbf{z} - \sigma_{ab,t} \\ &= \int_{\mathbb{R}^n} \left(\xi_t(z_a)\xi_t(z_b) - \sigma_{ab,t}\right)p_{t+1}(\mathbf{z})\mathrm{d}\mathbf{z} - \Delta\bar{z}_{a,t}\Delta\bar{z}_{b,t}.\end{aligned} \tag{B-12}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Substituting eq. (B-5) into the above, we obtain

$$\begin{aligned}\Delta\sigma_{ab,t} = &\int_{\mathbb{R}^n} \left(\xi_t(z_a)\xi_t(z_b) - \sigma_{ab,t}\right) W(\mathbf{z},\phi_t)\, p_t(\mathbf{z})\mathrm{d}\mathbf{z} \\ &+ \nu\left(\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_b)\, p_t(\mathbf{z}')\, \upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} - \sigma_{ab,t}\right) - \Delta\bar{z}_{a,t}\Delta\bar{z}_{b,t} + \mathcal{O}(\xi_t^5).\end{aligned} \tag{B-13}$$

The bracketed term in the second line of eq. (B-13), which captures the effects of mutations, can be simplified by first writing out the product of deviations in terms of parental phenotype as

$$\xi_t(z_a)\xi_t(z_b) = \xi_t(z_a')\xi_t(z_b') + (z_a - z_a')(z_b - z_b') + \xi_t(z_a)(z_b - z_b') + \xi_t(z_b)(z_a - z_a'), \tag{B-14}$$

and second, by noting that since mutations are assumed to be unbiased, the covariance between parental phenotype and mutation effect is zero:

$$\int_{\mathbb{R}^n} \left(\xi_t(z_a)(z_b - z_b') + \xi_t(z_b)(z_a - z_a')\right) \upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z} = 0. \tag{B-15}$$

Using eqs. (B-14)-(B-15), the effect of mutations in eq. (B-13) can then be written as

$$\begin{aligned}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_b)\, p_t(\mathbf{z}')\, \upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \left(\xi_t(z_a')\xi_t(z_b') + (z_a - z_a')(z_b - z_b')\right) p_t(\mathbf{z}')\, \upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} \\ &= \sigma_{ab,t} + \sigma_{ab}^{\mathrm{m}},\end{aligned} \tag{B-16}$$

where we have defined, $\sigma_{ab}^{\mathrm{m}} = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (z_a - z_a')(z_b - z_b')\, \upsilon(\mathbf{z}',\mathbf{z})\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z} \sim \mathcal{O}(\xi_t^2)$, as the (co)variance in mutational effects on traits a and b. Substituting eq. (B-16) into eq. (B-13), we obtain that the change in the (co)variance between two traits a and b over one time period is

$$\begin{aligned}\Delta\sigma_{ab,t} &= \nu\sigma_{ab}^{\mathrm{m}} + \int_{\mathbb{R}^n} \left(\xi_t(z_a)\xi_t(z_b) - \sigma_{ab,t}\right) W(\mathbf{z},\phi_t)\, p_t(\mathbf{z})\mathrm{d}\mathbf{z} - \Delta\bar{z}_{a,t}\Delta\bar{z}_{b,t} + \mathcal{O}(\xi_t^5) \\ &= \nu\sigma_{ab}^{\mathrm{m}} + \int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_b)\left(W(\mathbf{z},\phi_t) - 1\right) p_t(\mathbf{z})\mathrm{d}\mathbf{z} - \Delta\bar{z}_{a,t}\Delta\bar{z}_{b,t} + \mathcal{O}(\xi_t^5),\end{aligned} \tag{B-17}$$

where to go from the first to the second line, we have used the fact that mean lineage fitness is one: $\int_{\mathbb{R}^n} W(\mathbf{z},\phi_t)\, p_t(\mathbf{z})\mathrm{d}\mathbf{z} = 1$ (since the population size is constant).

Substituting eq. (B-1), and the change in mean, eq. (B-10), into eq. (B-17), we obtain after some re-arrangements that the one-generational change in phenotypic (co)variance between traits a and b is

$$\begin{aligned}\Delta\sigma_{ab,t} = &\nu\sigma_{ab}^{\mathrm{m}} + (W(\bar{\mathbf{z}},\phi_t) - 1)\sigma_{ab,t} + \sum_{c=1}^{n} \kappa_{abc,t} s_{c,t}(\bar{\mathbf{z}}_t) \\ &+ \frac{1}{2}\sum_{c=1}^{n}\sum_{d=1}^{n} \sigma_{abcd,t} h_{cd,t}(\bar{\mathbf{z}}_t) - \sum_{c=1}^{n}\sum_{d=1}^{n} \sigma_{ac,t}\sigma_{bd,t} s_{c,t}(\bar{\mathbf{z}}_t) s_{d,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t^5),\end{aligned} \tag{B-18}$$

which depends on the fourth central moment of the phenotypic distribution,

$$\sigma_{abcd,t} = \int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_b)\xi_t(z_c)\xi_t(z_d)\, p_t(\mathbf{z})\mathrm{d}\mathbf{z} \tag{B-19}$$

(in line with e.g., eq. 8b of Wakano and Iwasa, 2013 and eq. A24 b of Débarre and Otto, 2016 in well-mixed populations; eqs. B1-B8 of Wakano and Lehmann, 2014 for the island model with a single trait).

B.2 Closure assumptions

Finally, we close the dynamical system for the means and (co)variances (given by eqs. B-10 & B-18). We achieve this closure in two steps.

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

B.2.1 Normal closure

First, we assume that the phenotypic distribution, $p_t(\mathbf{z})$, is normal. Under this assumption, the skew in the phenotypic distribution is zero, $\kappa_{abc,t} = 0$, and the fourth central moments can be expressed in terms of the (co)variances, $\sigma_{abcd,t} = \sigma_{ab,t}\sigma_{cd,t} + \sigma_{ac,t}\sigma_{bd,t} + \sigma_{ad,t}\sigma_{bc,t}$. Substituting these relationships into eqs (B-10) and (B-18), we obtain that the one-generational changes in means and covariances are respectively given by

$$\begin{aligned}\Delta\bar{z}_{a,t} &= \sum_{b=1}^{n}\sigma_{ab,t}s_{b,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t^4)\\ \Delta\sigma_{ab,t} &= \nu\sigma_{ab}^{\mathrm{m}} - \sum_{c=1}^{n}\sum_{d=1}^{n}\sigma_{ac,t}\sigma_{bd,t}s_{c,t}(\bar{\mathbf{z}}_t)s_{d,t}(\bar{\mathbf{z}}_t) + \frac{1}{2}\sum_{c=1}^{n}\sum_{d=1}^{n}\left(\sigma_{ac,t}\sigma_{bd,t} + \sigma_{ad,t}\sigma_{bc,t}\right)h_{cd,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t^5).\end{aligned} \tag{B-20}$$

Since we make no assumption about the order of the fitness effects of traits (i.e., $s_{a,t}(\bar{\mathbf{z}}_t)$ and $h_{ab,t}(\bar{\mathbf{z}}_t)$ can be of order $\mathcal{O}(1)$), the magnitude of a one-generational change in mean trait $\bar{z}_{a,t}$ and (co)variance $\sigma_{ab,t}$ are respectively of order $\mathcal{O}(\xi_t^2)$ and $\mathcal{O}(\xi_t^4)$. In vector and matrix form, eq. (B-20) corresponds to eq. (3) of the main text, except that in eq. (B-20), the selection coefficients depend on time t (due to time-dependent relatedness coefficients, $r_{2,t}^{\circ}(\bar{\mathbf{z}}_t)$, $r_{3,t}^{\circ}(\bar{\mathbf{z}}_t)$, and $\partial r_{2,t}(\mathbf{z})/\partial z_a$). We get rid off of this dependency and finally achieve closure in the next section.

B.2.2 Quasi-equilibrium

Our second step to close the dynamical system eq. (B-20) is to assume that dispersal is strong enough (relative to selection) so that genetic associations between individuals within groups reach their steady-state values before any significant changes has occurred in the phenotypic distribution, $p_t(\mathbf{z})$, at the population level. This *quasi-equilibrium* assumption, which is frequently used in population genetic and social evolution theory (e.g., Kimura, 1965a, Nagylaki, 1993, Kirkpatrick et al., 2002, Roze and Rousset, 2005, 2008) is in line with our assumption that selection is weak. It entails that we can evaluate $r_{2,t}^{\circ}(\bar{\mathbf{z}}_t)$, $r_{3,t}^{\circ}(\bar{\mathbf{z}}_t)$, and $\partial r_{2,t}(\mathbf{z})/\partial z_a$ in eqs. (B-3)-(B-4) at their quasi-equilibrium, i.e., we take the limits $\lim_{\tau\to\infty} r_{2,\tau}^{\circ}(\bar{\mathbf{z}}_t) = r_2^{\circ}(\bar{\mathbf{z}}_t)$, $\lim_{\tau\to\infty} r_{3,\tau}^{\circ}(\bar{\mathbf{z}}_t) = r_3^{\circ}(\bar{\mathbf{z}}_t)$, and $\lim_{\tau\to\infty}\partial r_{2,\tau}(\mathbf{z})/\partial z_a|_{\mathbf{z}=\bar{\mathbf{z}}_t} = \partial r_2(\mathbf{z})/\partial z_a|_{\mathbf{z}=\bar{\mathbf{z}}_t}$, while holding $p_t(\mathbf{z})$ constant (we thus denote by $r_2^{\circ}(\bar{\mathbf{z}}_t)$,$r_3^{\circ}(\bar{\mathbf{z}}_t)$, and $\partial r_2(\mathbf{z})/\partial z_a|_{\mathbf{z}=\bar{\mathbf{z}}_t}$, the steady-state values of neutral pairwise relatedness, neutral three-way relatedness, and the first-order perturbation of pairwise relatedness, respectively). Substituting these steady-states into the selection coefficients eqs. (B-3)-(B-4) (now independent of time so written as $s_a(\bar{\mathbf{z}}_t)$ and $h_{ab}(\bar{\mathbf{z}}_t)$), which are in turn substituted into eq. (B-20), we finally obtain the closed dynamical eqs. (3) of the main text.

Computing relatedness coefficients. Computing relatedness coefficients under neutrality (i.e., $r_2^{\circ}(\bar{\mathbf{z}}_t)$, $r_3^{\circ}(\bar{\mathbf{z}}_t)$) is standard in population genetics (e.g., Karlin, 1968, Rousset, 2004). When generations are non-overlapping (i.e., a Wright-Fisher life cycle), for example, the relevant relatedness coefficients for our approach are given by

$$\begin{aligned}r_2^{\circ}(\bar{\mathbf{z}}_t) &= \frac{(1-m)^2}{N-(N-1)(1-m)^2}\\ r_3^{\circ}(\bar{\mathbf{z}}_t) &= \frac{(1-m)^3(1+3(N-1)r_2^{\circ}(\bar{\mathbf{z}}_t))}{N^2-(N-1)(N-2)(1-m)^3},\end{aligned} \tag{B-21}$$

(e.g., eqs. 12a & 12b of Ohtsuki, 2010; see also Table 1 of Mullon et al., 2016 for the Moran model). Calculating the first-order effect of selection on pairwise relatedness, $\partial r_2(\mathbf{z})/\partial z_a$, however, is more complicated. Under the quasi-equilibrium assumption, a perturbation of genetic associations between individuals will depend on first-order perturbations of individual fitness and neutral relatedness coefficients (see Roze and Rousset, 2008 for a general treatment, in particular their eq. 67). So far, the first-order effect of selection on pairwise relatedness, $\partial r_2(\mathbf{z})/\partial z_a$, has been explicitly derived for two standard life-cycles, the semelparous Wright-Fisher life-cycle (in which all adults die after reproduction; see eq. 18 of Ajar, 2003 and eq. 28 of Wakano and Lehmann, 2014) and the iteroparous birth-death Moran life-cycle (in which a single adult dies after reproduction in each

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

group; see eq. 14 of Mullon et al., 2016). In both cases, the effect of selection on relatedness can be written as

$$\left.\frac{\partial r_2(\mathbf{z})}{\partial z_a}\right|_{\mathbf{z}=\bar{\mathbf{z}}_t} = \kappa \frac{r_2^{\circ}(\bar{\mathbf{z}}_t)}{1-m}\left[\left[1+(N-1)r_2^{\circ}(\bar{\mathbf{z}}_t)\right]\left.\frac{\partial w^{\mathrm{P}}(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{i,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}} + \left[2r_2^{\circ}(\bar{\mathbf{z}}_t)+(N-2)r_3^{\circ}(\bar{\mathbf{z}}_t)\right](N-1)\left.\frac{\partial w^{\mathrm{P}}(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)}{\partial z_{j,a}}\right|_{\substack{\mathbf{z}_i=\bar{\mathbf{z}}_t\\ \mathbf{z}_{-i}=\bar{\mathbf{z}}_t}}\right], \tag{B-22}$$

where $w^{\mathrm{P}}(\mathbf{z}_i,\mathbf{z}_{-i},\bar{\mathbf{z}}_t)$ is the expected number of offspring of a focal individual (with phenotype $\mathbf{z}_i$, with neighbours $\mathbf{z}_{-i}$, and individuals in other groups with phenotype $\bar{\mathbf{z}}_t$) that successfully establish in their natal group; and $\kappa = 2$ for the Wright-fisher and $\kappa = 1$ for the Moran life cycle (watch out for an unfortunate typo in eq. 15 of Mullon et al., 2018, which has "$\kappa = N$" under the Moran life cycle).

C Second-order approximation of lineage fitness

Here, we derive the second-order Taylor expansion of lineage fitness, $W(\mathbf{z},\phi_t)$, around the population mean phenotype $\bar{\mathbf{z}}_t$ (i.e., we derive eqs. B-1 – B-4 of Appendix B). Let us first recall the definition of lineage fitness,

$$W(\mathbf{z},\phi_t) = \int_{\mathcal{S}} w_\mu(\mathbf{z},\phi_t)\, q(\mu|\mathbf{z},\phi_t)\,\mathrm{d}\mu, \tag{C-1}$$

where $w_\mu(\mathbf{z},\phi_t)$ is the fitness of an individual with phenotype $\mathbf{z}$ in a group in state μ. Our approach is to develop a second-order Taylor expansion of the individual fitness function $w_\mu(\mathbf{z},\phi_t)$, which we then plug back into eq. (C-1) to average it over group composition, $q(\mu|\mathbf{z},\phi_t)$, and thus obtain lineage fitness.

Our starting point is to rewrite individual fitness as

$$w_\mu(\mathbf{z},\phi_t) = w(\mathbf{z},\mathrm{n}_\mu(\mathbf{z}),\mathrm{e}(\phi_t)), \tag{C-2}$$

i.e., as a function that depends explicitly on all relevant phenotypes in the population: (1) $\mathbf{z}\in\mathbb{R}^n$, the phenotype of the focal individual (the individual whose fitness is under scrutiny); (2) $\mathrm{n}_\mu(\mathbf{z})\in\mathbb{R}^{(N-1)\times n}$, the set of phenotypes of the $N-1$ neighbors of the focal individual; and (3) $\mathrm{e}(\phi_t)\in\mathbb{R}^{N\times n}$, the set of N phenotypes from a representative (or average) group other than the one the focal resides in (which depends on the population state ϕ_t).

From these dependencies, the Taylor expansion of individual fitness, $w(\mathbf{z},\mathrm{n}_\mu(\mathbf{z}),\mathrm{e}(\phi_t))$, around $\bar{\mathbf{z}}_t$ has the generic form,

$$w(\mathbf{z},\mathrm{n}_\mu(\mathbf{z}),\mathrm{e}(\phi_t)) = 1 + \underbrace{\sum_{a=1}^{n}\sum_{\hat{z}_a\in\mathcal{Z}}\xi_t(\hat{z}_a)\frac{\partial w}{\partial \hat{z}_a}}_{\beta_{a,t}} + \frac{1}{2}\underbrace{\sum_{a=1}^{n}\sum_{b=1}^{n}\sum_{\hat{z}_a\in\mathcal{Z}}\sum_{\hat{z}_b\in\mathcal{Z}}\xi_t(\hat{z}_a)\xi_t(\hat{z}_b)\frac{\partial^2 w}{\partial\hat{z}_a\partial\hat{z}_b}}_{\beta_{ab,t}} + \mathcal{O}(\xi_t^3), \tag{C-3}$$

where 1 is individual fitness in a monomorphic population (i.e., when all individuals have phenotype $\bar{\mathbf{z}}_t$); $\xi_t(\hat{z}_a) = \hat{z}_a - \bar{z}_{a,t}$ denotes the difference between a value $\hat{z}_a$ and the average trait value a (and ξ_t in $\mathcal{O}(\xi_t^3)$ is the maximum deviation between individual trait value in the population and the population mean trait value at time t); and the set $\mathcal{Z} = \{\mathbf{z}\}\cup\mathrm{n}_\mu(\mathbf{z})\cup\mathrm{e}(\phi_t)$ collects the phenotypes of the focal, its neighbors, and those in other groups (so that it has $2N$ elements). The term $\beta_{a,t}$ in eq. (C-3) collects the marginal effects of a change in trait a on focal fitness: it sums the marginal effects of changing trait a ($\partial w/\partial\hat{z}_a$, where here and hereafter all derivatives are evaluated when all individuals have mean phenotype $\bar{\mathbf{z}}_t$) across all individuals that belong to $\mathcal{Z}$. Similarly, $\beta_{ab,t}$ collects the interaction effects of changes in traits a and b on focal fitness (summing the interaction effects of changes in traits a and b within, and between, all individuals that are in $\mathcal{Z}$).

We will develop these marginal ($\beta_{a,t}$ in section C.1) and interaction ($\beta_{ab,t}$ in section C.2) effects on individual fitness, and then average them over the group distribution an individual can reside in, $q(\mu|\mathbf{z},\phi_t)$, to obtain

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

lineage fitness (eq. C-1). But first, note that lineage fitness in a monomorphic population is,

$$W(\mathbf{z},\phi_t) = \int_{\mathcal{S}} 1 \times q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu + \mathcal{O}(\xi_t) = 1 + \mathcal{O}(\xi_t), \tag{C-4}$$

i.e., also one.

C.1 Marginal effects

We first develop the marginal effects, $\beta_{a,t}$, of varying trait a on individual fitness. To distinguish the effects of varying the trait in different individuals, we will use the symbols

$$\mathbf{z}_{\mathrm{n}} = \{z_{\mathrm{n},a},\ldots,z_{\mathrm{n},n}\} \in \mathrm{n}_\mu(\mathbf{z}), \quad \mathbf{z}_{\mathrm{m}} = \{z_{\mathrm{m},a},\ldots,z_{\mathrm{m},n}\} \in \mathrm{n}_\mu(\mathbf{z}), \tag{C-5}$$

to denote the phenotypes of two distinct individuals from the focal group (and distinct from the focal individual), and

$$\mathbf{z}_{\mathrm{nn}} = \{z_{\mathrm{nn},a},\ldots,z_{\mathrm{nn},n}\} \in \mathrm{e}(\phi_t), \quad \mathbf{z}_{\mathrm{mm}} = \{z_{\mathrm{mm},a},\ldots,z_{\mathrm{mm},n}\} \in \mathrm{e}(\phi_t), \tag{C-6}$$

to denote the phenotypes of two distinct individuals from a group different to the focal.

With these notations, $\beta_{a,t}$ can be decomposed into

$$\beta_{a,t} = \xi_t(z_a)\frac{\partial w}{\partial z_a} + \sum_{\mathbf{z}_{\mathrm{n}}\in\mathrm{n}_\mu(\mathbf{z})} \xi_t(z_{\mathrm{n},a})\frac{\partial w}{\partial z_{\mathrm{n},a}} + \sum_{\mathbf{z}_{\mathrm{nn}}\in\mathrm{e}(\phi_t)} \xi_t(z_{\mathrm{nn},a})\frac{\partial w}{\partial z_{\mathrm{nn},a}}, \tag{C-7}$$

where the first, second and third summands capture the marginal effect of varying the trait in: the focal; its neighbors; and individual in other groups, respectively. Because the fitness function $w(\mathbf{z},\mathrm{n}_\mu(\mathbf{z}),\mathrm{e}(\phi_t))$ is invariant to permutations of elements within the sets $\mathrm{n}_\mu(\mathbf{z})$ and $\mathrm{e}(\phi_t)$ (i.e., it does not matter to individual fitness which precise neighbor or individual from another group expresses which phenotype), the derivatives in eq. (C-7) can be taken out of their sums,

$$\beta_{a,t} = \xi_t(z_a)\frac{\partial w}{\partial z_a} + \left(\sum_{\mathbf{z}_{\mathrm{n}}\in\mathrm{n}_\mu(\mathbf{z})} \xi_t(z_{\mathrm{n},a})\right)\frac{\partial w}{\partial z_{\mathrm{n},a}} + \left(\sum_{\mathbf{z}_{\mathrm{nn}}\in\mathrm{e}(\phi_t)} \xi_t(z_{\mathrm{nn},a})\right)\frac{\partial w}{\partial z_{\mathrm{nn},a}}. \tag{C-8}$$

In order to make the sums in eq. (C-8) more convenient for averaging over $q(\mu|\mathbf{z},\phi_t)$, we seek to express them in terms of the counting measure $\mu(\mathbf{z})$ that counts the number of individuals with phenotype $\mathbf{z}$ in a group in state μ (see eq. A-2 in Appendix A). This is easily achieved for the last sum in eq. (C-8), which turns out to be zero:

$$\sum_{\mathbf{z}_{\mathrm{nn}}\in\mathrm{e}(\phi_t)} \xi_t(z_{\mathrm{nn},a}) = \int_{\mathbb{R}^n}\int_{\mathcal{S}} \xi_t(z_{\mathrm{nn},a})\mu(\mathbf{z}_{\mathrm{nn}})\phi_t(\mu)\mathrm{d}\mu\mathrm{d}\mathbf{z}_{\mathrm{nn}} = N\int_{\mathbb{R}^n} \xi_t(z_{\mathrm{nn},a})p_t(\mathbf{z}_{\mathrm{nn}})\mathrm{d}\mathbf{z}_{\mathrm{nn}} = 0 \tag{C-9}$$

(this reflects that since there is an infinite number of groups, the average deviation from the mean in groups other than the focal is zero).

For the first sum between parenthesis in eq. (C-8), we define the conditional counting measure,

$$\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}}) = \delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z})(\mu(\mathbf{z})-1) + (1-\delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z})))\mu(\mathbf{z}_{\mathrm{n}}), \tag{C-10}$$

where $\delta(.)$ is the Dirac-Delta function, so that $\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}})$ counts the number of neighbors of the focal that have phenotype $\mathbf{z}_{\mathrm{n}}$, given that the focal individual has phenotype $\mathbf{z}$. With eq. (C-10), we can write the first sum of eq. (C-8) as

$$\sum_{\mathbf{z}_{\mathrm{n}}\in\mathrm{n}_\mu(\mathbf{z})} \xi_t(z_{\mathrm{n},a}) = \xi_t(z_a)(\mu(\mathbf{z})-1) + \sum_{\substack{\mathbf{z}_{\mathrm{n}}\in\mathrm{n}_\mu(\mathbf{z})\\ \mathbf{z}_{\mathrm{n}}\neq\mathbf{z}}} \xi_t(z_{\mathrm{n},a})\mu(\mathbf{z}_{\mathrm{n}}) = \int_{\mathbb{R}^n} \xi_t(z_{\mathrm{n},a})\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}})\mathrm{d}\mathbf{z}_{\mathrm{n}}. \tag{C-11}$$

Substituting eqs. (C-9) & (C-11) into eq. (C-8) then gives

$$\beta_{a,t} = \xi_t(z_a)\frac{\partial w}{\partial z_a} + \left(\int_{\mathbb{R}^n} \xi_t(z_{\mathrm{n},a})\frac{\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}})}{N-1}\mathrm{d}\mathbf{z}_{\mathrm{n}}\right)(N-1)\frac{\partial w}{\partial z_{\mathrm{n},a}}. \tag{C-12}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

C.1.1 Average marginal effects

We proceed to average the marginal effects of a change in one trait, $\beta_{a,t}$, over the group distribution an individual can reside in, $q(\mu|\mathbf{z},\phi_t)$, which is necessary to obtain lineage fitness (eq. C-1). From eq. (C-12), this average can be written as,

$$\int_{\mathcal{S}} \beta_{a,t} q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu = \xi_t(z_a)\frac{\partial w}{\partial z_a} + \left(\int_{\mathbb{R}^n} \xi_t(z_{\mathrm{n},a}) q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)\mathrm{d}\mathbf{z}_\mathrm{n}\right)(N-1)\frac{\partial w}{\partial z_{\mathrm{n},a}}, \tag{C-13}$$

where we have defined

$$q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t) = \int_{\mathcal{S}} \frac{\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_\mathrm{n})}{N-1} q(\mu|\mathbf{z}_\mathrm{n},\phi_t)\mathrm{d}\mu \tag{C-14}$$

as the p.d.f. for the event of sampling an individual with phenotype $\mathbf{z}_\mathrm{n}$ within a group, given that within this group a focal individual with phenotype $\mathbf{z}$ has already been sampled (and removed).

C.1.2 Pairwise relatedness

The p.d.f. $q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)$ can be connected to the notion of pairwise relatedness by noting that two neighbors with the same phenotype may have a common ancestor who resided in the same group, which in the infinite island model is equivalent to the event that these individuals are identical-by-descent (IBD, e.g., Rousset, 2002). To make this connection explicit, we decompose $q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)$ as,

$$q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t) = r_{2,t}(\mathbf{z})\delta(\mathbf{z}_\mathrm{n}-\mathbf{z}) + k_t(\mathbf{z},\mathbf{z}_\mathrm{n})p_t(\mathbf{z}_\mathrm{n}), \tag{C-15}$$

where $r_{2,t}(\mathbf{z})$ is the conditional p.d.f for the event that, given a focal individual has phenotype $\mathbf{z}$ at time t, a randomly sampled individual among its neighbors is IBD to this focal (this p.d.f. depends on the whole phenotypic distribution, ϕ_t, which is captured by the time index in $r_{2,t}(\mathbf{z})$). We refer to $r_{2,t}(\mathbf{z})$ as pairwise relatedness. The two summands in eq. (C-15) respectively capture two complementary events: the sampled neighbor is either (1) IBD with the focal (and thus must have the same phenotype as the focal, $\mathbf{z}_\mathrm{n} = \mathbf{z}$); or (2) not IBD with the focal and has phenotype $\mathbf{z}_\mathrm{n}$ (which may or may not be equal to $\mathbf{z}$). We have written the p.d.f. for this latter event as $k_t(\mathbf{z},\mathbf{z}_\mathrm{n})p_t(\mathbf{z}_\mathrm{n})$, where $p_t(\mathbf{z}_\mathrm{n})$ is the marginal probability of sampling an individual with phenotype $\mathbf{z}_\mathrm{n}$ from the global population. The function $k_t(\mathbf{z},\mathbf{z}_\mathrm{n})$ can therefore be viewed as the multiplicative effect of having already sampled an individual with phenotype $\mathbf{z}$ in the group on this marginal probability.

In our endeavour to obtain an expression for lineage fitness up to the order ξ_t^2, we seek to express $q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)$ to the order of ξ_t (as it multiplies $\xi_t(z_{\mathrm{n},a})$, which is of order ξ_t, in eq. C-13). We thus Taylor expand both $r_{2,t}(\mathbf{z})$ and $k_t(\mathbf{z},\mathbf{z}_\mathrm{n})$ in eq. (C-15) to the first-order around $\bar{\mathbf{z}}_t$, and obtain

$$\begin{aligned} q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t) = &\left(r_{2,t}^{\circ}(\bar{\mathbf{z}}_t) + \sum_{a=1}^{n}\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}\xi_t(z_a)\right)\delta(\mathbf{z}_\mathrm{n}-\mathbf{z}) \\ &+\left(k_t(\bar{\mathbf{z}}_t,\bar{\mathbf{z}}_t) + \sum_{a=1}^{n}\frac{\partial k_t(\mathbf{z},\mathbf{z}_\mathrm{n})}{\partial z_a}\xi_t(z_a) + \sum_{a=1}^{n}\frac{\partial k_t(\mathbf{z},\mathbf{z}_\mathrm{n})}{\partial z_{\mathrm{n},a}}\xi_t(z_{\mathrm{n},a})\right)p_t(\mathbf{z}_\mathrm{n}) + \mathcal{O}(\xi_t^2), \end{aligned} \tag{C-16}$$

where $r_{2,t}^{\circ}(\bar{\mathbf{z}}_t)$ is the probability that two randomly sampled individuals within a group at time t are IBD under neutrality (i.e., when the population is monomorphic for $\bar{\mathbf{z}}_t$).

Next, we use the fact that from the definition of $q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)$ (eq. C-14), $q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)p_t(\mathbf{z})$ is the (unconditional) p.d.f. for the event of sampling one individual with phenotype $\mathbf{z}$ and another with $\mathbf{z}_\mathrm{n}$ *without* replacement

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

from a group. Thus, three identities must hold:

$$\int_{\mathbb{R}^n} q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)p_t(\mathbf{z})\mathrm{d}\mathbf{z}_\mathrm{n} = p_t(\mathbf{z}) \tag{C-17}$$

$$\int_{\mathbb{R}^n} q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)p_t(\mathbf{z})\mathrm{d}\mathbf{z} = p_t(\mathbf{z}_\mathrm{n}), \tag{C-18}$$

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)p_t(\mathbf{z})\mathrm{d}\mathbf{z}_\mathrm{n}\mathrm{d}\mathbf{z} = 1, \tag{C-19}$$

which can be used to express $k_t(\mathbf{z},\mathbf{z}_\mathrm{n})$ in terms of $r_{2,t}(\mathbf{z})$. In fact, substituting eq. (C-16) into eq. (C-19), we obtain

$$r^\circ_{2,t}(\bar{\mathbf{z}}_t) + k_t(\bar{\mathbf{z}}_t,\bar{\mathbf{z}}_t) = 1 \Rightarrow k_t(\bar{\mathbf{z}}_t,\bar{\mathbf{z}}_t) = 1 - r^\circ_{2,t}(\bar{\mathbf{z}}_t). \tag{C-20}$$

Substituting eq. (C-20) into eq. (C-16), which is in turn substituted into eq. (C-17) yields,

$$\begin{aligned} &p_t(\mathbf{z}) + \sum_{a=1}^n \frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}\xi_t(z_a)p_t(\mathbf{z}) + \sum_{a=1}^n \frac{\partial k_t(\mathbf{z},\mathbf{z}_\mathrm{n})}{\partial z_a}\xi_t(z_a)p_t(\mathbf{z}) = p_t(\mathbf{z}) \\ &\Rightarrow \sum_{a=1}^n \left(\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a} + \frac{\partial k_t(\mathbf{z},\mathbf{z}_\mathrm{n})}{\partial z_a}\right)\xi_t(z_a)p_t(\mathbf{z}) = 0. \end{aligned} \tag{C-21}$$

As the above equality holds for all possible $\xi_t(z_a)$ and for all $a = 1,\ldots,n$, we must have

$$\frac{\partial k_t(\mathbf{z},\mathbf{z}_\mathrm{n})}{\partial z_a} = -\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}. \tag{C-22}$$

Substituting eq. (C-20) into eq. (C-16), which is in turn substituted into (C-18) and using a similar argument, we obtain that,

$$\frac{\partial k_t(\mathbf{z},\mathbf{z}_\mathrm{n})}{\partial z_{\mathrm{n},a}} = -\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}. \tag{C-23}$$

Plugging eqs. (C-20), (C-22) and (C-23) into eq. (C-16) then gives us

$$q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t) = q^\circ_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t) + \left(\sum_{a=1}^n \xi_t(z_a)\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}\right)\delta(\mathbf{z}_\mathrm{n}-\mathbf{z}) - \left(\sum_{a=1}^n (\xi_t(z_a)+\xi_t(z_{\mathrm{n},a}))\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}\right)p_t(\mathbf{z}_\mathrm{n}) + \mathcal{O}(\xi_t^2), \tag{C-24a}$$

where

$$q^\circ_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t) = r^\circ_{2,t}(\bar{\mathbf{z}}_t)\delta(\mathbf{z}_\mathrm{n}-\mathbf{z}) + \left(1 - r^\circ_{2,t}(\bar{\mathbf{z}}_t)\right)p_t(\mathbf{z}_\mathrm{n}), \tag{C-24b}$$

is the neutral (conditional) p.d.f. for the event of sampling an individual with phenotype $\mathbf{z}_\mathrm{n}$ within the neighborhood of an individual with phenotype $\mathbf{z}$.

Regression definition of relatedness. It is noteworthy that the definition of pairwise relatedness as the probability of IBD between two randomly sampled individuals within a group that we use, aligns with the "regression definition of relatedness" (e.g., Grafen, 1985, eq. 2.13 of Frank, 1998). Under this latter definition, relatedness is the regression of neighbor phenotype on focal phenotype: it is the ratio of phenotypic covariance among neighbors,

$$\varphi_{ab,t} = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_{\mathrm{n},b})q_2(\mathbf{z}_\mathrm{n}|\mathbf{z},\phi_t)p_t(\mathbf{z})\mathrm{d}\mathbf{z}_\mathrm{n}\mathrm{d}\mathbf{z}, \tag{C-25}$$

to the covariance within individuals, $\sigma_{ab,t} = \int_{\mathbb{R}^n}\xi_t(z_a)\xi_t(z_b)p_t(\mathbf{z})\mathrm{d}\mathbf{z}$. Substituting eq. (C-24) into eq. (C-25) to compute this ratio, we obtain

$$\frac{\varphi_{ab,t}}{\sigma_{ab,t}} = r^\circ_{2,t}(\bar{\mathbf{z}}_t) + \sum_{c=1}^n \frac{\kappa_{abc,t}}{\sigma_{ab,t}}\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_c} + \mathcal{O}(\xi_t^2) \tag{C-26}$$

where $\kappa_{abc,t} = \int_{\mathbb{R}^n}\xi_t(z_a)\xi_t(z_b)\xi_t(z_c)p_t(\mathbf{z})\mathrm{d}\mathbf{z}$ is the skew between traits a, b and c within individuals. Our eq. (C-26) is thus in line with the result that in the infinite island model and under neutrality, relatedness as a regression coefficient is equal to the probability of identity-by-descent (i.e., $\varphi_{ab,t}/\sigma_{ab,t} = r^\circ_{2,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t)$, Rousset, 2002; for equivalent expressions to eq. C-26, see for e.g. eq. 7 of Queller, 1992a; eq. 2.13 of Frank, 1998; eq. 11 of Wakano and Lehmann, 2014).

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Average marginal effects. We can now return to our calculation of the average marginal effects of a change in one trait (eq. C-13). Substituting eq. (C-24) into eq. (C-13), we in fact obtain

$$\int_{\mathcal{S}} \beta_{a,t} q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu = \xi_t(z_a)\left(\frac{\partial w}{\partial z_a} + (N-1) r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\frac{\partial w}{\partial z_{\mathrm{n},a}}\right) + \sum_{b=1}^{n}\left(\xi_t(z_a)\xi_t(z_b) - \sigma_{ab,t}\right)\frac{\partial r_{2,t}(\mathbf{z})}{\partial z_b}(N-1)\frac{\partial w}{\partial z_{\mathrm{n},a}} + \mathcal{O}(\xi_t^3), \tag{C-27}$$

for the average marginal effect of trait a.

C.2 Interaction effects

We now tackle the interaction effects on focal fitness, $\beta_{ab,t}$ (eq. C-3), which we will then average over $q(\mu|\mathbf{z},\phi_t)$. Using notation eqs. (C-5)-(C-6), we first decompose $\beta_{ab,t}$ as

$$\beta_{ab,t} = \beta^{\text{same}}_{ab,t} + \beta^{\text{neigh}}_{ab,t} + \beta^{\text{cross isle}}_{ab,t}, \tag{C-28}$$

where

$$\beta^{\text{same}}_{ab,t} = \xi_t(z_a)\xi_t(z_b)\frac{\partial^2 w}{\partial z_a \partial z_b} + \sum_{\mathbf{z}_{\mathrm{n}}\in \mathrm{n}_\mu(\mathbf{z})} \xi_t(z_{\mathrm{n},a})\xi_t(z_{\mathrm{n},b})\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{n},b}} + \sum_{\mathbf{z}_{\mathrm{nn}}\in \mathrm{e}(\phi_t)} \xi_t(z_{\mathrm{nn},a})\xi_t(z_{\mathrm{nn},b})\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nn},b}}, \tag{C-29}$$

collects the interaction effects of traits a and b within individuals,

$$\beta^{\text{neigh}}_{ab,t} = \sum_{\mathbf{z}_{\mathrm{n}}\in \mathrm{n}_\mu(\mathbf{z})} \xi_t(z_a)\xi_t(z_{\mathrm{n},b})\frac{\partial^2 w}{\partial z_a \partial z_{\mathrm{n},b}} + \sum_{\mathbf{z}_{\mathrm{n}}\in \mathrm{n}_\mu(\mathbf{z})} \xi_t(z_b)\xi_t(z_{\mathrm{n},a})\frac{\partial^2 w}{\partial z_b \partial z_{\mathrm{n},a}} + \sum\sum_{\mathbf{z}_{\mathrm{n}},\mathbf{z}_{\mathrm{m}}\in \mathrm{n}_\mu(\mathbf{z})} \xi_t(z_{\mathrm{n},a})\xi_t(z_{\mathrm{m},b})\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{m},b}} + \sum\sum_{\mathbf{z}_{\mathrm{nn}},\mathbf{z}_{\mathrm{nm}}\in \mathrm{e}(\phi_t)} \xi_t(z_{\mathrm{nn},a})\xi_t(z_{\mathrm{nm},b})\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nm},b}}, \tag{C-30}$$

collects the interaction effects between neighbors, and

$$\beta^{\text{cross isle}}_{ab,t} = \sum_{\mathbf{z}_{\mathrm{nn}}\in \mathrm{e}(\phi_t)} \xi_t(z_a)\xi_t(z_{\mathrm{nn},b})\frac{\partial^2 w}{\partial z_a \partial z_{\mathrm{nn},b}} + \sum_{\mathbf{z}_{\mathrm{nn}}\in \mathrm{e}(\phi_t)} \xi_t(z_b)\xi_t(z_{\mathrm{nn},a})\frac{\partial^2 w}{\partial z_b \partial z_{\mathrm{nn},a}} + \sum\sum_{\substack{\mathbf{z}_{\mathrm{n}}\in \mathrm{n}_\mu(\mathbf{z})\\ \mathbf{z}_{\mathrm{nn}}\in \mathrm{e}(\phi_t)}} \xi_t(z_{\mathrm{n},a})\xi_t(z_{\mathrm{nn},b})\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{nn},b}} + \sum\sum_{\substack{\mathbf{z}_{\mathrm{n}}\in \mathrm{n}_\mu(\mathbf{z})\\ \mathbf{z}_{\mathrm{nn}}\in \mathrm{e}(\phi_t)}} \xi_t(z_{\mathrm{n},b})\xi_t(z_{\mathrm{nn},a})\frac{\partial^2 w}{\partial z_{\mathrm{n},b}\partial z_{\mathrm{nn},a}}, \tag{C-31}$$

the interaction effects between individuals between groups. We proceed to specify each of these in terms of counting measures and average them over $q(\mu|\mathbf{z},\phi_t)$ sequentially.

C.2.1 Interaction effects within individuals

We start with interaction effects within individuals, $\beta^{\text{same}}_{ab,t}$. Using eq. (C-10), eq. (C-29) can be expressed as

$$\beta^{\text{same}}_{ab,t} = \xi_t(z_a)\xi_t(z_b)\frac{\partial^2 w}{\partial z_a \partial z_b} + \left(\int_{\mathbb{R}^n} \xi_t(z_{\mathrm{n},a})\xi_t(z_{\mathrm{n},b})\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}})\mathrm{d}\mathbf{z}_{\mathrm{n}}\right)\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{n},b}} + \left(\int_{\mathbb{R}^n}\int_{\mathcal{S}} \xi_t(z_{\mathrm{nn},a})\xi_t(z_{\mathrm{nn},b})\mu(\mathbf{z}_{\mathrm{nn}})\phi_t(\mu)\mathrm{d}\mathbf{z}_{\mathrm{nn}}\mathrm{d}\mu\right)\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nn},b}}, \tag{C-32}$$

where the first summand consists of the interaction effect within the focal individual, the second within neighbors of the focal, and the third within individuals from other groups than the focal. To simplify these, we use the fact that because the population size remains constant, we have

$$\frac{\partial^2 w}{\partial z_a \partial z_b} + (N-1)\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{n},b}} + N\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nn},b}} = 0 \tag{C-33}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(see eq. B.13 of Wakano and Lehmann, 2014); and that the third term of eq. (C-32) can be expressed as

$$\int_{\mathbb{R}^n}\int_{\mathcal{S}} \xi_t(z_{\mathrm{nn},a})\xi_t(z_{\mathrm{nn},b})\mu(\mathbf{z}_{\mathrm{nn}})\phi_t(\mu)\mathrm{d}\mu\mathrm{d}\mathbf{z}_{\mathrm{nn}} = N\int_{\mathbb{R}^n} \xi_t(z_{\mathrm{nn},a})\xi_t(z_{\mathrm{nn},b})p_t(\mathbf{z}_{\mathrm{nn},a})\mathrm{d}\mathbf{z}_{\mathrm{nn}} = N\sigma_{ab,t}. \tag{C-34}$$

Substituting eqs. (C-33) and (C-34) into eq. (C-32), we find that the interaction effects of a and b within individuals can be written as

$$\beta^{\mathrm{same}}_{ab,t} = (\xi_t(z_a)\xi_t(z_b) - \sigma_{ab,t})\frac{\partial^2 w}{\partial z_a \partial z_b} + \left(\int_{\mathbb{R}^n} \xi_t(z_{\mathrm{n},a})\xi_t(z_{\mathrm{n},b})\frac{\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}})}{N-1}\mathrm{d}\mathbf{z}_{\mathrm{n}} - \sigma_{ab,t}\right)(N-1)\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{n},b}}. \tag{C-35}$$

Averaging these interaction effects within individuals over $q(\mu|\mathbf{z},\phi_t)$ then gives

$$\int_{\mathcal{S}} \beta^{\mathrm{same}}_{ab,t} q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu = (\xi_t(z_a)\xi_t(z_b) - \sigma_{ab,t})\left(\frac{\partial^2 w}{\partial z_a \partial z_b} + (N-1)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{n},b}}\right) + \mathcal{O}(\xi_t^3), \tag{C-36}$$

where we used definition eq. (C-14) and eq. (C-24).

C.2.2 Interaction effects between neighbors

Let us now turn to interaction effects between neighbors, $\beta^{\mathrm{neigh}}_{ab,t}$ (eq. C-30). It is composed of four summands: the first two capture the interaction effects between the focal and its neighbors; the third, between two neighbors of the focal; and the fourth, between two neighbors from another group than the focal. We consider these separately below.

Interaction effects between the focal and its neighbors. The first summand of $\beta^{\mathrm{neigh}}_{ab,t}$ (eq. C-30) corresponds to the interaction effect between trait a in the focal and trait b in its neighbors. Using eq. (C-10), it can be expressed as

$$\sum_{\mathbf{z}_{\mathrm{n}}\in \mathrm{n}_{\mu}(\mathbf{z})} \xi_t(z_a)\xi_t(z_{\mathrm{n},b})\frac{\partial^2 w}{\partial z_a \partial z_{\mathrm{n},b}} = \left(\int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_{\mathrm{n},b})\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}})\mathrm{d}\mathbf{z}_{\mathrm{n}}\right)\frac{\partial^2 w}{\partial z_a \partial z_{\mathrm{n},b}}. \tag{C-37}$$

Averaging eq. (C-37) over $q(\mu|\mathbf{z},\phi_t)$ then reads as

$$\int_{\mathbb{R}^n} \xi_t(z_a)\xi_t(z_{\mathrm{n},b})q_2(\mathbf{z}_{\mathrm{n}}|\mathbf{z},\phi_t)\mathrm{d}\mathbf{z}_{\mathrm{n}}(N-1)\frac{\partial^2 w}{\partial z_a \partial z_{\mathrm{n},b}} = \xi_t(z_a)\xi_t(z_b)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)(N-1)\frac{\partial^2 w}{\partial z_a \partial z_{\mathrm{n},b}} + \mathcal{O}(\xi_t^3), \tag{C-38}$$

where we used eq. (C-24). Similarly, averaging the second summand of $\beta^{\mathrm{neigh}}_{ab,t}$ (eq. C-30), which corresponds to the interaction effect between trait b in the focal and trait a in its neighbors, yields

$$\int_{\mathcal{S}}\left[\sum_{\mathbf{z}_{\mathrm{n}}\in \mathrm{n}_{\mu}(\mathbf{z})} \xi_t(z_b)\xi_t(z_{\mathrm{n},a})\frac{\partial^2 w}{\partial z_b \partial z_{\mathrm{n},a}}\right]q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu = \xi_t(z_a)\xi_t(z_b)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)(N-1)\frac{\partial^2 w}{\partial z_b \partial z_{\mathrm{n},a}} + \mathcal{O}(\xi_t^3). \tag{C-39}$$

Interaction effects between neighbors of the focal. The third summand of $\beta^{\mathrm{neigh}}_{ab,t}$ (eq. C-30) collects the interaction effects between neighbors of the focal. To express these in terms of the counting measure $\mu(\mathbf{z})$ and average them over $q(\mu|\mathbf{z},\phi_t)$, we introduce one further conditional counting measure,

$$\begin{aligned}\mu_{\mathrm{n},\mathbf{z}}(\mathbf{z}_{\mathrm{n}},\mathbf{z}_{\mathrm{m}}) &= \delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z})\delta(\mathbf{z}_{\mathrm{m}}-\mathbf{z}_{\mathrm{n}})(\mu(\mathbf{z})-1)(\mu(\mathbf{z})-2)\\ &+\delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z})(1-\delta(\mathbf{z}_{\mathrm{m}}-\mathbf{z}))(\mu(\mathbf{z})-1)\mu(\mathbf{z}_{\mathrm{m}}) + \delta(\mathbf{z}_{\mathrm{m}}-\mathbf{z})(1-\delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z}))(\mu(\mathbf{z})-1)\mu(\mathbf{z}_{\mathrm{n}})\\ &+\delta(\mathbf{z}_{\mathrm{m}}-\mathbf{z}_{\mathrm{n}})(1-\delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z}))\mu(\mathbf{z}_{\mathrm{n}})(\mu(\mathbf{z}_{\mathrm{n}})-1)\\ &+(1-\delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z}))(1-\delta(\mathbf{z}_{\mathrm{m}}-\mathbf{z}))(1-\delta(\mathbf{z}_{\mathrm{n}}-\mathbf{z}_{\mathrm{m}}))\mu(\mathbf{z}_{\mathrm{n}})\mu(\mathbf{z}_{\mathrm{m}})\end{aligned} \tag{C-40}$$

which counts the number of unordered pairs of neighbors that have phenotypes $\mathbf{z}_{\mathrm{n}}$ and $\mathbf{z}_{\mathrm{m}}$, given that the focal individual has phenotype $\mathbf{z}$. Integrated over $\mathbb{R}^n$, the first line of eq. (C-40) counts the number of pairs of neighbors that have focal phenotype $\mathbf{z}$; the second line, the number of pairs of neighbors in which only one has the focal phenotype $\mathbf{z}$; the third line, the number of pairs in which both neighbors have the same phenotype

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

that is different to the the focal; and the fourth line, the number of pairs in which neighbors have phenotypes that are different to one another, and to the focal.

Using eq. (C-40), we can then re-write the third summand of $\beta^{\text{neigh}}_{ab,t}$ as

$$\sum_{\mathbf{z}_{\text{n}},\mathbf{z}_{\text{m}}\in \text{n}_{\mu}(\mathbf{z})}\sum \xi_t(z_{\text{n},a})\xi_t(z_{\text{m},b})\frac{\partial^2 w}{\partial z_{\text{n},a}\partial z_{\text{m},b}} = \left(\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\mu_{\text{n},\mathbf{z}}(\mathbf{z}_{\text{n}},\mathbf{z}_{\text{m}})\text{d}\mathbf{z}_{\text{n}}\text{d}\mathbf{z}_{\text{m}}\right)\frac{\partial^2 w}{\partial z_{\text{n},a}\partial z_{\text{m},b}}, \tag{C-41}$$

which averaged over $q(\mu|\mathbf{z},\phi_t)$ reads as

$$\left[\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\xi_t(z_{\text{n},a})\xi_t(z_{\text{m},b})q_3(\mathbf{z}_{\text{n}},\mathbf{z}_{\text{m}}|\mathbf{z},\phi_t)\text{d}\mathbf{z}_{\text{n}}\text{d}\mathbf{z}_{\text{m}}\right](N-1)(N-2)\frac{\partial^2 w}{\partial z_{\text{n},a}\partial z_{\text{m},b}}, \tag{C-42}$$

where

$$q_3(\mathbf{z}_{\text{n}},\mathbf{z}_{\text{m}}|\mathbf{z},\phi_t) = \int_{\mathcal{S}}\frac{\mu_{\text{n},\mathbf{z}}(\mathbf{z}_{\text{n}},\mathbf{z}_{\text{m}})}{(N-1)(N-2)}q(\mu|\mathbf{z},\phi_t)\text{d}\mu \tag{C-43}$$

is the conditional p.d.f. for the event of sampling two individuals without replacement, one with phenotype $\mathbf{z}_{\text{n}}$ and another with phenotype $\mathbf{z}_{\text{m}}$ in a group, given that a focal individual with phenotype $\mathbf{z}$ has already been sampled in that group at time t. This p.d.f. can be connected to notions of relatedness as individuals that have the same phenotype may be IBD. In fact, using a coalescence argument, we can rewrite $q_3(\mathbf{z}_{\text{n}},\mathbf{z}_{\text{m}}|\mathbf{z},\phi_t)$ as

$$\begin{aligned} q_3(\mathbf{z}_{\text{n}},\mathbf{z}_{\text{m}}|\mathbf{z},\phi_t) &= \delta(\mathbf{z}_{\text{n}}-\mathbf{z})\delta(\mathbf{z}_{\text{m}}-\mathbf{z})\, r^{\circ}_{3,t}(\bar{\mathbf{z}}_t) \\ &+\delta(\mathbf{z}_{\text{n}}-\mathbf{z})\left(r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)-r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)\right)p_t(\mathbf{z}_{\text{m}})+\delta(\mathbf{z}_{\text{m}}-\mathbf{z})\left(r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)-r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)\right)p_t(\mathbf{z}_{\text{n}}) \\ &+\delta(\mathbf{z}_{\text{n}}-\mathbf{z}_{\text{m}})\left(r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)-r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)\right)p_t(\mathbf{z}_{\text{n}}) \\ &+\left(1-3r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)+2r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)\right)p_t(\mathbf{z}_{\text{n}})p_t(\mathbf{z}_{\text{m}})+\mathcal{O}(\xi_t), \end{aligned} \tag{C-44}$$

where $r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)$ is the probability of sampling three individuals without replacement from a group are IBD in the absence of selection at time t (i.e., three-way relatedness). Each summand of eq. (C-44) capture a different possible relationship among the sampled ($\mathbf{z}_{\text{n}}$ and $\mathbf{z}_{\text{m}}$) and focal ($\mathbf{z}$) phenotype. The first line of eq. (C-44) captures the event of sampling two individuals that are both IBD with the focal (and thus with the same phenotype as the focal, $\mathbf{z}=\mathbf{z}_{\text{n}}=\mathbf{z}_{\text{m}}$); the next line, the event of sampling only one individual IBD with the focal (and thus with the same phenotype: the two summands respectively capture the cases $\mathbf{z}=\mathbf{z}_{\text{n}}$ and $\mathbf{z}=\mathbf{z}_{\text{m}}$); the third line, the event of sampling two individuals that are IBD together (and thus with phenotypes $\mathbf{z}_{\text{n}}=\mathbf{z}_{\text{m}}$) but not with the focal; and the last line, the event of sampling two individuals that are not IBD with one another or with the focal.

Substituting eq. (C-44) into eq. (C-42) then gives us

$$\left[\xi_t(z_a)\xi_t(z_b)r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)+\sigma_{ab,t}\left(r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)-r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)\right)\right](N-1)(N-2)\frac{\partial^2 w}{\partial z_{\text{n},a}\partial z_{\text{m},b}}+\mathcal{O}(\xi_t^3) \tag{C-45}$$

for the average interaction effects between neighbors of the focal.

Interaction effects between neighbors from other groups. The fourth and final summand of $\beta^{\text{neigh}}_{ab,t}$ (eq. C-30), which collects the interaction effects between neighbors from other groups, can be expressed using eq. (C-10) as

$$\begin{aligned} &\sum_{\mathbf{z}_{\text{nn}},\mathbf{z}_{\text{nm}}\in \text{e}(\phi_t)}\sum \xi_t(z_{\text{nn},a})\xi_t(z_{\text{nm},b})\frac{\partial^2 w}{\partial z_{\text{nn},a}\partial z_{\text{nm},b}} \\ &= N(N-1)\underbrace{\left(\int_{\mathcal{S}}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\xi_t(z_{\text{nn},a})\xi_t(z_{\text{nm},b})\frac{\mu_{\text{n},\mathbf{z}_{\text{nn}}}(\mathbf{z}_{\text{nm}})\mu(\mathbf{z}_{\text{nn}})}{N(N-1)}\phi_t(\mu)\text{d}\mathbf{z}_{\text{nm}}\text{d}\mathbf{z}_{\text{nn}}\text{d}\mu\right)}_{\varphi_{ab,t}}\frac{\partial^2 w}{\partial z_{\text{nn},a}\partial z_{\text{nm},b}}, \end{aligned} \tag{C-46}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

where $\varphi_{ab,t}$ is the covariance among traits a and b between individuals within groups (see eq. C-25). We can then use eq. (C-26) to specify this covariance and obtain,

$$\sum_{\mathbf{z}_{\mathrm{nn}},\mathbf{z}_{\mathrm{nm}}\in \mathrm{e}(\phi_t)}\sum \xi_t(z_{\mathrm{nn},a})\xi_t(z_{\mathrm{nm},b})\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nm},b}} = N(N-1)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\sigma_{ab,t}\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nm},b}} + \mathcal{O}(\xi_t^3). \tag{C-47}$$

Since this expression does not depend on group composition μ, it will be invariant to averaging over $q(\mu|\mathbf{z},\phi_t)$.

Average interaction effects between neighbors. Hence, the average interaction effects between neighbors ($\int_{\mathcal{S}} \beta^{\mathrm{neigh}}_{ab,t} q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu$), which is given by the sum of eqs. (C-38), (C-39), (C-45) and (C-47), reads as

$$\begin{aligned}\int_{\mathcal{S}} \beta^{\mathrm{neigh}}_{ab,t} q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu &= \xi_t(z_a)\xi_t(z_b)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)(N-1)\left(\frac{\partial^2 w}{\partial z_a\partial z_{\mathrm{n},b}} + \frac{\partial^2 w}{\partial z_b\partial z_{\mathrm{n},a}}\right)\\ &+\left[\xi_t(z_a)\xi_t(z_b)r^{\circ}_{3,t}(\bar{\mathbf{z}}_t) + \sigma_{ab,t}(r^{\circ}_{2,t}(\bar{\mathbf{z}}_t) - r^{\circ}_{3,t}(\bar{\mathbf{z}}_t))\right](N-1)(N-2)\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{m},b}}\\ &+N(N-1)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\sigma_{ab,t}\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nm},b}} + \mathcal{O}(\xi_t^3).\end{aligned} \tag{C-48}$$

This can be further simplified by using the fact than since the total population size is constant, the following holds,

$$(N-1)\frac{\partial^2 w}{\partial z_a\partial z_{\mathrm{n},b}} + (N-1)\frac{\partial^2 w}{\partial z_b\partial z_{\mathrm{n},a}} + (N-1)(N-2)\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{m},b}} + N(N-1)\frac{\partial^2 w}{\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nm},b}} = 0. \tag{C-49}$$

(see eq. B.14 of Wakano and Lehmann, 2014). Substituting for $\partial^2 w/(\partial z_{\mathrm{nn},a}\partial z_{\mathrm{nm},b})$ using eq. (C-49) into eq. (C-48) finally gives

$$\begin{aligned}\int_{\mathcal{S}} \beta^{\mathrm{neigh}}_{ab,t} q(\mu|\mathbf{z},\phi_t)\mathrm{d}\mu &= (\xi_t(z_a)\xi_t(z_b) - \sigma_{ab,t})(N-1)\\ &\times\left(r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\left(\frac{\partial^2 w}{\partial z_a\partial z_{\mathrm{n},b}} + \frac{\partial^2 w}{\partial z_b\partial z_{\mathrm{n},a}}\right) + (N-2)r^{\circ}_{3,t}(\bar{\mathbf{z}}_t)\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{m},b}}\right) + \mathcal{O}(\xi_t^3).\end{aligned} \tag{C-50}$$

C.2.3 Interaction effects between individuals between groups

The final relevant interaction effect, $\beta^{\text{cross isle}}_{ab,t}$ (eq. C-31), is the interaction between individuals that belong to different groups. However, it is straightforward to show that these effects vanish in the infinite island model of dispersal. For example, consider the first summand of eq. (C-31), which measures the effect of a change in trait a of the focal and a trait b in an individual from a group other than the focal:

$$\begin{aligned}\sum_{\mathbf{z}_{\mathrm{nn}}\in \mathrm{e}(\phi_t)} \xi_t(z_a)\xi_t(z_{\mathrm{nn},b})\frac{\partial^2 w}{\partial z_a\partial z_{\mathrm{nn},b}} &= \left(\int_{\mathbb{R}^n}\int_{\mathcal{S}} \xi_t(z_a)\xi_t(z_{\mathrm{nn},b})\mu(\mathbf{z}_{\mathrm{nn}})\phi_t(\mu)\mathrm{d}\mu\mathrm{d}\mathbf{z}_{\mathrm{nn}}\right)\frac{\partial^2 w}{\partial z_a\partial z_{\mathrm{nn},b}}\\ &= \left(N\xi_t(z_a)\int_{\mathbb{R}^n} \xi_t(z_{\mathrm{nn},b})p_t(\mathbf{z}_{\mathrm{nn}})\mathrm{d}\mathbf{z}_{\mathrm{nn}}\right)\frac{\partial^2 w}{\partial z_a\partial z_{\mathrm{nn},b}} = 0\times\frac{\partial^2 w}{\partial z_a\partial z_{\mathrm{nn},b}}.\end{aligned} \tag{C-51}$$

Similar arguments show that all the other summands of $\beta^{\text{cross isle}}_{ab,t}$ are also zero.

C.3 Putting it all together

Summing eqs. (C-4), (C-27), (C-36), (C-50), we finally obtain the second order expansion of lineage fitness,

$$W(\mathbf{z},\phi_t) = 1 - \frac{1}{2}\sum_{a=1}^{n}\sum_{b=1}^{n}\sigma_{ab,t}h_{ab,t}(\bar{\mathbf{z}}_t) + \sum_{a=1}^{n}\xi_t(z_a)s_{a,t}(\bar{\mathbf{z}}_t) + \frac{1}{2}\sum_{a=1}^{n}\sum_{b=1}^{n}\xi_t(z_a)\xi_t(z_b)h_{ab,t}(\bar{\mathbf{z}}_t) + \mathcal{O}(\xi_t^3), \tag{C-52a}$$

where

$$s_{a,t}(\bar{\mathbf{z}}_t) = \frac{\partial w}{\partial z_a} + (N-1)r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\frac{\partial w}{\partial z_{\mathrm{n},a}}, \tag{C-52b}$$

bioRxiv preprint doi: https://doi.org/10.1101/393538; this version posted May 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and

$$h_{ab,t}(\bar{\mathbf{z}}_t) = h_{\mathrm{w},ab,t}(\bar{\mathbf{z}}_t) + h_{\mathrm{r},ab,t}(\bar{\mathbf{z}}_t), \tag{C-52c}$$

with

$$\begin{aligned} h_{\mathrm{w},ab,t}(\bar{\mathbf{z}}_t) &= \frac{\partial^2 w}{\partial z_a \partial z_b} + (N-1)\left[r^{\circ}_{2,t}(\bar{\mathbf{z}}_t)\left(\frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{n},b}} + \frac{\partial^2 w}{\partial z_a \partial z_{\mathrm{n},b}} + \frac{\partial^2 w}{\partial z_b \partial z_{\mathrm{n},a}} \right) + (N-2) r^{\circ}_{3,t}(\bar{\mathbf{z}}_t) \frac{\partial^2 w}{\partial z_{\mathrm{n},a}\partial z_{\mathrm{m},b}} \right], \\ h_{\mathrm{r},ab,t}(\bar{\mathbf{z}}_t) &= \frac{\partial r_{2,t}(\mathbf{z})}{\partial z_a}(N-1)\frac{\partial w}{\partial z_{\mathrm{n},b}} + \frac{\partial r_{2,t}(\mathbf{z})}{\partial z_b}(N-1)\frac{\partial w}{\partial z_{\mathrm{n},a}}. \end{aligned} \tag{C-52d}$$

Eq. (C-52) is equivalent to eq. B-1 – B-4 of Appendix B, in which we write the derivatives of individual fitness $w(\mathbf{z}, \mathrm{n}_\mu(\mathbf{z}), \mathrm{e}(\phi_t))$ with respect to z_a, $z_{\mathrm{n},a}$, and $z_{\mathrm{nn},a}$, in terms of the derivatives of the individual fitness function $w(\mathbf{z}_i, \mathbf{z}_{-i}, \bar{\mathbf{z}}_t)$ (eq. 5 of main text) with respect to $z_{i,a}$, $z_{j,a}$, and $z_{k,a}$, respectively; and add evaluation signs to all derivatives at the population mean $\bar{\mathbf{z}}_t$.

D Individual-based simulations

We performed individual based simulations for a population composed of N_d groups, each populated by N individuals, using Mathematica 11.0.1.0 (Wolfram Research, 2016). Starting with a monomorphic population, we track the evolution of the multidimensional phenotypic distribution under the constant influx of mutations. Each individual $i \in \{1, \ldots, N_\mathrm{d} N\}$ is characterized by two traits $(z_{i,1}, z_{i,2})$. At the beginning of a generation, we calculate the fecundity f_i of each individual according to its traits and those of its neighbors (using eq. 14). Then, we form the next generation of adults by sampling N individuals in each group with replacement according to parental fecundity, but to capture limited dispersal, the fecundity of each individual from the parental generation is weighted according to whether or not they belong to the group on which the breeding spot is filled: if an individual belongs to the same group in which a breeding spot is filled, its weighted fecundity is $f_i(1-m)$, where m is the dispersal probability; if it belongs to another group, its weighted fecundity is $f_i m/(N_\mathrm{d}-1)$ (as a disperser is equally likely to reach any other group, it lands with probability $1/(N_\mathrm{d}-1)$ in a focal group). Once an individual is chosen to fill the breeding spot, it mutates with probability ν, in which case we add to parental values a perturbation that is sampled from a multivariate normal distribution with mean $(0,0)$ and variance-covariance matrix $\begin{pmatrix} \sigma^{\mathrm{m}}_{11} & \sigma^{\mathrm{m}}_{12} \\ \sigma^{\mathrm{m}}_{12} & \sigma^{\mathrm{m}}_{22} \end{pmatrix}$. The resulting phenotypic values are truncated to remain between 0 and 4. We repeat the procedure for a fixed number of generations (see Figures for parameter values).