bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Evolutionary stability of jointly evolving traits in subdivided populations

Charles Mullon, Laurent Keller, and Laurent Lehmann

Department of Ecology and Evolution, University of Lausanne, 1004 Lausanne, Switzerland

Abstract

The evolutionary stability of quantitative traits depends on whether a population can resist invasion by any mutant. While uninvadability is well understood in well-mixed populations, it is much less so in subdivided populations when multiple traits evolve jointly. Here, we investigate whether a spatially subdivided population at a monomorphic equilibrium for multiple traits can withstand invasion by any mutant, or is subject to diversifying selection. Our model also explores the among traits correlations arising from diversifying selection and how they depend on relatedness due to limited dispersal. We find that selection favours a positive (negative) correlation between two traits, when the selective effects of one trait on relatedness is positively (negatively) correlated to the indirect fitness effects of the other trait. We study the evolution of traits for which this matters: dispersal that decreases relatedness, and helping that has positive indirect fitness effects. We find that when dispersal cost is low and the benefits of helping accelerate faster than its costs, selection leads to the coexistence of mobile defectors and sessile helpers. Otherwise, the population evolves to a monomorphic state with intermediate helping and dispersal. Overall, our results highlight the importance of population subdivision for evolutionary stability and correlations among traits.

Keywords: Social evolution, Infinite-island model, multi-traits phenotypes, evolutionary game theory, behavioural syndromes, dispersal syndromes

Introduction

One of the main goal of evolutionary biology is to explain patterns of phenotypic diversity, not only between but also within species. Among the most variable and striking phenotypes are social traits, i.e., traits that affect their carriers as well as other individuals in the population, like cooperative or aggressive behaviour. Recent years have witnessed an accumulation of evidence for the co-existence of individuals exhibiting diverse correlated quantitative social traits within populations of insects, mammals, birds, reptiles and fishes (e.g. Sih et al., 2004a; Pruitt et al., 2008; Ducrest et al., 2008; Carter et al., 2010; Cote et al., 2010a; Edenbrow and Croft, 2011; Wang et al., 2013). In sticklebacks, for example, social individuals that tend to be exploratory and aggressive to intruders co-exist with individuals that are less social, but also less exploratory and less aggressive (Laskowski and Bell, 2014). In addition, non-random associations among traits often influence the reproductive success of individuals (reviewed in Sih et al., 2004a; Pruitt et al., 2008), and are to some extent heritable (Drent et al., 2003; van Oers et al., 2004; Sinn et al., 2006; Ariyomo et al., 2013; Wang et al., 2013; Purcell et al., 2014; Dochtermann et al., 2015). These empirical findings have generated considerable interest, yet how selection either leads a population to exhibit little phenotypic variation, or favours phenotypic diversification maintaining correlations among traits, remains in general poorly understood.

The effects of selection on quantitative traits can be studied by looking at the adaptive dynamics of traits, which are the gradual phenotypic changes displayed by a population under the constant influx of mutations (e.g., Eshel, 1983; Parker and Maynard Smith, 1990; Christiansen, 1991; Grafen, 1991; Abrams et al., 1993; Eshel, 1996; Geritz et al., 1998). Unfortunately, a detailed description of adaptive dynamics is in general complicated to achieve. In order to understand the long-term effects of selection on jointly

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

evolving traits, it is therefore more fruitful to study the equilibria of the adaptive dynamics and their evolutionary stability. Importantly, investigating evolutionary stability provides insight into the fundamental features of adaptation in the absence of genetic constraints (Parker and Maynard Smith, 1990), and into the conditions that lead to phenotypic diversification (e.g., random environments, sexual selection, niche partitioning, trophic interactions and social behaviour, van Doorn et al., 2004; de Mazancourt and Dieckmann, 2004; Leimar, 2005; Dercole and Rinaldi, 2008; Brännström et al., 2010).

Evolutionary stability rests on two questions about trait values that are equilibria of adaptive dynamics (also known as singular values, Eshel, 1983; Taylor, 1989; Christiansen, 1991). The first is whether a mutant whose trait is closer to the singular value than the trait of the monomorphic population it arose in will invade. If that is the case, the population will converge to the singular value through recurrent substitutions, and such evolutionary attractors have been coined convergence stable. The second question asks whether a population that is monomorphic for a singular value can resist invasion by any mutant whose trait is close to the singular value. When that is the case, the singular value is said to be locally uninvadable. A singular trait value that is both convergence stable and locally uninvadable is an evolutionary end-point: a population will gradually converge to it and, in the absence of genetic drift or exogenous changes, remain there forever (Eshel, 1983). Alternatively, a convergence stable trait value may also be locally invadable. In that case, the population approaches the convergence stable value but then diversifies, possibly undergoing evolutionary branching, whereby a unimodal phenotypic distribution becomes bimodal, leading to the stable coexistence of highly differentiated morphs (Geritz et al., 1998). Evolutionary branching has been shown to occur in a number of scenarios of social interactions like mutualism, helping or competition (Ferriere et al., 2002; Doebeli et al., 2004; Dercole and Rinaldi, 2008, and references therein). Understanding both evolutionary convergence and uninvadability are therefore necessary to understand how selection leads to phenotypic diversification in social traits.

The evolutionary convergence and local uninvadability of traits evolving in isolation from one another are well-understood in well-mixed populations (or panmixia, e.g., Eshel, 1983; Taylor, 1989; Christiansen, 1991). But to understand patterns of diversification and covariation among multiple traits requires looking at the evolutionary stability of jointly evolving traits. When multiple traits are under selection, selection on one trait may affect selection on other traits (Lande and Arnold, 1983; Phillips and Arnold, 1989; Brodie et al., 1995), and when a population becomes polymorphic for multiple traits, selection can then preferentially target certain combinations of traits over others, thereby creating phenotypic correlations (Phillips and Arnold, 1989). The evolutionary stability of phenotypes that consist of multiple traits is also well-understood in well-mixed populations (e.g., Lessard, 1990; Leimar, 2009). Studies on the uninvadability of multiple traits in well-mixed populations have for instance shown that interactions among traits can facilitate evolutionary phenotypic diversification (Leimar, 2009; Doebeli and Ispolatov, 2010; Svardal et al., 2014; Débarre et al., 2014; Ito and Dieckmann, 2014).

The vast majority of natural populations, however, are spatially-structured. When dispersal is limited, spatial structure creates genetic structure: interacting individuals are more likely to carry identical alleles than individuals sampled at random from the population. As a result, selection depends on the genetic structure and indirect fitness effects (i.e., the effects that traits in neighbouring individuals have on the fitness of a focal individual, Hamilton, 1964; Eshel, 1972; Rousset, 2004). Evolutionary convergence in subdivided populations has been well-studied, and whether a singular trait value is convergent stable can be determined solely from Hamilton's selection gradient, which uses the probability that two neutral genes are identical-by-descent (i.e., genetic relatedness) as a measure of genetic structure (Frank, 1998; Day, 2001; Rousset, 2004). The joint study of Hamilton's selection gradients on multiple traits then informs on the convergence stability of multiple traits (Brown and Taylor, 2010, for a review), and this type of analysis has yielded intuitive insights into the evolutionary convergence of many co-evolving social traits, such as dispersal and sex-ratio, altruism and dispersal, altruism and punishment, or altruism and kin-recognition (e.g., Gandon, 1999; Perrin and Mazalov, 2000; Reuter and Keller, 2001; Rousset and Gandon, 2002; Lehmann and Perrin, 2002; Gardner and West, 2004; Leturque and Rousset, 2004).

By contrast, local uninvadability in subdivided populations has received far less attention and is significantly more challenging to study than in well-mixed populations (Day, 2001; Metz and Gyllenberg, 2001; Ajar, 2003; Massol et al., 2009; Wakano and Lehmann, 2014; Svardal et al., 2015). Computational methods that determine uninvadability in subdivided populations are available (Metz and Gyllenberg, 2001; Massol et al., 2009), but they do not straightforwardly reveal how uninvadability depends on genetic structure and fitness effects, which are central components of biological evolution (e.g., Hamilton, 1964;

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Frank, 1998; Rousset, 2004; Wenseleers et al., 2010, but see Svardal et al., 2015, for environmental effects on uninvadability when local populations are infinitely large). Two studies so far have helped understand the influence of genetic structure and fitness effects by interpreting the uninvadability of a singular trait value in terms of relatedness and indirect fitness effects when local populations are small (Ajar, 2003; Wakano and Lehmann, 2014). In particular, it was shown that whether a trait value is uninvadable depends on how selection on the trait affects genetic structure. However, the works of Ajar (2003) and Wakano and Lehmann (2014) looked at a trait that is evolving in isolation from any other traits. Thus, how population subdivision, which is so important to evolution, influences the diversification and the maintenance of correlations among jointly evolving traits is still poorly understood.

In this paper, we investigate mathematically when multiple evolving traits in a subdivided population are locally uninvadable, or alternatively, when diversification occurs. In the case of diversification, we also study the type of correlations among traits that are favoured by selection. In order to perform our analysis of selection, we use the growth rate of a mutation when rare. We highlight the influence of spatial genetic structure by expressing the conditions for uninvadability, diversification and the among-traits correlations favoured by selection in terms of genetic relatedness coefficients and indirect fitness effects.

Model

Life-cycle. We consider a haploid population divided into an infinite number of patches, each with N adult individuals (Wright, 1931's island model). The life cycle is as follows. (1) Patches may go extinct, and do so independently of one another. (2) Each of the N adults in a surviving patch produces offspring (in sufficient numbers for each patch to be always of size N at the beginning of stage (1) of the life cycle), and then either survives or dies. (3) Dispersal and density-dependent competition for vacated breeding spots occur.

This life-cycle allows for one, several, or all adults to die per life-cycle iteration, thereby encompassing overlapping and non-overlapping generations, as well as meta-population processes where whole patches go extinct before reproduction. We assume that each offspring has a non-zero probability of dispersal, so that patches are not completely isolated from each other, and that dispersal occurs to a randomly chosen patch (i.e., there is no isolation-by-distance). However, dispersal is allowed to occur in groups and before or after density-dependent competition, so that more than one offspring from the same natal patch can establish in the same non-natal patch.

Multidimensional phenotypes. Each individual expresses a genetically determined multidimensional phenotype that consists of n continuous traits. These traits can affect any event of the life cycle, under the assumption that the expression of these traits and their effects are independent of age. For instance, the fertility or mortality of an individual is assumed to be independent from its age and that of any other individual who may affect it.

Uninvadability of a multidimensional phenotype. A resident phenotype $\mathbf{z} = (z_1, z_2, ..., z_n)$, where z_p is the value of the p^{th} quantitative trait is said to be uninvadable if any mutation that arises as a single copy in the population and causes the expression of phenotype $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, ..., \zeta_n)$, goes extinct with probability one. The uninvadability of a resident phenotype $\mathbf{z}$ is therefore assessed by considering the evolutionary success of all possible mutations $\boldsymbol{\zeta}$ that would arise when the population is monomorphic for the resident.

Lineage fitness, uninvadability and selection on multiple traits

Lineage fitness and global uninvadability

In order to measure the evolutionary success of a mutation, we use the fact that it is very rare in the population when it arises as a single copy. Because the total number of patches is infinite, it will initially continue to remain rare, even if it increases in frequency due to selection and/or genetic drift and starts

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

to spread to other patches. As a result, interactions among mutants from different patches are very unlikely and can be neglected in the initial growth phase of the gene lineage initiated by the mutation. The evolutionary success of a mutation can therefore be assessed by measuring the individual fitness of a carrier of the mutation that is randomly sampled from the mutantlineage, when the rest of the population is monomorphic for $\mathbf{z}$. We define the lineage fitness of a mutant $\boldsymbol{\zeta}$ which arises as a single copy in a resident $\mathbf{z}$ population as

$$v(\boldsymbol{\zeta}, \mathbf{z}) = \sum_{k=1}^{N} w_k(\boldsymbol{\zeta}, \mathbf{z}) q_k(\boldsymbol{\zeta}, \mathbf{z}), \tag{1}$$

where $w_k(\boldsymbol{\zeta}, \mathbf{z})$ is an individual fitness function that gives the expected number of adult offspring produced by an individual carrying the mutation when there are k mutants in a patch (including the individual itself if generations overlap). The quantity $q_k(\boldsymbol{\zeta}, \mathbf{z})$ is the probability that a randomly drawn member of the mutant lineage resides in a patch with a total number k of mutants, and which is calculated over the quasi-stationary distribution of mutant patch types in the initial growth phase of the mutant lineage (before extinction or eventual invasion). The lineage fitness of a mutation therefore gives the expected individual fitness of a randomly sampled carrier of the mutation from its lineage (Day, 2001; Lehmann et al., 2015).

If the lineage fitness of a mutation is greater than one, then we expect it to invade the population because in that case, the individual fitness of an average carrier of the mutation is greater than the individual fitness of a resident, which is one. In fact, building on the branching process approach of Wild (2011), we find that $v(\boldsymbol{\zeta}, \mathbf{z})$ provides an exact assessment of the evolutionary success of a mutation because if, and only if $v(\boldsymbol{\zeta}, \mathbf{z}) \leq 1$, the mutation goes extinct with probability one, but if $v(\boldsymbol{\zeta}, \mathbf{z}) > 1$, there is a nonzero probability that the mutant persists in the population (Appendix A). We can therefore formulate the following uninvadability condition:

$$\mathbf{z} \text{ is uninvadable} \iff v(\boldsymbol{\zeta}, \mathbf{z}) \leq 1 \text{ for all } \boldsymbol{\zeta} \in \mathbb{R}^n, \tag{2}$$

i.e., a phenotype $\mathbf{z}$ is uninvadable if, and only if, a mutation that causes its bearers to express $\mathbf{z}$ has the greatest lineage fitness when all other individuals in the population also express $\mathbf{z}$ (see also, Lehmann et al., 2015).

Lineage fitness and other measures of fitness.

Lineage fitness can be linked to other measures of fitness that have been used to study mutant invasion. The links depend on the properties of the mutant lineage, which are captured by the patch profile distribution, $q_k(\boldsymbol{\zeta}, \mathbf{z})$.

Lineage fitness, average growth rate and basic reproductive number R_0. In general, $q_k(\boldsymbol{\zeta}, \mathbf{z})$ can always be expressed as the probability that a member of the mutant lineage, randomly drawn from the quasi-stationary distribution of the lineage, resides in a patch with k mutants (i.e., in terms of the eigenvectors of the matrix describing the mean of the branching process before extinction or eventual invasion, Appendix eq. A-8 and Harris, 1963, p.44). In that case, lineage fitness is also equal to the average growth rate of the mutant (Appendix eq. A-2; Caswell, 2001; Tuljapurkar et al., 2003). Condition (2) is then equivalent to the condition that the leading eigenvalue of the linearised deterministic dynamical system describing the growth rate of the mutant when rare in the population is less than one (Caswell, 2001; or less than zero in continuous time, Metz, 2010). In addition, because the basic reproductive number R_0 – the leading eigenvalue of the matrix that gives the total expected offspring production over the local lifetime of the mutant lineage in a patch – is a proxy to the growth rate (i.e., R_0 is sign equivalent to the growth rate around one; Ellner and Rees, 2006), condition (2) is also equivalent to the condition that $R_0 \leq 1$ (Appendix eq. A-11).

Lineage fitness proxy and R_m. When a local lineage mutant lineage (i.e., confined to a single patch) can only be initiated by a single founding mutant, it is possible to obtain a proxy for the average growth rate (i.e., an invasion fitness proxy) that is of the same functional form as lineage fitness (Appendix eqs. A-9 - A-21). In that case, the only difference with lineage fitness is that the probability $q_k(\boldsymbol{\zeta}, \mathbf{z})$ stands for the probability that a randomly drawn member of a local mutant lineage resides in the focal patch when there are k mutants (see eq. (20) of the Appendix and Box 2 of Lehmann et al., 2015).

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Written as an invasion fitness proxy, lineage fitness is easier to evaluate explicitly because it requires only a matrix inversion rather than the explicit computation of eigenvectors. The expected number R_m of emigrants produced by a local lineage founded by a single immigrant is also an invasion fitness proxy (Appendix eq. A-14; Metz and Gyllenberg, 2001; Massol et al., 2009; Mullon and Lehmann, 2014). Although sign-equivalent to lineage fitness proxy around one, R_m is not equal to lineage fitness proxy. In fact, R_m is equal to R_0 when local lineages can only be founded by a single mutant. Although lineage fitness proxy and R_m can both be used to study uninvadability when local lineages are initiated by a single mutant, lineage fitness proxy is expressed directly in terms of individual fitness w_k, which in some cases make it easier to interpret or manipulate.

Evolution when mutations have weak phenotypic effects

When mutations have weak phenotypic effects, the lineage fitness of a mutation can be approximated in terms of the deviation between resident and mutant phenotype $(\boldsymbol{\zeta} - \mathbf{z})$, by way of a Taylor expansion

$$v(\boldsymbol{\zeta}, \mathbf{z}) = 1 + (\boldsymbol{\zeta} - \mathbf{z})^{\mathrm{T}} \mathbf{s}(\mathbf{z}) + \frac{1}{2}(\boldsymbol{\zeta} - \mathbf{z})^{\mathrm{T}} \mathbf{H}(\mathbf{z})(\boldsymbol{\zeta} - \mathbf{z}) + O\left(\|\boldsymbol{\zeta} - \mathbf{z}\|^3\right), \tag{3}$$

where $O(\|\boldsymbol{\zeta} - \mathbf{z}\|^3)$ is a remainder of order $\|\boldsymbol{\zeta} - \mathbf{z}\|^3$. Here, the n-dimensional vector $\mathbf{s}(\mathbf{z})$ is the selection gradient at $\mathbf{z}$, i.e., each of its entry p,

$$s_p(\mathbf{z}) = \frac{\partial v(\boldsymbol{\zeta}, \mathbf{z})}{\partial \zeta_p}, \tag{4}$$

measures the change in lineage fitness due to varying only the trait at position p, henceforth referred to as trait p, when the population is monomorphic for $\mathbf{z}$ (all derivatives here and thereafter are evaluated at the resident value $\mathbf{z}$). Each (p, q) entry of the $n \times n$ Hessian matrix $\mathbf{H}(\mathbf{z})$,

$$h_{pq}(\mathbf{z}) = \frac{\partial^2 v(\boldsymbol{\zeta}, \mathbf{z})}{\partial \zeta_p \partial \zeta_q}, \tag{5}$$

measures how simultaneously varying traits p and q affect lineage fitness (i.e., the non-additive or interaction effects of p and q on lineage fitness).

Singular phenotype. A multidimensional phenotype $\mathbf{z}$ is called singular if the selection gradient vanishes at $\mathbf{z}$, namely if

$$\mathbf{s}(\mathbf{z}) = \mathbf{0}. \tag{6}$$

Singularity is a necessary condition for an interior phenotype to be convergent stable and/or uninvadable.

Convergence stability. When the difference between a non-singular resident ($\mathbf{s}(\mathbf{z}) \neq \mathbf{0}$) and the mutant is small ($\|\boldsymbol{\zeta} - \mathbf{z}\| << 1$), the selection gradient in the island model is sufficient to determine whether the mutant will go extinct or fix in the population (Rousset, 2004). When the mutation rate is very low, fixation of a new mutation occurs before another mutant arises – as traditionally assumed in the adaptive dynamics framework (Dercole and Rinaldi, 2008), or the weak-mutation strong-selection regime of population genetics (Gillespie, 1991), and evolution proceeds by a trait substitution sequence whereby the population jumps from one monomorphic state to another. A singular phenotype $\mathbf{z}$ will then be approached by gradual evolution, i.e., is convergence stable (Leimar, 2009), if the $n \times n$ Jacobian $\mathbf{J}(\mathbf{z})$ matrix with (p, q) entry

$$(\mathbf{J}(\mathbf{z}))_{pq} = \frac{\partial s_p(\mathbf{z})}{\partial z_q} \tag{7}$$

is negative-definite at $\mathbf{z}$, or equivalently if all its eigenvalues are negative.

Local uninvadability. At a singular resident ($\mathbf{s}(\mathbf{z}) = \mathbf{0}$), $\mathbf{H}(\mathbf{z})$ determines whether the resident phenotype is locally uninvadable. From eq. (3), $\mathbf{z}$ is uninvadable when $(\boldsymbol{\zeta} - \mathbf{z})^{\mathrm{T}} \mathbf{H}(\mathbf{z})(\boldsymbol{\zeta} - \mathbf{z}) < 0$ for all $\boldsymbol{\zeta}$, i.e., when the matrix $\mathbf{H}(\mathbf{z})$ is negative-definite, or equivalently when its eigenvalues are negative (e.g., Horn and Johnson, 1985, p. 104). Denoting $\lambda_1(\mathbf{z})$ as the dominant eigenvalue of $\mathbf{H}(\mathbf{z})$, we can sum up that $\mathbf{z}$

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

is locally uninvadable when

$$\begin{cases} \mathbf{s}(\mathbf{z}) = \mathbf{0} \text{ and,} \\ \lambda_1(\mathbf{z}) < 0. \end{cases} \tag{8}$$

The eigenvalue $\lambda_1(\mathbf{z})$ determines the maximal lineage fitness of a mutant, $1+\|\boldsymbol{\zeta}-\mathbf{z}\|^2\lambda_1(\mathbf{z})/2$, at a singular phenotype when the mutant deviates by small magnitude $\|\boldsymbol{\zeta}-\mathbf{z}\|$ from the resident (Appendix B).

Diversifying and stabilising selection. Because $\mathbf{H}(\mathbf{z})$ is symmetric and composed of real entries, if any diagonal entry of $\mathbf{H}(\mathbf{z})$ is positive, then $\mathbf{H}(\mathbf{z})$ has at least one positive eigenvalue (Horn and Johnson, 1985, p. 398). It is hence necessary that all the diagonal entries of $\mathbf{H}(\mathbf{z})$ are non-positive for $\mathbf{z}$ to be uninvadable. The evolutionary significance of these diagonal entries can be seen by considering the lineage fitness of a mutation that only changes the value of trait p by a small amount $\delta_p = (\boldsymbol{\zeta}-\mathbf{z})_p$. From eq. (3), the lineage fitness of such a mutation at a singular phenotype is $v(\boldsymbol{\zeta},\mathbf{z}) = 1+\delta_p^2 h_{pp}(\mathbf{z})/2$. When $h_{pp}(\mathbf{z}) > 0$, selection favours the invasion of any mutation that changes the value of trait p (since $\delta_p^2 > 0$), i.e., selection on trait p is diversifying or disruptive. Conversely, when $h_{pp}(\mathbf{z}) < 0$, selection on trait p is stabilising. Hence, the sign of the diagonal entries of $\mathbf{H}(\mathbf{z})$ reflects whether selection on each isolated trait is either diversifying or stabilising (also referred to as concave and convex selection respectively, Phillips and Arnold, 1989).

Synergy among traits. As eq. (5) shows, each off-diagonal entry of $\mathbf{H}(\mathbf{z})$ capture the synergistic effects among pairs of traits on lineage fitness. If $h_{pq}(\mathbf{z})$ is positive, then the effects of trait p and trait q on lineage fitness are synergistically positive and a joint increase or decrease in both trait values increase lineage fitness. Conversely, if $h_{pq}(\mathbf{z})$ is negative, opposite changes in trait values increase lineage fitness. The mathematical relationship between the eigenvalues of $\mathbf{H}(\mathbf{z})$ and its off-diagonal entries is less straightforward than with its diagonal entries, but negative eigenvalues, and thus uninvadability, tend to be associated with off-diagonal entries that are close to zero (using results for positive-definite matrices, Horn and Johnson, 1985, p. 398). Therefore, uninvadability is associated with weak synergy among traits (as found in well-mixed populations, Svardal et al., 2014; Débarre et al., 2014).

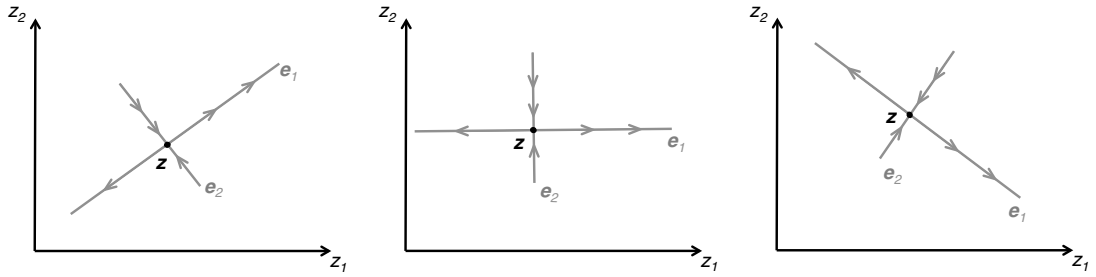
Predicting the build-up of correlations among traits. The synergistic effects among traits on lineage fitness indicate whether selection favours joint or opposite changes in pairs of traits, i.e., when $h_{pq}(\mathbf{z})$ is positive, selection favours a positive correlation among p and q, and conversely, when $h_{pq}(\mathbf{z})$ is negative, selection favours a negative correlation. This type of selection has thus been referred to as correlational selection (Phillips and Arnold, 1989). From $\mathbf{H}(\mathbf{z})$, it is possible to predict how diversifying selection leads to the build-up of phenotypic correlations among n traits under scrutiny. As shown in Appendix B, diversifying selection is greatest along the right eigenvector $e_1(\mathbf{z})$ that is associated with $\lambda_1(\mathbf{z})$, i.e., the vector such that

$$\lambda_1(\mathbf{z})\mathbf{e}_1(\mathbf{z}) = \mathbf{H}(\mathbf{z})\mathbf{e}_1(\mathbf{z}). \tag{9}$$

Thus, mutations that are most likely to invade when the resident population expresses a singular phenotype lie on $\mathbf{e}_1(\mathbf{z})$ in phenotypic space (see Fig. 1 and Appendix B, and as implied by the calculations of Doebeli and Ispolatov, 2010, in well mixed-populations). Thus, the correlations among n traits that are most likely to develop are those given by the direction of $\mathbf{e}_1(\mathbf{z})$ (Fig. 1).

The interpretation of $\mathbf{H}(\mathbf{z})$ in terms of the mode of selection on isolated, and of synergy among, traits mirrors the interpretation of the matrix of second-order effects of selection in well-mixed populations (sometimes denoted $\boldsymbol{\gamma}$ and referred to as the matrix of quadratic selection coefficients, Lande and Arnold, 1983; Phillips and Arnold, 1989; Lessard, 1990; Leimar, 2009; Doebeli and Ispolatov, 2010; Svardal et al., 2014; Débarre et al., 2014). However, it should be noted that in well-mixed populations, the matrix of second-order effects of selection matrix collects the second order effects of traits on individual fitness only (effects on w only). By contrast, $\mathbf{H}(\mathbf{z})$ here summarises the second order effects on lineage fitness, which depend on individual fitness but also on population structure and local demography. In the next section, we highlight how population subdivision affect selection by expressing the second order effects on lineage fitness in terms of individual fitness and relatedness.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



(a) Positive correlation. **(b)** No correlation. **(c)** Negative correlation.

Figure 1: Leading eigenvector and phenotypic correlations favoured by selection. The multi-dimensional phenotype consists of two traits, z_1 and z_2. The population is monomorphic for a singular phenotype $\mathbf{z}$ (black filled circle). The eigenvectors of the Hessian matrix, $\mathbf{e}_1$ and $\mathbf{e}_2$ (grey lines), are positioned to intersect at $\mathbf{z}$. A positive eigenvalue, $\lambda_1 > 0$, indicates that selection along its associated eigenvector $\mathbf{e}_1$ is diversifying, as shown by the outward arrows. In contrast, a negative eigenvalue, $\lambda_2 < 0$, tells us that selection along $\mathbf{e}_2$ is stabilising, as shown by the inward arrows. Selection on phenotypic correlations within individuals depends on the direction of $\mathbf{e}_1$. In (a), the direction of $\mathbf{e}_1$ indicates that selection favours a positive correlation; in (b), no correlation; and in (c), a negative correlation.

Uninvadability and relatedness

In order to gain greater insight into the effects of population subdivision on selection on jointly evolving traits and uninvadability, as well as connect our results to social evolutionary theory (e.g., Hamilton, 1964; Frank, 1998; Rousset, 2004; Wenseleers et al., 2010), we seek to express the selection gradient and the Hessian matrix in terms of individual fitness and relatedness.

Individual fitness and relatedness can both be recovered from lineage fitness. Lineage fitness depends on $w_k(\boldsymbol{\zeta}, \mathbf{z})$, which is the individual fitness of a mutant when there are k mutants in its patch (eq. 1). More generally, we can write the fitness of mutant and resident alike as $w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$, which is the fitness of individual indexed $i \in \{1, \ldots, N\}$ in a focal patch with phenotype $\mathbf{z}_i$, when the vector of phenotypes among its $N-1$ neighbours is $\mathbf{z}_{-i} = (\mathbf{z}_1, \ldots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \ldots, \mathbf{z}_N)$ and the phenotype carried by all other individuals in the population is the resident phenotype $\mathbf{z}$. Then, the fitness of a mutant when there are k mutants in its patch that appears in lineage fitness (eq. 1) is

$$w_k(\boldsymbol{\zeta}, \mathbf{z}) = w(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}), \tag{10}$$

where $\tilde{\mathbf{z}}_{k-1}$ is a vector of multidimensional phenotypes consisting of $k-1$ entries with phenotype $\boldsymbol{\zeta}$, and $N-k$ entries with phenotype $\mathbf{z}$. Note that because the population under consideration is not class structured, the fitness of a focal individual is not affected by which precise individual in the patch is a mutant, what matters is how many residents and how many mutants there are in a patch (i.e., $w(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z})$ is invariant under permutations of the entries of the vector $\tilde{\mathbf{z}}_{k-1}$). In order to illustrate what an individual fitness function typically looks like, and simultaneously provide a basis for examples to come, we present in Box 1 a fitness function for an iteroparous population.

Lineage fitness also depends on the probability $q_k(\boldsymbol{\zeta}, \mathbf{z})$ that a randomly sampled member of the mutant lineage has $k-1$ patch neighbours that are also members of the mutant lineage. The probability mass function $q_k(\boldsymbol{\zeta}, \mathbf{z})$ characterises identity-by-descent within a patch and therefore relatedness. In fact, the function

$$r_l(\boldsymbol{\zeta}, \mathbf{z}) = \sum_{k=1}^{N} \prod_{i=1}^{l-1} \frac{k-i}{N-i} q_k(\boldsymbol{\zeta}, \mathbf{z}), \tag{11}$$

gives the probability that $l-1$ randomly drawn neighbours without replacement of a randomly sampled mutant from its lineage are also mutants (i.e., that they all descend from the founder of the lineage; for $2 \leq l < N$). For example, $r_2(\boldsymbol{\zeta}, \mathbf{z})$ is the probability of sampling a mutant among the neighbours of a

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

random mutant individual, and thus provides a measure of pairwise relatedness between patch members. Under neutrality, all individuals in the patch have the same phenotype ($\boldsymbol{\zeta} = \mathbf{z}$), and therefore $r_l(\mathbf{z},\mathbf{z})$ reduces to the probability of sampling l individuals without replacement whose lineages are identical-by-descent, which is the standard l^{th} order measures of relatedness for the island model (e.g, Roze and Rousset, 2008, eqs. 22–27). Using the relationships (10)–(11), the selection gradient and Hessian matrix can be expressed in terms of individual fitness and relatedness.

Uninvadability in subdivided populations

The selection gradient: classical kin selection effects. First, we find that the selection gradient on trait p can be written as

$$s_p(\mathbf{z}) = \frac{\partial w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}} + r_2(\mathbf{z},\mathbf{z})\,(N-1)\,\frac{\partial w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}} \tag{12}$$

(Appendix C). The first derivative measures the change in fitness of a focal individual as a result of a change in its own trait p (i.e., it measures the direct fitness effects of trait p). The second derivative measures the change in fitness of the focal due to the change in trait p in a patch neighbour (i.e., it measures the indirect fitness effect of trait p). Owing to the permutation invariance of patch members on focal fitness (eq. 10), we arbitrarily choose this neighbour to be individual $j \neq i$. The indirect fitness effect is weighted by the neutral coefficient of relatedness $r_2(\mathbf{z},\mathbf{z})$ between two neighbours (eq. 11). Hence, eq. (12) is the usual selection gradient on a single trait for the island model of dispersal, which is Hamilton (1964)'s selection gradient on trait p, or the so-called inclusive fitness effect of trait p (Rousset, 2004).

The Hessian matrix: kin selection effects beyond neutral relatedness. Next, we find that the (p,q) entry of $\mathbf{H}(\mathbf{z})$ can be decomposed as the sum of two terms,

$$h_{pq}(\mathbf{z}) = h_{\mathrm{w},pq}(\mathbf{z}) + h_{\mathrm{r},pq}(\mathbf{z}) \tag{13a}$$

(Appendix D). The first term, $h_{\mathrm{w},pq}(\mathbf{z})$, measures the effect that joint changes of traits p and q have on individual fitness, while holding the distribution of mutants at neutrality. The second term, $h_{\mathrm{r},pq}(\mathbf{z})$, captures the effect that a change in each trait p and q has on the local distribution of mutants. As we next see, both terms depend on population subdivision and demography.

The first term of eq. (13a) is

$$\begin{aligned} h_{\mathrm{w},pq}(\mathbf{z}) = {} & \frac{\partial^2 w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}\partial z_{iq}} \\ & + r_2(\mathbf{z},\mathbf{z})(N-1)\left(\frac{\partial^2 w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}\partial z_{jq}} + \frac{\partial^2 w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{iq}\partial z_{jp}} + \frac{\partial^2 w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}\partial z_{jq}}\right) \\ & + r_3(\mathbf{z},\mathbf{z})(N-1)(N-2)\frac{\partial^2 w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}\partial z_{hq}}. \end{aligned} \tag{13b}$$

The first derivative measures the change in fitness of a focal individual as a result of a joint change in its traits p and q. The first and second derivatives on the second line of eq. (13b) measure the change in focal fitness due to a change in one trait of the focal, and a joint change in the other trait of a neighbour. The third derivative on the second line is the change in focal fitness due to a neighbour expressing joint changes in p and q. Finally, the last derivative of eq. (13b) is the change in focal fitness due to a change in one trait of a neighbour (j), and a joint change in the other trait of another neighbour ($h \neq j$). The fitness effects arising from phenotypic changes in a single neighbour are weighted by $r_2(\mathbf{z},\mathbf{z})$, and in two different neighbours by $r_3(\mathbf{z},\mathbf{z})$, which is the probability of sampling two mutants without replacement among the neighbours of a random mutant individual (which is also equal to the probability of sampling three mutants without replacement from a random patch under neutrality, $\boldsymbol{\zeta} = \mathbf{z}$).

Overall, the expression for $h_{\mathrm{w},pq}(\mathbf{z})$ measures the direct and indirect fitness effects due to a joint change in

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the values of traits p and q while holding the demography constant. It shows how synergistic effects among two traits can arise due to indirect fitness effects alone: from the fitness effects due to one genetically related neighbour changing both traits, and from those due to two different genetically related neighbours, each changing different traits.

The second term in eq. (13a) is

$$h_{\mathrm{r},pq}(\mathbf{z}) = (N-1)\left(\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}\frac{\partial r_2(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_q} + \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jq}}\frac{\partial r_2(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p}\right) \tag{13c}$$

where the first (second) product between derivatives measures the change in fitness to the focal resulting from a patch neighbour changing its trait p (q), weighted by the change in the probability that a randomly sampled neighbour of a mutant is also a mutant due to a change in trait q (p).

Eq. (13c) shows that synergy among traits can be due to the combination of indirect fitness effects of one trait, and of a change in mutant relatedness due to a change in another trait value. For example, when a change in p in a neighbour causes an increase in a focal mutant's fitness, and simultaneously a change in q increases the probability that the neighbour carry the mutation, eq. (13c) shows that this causes synergy between p and q to increase, and as a result, so does selection for mutations that change p and q jointly. Therefore, in subdivided populations, selection can favour a correlation between a trait with indirect fitness effects and another other trait that affect local demography if overall, it results in related individuals reaping greater benefits than unrelated ones.

Eq. (13) for a single trait ($p = q$) substituted into the condition for uninvadability (8) reduces to eq. (B.22) of Lehmann et al. (2015), which was derived under the assumption that only a single mutant can initiate a local lineage. In addition, eq. (13c) for a single trait ($p = q$) is consistent with previous interpretation that the second order effects of selection on an isolated trait depends on how selection affects relatedness (eq. 29 of Wakano and Lehmann, 2014, and under the assumption that generations do not overlap, see second line of eq. 9 in Ajar, 2003). Hence, our analysis not only confirms previous conditions for uninvadability of a single trait, but also shows that these hold under more general conditions, allowing for instance local patch extinctions or dispersal in groups from the same patch. But more importantly, we have extended previous analyses to consider selection on multiple traits under limited dispersal. This highlights that interactions among traits, which are important to uninvadability and selection on phenotypic correlations, can arise through the effects of traits on genetic structure and on the fitness of neighbours. We now proceed to study how uninvadability is computed explicitly, in particular showing how the effect of selection on relatedness is evaluated (i.e., $\partial r_2(\boldsymbol{\zeta},\mathbf{z})/\partial\zeta_p$).

The Moran process and weak selection

Uninvadability under a Moran process

If the population follows a Birth-Death Moran process, individual fitness is as in Box 1. Since offspring disperse independently from one another in this model, local lineages can only ever be initiated by a single individual in a patch (see below eq. A-12 in Appendix A). We can therefore use lineage fitness proxy ("**Lineage fitness proxy and $\boldsymbol{R}_m$**" section). For simplicity, we also assume in this section that there is no patch extinction. Pairwise and three-way neutral relatedness are found using standard techniques (e.g., Karlin, 1968) and are given in Table 1. Together with the fitness function (eq. I-a in Box 1), they allow for the evaluation of the selection gradient $s_p(\mathbf{z})$ (eq. 12), and for the $h_{\mathrm{w},pq}(\mathbf{z})$ (eq. 13b) component of the Hessian matrix.

The remaining term necessary for the second order effects of selection, $h_{\mathrm{r},pq}(\mathbf{z})$ (eq. 13c), depends on the

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$m(\mathbf{z})$	$\frac{s(\mathbf{z},\mathbf{z},\mathbf{z})d(\mathbf{z},\mathbf{z},\mathbf{z})}{1-d(\mathbf{z},\mathbf{z},\mathbf{z})+s(\mathbf{z},\mathbf{z},\mathbf{z})d(\mathbf{z},\mathbf{z},\mathbf{z})}$
$r_2(\mathbf{z},\mathbf{z})$ †	$\frac{1-m(\mathbf{z})}{1+m(\mathbf{z})(N-1)}$
$r_3(\mathbf{z},\mathbf{z})$ †	$\frac{2(1-m(\mathbf{z}))^2}{(2+m(\mathbf{z})(N-2))(1+m(\mathbf{z})(N-1))}$
$a(\mathbf{z})$	$\frac{(N-1)(1-m(\mathbf{z}))^2}{N-(1-m(\mathbf{z}))^2}$

Table 1: Backward dispersal, neutral relatedness and spatial scale of competition for a Moran life-cycle in the absence of patch extinction. $m(\mathbf{z})$ refers to the backward probability of dispersal (i.e., the probability that a breeding spot is filled by a dispersing offspring in a monomorphic population Gandon, 1999), which may depend on the evolving phenotype $\mathbf{z}$. † See Appendix F for derivation. The coefficient $a(\mathbf{z})$ is the spatial scale of competition when the evolving trait affect payoffs, which in turn affect fertility. It is found by Taylor expanding individual fitness to the first order of ϵ, and re-arranging to take the form of Appendix eq. (G-1). If payoffs affect other life history traits, like adult survival, then $a(\mathbf{z})$ will take a different form.

first order effect of trait p on pairwise relatedness, which we find is

$$\frac{\partial r_2(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} = \frac{\alpha(\mathbf{z})}{\mu(\mathbf{z},\mathbf{z},\mathbf{z})}\left(-\frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}}+\frac{\partial \mu(\mathbf{z}_j,\mathbf{z}_{-j},\mathbf{z})}{\partial z_{ip}}\right) + \beta(\mathbf{z})\frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}}+\gamma(\mathbf{z})\frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}, \tag{14}$$

where $\mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})$ is the death rate of individual i and philopatric fitness $w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})$ is its expected number of offspring that remain in the natal patch (Box 1, Appendix E for derivation of eq. 14). The functions $\alpha(\mathbf{z})$, $\beta(\mathbf{z})$ and $\gamma(\mathbf{z})$ are positive and decrease with the neutral measures of relatedness (expressed in terms of demographic parameters in Table 2). The second line of eq. (14) shows that mutant relatedness increases if the trait causes the mutant lineage to locally grow faster than a resident lineage, consistent with the effect on relatedness in a Wright Fisher model (Appendix eq. E-23 for link between eq. 14 and Wright-Fisher model). Because generations overlap in the Moran model, relatedness among mutants is different than among residents when the mutant affects the chance of individuals to survive from one generation to the next and this is captured by the first line of eq. (14). This first line shows that relatedness increases if a change in a trait decreases the death rate of its carrier, but increases that of a randomly sampled patch neighbour. Such a trait results in the longer coexistence of multiple mutant generations in the same patch, and therefore increases relatedness between mutants.

$\alpha(\mathbf{z})$	$\frac{N(N-1)m(\mathbf{z})(1-m(\mathbf{z}))}{(2+m(\mathbf{z})(N-2))(1+m(\mathbf{z})(N-1))^2}$
$\beta(\mathbf{z})$	$\frac{N}{(1+m(\mathbf{z})(N-1))^2}$
$\gamma(\mathbf{z})$	$\frac{2N(N-1)(1-m(\mathbf{z}))}{(2+m(\mathbf{z})(N-2))(1+m(\mathbf{z})(N-1))^2}$

Table 2: Weights for the first order effects of traits on relatedness for a Moran life-cycle in the absence of patch extinction (eq. 14). See Appendix G for derivation.

Eqs. (12)-(13) and (14) provide all the necessary components to characterise the uninvadability of multi-

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

dimensional phenotypes in subdivided populations under a Moran life-cycle. We note that if the evolving traits affect only adult fertility and/or offspring survival, philopatric fitness can be written as

$$w_{\mathrm{p}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = \frac{(1-d) f(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{(1-d) \sum_{j=1}^{N} f(\mathbf{z}_j, \mathbf{z}_{-j}, \mathbf{z})/N + ds\, f(\mathbf{z}, \mathbf{z}, \mathbf{z})}, \tag{15}$$

where $f(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$ is the number of offspring produced by individual i, d is the probability that an offspring disperses and s is the probability that it survives dispersal (Box 1). Substituting eq. (15) into eq. (14) shows that when $\mathbf{z}$ is a singular phenotype, traits have no effects on relatedness, i.e.,

$$\frac{\partial r_2(\boldsymbol{\zeta}, \mathbf{z})}{\partial \zeta_p} = 0 \tag{16}$$

(Appendix eq. E-25). So, when assessing the uninvadability of $\mathbf{z}$ under a Moran process, the effect of traits on relatedness can be ignored if there are only fecundity effects, thereby facilitating mathematical analysis. As we show in the next section, uninvadability conditions can also be made simpler for a large class of demographic models when traits have weak effects.

Uninvadability under weak selection

Within an arbitrary life cycle as laid out in the Model section, suppose that as a result of social interactions within patches, individuals receive a material payoff, like food shelter, immunity from pathogens, or some material resources. The expected material payoff obtained by individual i in a focal patch during social interactions is written $\pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$ if its phenotype is $\mathbf{z}_i$, its $N-1$ neighbours have phenotypes $\mathbf{z}_{-i}$, and the remainder of the population has resident phenotype $\mathbf{z}$. We assume that the evolving traits only affect the payoffs received during social interactions within patches, and that the payoffs affect only weakly life history traits, such as fecundity, adult survival or dispersal.

A given life history trait, written as $g(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$ for a focal individual with phenotype $\mathbf{z}_i$, can then be expressed as

$$g(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = g_{\mathrm{b}}(\mathbf{z}) + \epsilon \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) + O(\epsilon^2), \tag{17}$$

where $g_{\mathrm{b}}(\mathbf{z})$ is a baseline value for all individuals that may depend on the resident population, and $\epsilon > 0$ is the effect of payoff on the life history traits, which is assumed to be small. We assume that the fitness $w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$ of the focal individual increases with it material payoff, but decreases or is unaffected by the material payoffs of its neighbours. Also, it is assumed that the effect on the fitness of the focal of changing the payoffs of a single of its neighbour is weaker than the effect of changing the payoffs of the focal.

When the above assumptions holds, individual fitness can be expressed as a linear function of ϵ (Appendix eq. G-1). As a consequence, we find that the entries of the selection gradient $\mathbf{s}(\mathbf{z})$ are

$$s_p(\mathbf{z}) = \epsilon a_{\mathrm{f}}(\mathbf{z}) \left(\frac{\partial \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}} + \kappa(\mathbf{z}) (N-1) \frac{\partial \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}} \right) \tag{18}$$

(where $a_{\mathrm{f}}(\mathbf{z})$ is positive and model-dependent, see Appendix eqs. G-1 - G-6). This expression closely resembles the general selection gradient eq. (12), with the first term within brackets capturing direct fitness effects, and the second, indirect effects. However, instead of being expressed in terms of the individual fitness function ($w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$), eq. (18) depends directly on the payoff function, and instead of $r_2(\mathbf{z})$, the indirect effects of selection are weighted by the quantity $\kappa(\mathbf{z})$, which is a scaled measure of relatedness among two individuals that balances the effects of relatedness and local competition (e.g., Queller, 1994; Lehmann and Rousset, 2010; Akcay and Van Cleve, 2012; Van Cleve, 2015). Relatedness and local competition tend to have opposite effects on selection on social traits as the former promotes the evolution of traits with positive indirect fitness effects whilst the latter favours the evolution of traits with negative indirect fitness effects. Scaled relatedness is therefore useful towards understanding how the balance between relatedness and local competition affect social evolution (Queller, 1994; Lehmann

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and Rousset, 2010). The explicit expression and interpretation of $\kappa(\mathbf{z})$ are given in Box 2 (eq. II-b).

Importantly, we find that when traits have weak effects on fitness, they have no effect on pairwise relatedness, $\partial r_2(\boldsymbol{\zeta}, \mathbf{z})/\partial \zeta_p = 0$. As a result, the second component of the Hessian matrix vanishes ($h_{\mathrm{r},pq}(\mathbf{z}) = 0$), and the (p, q) entry of $\mathbf{H}(\mathbf{z})$ is

$$h_{pq}(\mathbf{z}) = \epsilon a_{\mathrm{f}}(\mathbf{z})\left(\frac{\partial^2 \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}\partial z_{iq}} + \iota(\mathbf{z})(N-1)\left(\frac{\partial^2 \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}\partial z_{jq}} + \frac{\partial^2 \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{iq}\partial z_{jp}}\right)\right.$$
$$\left. + \kappa(\mathbf{z})(N-1)\frac{\partial^2 \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}\partial z_{jq}} + \eta(\mathbf{z})(N-1)(N-2)\frac{\partial^2 \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}\partial z_{kq}}\right) \tag{19}$$

(Appendix eqs. G-7 - G-8). This equation bears close resemblance to eq. (13b), and its elements can be interpreted similarly. However, eq. (19) depends directly on the payoff instead of fitness functions, and on two additional scaled relatedness measures, which, like $\kappa(\mathbf{z})$, balance the effects of relatedness and local competition: $\iota(\mathbf{z})$, between two randomly sampled individuals; and $\eta(\mathbf{z})$ between three individuals (see Box 2 for a detailed interpretation of these coefficients; and when there is only one evolving trait, $n = 1$, and interactions are pairwise the term in parentheses in eq. 19 reduces to eq. 37 of Wakano and Lehmann, 2014).

Eqs. (18)-(19), along with Box 2, is all that is necessary to evaluate the uninvadability of multidimensional phenotypes in subdivided populations, provided traits only affect the payoffs received during interactions, and selection is weak. Because eqs. (18)-(19) depend on the payoff rather than the fitness function, they tend to be easier to explore mathematically. In addition, the expressions for scaled relatedness reveal more clearly the effects of demography on two antagonistic forces on social traits: relatedness and competition among kin.

However, one should be cautious that the equilibrium values found for life history traits using eq. (17) may not correspond to the equilibrium values that would be found by considering the evolution of the life history trait itself. For instance, while it is possible to model the evolution of game strategies when payoffs affect the ability to disperse using eqs. (17)–(19), equilibrium strategies found in this case may not predict the same equilibrium values for dispersal if we modelled the evolution of dispersal itself. When directly modelling the evolution of dispersal, it is not possible to use the simpler eqs. (18)-(19) because dispersal typically affects the probability that individuals of the same patch carry the same mutation (i.e., $\partial r_2(\boldsymbol{\zeta}, \mathbf{z})/\partial \zeta_p \neq 0$). We show this explicitly in the next section by considering the evolution of dispersal itself, together with another social trait.

The joint evolution of helping and dispersal

In order to illustrate the above results and how to apply them, we now study the joint evolution of helping and dispersal. The evolutionary paths of helping and dispersal are intimately intertwined (Lehmann and Perrin, 2002; Le Galliard et al., 2005; Hochberg et al., 2008; El Mouden and Gardner, 2008; Purcell et al., 2012; Parvinen, 2013) because in subdivided populations, the level of dispersal determines relatedness, and therefore tunes selection on helping traits (e.g., Rousset, 2004). Simultaneously, dispersal evolution directly responds to the level of kin competition within a patch (Hamilton and May, 1977).

The complicated interaction between helping and dispersal has meant that so far, theoretical studies of the evolution of these two traits have either focused on evolutionary convergence and ignored the problem of uninvadability (Lehmann and Perrin, 2002; Le Galliard et al., 2005; Hochberg et al., 2008; El Mouden and Gardner, 2008), or relied on simulations and numerical methods to study invadability (Purcell et al., 2012; Parvinen, 2013). Using our framework, we are able to analytically study not only the joint invadability of helping and dispersal, but also the correlation among those two traits that are favoured by selection when evolutionary branching occurs.

Biological scenario. The life cycle is assumed to follow the Moran process in the absence of patch extinction (see section Moran process and Box 1). During the adult stage, individuals pair up randomly and engage with one another in the so called "continuous snow drift game", which is a model of helping

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

behaviour that can lead to evolutionary branching among helpers and defectors (Doebeli et al., 2004; Wakano and Lehmann, 2014). In this model, the payoff received by individual i when it interacts with j is expressed as

$$b_1(x_i + x_j) + b_2(x_i + x_j)^2 - c_1 x_i - c_2 x_i^2, \tag{20}$$

where $0 \leq x_i \leq 1$ is the level of helping expressed by individual i, and x_j that of individual j. The constants b_1 and b_2 tune the benefit to i of both interacting partners investing into helping, while c_1 and c_2 tune the cost to i. Individual fertility increases with the average payoff received,

$$f(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = 1 + \epsilon\pi(x_i, \mathbf{x}_{-i}) = 1 + \epsilon \sum_{j \neq i}^{N} \frac{b_1(x_i + x_j) + b_2(x_i + x_j)^2 - c_1 x_i - c_2 x_i^2}{N - 1}, \tag{21}$$

where the parameter $\epsilon > 0$ measures the effect of payoffs on fertility. After reproduction, an adult in each patch is selected at random to die, with each adult having the same death rate. The offspring of an individual i disperses with probability $d_i \in (0, 1]$. Dispersing offspring survive during dispersal with probability s.

Set-up. From these assumptions, the fitness of individual i is given by Box 1, with no patch extinction ($e(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = 0$), constant adults survival ($\mu(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = \mu$), and constant survival during dispersal ($s(\mathbf{z}, \mathbf{z}, \mathbf{z}) = s$). The vector of trait values of an individual i consists of its level of helping and dispersal probability, $\mathbf{z}_i = (x_i, d_i)$. The vector of phenotypes in the rest of the patch is $\mathbf{z}_{-i} = ((x_1, d_1), \ldots, (x_{i-1}, d_{i-1}), (x_{i+1}, d_{i+1}), \ldots, (x_N, d_N))$, and $\mathbf{z} = (x, d)$ stands for resident strategies. Hence, trait $p = 1$ corresponds to helping and trait $p = 2$ corresponds to dispersal. The neutral relatedness functions, $r_2(\mathbf{z}, \mathbf{z})$ and $r_3(\mathbf{z}, \mathbf{z})$, are those given in Table 1. We now have all the elements to use our framework, and start by considering the evolution of each trait in isolation, and then study their coevolution.

Uninvadability of helping with fixed dispersal. Substituting fitness (eq. I-a, Box 1) into the selection gradient (eq. 12) for helping, we find that when dispersal is fixed at d for all individuals, there is a unique singular helping strategy for which the selection gradient vanishes, which is

$$x = \frac{c_1(N + 1 - m) - b_1(N + 2 - 2m)}{4b_2(N + 2 - 2m) - 2c_2(N + 1 - m)}, \tag{22}$$

where $m = m(\mathbf{z})$ is the probability that a breeding spot is filled by a dispersing offspring (i.e., the backward dispersal probability, see Table 1). The invadability of this singular point is assessed by calculating $h_{11}(\mathbf{z})$, which is found by substituting fitness (eq. I-a, Box 1) into eq. (13) and evaluating it at eq. (22). Although such a computation yields an analytical expression for $h_{11}(\mathbf{z})$, it is too complicated to generate useful insights. We therefore first study the invadability of helping under weak selection, i.e., with ϵ small in eq. (21), and using eq. (19) to compute $h_{11}(\mathbf{z})$ at eq. (22). Using the spatial scale of competition given in Table 1, and the scaled relatedness coefficients of Box 2, fitness (eq. I-a, Box 1) into eq. (19) gives

$$h_{11}(\mathbf{z}) \propto \left(-2c_2 + 2b_2 \frac{(4 + m(N - 4))(N + 2 - 2m)}{(2 + m(N - 2))(N + 1 - m)} \right), \tag{23}$$

where we have ignored ϵ and $a_f(\mathbf{z})$ since they are both positive and we are only interested into whether $h_{11}(\mathbf{z}) < 0$ or not. Eq. (19) shows that the singular point (eq. 22) is uninvadable, i.e., $h_{11} < 0$, only if

$$b_2 < c_2 \frac{(2 + m(N - 2))(N + 1 - m)}{(4 + m(N - 4))(N + 2 - 2m)}, \tag{24}$$

where the factor of c_2 on the right hand side is positive and decreases with backward dispersal (m) and patch size (N), so that it correlates positively with patch relatedness.

When dispersal is complete ($d = 1$, so that $m = 1$), eq. (24) reduces to the results previously found for well-mixed populations: the singular helping strategy (eq. 22) is uninvadable only if $b_2 < c_2$ (Doebeli et al., 2004). In other words, the benefits of helping should accelerate at the same rate or slower than its cost. Otherwise, the accelerating returns of helping favour the diversification of helping strategies, and leads to the evolutionary branching between helpers and defectors (Fig. 3, top left panel).

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

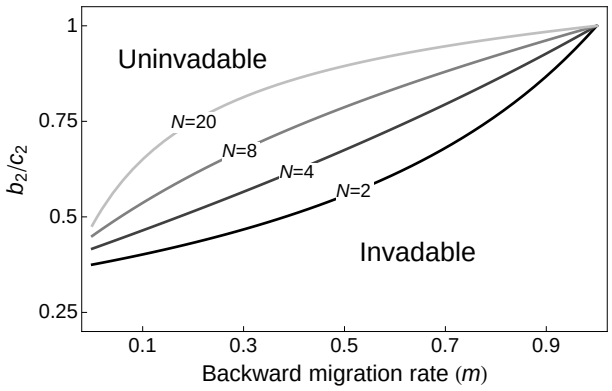


Figure 2: Uninvadability of helping in subdivided populations. Region above the curves gives combination of values of b_2/c_2 and backward migration rate m for which helping is uninvadable, while region under the curve gives values for which helping is invadable. Different curves correspond to different patch sizes. Thus, small patch size and low migration rate stabilise helping in subdivided populations.

As dispersal decreases ($d < 1$, so that $m < 1$), relatedness among individuals within patches increases, and indirect effects become increasingly important in the fate of newly arising mutations, consequently affecting the stable level of helping. Insights into the effects of population subdivision on the stability of helping can be obtained by expressing $h_{11}(\mathbf{z})$ close to full dispersal ($m \sim 1$),

$$h_{11}(\mathbf{z}) \propto 2(b_2 - c_2) + \frac{6(1-m)}{N} b_2 + O((1-m)^2). \tag{25}$$

The first and second summands of eq. (25) respectively capture direct and indirect fitness effects. The latter increases with relatedness, here measured by $(1-m)/N$. Indirect effects also increases with b_2 because the payoffs to a focal individual increase by a factor of b_2 whenever neighbours change their helping strategy. So if b_2 is positive, indirect effects favour the invasion of mutations that change the degree of helping, and thereby destabilise the singular helping strategy. Conversely, if b_2 is negative, then focal payoffs decrease, and indirect effects disfavour any change in helping strategy, stabilising the existing degree of helping.

For a given patch size, eq. (24) shows that there is a threshold value for dispersal under which helping, otherwise invadable in well-mixed populations, becomes uninvadable. For example, when $N = 4$, if $m < 2 + 2(b_2/c_2) + \sqrt{9 - 4(1 - (b_2/c_2))(b_2/c_2)}$, helping is uninvadable in subdivided populations but invadable in well-mixed populations (Fig. 2). This is consistent with previous results found for the Wright-Fisher process (Wakano and Lehmann, 2014), and highlights that when the benefits of helping are decelerating ($b_2 < 0$), indirect fitness effects combined with high relatedness promote the stability of helping in the population. Because relatedness is greater under the Moran model, the threshold value for dispersal is greater than under the Wright-Fisher model.

In order to check the robustness of our weak selection conclusions against increased selection strength, we generated random values for the model parameters, and asked whether the sign of h_{11} given by eq. (23) were of the same the sign as h_{11} calculated under strong selection ($\epsilon = 1$). Both expressions showed the same sign for all combinations of values tested (100% of 465943 trials, Appendix H), suggesting that the sign of h_{11} given by eq. (23) is an excellent predictor of the sign of h_{11} under strong selection. In addition, individual based simulations with strong selection behave as predicted by the analysis performed under weak selection (Fig. 3). This suggests that uninvadability condition computed under weak selection for traits that affect fertility is robust to increases in selection strength.

Uninvadability of dispersal with fixed helping. Substituting fitness (eq. I-a, Box 1) into the selection gradient eq. (12) with its derivatives with respect to the probability of dispersal, we find a single singular dispersal strategy,

$$d = \frac{1}{1 + N(1-s)}. \tag{26}$$

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

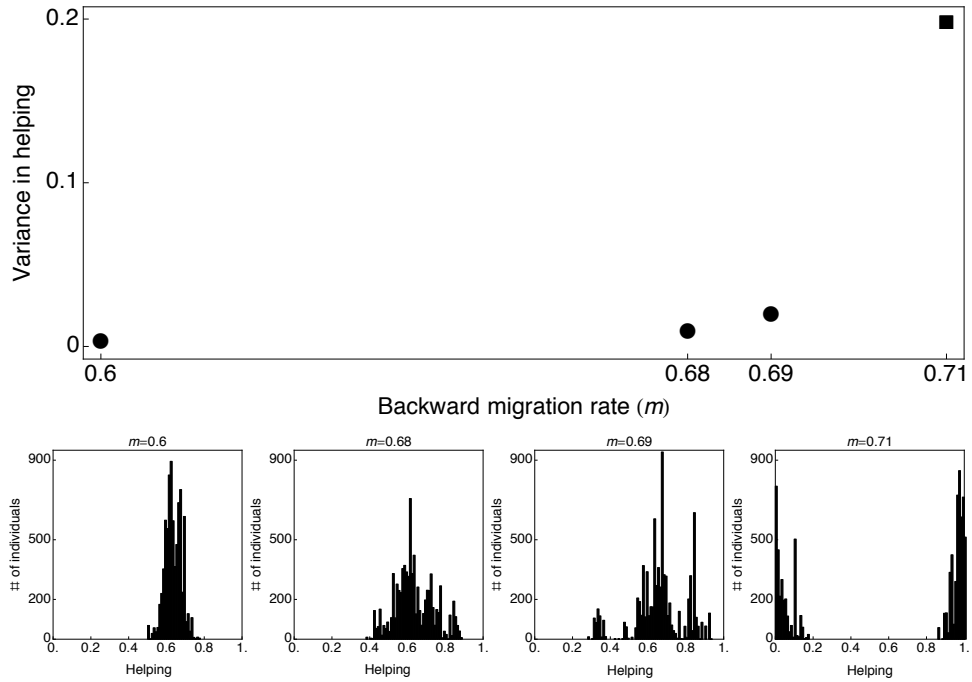


Figure 3: Evolution of helping in a subdivided population. The top panel shows the phenotypic variance of helping in a simulated population for different values of m (other parameters were held at $N = 8$, $b_1 = 6$, $b_2 = -1.4$, $c_1 = 4.56$, $c_2 = -1.6$, number of patches $= 1000$, Appendix I for details on simulations). The variance is averaged over 5×10^3 generations after 1.45×10^5 generations of evolution. Circles indicate variance for parameter values under which h_{11} (eq. 23) is negative, predicting a stable monomorphic population, while squares correspond to a positive h_{11}, predicting a polymorphic population. The bottom panel shows snapshots of the the level of helping in a simulated population after 1.5×10^5 generations for four different values of m.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Eq. (26) does not depend on helping because all individuals exhibit the same level of helping and there is thus no variation in offspring production among patches. Since kin competition decreases with patch size (N), the candidate dispersal strategy decreases with N and since the cost to dispersal decreases with the survival rate during dispersal (s), the candidate dispersal strategy increases with s. Qualitatively, eq. (26) is therefore the same as the one obtained by the classical models of dispersal evolution, which assume a Wright-Fisher reproductive process (Frank, 1998; Gandon and Rousset, 1999). However, because there is more competition among kin when generations overlap, the singular dispersal strategy under the Moran model is always greater than under the Wright-Fisher model.

Computing $h_{22}(\mathbf{z})$ from eq. (13) and evaluating it at eq. (26), we find that the singular strategy (eq. 26) is uninvadable only if

$$h_{22}(\mathbf{z}) = -\frac{2s(N(1-s)+1)^3}{N^2(N(1-s)+s)(2-s)} < 0, \tag{27}$$

which is always true. Therefore, when dispersal is the only evolving trait, dispersal is always stable in a Moran population. This corroborates the results found under the Wright-Fisher process (Ajar, 2003), but the result for the Moran process is more straightforward, with eq. (27) simpler than the Wright-Fisher condition (eq. 15 in Ajar, 2003). Interestingly, in a Moran population monomorphic for the uninvadable dispersal level, pairwise relatedness is simply

$$r_2(d,d) = 1 - s, \tag{28}$$

the probability of surviving dispersal.

The coevolution of helping and dispersal. When helping and dispersal evolve jointly, singular strategies are found by solving simultaneously for vanishing selection gradients for both traits, $s_1(\mathbf{z}) = 0$ and $s_2(\mathbf{z}) = 0$, which produces the singular point

$$(x,d) = \left(\frac{c_1(N(1-s)+1) - b_1(N(1-s)+2-s)}{4b_2(N(1-s)+2-s) - 2c_2(N(1-s)+1)}, \frac{1}{1+N(1-s)}\right). \tag{29}$$

The singular dispersal strategy is the same as in eq. (26), while the singular helping strategy is eq. (22) with $m = m(\mathbf{z})$ as in Table 1, and dispersal d given by the singular phenotype eq. (26).

The joint stability of helping and dispersal is deduced from the leading eigenvalue of $\mathbf{H}(\mathbf{z})$ (eq. 13), which is a complicated function of the model parameters. However, when the effects of helping on fertility are weak (ϵ small in eq. 21), it is possible to express the eigenvalues of $\mathbf{H}(\mathbf{z})$ as a perturbation of the simpler eigenvalues of $\mathbf{H}(\mathbf{z})$ with $\epsilon = 0$ (see Appendix J). Using this method, we find that the eigenvalues of $\mathbf{H}(\mathbf{z})$ under weak selection on helping are

$$\begin{aligned}\lambda_1(\mathbf{z}) &\simeq \epsilon\frac{s(N(1-s)+1)}{N^2(1-s)+Ns}\left(-2c_2 + 2b_2\frac{(2-s+N(1-s))(4-3s)}{(N(1-s)+1)(2-s)}\right)\\ \lambda_2(\mathbf{z}) &\simeq -\frac{2s(N(1-s)+1)^3}{N^2(2-s)(N(1-s)+s)} < 0,\end{aligned} \tag{30}$$

which, respectively, are proportional to eq. (23) and h_{22} (eq. 27) at the singular dispersal strategy (i.e., with m and d as in Table 1 and eq. 29 respectively). Therefore, when the effects of helping on fertility are weak, the condition for helping and dispersal to be jointly stable is the same as the condition for helping to be stable when dispersal is held fixed at the singular dispersal strategy (eq. 24).

Numerical comparisons between the sign of λ_1 in eq. (30) and the sign of the exact leading eigenvalue of $\mathbf{H}(\mathbf{z})$ calculated under strong selection (with $\epsilon = 1$) shows that the former is a very good predictor of the latter, with both having the same sign in 99.86% (of 464686 cases, Appendix H), and therefore suggests that the uninvadability condition (24) should also hold when the effects of helping on fertility are strong. In addition, individual based simulations with strong selection ($\epsilon = 1$) match theoretical predictions very well (Fig. 4).

Our results therefore show that invadability of helping strategy also leads to invadability in dispersal

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

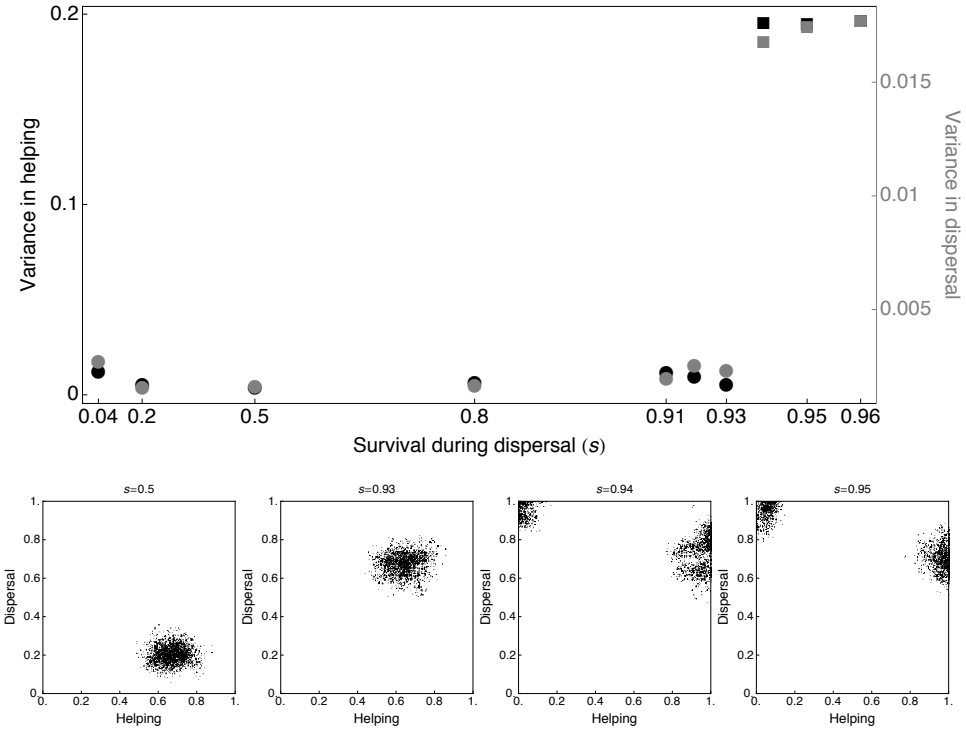


Figure 4: Coevolution of helping and dispersal in a subdivided population. The top panel shows the phenotypic variance of helping (black, left scale) and dispersal (grey, right scale) in a simulated population for different values of s (other parameters are as in Fig 3, Appendix I for details on simulations). The variances are averaged over 5×10^3 generations after 1.45×10^5 generations of evolution. Circles indicate variance for parameter values under which the leading eigenvalue of the Hessian matrix is negative, predicting a stable monomorphic population, while squares correspond to a positive eigenvalue, predicting a polymorphic population. The bottom panel shows snapshots of the phenotypes in a simulated population after 1.5×10^5 generations for four different values of s. Each point represents the phenotypic values in helping and dispersal of an individual.

strategy. In order to predict whether helpers and defectors evolve different dispersal strategies, we now study the level of correlation among helping and dispersal that is favoured by selection at the singular phenotype. Recall that among-traits correlation are given by the direction of the eigenvector associated with the leading positive eigenvalue of $\mathbf{H}(\mathbf{z})$ (section "**Predicting the build-up of correlations among traits**"). Since $\lambda_2(\mathbf{z})$ (eq. 30) is always negative, $\lambda_1(\mathbf{z})$ is the leading eigenvalue. If the effects of helping are weak (small ϵ), then it is possible to express the eigenvector associated with $\lambda_1(\mathbf{z})$ as a perturbation of the eigenvectors of $\mathbf{H}(\mathbf{z})$ in the absence of helping ($\epsilon = 0$, Appendix J). We find that in this case, the leading eigenvector has direction

$$\mathbf{e}_1(\mathbf{z}) \propto \begin{pmatrix} 1 \\ -\dfrac{h_{12}}{h_{22}} \end{pmatrix} = \begin{pmatrix} 1 \\ -\dfrac{(2b_2c_1 - b_1c_2)s}{4b_2(N(1-s)+2-s) - 2c_2(N(1-s)+1)}. \end{pmatrix} \tag{31}$$

Eq. (31) shows that, since $h_{22} < 0$, selection promotes a positive (negative) correlation among helping and dispersal if the synergy h_{12} among them is positive (negative). In order to garner greater insight into the synergy h_{12}, we approximate it as

$$h_{12}(\mathbf{z}) \simeq \epsilon \frac{2}{N} \left(\frac{2b_2c_1 - b_1c_2}{2b_2 - c_2} \right) s(1-s) \left(\frac{1}{2} - \frac{1}{2-s} \right) \tag{32}$$

to the first order of ϵ, and neglecting terms of order $O(1/N^2)$ (Appendix eq. H-1). Looking at eq. (32),

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

we see that the term inside the first parentheses is the effect that a change in the degree of helping of neighbours has on focal payoff at the singular level of helping: $(N-1)\partial\pi(x_i,\mathbf{x}_{-i})/\partial x_j = (2b_2c_1 - b_1c_2)/(2b_2 - c_2) + O(1/N)$; i.e., the indirect effects on focal's payoff. Since helping always has positive indirect effects, this term is positive. Then, because the rest of eq. (32) is negative, synergy among helping and dispersal on fitness is always negative.

Since synergy among helping and dispersal is negative, we expect that if polymorphism arises in the population, selection will favour a negative correlation among helping and dispersal. A closer look at the synergy reveals why this is so. Eq. (32) is negative due to the large negative term $-1/(2-s)$, which stems from the fact that an increase in dispersal has a negative effect on relatedness,

$$\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial x_j}\frac{\partial r_2(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_2} = \frac{4b_2c_1 - 2b_1c_2}{2b_2 - c_2}s(1-s)\left(-\frac{1}{2-s}\right) + O(\epsilon^2, 1/N^2). \tag{33}$$

Therefore, selection promotes a negative correlation among helping and dispersal because a simultaneous increase in helping and dispersal leads to greater benefits to unrelated individuals, and, conversely, a simultaneous decrease in helping and dispersal leads to lesser benefits to related individuals. By contrast, lesser dispersal coupled with greater helping leads to greater benefits to related individuals. We find by numerical simulations that the sign of eq. (32) is a very good predictor of the sign of the exact $h_{12}(\mathbf{z})$ value with arbitrary population size and under strong selection ($\epsilon = 1$, Appendix G), so that we also expect a negative correlation among helping and dispersal under strong selection. In fact, individual based simulations under strong selection show that when evolutionary branching occurs, the population splits into very mobile defectors and more sessile helpers (Fig. 4), thereby befitting our analytical predictions.

Our analytical finding of a negative correlation among helping and dispersal corroborates previous simulation results when patches do not vary in size (Purcell et al., 2012; Parvinen, 2013). When patches experience different demographies, simulations show that it is possible for selection to favour a positive correlation among helping and dispersal because helpers and defectors experience different benefits from dispersing: helpers benefit from invading patches with few individuals whereas defectors benefit from invading patches with a large number of individuals (Parvinen, 2013). Other works that have looked at the correlation among helping and dispersal have done so by studying either the convergence stable level of helping according to dispersal strategy (El Mouden and Gardner, 2008), or the convergence stable level of dispersal according to helping strategy (Hochberg et al., 2008), and found results similar to the ones mentioned here. However, it should be noted that in those studies, the presence of helpers and defectors or of disperser and non-dispersers was a starting point and a built-in assumption, rather than a product of evolution like in our analysis.

Discussion

Understanding uninvadability is key to understand the evolution of quantitative traits because uninvadability determines whether selection favours a population to remain monomorphic or become polymorphic (e.g., Eshel, 1983; Taylor, 1989; Christiansen, 1991; Geritz et al., 1998). So far, how genetic structure and indirect fitness effects influence uninvadability had only been studied when single isolated traits evolve (Ajar, 2003; Wakano and Lehmann, 2014), but organisms consist of a multitude of traits that rarely, if ever, evolve in isolation from one another (Lande and Arnold, 1983; Phillips and Arnold, 1989). Here, we have presented a framework to study the uninvadability of phenotypes that consist of multiple quantitative traits, as well as the among traits correlations arising from diversifying selection, in subdivided populations when local populations can be of any size and are connected by limited dispersal.

In order to analyse the effects of selection under limited dispersal, we used the lineage fitness of a mutation, which is the expected number of mutant copies in the offspring generation that are produced by an individual randomly drawn from the mutant lineage (see Lehmann et al., 2015, Akçay and Van Cleve, 2016 and Lehmann et al., 2016 for other applications of this fitness concept). Lineage fitness thus gives an average individual fitness over the possible genetic backgrounds in which a carrier of the mutation can reside. It allowed us to reveal the effects of genetic structure through relatedness coefficients and of indirect fitness effects on the uninvadability of multiple traits. We found that relatedness and

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

indirect fitness effects influence the synergy among traits (or strength of correlational selection). Because strong synergy tends to disfavour uninvadability and thus promote diversification, relatedness and indirect fitness effects may be critical for the evolution of multiple traits under limited dispersal. In particular, we showed that synergy among two traits can arise through the effects of one trait on genetic structure and the indirect fitness effects of the other trait. When positively correlated change in two traits results in related individuals reaping greater fitness benefits than unrelated ones, synergy tends to be positive, and conversely, when negatively correlated changes in two traits result in related individuals reaping greater fitness benefits, synergy tends to be negative.

We further found that, since the synergy among pairs of traits determines the among-traits correlations that develop when polymorphism arises, relatedness and indirect fitness effects also influence the evolution of phenotypic correlations. Because behavioural traits tend to have the greatest indirect fitness effects, our results suggest that synergy due to the combination of traits' effects on genetic structure and indirect fitness effects is important to the evolution of behavioural syndromes (correlations among behavioural traits within individuals, Sih et al., 2004a,b). Previous studies have suggested that behavioural syndromes may be maintained mechanistically by pleiotropic mutations (Ducrest et al., 2008), and by fitness trade-offs between life-history traits (Wolf et al., 2007; Réale et al., 2010). Here, our model shows that selection promotes a positive correlation among two traits when one trait has positive indirect benefits, and the other trait increases pairwise relatedness (i.e., when two individuals that show an increase in the value of the trait have a greater probability of being related than two resident individuals).

In this context, dispersal syndromes, which refer to patterns of covariation between the tendency to disperse and other traits, are of particular interest (Clobert et al., 2009). Dispersal syndromes are ecologically and evolutionarily relevant as they influence the demographic and genetic consequences of movement (Edelaar and Bolnick, 2012; Jean Clobert et al., 2012). In Western bluebirds, for example, individuals that disperse further away from their natal site also tend to be more aggressive towards conspecifics and towards sister species, and this has caused a shift in the range of these two species (Duckworth and Badyaev, 2007; Duckworth and Kruuk, 2009). Since dispersal decreases relatedness, our model predicts that the tendency to disperse should be negatively correlated with behaviours that have positive indirect fitness effects. In our model of helping, we indeed found that when evolutionary branching occurs, the population splits between full defectors that always disperse and full helpers that are more sessile. This negative correlation among helping and dispersal also matched previous simulation results when patches do not vary in size (Purcell et al., 2012; Parvinen, 2013). Similarly, our model predicts dispersal to be positively correlated with behaviours that have negative indirect fitness effects, like aggressiveness. Interestingly, both a negative correlation between dispersal and pro-social behaviour (Ims and Ims, 1990; Mehlman et al., 1995; O'Riain et al., 1996; Sinervo and Clobert, 2003) and a positive correlation between dispersal and aggressive behaviour have been observed in natural populations of voles, mole rats, rhesus macaques, mosquitofish and side-blotched lizards (Myers and Krebs, 1971; Mehlman et al., 1995; Cote et al., 2010b; Aguillon and Duckworth, 2015, see also Cote et al., 2010a for review).

Our work also paves the way for more in-depth understanding of uninvadability and evolutionary branching in subdivided populations. A difficulty in studying uninvadability in subdivided population stems from the necessity to calculate the effects that selection on traits has on pairwise relatedness. In this endeavour, our work is helpful in three ways. First, we have explicitly calculated the effects of selection on relatedness for the Moran process, therefore opening the door to studying uninvadability under this standard model of reproduction. Second, we have shown that the leading effects of selection on relatedness are zero at a singular phenotype when the evolving traits only affect adult fertility or offspring survival before dispersal under the Moran process, thus facilitating stability analysis for this case. Third, and more generally, the effect of selection on relatedness has no bearings on uninvadability when traits influence material payoffs that in turn weakly affect fitness (i.e., weak selection), and this holds for any demographical model that fits our general assumptions (see "Main assumptions" section). As illustrated by our study of the joint evolution of helping and dispersal, this method is useful to reach analytical results. In addition, we were able to relate the uninvadability of traits directly to the effects of traits on payoff and to the level of local competition when fitness effects are weak. Since relatedness and local competition have antagonistic effects on the evolution of social traits, our decomposition is useful to understand the forces at play for the uninvadability of social traits. We expect that our method to study uninvadability under weak selection should be particularly useful to understand adaptation in multiple strategies within complex, multi-strategies, games in the presence of genetic structure. Despite our progress, calculating the effect of selection on pairwise relatedness for arbitrary selection and demographical models under

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

our general assumptions remains challenging, and explicitly characterising uninvadability may require extensive additionalcomputations depending on model choice.

In order to reach tractable results, we have made a number of assumptions. Many of them – infinite population size, clonal reproduction of haploid genomes, rare mutation with weak effects – are common to those of the adaptive dynamics framework, and have been extensively discussed elsewhere (e.g., Geritz and Kisdi, 2000; Geritz and Gyllenberg, 2005; Champagnat et al., 2006; Dercole and Rinaldi, 2008). We have also assumed that the population is subdivided according to Wright (1931)'s classic infinite island model. This model ignores important biological realities like isolation-by-distance or temporal and spatial heterogeneity in environment. By changing genetic structure in space, isolation-by-distance makes selection on social traits more complicated (Rousset, 2004). Yet, most relevant insights on how selection moulds social traits are revealed by the infinite island model (like for cooperation and altruism, Lehmann and Rousset, 2010), and we therefore expect results qualitatively similar to ours under isolation-by-distance, at least when patches are homogeneous. It would be interesting to extend our framework to study the effects temporal and spatial heterogeneity in environment on the uninvadability of traits, in particular when traits can themselves influence the environment locally (Rousset and Ronce, 2004). As shown by simulations, temporal heterogeneity in patch sizes due to dispersal patterns may be important for the synergy among helping and dispersal (Parvinen, 2013). More generally, in order to study age-, sex- or caste-specific effects, or demographic stochasticity, it is necessary to incorporate class-structure (i.e., when the effect of a trait on the fitness of an individual is contingent on the class the individual belongs to). Like the vast majority of stability analyses of adaptive dynamics, our model falls short of predicting the course of evolution once evolutionary branching has occurred. This would require describing the post-branching phenotypic distribution in the population (as in Sasaki and Dieckmann, 2011).

In conclusion, we have provided a framework to study the evolutionary stability of intertwined traits in subdivided populations. In particular, with uninvadability expressed in terms of relatedness coefficients and indirect fitness effects, our model sheds light onto the consequences of population structure for evolutionary stability, diversification and the evolution of correlations among traits.

Literature cited

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

Aguillon, S. M. and R. a. Duckworth. 2015. Kin aggression and resource availability influence phenotype-dependent dispersal in a passerine bird. *Behav. Ecol. Sociobiol.* 69:625–633.

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

Akcay, E. and J. Van Cleve. 2012. Behavioral responses in structured populations pave the way to group optimality. *Am. Nat.* 179:257–269.

Akçay, E. and J. Van Cleve. 2016. There is no fitness but fitness, and the lineage is its bearer. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150085.

Ariyomo, T. O., M. Carter, and P. J. Watt. 2013. Heritability of boldness and aggressiveness in the zebrafish. *Behav. Genet.* 43:161–167.

Brännström, Å., N. Loeuille, M. Loreau, and U. Dieckmann. 2010. Emergence and maintenance of biodiversity in an evolutionary food-web model. *Theor. Ecol.* 4:467–478.

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. Biol. Sci.* 277:415–422.

Carter, A. J., A. W. Goldizen, and S. a. Tromp. 2010. Agamas exhibit behavioral syndromes: Bolder males bask and feed more but may suffer higher predation. *Behav. Ecol.* 21:655–661.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Caswell, H. 2001. Matrix population models: construction, analysis, and interpretation, vol. 2nd. Sinauer Associates, Inc.

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

Christiansen, F. B. 1991. On Conditions for Evolutionary Stability for a Continuously Varying Character. *Am. Nat.* 138:37.

Clobert, J., J. F. Le Galliard, J. Cote, S. Meylan, and M. Massot. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecol. Lett.* 12:197–209.

Cote, J., J. Clobert, T. Brodin, S. Fogarty, and a. Sih. 2010a. Personality-dependent dispersal: characterization, ontogeny and consequences for spatially structured populations. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365:4065–4076.

Cote, J., S. Fogarty, K. Weinersmith, T. Brodin, and A. Sih. 2010b. Personality traits and dispersal tendency in the invasive mosquitofish (Gambusia affinis). *Proc. Biol. Sci.* 277:1571–1579.

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

de Mazancourt, C. and U. Dieckmann. 2004. Trade Off Geometries and Frequency Dependent Selection. *Am. Nat.* 164:765–778.

Débarre, F., S. L. Nuismer, and M. Doebeli. 2014. Multidimensional (Co)Evolutionary Stability. *Am. Nat.* 184:158–171.

Dercole, F. and S. Rinaldi. 2008. Analysis of Evolutionary Processes: The Adaptive Dynamics Approach and Its Applications. Princeton University Press.

Dochtermann, N. A., T. Schwab, and A. Sih. 2015. The contribution of additive genetic variation to personality variation: heritability of personality. *Proc. Biol. Sci.* 282:20142201.

Doebeli, M., C. Hauert, and T. Killingback. 2004. The evolutionary origin of cooperators and defectors. *Science* 306:859–862.

Doebeli, M. and I. Ispolatov. 2010. Complexity and diversity. *Science* 328:494–7.

Drent, P. J., K. van Oers, and A. J. van Noordwijk. 2003. Realized heritability of personalities in the great tit (Parus major). *Proc. Biol. Sci.* 270:45–51.

Duckworth, R. A. and A. V. Badyaev. 2007. Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. *Proc. Natl. Acad. Sci. U. S. A.* 104:15017–22.

Duckworth, R. a. and L. E. B. Kruuk. 2009. Evolution of genetic integration between dispersal and colonization ability in a bird. *Evolution (N. Y).* 63:968–977.

Ducrest, A. L., L. Keller, and A. Roulin. 2008. Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends Ecol. Evol.* 23:502–510.

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

Edenbrow, M. and D. P. Croft. 2011. Behavioural types and life history strategies during ontogeny in the mangrove killifish, Kryptolebias marmoratus. *Anim. Behav.* 82:731–741.

El Mouden, C. and A. Gardner. 2008. Nice natives and mean migrants: the evolution of dispersal-dependent social behaviour in viscous populations. *J. Evol. Biol.* 21:1480–91.

Ellner, S. P. and M. Rees. 2006. Integral projection models for species with complex demography. *Am. Nat.* 167:410–428.

Eshel, I. 1972. On the Neighbor Effect and the Evolution. *Theor. Popul. Biol.* 277:258–277.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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

Eshel, I. 1996. On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *J. Math. Biol.* 34:485–510.

Ewens, W. 2004. Mathematical Population Genetics. Springer.

Ferriere, R., J. L. Bronstein, S. Rinaldi, R. Law, and M. Gauduchon. 2002. Cheating and the evolutionary stability of mutualisms. *Proc. Biol. Sci.* 269:773–80.

Frank, S. 1998. Foundations of social evolution. Princeton University Press.

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

Gandon, S. and F. Rousset. 1999. Evolution of stepping-stone dispersal rates. *Proc. Biol. Sci.* 266:2507–2513.

Gardner, A. and S. West. 2004. Cooperation and Punishment, Especially in Humans. *Am. Nat.* 164:753–764.

Geritz, S. A. H. and M. Gyllenberg. 2005. Seven answers from adaptive dynamics. *J. Evol. Biol.* 18:1174–1177.

Geritz, S. A. H. and E. Kisdi. 2000. Adaptive dynamics in diploid, sexual populations and the evolution of reproductive isolation. *Proc. R. Soc. B Biol. Sci.* 267:1671–1678.

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.

Gillespie, J. H. 1991. The Causes of Molecular Evolution. Oxford University Press.

Golub, G. H. and C. F. Van Loan. 1996. Matrix computations. JHU Press.

Grafen, A. 1991. Modelling in behavioural ecology. *Behav. Ecol. An Evol. Approach* pp. 5–31.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. *J. Theor. Biol.* 7:1–52.

Hamilton, W. D. and R. M. May. 1977. Dispersal in stable habitats. *Nature* 269:578–581.

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

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. 1985. Matrix Analysis, vol. 169. Cambridge University Press.

Ims, R. A. and R. A. Ims. 1990. Determinants of natal dispersal and space use in grey-sided voles, Clethrionomys rufocanus: a combined field and laboratory experiment. *Oikos* 57:106.

Ito, H. C. and U. Dieckmann. 2014. Evolutionary branching under slow directional evolution. *J. Theor. Biol.* 360:290–314.

Jean Clobert, M. Baguette, T. G. Benton, J. M. Bullock, S. Ducatez, J. Clobert, T. G. Benton, and J. M. Bullock. 2012. Dispersal Ecology and Evolution. Oxford University Press, USA.

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.

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

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

Laskowski, K. L. and A. M. Bell. 2014. Strong personalities, not social niches, drive individual differences in social behaviours in sticklebacks. *Anim. Behav.* 90:287–295.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Le Galliard, J.-F., R. Ferrière, and U. Dieckmann. 2005. Adaptive evolution of social traits: origin, trajectories, and correlations of altruism and mobility. *Am. Nat.* 165:206–224.

Lehmann, L., I. Alger, and J. Weibull. 2015. Does evolution lead to maximizing behavior? *Evolution (N. Y).* 69:1858–73.

Lehmann, L., C. Mullon, E. Akcay, and J. Van Cleve. 2016. Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. Tech. rep.

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

Lehmann, L. and F. Rousset. 2010. How life history and demography promote or inhibit the evolution of helping behaviours. *Philos. Trans. R. Soc. B Biol. Sci.* 365:2599–2617.

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.

Lessard, S. 1990. Evolutionary stability: One concept, several meanings. *Theor. Popul. Biol.* 37:159–170.

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

Massol, F., V. Calcagno, and J. Massol. 2009. The metapopulation fitness criterion: Proof and perspectives. *Theor. Popul. Biol.* 75:183–200.

Mehlman, P. T., J. D. Higley, I. Faucher, a. a. Lilly, D. M. Taub, J. Vickers, S. J. Suomi, and M. Linnoila. 1995. Correlation of CSF 5-HIAA concentration with sociality and the timing of emigration in free-ranging primates. *Am. J. Psychiatry* 152:907–913.

Metz, J. a. J. 2010. Thoughts on the geometry of meso-evolution collecting mathematical elements for a postmodern synthesis. In Chalub, F. A. C. C. and J. Rodrigues (eds.), *Math. Darwin's Leg.*, Mathematics and biosciences in interaction, p. in press. Birkh{ä}user, Basel.

Metz, J. A. J. and M. Gyllenberg. 2001. How should we define fitness in structured metapopulation models? including an application to the calculation of evolutionary stable dispersal strategies. *Proc. R. Soc. London Ser. B-Biological Sci.* 268:499—-508.

Mullon, C. and L. Lehmann. 2014. The robustness of the weak selection approximation for the evolution of altruism against strong selection. *J. Evol. Biol.* 27:2272–2282.

Myers, J. and C. Krebs. 1971. Genetic, behavioral, and reproductive attributes of dispersing field voles Microtus pennsylvanicus and Microtus ochrogaster. *Ecol. Monogr.* 41:53–78.

O'Riain, M. J., J. U. Jarvis, and C. G. Faulkes. 1996. A dispersive morph in the naked mole-rat. *Nature* 380:619–621.

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

Parvinen, K. 2013. Joint evolution of altruistic cooperation and dispersal in a metapopulation of small local populations. *Theor. Popul. Biol.* 85:12–19.

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

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

Pruitt, J. N., S. E. Riechert, and T. C. Jones. 2008. Behavioural syndromes and their fitness consequences in a socially polymorphic spider, Anelosimus studiosus. *Anim. Behav.* 76:871–879.

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.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Purcell, J., A. Brelsford, Y. Wurm, N. Perrin, and M. Chapuisat. 2014. Convergent Genetic Architecture Underlies Social Organization in Ants. *Curr. Biol.* 24:2728–2732.

Queller, D. C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* 8:70–73.

Réale, D., N. J. Dingemanse, A. J. N. Kazem, and J. Wright. 2010. Evolutionary and ecological approaches to the study of personality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365:3937–46.

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

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.

Rousset, F. and O. Ronce. 2004. Inclusive fitness for traits affecting metapopulation demography. *Theor. Popul. Biol.* 65:127–141.

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

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

Sih, A., A. Bell, and J. C. Johnson. 2004a. Behavioral syndromes: An ecological and evolutionary overview. *Trends Ecol. Evol.* 19:372–378.

Sih, A., A. M. Bell, J. C. Johnson, and R. E. Ziemba. 2004b. Behavioral Syndromes: An Integrative Overview. *Q. Rev. Biol.* 79:241–277.

Sinervo, B. and J. Clobert. 2003. Morphs, dispersal behavior, genetic similarity, and the evolution of cooperation. *Science* 300:1949–1951.

Sinn, D. L., L. a. Apiolaza, and N. a. Moltschaniwskyj. 2006. Heritability and fitness-related consequences of squid personality traits. *J. Evol. Biol.* 19:1437–1447.

Svardal, H., C. Rueffler, and M. Doebeli. 2014. Organismal Complexity and the Potential for Evolutionary Diversification. *Evolution (N. Y).* 68:3248–3259.

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.

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

Thieme, H. R. 2009. Spectral Bound and Reproduction Number for Infinite-Dimensional Population Structure and Time Heterogeneity. *SIAM J. Appl. Math.* 70:188–211.

Tuljapurkar, S., C. C. Horvitz, and J. B. Pascarella. 2003. The many growth rates and elasticities of populations in random environments. *Am. Nat.* 162:489–502.

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

van Doorn, G. S., U. Dieckmann, and F. J. Weissing. 2004. Sympatric speciation by sexual selection: a critical reevaluation. *Am. Nat.* 163:709–25.

van Oers, K., P. J. Drent, P. de Goede, and A. J. van Noordwijk. 2004. Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proc. Biol. Sci.* 271:65–73.

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

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Wang, J., Y. Wurm, M. Nipitwattanaphon, O. Riba-Grognuz, Y.-C. Huang, D. Shoemaker, and L. Keller. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* 493:664–8.

Wenseleers, T., a. Gardner, and K. R. Foster. 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.

Wild, G. 2011. Inclusive Fitness from Multitype Branching Processes. *Bull. Math. Biol.* 73:1028–1051.

Wolf, M., G. S. van Doorn, O. Leimar, and F. J. Weissing. 2007. Life-history trade-offs favour the evolution of animal personalities. *Nature* 447:581–584.

Wolfram Research, I. 2014. Mathematica. Wolfram Research, Inc., Champaign, Illinois, USA.

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

Appendix

A. Lineage fitness and uninvadability

We show here that a mutation causing the expression of phenotype $\boldsymbol{\zeta}$ that initially appears as a single copy in a resident population expressing $\mathbf{z}$ will vanish with probability one if, and only if, the lineage fitness of this mutation is less or equal to one ($v(\boldsymbol{\zeta},\mathbf{z}) \leq 1$, eq. 1).

Multitype branching process. We denote by $X_i(t)$ the random number of patches in the population with $i \in I = \{1,2,...,N\}$ mutants at time t, which are collected into the vector $\mathbf{X}(t) = (X_1(t),\ldots,X_N(t))$. We assume that the stochastic process $\{\mathbf{X}(t)\}_{t\geq 0}$ is a multi-type branching process (as in Wild, 2011), which is equivalent to assuming that when the mutant is globally rare, only residents can immigrate to a patch that already contains mutants. Under this assumption, the expected change in $\mathbf{X}(t)$ over one generation is governed by a matrix $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$, called the mean matrix, whose (i,j) element, denoted $a_{ij}(\boldsymbol{\zeta},\mathbf{z})$, gives the expected number of patches with $i \in I$ mutants that are generated by a patch when it has $j \in I$ mutants, and the population is otherwise monomorphic for $\mathbf{z}$,

$$\mathrm{E}[\mathbf{X}(t+1)|\mathbf{X}(t)] = \mathbf{A}(\boldsymbol{\zeta},\mathbf{z})\mathbf{X}(t), \tag{A-1}$$

where the expectation is taken over replicates of the one-generational demographic process (i.e., over one life cycle iteration).

It follows from standard results in stochastic demography that the average growth rate of the mutant population is the leading eigenvalue of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ (Caswell, 2001; Tuljapurkar et al., 2003). More precisely, the leading eigenvalue of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$, denoted $r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))$, gives the time-averaged mean cumulative growth over different replicates or sample paths of the invasion dynamics, i.e.,

$$\log\left(r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))\right) = \lim_{t\to\infty}\frac{1}{t}\log\left(\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\frac{L_j(t)}{L(0)}\right), \tag{A-2}$$

where $L_j(t)$ is the total random number of patches with at least one mutant at generation t for replicate population j (i.e., the sum of the $X_i(t)$'s for a given replicate j), n is the total number of replicate populations, and $L(0) = 1$ when the mutation arises as a single copy. Since the expected number of patches of each type grows asymptotically with $r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))$, the expected total mutant lineage size, $\mathrm{E}[\sum_{i=1}^{N} iX_i(t)]$, also grows asymptotically at rate $r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))$. Furthermore, the mutant lineage goes extinct at some time $t > 0$ with probability one (i.e., $\Pr\{\mathbf{X}(t) = \mathbf{0} \text{ for some } t \,|\mathbf{X}(0) = (1,0,\ldots,0)\} = 1$) if, and only if the leading eigenvalue of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ is less or equal to one (Karlin and Taylor, 1975; Harris, 1963, p. 41), i.e., if, and only if $r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})) \leq 1$.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Lineage fitness as the average growth rate of the mutation. We now show how the growth rate $r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))$ can be expressed as lineage fitness (eq. 1). Writing $\mathbf{u}(\boldsymbol{\zeta},\mathbf{z})$ as the right eigenvector of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ associated with $r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))$, we have by definition

$$r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))\mathbf{u}(\boldsymbol{\zeta},\mathbf{z}) = \mathbf{A}(\boldsymbol{\zeta},\mathbf{z})\mathbf{u}(\boldsymbol{\zeta},\mathbf{z}). \tag{A-3}$$

The right eigenvector is normalised such that its entries sum to one, $\sum_{j=1}^{N} u_j(\boldsymbol{\zeta},\mathbf{z}) = 1$. The eigenvector $\mathbf{u}(\boldsymbol{\zeta},\mathbf{z})$ is the quasi-stationary distribution of mutant patch types as it is invariant to multiplication by $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ (eq. A-3) (Harris, 1963, p. 44). Thus, $u_i(\boldsymbol{\zeta},\mathbf{z})$ can be interpreted as the frequency of patches with i mutants among patches with at least one mutant. Left multiplying eq. (A-3) by $\mathbf{x}^{\mathrm{T}} = (1,2,\ldots,N)$ gives

$$r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))\mathbf{x}^{\mathrm{T}}\mathbf{u}(\boldsymbol{\zeta},\mathbf{z}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{u}(\boldsymbol{\zeta},\mathbf{z}) = \sum_{i=1}^{N}\sum_{j=1}^{N} i a_{ij}(\boldsymbol{\zeta},\mathbf{z})u_j(\boldsymbol{\zeta},\mathbf{z}) \tag{A-4}$$

where $\mathbf{x}^{\mathrm{T}}\mathbf{u}(\boldsymbol{\zeta},\mathbf{z}) = \sum_{i\in I} i u_i(\boldsymbol{\zeta},\mathbf{z}) > 0$ is the average number of mutants among mutant patches (i.e., patches that contain at least one mutant). Then, note that the expected total number of mutants produced over one generation by the mutants residing in a patch with j mutants may be written in two ways,

$$\sum_{i=1}^{N} i a_{ij}(\boldsymbol{\zeta},\mathbf{z}) = w_j(\boldsymbol{\zeta},\mathbf{z})j, \tag{A-5}$$

where $w_j(\boldsymbol{\zeta},\mathbf{z})$ is the expected total number of adult offspring produced by a mutant over one iteration of the life cycle when there are j mutants in its patch and the remaining individuals in the population have phenotype $\mathbf{z}$. So,

$$r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} i a_{ij}(\boldsymbol{\zeta},\mathbf{z})u_j(\boldsymbol{\zeta},\mathbf{z})}{\sum_{i\in I} i u_i(\boldsymbol{\zeta},\mathbf{z})} = \frac{\sum_{j=1}^{N} j w_j(\boldsymbol{\zeta},\mathbf{z})u_j(\boldsymbol{\zeta},\mathbf{z})}{\sum_{i\in I} i u_i(\boldsymbol{\zeta},\mathbf{z})}. \tag{A-6}$$

The growth rate of the mutation can therefore be written in the form of lineage fitness

$$r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})) = \sum_{j=1}^{N} w_j(\boldsymbol{\zeta},\mathbf{z})q_j(\boldsymbol{\zeta},\mathbf{z}), \tag{A-7}$$

where

$$q_j(\boldsymbol{\zeta},\mathbf{z}) = \frac{j u_j(\boldsymbol{\zeta},\mathbf{z})}{\sum_{i\in I} i u_i(\boldsymbol{\zeta},\mathbf{z})}, \tag{A-8}$$

is the probability that a randomly sampled mutant from the mutant lineage belongs to a patch with j mutants.

Lineage fitness as proxy to the average growth rate and other proxies. The expression for $r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}))$ in eq. (A-7) depends on the eigenvectors of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$, which are difficult to evaluate in practice. In order to circumvent this problem, we seek a proxy whose sign around one is equivalent to that of the average growth rate, and that is also easier to evaluate. The usual fitness proxy in evolutionary biology is the basic reproductive number R_0 (Ellner and Rees, 2006) which will be the starting point to our derivation.

The basic reproductive number R_0 is obtained from the matrix describing mutant growth when rare, here $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$, by noting that it can be decomposed as

$$\mathbf{A}(\boldsymbol{\zeta},\mathbf{z}) = \mathbf{Q}(\boldsymbol{\zeta},\mathbf{z}) + \mathbf{B}(\boldsymbol{\zeta},\mathbf{z}), \tag{A-9}$$

where $\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})$ is a matrix whose element (i,j) gives the probability that the focal patch with $j \in I$ mutants turns into a patch with $i \in I$ mutants. By the properties of multi-type branching processes, the transition probabilities are independent of the state of the process $\{\mathbf{X}(t)\}$. Therefore, $\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})$ is the transient matrix of the Markov chain describing the subpopulation of mutants in the focal patch on the state space I, which has the local extinction of the mutant as its only absorbing state since only

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

residents immigrate into the patch. Meanwhile, the (i,j) entry of matrix $\mathbf{B}(\boldsymbol{\zeta},\mathbf{z})$ is the expected number of patches with i mutants that are produced by mutant emigration from the focal patch when the latter is in state j (i.e., with j mutants).

Then, the basic reproductive number $R_0(\boldsymbol{\zeta},\mathbf{z})$ is defined as the leading eigenvalue of the transition kernel giving lifetime reproduction (Caswell, 2001; Ellner and Rees, 2006), i.e., the leading eigenvalue of the matrix

$$\mathbf{R} = \mathbf{B}(\boldsymbol{\zeta},\mathbf{z})(\mathbf{I} - \mathbf{Q}(\boldsymbol{\zeta},\mathbf{z}))^{-1} \tag{A-10}$$

(Ellner and Rees, 2006). By construction of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ (eq. A-9) and the properties of $\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})$ and $\mathbf{B}(\boldsymbol{\zeta},\mathbf{z})$, the next-generation-theorem (Thieme, 2009, Theorem 1) implies the equivalence

$$r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})) \leq 1 \iff R_0(\boldsymbol{\zeta},\mathbf{z}) = r(\mathbf{R}(\boldsymbol{\zeta},\mathbf{z})) \leq 1, \tag{A-11}$$

which shows that $R_0(\boldsymbol{\zeta},\mathbf{z})$ is a proxy for the average growth rate, which can be computed by way of matrix inversion only.

We can further simplify the expression for $R_0(\boldsymbol{\zeta},\mathbf{z})$ when only a single mutant can establish into a resident patch by immigration. In that case, only the first row of $\mathbf{B}(\boldsymbol{\zeta},\mathbf{z})$ is non-zero, i.e.,

$$\mathbf{B}(\boldsymbol{\zeta},\mathbf{z}) = \begin{pmatrix} b_1(\boldsymbol{\zeta},\mathbf{z}) & b_2(\boldsymbol{\zeta},\mathbf{z}) & \dots & b_N(\boldsymbol{\zeta},\mathbf{z}) \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \tag{A-12}$$

where $b_j(\boldsymbol{\zeta},\mathbf{z})$ is the expected number of patches with one mutant that are produced by mutant emigration from the focal patch when in state j. The elements of the matrix $\mathbf{B}(\boldsymbol{\zeta},\mathbf{z})$ below its first line are zero since the probability that two or more offspring from the same patch settle in the same patch through dispersal is zero. Note that in the infinite island model of dispersal, only a single mutant can establish into a resident patch by immigration when offspring disperse independently from one another. This can be understood by noting that if the number of patches is N_d and migration probability of mutant offspring is m, then the probability that a given breeding spot on a given patch is settled through dispersal by an offspring from the focal patch is of the order $O(m/(N_\mathrm{d}))$, and the probability that two or more such offspring settle in the same patch is of order $O(m^2/(N_\mathrm{d})^2)$. Summing over all patches, the probability that two or more offspring from the same individual settle on the same patch through dispersal is at most of order $O(m^2/(N_\mathrm{d}))$, which goes to zero as $N_\mathrm{d} \to \infty$. Hence, the focal patch with j mutants can only turn a patch with zero mutants into a patch with a single mutant.

With $\mathbf{B}(\boldsymbol{\zeta},\mathbf{z})$ as in eq. (A-12), the matrix $\mathbf{R}(\boldsymbol{\zeta},\mathbf{z})$ is zero everywhere except in its first row. Since the eigenvalues of a triangular matrix are its diagonal entries, the only eigenvalue of $\mathbf{R}(\boldsymbol{\zeta},\mathbf{z})$ that may be greater than one is simply its first diagonal element (since all the other eigenvalues are zero). In this case, the leading eigenvalue of $\mathbf{R}(\boldsymbol{\zeta},\mathbf{z})$ (eqs. A-10-A-11) is

$$R_0(\boldsymbol{\zeta},\mathbf{z}) = \sum_{k=1}^{N} b_k(\boldsymbol{\zeta},\mathbf{z}) t_k(\boldsymbol{\zeta},\mathbf{z}), \tag{A-13}$$

where $t_k(\boldsymbol{\zeta},\mathbf{z})$ denotes the expected number of generations a patch that started with a single mutant spends with k mutants. Thus, when offspring disperse independently from one another, $R_0(\boldsymbol{\zeta},\mathbf{z})$ is equal to the expected number of successful emigrant mutants produced by a patch that started with a single mutant, which is the definition of the proxy $R_m(\boldsymbol{\zeta},\mathbf{z})$ by Metz and Gyllenberg (2001), and

$$R_0(\boldsymbol{\zeta},\mathbf{z}) = R_m(\boldsymbol{\zeta},\mathbf{z}) = \sum_{k=1}^{N} b_k(\boldsymbol{\zeta},\mathbf{z}) t_k(\boldsymbol{\zeta},\mathbf{z}) \leq 1 \iff r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})) \leq 1. \tag{A-14}$$

We now seek to obtain a fitness proxy expressed in terms of individual fitness, which will turn out to be of the form of lineage fitness (eq. A-7). To that end, note that the expected number of successful emigrants from a patch may be expressed as $b_k(\boldsymbol{\zeta},\mathbf{z}) = k e_k(\boldsymbol{\zeta},\mathbf{z})$, where $e_k(\boldsymbol{\zeta},\mathbf{z})$ is the expected number

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

of successful emigrants produced by a single mutant in a patch with k mutants when the rest of the population is monomorphic for $\mathbf{z}$. So, we find

$$\begin{aligned}\sum_{k=1}^{N} b_k(\boldsymbol{\zeta},\mathbf{z})t_k(\boldsymbol{\zeta},\mathbf{z}) \leq 1 &\iff \sum_{k=1}^{N} ke_k(\boldsymbol{\zeta},\mathbf{z})t_k(\boldsymbol{\zeta},\mathbf{z}) \leq 1 \\ &\iff \sum_{k=1}^{N} ke_k(\boldsymbol{\zeta},\mathbf{z})t_k(\boldsymbol{\zeta},\mathbf{z}) - 1 \leq 0.\end{aligned} \tag{A-15}$$

As shown in Mullon and Lehmann (2014), we can rewrite $\sum_{k=1}^{N} ke_k(\boldsymbol{\zeta},\mathbf{z})t_k(\boldsymbol{\zeta},\mathbf{z}) - 1$ in terms of the fitness of a mutant type $\boldsymbol{\zeta}$, $w_k(\boldsymbol{\zeta},\mathbf{z})$, which is the expected total number of adult offspring produced by a mutant, when there are k mutants in the patch and the remaining individuals in the population have phenotype $\mathbf{z}$:

$$\sum_{k=1}^{N} ke_k(\boldsymbol{\zeta},\mathbf{z})t_k(\boldsymbol{\zeta},\mathbf{z}) - 1 = \sum_{k=1}^{N} \left(w_k(\boldsymbol{\zeta},\mathbf{z}) - 1\right) kt_k(\boldsymbol{\zeta},\mathbf{z}). \tag{A-16}$$

So, condition (A-15) may be expressed as

$$\begin{aligned}\sum_{k=1}^{N} ke_k(\boldsymbol{\zeta},\mathbf{z})t_k(\boldsymbol{\zeta},\mathbf{z}) - 1 \leq 0 &\iff \sum_{k=1}^{N} \left(w_k(\boldsymbol{\zeta},\mathbf{z}) - 1\right) kt_k(\boldsymbol{\zeta},\mathbf{z}) \leq 0 \\ &\iff \bar{k}(\boldsymbol{\zeta},\mathbf{z})\left(\sum_{k=1}^{N} w_k(\boldsymbol{\zeta},\mathbf{z})\frac{kt_k(\boldsymbol{\zeta},\mathbf{z})}{\bar{k}(\boldsymbol{\zeta},\mathbf{z})} - 1\right) \leq 0\end{aligned} \tag{A-17}$$

where

$$\bar{k}(\boldsymbol{\zeta},\mathbf{z}) = \sum_{k} kt_k(\boldsymbol{\zeta},\mathbf{z}) > 0, \tag{A-18}$$

is the expected number of mutants present in a mutant patch. Now, since $\bar{k}(\boldsymbol{\zeta},\mathbf{z}) > 0$, (A-17) is equivalent to

$$\bar{k}(\boldsymbol{\zeta},\mathbf{z})\left(\sum_{k=1}^{N} w_k(\boldsymbol{\zeta},\mathbf{z})\frac{kt_k(\boldsymbol{\zeta},\mathbf{z})}{\bar{k}(\boldsymbol{\zeta},\mathbf{z})} - 1\right) \leq 0 \iff \sum_{k=1}^{N} w_k(\boldsymbol{\zeta},\mathbf{z})\frac{kt_k(\boldsymbol{\zeta},\mathbf{z})}{\bar{k}(\boldsymbol{\zeta},\mathbf{z})} \leq 1. \tag{A-19}$$

Then, note that

$$q_k(\boldsymbol{\zeta},\mathbf{z}) = \frac{kt_k(\boldsymbol{\zeta},\mathbf{z})}{\bar{k}(\boldsymbol{\zeta},\mathbf{z})}, \tag{A-20}$$

is a probability mass function which returns the probability that a randomly drawn member of the mutant local lineage resides in a patch with a total of $k \in \{1,2,...,N\}$ mutants. Therefore, a mutation coding for phenotype $\boldsymbol{\zeta}$ in a resident population with phenotype $\mathbf{z}$ will eventually go extinct with probability one if, and only if,

$$r(\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})) \leq 1 \iff \sum_{k=1}^{N} w_k(\boldsymbol{\zeta},\mathbf{z})q_k(\boldsymbol{\zeta},\mathbf{z}) \leq 1. \tag{A-21}$$

This shows that when only a single mutant can establish into a resident patch by immigration, the conditions that cause the leading eigenvalue of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ to be less or equal to one are the same as those that cause $v(\boldsymbol{\zeta},\mathbf{z})$ to be less or equal to one when $q_k(\boldsymbol{\zeta},\mathbf{z})$ is defined in terms of sojourn times (eq. A-20) rather than the eigenvector of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ (eq. A-8). While $q_k(\boldsymbol{\zeta},\mathbf{z})$ in terms of the eigenvector of $\mathbf{A}(\boldsymbol{\zeta},\mathbf{z})$ takes into account all possible patch composition starting with different numbers of mutant immigrants, $q_k(\boldsymbol{\zeta},\mathbf{z})$ in terms of sojourn times only considers a typical local lineage that starts with a single mutant. Lineage fitness $v(\boldsymbol{\zeta},\mathbf{z})$ with $q_k(\boldsymbol{\zeta},\mathbf{z})$ given by eq. (A-20) (i.e., lineage fitness as an invasion fitness proxy) is then sufficient to characterise uninvadability.

B. The eigenvectors of $\mathbf{H}(\mathbf{z})$ and the moulding of phenotypic correlations by selection

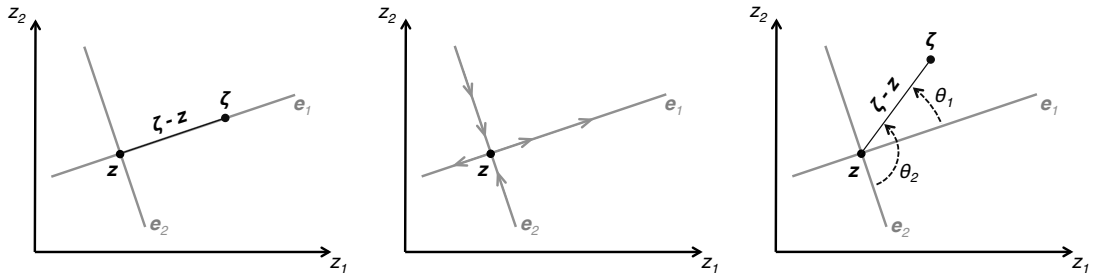
Here, we show that when a singular phenotype is invadable, the phenotypic correlations that are most likely to emerge are given by the eigenvector associated with the greatest positive eigenvalue of $\mathbf{H}(\mathbf{z})$. Each eigenvector $\mathbf{e}_l(\mathbf{z})$ for $l = 1,2,...,n$ of $\mathbf{H}(\mathbf{z})$ is a linear combination of the n traits. In addition, because $\mathbf{H}(\mathbf{z})$ is Hessian, these eigenvectors are perpendicular to one another (Horn and Johnson, 1985,

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

p.104), i.e., $\mathbf{e}_l(\mathbf{z}) \cdot \mathbf{e}_m(\mathbf{z}) = 0$ for $l \neq m$ ($\cdot$ refers to the dot product). They can therefore be represented as perpendicular lines in multi-dimensional phenotypic space (Supplemental Fig. 5a). The unique feature of each eigenvector $\mathbf{e}_l(\mathbf{z})$ is that, at a singular phenotype, both the strength and direction of selection along it is only determined by its associated eigenvalue $\lambda_l(\mathbf{z})$. To see this, consider a mutation that appears precisely along $\mathbf{e}_1(\mathbf{z})$ (Supplemental Fig. 5a). With each eigenvector scaled to have unit length ($\|\mathbf{e}_l(\mathbf{z})\| = \sqrt{\mathbf{e}_l(\mathbf{z}) \cdot \mathbf{e}_l(\mathbf{z})} = 1$), the vector $\boldsymbol{\zeta} - \mathbf{z}$ that connects the mutant $\boldsymbol{\zeta}$ to the resident $\mathbf{z}$ can be expressed as $\boldsymbol{\zeta} - \mathbf{z} = \|\boldsymbol{\zeta} - \mathbf{z}\|\mathbf{e}_1(\mathbf{z})$. From eq. (3) and the above identities, the lineage fitness of this mutation is

$$v(\boldsymbol{\zeta}, \mathbf{z}) = 1 + \frac{1}{2}\|\boldsymbol{\zeta} - \mathbf{z}\|^2 \mathbf{e}_1(\mathbf{z})^{\mathrm{T}} \mathbf{H}(\mathbf{z}) \mathbf{e}_1(\mathbf{z}) = 1 + \frac{1}{2}\|\boldsymbol{\zeta} - \mathbf{z}\|^2 \lambda_1(\mathbf{z}). \tag{B-1}$$

Since $\|\boldsymbol{\zeta} - \mathbf{z}\|^2 > 0$, whether eq. (B-1) is greater than one only depends on the sign of the eigenvalue $\lambda_1(\mathbf{z})$, and in addition, the magnitude of lineage fitness is directly related to the magnitude of $\lambda_1(\mathbf{z})$.



(a) Mutation appears along an eigenvector.

(b) Stabilising and diversifying selection along eigenvectors given by eigenvalues.

(c) Arbitrary mutation.

Figure 5: Selection close to a singular phenotype in two-traits space. The multi-dimensional phenotype consists of two traits, z_1 and z_2. The population is monomorphic for a singular phenotype $\mathbf{z}$ (black filled circle). The eigenvectors of the Hessian matrix, $\mathbf{e}_1$ and $\mathbf{e}_2$ (grey lines), are positioned to intersect at $\mathbf{z}$. (a) A mutation appears along the eigenvector $\mathbf{e}_1$, causing the expression of trait values $\boldsymbol{\zeta}$ (black filled circle). The vector that connects $\boldsymbol{\zeta}$ to $\mathbf{z}$ is shown in black. (b) An example of selection direction for an invadable singular phenotype is shown. A positive eigenvalue, $\lambda_1 > 0$, indicates that selection along its associated eigenvector $\mathbf{e}_1$ is diversifying, as shown by the outward arrows. In contrast, a negative eigenvalue, $\lambda_2 < 0$, tells us that selection along $\mathbf{e}_2$ is stabilising, as shown by the inward arrows. (c) Selection on mutations away from eigenvectors. The vector that connects $\boldsymbol{\zeta}$ to $\mathbf{z}$ is shown in black. The angles between this vector and both eigenvectors, θ_1 and θ_2, are represented by dashed curved arrows.

The eigenvectors of $\mathbf{H}(\mathbf{z})$, along with their associated eigenvalue, therefore provide a geometric representation of the direction and intensity of selection in the neighbourhood of a singular phenotype. The eigenvectors associated with negative eigenvalues give lines along which any mutation is counter-selected, whilst those associated with positive eigenvalues give lines along which selection favours mutation invasion (Supplemental Fig. 5b). In addition, the intensity of selection along an eigenvector, whether it is stabilising or diversifying, is reflected in the absolute value of its associated eigenvalue. This can be seen from the lineage fitness of a mutation that causes the expression of phenotype ζ, which can be written as a composite sum of the selection acting on all eigenvectors, each weighted according to the proximity of the mutation to the eigenvector,

$$v(\boldsymbol{\zeta}, \mathbf{z}) = 1 + \frac{1}{2}\|\boldsymbol{\zeta} - \mathbf{z}\|^2 \sum_{l=1}^{n} \lambda_l(\mathbf{z}) \left(\cos\theta_l(\boldsymbol{\zeta}, \mathbf{z})\right)^2 \tag{B-2}$$

(using the eigenvalue decomposition of $\mathbf{H}(\mathbf{z})$, e.g., Horn and Johnson, 1985, p. 104), where $\theta_l(\boldsymbol{\zeta}, \mathbf{z})$ is the angle between the eigenvector of $\mathbf{H}(\mathbf{z})$ that is associated with $\lambda_l(\mathbf{z})$, and the vector that connects the mutant phenotype $\boldsymbol{\zeta}$ to the resident $\mathbf{z}$ in phenotypic space (i.e., $\cos\theta_l(\boldsymbol{\zeta}, \mathbf{z}) = e_l(\mathbf{z}).(\boldsymbol{\zeta} - \mathbf{z})/\|\boldsymbol{\zeta} - \mathbf{z}\|$, see Supplemental Fig. 5c). The squared cosine of the angle $\theta_l(\boldsymbol{\zeta}, \mathbf{z})$ in eq. (B-2), $0 \leq (\cos\theta_l(\boldsymbol{\zeta}, \mathbf{z}))^2 \leq 1$, measures how closely the l^{th} eigenvector is aligned to the vector that connects the mutant phenotype $\boldsymbol{\zeta}$ to the resident $\mathbf{z}$: when it is zero, they are perpendicular; and when it is one, they are parallel.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Because all eigenvectors of a Hessian matrix are perpendicular to one another, the sum of weights is equal to one $\sum_{l=1}^{n}(\cos\theta_l(\boldsymbol{\zeta},\mathbf{z}))^2 = \sum_{l=1}^{n}(e_l(\mathbf{z}).(\boldsymbol{\zeta}-\mathbf{z})/\|\boldsymbol{\zeta}-\mathbf{z}\|)^2 = 1$. As a consequence, the maximum force that diversifying selection on a mutation can take, measured by $\sum_{l=1}^{n}\lambda_l(\mathbf{z})\left(\cos\theta_l(\boldsymbol{\zeta},\mathbf{z})\right)^2$, is the largest eigenvalue among the $\lambda_l(\mathbf{z})$'s, and the only mutations that will be affected by the maximum force of diversifying selection are those that appear on the eigenvector associated with the leading eigenvalue.

C. First-order effects of selection

We derive here eq. (12) of the main text. To do so, we first reformulate the fitness function of a mutant $w_k(\boldsymbol{\zeta},\mathbf{z})$, as a fitness function that depends explicitly on phenotypic values of all individuals in the population, and set $w_k(\boldsymbol{\zeta},\mathbf{z}) = w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})$ (see eq. 10). The first argument of this function is the phenotype of a focal individual whose fitness is under scrutiny, here a mutant with phenotype $\boldsymbol{\zeta}$. The second argument, $\tilde{\mathbf{z}}_{k-1}$, is the collection of $N-1$ phenotypes that express the neighbours of the focal, which is composed of $k-1$ other mutants and $N-k$ residents. The last argument is the resident phenotype $\mathbf{z}$ expressed in the rest of the population. Therefore, from eq. (1), we have

$$s_p(\mathbf{z}) = \frac{\partial v(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_p} = \sum_{k=1}^{N}\left(\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p}q_k(\mathbf{z},\mathbf{z}) + \frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_p}\right), \tag{C-1}$$

where all derivatives, here and throughout the entire appendix, are evaluated at the resident value $\mathbf{z}$. Then, because the total probability $\sum_{k=1}^{N}q_k(\boldsymbol{\zeta},\mathbf{z}) = 1$ is constant, we are left with

$$s_p(\mathbf{z}) = \sum_{k=1}^{N}\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p}q_k(\mathbf{z},\mathbf{z}). \tag{C-2}$$

The derivative of mutant fitness with respect to mutant phenotype can be expressed in terms of individual fitness derived with respect to individual and representative neighbouring phenotypes as follows

$$\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p} = \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}(k-1), \tag{C-3}$$

where $w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})$ is the same function as $w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})$ but now refers to the fitness of a focal individual, arbitrarily indexed as the individual $i\in\{1,\ldots,N\}$, which depends on its own phenotype $\mathbf{z}$, the collection of phenotypes of its $N-1$ neighbours $\mathbf{z}_{-i} = (\mathbf{z}_1,\ldots,\mathbf{z}_{i-1},\mathbf{z}_{i+1},\ldots,\mathbf{z}_N)$, as well as the resident phenotype $\mathbf{z}$ expressed in the rest of the population (e.g., Rousset, 2004). The first derivative in eq. (C-3) measures the effect of changing trait p in the focal i, and the second derivative measures the effect of changing trait p in a representative neighbour $j\neq i$ of the focal. Substituting eq. (C-3) into eq. (C-2), we have

$$\begin{aligned} s_p(\mathbf{z}) &= \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}\sum_{k=1}^{N}(k-1)q_k(\mathbf{z},\mathbf{z}) \\ &= \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + (N-1)\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}\sum_{k=1}^{N}\frac{(k-1)}{(N-1)}q_k(\mathbf{z},\mathbf{z}) \\ &= \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + (N-1)\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}r_2(\mathbf{z},\mathbf{z}), \end{aligned} \tag{C-4}$$

where the last line follows from the definition of $r_2(\mathbf{z},\mathbf{z})$ (eq. 11) and thus gives eq. (12) (see also Day, 2001, and Lehmann et al., 2015, for similar arguments).

D. Second-order effects of selection

We derive here eq. (13) of the main text. The derivative of $v(\boldsymbol{\zeta},\mathbf{z})$ with respect to ζ_p and ζ_q at $\boldsymbol{\zeta} = \mathbf{z}$ reads

$$\frac{\partial^2 v(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_p\partial\zeta_q} = \sum_{k=1}^{N}\left(\frac{\partial^2 w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p\partial\zeta_q}q_k(\mathbf{z},\mathbf{z}) + \frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p}\frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_q} + \frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_q}\frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_p}\right). \tag{D-1}$$

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Similarly to the first derivative (eq. C-3), the second derivative of mutant fitness can be expressed in terms of individual fitness derivatives as

$$\begin{aligned}\frac{\partial^2 w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p\partial\zeta_q} &= \frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}\partial z_{iq}} \\ &+\left(\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}\partial z_{jq}}+\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}\partial z_{jq}}+\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{iq}\partial z_{jp}}\right)(k-1) \\ &+\frac{\partial^2 w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}\partial z_{hq}}(k-1)(k-2).\end{aligned} \tag{D-2}$$

Substituting eq. (D-2) into

$$h_{\mathrm{w},pq}(\mathbf{z}) = \sum_{k=1}^{N}\frac{\partial^2 w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p\partial\zeta_q}q_k(\mathbf{z},\mathbf{z}), \tag{D-3}$$

and using the definition of relatedness eq. (11), yields eq. (13b).

Substituting eq. (C-3) into eq. (D-1) and using the fact that $\sum_{k=1}^{N} q_k(\boldsymbol{\zeta},\mathbf{z}) = 1$, the second term of eq. (D-1) can be written as

$$\begin{aligned}\sum_{k=1}^{N}\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p}\frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_q} &= \frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}\sum_{k=1}^{N}(k-1)\frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_q} \\ &= (N-1)\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}\frac{\partial}{\partial\zeta_q}\left(\sum_{k=1}^{N}\frac{k-1}{N-1}q_k(\boldsymbol{\zeta},\mathbf{z})\right).\end{aligned} \tag{D-4}$$

So, using the definition of $r_2(\boldsymbol{\zeta},\mathbf{z})$ (eq. 11), we have

$$\sum_{k=1}^{N}\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p}\frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_q} = (N-1)\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}\frac{\partial r_2(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_q}. \tag{D-5}$$

Similarly, the second term of eq. (D-1) is

$$\sum_{k=1}^{N}\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_q}\frac{\partial q_k(\boldsymbol{\zeta},z)}{\partial\zeta_p} = (N-1)\frac{\partial w(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jq}}\frac{\partial r_2(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_p}. \tag{D-6}$$

Substituting the last two equations into

$$h_{\mathrm{r},pq}(\mathbf{z}) = \sum_{k=1}^{N}\left(\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p}\frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_q}+\frac{\partial w(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_q}\frac{\partial q_k(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_p}\right) \tag{D-7}$$

gives eq. (13c).

E. First order effects on pairwise relatedness under a Moran process

Here, we calculate the first order effects of trait p on pairwise relatedness, $\partial r_2(\boldsymbol{\zeta},\mathbf{z})/\partial\zeta_p$ when the population follows a Moran life cycle. In this case, uninvadability can be characterised using the lineage fitness in the form of a proxy for invasion fitness. We therefore calculate $\partial r_2(\boldsymbol{\zeta},\mathbf{z})/\partial\zeta_p$ with the local mutant distribution $q_k(\boldsymbol{\zeta},\mathbf{z})$ given by eq. (A-20). Substituting, eq. (A-20) into eq. (11), relatedness between mutants can be expressed as

$$r_2(\boldsymbol{\zeta},\mathbf{z}) = \frac{\bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z})-\bar{k}(\boldsymbol{\zeta},\mathbf{z})}{(N-1)\bar{k}(\boldsymbol{\zeta},\mathbf{z})} = \frac{1}{(N-1)}\left(\frac{\bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z})}{\bar{k}(\boldsymbol{\zeta},\mathbf{z})}-1\right), \tag{E-1}$$

where

$$\bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z}) = \sum_{k}k^2 t_k(\boldsymbol{\zeta},\mathbf{z}). \tag{E-2}$$

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Taking the derivative of eq. (E-1) with respect to mutant effect ζ_p then reads

$$\frac{\partial r_2(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} = \frac{1}{(N-1)\bar{k}(\mathbf{z},\mathbf{z})}\left(\frac{\partial \bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} - \frac{\bar{\bar{k}}(\mathbf{z},\mathbf{z})}{\bar{k}(\mathbf{z},\mathbf{z})}\frac{\partial \bar{k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p}\right), \tag{E-3}$$

which depends on mean sojourn times under neutrality given that the patch started with a single mutant present ($t_k(\mathbf{z},\mathbf{z})$) and under selection ($t_k(\boldsymbol{\zeta},\mathbf{z})$). We derive those below for the Moran process.

The general mean sojourn times $t_{ij}(\boldsymbol{\zeta},\mathbf{z})$ spent with i mutants before absorption starting with j mutants (so that $t_k(\boldsymbol{\zeta},\mathbf{z}) = t_{k1}(\boldsymbol{\zeta},\mathbf{z})$) are obtained as the elements of

$$\mathbf{T}(\boldsymbol{\zeta},\mathbf{z}) = (\mathbf{I} - \mathbf{Q}(\boldsymbol{\zeta},\mathbf{z}))^{-1}, \tag{E-4}$$

where $\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})$ is as in eq. (A-9). For a Moran process (a birth and death process), the elements of the transient matrix $\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})$, i.e., the transient transition probability $q_{ij}(\boldsymbol{\zeta},\mathbf{z})$ from j to i, can be written as

$$q_{ij}(\boldsymbol{\zeta},\mathbf{z}) = \begin{cases} b_j(\boldsymbol{\zeta},\mathbf{z}), & \text{if } i = j+1 \text{ (“birth” of a mutant)},\\ d_j(\boldsymbol{\zeta},\mathbf{z}), & \text{if } i = j-1 \text{ (“death” of a mutant)},\\ 1 - b_j(\boldsymbol{\zeta},\mathbf{z}) - d_j(\boldsymbol{\zeta},\mathbf{z}), & \text{if } i = j \text{ (“no change”)},\\ 0, & \text{otherwise.}\end{cases} \tag{E-5}$$

In that case, eq. (E-4) reads

$$t_{ij}(\boldsymbol{\zeta},\mathbf{z}) = \begin{cases} \frac{1}{d_i(\boldsymbol{\zeta},\mathbf{z})}\left(1 + \sum_{h=1}^{i-1}\prod_{k=1}^{h}\frac{b_{i-k}(\boldsymbol{\zeta},\mathbf{z})}{d_{i-k}(\boldsymbol{\zeta},\mathbf{z})}\right) & \text{if } i = 1,2,...,j\\ t_{jj}\left(\prod_{k=j}^{i-1}\frac{b_k(\boldsymbol{\zeta},\mathbf{z})}{d_{k+1}(\boldsymbol{\zeta},\mathbf{z})}\right) & \text{if } i = j+1,2,...,N\end{cases} \tag{E-6}$$

(e.g., Ewens, 2004, eq. 2.160). In particular, the mean number of generations with i mutants starting with a single mutant ($j = 1$) is

$$t_i(\boldsymbol{\zeta},\mathbf{z}) = t_{i1}(\boldsymbol{\zeta},\mathbf{z}) = \frac{1}{d_1(\boldsymbol{\zeta},\mathbf{z})}\prod_{k=1}^{i-1}\frac{b_k(\boldsymbol{\zeta},\mathbf{z})}{d_{k+1}(\boldsymbol{\zeta},\mathbf{z})}. \tag{E-7}$$

For the life cycle described in the main text (section “Uninvadability under a Moran process”), the birth and death probabilities when k mutants are present are

$$b_k(\boldsymbol{\zeta},\mathbf{z}) = \frac{(N-k)\mu(\mathbf{z},\tilde{\mathbf{z}}_k,\mathbf{z})}{(N-k)\mu(\mathbf{z},\tilde{\mathbf{z}}_k,\mathbf{z}) + k\mu(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}\left(w_{\mathrm{p}}(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})\frac{k}{N}\right) \tag{E-8a}$$

$$d_k(\boldsymbol{\zeta},\mathbf{z}) = \frac{k\mu(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{(N-k)\mu(\mathbf{z},\tilde{\mathbf{z}}_k,\mathbf{z}) + k\mu(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}\left(1 - w_{\mathrm{p}}(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})\frac{k}{N}\right). \tag{E-8b}$$

They are explained as follows and use the notation given in Box 1 in the main text. For the birth of a mutant (eq. E-8a), two events must occur. First, a resident dies, with a probability that is given by the first term of eq. (E-8a). Secondly, the offspring who settles in the vacated breeding spot descends from a mutant in that patch, which occurs with a probability that is given by the second term of eq. (E-8a). Similarly, the death of a mutant (eq. E-8b) requires the death of a mutant and replacement by a resident.

The neutral sojourn times $t_k(\mathbf{z},\mathbf{z})$ are found using eq. (E-8) substituted into eq. (E-7) evaluated at $\boldsymbol{\zeta} = \mathbf{z}$. Doing so, we find

$$\begin{aligned}\bar{k}(\mathbf{z},\mathbf{z}) &= \sum_k k t_k(\mathbf{z},\mathbf{z}) = \frac{N}{m(\mathbf{z})}\\ \bar{\bar{k}}(\mathbf{z},\mathbf{z}) &= \sum_k k^2 t_k(\mathbf{z},\mathbf{z}) = \frac{N^2}{m(\mathbf{z})\left(1 + m(\mathbf{z})(N-1)\right)},\end{aligned} \tag{E-9}$$

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where $m(\mathbf{z}) = 1 - w_{\mathrm{p}}(\mathbf{z},\mathbf{z},\mathbf{z})$ is the backward migration rate.

The first order effects of trait p on pairwise relatedness (eq. E-3) also depends on the sojourn times under selection, $\bar{k}(\boldsymbol{\zeta},\mathbf{z})$ (eq. A-18) and $\bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z})$ (eq. E-2). These terms can be rewritten as the following matrix operation,

$$\begin{aligned}\bar{k}(\boldsymbol{\zeta},\mathbf{z}) &= \mathbf{u}\mathbf{T}(\boldsymbol{\zeta},\mathbf{z})\mathbf{n}_1\\ \bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z}) &= \mathbf{u}^2\mathbf{T}(\boldsymbol{\zeta},\mathbf{z})\mathbf{n}_1,\end{aligned} \tag{E-10}$$

where $\mathbf{u} = (1,2,\ldots,N)$, $\mathbf{u}^2 = (1,4,\ldots,N^2)$ and $\mathbf{n}_1 = (1,0,\ldots,0)^{\mathrm{T}}$. The first order effect of a trait on relatedness (eq. E-3) depends on the derivatives of eq. (E-10) with respect to mutant effect. Using eq. (E-4), and the fact that the derivative of the inverse of a matrix $\mathbf{A}$ is given by

$$\frac{\partial \mathbf{A}^{-1}}{\partial \zeta_p} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \zeta_p}\mathbf{A}^{-1}, \tag{E-11}$$

the derivatives of eq. (E-10) with respect to mutant effect ζ_p are

$$\begin{aligned}\frac{\partial \bar{k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \mathbf{u}\frac{\partial \mathbf{T}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p}\mathbf{n}_1 = \mathbf{u}\mathbf{T}(\mathbf{z},\mathbf{z})\frac{\partial}{\partial \zeta_p}\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})\mathbf{T}(\mathbf{z},\mathbf{z})\mathbf{n}_1\\ \frac{\partial \bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \mathbf{u}^2\frac{\partial \mathbf{T}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p}\mathbf{n}_1 = \mathbf{u}^2\mathbf{T}(\mathbf{z},\mathbf{z})\frac{\partial}{\partial \zeta_p}\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})\mathbf{T}(\mathbf{z},\mathbf{z})\mathbf{n}_1.\end{aligned} \tag{E-12}$$

Now, using eqs. (E-6)-(E-8) at neutrality ($\boldsymbol{\zeta} = \mathbf{z}$), some algebraic manipulations show that

$$\begin{aligned}\left(\mathbf{u}\mathbf{T}(\mathbf{z},\mathbf{z})\right)_j &= \sum_{i=1}^{N} i t_{ij}(\mathbf{z},\mathbf{z}) = \frac{N}{m(\mathbf{z})}j\\ \left(\mathbf{u}^2\mathbf{T}(\mathbf{z},\mathbf{z})\right)_j &= \sum_{i=1}^{N} i^2 t_{ij}(\mathbf{z},\mathbf{z}) = \frac{N^2}{m(\mathbf{z})\left(1+m(\mathbf{z})(N-1)\right)}\left[\left(1-\frac{m(\mathbf{z})}{2}\right)j + \frac{m(\mathbf{z})}{2}j^2\right],\end{aligned} \tag{E-13}$$

which in matrix form reads

$$\begin{aligned}\mathbf{u}\mathbf{T}(\mathbf{z},\mathbf{z}) &= \frac{N}{m(\mathbf{z})}\mathbf{u}\\ \mathbf{u}^2\mathbf{T}(\mathbf{z},\mathbf{z}) &= \frac{N^2}{m(\mathbf{z})\left(1+m(\mathbf{z})(N-1)\right)}\left[\left(1-\frac{m(\mathbf{z})}{2}\right)\mathbf{u} + \frac{m(\mathbf{z})}{2}\mathbf{u}^2\right].\end{aligned} \tag{E-14}$$

Substituting eq. (E-14) into eq. (E-12) then gives

$$\begin{aligned}\frac{\partial \bar{k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \frac{N}{m(\mathbf{z})}\frac{\partial}{\partial \zeta_p}\left(\mathbf{u}\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})\mathbf{T}(\mathbf{z},\mathbf{z})\mathbf{n}_1\right)\\ \frac{\partial \bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \frac{N^2}{m(\mathbf{z})\left(1+m(\mathbf{z})(N-1)\right)}\frac{\partial}{\partial \zeta_p}\left[\left(\left(1-\frac{m(\mathbf{z})}{2}\right)\mathbf{u} + \frac{m(\mathbf{z})}{2}\mathbf{u}^2\right)\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})\mathbf{T}(\mathbf{z},\mathbf{z})\mathbf{n}_1\right].\end{aligned} \tag{E-15}$$

Writing $\phi_{1,j}(\boldsymbol{\zeta},\mathbf{z})$ and $\phi_{2,j}(\boldsymbol{\zeta},\mathbf{z})$ as the first and second moments of the number of mutants in the focal patch after a generation conditional on there being j mutants before the transition, we have

$$\begin{aligned}\left(\mathbf{u}\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})\right)_j &= \sum_{i=1}^{N} i q_{ij}(\boldsymbol{\zeta},\mathbf{z}) = \phi_{1,j}(\boldsymbol{\zeta},\mathbf{z})\\ \left(\mathbf{u}^2\mathbf{Q}(\boldsymbol{\zeta},\mathbf{z})\right)_j &= \sum_{i=1}^{N} i^2 q_{ij}(\boldsymbol{\zeta},\mathbf{z}) = \phi_{2,j}(\boldsymbol{\zeta},\mathbf{z}),\end{aligned} \tag{E-16}$$

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and by definition,

$$\left(\mathbf{T}(\mathbf{z},\mathbf{z})\mathbf{n}_1\right)_j = t_j(\mathbf{z},\mathbf{z}). \tag{E-17}$$

Thus, eqs. (E-16)-(E-17) substituted into eq. (E-15) give

$$\begin{aligned}
\frac{\partial \bar{k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \frac{N}{m(\mathbf{z})}\sum_k \frac{\partial \phi_{1,k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} t_k(\mathbf{z},\mathbf{z}) \\
\frac{\partial \bar{\bar{k}}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \frac{N^2}{m(\mathbf{z})\left(1+m(\mathbf{z})(n-1)\right)} \\
&\quad \times \sum_k \left(\left(1-\frac{m(\mathbf{z})}{2}\right)\frac{\partial \phi_{1,k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} + \frac{m(\mathbf{z})}{2}\frac{\partial \phi_{2,k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p}\right) t_k(\mathbf{z},\mathbf{z}).
\end{aligned} \tag{E-18}$$

We now find an expression for the derivatives of $\phi_{1,k}(\boldsymbol{\zeta},\mathbf{z})$ and $\phi_{2,k}(\boldsymbol{\zeta},\mathbf{z})$. Conditional on k, the first and second moments of the number of mutants in the focal patch after a life cycle iteration are

$$\begin{aligned}
\phi_{1,k}(\boldsymbol{\zeta},\mathbf{z}) &= (k+1)b_k(\boldsymbol{\zeta},\mathbf{z}) + (k-1)d_k(\boldsymbol{\zeta},\mathbf{z}) + k\left(1-b_k(\boldsymbol{\zeta},\mathbf{z})-d_k(\boldsymbol{\zeta},\mathbf{z})\right) \\
\phi_{2,k}(\boldsymbol{\zeta},\mathbf{z}) &= (k+1)^2 b_k(\boldsymbol{\zeta},\mathbf{z}) + (k-1)^2 d_k(\boldsymbol{\zeta},\mathbf{z}) + k^2\left(1-b_k(\boldsymbol{\zeta},\mathbf{z})-d_k(\boldsymbol{\zeta},\mathbf{z})\right).
\end{aligned} \tag{E-19}$$

In order to get the derivatives of eq. (E-19) with respect to ζ_p, we substitute eq. (E-8), use the chain rule and the following identities,

$$\begin{aligned}
\frac{\partial \mu(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial \zeta_p} &= \frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + (k-1)\frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}} \\
\frac{\partial \mu(\mathbf{z},\tilde{\mathbf{z}}_k,\mathbf{z})}{\partial \zeta_p} &= k\frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}} \\
\frac{\partial w_{\mathrm{p}}(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial \zeta_p} &= \frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + (k-1)\frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}},
\end{aligned} \tag{E-20}$$

and eventually obtain

$$\begin{aligned}
\frac{\partial \phi_{1,k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \frac{k(N-k)}{N^2}\frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + \frac{k(k-1)(N-k)}{N^2}\frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}} \\
&\quad + \frac{k}{n}\frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} + \frac{k(k-1)}{N}\frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}} \\
\frac{\partial \phi_{2,k}(\boldsymbol{\zeta},\mathbf{z})}{\partial \zeta_p} &= \frac{k(N-k)(2k(N+1-m)-N)}{N^3}\frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} \\
&\quad + \frac{k(k-1)(N-k)(2k(N+1-m)-N)}{N^3}\frac{\partial \mu(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}} \\
&\quad + \frac{k(2k(N-1)+N)}{N^2}\frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}} \\
&\quad + \frac{k(k-1)(2k(N-1)+N)}{N^2}\frac{\partial w_{\mathrm{p}}(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}}.
\end{aligned} \tag{E-21}$$

Finally, we substitute eq. (E-21) into eq. (E-18), which is in turn substituted into eq. (E-3), and we use the first and second moments of sojourn times (eq. E-9), as well as the third and fourth moments,

$$\begin{aligned}
\sum_k k^3 t_k(\mathbf{z},\mathbf{z}) &= \frac{N^3}{m(\mathbf{z})}\frac{2-m(\mathbf{z})}{\left(1+m(\mathbf{z})(N-1)\right)\left(2+m(\mathbf{z})(N-2)\right)} \\
\sum_k k^4 t_k(\mathbf{z},\mathbf{z}) &= \frac{N^3}{m(\mathbf{z})}\frac{6N-m(\mathbf{z})\left(1-m(\mathbf{z})+N(6-m(\mathbf{z}))\right)}{\left(1+m(\mathbf{z})(N-1)\right)\left(2+m(\mathbf{z})(N-2)\right)\left(3+m(\mathbf{z})(N-3)\right)},
\end{aligned} \tag{E-22}$$

to obtain the result of eq. (14) and Table 2 of the main text.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The first order effects of a trait on pairwise relatedness that we have derived for the Moran process (eq. 14) can be written in a form similar to the one that has been used for the Wright Fisher process (Ajar, 2003; Wakano and Lehmann, 2014). In the absence of variation among death rates, $\mu(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = \mu$ for all i, eq. (14) with Table 2 can be expressed as

$$\frac{\partial r_2(\boldsymbol{\zeta}, \mathbf{z})}{\partial \zeta_p} = r_2^{\mathrm{R}}(\mathbf{z}, \mathbf{z}) \left(N r_2^{\mathrm{R}}(\mathbf{z}, \mathbf{z}) \frac{\partial w_{\mathrm{p}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}} + \left(N^2 r_3^{\mathrm{R}}(\mathbf{z}, \mathbf{z}) - N r_2^{\mathrm{R}}(\mathbf{z}, \mathbf{z}) \right) \frac{\partial w_{\mathrm{p}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}} \right), \tag{E-23}$$

where

$$\begin{aligned} r_2^{\mathrm{R}}(\mathbf{z}, \mathbf{z}) &= \frac{1}{1 + m(N-1)} \\ r_3^{\mathrm{R}}(\mathbf{z}, \mathbf{z}) &= \frac{2 - m}{(1 + m(N-1))\,(2 + m(N-2))}, \end{aligned} \tag{E-24}$$

are the probabilities of sampling two and three individuals with replacement that are identical-by-descent respectively. Eq. (E-23) bears close to resemblance to the parts of eq. (18) of Ajar (2003) and eq. (C.2) of Wakano and Lehmann (2014) that correspond to the first order effects of a trait on pairwise relatedness for the Wright Fisher process. Differences with eq. (E-23) are due to the fact that generations overlap in the Moran process, but not in the Wright Fisher process.

We now show that the first order effects of a trait on pairwise relatedness vanishes at singular phenotypes if the evolving traits only affect adult fertility or offspring survival. If that is the case, then the death rate $\mu(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = \mu$ is constant for all individuals in the population, and the philopatric component of fitness may be written as eq. (15). Substituting eq. (15) into eq. (14) and using Table 2, we find that

$$\frac{\partial r_2(\boldsymbol{\zeta}, \mathbf{z})}{\partial \zeta_p} = \frac{(1-m)N^2}{(2 + m(N-2))(1 + m(N-1))} s_p(\mathbf{z}) = 0, \tag{E-25}$$

since at a singular phenotype, $s_p(\mathbf{z}) = 0$.

F. Neutral pairwise and three-way relatedness for a Moran process

The stability criteria also depend on the probabilities of that two and three individuals are related in the resident population. These are found using standard identity-by-descent arguments (e.g., Karlin, 1968), by solving the recursions

$$\begin{aligned} r_2(\mathbf{z}, \mathbf{z}) &= \frac{N-2}{N} r_2(\mathbf{z}, \mathbf{z}) + \frac{2}{N}(1 - m(\mathbf{z})) \left(\frac{1}{N} + \frac{N-1}{N} r_2(\mathbf{z}, \mathbf{z}) \right) \\ r_3(\mathbf{z}, \mathbf{z}) &= \frac{N-3}{N} r_3(\mathbf{z}, \mathbf{z}) + \frac{3}{N}(1 - m(\mathbf{z})) \left(\frac{2}{N} + \frac{N-2}{N} r_2(\mathbf{z}, \mathbf{z}) \right), \end{aligned} \tag{F-1}$$

for a Moran process, whose solutions are given in Table 1.

G. Uninvadability under weak effects

In order to derive the selection gradient and Hessian matrix when payoffs have weak effects (eqs. 18 & 19), we follow the line of arguments developed in Lehmann et al., 2015 and first observe that under weak effects (i.e., $\epsilon \to 0$ in eq. 17), the fitness of a focal mutant in the focal patch with k mutants can be expressed as

$$\begin{aligned} w(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) \approx 1 + \epsilon \Bigg(& a_{\mathrm{f}}(\mathbf{z}) \left(\pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z}) \right) \\ & - a_{\mathrm{n}}(\mathbf{z}) \left(\frac{k-1}{N-1} \pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) + \frac{N-k}{N-1} \pi(\mathbf{z}, \tilde{\mathbf{z}}_k, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z}) \right) \Bigg), \end{aligned} \tag{G-1}$$

to the first order of ϵ (and where we used the zero-sum effect of selection on fitness in populations of

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

constant size, see p. 96 of Rousset, 2004, for explanation). The difference $\pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z})$ is the difference between the payoff received by the focal and the payoff of a resident from another patch, and $(k-1)/(N-1)\pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) + (N-k)/(N-1)\pi(\mathbf{z}, \tilde{\mathbf{z}}_k, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z})$ is the difference between the average payoff in the focal patch, excluding the focal individual, and the payoff of a resident from another patch. So, the difference between these two differences measures how well the focal does compared to the rest of her patch, relative to an individual from another patch. Focal fitness depends on the difference between these two, but where each is weighted by a coefficient, $a_{\mathrm{f}}(\mathbf{z})$ and $a_{\mathrm{n}}(\mathbf{z})$ respectively, that together measure the spatial scale of competition in the resident population (Frank, 1998).

The spatial scale of competition is given by $0 \leq a(\mathbf{z}) = a_{\mathrm{n}}(\mathbf{z})/a_{\mathrm{f}}(\mathbf{z}) \leq 1$ (since $0 < a_{\mathrm{f}}(\mathbf{z}) \leq 1$ and $a_{\mathrm{n}}(\mathbf{z}) \leq a_{\mathrm{f}}(\mathbf{z})$ according to our assumptions listed in the main text, see "Uninvadability under weak selection" section). To see how $a(\mathbf{z})$ measures the spatial scale of competition, we rewrite eq. (G-1) as

$$\begin{aligned} w(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) \approx 1 + \epsilon a_{\mathrm{f}}(\mathbf{z}) \Bigg(& (1 - a(\mathbf{z})) \left(\pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z})\right) \\ & + a(\mathbf{z}) \left(\frac{k-1}{N-1} \pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) - \frac{N-k}{N-1} \pi(\mathbf{z}, \tilde{\mathbf{z}}_k, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z}) \right) \Bigg), \end{aligned} \tag{G-2}$$

which shows that if $a(\mathbf{z})$ is small, then the payoff of neighbours has little effect on focal fitness compared to the payoff of individuals from other patches, i.e., competition tends to occur globally. This would occur, for instance, if dispersal is strong. Conversely, if $a(\mathbf{z})$ is close to one, then competition tends to occur locally. The coefficients $a_{\mathrm{f}}(\mathbf{z})$ and $a_{\mathrm{n}}(\mathbf{z})$, and therefore $a(\mathbf{z})$, are model specific, and will depend on the demographic properties of the population, such as patch size, dispersal rate, or the number of open breeding spots at each generation, as well as the evolving trait. The coefficients $a_{\mathrm{f}}(\mathbf{z})$ and $a_{\mathrm{n}}(\mathbf{z})$ are found by Taylor expanding the fitness function in the form of eq. (G-1).

Lineage fitness (eq. 1) is found by marginalising eq. (G-1) over the distribution of mutants $q_k(\boldsymbol{\zeta}, \mathbf{z})$. However, since ϵ is small and eq. (G-1) is of $O(\epsilon)$, it is sufficient to marginalise over the probability mass function for the number of mutants under neutrality, $q_k^\circ = q_k(\mathbf{z}, \mathbf{z})$,

$$q_k(\boldsymbol{\zeta}, \mathbf{z}) = q_k^\circ + O(\epsilon), \tag{G-3}$$

where q_k° is independent of the mutant type. Substituting eq. (G-1) and eq. (G-3) into lineage fitness eq. (1) gives to the order of $O(\epsilon)$

$$\begin{aligned} v(\boldsymbol{\zeta}, \mathbf{z}) = 1 + \epsilon a_{\mathrm{f}}(\mathbf{z}) \sum_{k=1}^{N} \Bigg(& \pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z}) \\ & - a(\mathbf{z}) \left(\frac{k-1}{N-1} \pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z}) + \frac{N-k}{N-1} \pi(\mathbf{z}, \tilde{\mathbf{z}}_k, \mathbf{z}) - \pi(\mathbf{z}, \mathbf{z}, \mathbf{z}) \right) \Bigg) q_k^\circ. \end{aligned} \tag{G-4}$$

First-order condition. Taking the derivative of eq. (G-4) with respect to ζ_p at $\boldsymbol{\zeta} = \mathbf{z}$ reads

$$\frac{\partial v(\boldsymbol{\zeta}, \mathbf{z})}{\partial \zeta_p} = \epsilon a_{\mathrm{f}}(\mathbf{z}) \sum_{k=1}^{N} \left[\frac{\partial \pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z})}{\partial \zeta_p} - a(\mathbf{z}) \left(\frac{k-1}{N-1} \frac{\partial \pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z})}{\partial \zeta_p} + \frac{N-k}{N-1} \frac{\partial \pi(\mathbf{z}, \tilde{\mathbf{z}}_k, \mathbf{z})}{\partial \zeta_p} \right) \right] q_k^\circ. \tag{G-5}$$

Then, after substituting for the derivatives,

$$\begin{aligned} \frac{\partial \pi(\boldsymbol{\zeta}, \tilde{\mathbf{z}}_{k-1}, \mathbf{z})}{\partial \zeta_p} &= \frac{\partial \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{ip}} + (k-1) \frac{\partial \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}} \\ \frac{\partial \pi(\mathbf{z}, \tilde{\mathbf{z}}_k, \mathbf{z})}{\partial \zeta_p} &= k \frac{\partial \pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\partial z_{jp}}, \end{aligned} \tag{G-6}$$

some re-arrangement, and using the definition of relatedness eq. (11), eq. (G-5) gives the result eq. (18).

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Second-order condition. Taking the derivative of eq. (G-4) with respect to ζ_p and ζ_q at $\boldsymbol{\zeta} = \mathbf{z}$ yields

$$\frac{\partial^2 v(\boldsymbol{\zeta},\mathbf{z})}{\partial\zeta_p\partial\zeta_q} = \epsilon a_{\mathrm{f}}(\mathbf{z})\sum_{k=1}^{N}\left[\frac{\partial^2\pi(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p\partial\zeta_q} - a(\mathbf{z})\left(\frac{k-1}{N-1}\frac{\partial^2\pi(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p\partial\zeta_q} + \frac{N-k}{N-1}\frac{\partial^2\pi(\mathbf{z},\tilde{\mathbf{z}}_{k},\mathbf{z})}{\partial\zeta_p\partial\zeta_q}\right)\right]q_k^{\circ}. \quad \text{(G-7)}$$

The derivatives can be expressed as

$$\begin{aligned}\frac{\partial^2\pi(\boldsymbol{\zeta},\tilde{\mathbf{z}}_{k-1},\mathbf{z})}{\partial\zeta_p\partial\zeta_q} &= \frac{\partial^2\pi(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}\partial z_{iq}} + (k-1)\left(\frac{\partial^2\pi(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}\partial z_{jq}} + \frac{\partial^2\pi(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{ip}\partial z_{jq}} + \frac{\partial^2\pi(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{iq}\partial z_{jp}}\right)\\ &\quad + (k-1)(k-2)\frac{\partial^2\pi(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}\partial z_{hq}} \qquad \text{(G-8)}\\ \frac{\partial^2\pi(\mathbf{z},\tilde{\mathbf{z}}_{k},\mathbf{z})}{\partial\zeta_p\partial\zeta_q} &= k\frac{\partial^2\pi(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}\partial z_{jq}} + k(k-1)\frac{\partial^2\pi(\mathbf{z}_i,\mathbf{z}_{-i},\mathbf{z})}{\partial z_{jp}\partial z_{hq}}.\end{aligned}$$

Substituting eq. (G-8) into eq. (G-7) and using the definitions of relatedness eqs. (11) and Table 1 generates the result eq. (19).

H. Agreement between weak and strong selection modes

Here, we show that the sign of $h_{11}(\mathbf{z})$ derived under weak selection (eq. 23) is the same as the sign of $h_{11}(\mathbf{z})$ for strong selection (equation not shown). In order to do so, we used Mathematica 10.0.1.0 (Wolfram Research, 2014) to generate random values for the parameters b_1, b_2, c_1 and c_2 between -100 and 100, for the patch size N between 2 and 1000 and for effective dispersal m between 0 and 1 (code available on request). Then, if these values generated admissible singular helping strategies eq. 22, (i.e., between 0 and 1), and admissible individual fertility at singular helping strategy eq. 21, (i.e., positive), then we compared the sign of $h_{11}(\mathbf{z})$ derived under the weak and strong selection modes. Out of 2.10^6 random sample, 465943 produced admissible singular strategies and individual fertility and, in all cases, the sign of $h_{11}(\mathbf{z})$ between weak and strong selection modes matched.

We show similarly that the sign of h_{11} given by eq. 23 and the sign of the leading eigenvalue of $\mathbf{H}(\mathbf{z})$ with $\epsilon = 1$ (equation not shown) are very often equivalent. Out of 2.10^6 random sample, 465359 produced admissible singular strategies and individual fertility and, in 464686 cases (99.86%), the sign of $h_{12}(\mathbf{z})$ given by eq. (32) matched the sign of $h_{12}(\mathbf{z})$ for strong selection and arbitrary patch size.

We also show that the sign of $h_{12}(\mathbf{z})$ derived under weak selection and large patch size (eq. 32) is very often the same as the sign of $h_{12}(\mathbf{z})$ for strong selection and arbitrary patch size (equation not shown). Out of 2.10^6 random sample, 465600 produced admissible singular strategies and individual fertility and, in 465589 cases (99.998%), the sign of $h_{12}(\mathbf{z})$ given by eq. (32) matched the sign of $h_{12}(\mathbf{z})$ for strong selection and arbitrary patch size. We remark here that $h_{12}(\mathbf{z})$ to the first order of selection ϵ but with arbitrary patch size is

$$\begin{aligned}h_{12}(\mathbf{z}) &= -\epsilon\frac{(2b_2c_1 - b_1c_2)s}{4b_2(N(1-s)+2-s) - 2c_2(N(1-s)+1)}\frac{2s(N(1-s)+s)^3}{N^2(N(1-s)+s)(2-s)} + O(\epsilon^2)\\ &= \epsilon\frac{(2b_2c_1 - b_1c_2)s}{4b_2(N(1-s)+2-s) - 2c_2(N(1-s)+1)}h_{22}(\mathbf{z}) + O(\epsilon^2).\end{aligned} \quad \text{(H-1)}$$

I. Individually-based simulations

We used Mathematica 10.0.1.0 (Wolfram Research, 2014) to simulate the joint evolution of helping and dispersal in a population with $N_p = 1000$ patches, each populated by $N = 8$ individuals (code available on request). Starting with a population monomorphic for singular helping and dispersal strategies, we track the evolution of the multidimensional phenotypic distribution as small mutations continuously arise. Each individual is characterised by a level of helping x_i and dispersal probability d_i. At the beginning of a generation, we calculate the fertility f_i of each individual according to its helping strategy and that of its neighbours (eq. 21 with $\epsilon = 1$). We use parameter values for the benefit and cost of helping ($b_1 = 6$, $b_2 = -1.4$, $c_1 = 4.56$, $c_2 = -1.6$) that are known to lead to evolutionary branching in infinite well-mixed

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

populations (Doebeli et al., 2004). Then, in each patch, an individual is chosen at random to die. We replace it by means of the weighted sampling of an individual in the population, where each individual is weighted according to whether they belong to the patch on which the breeding spot is filled or not. If an individual belongs to the same patch in which a breeding spot is filled, then its weight is $f_i(1-d_i)$. If it belongs to another patch, then its weight is $f_i d_i/(N_p-1)$. Once an individual is chosen to fill the breeding spot, its phenotypic values mutate with probability 0.01. If they do not mutate, then the offspring has the same phenotypic values as its parents. If they mutate, then we add small perturbations to the parental level of helping and dispersal probability that are sampled from a binormal distribution with mean $(0,0)$, variance 0.02^2 in each trait, and no covariance. The resulting phenotypic values are controlled to remain between 0 and 1. Once a breeding spot has been opened and filled in each patch, the generation is over and we repeat the iteration. For each values of $s \in \{0.04, 0.2, 0.5, 0.8, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96\}$, the population is initially monomorphic for the singular strategies and we simulate 1.5×10^5 generations.

To simulate the evolution of helping alone, the same procedure as above is used, except that survival in dispersal is fixed at $s = 1$ and dispersal is held constant at $m \in \{0.6, 0.68, 0.69, 0.71\}$ for each individual. In the case of a mutation, only helping is perturbed by an amount sampled from a Normal distribution with mean 0 and variance 0.02^2.

J. First order perturbation of the eigenvectors and eigenvalues of a matrix

Here, we derive eqs. 30 and 31 from the main text. To do this, we note that if we can express the Hessian matrix as $\mathbf{H}(\mathbf{z}) = \mathbf{H}^*(\mathbf{z}) + \epsilon\mathbf{H}^\bullet(\mathbf{z})$ where ϵ is small, and we label $\lambda_l^*(\mathbf{z})$ and $\mathbf{e}_l^*(\mathbf{z})$, the l^{th} eigenvalue of $\mathbf{H}^*(\mathbf{z})$ and its associated eigenvector respectively, then the corresponding eigenvalues and eigenvectors of $\mathbf{H}(\mathbf{z})$ are approximately

$$\begin{aligned}\lambda_l(\mathbf{z}) &\approx \lambda_l^*(\mathbf{z}) + \mathbf{e}_l^*(\mathbf{z})^{\mathrm{T}}\epsilon\mathbf{H}^\bullet(\mathbf{z})\mathbf{e}_l^*(\mathbf{z}) \\ \mathbf{e}_l(\mathbf{z}) &\approx \mathbf{e}_l^*(\mathbf{z}) + \sum_{k\neq l}\frac{\mathbf{e}_k^*(\mathbf{z})^{\mathrm{T}}\epsilon\mathbf{H}^\bullet(\mathbf{z})\mathbf{e}_l^*(\mathbf{z})}{\lambda_l^*(\mathbf{z}) - \lambda_k^*(\mathbf{z})}\mathbf{e}_k^*(\mathbf{z})\end{aligned} \tag{J-1}$$

(Golub and Van Loan, 1996, p. 323). In our example of the joint evolution of helping and dispersal, we find that substituting fitness (eq. I-a, Box 1) into eq. (13) together with the perturbation of relatedness eq. (14) and evaluating it at eq. (29), to the first order of ϵ the Hessian matrix can be decomposed as $\mathbf{H}(\mathbf{z}) = \mathbf{H}^*(\mathbf{z}) + \epsilon\mathbf{H}^\bullet(\mathbf{z})$ where

$$\mathbf{H}^*(\mathbf{z}) = \begin{pmatrix} 0 & 0 \\ 0 & -\dfrac{2s(N(1-s)+1)^3}{N^2(N(1-s)+s)(2-s)} \end{pmatrix} \tag{J-2}$$

and

$$\mathbf{H}^\bullet(\mathbf{z}) = \frac{s(N(1-s)+1)}{N^2(1-s)+Ns}\begin{pmatrix} -2c_2 + 2b_2\dfrac{(2-s+N(1-s))(4-3s)}{(N(1-s)+1)(2-s)} & \dfrac{s(1+N(1-s))(2b_2c_1-b_1c_2)}{N(2-s)\left(c_2 - \frac{2b_2(2-s+N(1-s))}{1+N(1-s)}\right)} \\ \dfrac{s(1+N(1-s))(2b_2c_1-b_1c_2)}{N(2-s)\left(c_2 - \frac{2b_2(2-s+N(1-s))}{1+N(1-s)}\right)} & 0 \end{pmatrix}. \tag{J-3}$$

The eigenvalues and eigenvectors of $\mathbf{H}^*(\mathbf{z})$ are then

$$\begin{aligned}\lambda_1^*(\mathbf{z}) &= 0, & \mathbf{e}_1^*(\mathbf{z}) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \lambda_2^*(\mathbf{z}) &= -\frac{2s(N(1-s)+1)^3}{N^2(N(1-s)+s)(2-s)}, & \mathbf{e}_2^*(\mathbf{z}) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}\end{aligned} \tag{J-4}$$

which substituted into eq. (J-1) give the approximate results of the main text (eqs. 30 and 31).

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

List of boxes

Box I. Individual fitness under a Moran process. A patch goes extinct with probability $e(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$. In non-extinct patches, individual i produces a large number $f(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$ of offspring. Then, exactly one adult individual dies on each non-extinct patch. Individual i has a death rate function $\mu(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$, so that it dies with probability $\mu(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) / \sum_{j=1}^{N} \mu(\mathbf{z}_j, \mathbf{z}_{-j}, \mathbf{z})$ in the current time step. Its offspring disperse independently from one another with probability $d(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$, during which they survive with probability $s(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})$. Then, the fitness of individual i is

$$w(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = (1 - e(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})) \times \left(1 - \frac{\mu(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{\sum_{j=1}^{N} \mu(\mathbf{z}_j, \mathbf{z}_{-j}, \mathbf{z})} + \frac{1}{N} \left(w_{\mathrm{p}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) + w_{\mathrm{d}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})\right)\right), \tag{I-a}$$

where the term on the first line is the probability that the patch of individual i does not go extinct. The first term on the second line is the probability that individual i survives to the next generation, while the second term is the expected number of offspring colonising vacant breeding spots, which is decomposed between those that remain in the philopatric patch $w_{\mathrm{p}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})/N$, and those that disperse and establish into other patches $w_{\mathrm{d}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})/N$. The philopatric component of fitness is given in terms of life-history traits by

$$w_{\mathrm{p}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = \frac{f(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})(1 - d(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}))}{\sum_{j=1}^{N} f(\mathbf{z}_j, \mathbf{z}_{-j}, \mathbf{z})(1 - d(\mathbf{z}_j, \mathbf{z}_{-j}, \mathbf{z}))/N + (1 - e(\mathbf{z}, \mathbf{z}, \mathbf{z})) f(\mathbf{z}, \mathbf{z}, \mathbf{z}) d(\mathbf{z}, \mathbf{z}, \mathbf{z}) s(\mathbf{z}, \mathbf{z}, \mathbf{z})}, \tag{I-b}$$

where the numerator is the expected number of offspring of individual i that stay in their natal patch and the denominator is the competition faced by an offspring of individual i for a philopatric breeding spot. The allopatric component of fitness is given by

$$\begin{aligned} w_{\mathrm{d}}(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) = {} & (1 - e(\mathbf{z}, \mathbf{z}, \mathbf{z})) \frac{f(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) d(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) s(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{f(\mathbf{z}, \mathbf{z}, \mathbf{z})(1 - d(\mathbf{z}, \mathbf{z}, \mathbf{z}) + (1 - e(\mathbf{z}, \mathbf{z}, \mathbf{z})) d(\mathbf{z}, \mathbf{z}, \mathbf{z}) s(\mathbf{z}, \mathbf{z}, \mathbf{z}))} \\ & + N e(\mathbf{z}, \mathbf{z}, \mathbf{z}) \frac{f(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) d(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z}) s(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})}{(1 - e(\mathbf{z}, \mathbf{z}, \mathbf{z})) f(\mathbf{z}, \mathbf{z}, \mathbf{z}) d(\mathbf{z}, \mathbf{z}, \mathbf{z}) s(\mathbf{z}, \mathbf{z}, \mathbf{z})}, \end{aligned} \tag{I-c}$$

which is the sum between the expected number offspring that colonise non-extinct patches (first line) and those that colonise extinct patches (second line). In non-extinct patches, offspring compete for a single breeding spot with offspring that from that patch, and those that come from other patches. In extinct patches, they compete for N breeding spots, which explains the factor N in the second summand, and only against offspring that come from other patches.

bioRxiv preprint doi: https://doi.org/10.1101/037887; this version posted January 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Box II. Scaled relatedness and competition. Each scaled relatedness term, $\kappa(\mathbf{z})$, $\iota(\mathbf{z})$, and $\eta(\mathbf{z})$, balances relatedness with local competition as,

$$\frac{1}{1-a(\mathbf{z})r_2}\Big(\text{relatedness} - (\text{direct competition} + \text{indirect competition})\Big), \tag{II-a}$$

where $0 \leq a(\mathbf{z}) \leq 1$ measures how much of the competition is between individuals of the same patch. When $a(\mathbf{z})$ is small, competition is mostly global, when $a(\mathbf{z})$ is close to 1, competition is mostly local. It is found according to the fitness function (Appendix G, eq. G-1). Scaled relatedness therefore discounts the effects of local competition from the effect of relatedness on selection. In

$$\kappa(\mathbf{z}) = \frac{1}{1-a(\mathbf{z})r_2}\left(r_2 - \frac{a(\mathbf{z})}{N-1}(1+(N-2)r_2)\right), \tag{II-b}$$

the factor 1 within $1+(N-2)r_2$ captures direct competition among two individuals within a patch because when the focal interacts with an individual j, it directly affects its payoff, and thereby directly affects the competition for its own offspring. E.g., if $\mathbf{z}$ is an altruistic trait, $\partial\pi(\mathbf{z}_i, \mathbf{z}_{-i}, \mathbf{z})/\partial z_{jp}$ is positive, so the focal increases the fertility of j and increases the local competition for its own offspring. The term $(N-2)r_2$ reflects that the neighbour j also affects the payoff of the other $(N-2)$ individuals of the patch, and if those are related to the focal (with probability r_2), this indirectly affects competition among related mutants. The term

$$\iota(\mathbf{z}) = \frac{1}{1-a(\mathbf{z})r_2}\left(r_2 - \frac{a(\mathbf{z})}{N-1}(r_2+(N-2)r_3)\right), \tag{II-c}$$

is a scaled measure of relatedness term among two individuals that incorporates the local competition arising from the interactions among two mutants. The term r_2 within $r_2+(N-2)r_3$ reflects the competition arising from the effects of the focal on the payoffs of j when they are related. The second term within the bracket reflects that j also changes the payoffs of the other individuals in the patch related to itself, and that in turn this affects the competition for the focal if they are all related to the focal, which occurs with probability r_3. Similarly, in

$$\eta(\mathbf{z}) = \frac{1}{1-a(\mathbf{z})r_2}\left(r_3 - \frac{a(\mathbf{z})}{(N-1)(N-2)}(2(N-2)r_2+(N-2)(N-3)r_3)\right), \tag{II-d}$$

which is a scaled measure of relatedness term among three individuals, the term $2(N-2)r_2$ is the effect of the competition that emerges from the focal interacting with individuals, and that interaction affects the payoffs of a third individual. Finally, the term $(N-2)(N-3)r_3$ is the indirect competition among related mutants that arises as consequence of neighbours related to the focal interacting with two other individuals related to the focal.