Evolution of quantitative traits: exploring the ecological, social and genetic bases of adaptive polymorphism

Laurent Lehmann* and Charles Mullon†

Abstract. Adaptive polymorphism in quantitative traits, characterised by the coexistence of genetically distinct phenotypes maintained by selection, is a significant contributor to biological diversity. Building on established theoretical results from invasion analysis, we analyse here conditions under which intraspecific adaptive polymorphism emerges and persists through long-term gradual evolution. We highlight that identifying adaptive polymorphisms of a quantitative trait is, in general, a computationally hard problem, but that invasion analyses can straightforwardly provide necessary conditions for both the evolutionary instability of monomorphic populations and the long-term stability of oligomorphic equilibria. We refine two core principles that underpin adaptive polymorphism across biological contexts: (i) the necessity of negative trait-dependent selection, whereby population trait expression reduces directional selection on the trait, for polymorphism to emerge; (ii) the necessity of qualitative dissimilarity among morphs in their fitness responses to the frequency of segregating morphs to ensure long-term coexistence. To clarify these conditions, we present a structured framework that distinguishes production factors (inputs to survival and reproduction that are endogenous to individuals) from regulating factors (genetic, social or ecological inputs to fitness that are influenced in part by conspecifics). We apply this decomposition to elucidate negative trait-dependent selection across both established and new scenarios of adaptive polymorphism across the three broad categories of ecologically, socially, and genetically mediated interactions. By integrating theoretical insights from population genetics, evolutionary ecology, adaptive dynamics, game theory, and inclusive fitness theory, we aim to provide a coherent synthesis of the multiple pathways leading to adaptive polymorphisms in quantitative traits.

Keywords: adaptive polymorphism, adaptive dynamics, environmental feedbacks, environmental fluctuations, life-history, preference evolution

*Department of Ecology and Evolution, University of Lausanne, Switzerland.
†Department of Ecology and Evolution, University of Lausanne, Switzerland.

1 Introduction

Polymorphism — the coexistence of multiple genetically determined phenotypes within a population — is widespread in nature, from discrete colour morphs in many flowering plants and animals (e.g. Gigord et al., 2001; Sinervo and Lively, 1996), to alternative mating strategies in vertebrates (e.g. Gross, 1996; Oliveira et al., 2008), sexual dimorphism in many species (e.g. Lande, 1980; Fairbairn, 2013), and allelic diversity at metabolic loci (e.g. Gillespie, 1991; Tishkoff et al., 2001). Unlike polyphenism, where distinct phenotypes are environmentally induced by the same genotype (West-Eberhard, 2003), polymorphism necessarily involves genetic variation, maintained by the segregation of different alleles. Although non-selective processes such as mutation, recombination, and genetic drift shape genetic variation (Crow and Kimura, 1970; Lynch and Walsh, 2018), natural selection is the only evolutionary force that can maintain cohesive and clearly differentiated phenotypes over evolutionary timescales. Here, we use "adaptive polymorphism" to refer to an intraspecific polymorphism that is underlain by the segregation of different alleles whose coexistence is favoured by natural selection and by natural selection alone (thus resulting in the stable maintenance of phenotypic variation in the absence of mutation). Understanding the conditions that favour adaptive polymorphism is important to identifying the mechanisms that promote and maintain the diversity of forms and functions so characteristic of the natural world.

Different theoretical frameworks, including population genetics, evolutionary ecology, and evolutionary game theory, have proposed a variety of mechanisms that favour intraspecific adaptive polymorphism. These include temporally or spatially varying selection (e.g. Levene, 1953; Gillespie, 1991; Svardal et al., 2015; Yamamichi et al., 2023), pleiotropy and trade-offs between competitive traits due to functional constraints (e.g. de Mazancourt and Dieckmann, 2004; Ito and Sasaki, 2016a), niche differentiation (e.g. Karlin and McGregor, 1972), strategic interactions producing frequency-dependent fitness effects in social contexts (e.g. Maynard Smith, 1982; McNamara and Leimar, 2020), and balancing selection via heterozygote advantage (e.g. Karlin and McGregor, 1972; Lewontin, 1974; Hartl and Clark, 2007; Sellis et al., 2011). This diversity of mechanisms can make it difficult, when analysing a specific biological scenario, to determine which intuition to apply or what general conditions, if any, govern the emergence and stability of adaptive polymorphism.

A common feature unifying these diverse models is that some form of negative intraspecific population feedback must act to favour diversity. However, the biological origin of this feedback differs across theoretical frameworks. Evolutionary ecology models often invoke environmental feedbacks, mediated by density-dependent resource depletion, co-adapted parasitism, or prey- predator interactions (e.g. Geritz et al., 1998; Pásztor et al., 2016). Game-theoretic models emphasise social feedback and frequency-dependent selection, where the payoff to behavioural strategies, such as aggression, cooperation, or signalling, depends on their frequency in the population (e.g. Hofbauer and Sigmund, 1998; McNamara and Leimar, 2020). Population genetics, meanwhile, highlights genetic feedbacks, notably overdominance in diploids, where allelic fitness depends directly on allele frequencies and mating patterns (e.g. Crow and Kimura, 1970; Karlin and McGregor,

1972; Ewens, 2004).

How do these various population feedbacks influence the emergence and long-term maintenance of adaptive polymorphism when new phenotypes regularly arise by mutation? We address this question in the context of quantitative traits evolving gradually through Darwinian adaptive dynamics, i.e. through the many successive small trait modifications. We build on the established theory of analysing the emergence of polymorphism under gradual evolution (e.g. Taylor, 1989; Christiansen, 1991; Metz et al., 1996; Eshel et al., 1997; Vincent et al., 1996; Geritz et al., 1998; Ajar, 2003; Rousset, 2004; Vincent and Brown, 2005; Otto and Day, 2007; Dercole and Rinaldi, 2008; Sasaki and Dieckmann, 2011; Brännström et al., 2013; Kisdi and Geritz, 2016; Mullon et al., 2016; Avila and Mullon, 2023), and adopt evolutionary invasion analysis or long-term evolution theory ("ESS theory", see Van Cleve, 2023, for a comprehensive review). This provides a unifying framework to study conditions under which adaptive polymorphisms initially emerge from ecological, social, and genetic interactions, as well as their eventual evolutionary endpoints.

Our study has five goals. The first is to formalize the concept of an adaptive polymorphism and highlight its useful relationship to the concept of Nash equilibrium. The second is to revisit the conditions under which monomorphic populations become evolutionarily unstable, and clarify when evolution converges towards stable oligomorphic outcomes (polymorphism consisting of a discrete number of morphs). In doing so, we highlight the important role of negative trait-dependent selection, which occurs when mutant fitness increases as mutant and resident trait values diverge. Negative trait-dependent selection is a key necessary condition for adaptive polymorphism to emerge in quantitative traits. Third, we identify the sources of the population feedback that generates this negative trait-dependent selection. We approach this using the concept of fitness factors, which links heritable trait values to survival and reproduction through intermediate layers such as phenotype and payoff, and integrates influences from the traits and behaviours of other conspecifics and the broader ecological context. Fourth, we leverage this decomposition into fitness factors to systematically analyse pathways to adaptive polymorphism across ecological, social, and genetic interactions. For each category of interaction, we specify conditions necessary for polymorphism to emerge, which allow us to understand the sources of trait-dependent selection across different biological scenarios (established and new). Finally, we relate the insights gained to classical ecological and population genetic principles of coexistence. Overall, our broader objectives are therefore to: identify general conditions that promote long-term adaptive polymorphism in quantitative traits, sharpen intuition about when adaptive diversity is expected to arise in natural populations, and to interweave several distinct threads of the literature.

2 Evolutionary model

2.1 Biological assumptions

We consider a large (ideally infinite) density-dependent regulated population. We start by assuming that individuals are haploid and reproduce asexually (this seemingly strong assumption is relaxed in section 6.1, where we consider diploidy and sexual reproduction). Individuals can belong to different classes (e.g. be of different ages or be in different physiological state) and can be distributed in space provided there is no isolation by distance (e.g. group or patch structure under the island model of dispersal, Wright, 1931; Rousset, 2004). The population may also experience environmental fluctuations in space and time as long as the biotic and abiotic processes driving these fluctuations are ergodic. We census this population at discrete demographic time steps between which individuals can survive, reproduce and interact.

Each individual expresses a quantitative or continuous trait; that is, a trait whose heritable value belongs to a subset $\mathcal{X} \subseteq \mathbb{R}$ of the real numbers. This trait can influence any aspect of an individual's morphology, physiology, or behavior. We assume that the trait evolves gradually via the constant influx of rare genetic mutations of small effects.

2.2 Adaptive polymorphism

2.2.1 Definition

We are interested in whether gradual evolution leads to an adaptive polymorphism in the trait. By an adaptive polymorphism, we mean different trait values or morphs that once in the population are both: (a) maintained in the population by natural selection alone (and not any other process such as mutation or recombination); and (b) protected from the invasion of any new genetic mutation by natural selection. To specify this notion, let us introduce two concepts.

First, we denote by p the probability density function of trait values in the population such that $p(x)$ is the probability density of trait $x \in \mathcal{X}$, i.e., $\int_{\mathcal{X}} p(x)\,\mathrm{d}x = 1$ where $p(x)\,\mathrm{d}x$ can be thought of as the frequency of individuals in the population with trait value in the interval $[x, x + \mathrm{d}x)$. Let $\mathcal{P}(p)$ stand for the support of the trait distribution p; namely, $\mathcal{P}(p)$ is the set of trait values that segregate in the population (formally $\mathcal{P}(p) = \{x \in \mathcal{X} : p(x) > 0\}$). If the population is monomorphic for some value x, then the support contains only this value, $\mathcal{P}(p) = \{x\}$. If two trait values x_1 and x_2 coexist, then the support contains these two values, $\mathcal{P}(p) = \{x_1, x_2\}$. More generally, the support of the trait distribution can be finite, consisting of a finite number of traits (i.e. an oligomorphism *sensu* Sasaki and Dieckmann, 2011), or infinite consisting of either a continuum of traits or an infinite number of discrete traits. A population will be said to be polymorphic

when the support $\mathcal{P}(p)$ contains two or more traits (or morphs). In a polymorphic population, individual reproduction and survival can depend on both the frequency distribution p and the segregating trait values $\mathcal{P}(p)$.

Second, consider a mutant type x introduced as a single copy in a resident population with trait distribution p and support $\mathcal{P}(p)$. Under our biological assumptions of section 2.1, the dynamics of the number of mutant copies descending from the initial mutant can be modelled as a multitype branching process (see Harris, 1963; Mode, 1971; Tanny, 1981 for general theory and Lehmann et al., 2016; Ohtsuki et al., 2020 for applications under our biological assumptions), whose defining feature is that a new copy of the process starts each time a newborn is produced (owing to the fact that mutant-mutant interactions can be neglected at the global population level). Accordingly, the mutant lineage goes extinct with probability one if and only if $\tilde{\rho}(x, \mathcal{P}(p), p) \leq 1$, where $\tilde{\rho}(x, \mathcal{P}(p), p)$ denotes the invasion fitness of mutant x in a resident population with trait distribution p and support $\mathcal{P}(p)$. Invasion fitness $\tilde{\rho}(x, \mathcal{P}(p), p)$ is the geometric growth ratio (or Malthusian parameter) of the lineage of mutants, i.e. it is the per capita number of mutant copies produced by individuals of the mutant lineage over one demographic time step, when the mutant reproductive process has reached stationarity in the resident population characterised by p and $\mathcal{P}(p)$. Invasion fitness is an appropriate representation of Darwinian fitness for large populations and the standard concept of fitness used in invasion analyses (e.g., Fisher, 1930; Eshel and Feldman, 1984; Altenberg and Feldman, 1987; Liberman, 1988; Lessard, 1990; Tuljapurkar, 1990; Metz et al., 1992; Charlesworth, 1994; Arnold et al., 1994; Rand et al., 1994; Ferrière and Gatto, 1995; Metz et al., 1996; Otto and Day, 2007; Lehmann et al., 2016; Altenberg et al., 2017; Van Cleve, 2023). Because the trait distribution p induces the support $\mathcal{P}(p)$, it is useful to use the following short-hand notation

$$\rho(x, p) = \tilde{\rho}(x, \mathcal{P}(p), p) \tag{1}$$

for the invasion fitness of mutant x.

The above allows us to define an adaptive polymorphism more formally as : a trait distribution p^* such that (i) its support $\mathcal{P}(p^*)$ contains two or more morphs; and (ii) p^* constitutes an uninvadable coalition of morphs that satisfies

$$\rho(x, p^*) = 1 \qquad \text{for all } x \in \mathcal{P}(p^*) \tag{2a}$$

$$\rho(x, p^*) \leq 1 \qquad \text{for all } x \in \mathcal{X} \setminus \mathcal{P}(p^*). \tag{2b}$$

Condition (2a) entails that each morph $x \in \mathcal{P}(p^*)$ in the uninvadable coalition exactly replaces itself through survival and reproduction, and is therefore maintained in the population. Condition (2b), meanwhile, ensures that any type $x \in \mathcal{X} \setminus \mathcal{P}(p^*)$ outside of the coalition goes extinct if introduced as a single copy into the population. Therefore, when both conditions (2a)–(2b) hold, the polymorphism can no longer be altered through natural selection on the introduction of alternative morphs by mutation. Hence, the polymorphism is an equilibrium and each morph in the coalition satisfies the biological concept of an adaptation: "a phenotypic

variant that results in the highest fitness among a specified set of variants in a given environment" (Reeve and Sherman, 1993, p. 9).

2.2.2 The challenge of characterising an adaptive polymorphism

Ideally, one would be able to characterize an adaptive polymorphism in a quantitative way for any given biological scenario; that is, one would be able to tell the number of morphs in the coalition, the traits they express, and their frequency in the population. Moreover one would like to show that such an equilibrium is evolutionarily attainable, i.e. is convergence stable (Eshel, 1983; Eshel et al., 1997; Leimar, 2009). However, working towards these general goals may prove challenging. In fact, we show in Box I that an uninvadable coalition is equivalent to the Nash equilibrium of a population game. This connection to game theory is interesting and provides us with two key insights about an adaptive polymorphism. First, an uninvadable coalition always exists under the assumptions that invasion fitness is continuous in p and $\mathcal{X}$ is a closed interval (Box I). Second, characterizing the uninvadable coalition – i.e. computing the Nash equilibrium – is a computationally hard problem. It thus follows that, on the positive side, a long-term adaptive equilibrium (monomorphic or polymorphic) should exist under most biological scenarios. On the negative side, however, characterising this equilibrium can be intractable—even leaving aside the question of the evolutionary convergence towards adaptive polymorphism.

To circumvent this difficulty, we will instead focus on two relevant segments of the evolutionary dynamics and investigate the conditions that are required for: (i) adaptive polymorphism to emerge from a population initially consisting of a single morph (in section 2.3); and (ii) a discrete number of morphs, i.e. an oligomorphism, to be evolutionarily maintained in the long term (in section 2.4). In other words, we aim to understand the conditions in which monomorphic population states are unstable under gradual evolution, since such instability gives rise to polymorphism, and to assess the long-term evolutionary stability of oligomorphism (our focus on oligomorphism rather than a continuum of morphs is motivated below).

2.3 Monomorphic instability

2.3.1 Gradual emergence of polymorphism

The mathematical approach to determining when a monomorphic population becomes polymorphic through gradual evolution is well-established (e.g. Taylor, 1989; Christiansen, 1991; Metz et al., 1996; Eshel et al., 1997; Geritz et al., 1998; Rousset, 2004; Otto and Day, 2007; Dercole and Rinaldi, 2008; Sasaki and Dieckmann, 2011; Brännström et al., 2013; Kisdi and Geritz, 2016; Mullon et al., 2016). It has been repeatedly and systematically confirmed by individual-based simulations across a wide range of scenarios, from simple models of local adaptation, anisogamy, and social interactions (e.g., Kisdi and Geritz, 1999; Maire et al., 2001;

Wakano and Iwasa, 2013) to more complex ones involving community assembly, non-equilibrium population dynamics, ecological inheritance, and immune diversity (e.g., Ito and Dieckmann, 2007; Nurmi and Parvinen, 2017; Prigent and Mullon, 2023; Siljestam and Rueffler, 2024). The basis to the approach is the invasion fitness $\rho(x, y)$ of a mutant with heritable trait value x in a resident population monomorphic for trait value y. Since the resident population is monomorphic here (i.e. $\mathcal{P}(p) = \{y\}$), we use y directly as the second argument in invasion fitness (instead of $\rho(x, p)$ as in eq. 1). We assume throughout that $\rho(x, y)$ is differentiable with respect to both of its arguments. Owing to density-dependent regulation, the invasion fitness of a single copy of a resident y must be one, $\rho(y, y) = 1$; that is, each gene copy on average replaces itself in a monomorphic population at its demographic attractor (this turns out to be a useful property for subsequent analyses).

Whether a population monomorphic for y is evolutionarily unstable and whether such instability favours polymorphism can be examined from the following three coefficients of selection:

$$S(y) = \left.\frac{\partial \rho(x, y)}{\partial x}\right|_{x=y}, \quad H(y) = \left.\frac{\partial^2 \rho(x, y)}{\partial x^2}\right|_{x=y} \text{ and } \quad I(y) = \left.\frac{\partial^2 \rho(x, y)}{\partial x \partial y}\right|_{x=y}, \tag{3}$$

which we refer to as the selection gradient (S), the selection Hessian (H), and the selection interaction coefficient (I), respectively. The selection gradient $S(y)$ indicates the direction of selection on the trait in a population at y. When $S(y) > 0$ ($S(y) < 0$), a mutation with small positive (negative) effect on trait value has a positive probability of invading, and if it does invade, it fixes and substitutes the resident (Eshel et al., 1997; Rousset, 2004; Cai and Geritz, 2020; Priklopil and Lehmann, 2021). Therefore, $S(y)$ is sufficient to characterise gradual evolution under the constant but rare input of mutations with small effects: the population will converge either to a boundary point of the state space $\mathcal{X}$, or to a trait value x^* lying in the interior of $\mathcal{X}$ such that

$$S(x^*) = 0 \tag{4}$$

$$J(x^*) = \left.\frac{\mathrm{d}S(y)}{\mathrm{d}y}\right|_{y=x^*} = H(x^*) + I(x^*) < 0, \tag{5}$$

(e.g. Eshel, 1983; Taylor, 1989; Eshel et al., 1997; Leimar, 2009), where $J(x^*)$ will be referred to as the selection Jacobian. A trait value x^* satisfying the condition in eq. (4) is known as a singular trait value, while those satisfying also the condition in eq. (5) are said to be convergence stable.

When all individuals in a population express a singular trait value x^*, $H(x^*)$ indicates whether selection is stabilizing or disruptive. When $H(x^*) < 0$, any change (increase or decrease) in trait value decreases invasion fitness, so that selection ousts any mutant whose value deviates locally from x^*, i.e. selection is stabilizing for x^*. In contrast, when $H(x^*) > 0$ any trait change increases fitness so that selection favours any mutant whose value deviates locally from x^*, i.e. selection is disruptive. Two conditions are therefore necessary for a monomorphic population to gradually evolve an adaptive polymorphism: $H(x^*) + I(x^*) < 0$ and $H(x^*) > 0$. When those are satisfied, the population will first converge to x^* and then become dimorphic (Eshel et al.,

1997; Geritz et al., 1998) in a process often referred to as evolutionary branching.

2.3.2 Negative-trait dependent selection (NTDS) is necessary for polymorphism

It follows from the considerations of section 2.3.1 that the gradual emergence of adaptive polymorphism requires that

$$I(x^*) < 0, \tag{6}$$

namely, that the selection interaction coefficient is negative at x^*. The selection interaction coefficient measures how directional selection on the trait $(\partial\rho(x, y)/\partial x)$ varies with a change in the trait value y expressed by the resident population (recall eq. 3). It can thus be thought of as a measure of "trait-dependent selection". Condition (6) says that negative trait-dependent selection (NTDS) is necessary for adaptive polymorphism to emerge by gradual evolution. When condition (6) holds, invasion fitness increases as a result of mutant and population trait values moving in opposite directions close to a singular point. This favours trait diversification because if individuals in the population do more of something (e.g. increase y), the mutant's best response is to do less (e.g. decrease x), thus favouring the existence of a polymorphic equilibrium. This is reflected in the condition (6) also being necessary for two types that are weakly diverged around a singularity x^* to mutually invade each other and thus to coexist in a protected polymorphism (Geritz et al., 1998, and see Appendix A.2 here). That is, condition (6) is necessary for negative-frequency dependent selection close to x^*, where the growth rate of each morph decreases as its frequency increases (Maynard Smith, 1998; Heino et al., 1998).

We will use the interaction coefficient $I(x^*)$ as a basis to identify conditions that favour the emergence of adaptive polymorphism. The importance of condition (6), beyond its biological interpretation and computational attractiveness (to be illustrated in examples below), also lies in the fact that once it is satisfied, one can in principle always construct a trade-off for the trait within the model such that $H(x^*) > 0$ holds (Kisdi, 2006, 2015). Selection then favours divergence of two types away from x^*, possibly leading to multiple diversification events later (e.g. Geritz et al., 1998; Ito and Dieckmann, 2007; Sasaki and Dieckmann, 2011; Svardal et al., 2015; Siljestam and Rueffler, 2024).

This then raises the question: where does the evolutionary process end up in the long-term? The computational hardness of determining an uninvadable coalition (Box 1) suggests that answering this question is difficult, especially because an uninvadable coalition may consist of an infinite number of morphs. Several results however suggest that an uninvadable coalition should consist of a finite number of morphs, that is, an oligomorphism. In particular, Hingu et al. (2008) show that the equilibria of the continuum replicator dynamics always consist of a finite number of coexisting morphs in panmictic populations without class structure. More generally, Gyllenberg and Meszéna (2005) show that the coexistence of an infinite number of morphs along a continuous range of trait values is not robustly stable: it is sensitive to changes in parameters (a qualitative result first observed in the context of polyphenism, Sasaki and Ellner, 1995). The

adaptive polymorphism (if any) to which gradual evolution converges in the long run is therefore likely an oligomorphism. We therefore turn to analysing the conditions under which an oligomorphism is evolutionary stable, i.e. convergence stable and uninvadable, in the next section.

2.4 Oligomorphic stability

2.4.1 Local uninvadable oligomorphism

To characterise the evolutionary stability of oligomorphism, consider a resident population with a discrete number n of segregating morphs. Let us now set $p = \mathbf{p} = (p_1, p_2, ..., p_n)$ to denote the joint probability mass function of these morphs in the population ($\sum_{i=1}^{n} p_i = 1$), with support $\mathcal{P}(\mathbf{p}) = \mathbf{y} = (y_1, y_2, ..., y_n)$. At an equilibrium of the dynamics of morph frequencies, each morph $y_i \in \mathcal{P}(\mathbf{p})$ must have invasion fitness equal to one so that the frequency vector $\mathbf{p}$ must satisfy:

$$\tilde{\rho}(y_i, \mathbf{y}, \mathbf{p}) = 1 \qquad \text{for all } y_i \in \mathcal{P}(\mathbf{p}). \tag{7}$$

Assume such an equilibrium exists and write its solution as $\mathbf{p}(\mathbf{y})$ to emphasise the dependency of $\mathbf{p}$ on $\mathbf{y}$. In this oligomorphic resident population, the invasion fitness of a mutant $x \in \mathcal{X}$ can then be written as

$$\rho(x, \mathbf{y}) = \tilde{\rho}(x, \mathbf{y}, \mathbf{p}(\mathbf{y})). \tag{8}$$

From this, we can define the selection gradient and selection Hessian for a mutant deviating from morph $i \in \{1, \ldots, n\}$ with trait y_i as

$$S_i(\mathbf{y}) = \left.\frac{\partial \rho(x, \mathbf{y})}{\partial x}\right|_{x=y_i} \qquad \text{and} \qquad H_i(\mathbf{y}) = \left.\frac{\partial^2 \rho(x, \mathbf{y})}{\partial x^2}\right|_{x=y_i}. \tag{9}$$

A sufficient condition for an oligomorphic resident population with support $\mathcal{P}(\mathbf{p}^*) = \mathbf{x}^*$ and equilibrium frequency $\mathbf{p}^*$ to be locally uninvadable with trait values in the interior of the state space is thus that

$$S_i(\mathbf{x}^*) = 0 \tag{10}$$

$$H_i(\mathbf{x}^*) < 0 \tag{11}$$

for all $i \in \{1, 2, ...n\}$. A necessary condition for local uninvadability is that eq. (11) holds with weak inequality. An oligomorphism satisfying only condition (10) is referred to as a singular coalition, and those satisfying both conditions (10) and (11) to as locally uninvadable coalitions. Conditions (10)–(11) correspond to those provided previously for an oligomorphism to be locally evolutionary stable in panmictic populations (Vincent et al., 1996, eq. 17, Geritz et al., 1998, eq. 15, Sasaki and Dieckmann, 2011, eq. 36 for the unstructured case and see Lion et al., 2022 for the structured case).

2.4.2 Convergence stable and hyperbolically stable oligomorphism

When conditions (10) and (11) hold, a coalition with morphs $\mathbf{x}^*$ and morphs frequencies $\mathbf{p}^*$, if established, will no longer change. But will such a pair ($\mathbf{x}^*$, $\mathbf{p}^*$) establish through gradual evolution; namely, is ($\mathbf{x}^*$, $\mathbf{p}^*$) an attractor of the evolutionary dynamics? Convergence stability here means that for any coalition $\mathbf{y} \neq \mathbf{x}^*$ in a small neighborhood of $\mathbf{x}^*$, the evolutionary dynamics should gradually converge to ($\mathbf{x}^*$, $\mathbf{p}^*$) by small step mutations. In a population of morphs $\mathbf{y} \neq \mathbf{x}^*$ within some neighborhood of the uninvadable coalition, the associated equilibrium morph frequencies $\mathbf{p} \neq \mathbf{p}^*$ will differ from that of the uninvadable coalition. If $\mathbf{x}^*$ is favoured in such a population, the frequency of morphs should eventually be attracted by $\mathbf{p}^*$, which requires that this is a local asymptotically stable fixed point of morph frequency dynamics. In addition, for such a polymorphic equilibrium to be biologically relevant, it must not depend on finely tuned parameters. It particular, it should retain its stability properties upon changes in parameters affecting the relationship between traits, morphs frequencies and fitness, otherwise "it has probability zero of occurring in the real word" (Meszéna et al., 2006, p. 70). We therefore require that an uninvadable coalition $\mathbf{x}^*$ is robustly convergence stable in the sense that it is a phenotypic attractor and that $\mathbf{p}^*$ is a hyperbolically stable fixed point of morph frequency dynamics. If $\mathbf{p}^*$ were not hyperbolically stable, then morph frequencies would either not converge towards the equilibrium $\mathbf{p}^*$ once $\mathbf{x}^*$ arises by mutation in a $\mathbf{y}$ population, or the equilibrium $\mathbf{p}^*$ would be destroyed by any small environmental change.

In Appendix A.2.1, we further detail the notion of hyperbolic stability of $\mathbf{p}^*$ and show that a necessary condition for it is that

$$\det \mathbf{J}^*_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0, \tag{12}$$

where matrix $\mathbf{J}^*_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$ has (i,j)-entry given by $\partial\tilde{\rho}(x_i^*, \mathbf{x}^*, \mathbf{p})/\partial p_j|_{\mathbf{p}=\mathbf{p}(\mathbf{x}^*)}$, which is the effect of a change in the frequency of morph j on the invasion fitness of morph i holding trait values constant, and where the vector $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}(y_1, \mathbf{y}, \mathbf{p}(\mathbf{y})), \tilde{\rho}(y_2, \mathbf{y}, \mathbf{p}(\mathbf{y})), ..., \tilde{\rho}(y_n, \mathbf{y}, \mathbf{p}(\mathbf{y})))$ collects the fitness of the different morphs (throughout $\mathbf{J}^*_{\mathbf{z}}(\mathbf{g})$ denotes the derivative of the vector valued function $\mathbf{g}$ with respect to the vector $\mathbf{z}$ with all quantities evaluated at $\mathbf{x}^*$). When condition (12) does not hold, this indicates that the invasion fitnesses of the n morphs are functionally dependent: the fitness of at least one morph can be written as a function of the others (by the Theorem of local functional dependence, Sydsaeter et al., 2008, p. 91) so that system (7) contains redundant equations. Biologically, this indicates that the invasion fitnesses of different morphs respond similarly to their frequency. Condition (12) can therefore be understood as requiring qualitative dissimilarity among coexisting morphs in how their fitness responds to intraspecific frequency driven changes. We return to this biological interpretation in section 3.3.

Condition (12) also provides a useful criterion for determining the candidate number n^* of coexisting morphs in an uninvadable coalition (i.e., the number of elements of $\mathbf{x}^*$ and $\mathbf{p}^*$). If for $n = n^* + 1$, $\det \mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = 0$, then the matrix is generically rank-deficient, and adding a row and column is unlikely to restore full rank (unless this new row and column have a very specific structure, see also Appendix A.2.4). In practice, the

smallest value of n such det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = 0$ therefore gives an upper bound for the maximum number of morphs that can stably coexist.

Hyperbolic stability of an uninvadable coalition guarantees that any of the finite number of fixed point of eq. (7) retains its stability properties upon changes of morph $\mathbf{y}$ values around a singularity $\mathbf{x}^*$ and ensures that the equilibrium $\mathbf{p}(\mathbf{y})$ of eq. (7) is a continuous function of $\mathbf{y}$ (from the implicit function theorem Sydsaeter et al., 2008, p.84 and Principle 3 of Karlin and McGregor, 1972). Then, in a population at $\mathbf{y}$ close to an uninvadable $\mathbf{x}^*$ with hyperbolically stable $\mathbf{p}^*$, the selection gradient $S_i(\mathbf{y})$ is sufficient to describe the direction of long-term evolution. Specifically, $S_i(\mathbf{y})$ determines whether a mutant $y_i + \delta$ arising from morph y_i with $|\delta| \ll 1$ in population state $\mathbf{y}$ will go extinct or substitute it ancestral type—leading to a new polymorphic population state $(y_1, y_{i-1}, y_i + \delta, y_{i+1}, ..., y_n)$ (Cai and Geritz, 2020; Priklopil and Lehmann, 2020, 2021). Since mutations occur independently in each morph, the rate of trait change of morph i is proportional to $S_i(\mathbf{y})$ scaled by some measure of phenotypic variance in the population, and the morphs together evolve in the direction described by the selection gradient vector $\mathbf{s}(\mathbf{y}) = (S_1(\mathbf{y}), S_2(\mathbf{y}), ..., S_n(\mathbf{y}))$ (Appendix A.2.2 for more details). Therefore, a necessary condition for a polymorphic singularity $\mathbf{x}^*$ satisfying $\mathbf{s}(\mathbf{x}^*) = \mathbf{0}$ to be robustly convergence stable–to be attracted towards $\mathbf{x}^*$ regardless of the scaling factor–is that the Jacobian matrix

$$\mathbf{J}^*_{\mathbf{y}}(\mathbf{s}) = \begin{pmatrix} \dfrac{\partial S_1(\mathbf{y})}{\partial y_1} & \cdots & \dfrac{\partial S_1(\mathbf{y})}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial S_n(\mathbf{y})}{\partial y_1} & \cdots & \dfrac{\partial S_n(\mathbf{y})}{\partial y_n} \end{pmatrix}_{\mathbf{y}=\mathbf{x}^*} \tag{13}$$

has all its eigenvalues with negative real parts. Sufficient conditions for convergence stability include (i) that $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ is a negative definite matrix (i.e., all eigvenvalues of the symmetric part of $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ are negative) and (ii) that $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ is a sign stable matrix, which has the attractive property that convergence stability can be predicated only upon qualitative properties of the signs of the entries of $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ (see Appendix A.2.3 for details).

A singular coalition that is locally uninvadable and robustly convergence stable will be called a locally evolutionary stable coalition. It is a local end-point of gradual evolution.

2.4.3 Saturated convergence stable oligomorphism

A biologically relevant way to look at convergence stability comes from observing that the Jacobian eq. (13) can be decomposed as

$$\mathbf{J}^*_{\mathbf{y}}(\mathbf{s}) = \mathbf{H}^* + \boldsymbol{I}^* \tag{14}$$

where $\mathbf{H}^*$ is a diagonal matrix with (i,i)-element $H_i(\mathbf{x}^*)$ capturing disruptive selection on morphs i (eq. 9), while matrix $\boldsymbol{I}^*$ has (i,j)-element given by $\partial^2\tilde{\rho}(x_i, \mathbf{y}, \mathbf{p}(\mathbf{y}))/\partial x_i \partial y_j\big|_{x_i=x_i^*, \mathbf{y}=\mathbf{x}^*}$, which is the effect of a change in the trait of morph j on selection in morph i. As such, matrix $\boldsymbol{I}^*$ captures trait-dependent selection in an

oligomorphic population (Appendix A.2.2 for derivation). This latter matrix can be further decomposed as,

$$\boldsymbol{I}^* = \boldsymbol{I}^*_{\mathbf{y}} - \boldsymbol{I}^*_{\mathbf{p}}\,\mathbf{M}^*, \tag{15}$$

where $\boldsymbol{I}^*_{\mathbf{y}}$ is a matrix with (i,j)-element $\partial^2\tilde{\rho}(x_i,\mathbf{y},\mathbf{p}(\mathbf{x}^*))/\partial x_i\partial y_j\big|_{x_i=x_i^*,\mathbf{y}=\mathbf{x}^*}$, which is trait-dependent selection on morph i stemming from varying resident trait of morph j, while holding morph frequencies constant; $\boldsymbol{I}^*_{\mathbf{p}}$ is a matrix with (i,j)-element $\partial^2\tilde{\rho}(x_i,\mathbf{x}^*,\mathbf{p})/\partial x_i\partial p_j\big|_{x_i=x_i^*,\mathbf{p}=\mathbf{p}(\mathbf{x}^*)}$, which is frequency-dependent selection on morph i stemming from varying the frequency of morph j, while holding trait values constant; and

$$\mathbf{M}^* = \mathbf{J}^*_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})^{-1}\mathbf{J}^*_{\mathbf{y}}(\tilde{\boldsymbol{\rho}}), \tag{16}$$

where matrix $\mathbf{J}^*_{\mathbf{y}}(\tilde{\boldsymbol{\rho}})$ has (i,j)-element $\partial\tilde{\rho}(x_i^*,\mathbf{y},\mathbf{p}(\mathbf{x}^*)/\partial y_j|_{\mathbf{y}=\mathbf{x}^*}$, which is the effect of a change in the trait of morph j on the fitness of morph i, holding morph frequencies constant. This matrix $\mathbf{M}^*$ can be interpreted as a conversion factor, measuring how much frequency-dependent effects on fitness can be substituted for trait dependent effects, holding fitness constant.

The negative sign in eq. (15) indicates that the presence of multiple morphs reduces the scope for further diversification via trait-dependent selection. This is because trait-dependent selection favours diversification with mutant fitness increasing as mutant and resident traits diverge. However, once several distinct morphs have evolved and coexist in a polymorphic equilibrium, the diversification is already partially realised. Consequently, the selective advantage of further divergence is diminished. The term $-\boldsymbol{I}^*_{\mathbf{p}}\,\mathbf{M}^*$ captures this effect, as it quantifies how frequency-dependent selection (represented by $\boldsymbol{I}^*_{\mathbf{p}}$) acts to offset the potential for further trait diversification (represented by $\boldsymbol{I}^*_{\mathbf{y}}$). In the limiting case where trait-dependent selection is fully offset by frequency dependence, namely when

$$\boldsymbol{I}^*_{\mathbf{y}} = \boldsymbol{I}^*_{\mathbf{p}}\,\mathbf{M}^* \tag{17}$$

the existing polymorphism leaves no further scope for diversification. Under these conditions, the criteria for convergence stability collapse to that of local uninvadability (together with the hyperbolic stability condition on morph frequencies). In such a scenario, we say the polymorphism is saturated, a concept we connect to existing literature in section 3.3.

3 Fitness factors: production, regulation, and their interactions

The analysis outlined in section 2 provides a recipe for assessing whether adaptive polymorphism can emerge under gradual evolution, as well as whether an uninvadable coalition is evolutionary stable. In particular, polymorphism requires a negative intraspecific feedback on fitness via the trait of the population (condition 6). To characterise such negative feedback and understand better the biological processes via which it can be generated, we need to be more explicit about what factors influence individual survival and reproduction –

"fitness factors" for short – and thus contribute to invasion fitness. From its definition given in section 2.3, invasion fitness $\rho(x, p)$ can be seen as the individual fitness of an "average" individual, i.e. randomly sampled from the lineage of individuals carrying trait value x, when the mutant lineage is rare in a resident population at its demographic attractor (Lehmann et al., 2016; Van Cleve, 2023). This average, which may be arithmetic, geometric, or a combination of the two depending on the biological situation, needs to be taken over all the relevant contexts in which the gene that code for x may find itself; that is, all types of genetic (e.g. genomic background), demographic (e.g. population size), ecological (e.g. prey density), and environmental states (e.g. temperature, see examples below and Box 2). Invasion fitness may thus be influenced by an untold number of factors. Yet, these can often be usefully grouped into different categories, as we next outline.

3.1 Limiting, regulating, and production factors

Fitness factors that have always been considered fundamental are those limiting survival and reproduction, i.e. limiting factors (Grovers, 1997). These fall into two types. On the one hand, limiting factors may be exogenous, such as temperature, sunlight, or humidity. On the other hand, limiting factors can be endogenous, meaning that their availability depends on conspecifics, conspecifics' abundance, and the traits they express. Examples include available space or nutrients. When limiting factors depend endogenously on the evolving population, they become special cases of a broader category usually referred to as "regulating factors" (e.g., May, 1973; Armstrong and McGehee, 1980; Grovers, 1997; Pásztor et al., 2016). Regulating factors are defined here as inputs to individual survival and reproduction that depend directly or indirectly on the traits expressed by *other* individuals in the population. Typically, regulating factors have negative effects on individual fitness—for instance, through competition for limited resources. However, regulating factors may sometimes enhance fitness, as occurs when mutualistic interactions improve resource availability. In this paper, regulating factors represent the first of two broad categories of fitness factors we explicitly consider.

The second category, which we refer to as "production factors", comprises inputs to individual fitness that depend, at least partially, on the trait x of the focal individual. Unlike regulating factors, which depend solely on others, production factors reflect intrinsic characteristics of the focal individual, such as baseline fecundity in the absence of competition, germination probability, energy absorption rates, or behavioral predisposition. As these examples suggest, production factors are typically associated with improved survival and/or reproduction. However, production outputs of one individual often affect the regulating factors experienced by others—creating indirect interactions among individuals. These interactions between regulating and production factors may combine in complex ways to shape overall individual reproduction and survival. To disentangle these effects and explore their evolutionary implications, we next introduce a model explicitly linking trait x to its invasion fitness through a structured hierarchy of regulating and production factors.

3.2 Invasion fitness in terms of regulating and production factors

3.2.1 Phenotype, payoff, and the fitness of an average individual

The first production factor we consider is the phenotype, defined broadly here as any physiological, morphological, or behavioural phenotype expressed by an individual (e.g. nutrient uptake rate, coat colour, effort into alloparental care). The phenotype expressed by an individual carrying trait x may depend not only on x but also on its direct or indirect interactions with others in the population. To capture this idea, we write the n_{z}-dimensional phenotype $\mathbf{z}$ of an individual as a function $\mathbf{z}(x, \mathbf{e}^{\mathrm{z}}(x,p)) \in \mathbb{R}^{n_{\mathrm{z}}}$ that depends on: (i) the heritable trait value x it carries (first argument); and (ii) a set of n_{z} regulating factors, which can be formalised as a function $\mathbf{e}^{\mathrm{z}}(x,p) \in \mathbb{R}^{n_{\mathrm{z}}}$ of mutant trait values x and resident distribution p. Here, and throughout, dependence on the resident distribution p implies that the regulating factors depend both on the frequency distribution of traits and on the trait values themselves (i.e. the support $\mathcal{P}(p)$).

For example, in a model of social interactions where individuals can respond to each other's actions (e.g. Akzcay and Van Cleve, 2012), the behaviour of a mutant carrier depends on the trait of its social partners (section 5.2 for examples). These interaction partners may be residents (under random interactions), or possibly other mutants (under preferential interactions among relatives or limited dispersal; section 6.2 for examples). Similarly, in diploid populations, this formulation allows for gene interactions within individuals on the phenotype, where mutant carriers are heterozygotes (carrying both x and some trait in the support $\mathcal{P}(p)$) under random mating but may be homozygotes (carrying two copies of x) under inbreeding (section 6.1 for more details on diploids). More generally, $\mathbf{z}(x, \mathbf{e}^{\mathrm{z}}(x,p))$ allows us to capture biotic and abiotic regulating factors that determine the expression of quantitative traits, including environmental effects, non-additive genetic effects, gene-by-environment effects, and indirect genetic effects (Lynch and Walsh, 2018).

The second production factor we consider explicitly is termed payoff, an umbrella term for all the material and informational inputs that are acquired by the individual and that serve for its survival or reproduction (e.g., number of mates, prey caught, quality of breeding spot, calories, knowledge). Using "payoff" in this broad sense has a long tradition in evolutionary biology (Maynard Smith, 1982; Stephens and Krebs, 1987; Houston and McNamara, 1999). We conceive the n_{π}-dimensional payoff of an individual as a function $\boldsymbol{\pi}(\mathbf{z}(x, \mathbf{e}^{\mathrm{z}}(x,p)), \mathbf{e}^{\pi}(x,p)) \in \mathbb{R}^{n_{\pi}}$ that depends on: (i) the phenotype $\mathbf{z}(x, \mathbf{e}^{\mathrm{z}}(x,p))$ it expresses (first argument); and (ii) a number n_{π} of regulating factors, given by a function $\mathbf{e}^{\pi}(x,p) \in \mathbb{R}^{n_{\pi}}$ of mutant trait values x and resident distribution p (and thus support $\mathcal{P}(p)$). In evolutionary games for instance, such regulating factors would include the behavioural phenotype of others that the individual interacts with, who may also be mutant carriers where interactions occurs in the family or in dispersal-limited groups, or whose behaviour depend on the individual's own trait under social responsiveness.

Finally, the translation of payoff into actual survival and reproduction may itself depend on further regulating factors, that we collect in $\mathbf{e}^{\mathrm{w}}(x,p) \in \mathbb{R}^{n_{\mathrm{w}}}$. These additional regulating factors include every biotic and

abiotic factors outside the individual that may influence its survival and reproduction and that are affected by conspecifics (holding the phenotype and payoff of the individual fixed). Examples include population density, the payoff of competitors, or prey density. In a well-mixed population, such regulating factors would typically depend on the resident traits $\mathcal{P}(p)$ and distribution p only, but under limited dispersal, they may also depend on the mutant x (e.g., Rousset and Ronce, 2004; Mullon and Lehmann, 2018). Putting it all together, we obtain that the invasion fitness of the mutant x in a resident population characterised by p can be written as

$$\rho(x,p) = W\big(\boldsymbol{\pi}(\mathbf{z}(x,\mathbf{e}^{\mathrm{z}}(x,p)),\mathbf{e}^{\pi}(x,p)),\mathbf{e}^{\mathrm{w}}(x,p)\big), \tag{18}$$

where the function W has as its first argument the payoff $\boldsymbol{\pi}$ and second argument the regulating factors $\mathbf{e}^{\mathrm{w}}(x,p)$ (see Fig. 1 for an illustration).

3.2.2 Intensive versus extensive regulating factors and their trait dependence

To connect to the literature, it is useful to recognise that a regulating factor can be one of two types. The first type a regulating factor can be is an extensive variable, which is a variable indicating the size of a system under focus (Kondepuli and Prigogine, 2015), such as population size, number of competitors, or total resource availability. In evolutionary ecology theory, such variables are often treated as state variables whose equilibrium values are determined by the demographic structure or dynamics of the resident population (e.g., Diekmann et al., 2003; Pásztor et al., 2016). We illustrate this notion in Box 2, where regulating factors are determined by resident equilibrium equations (see also Diekmann et al., 2003, eq. 2.1; Kisdi and Geritz, 2016, eq. 8).

Alternatively, a regulating factor can be an intensive variable, which is a variable that is not indicative of the size of the system (Kondepuli and Prigogine, 2015), i.e. variables that do not scale with system size. Instead, these reflect per capita quantities—such as average survival, average fecundity, or local competitive pressure. This perspective is particularly common in population and quantitative genetics, or evolutionary game theory models with constant population size, where fitness is defined in terms of relative success. This is the case for instance in the classical Wright-Fisher and Moran models of population genetics, where an individual's fitness is expressed directly as a function of the survival probability and/or fecundity of others (e.g. Maynard Smith, 1982; Gillespie, 2004; Hartl and Clark, 2007; Ewens, 2004; Lynch and Walsh, 2018).

Most classical approaches in ecology and evolutionary ecology treat regulating variables, whether they are extensive or intensive variables, as fixed by the resident population alone, i.e., assume that $\mathbf{e}(x,p) = (\mathbf{e}^{\mathrm{z}}(x,p),\mathbf{e}^{\pi}(x,p),\mathbf{e}^{\mathrm{w}}(x,p))$ is independent of x (such that $\mathbf{e}(x,p) = \mathbf{e}(p)$ e.g. Chesson, 2000a; Metz et al., 2008; Metz, 2011; Pásztor et al., 2016; Lion, 2018). But as we've alluded to above, this assumption holds only under panmixia and in the absence of direct social interactions. Any non-random interaction or interactions where individuals respond to each other, owing to e.g. limited dispersal, mate choice, behavioural feedback, or communication, entails that the factors regulating the fitness of a mutant trait x also depends on x.

This will be made clearer when we visit some examples with social responses (section 5.2) and non-random interactions (section 6.2).

3.3 Fitness factors, population (in)stability, and saturated polymorphisms

The expression of invasion fitness in terms of regulating factors (eq. 18) can already provide some general biological insights into the instability of monomorphism (section 2.3). Indeed, eq. (18) shows that mutant and resident traits can interact in diverse ways – in the sense that $I(y)$ as in eq. (3) is non-zero. Because fitness emerges from a hierarchy of nested interactions (trait $\rightarrow$ phenotype $\rightarrow$ payoff $\rightarrow$ fitness), x and y can influence $\rho(x, y)$ negatively through multiple biological pathways. In fact, NTDS, $I(y) < 0$, and thus monomorphic instability may well be the rule rather than an exception in natural populations where competition is so prevalent.

Eq. (18) also provides a useful basis to specify whether a stable oligomorphism is saturated or not (recall section 2.4). Consider a resident populations consisting of a finite number of morphs with distribution $p = \mathbf{p} = (p_1, p_2, ..., p_n)$ and support $\mathcal{P}(\mathbf{p})$. The impact of these different morphs on the invasion fitness of a mutant is then mediated by the collection of regulating variables $\mathbf{e}(x, p) = (\mathbf{e}^{\mathrm{z}}(x, p), \mathbf{e}^{\pi}(x, p), \mathbf{e}^{\mathrm{w}}(x, p))$. If this collection is independent of x, i.e. $\mathbf{e}(x, \mathbf{p}) = \mathbf{e}(\mathbf{p})$, then the competitive exclusion principle of ecology theory applies directly: the maximum number of morphs n^* that can be maintained cannot exceed the number of regulating factors, i.e.

$$n^* \leq n_{\mathrm{z}} + n_{\pi} + n_{\mathrm{w}}, \tag{19}$$

(Levin, 1970; Haigh and Maynard Smith, 1972; Meszéna et al., 2006; Pásztor et al., 2016). When the number of coexisting morphs in an uninvadable coalitions equals the number of regulating factors assumed to be independent, $n^* = n_{\mathrm{z}} + n_{\pi} + n_{\mathrm{w}}$, then the polymorphism is said to be saturated by (Kisdi and Geritz, 2016). They further show that trait-dependent selection vanishes entirely at such a polymorphism (specifically, $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s}) = \mathbf{H}^*$ so that eq. 17 holds). Hence, the concept of saturated polymorphism defined by eq. (17) is consistent with that of (Kisdi and Geritz, 2016) and further allows us to consider scenarios in which regulating factors may explicitly depend on the mutant trait, i.e. in which $\mathbf{e}(x, \mathbf{p})$ varies with x.

Beyond the specific case of saturated polymorphisms, eq. (18) combined with the condition (12) highlights the fundamental requirements multiple morphs to be maintained by selection (Appendix A.2.4 for analyses). Specifically, coexisting morphs must differ qualitatively in two respects: first, in how their fitness depends on regulating factors; second, in how their own frequencies influence these regulating factors. Taken together, these conditions imply a need for morph-specific feedbacks, meaning that each morph experiences a different population feedback when its frequency changes. Condition (12) thus ensures a sufficient degree of qualitative dissimilarity among morphs – dissimilarity both in their responses to regulating factors and in their frequency-dependent impacts on these factors. This condition is nothing else than the limiting similarity principle

(Meszéna et al., 2006, p. 78, Pásztor et al., 2016.

For the remainder of this paper, we explore various instances of eq. (18) from which NTDS emerges, opening the door to adaptive polymorphism. While several of these scenarios have been previously described, others are novel. Rather than providing a definitive classification, our aim is to illustrate the broad range of factors that can favour polymorphism in natural populations. The form of eq. (18) suggests three main biological pathways leading to NTDS: when interactions are (i) ecologically, (ii) socially, or (iii) genetically mediated. For each, we provide several examples and analyse the conditions under which monomorphism is unstable (as explained in section 2.3). In cases where selection leads to evolutionary branching, we examine the resulting oligomorphic stability (following the analysis detailed in section 2.4); we do so in the Appendix to keep the main text more fluent and focused.

4 Ecologically mediated interactions

4.1 Ecological setup

Many interactions among individuals are indirect, occurring through environmental change, e.g., through resource consumption or habitat modification. In such cases, trait dependence arises via ecological feedbacks, where changes in population trait values alter environmental conditions, which in turn impact individual reproduction and survival. To formalise this, we set $\boldsymbol{\pi}(\mathbf{z}(x, \mathbf{e}^{\mathrm{z}}(x, y)), \mathbf{e}^{\pi}(x, y)) = \mathbf{z}(x)$ and $\mathbf{e}^{\mathrm{w}}(x, y) = \mathbf{e}(y)$ for a monomorphic resident population (i.e. $p = y$) in eq. (18) so that an individual's payoff depends only on its own trait x, while the environment is determined solely by the resident trait y. Invasion fitness then simplifies to $\rho(x, y) = W(\mathbf{z}(x), \mathbf{e}(y))$. We refer to "ecologically mediated interactions" as cases where fitness takes this form, with interactions mediated through a regulating factor $\mathbf{e}(y)$ exogenous to the individual. These interactions can nevertheless involve complex ecological feedback loops, potentially spanning multiple environmental state variables (e.g., multitrophic consumer-resource models). In the ecological literature, the production factors $\mathbf{z}(x)$ are sometimes referred to as environmentally dependent parameters, while regulating factors $\mathbf{e}(y)$ are called competition parameters (Chesson, 2000a, p. 212-213).

For simplicity, we further assume that

$$\rho(x, y) = W(\underbrace{z_1(x), z_2(x)}_{=\mathbf{z}(x)}, \underbrace{e_1(y), e_2(y)}_{=\mathbf{e}(y)}), \tag{20}$$

so that the function W depends on only four arguments. This minimal set-up is sufficient to illustrate the standard pathways to trait-dependent selection under ecological interactions. The first two arguments, $z_1(x)$ and $z_2(x)$, are production factors, assumed to never decrease fitness such as fecundity or survival ($\partial W/\partial z_1 \geq 0$ and $\partial W/\partial z_2 \geq 0$). The second two, $e_1(y)$ and $e_2(y)$, are regulating factors, which are assumed

to only feedback negatively on fitness ($\partial W/\partial e_1 \leq 0$ and $\partial W/\partial e_2 \leq 0$). As discussed in section 3.2.2, the regulating factors can be conceptualized as intensive variables, in which case one can set $e_1(y) = z_1(y)$ and $e_2(y) = z_2(y)$, or as extensive variables, which satisfy an equilibrium system of equations as detailed in Box 2.

Substituting eq. (20) into eq. (4), shows that any singularity x^* must satisfy

$$\frac{\partial W}{\partial z_1}\frac{\mathrm{d}z_1(x)}{\mathrm{d}x}\bigg|_{x=y=x^*} + \frac{\partial W}{\partial z_2}\frac{\mathrm{d}z_2(x)}{\mathrm{d}x}\bigg|_{x=y=x^*} = 0. \tag{21}$$

(here and throughout, we drop the argument of functions with multiple arguments when the notation becomes too cumbersome and there is no risk of confusion). Eq. (21) shows that for an evolutionary equilibrium x^* to exist, there must be a trade-off between the fitness returns obtained via the two production factors (if the production factors are monotonic functions of the trait). Specifically, since fitness increases with both production factors ($\partial W/\partial z_1 \geq 0$ and $\partial W/\partial z_2 \geq 0$), an evolving trait cannot increase both factors indefinitely—otherwise, selection would continuously favour larger trait values. Instead, the trait must increase one production factor while reducing the other, with these opposing effects on fitness balanced at x^*.

Because invasion fitness in a monomorphic population must equal one, i.e. $W(z_1(y), z_2(y), e_1(y), e_2(y)) = 1$, we also have

$$\frac{\partial W}{\partial e_1}\frac{\mathrm{d}e_1(y)}{\mathrm{d}y}\bigg|_{y=x^*} + \frac{\partial W}{\partial e_2}\frac{\mathrm{d}e_2(y)}{\mathrm{d}y}\bigg|_{y=x^*} = 0 \tag{22}$$

(using eq. 21). Since $\partial W/\partial e_1 \leq 0$ and $\partial W/\partial e_2 \leq 0$, eq. (22) indicates that $\mathrm{d}e_1(y)/\,\mathrm{d}y$ and $\mathrm{d}e_2(y)/\,\mathrm{d}y$ must have opposite signs at $y = x^*$. In other words, a trait increase in the population must be associated with opposite effects on regulating variables 1 and 2. This result expresses a fundamental ecological constraint: due to density-dependent regulation, an increase in competition for one resource must be offset by a decrease in competition for another at equilibrium. More intuitively perhaps, if, as a result of their trait expression, individuals start competing more intensely for one factor—such as food, nesting sites, or mates—they compete less for another. This will be useful to keep in mind to understand some of our upcoming results.

4.2 Interaction coefficient

We now turn to specifying the interaction coefficient $I(x^*)$ under the above setup. Substituting eq. (20) into eq. (3) and using eqs. (21)–(22), we obtain

$$I(x^*) = L(x^*)K(x^*), \tag{23}$$

where $L(x^*) = (\mathrm{d}z_1(y)/\,\mathrm{d}y)(\mathrm{d}e_1(y)/\,\mathrm{d}y)|_{y=x^*}$ is the product of the marginal changes in the first production and regulating factors; and

$$K(x^*) = \left[\left(\frac{\partial^2 W}{\partial z_1 \partial e_1} - M_\mathrm{z}(x^*)\frac{\partial^2 W}{\partial z_2 \partial e_1}\right) - M_\mathrm{e}(x^*)\left(\frac{\partial^2 W}{\partial z_1 \partial e_2} - M_\mathrm{z}(x^*)\frac{\partial^2 W}{\partial z_2 \partial e_2}\right)\right]_{x=y=x^*}, \tag{24}$$

in which $M_\mathrm{z}(x^*)$ and $M_\mathrm{e}(x^*)$ are marginal rates of factor substitution (Jehle and Reny, 2011, p. 128), ensuring that all terms in $K(x^*)$ are in the same units as $\partial^2 W/\partial z_1 \partial e_1$ (see Appendix B.1 for calculations). Specifically, $M_\mathrm{z}(x^*) = (\partial W/\partial z_1)/(\partial W/\partial z_2)|_{x=y=x^*} \geq 0$ is the rate at which production factor z_2 can be substituted for z_1 without changing fitness (i.e. it measures the change in the number of units of z_2 required to compensate for a one-unit increase in z_1 in terms of fitness effects). Similarly, $M_\mathrm{e}(x^*) = (\partial W/\partial e_1)/(\partial W/\partial e_2)|_{x=y=x^*} \geq 0$ is the rate at which limiting factor e_2 can be substituted for e_1 without changing fitness.

Before interpreting $K(x^*)$, let us remark that for most ecologically relevant models, $L(x^*)$ will be nonzero ($L(x^*) \neq 0$), since in the presence of a trade-off, neither a production factor nor a regulating factor will be pushed to a maximum by evolution (i.e. $\mathrm{d}z_1(y)/\,\mathrm{d}y \neq 0$ and $\mathrm{d}e_1(y)/\,\mathrm{d}y \neq 0$ at $y = x^*$). In fact, one can typically construct a model and define fitness factors such that $L(x^*) > 0$, since when a trait increases the production factor for one individual (e.g., a trait increases its fecundity), this is likely to increase a regulating factor when expressed by others (e.g., greater density of offspring before density-dependent competition for recruitment). In particular, when regulating factors are taken as intensive variables (such that $e_1(y) = z_1(y)$, $e_2(y) = z_2(y)$), this leads to $L(x^*) = (\mathrm{d}z_2(x^*)/\,\mathrm{d}x^*)^2 \geq 0$. Either way, condition (6) simplifies to

$$K(x^*) < 0, \tag{25}$$

which therefore provides the condition for NTDS.

To understand eq. (24) biologically, consider that the first summand can be expressed as: $\partial/\partial e_1\,[\partial W/\partial z_1 - M_\mathrm{z}(x^*)\partial W/\partial z_2]$. The term between squared brackets can be thought of as a selection coefficient that represents the net fitness change due to an increase in mutant trait value, taking into account the constraint that an increase in one production factor (z_1) necessarily entails a reduction in the other (z_2) close to x^* (recall eq. 21 or see e.g., Heino et al., 1998; Pásztor et al., 2016). The outer derivative with respect to e_1 then captures how this fitness change is influenced by an increase in the first regulating factor, where such an increase would result from a higher trait value in the resident population. This expression therefore quantifies trait-dependent selection through regulating factor 1. If negative, this indicates that as mutant and resident traits vary in the same direction, competition (or another regulatory constraint associated with e_1) intensifies through regulating factor 1.

Whether this ultimately leads to negative trait-dependent selection depends on the second summand in eq. (24), which can be rewritten as $-M_\mathrm{e}(x^*) \cdot \partial/\partial e_2\,[\partial W/\partial z_1 - M_\mathrm{z}(x^*)\partial W/\partial z_2]$, and quantifies trait-dependent selection through regulating factor 2 (measured in the same units as trait-dependent selection

through regulating factor 1 thanks to $-M_{\mathrm{e}}(x^*)$). The reason trait-dependent selection through regulating factor 2 depends on the negative term $-M_{\mathrm{e}}(x^*)$ can be traced back to their inherent zero-sum relationship between regulating factors 1 and 2 (recall eq. 22): increased competition via one regulating factor necessarily reduces competition via the other. Thus, negative trait dependent selection through, say regulating factor 1 ($\partial/\partial e_1\,[\partial W/\partial z_1 - M_{\mathrm{z}}(x^*)\partial W/\partial z_2]<0$), can be offset by positive trait dependent selection through regulating factor 2 ($-M_{\mathrm{e}}(x^*)\cdot\partial/\partial e_2\,[\partial W/\partial z_1 - M_{\mathrm{z}}(x^*)\partial W/\partial z_2] > 0$) where this compensation is strong. Then, $K(x^*) = 0$ and any fitness loss experienced by the mutant due to population feedback via one regulating factor is precisely offset by a fitness gain via the other regulating factor. If, however, this compensation is incomplete, or NTDS occurs through both regulating variable, then $K(x^*) < 0$, which results in net NTDS.

A key corollary of this analysis is that polymorphism cannot emerge when invasion fitness depends on a single regulating factor, since in this case, $K(x^*) = 0$ and the necessary condition for polymorphism can never be met. Specifically, when only one regulating factor, say e_2 affects fitness while the other (e_1) does not, it follows that $M_{\mathrm{e}}(x^*) = 0$ and $\partial^2 W/(\partial z_1 \partial e_1) = \partial^2 W/(\partial z_2 \partial e_1) = 0$. This conforms to the standard precept in ecology that at least two regulating factors are necessary to sustain coexistence between two morphs (recall eq. (19)). Usefully, eq. (24) does not rely on explicitly modelling how traits influence individual fitness factors. This generality makes eq. (24) valuable as it provides a direct means to identify conditions that promote polymorphism when applied to a specific scenario. We next illustrate these properties under two different classes of scenarios: (1) within-generation heterogeneity in fitness factors, and (2) between-generation heterogeneity in these factors.

4.3 Within generation heterogeneity in production factors

4.3.1 Resource specialisation under panmixia

We begin by considering a classic scenario widely used to explain how polymorphism can arise from environmental heterogeneity. We consider a situation involving two distinct habitats, each providing a fixed number of breeding spots. The life cycle unfolds in three steps: (a) Each adult, regardless of the habitat it occupies, produces a large number of juveniles and then dies. (b) Each juvenile disperses randomly and encounters each habitat according to their frequency. (c) Density-dependent regulation occurs for access to the breeding spots in each habitat and the probability of surviving this regulation depends on habitat specific phenotypic expression. For this life-cycle, invasion fitness $\rho(x, y) = W(z_1(x), z_2(x), e_1(y), e_2(y))$ is given by

$$\rho(x, y) = q\frac{z_1(x)}{e_1(y)} + (1-q)\frac{z_2(x)}{e_2(y)}, \tag{26}$$

where q is the frequency of habitat 1 (and $1-q$ of habitat 2), and $z_i(x)$ denotes the competitive ability of an individual with trait x in habitat $i \in \{1, 2\}$, e.g. how effectively it extracts resources or germinates. The regulating factors in this example correspond directly to the competitive abilities of the resident population,

so that $e_1(y) = z_1(y)$ and $e_2(y) = z_2(y)$, making them intensive variables. Many evolutionary models of local adaptation have fitness functions that explicitly or implicitly follow the form of eq. (26) (e.g., Levene, 1953; Gillespie, 1991; Kisdi and Geritz, 1999; Svardal et al., 2015). More generally, invasion fitness in this form encapsulates a broad class of biologically relevant scenarios that lead to a weighted average over two fitness components, each involving contest competition for a limiting resource, expressed as a ratio $z_i(x)/z_i(y)$. This is the case for instance of models of sexual antagonism where a trait has different optima in males and females and competition occurs within sexes (with $q = 1/2$ under equal sex ratio, Van Dooren et al., 2004; Flintham et al., 2024).

Substituting eq. (26) into eq. (21), shows that x^* solves

$$\frac{q}{z_1(x^*)}\frac{\mathrm{d}z_1(x)}{\mathrm{d}x}\bigg|_{x=y=x^*} + \frac{1-q}{z_2(x^*)}\frac{\mathrm{d}z_2(x)}{\mathrm{d}x}\bigg|_{x=y=x^*} = 0. \tag{27}$$

This equation expresses the fact that at x^*, selection across the two habitats must balance: any increase in fitness in one habitat due to a change in trait value is offset by a corresponding fitness decrease in the other. Substituting eq. (26) into eq. (24) (and keeping eq. 20 in mind) allows us to see that

$$K(x^*) = -\frac{q}{(1-q)z_1(x^*)^2} \leq 0, \tag{28}$$

whenever $0 < q < 1$. This indicates that as long as fitness depends on both habitats, the necessary condition for NTDS is satisfied. The ease with which NTDS emerges in this model stems from the absence of interference between trait-dependent selection in each habitat. Formally, this is captured by $\partial^2 W/(\partial z_2 \partial e_1) = \partial^2 W/(\partial z_2 \partial e_1) = 0$, such that $K(x^*) = \partial^2 W/(\partial z_1 \partial e_1) + M_{\mathrm{e}}(x^*)M_{\mathrm{z}}(x^*)\partial^2 W/(\partial z_2 \partial e_2) < 0$, which, owing to the fact that $\partial^2 W/(\partial z_i \partial e_i) < 0$, is the sum of NTDS experienced across both habitats.

A typical form of the production factor under fitness model (26) is $z_i(x) = e^{-(x-o_i)^2/(2\sigma^2)}$, where o_i is the optimal trait value in habitat i and σ^2 parametrises the strength of stabilizing selection around this optimum. With this formulation, it is straightforward to show that for a range of parameter values, monomorphism is evolutionarily unstable, leading to evolutionary branching, a well-established result (e.g., Christiansen, 1991; Kisdi and Geritz, 1999; Svardal et al., 2015). If so, gradual evolution results in the differentiation of two morphs, each better adapted to one habitat, leading to a convergence-stable and uninvadable saturated dimorphism. We recapitulate and further detail these classical results in Appendix B.2.1. Fitness model (26) easily generalises to an arbitrary number of habitats and thus regulating factors (eq. A-32 for invasion fitness in that case and Svardal et al., 2015 for a general analysis). In Fig. 2, we show the evolution of long-term saturated and unsaturated polymorphism from individual-based simulations within such a model.

4.3.2 Life-history under limited dispersal

We now turn to another classical scenario: the trade-off between fecundity and survival. While this trade-off has been widely studied in the context of life-history evolution (e.g. Charnov, 1997; Case, 2000), its role in maintaining polymorphism under limited dispersal remains unexplored. Here, we use our framework to fill this gap. We consider a subdivided iteroparous population, where individuals undergo the following life-cycle events: (a) Each adult individual occupies a site where it produces a large number of juveniles and then either survives or dies. (b) Each juvenile either migrates out of the natal site with probability m or remains philopatric with complementary probability $1-m$. (c) Density-dependent competition occurs locally among juveniles for the breeding spots left vacant by the death of adults, such that the population is regulated back to the fixed adult size.

With $z_1(x)$ as the survival probability of an adult individual expressing trait x and $z_2(x)$ as its fecundity, invasion fitness for this model is

$$\rho(x,y) = z_1(x) + (1-z_1(x))\frac{(1-m)z_2(x)}{(1-m)z_2(x)+me_2(y)} + (1-e_1(y))\frac{mz_2(x)}{e_2(y)}, \tag{29}$$

where the regulating factors are $e_1(y) = z_1(y)$ and $e_2(y) = z_2(y)$, i.e. the survival and fecundity of the resident, and are thus intensive variables. Invasion fitness eq. (29) corresponds exactly to the model of Pen (2000) with one individual per patch and to that of Ronce and Promislow (2010) in the absence of senescence. With complete dispersal $m = 1$, it reduces to the standard reproductive effort model from life-history theory in panmictic populations (Charnov, 1997; Case, 2000).

Substituting eq. (29) into eq. (21), shows that any singularity x^* must solve

$$\left.\frac{\mathrm{d}z_1(x)}{\mathrm{d}x}\right|_{x=y=x^*} + \frac{(2-m)(1-z_1(x^*))}{z_2(x^*)}\left.\frac{\mathrm{d}z_2(x)}{\mathrm{d}x}\right|_{x=y=x^*} = 0, \tag{30}$$

which expresses the trade-off between investing in survival (first term) and reproduction (second term). The coefficient $(2-m)$ in eq. (30) determines the strength of this trade-off: the smaller this coefficient, the larger the investment in fecundity required to compensate for a given loss of survival. Since $(2-m)$ is smallest when dispersal is unlimited ($m = 1$), individuals in this case evolve to invest less in reproduction than under limited dispersal (all else equal). This is because limited dispersal increases competition among kin, requiring the production of more offspring to be offset (kin competition is captured by $(1-m)z_2(x)$ in the denominator of eq. (29)).

On substituting eq. (29) into eq. (24), we obtain

$$K(x^*) = -\frac{(1-m)m^2}{(2-m)^2[1-z_1(x^*)]} \leq 0. \tag{31}$$

This indicates that when dispersal is unlimited, $m = 1$, $K(x^*) = 0$ so that selection is never negative

frequency-dependent meaning no polymorphism can evolve under gradual evolution. In contrast, when dispersal is limited, $K(x^*) < 0$, which results in NTDS. To see why, it is useful to consider mutant and residents both investing more in reproduction and less in survival. When $m = 1$, juveniles mix freely. An increase in competition experienced by the mutant for each available breeding spots is then exactly compensated by the availability of more breeding spots, since investment into survival decreases (the two terms in eq. 24 exactly cancel each other out, the first capturing positive trait-dependent selection while the other negative trait-dependent selection). In contrast, when $m < 1$, the increased competition for available breeding spots is exacerbated locally due to kin competition. This creates an asymmetry in the response to both regulating factors and entails that an individual investing more in survival gains a fitness advantage. This NTDS should therefore enable the coexistence of different life-history strategies, given an appropriate trade-off between survival and reproduction.

We investigated this specifically with a trade off given by $z_1(x) = x^\alpha$ and $z_2(x) = (1-x)^\beta$ with parameters $\alpha < 1$ and $\beta < 1$. Our analysis in Appendix B.2.2 confirms that, for a range of parameter values, polymorphism in resource allocation can indeed be favoured by gradual evolution as long as $m < 1$. Interestingly, the end-point of evolution consists of one morph investing entirely in survival (becoming effectively immortal), while the other morphs can consist of any mixed allocation to survival and fecundity. Once immortality has evolved, mortality-driven competition vanishes and thus selection on both types ceases. Since this form of adaptive polymorphism in life-history has not been previously reported, we conducted individual-based simulations to confirm our findings. As illustrated Fig. 3, individual-based simulations reveal the gradual emergence of a quasi immortal morph (mutations preclude converging to exactly $x = 1$ for the immortal morph), coexisting with a second type that invests in both survival and reproduction. The population eventually shows a marked dimorphism in life-history strategies (see Appendix B.2.2 for more details).

4.4 Between generation heterogeneity in production

4.4.1 Resource specialization under panmixia and the storage effect

Like within-generation (spatial) heterogeneity, between-generation (temporal) heterogeneity has long been recognised as a key driver of polymorphism. We explore this in a well-mixed population that experiences two environmental states, say a "wet" and a "dry" environment, occurring independently in each generation with probabilities q and $1 - q$, respectively. At each demographic time step, individuals reproduce with a fecundity that depends on both their trait and the environmental state, where $z_i(x)$ denotes the fecundity of an individual with trait x in state $i \in \{1.2\}$. Adults then either survive to the next time step with probability s or die, freeing up breeding spots. Offspring compete randomly for these vacant spots, with the total number of breeding spots remaining fixed. Invasion fitness for this scenario then takes the form of a geometric mean

over environmental states, i.e.

$$\rho(x,y) = \left(s + (1-s)\frac{z_1(x)}{e_1(y)}\right)^q \left(s + (1-s)\frac{z_2(x)}{e_2(y)}\right)^{1-q}, \tag{32}$$

with regulating factors $e_1(y) = z_1(y)$ and $e_2(y) = z_2(y)$. Eq. (32) is a two environmental state version of the haploid model of Turelli et al. (2001) and that of Svardal et al. (2015) under complete migration.

Substituting eq. (32) into eq. (21) shows that any singularity x^* satisfies eq. (27), meaning that x^* does not depend on survival s or the nature of heterogeneity. When substituting eq. (32) into eq. (24), we obtain

$$K(x^*) = -\frac{s(1-s)q}{(1-q)z_1(x^*)^2} \leq 0, \tag{33}$$

indicating that NTDS requires overlapping generations $s > 0$. When $s = 0$, all adults die after reproducing, ensuring that direct competition occurs only among newly produced juveniles each generation. As a result, when mutant and resident traits change similarly to produce more offspring in environment 1, fitness effects are exactly cancelled out by the loss of producing offspring in environment 2 (i.e., the two terms in eq. 24 exactly cancel each other out). Consequently, NTDS cannot arise. In contrast, when $s > 0$, some adults survive between generations. So when mutant and resident both produce more offspring in environment 1, the increase in competition is carried over across generation because the offspring of surviving individuals then compete more intensely among themselves when environment 1 recurs. This overlap in generations creates the necessary asymmetry in regulating factors for individuals producing more offspring in environment 2 to gain a fitness advantage, thereby enabling NTDS. For instance, if $z_i(x) = \exp(-(x-o_i)^2/(2\sigma^2))$, it can readily be shown that with fitness eq. (32) there exists a range of parameter values leading to evolutionary branching, and that gradual evolution leads to a convergence stable and uninvadable saturated dimorphism, where each morph is better adapted to one of the two environments (Appendix B.3).

These results are closely related to the well-known storage effect of ecological theory, which explains species coexistence in fluctuating environments through three key biological features (Chesson, 2000a). (i) A covariance between production and regulating factors, which in our model is captured by the fact that production directly determines the regulating factors ($e_1(y) = z_1(y)$ and $e_2(y) = z_2(y)$). (ii) Trait-specific responses that decouple production and regulation across different environments. In our setup, the trait x has distinct environmental optima and the regulating factors $e_1(y) = z_1(y)$ and $e_2(y) = z_2(y)$ vary accordingly. (iii) A buffering mechanism against unfavourable conditions, provided here by overlapping generations ($s > 0$). Together, these three features ensure that when mutant and resident traits are varied in opposite directions competition is reduced and fitness is enhanced, generating NTDS.

4.4.2 Limited dispersal and the mitigation of the storage effect

As a final scenario of ecologically mediated interactions, we consider a combination of between-generation environmental fluctuations (eq. 32) with limited dispersal (eq. 29). Each adult occupies a site where it produces a large number of juveniles, with fecundity $z_i(x)$ depending on both its trait x and the environmental state $i \in \{1.2\}$. Adults then either survive with probability s or die, freeing up breeding spots. Each juvenile either disperses with probability m or remains philopatric. Density-dependent competition among juveniles occurs locally for the breeding spots vacated by adult mortality.

Under this scenario, invasion fitness is

$$W(z_1(x), z_2(x), e_1(y), e_2(y)) = \left(s + (1-s)\left[\frac{(1-m)z_2(x)}{(1-m)z_2(x) + me_2(y)} + \frac{mz_2(x)}{e_2(y)}\right]\right)^q \times \left(s + (1-s)\left[\frac{(1-m)z_1(x)}{(1-m)z_1(x) + me_1(y)} + \frac{mz_1(x)}{e_1(y)}\right]\right)^{1-q}, \quad (34)$$

where $e_1(y) = z_1(y)$ and $e_2(y) = z_2(y)$. Substituting eq. (34) into eq. (24) and using $S(x^*) = 0$ from eq. (21) yields

$$K(x^*) = -\left[s(2-m)^2 - (1-m)^2\right] \frac{q\, m^2(1-s)}{(1-q)z_1(x^*)^2}.$$

Whether $K(x^*) < 0$, now hinges on the condition that $s > (1-m)^2/(2-m)^2$. This condition reflects the antagonistic effects that adult survival and limited dispersal have on NTDS here. On the one hand, overlapping generations combined with between-generation heterogeneity favour NTDS (as under the storage effect discussed above). On the other hand, limited dispersal increases kin competition, which opposes NTDS. Kin competition reduces NTDS here by weakening the advantage of using a different environment than other individuals in the population. This occurs because offspring that share the same traits compete primarily among themselves when dispersal is limited, reducing the advantage of an individual producing offspring in a different environment than the population. In other words, kin competition erodes the asymmetry necessary for NTDS by limiting the fitness advantage mutants gain by switching environments. This outcome aligns with earlier findings that kin competition tends to inhibit evolutionary branching under environmental heterogeneity (Day, 2001; Ajar, 2003). Together, the results of this section and section 4.3.2 show that kin competition can both favor or inhibit NTDS, depending on the biological scenario. The role of kin selection in polymorphism is explored further in section 6.2.

5 Socially mediated interactions

So far, we examined trait-dependent selection arising through ecological feedbacks, where traits alter environmental conditions that, in turn, influence selection. We now turn to another fundamental class of trait-dependent interactions: those mediated directly by the payoffs individuals receive through social in-

teractions. Classic examples include contests for mates or territories, public-good investments, cooperative breeding, and various forms of aggressive or cooperative encounters, all well studied within the framework of evolutionary game theory (e.g., Maynard Smith, 1982; Vincent and Brown, 2005; McNamara and Leimar, 2020). Here, an individual's payoff explicitly depends on the traits expressed by others in the population, creating a social feedback loop that may lead to stable polymorphisms.

5.1 Selection and interaction coefficients at the payoff level

To capture socially mediated interactions in a simple yet general way, we assume that the payoff to an individual is one-dimensional, given by $\boldsymbol{\pi}(\mathbf{z}(x, \mathbf{e}^{\mathrm{z}}(x,y)), \mathbf{e}^{\pi}(x,y)) = \pi(x,y)$, and depends explicitly on both mutant and resident traits. Additionally, we assume that the regulating variable $\mathbf{e}^{\mathrm{w}}(x,y) = e(y)$ is also one dimensional and depends solely on the resident population. Under these assumptions, invasion fitness (18) takes the form

$$\rho(x,y) = W(\pi(x,y), e(y)), \tag{35}$$

for some mapping W. We assume that invasion fitness is monotonically increasing with payoff (i.e. W increases with its first argument). Due to the constraint $\rho(y,y) = 1$, it then necessarily follows that W decreases with regulating variable $e(y)$ (the second argument). Eq. (35) represents the standard fitness formulation that is implicitly used in evolutionary game theory, encompassing both pairwise interactions and cases where individuals interact collectively ("playing against the field"; e.g., Maynard Smith, 1982; Vincent and Brown, 2005; McNamara and Leimar, 2020). This formulation captures the essence of socially mediated interactions, where fitness depends explicitly on payoffs—determined by the traits of interacting individuals—while environmental factors influence selection only indirectly through their effect on population-wide competition or regulation.

Substituting eq. (35) into eq. (3), we show in Appendix C.1 that the invasion analysis can be fully worked out from the payoff function. In particular, any singular strategy x^* must be such that

$$\left.\frac{\partial \pi(x,y)}{\partial x}\right|_{x=y=x^*} = 0, \tag{36}$$

and the necessary condition for protected polymorphism (eq. 6) reduces to

$$I(x^*) \propto I_\pi(x^*) = \left.\frac{\partial^2 \pi(x,y)}{\partial x \partial y}\right|_{x=y=x^*} < 0. \tag{37}$$

When condition (37) holds, this indicates that a mutant's best response for payoff is to do the opposite to a trait change in the population (e.g. decrease x when there's an increase in y). Since fitness is monotonically increasing with payoff, these payoff effects directly generate NTDS. We explore further how such selection can emerge owing to social interactions by making more specific assumptions concerning payoff below.

5.2 Action based feedback

To investigate how a social feedback can promote polymorphism, we consider a class of models in which an individual's payoff depends explicitly on actions taken during interactions. Specifically, we assume that the payoff can be expressed as:

$$\pi(x, y) = \Pi(a(x, y), a(y, x)), \tag{38}$$

for a mapping Π, where $a(x, y)$ and $a(y, x)$, are the actions of a mutant and a resident individual, respectively (formally actions here are equivalent to phenotypes z but for ease of interpretation in the social context, we'll refer to actions and use the variable a in this section). These actions may depend not only on an individual's own trait but also on the trait of its interaction partner.

This formulation covers two broad classes of interactions. The first is pairwise interactions, where $a(x, y)$ and $a(y, x)$ describe the actions of two individuals in a direct encounter. The dependence of an individual's action on its own trait and that of its partner allows us to model cases where actions arise as equilibrium outcomes of a dynamic behavioural process, such as learning, information exchange, or negotiation. This notion is widely used in models of evolving response rules and preference evolution (e.g., McNamara et al., 1999; André and Day, 2007; Heifetz et al., 2007a; Akzcay and Van Cleve, 2012; Alger and Lehmann, 2023). Of course, this also captures simpler pairwise interactions where one's action depends only on own trait, $a(x, y) = a(x)$, as is very commonly studied in evolutionary game theory (Maynard Smith, 1982; McNamara and Leimar, 2020). In the more general case, the resident's trait y in $a(x, y)$ corresponds to the regulating factor $\mathbf{e}^{z}(x, y) = y$ acting at the level of the phenotype (i.e. the action), while the resident's action $a(y, x)$ correspond to the regulating factor $\mathbf{e}^{\pi}(x, y) = a(y, x)$ at the level of payoff (recall eq. 18).

The second class of interactions covered by eq. (38) corresponds to "playing against the field" multiplayer interactions (Maynard Smith, 1982). In this case, actions depend only on the focal individual's own trait, such that $a(x, y) = a(x)$ and $a(y, x) = a(y)$. This extends beyond pairwise interactions, as multiple individuals expressing action $a(y)$ can simultaneously influence the focal individual's payoff. Here, there is no regulating factor at phenotype level, and the regulating factor at the payoff level simplifies to $\mathbf{e}^{\pi}(x, y) = a(y)$.

Substituting eq. (38) into eq. (36) reveals that any singularity x^* must be such that

$$\left.\frac{\partial \Pi(a, b)}{\partial a}\right|_{\substack{x=y=x^* \\ a=b=a(x^*,x^*)}} + \lambda(x^*) \left.\frac{\partial \Pi(a, b)}{\partial b}\right|_{\substack{x=y=x^* \\ a=b=a(x^*,x^*)}} = 0, \tag{39}$$

where

$$\lambda(x^*) = \frac{\partial a(y, x)}{\partial x} \Big/ \left.\frac{\partial a(x, y)}{\partial x}\right|_{x=y=x^*} \tag{40}$$

is a response coefficient, which measures how much the partner to a focal individual changes her action in response to a change in the action of the focal (a change that is due to a trait change in the focal). When $\lambda(x^*) > 0$, an increase in one's action due to a trait change leads to a corresponding increase in the partner's

action. Conversely, when $\lambda(x^*) < 0$, the response is negative, meaning that a trait-driven increase in one's action is associated with a decrease in the partner's action. Eq. (39) expresses that at equilibrium, selection balances two opposing effects on payoff: a direct effect due to changes in the focal individual's own action (first term) and an indirect effect due to the response of the partner's action (second term), weighted by the response coefficient $\lambda(x^*)$. This balance helps to understand the evolution of many well-known social interactions. The evolution of cooperation via reciprocity, for instance, may occur when $\lambda(x^*) > 0$ as this allows individuals who express cooperative behaviour to receive cooperative responses in return. More broadly, the coefficient $\lambda(x^*)$ plays a central role in different social evolution models (e.g., Akzcay and Van Cleve, 2012; Akzcay and Cleve, 2014; Lehmann, 2022; Alger and Lehmann, 2023).

Using eq. (39) and eq. (40) on substituting eq. (38) into eq. (37), we find that $I_\pi(y) = L_\pi(x^*)[K_\pi(x^*) + K_\lambda(x^*)]$, where $L_\pi(x^*) = (\partial a(x,y)/\partial x)^2\Big|_{x=y=x^*} \geq 0$, such that NTDS selection occurs when

$$K_\pi(x^*) + K_\lambda(x^*) < 0, \tag{41}$$

in which

$$K_\pi(x^*) = \left[\left(1+\lambda(x^*)^2\right)\frac{\partial^2\Pi(a,b)}{\partial a\partial b} + \lambda(x^*)\left(\frac{\partial^2\Pi(a,b)}{\partial a^2} + \frac{\partial^2\Pi(a,b)}{\partial b^2}\right)\right]_{\substack{a=a(x^*,x^*)\\ b=a(x^*,x^*)}} \tag{42a}$$

and

$$K_\lambda(x^*) = \underbrace{\frac{\partial^2 a(y,x)}{\partial x\partial y}\Big/\left(\frac{\partial a(x,y)}{\partial x}\right)^2\Bigg|_{x=y=x^*}}_{=A(x^*)} \cdot \left(\frac{\partial\Pi(a,b)}{\partial a} + \frac{\partial\Pi(a,b)}{\partial b}\right)_{\substack{a=a(x^*,x^*)\\ b=a(x^*,x^*)}}. \tag{42b}$$

The first term, $K_\pi(x^*)$, describes trait-dependent selection via direct payoff effects, which considers how selection on the mutant changes as the population trait is varied, while keeping the response coefficient $\lambda(x^*)$ unchanged. This is relevant in cases where the effects of actions on payoff are antagonistic, meaning that the actions of interacting individuals interfere in a way that reduces payoff, i.e. $\partial^2\Pi(a,b)/(\partial a\partial b) < 0$. When this holds, a mutant obtains lower payoffs when the actions of interacting individuals change similarly to the mutant.

The second term, $K_\lambda(x^*)$, describes trait-dependent selection via response effects, i.e. through changes in $\lambda(y)$. This is governed by $A(x^*)$, which measures how one's action changes in response to a similar trait change in others. When $A(x^*) < 0$ changing traits similarly reduces the response coefficient, whereas when $A(x^*) > 0$ it increases it. For example, if $\partial\Pi/\partial a + \partial\Pi/\partial b > 0$ as in cooperative interactions and $A(x^*) < 0$—for instance, due to interference in cooperative investments—an increase in population trait leads to weaker cooperation overall, favouring NTDS. We next illustrate how eq. (41) helps understand two more specific examples of social interactions generating NTDS.

5.3 Examples

5.3.1 Contests over resources

Consider a scenario where payoffs are determined by contest competition among N individuals for access to a resource of value v. Suppose the payoff function is

$$\Pi(a(x), b(y)) = \frac{a(x)}{a(x) + (N-1)b(y)} v - C(a(x)), \tag{43}$$

where the action $a(x)$ measures the fighting or competitive effort of an individual with trait x and $C(a(x))$ is the cost associated with such effort. The probability of winning the contest depends on an individual's relative effort, given by the ratio $a(x)/[a(x) + (N-1)b(y)]$, which is standard of conflict and contest competition (e.g., Hirshleifer, 1987; Skaperdas, 1996; Cant, 2012). This is a scenario of "playing against the field," where actions depend only on one's own trait and not on the trait of interacting partners. Thus, the response coefficient (eq. 40) is zero $\lambda(y) = 0$, which implies $A(x^*) = 0$ and as a result, $K_\lambda(x^*) = 0$ here.

Whether trait-dependent selection is negative therefore depends solely on:

$$K_\pi(x^*) = -\frac{(N-1)(N-2)v}{n^3 a(x)^2} \leq 0, \tag{44}$$

found by substituting eq. (43) into eq. (37). This equation shows that as long as $N > 2$ (meaning that at least three individuals compete for a resource), NTDS occurs. The ease with which NTDS arises under payoff model (43) exemplifies how conflicts in general provide a natural setting for polymorphism, since if one individual increases its competitive effort, this is likely be be mitigated if others in the population change similarly (otherwise there wouldn't be much competition). In fact, evolutionary diversification has been found to occur through gradual evolution in multiple conflict models (McNamara and Leimar, 2020).

As a concrete example of polymorphism emerging under fitness model (43), one can use the production function $a(x) = e^{\omega x}$ and the cost function $C(a(x)) = x^2$ for some parameter $\omega > 0$. For $N > 2$, it can be verified that there exists a range of parameter values leading to evolutionary branching with these functions. Over the long term, gradual evolution results in a convergence-stable and uninvadable dimorphism, with one morph investing heavily in competition while the other adopts a more moderate investment strategy. Further details on the analysis are provided in Appendix C.2.1, and Fig. 3 shows the emergence of this previously unreported polymorphism in individual-based simulations.

5.3.2 Evolution of preferences

The previous example had no behavioural response ($\lambda(y) = 0$); we now consider a case where individuals adjust their behaviour in response to their partner's traits ($\lambda(y) \neq 0$). We assume individuals interact in a

pairwise manner and aim to maximise a utility function, which evolves under a situation of complete information (see Alger, 2023 for discussion of information settings). This follows models of preference evolution, where selection acts not directly on the actions individuals take but on the underlying behavioral mechanism guiding their decision-making (e.g., Heifetz et al., 2007b; Akzcay and Van Cleve, 2012; Alger and Lehmann, 2023; see Alger, 2023 for a review).

Building on Heifetz et al. (2007a,b), we assume that each individual's utility consists of their own material payoff and a self-esteem bias, which distorts their perception of their own actions. The utility functions for a mutant with action a and trait x, and a resident with action b and trait y, are given by

$$\begin{aligned} u_x(a,b) &= \Pi(a,b) + xa \\ u_y(b,a) &= \Pi(b,a) + yb, \end{aligned} \tag{45}$$

respectively. Here, the self-esteem bias, modulated by the evolving trait (x in mutants, y in residents), determines whether individuals overestimate or underestimate the benefits of their own actions. When $x = 0$, a mutant individual simply seeks to maximise its material payoff. But when $x > 0$, the mutant overestimates the benefits of its own action, acting as if its action are more rewarding than they actually are. Conversely, when $x < 0$, the mutant underestimates its returns and behaves more cautiously, even when a greater action value would be materially beneficial. In this way, the evolving trait x governs an individual's "disposition" (Heifetz et al., 2007a,b), shaping whether they are overconfident optimists ($x > 0$) or underconfident pessimists ($x < 0$) in how they perceive and respond to interactions.

To complete the model, we assume that material payoffs follow a linear-quadratic structure, given by

$$\Pi(a,b) = (\alpha - \gamma a - \beta b)a, \tag{46}$$

for parameters $\alpha > 0$, $\gamma > 0$, and $\beta \geq \gamma$. Eq. (46) is consistent with a model where actions constitute the effort or energy allocated to the extraction of a biotic resource that has logistic growth and a linear functional response to extraction. In this case, parameters of eq. (46) can be interpreted as follows. The parameter α is proportional to the replenishment rate of the resource and γa to the rate of resource extraction by an individual with action a. Assuming fast resource dynamics, $(\alpha - \gamma a - \gamma b)a$ is then proportional to the amount of resources individual with action a extracts per unit time. Finally, we add to this $ab(\beta - \gamma)$, which can be seen as a cost stemming from interference competition. In particular, when $\beta > \gamma$, joint efforts lead to diminishing returns—similar to interference competition where competition over shared resources reduces overall resource uptake rate (e.g.,Begon and Townsend, 2021, chapter 5).

By maximising their utility given the action of their partner, each individual plays a Nash equilibrium strategy. The necessary first-order conditions for a Nash equilibrium require that neither individual can improve their utility by unilaterally changing their action, meaning $\partial u_x(a,b)/\partial a = 0$ and $\partial u_y(b,a)/\partial b = 0$. Using eqs. (45)

and (46), these conditions can be expressed as the system:

$$a = \frac{\alpha + x}{2\gamma} - b\frac{\beta}{2\gamma} \qquad (47)$$
$$b = \frac{\alpha + y}{2\gamma} - a\frac{\beta}{2\gamma}.$$

Solving these for $a = a(x, y)$ and $b = a(y, x)$ yields

$$a(x, y) = \frac{2\gamma(\alpha + x) - \beta(\alpha + y)}{4\gamma^2 - \beta^2}. \qquad (48)$$

This solution represents a Nash equilibrium action because it is unique and satisfies the second-order condition for stability: $\partial^2 u_x(a, b)/\partial a^2 = \partial^2 \Pi(a, b)/\partial a^2 = -2\gamma \leq 0$ and $\partial^2 u_y(b, a)/\partial b^2 = \partial^2 \Pi(b, a)/\partial b^2 = -2\gamma \leq 0$. Eq. (48) shows that an individual's action increases with their own optimism (x) but decreases with that of their partner (y). This is because when $\gamma > 0$ joint increase in action reduces individual resource extraction. To avoid these deleterious effects, since it is best to reduce one's own action when faced with more optimistic individuals (who increase their action as a result). This relationship is captured by the negative response coefficient $\lambda(x^*) = -\beta/(2\gamma)$ (found by plugging eq. 48 in eq. 40), indicating that an increase in the partner's action leads to a decrease in one's own action.

A key feature of eq. (47) is that each individual's action is a linear response to their partner's action. This structure is formally identical to models for the evolution of response rules, where individuals adjust their behaviour based on social interactions (e.g., McNamara et al., 1999; Taylor and Day, 2004; see Alger, 2023 for a discussion on the connection between preference evolution and response rule evolution models). This linearity means that, here and in these aforementioned models, $A(y) = 0$ (eq. 42b). As a result, there can be no trait-dependent selection via a feedback on behaviour rules (i.e. $K_\lambda(x^*) = 0$).

We now turn to the question of whether NTDS can instead arise at the level of payoffs, by checking whether $K_\pi(x^*) < 0$. Substituting eqs. (46) and (48) into eq. (42a), we obtain

$$K_\pi(x^*) = -\frac{\beta^3}{4\gamma^2}, \qquad (49)$$

which is negative because when both a focal individual and population member increase their optimism bias, payoff is reduced in the presence of interference competition (where $\beta > \gamma > 0$ according to our assumptions). It is then straightforward to show that evolutionary branching in this self-esteem trait can happen as long as $\beta > \sqrt{2}\gamma$. When it occurs, evolutionary branching leads to ever more "optimists" (with increasingly positive trait values) and ever more "pessimists" (with increasingly negative trait values) as both strategy fuel one another. This open ended evolution of exaggerated self-perception comes to a halt where there is a cost to overconfidence (Appendix C.2.2 for analysis and Fig. 3 for individual-based simulations).

The original study by Heifetz et al. (2007a,b) did not analyse the scope for polymorphism in self-esteem

biases, and in fact considered parameter values generating evolutionarily stable singularities only. Our re-examination of this model shows that the interaction coefficient (eq. 42) offers a straightforward method to detect the potential for polymorphism, and, in some cases, the approach describes in section 2.4 allows to fully characterise the resulting evolutionary stable coalition analytically (Appendix C.2.2 for details). Heifetz et al. (2007a,b) also considered an alternative model of preference, in which the evolving trait x measures an individual's other-regard. In this formulation, mutant and resident utilities take the form $u_x(a,b) = \Pi(a,b) + x\Pi(b,a)$ and $u_y(b,a) = \Pi(b,a) + y\Pi(a,b)$ (see also Alger, 2023). We analyse this model in Appendix C.2.3 using eq. (46) for payoff. We show that for this model NTDS arises through payoff effects ($K_\pi(x^*) < 0$) as well as via action response effects ($K_\lambda(x^*) < 0$). Thus, polymorphism can emerge in this model too.

Overall, the findings of this section highlight how the interaction coefficient helps to pinpoint the conditions under which traits—whether reflecting self-esteem biases or other-regard—can diversify. They also underscore that socially mediated interactions, in which individuals' payoffs directly depend on how others act and perceive, are fertile grounds for NTDS. Consequently, we suggest that adaptive polymorphism in preferences should emerge readily and be investigated more systematically.

6 Genetically mediated interactions

Up until now, we have assumed that individuals are haploid and that the population is well mixed. We relax here both of these assumptions in turn, and investigate their implications for adaptive polymorphism. We'll see that these implications depend on the nature of interactions between genes, respectively within and between individuals (hence "genetically mediated interactions").

6.1 Diploidy and heterozygote advantage

Let us consider now that individuals are diploid and that an individual's phenotype depends on the two homologous alleles it carries. To model this, we denote as $z_i(x,y)$ the phenotype i of an individual carrying a mutant allele with heritable trait value x and and a resident allele with heritable trait value y. As such an individual homozygote for the resident allelic value y expresses phenotype $z_i(y,y)$. We assume that mating is random so that the mutant allelic form x is only ever present in heterozygotic state. We also assume that the genotype-to-phenotype map is: (i) homogeneous (by which we mean $z_i(x,y) = z_i(y,x)$) and (ii) such that variation in allelic value always leads to variation in phenotype, i.e. that phenotypes always show some degree of heritability (which entails $\partial z_i(x,y)/\partial x|_{x=y} = \partial z_i(x,y)/\partial y|_{x=y} \neq 0$, where the first equality unfolds from assumption (i)). Where the phenotype of heterozygotes is a convex combination of the phenotypes of homozygotes, assumptions (i) and (ii) entail additive gene action (i.e. entail that $z_i(x,y) = z_i(x,x)/2 + z_i(y,y)/2$), but our forthcoming results apply more broadly.

6.1.1 No effect of diploidy on the conditions for polymorphism of a single phenotype

Let us first consider the case where invasion fitness depends on a single phenotype such that we can write

$$\rho(x, y) = W(z(x, y), z(y, y)), \tag{50}$$

where the function W gives the invasion fitness of a x/y heterozygote in a population of y homozygotes. Substituting eq. (50) into eqs. (3)-(4) and using our assumptions concerning $z(x, y)$, we show in Appendix D.1 that the population converges to a singular allelic value x^*, which generates phenotype $z^* = z(x^*, x^*)$, and subsequently undergoes evolutionary branching when the following holds:

$$\begin{aligned} &\left.\frac{\partial W(z,\bar{z})}{\partial z}\right|_{z=\bar{z}=z^*} = 0 \\ &\left.\frac{\partial^2 W(z,\bar{z})}{\partial z^2}\right|_{z=\bar{z}=z^*} + \left.\frac{\partial^2 W(z,\bar{z})}{\partial z \partial \bar{z}}\right|_{z=\bar{z}=z^*} < 0 \\ &\left.\frac{\partial^2 W(z,\bar{z})}{\partial z^2}\right|_{z=\bar{z}=z^*} > 0. \end{aligned} \tag{51}$$

These conditions are equivalent to those derived under haploidy (mirroring eq. 4 and conditions discussed thereafter). This shows that diploidy does not affect the conditions for polymorphism, at least when underlying a single phenotype. In particular, the emergence of polymorphism requires that the cross derivative of the fitness function with respect to mutant and resident phenotypes is negative, i.e. $\partial^2 W(z, \bar{z})/\partial z \partial \bar{z}\big|_{z=\bar{z}=z^*} < 0$, indicating that directional selection depends negatively on the phenotype expressed in the resident population (as in eq. 6 at the level of the trait value). Consequently, all our results in sections (4)–(5), which are based on a invasion fitness function of the form $\rho(x, y)$, still apply unchanged to diploids, with x now standing for the phenotypic value of a mutant individual (z) and y for the phenotypic value in the resident population ($\bar{z}$).

Although the conditions favoring polymorphism are left unchanged by diploidy under the above assumptions, evolution after the onset of evolutionary branching will likely differ between haploid and diploid populations owing to heterozygotes with intermediate phenotypes in the latter. Take for instance the two habitats (or two niches) model explored in section 4.3.1. In a haploid population, the evolutionary divergence of two alleles leads to the coexistence of two morphs, each better adapted to one of the habitats. In sexual diploids, by contrast, mating between homozygotes for these diverged alleles creates a third, intermediate heterozygote. Evolution in this case leads one of the three resulting types – either one of the two homozygotes or the heterozygote – to be maladapted to both habitats (which type evolves to be maladapted depends on model parameters; Kisdi and Geritz, 1999). As a result of the genetic constraint of additive effects, a proportion of the phenotypic polymorphism in the population is therefore non-adaptive. This in turn creates selection on the genetic architecture of the trait to reduce such genetic load, e.g. through the evolution of dominance (Van Dooren, 1999). More generally, genetic constraints that lead to a genetic load favour

the evolution modifiers of recombination, epistasis, or gene duplication, that should eventually lead to a phenotypic polymorphism mirroring that seen in haploid asexuals, where such genetic constraints are absent.

6.1.2 Average heterozygote advantage when there are multiple phenotypes

Next, we consider invasion fitness when it depends on multiple phenotypes that are influenced by the evolving locus. Specifically, we assume that the biological scenario is such that invasion fitness can be written as

$$\rho(x,y) = W(\pi(x,y), e(y)) \tag{52}$$

where

$$\pi(x,y) = \Pi(z_1(x,y), z_2(x,y)) \tag{53}$$

is the payoff to an heterozygote showing phenotypes $z_1(x,y)$ and $z_2(x,y)$, and $e(y)$ is a regulating variable determined by the resident population (where individuals are homozygotes for the resident allele y). This form of invasion fitness, which is functionally equivalent to eq. (35), could for instance capture a situation where the fecundity of an individual depends on two of its phenotypes $z_1(x,y)$ and $z_2(x,y)$ (e.g. expression levels of two genes). Whatever the specific scenario, $z_1(x,y)$ and $z_2(x,y)$ are genetically correlated because they both depend on allelic values x and y. We assume that this dependency is given by

$$z_i(x,y) = \frac{g_i(x) + g_i(y)}{2}, \tag{54}$$

where $g_i(h)$ are the effects of an allele with genetic value $h \in \{x,y\}$ on phenotype i. In other words, we assume additive genetic effects on the phenotype between homologous loci (which will allow us to connect to a number of existing results). These effects however may translate into non-additive effects on fitness.

Substituting eqs. (53)–(54) into eq. (36) shows that any singularity x^* must satisfy

$$\frac{\partial \Pi(z_1,z_2)}{\partial z_1}\frac{\mathrm{d}g_1(x)}{\mathrm{d}x}\bigg|_{\substack{x=x^*\\ z_1=z_1(x^*,x^*)\\ z_2=z_2(x^*,x^*)}} + \frac{\partial \Pi(z_1,z_2)}{\partial z_2}\frac{\mathrm{d}g_2(x)}{\mathrm{d}x}\bigg|_{\substack{x=x^*\\ z_1=z_1(x^*,x^*)\\ z_2=z_2(x^*,x^*)}} = 0. \tag{55}$$

This says that at x^*, the effects on fitness of varying allelic value via phenotypes 1 and 2 must cancel each other out. Consider for example the case where fitness increases with both phenotypes (i.e. $\partial\Pi/\partial z_1 > 0$ and $\partial\Pi/\partial z_2 > 0$). Eq. (55) then tells us that for a singular trait value x^* to exist, phenotypes 1 and 2 need to change in opposite directions with x, namely $g_1'(x)$ and $g_2'(x)$ have opposite signs, such that a change in x has antagonistic effects on fitness via phenotypes 1 and 2 – antagonistic effects which are balanced at x^*.

If we substitute eqs. (53)–(54) into eq. (37) and use eq. (55), we readily obtain that the interaction coefficient can be expressed as $I_\pi(y) = L_\pi(x^*)K_\pi(x^*)/2$ where $L_\pi(x^*) = g_2'(x^*)^2 \geq 0$, such that the sign of trait-

dependent selection is governed by

$$K_\pi(x^*) = \frac{1}{2}\left[M(x^*)^2\frac{\partial^2\Pi(z_1,z_2)}{\partial z_1^2} + \frac{\partial^2\Pi(z_1,z_2)}{\partial z_2^2} - M(x^*)\frac{\partial^2\Pi(z_1,z_2)}{\partial z_1\partial z_2}\right]_{\substack{y=x^* \\ z_1=z_1(y,y) \\ z_2=z_2(y,y)}}. \tag{56}$$

Here, $M(x^*) = (\partial\Pi/\partial z_2)/(\partial\Pi/\partial z_1)|_{y=x^*}$ gives the rate at which phenotype 1 can be substituted for phenotype 2 without changing payoff (i.e. the change in the number of units of phenotype 1 that correspond to the effect on payoff of a change in one unit of phenotype 2). Eq. (56) shows that NTDS tends to occur when the payoff function decelerates with phenotypes ($\partial^2\Pi/\partial z_i < 0$ for $i \in \{1,2\}$) and phenotypes have complementary effects on payoff ($\partial^2\Pi/\partial z_1\partial z_2 > 0$). This is because such situations lead to heterozygote advantage, i.e. where heterozygotes for divergent allelic effects close to the singular allelic value x^* have greater fitness that either homozygotes. This occurs because heterozygotes benefit from a form of averaging across phenotypes here. When alleles contribute additively to two phenotypes (eq. 54), and payoff decelerates with both phenotypes in a concave way, heterozygotes—who express an intermediate phenotype—can have greater fitness than either homozygote when averaged over both phenotypes (see Fig. 4 panel A for a graphical argument). When the payoff exhibit complementary effects, heterozygotes necessarily gain larger payoff (see Fig. 4 panel B for a graphical argument). Heterozygote advantage thus arises because the contribution of an allele to fitness is context-dependent here, meaning that its effect varies across different phenotypes.

This mechanism is in fact similar to the one described in the seminal SAS-CFF population genetic model ("Stochastic Additive Scale - Concave Fitness Function", Gillespie, 1978, 1991). That model assumes that alleles at an homologous locus contribute additively to physiological activity (which is a phenotype in our model, eq. 54) in a fashion that varies stochastically with environmental conditions (say phenotype 1 in environment 1 and phenotype 2 in environment 2); and that fitness increases in a decelerating – concave – manner with physiological activity. To see the connection with our approach here, consider that eq. (56) applies to this SAS-CFF scenario with Π taken as the geometric mean fecundity of the mutant and the regulating variable e as the geometric mean fecundity of a resident individual under a Wright-Fisher process (as assumed in Gillespie, 1978, 1991). While eq. (56) applies to a broader set of situations, its connection to the SAS-CFF model provides a couple of insights. Firstly, it reinforces the idea that heterozygote advantage should arise from context-dependent allelic effect and averaging over different contexts. In the original SAS-CFF model (Gillespie, 1978), this averaging is temporal (over changing environments), but the same principle applies to averaging across multiple phenotype or biological functions. Secondly, since the SAS-CFF model is closely linked to the storage effect (Schreiber, 2020), this connection extends to our framework and can be phrased as follows. When conditions vary, diploidy allows alleles that are poorly adapted to current conditions to persist by buffering them in heterozygotes. This buffering effect can prevent the loss of alleles that may be beneficial under different environmental or functional conditions. In turn, the adaptive polymorphism that evolves under the above model leads to the divergence of two morphs, each showing a better fit for these different environmental or functional conditions, showing heterozygote advantage on average.

To illustrate these concepts with a concrete example, we consider the scenario envisaged by Gillespie (1978) where the environment changes each generation, being in state 1 with probability q and in state 2 with probability $1-q$ so that the fecundity is given by $\Pi(z_1, z_2) = z_1^q z_2^{1-q}$. We let z_1 and z_2 be the physiological activity of a gene in environments 1 and 2, respectively; and assume that activity increases with allelic value in environment 1, say $g_1(y) = y^{u_1}$, and decreases with allelic value in environment 2, say $g_2(y) = (1-y)^{u_2}$ (with $0 < y < 1$). In that case, $K_\pi(x^*) = -[(qu_1+(1-q)u_2)^2]/[4(1-q)q] < 0$, and it can then be checked that there is a range of parameter values leading to evolutionary branching and the coexistence of two morphs: one that increases activity in environment 1 and another in environment 2, leading to heterozygote advantage on average over the two environments (see Appendix D.2 for more details). See also Siljestam and Rueffler (2024) for a conceptually similar model that leads to poly-allelic polymorphism when there are more than two biological functions (or environments) over which to average.

6.2 Genetic associations between individuals under limited dispersal

6.2.1 Negative trait-dependent kin selection

We now relax the assumption that the population is well-mixed and consider the case where owing to limited dispersal, individuals interacting with one another show genetic associations generated by identity-by-descent, i.e., individuals are related (Hamilton, 1964; Michod, 1982 and see Rousset, 2004 for a review). When relatedness is high, this means that a globally rare mutant is nevertheless likely to interacts with other mutants locally. Relatedness should therefore modulate trait-dependent selection depending on the fitness effects of mutant-mutant interactions. To understand this, we analyse $I(x^*)$ in the infinite island model of dispersal with two individuals per island (or patch or deme). For simplicity and to focus on genetic interactions among individuals, we assume individuals are haploid and reproduce clonally. Invasion fitness in this case can be written as

$$\rho(x, y) = r(x, y)w(x, x, y) + [1 - r(x, y)]w(x, y, y), \tag{57}$$

where $r(x, y)$ is relatedness, that is, the probability that a rare mutant x in a resident y population has another mutant as patch neighbor (conversely, it has a resident as patch neighbor with probability $1-r(x, y)$); $w(x, x, y)$ is the individual fitness of a focal mutant x (first argument of w) when its patch neighbor is a mutant x (second argument), and the rest of the population in other patches consists of residents y (third argument); similarly, $w(x, y, y)$ is the fitness of the mutant when its patch neighbor is a resident y (second argument). Eq. (57) is a special case of the invasion fitness in a patch structured population with arbitrary patch size (e.g., Mullon et al., 2016; Van Cleve, 2023; Box 1 in Alger and Lehmann, 2023).

Plugging eq. (57) into eq. (4), we first recover the classical result that the singular trait value x^* must be

such that the inclusive fitness effect, i.e. Hamilton's rule in marginal form, vanishes:

$$\underbrace{\left.\frac{\partial w(x,y,y)}{\partial x}\right|_{x=y=x^*}}_{-c} + r(x^*,x^*)\underbrace{\left.\frac{\partial w(x,y,x^*)}{\partial y}\right|_{x=y=x^*}}_{b} = 0 \tag{58}$$

(e.g., Rousset, 2004). Therefore at x^*, the direct fitness effects of the trait (often conceived as a cost $-c$) must balance the relatedness-weighted indirect effects (i.e. the marginal effect a trait change in a neighbour has on an individual's fitness; typically conceived as a benefit b), where relatedness here is evaluated in a population monomorphic for x^*. The equilibrium x^* can thus be seen as the result of a trade-off between different components of inclusive fitness: that is, between the effect an individual has on its own fitness and the effect it has on the fitness of relatives.

Substituting eq. (57) into eq. (3), we obtain that the interaction coefficient can be decomposed as

$$I(x^*) = I_{\mathrm{w}}(x^*) + I_{\mathrm{q}}(x^*), \tag{59}$$

where

$$\begin{aligned} I_{\mathrm{w}}(x^*) &= r(x^*,x^*)\frac{\partial^2 w(x^*,x,y)}{\partial x \partial y} + [1 - r(x^*,x^*)]\frac{\partial^2 w(x,y,x^*)}{\partial x \partial y} + \frac{\partial^2 w(x,x^*,y)}{\partial x \partial y} \\ I_{\mathrm{q}}(x^*) &= \frac{\partial w(x^*,y,x^*)}{\partial y}\left(\frac{\partial r(x^*,y)}{\partial y} - \frac{\partial r(x,x^*)}{\partial x}\right), \end{aligned} \tag{60}$$

each capturing a pathway to negative trait-dependent selection.

The first term, $I_{\mathrm{w}}(x^*)$, captures how invasion fitness changes when the mutant and resident traits are varied simultaneously in the same direction, while keeping relatedness constant– hence relatedness $r(x^*,x^*)$ being evaluated in a population monomorphic for x^*. This term shows NTDS is weakened in populations showing higher relatedness when joint trait changes in a focal and its neighbour have negative fitness effects (i.e. where $\partial^2 w(x,y,x^*)/\partial x \partial y < 0$). Such a fitness effect could arise for various reasons. One relevant instance would be when the trait mediates a social interaction among neighbours that shows complementary or negative synergistic effects (as given by eq. 37, e.g. snowdrift games - Doebeli et al., 2004; Wakano and Lehmann, 2014 - or trait based competition for limiting resources leading to character displacement - Ajar, 2003; Schmid et al., 2024). In that case, high relatedness makes the maintenance of polymorphism more difficult simply because it makes it more difficult for social partners to show differentiated phenotypes, and therefore enjoy the inherent benefit that being different yields under this type of interaction. Kin selection here opposes NTDS and this effect explains why several studies have found that limited dispersal tends to make evolutionary branching more difficult (e.g. Day, 2001; Ajar, 2003; Wakano and Lehmann, 2014; Mullon and Lehmann, 2019; Schmid et al., 2024).

In contrast, the second term of eq. (59), $I_{\mathrm{q}}(x^*)$, reveals how kin selection can contribute to NTDS through

relatedness mediated effects. Specifically, $I_{\mathrm{q}}(x^*)$ tells us that trait-dependent selection tends to be negative when

$$\frac{\partial w(x^*,y,x^*)}{\partial y}\frac{\partial r(x,x^*)}{\partial x} > \frac{\partial w(x^*,y,x^*)}{\partial y}\frac{\partial r(x^*,y)}{\partial y} \tag{61}$$

holds. Why this is can be understood by considering a trait that has positive indirect fitness effects, i.e. that increases the fitness of its neighbour ($\partial w(x^*,y,x^*)/\partial y > 0$). In this case, ineq. (61) reduces to $\partial r(x,x^*)/\partial x > \partial r(x^*,y)/\partial y$. This condition states that a trait change in the mutant increases relatedness among mutants more strongly (left-hand side) than a similar trait change in the resident population would (right-hand side). As a result, the inclusive fitness benefits that mutants obtain via indirect effects are reduced if the resident trait value changes similarly to the mutant, thus contributing to NTDS.

6.2.2 Example: evolution of a social-dispersal syndrome

As an example of relatedness modulated NTDS, let us consider a situation where the trait constitutes a form of helping that increases the fecundity of one's neighbour, and that its costs are over dispersal such that an individual who invests more resources to help one's neighbour has fewer resources to commit to its dispersing offspring. Trait expression therefore has positive indirect fitness benefits (via fecundity) and positive effects on relatedness (via dispersal), especially when rare. For simplicity, we assume a Moran type life cycle whereby exactly one individual dies per life-cycle iteration in each patch, such that individual fitness takes the form

$$w(x,x_{\mathrm{n}},y) = \frac{1}{2} + \frac{1}{2}\left[\frac{(1-z_1(x))z_2(x_{\mathrm{n}})}{\left[(1-z_1(x))z_2(x_{\mathrm{n}}) + (1-z_1(x_{\mathrm{n}}))z_2(x)\right]/2 + z_1(y)z_2(y)} + \frac{z_1(x)z_2(x_{\mathrm{n}})}{z_2(y)}\right], \tag{62}$$

where $z_1(x)$ is the probability of dispersal of an individual expressing trait value x (which we assume is a decreasing function) and $z_2(x_{\mathrm{n}})$ is the fecundity of an individual whose neighbour expresses trait x_{n} (an increasing function of x_{n}). Therefore, the first term within square bracket in eq. (62) is the philopatric component of fitness, i.e. the expected number of offspring that establish in the natal patch, and the second term is the dispersal component, i.e. the expected number of offspring that establish in other patches. Using Box 2 of Alger and Lehmann, 2023, we obtain that relatedness for this model can be expressed as

$$r(x,y) = \frac{(1-z_1(x))\Big((1-z_1(x))z_2(x) + z_1(y)z_2(y)\Big)}{z_2(x)\Big((1-z_1(x))^2 + z_1(y) - z_1(y)^2\Big) + 2z_1(y)z_2(y)\Big(1 - z_1(x) + z_1(y)\Big)} \tag{63}$$

(strictly speaking, this is a relatedness proxy but because it does not change the sign of invasion fitness around one, it is sufficient for our purpose here).

Analysis of this model does not give rise to a simple expression for the interaction coefficient $I(x^*)$ from which useful information can be extracted, but a straightforward numerical approach with $z_1(x) = 1 - x$ and $z_2(x_{\mathrm{n}}) = 3x_{\mathrm{n}}^2 + 0.3x_{\mathrm{n}} + 1.05$ shows that at $x^* = 0.38$, we have $I(x^*) < 0$ while $I_{\mathrm{q}}(x^*) < 0 < I_{\mathrm{w}}(x^*)$, i.e. the population experiences NTDS due to relatedness effects only. Additionally, the population undergoes

evolutionary branching leading to the coexistence of two differentiated morphs that show a social-dispersal syndrome: (i) one that helps and thus tends to remain philopatric (expressing $x \approx 0.65$); and (ii) another that does not help and disperses (expressing $x = 0$). Trait-dependent selection can be understood by considering the mutant and resident population both being more selfish and thus more prone to dispersal simultaneously. Such a mutant benefits by dispersing away and thereby avoiding interactions with other mutants. However, if the resident population simultaneously becomes more selfish and dispersal-prone, these benefits vanish because resident individuals help less anyway and also disperse, reducing the mutant's relative ability to preferentially interact with non-relatives. A handful models have reported the evolution of similar social-dispersal syndromes owing to kin selection (Mullon et al., 2018; Prigent and Mullon, 2023) but beyond these, there remain few analyses of how effects on relatedness can generate NTDS. This is in spite of the fact that there are conceivably many other pathways via which traits can influence relatedness and thus experience NTDS due to this effect, especially traits related to gene flow such as those that affect budding dispersal, dispersal success, or establishment probability, as well as those that modulate individual survival or group extinction.

6.3 The role of modifiers of genetic context in NTDS

We saw in the previous section how group structure and the heterogeneity it entails allow for a form of NTDS that is relatedness-mediated. There, it arose when trait expression increased individual fitness in one context (e.g. when being in a group with individuals expressing a similar trait due to identity-by-descent) and the population trait moving in the same direction reduces the probability of individuals being in that context (e.g. due to dispersal decreasing identity-by-descent). This mechanism rests on the existence of heterogeneity in the population: mutant individuals do not interact randomly but instead experience different fitness effects depending on the local genetic composition of their group or patch. More generally, population heterogeneity and the existence of different contexts in which individuals can find themselves open new pathways for NTDS: it can arise whenever a trait influences both the contexts an individual experiences and fitness within those contexts.

Box 3 formalises this idea by considering populations structured into discrete classes to capture heterogeneity either within or between generations. These classes may correspond to different genetic backgrounds, phenotypic states, environmental conditions, or stages in an individual's life cycle. The general form of the interaction coefficient in such populations (eqs. 3.E–3.F) shows that trait-dependent selection is influenced not only by how population trait expression affects fitness within each class but also how it modifies the class structure. In particular, NTDS emerges when an increase in the resident trait value reduces either (i) the probability that mutants occur in the class in which their fitness is increased or (ii) the reproductive value of individuals produced by mutants in that class (see Box 3 for further details). This notion should extend to a broader range of heterogeneities than those considered in Box 3 (since invasion fitness can usually be

expressed as a weighted average over the distribution of contexts in which mutant and resident individuals occur, Lehmann et al., 2016).

The mechanism we have described here should be especially relevant to the maintenance of polymorphism in traits that influence how individuals experience environmental or demographic variation, such as sex ratio, habitat choice, bet-hedging mechanisms, aging, or life-history switches. Consider for instance a trait influencing habitat choice and reproductive success in that habitat, such that mutants expressing greater trait value preferentially settle in a specific habitat that confers them higher fitness (Ravigné et al., 2009). However, if the resident population also expresses higher trait values and preferentially settles in the same habitat, this increases local competition and thus reduces the per capita benefit of the trait, leading to NTDS. More broadly, these considerations show how population heterogeneity should favor diversity by allowing different morphs to specialise in different contexts—favouring the coexistence of different trait values that associate their reproductive success with the conditions where they perform best (see also Avila and Mullon, 2023 for similar considerations of context-dependence for disruptive selection).

7 Discussion

Polymorphism is widespread in nature. Numerous models from population genetics, evolutionary ecology, and game theory have proposed mechanisms by which polymorphisms can be maintained by natural selection. Here, we have aimed to organise these mechanisms into a common framework that clarifies when and what kind of adaptive polymorphism can be expected under the gradual evolution of quantitative traits.

7.1 Monomorphic instability, the role of NTDS, and its underlying mechanisms

One general insight we have found useful is that an adaptive polymorphism, whose definition we based on considerations from branching processes (eq. 2), is formally equivalent to a Nash equilibrium of a population invasion fitness game (Box 1). This allows to lift results from game theory and indicates that identifying adaptive polymorphisms favoured by long-term evolution is, in general, a computationally hard problem. Consequently, it is often more productive to assess the evolutionary instability of monomorphism to identify conditions that favour the emergence of polymorphism. For such instability to arise, the selection interaction coefficient I must be negative (eq. 6). A negative value of I means that selection on a mutant is reduced when the population expresses a similar trait value. Put differently, invasion fitness of a mutant increases when the mutant and resident trait values diverge. This negative trait-dependent selection (NTDS) is a necessary condition for adaptive polymorphism (eq. 6). As such, NTDS provides a general pathway for the emergence of polymorphism under gradual evolution, independent of specific biological mechanisms.

To clarify the biological origins of NTDS, we structured invasion fitness around two classes of fitness factors

(section 3). Production factors are inputs to an individual's survival and reproduction that depend directly on its own trait (e.g. baseline fecundity, germination rate). Regulating factors, meanwhile, are inputs to survival and reproduction that depend on the traits expressed by conspecifics (e.g. local density, other's behaviour). These two types of factors can act at different biological levels: from phenotypes, to payoffs in interactions, to invasion fitness (Fig. 1).

When regulating factors at the level of invasion fitness depend solely on the resident trait (e.g. eq. 20), our model captures ecologically mediated interactions. Classical mechanisms promoting polymorphism through ecological interactions include spatial and temporal heterogeneity in fitness components (e.g. Levene, 1953; Karlin and McGregor, 1972; Gillespie, 1991; Barton et al., 2007; Begon and Townsend, 2021). In this context, we used our approach to confirm that temporal heterogeneity is generally less conducive to adaptive intraspecific polymorphism than spatial heterogeneity (compare eqs. 28 and 33). We additionally showed that polymorphism under ecologically mediated interactions requires trait-specific responses to production factors, and a decoupling between production and regulating factors across contexts (eq. 24). This decoupling enables niche partitioning, with different trait values adapting to different environments associated with different regulating variables. While all these conclusions are well established, previous demonstrations in population genetics and evolutionary ecology often relied on more involved arguments, such as diffusion approximations (Gillespie, 1991) or coexistence theory (Chesson, 2000a; Barabas et al., 2008), and those using invasion analysis did not analyze directly the interaction coefficient (Svardal et al., 2015).

Within ecologically mediated interactions, we uncovered a previously undescribed life-history polymorphism. Although the fecundity–survival trade-off is well recognised, it is rarely examined in the context of polymorphism in life-history strategies (e.g. Charnov, 1993; Case, 2000). We showed that limited dispersal, which generates local kin competition, enables the coexistence of a morph with low survival and high fecundity alongside a quasi-immortal morph with low fecundity. NTDS arises because increased fecundity raises local offspring density, intensifying kin competition and reducing the fitness of the more fecund morph. This favours allocation towards survival as an alternative strategy. Although we did not explicitly partition fitness in this way, one could therefore have interpreted the regulating factor in this example as local offspring density (the denominator in the middle term of eq. 29). On that view, the regulating factor would depend on the mutant trait.

Our general analysis in fact emphasises that regulating factors can depend on both resident and mutant traits (section 3.2.2), unlike earlier related formalisations of invasion fitness (e.g. Metz, 2011; Pásztor et al., 2016; Lion, 2018). This generalisation is important for capturing several biologically relevant scenarios: (1) behavioural responsiveness, where resident individuals adjust their actions in response to mutant traits; (2) diploid inheritance, where mutant alleles co-occur with resident alleles in heterozygotes; and (3) mutant–mutant interactions, as occur under limited dispersal, family interactions, or assortative mating. In such cases, NTDS can arise through interaction effects on regulating factors, without requiring direct effects on production factors. We illustrated these broader mechanisms through examples involving socially and

genetically mediated interactions.

Socially mediated interactions involve direct behavioural encounters between individuals (e.g. Maynard Smith, 1982; Akzcay and Van Cleve, 2012; McNamara and Leimar, 2020). When fitness increases with the payoff obtained from such interactions, and regulating factors at the fitness level depend only on the resident population (due to e.g. density dependent regulation), the interaction coefficient can be evaluated directly at the level of payoffs (eq. 37). However, the regulating variables that determine these payoffs may themselves depend on mutant traits. This occurs, for instance, when individuals modulate their behaviour in response to their partner's traits, i.e. under social responsiveness. Social responsiveness introduces additional pathways to NTDS, which can act independently of direct trait effects on payoffs (eq. 42).

These considerations over socially mediated interactions led us to identify novel adaptive polymorphisms in behavioural dispositions. Revisiting models of preference evolution (Heifetz et al., 2007a,b), we examined traits that bias how individuals evaluate the consequences of their own actions in pairwise interactions under complete information (section 5.3.2). These dispositions influence not only an individual's own behaviour but also, via social responsiveness, the behaviour of their partners. When interactions have antagonistic effects on payoff (as per eq. 46), we found that selection favours a dimorphism between "optimists," who overvalue their actions, and "pessimists," who undervalue them. A similar diversification occurs for other-regarding preferences, though the long-term evolution of such polymorphisms remain to be fully characterised (Appendix C.2.3). These types of preference evolution models warrant further investigation given their apparent potential to generate stable polymorphisms in motivations underlying social behaviour.

Finally, we considered how genetically mediated interactions can influence NTDS. One relevant instance of this are interactions between mutant and resident alleles within diploid individuals. Under random mating, conditions for the emergence of polymorphism depend solely on heterozygote fitness, making them equivalent to conditions identified in haploid or asexual systems (eq. 51). This implies that diploidy, in itself, does not change the ecological or social conditions under which gradual evolution leads to polymorphism. Therefore, if one's interest lies primarily in understanding adaptative polymorphism, explicit modelling of genetic architecture is unnecessary—a form of phenotypic gambit for adaptative trait variation. Nevertheless, diploidy can generate NTDS at the level of alleles when there are pleiotropic constraints over multiple production factors (e.g. survival and fecundity). In particular, such constraints can lead to overdominance, or heterozygote advantage, a classical mechanism maintaining genetic variation. Our analysis clarified that, under gradual evolution, complementary allelic effects within heterozygotes (rather than simply intermediate optimal phenotypes) are required for overdominance to drive diversification (eq. 56).

Another important type of genetic interaction arises when limited dispersal generates identity-by-descent between interacting individuals so that kin selection occurs (e.g. Hamilton, 1964; Michod, 1982; Rousset, 2004). Here, we revealed a form of trait-dependent kin selection, which when negative, can contribute to diversification. Specifically, our analysis demonstrates that NTDS can emerge when a trait simultaneously

increases the fitness of neighbouring individuals and local relatedness (eq. 61). Such traits effectively alter their own genetic context, creating conditions in which indirect fitness benefits are disproportionately high when the mutant and resident are diverged. More broadly, when population structure creates different contexts in which the trait can be expressed and influence fitness (e.g. sex, age or environmental structure), NTDS is increased if the trait expressed by an individual increases its reproductive success in a particular context and trait expression at the population level reduces the representation of individuals in this context (Box 3). This pathway, which appears to be underappreciated in the context of NTDS, should be particularly relevant for traits influencing habitat choice, senescence, and sex determination, as these traits naturally interface with context-dependent selection. In this sense, population heterogeneity creates distinct niches, allowing genetic effects to specialise within particular contexts.

Beyond all these specific cases, our analyses allow us to make some comments more broadly applicable to intraspecific polymorphism and to connect to conditions for maintenance of interspecific diversity, i.e. to coexistence theory of ecology. The starting point to these comments is that trade-offs between different biological functions or fitness factors, while commonly invoked as drivers of polymorphism , are not sufficient on their own (e.g. de Mazancourt and Dieckmann, 2004; Ito and Sasaki, 2016b). They are necessary, which is reflected by the selection gradient being zero at the singular strategy, inherently entailing a trade-off between fitness factors. However, as we emphasized throughout the emergence of polymorphism in quantitative traits also requires NTDS, which is captured by a negative interaction coefficient. These considerations connect directly to (modern) coexistence theory (e.g. Chesson, 2000b,a; Barabas et al., 2008), in which stable coexistence between species is shown to be facilitated by two complementary types of mechanisms. First, equalising mechanisms reduce average fitness differences between species, thereby slowing competitive exclusion. Second, stabilising mechanisms reduce the intensity of competition between different species relative to competition within species (e.g., resource partitioning, storage effect). Under gradual evolution, directional selection toward a singular strategy acts as an equalising mechanism: near such a singular point, all nearby morphs (analogous here to species in ecological theory) have nearly identical fitness. Meanwhile, a negative interaction coefficient captures the scope for stabilising mechanisms, since morphs having diverged trait values experience increased fitness and thus suffer less from competition among each other (eq. A-1 in Appendix for details).

7.2 Oligomorphic stability, limiting similarity, and saturated polymorphisms

Once selection favours polymorphism, an essential question is where the evolutionary dynamics will lead the population to in the long-run. Relatively little theoretical effort has been devoted to characterize these long-term outcomes of adaptive polymorphism in quantitative (or continuous) traits. Nevertheless, a useful insight is that such a polymorphism should be oligomorphic, that is only a discrete number of morphs should be able to coexist stably (Gyllenberg and Meszéna, 2005; Hingu et al., 2008).

Evolutionary convergence towards an uninvadable oligomorphism requires qualitative dissimilarities among morphs in how their fitness respond to population-level feedbacks. More specifically, for a coalition of morphs to act as a robust attractor of the evolutionary dynamics, each morph must experience feedback from individuals of its own type that differs qualitatively from the feedback it receives from other morphs (such that condition (12) is satisfied). This distinct feedback entails that the fitness of each morph reacts uniquely to variation in the environment brought by morph frequency changes. This condition (eq. 12) not only provides a practical means of predicting the number of morphs in a stable evolutionary coalition; it also makes it clear that a necessary condition for intraspecific adaptive polymorphism is a direct reflection of the limiting similarity principle from evolutionary ecology (the principle that stable coexistence between species depends on sufficient differentiation in their ecological niches or responses to regulating factors, Meszéna et al., 2006; Pásztor et al., 2016). Therefore, analogous to the conditions promoting monomorphic instability, the criteria for maintaining stable polymorphism within species recapitulate foundational principles that underlie interspecific coexistence.

In our illustrative examples of polymorphic evolution driven by ecological and social interactions, we often observed that the convergence stable coalition is saturated. Saturation implies that all opportunities for further trait-dependent diversification have been exhausted, meaning trait-dependent selection no longer drives evolutionary change (such that eq. 17 holds). While it has been established that saturation can occur by gradual evolution in scenarios where regulating factors are solely dependent on resident trait values (Kisdi and Geritz, 2016), our findings suggest that saturation can also obtain in contexts where regulating factors depend on mutant traits, as illustrated in our model on the evolution of behavioural dispositions (Appendix C.2.2 for details). Future work should therefore aim at establishing sufficient conditions under which adaptive evolutionary endpoints consist in saturated polymorphisms.

Knowing this would be particularly valuable because saturation implies that local uninvadability coincides precisely with convergence stability. Consequently, knowing that evolution leads to a saturated polymorphism would allow the evolutionarily stable coalition to be fully characterised by the selection gradient alone. Given that the most general biological representation of the selection gradient is the generalised marginal inclusive fitness effect for polymorphic populations (capable of capturing both class and group structure, Priklopil and Lehmann, 2021), this would solidify the unique role Hamilton's marginal rule plays in understanding the nature and process of adaptation of quantitative traits.

7.3 Conclusions

While many of the ideas concerning adaptive polymorphism discussed here are established in the literature, we hope that the framework developed in this paper clarifies some of their underlying logic and connects various threads of existing theory. By decomposing invasion fitness into production and regulating factors and focusing on negative trait-dependent selection (NTDS), we have provided necessary conditions for adaptive

polymorphism applicable across diverse biological systems. This approach allows for (i) a consistent criterion to identify the presence or absence of NTDS; (ii) a systematic way of categorising regulating factors at the phenotype, payoff, and fitness levels, clarifying their roles in driving NTDS; and (iii) a synthesis of ecological, social, and genetic scenarios that are conducive to NTDS and thus adaptive polymorphism. These illustrative scenarios, while not exhaustive, highlight the diverse origins of polymorphism. They reinforce the notion that given the hierarchical organisation of biological systems and the ubiquity of competition, negative trait-dependence and the resulting qualitative dissimilarity in fitness responses to regulating factors may be common rather than exceptional. If NTDS is indeed so common, it may hold the key to understanding why intra-specific trait diversity is so pervasive, continuously providing the fuel for evolutionary change.

Acknowledgments We thank Claus Rueffler for many useful discussions, providing references, and spotting an oddity within a previous version of one model.

Box 1. Uninvadable coalition, Nash equilibrium and computational complexity

Let us define

$$\overline{\rho}(p,q) = \int_{\mathcal{X}} \rho(x,q)p(x)\,\mathrm{d}x \tag{1.A}$$

as the average of the invasion fitness $\rho(x,q)$ over p when the distribution of types in the population is q (i.e., $\int_{\mathcal{X}} q(x)\,\mathrm{d}x = 1$). In the context of a population game (e.g. Sandholm, 2011; Anderson et al., 2013), $\rho(x,q)$ can be read as the payoff to a player with strategy x, when the population state is q, and $\overline{\rho}(p,q)$ the expected payoff to a player with mixed strategy $p \in \Delta(\mathcal{X})$ in that population (where $\Delta(\mathcal{X})$ denotes the set of probability measures over $\mathcal{X}$) when individuals in the population play mixed strategy $q \in \Delta(\mathcal{X})$). By definition, the mixed Nash equilibrium p^* to this game satisfies

$$p^* \in \underset{p\in\Delta(\mathcal{X})}{\arg\max}\, \overline{\rho}(p,p^*), \tag{1.B}$$

which entails that mixed strategy p^* is a best reply to self Anderson et al. (2013). The strategy p^* satisfies eq. (1.B) if and only if it satisfies eq. (2) (Anderson et al., 2013, Proposition 1, and eq. 1.B is sometimes taken as an alternative definition of Nash equilibrium). In the context of our model, this means that p^* is an uninvadable coalition if and only if it characterises a mixed strategy of a population game where payoff is given by average invasion fitness. This correspondence between the definition of an uninvadable coalition and Nash equilibrium allows us to lift two results from game theory to the problem of evaluating long-term evolutionary equilibria.

First, when $\mathcal{X}$ is compact and the payoff function $\rho(x,p) = \tilde{\rho}(x,\mathcal{P}(p),p)$ is continuous in p, then a Nash equilibrium is guaranteed to exist. This follows from applying established results to the best-reply correspondence $b(q) = \arg\max_{p\in\Delta(\mathcal{X})} \overline{\rho}(p,q)$ as is usually done in game theory for normal form games (Fudenberg and Tirole, 1991,Mas-Colell et al., 1995). When $\mathcal{X}$ is non-empty and compact, so will be $\Delta(\mathcal{X})$ (Alipantris and Border, 2006, Th. 15.11) and thus by Berge's maximum theorem (Alipantris and Border, 2006, Th. 17.31) $b(q)$ defines an upper hemi-continuous and non-empty compact-valued correspondence. Because, $b(q)$ is convex-valued (since $\overline{\rho}(p,q)$ is linear in p) and $\Delta(\mathcal{X})$ is a convex set (since the mixture of two probability distribution is a probability distribution), the conditions of Glicksberg's fixed point theorem are satisfied, and the set of equilibrium point $p^* \in b(p^*)$ is non-empty and compact (Alipantris and Border, 2006, Th. 17.55). Therefore, there is at least one uninvadable coalition in any evolutionary model that satisfy our main biological assumptions (see section 2), unless the set of traits $\mathcal{X}$ is unbounded. This uninvadable coalition may consist of an adaptive polymorphism or be a singleton trait, say x^*, in which case the uninvadable trait value satisfies $x^* \in \arg\max_{x\in\mathcal{X}} \rho(x,x^*)$ for invasion fitness $\rho(x,y)$ of a mutant x in a monomorphic resident population y. Note that this latter case corresponds to a symmetric Nash equilibrium in pure strategies and such an equilibrium exists if $\rho(x,y)$ is continuous, quasi-concave in its first argument, and $\mathcal{X}$ is compact (e.g. Fudenberg and Tirole, 1991, p. 34).

Second, finding a Nash equilibrium is computationally hard (e.g., Papadimitriou, 2007; Conitzer and Sandholm, 2008; Papadimitriou, 2014; Narahari, 2014, Theorem 2.1). Just finding the mixed Nash equilibrium in a two-player symmetric finite game belongs to the PPAD complexity class (e.g., Papadimitriou, 2007; Conitzer and Sandholm, 2008; Papadimitriou, 2014; Narahari, 2014, Theorem 2.1). There is strong evidence that problems in this class cannot be solved with a polynomial algorithm, i.e., PPAD≠P (Papadimitriou, 2007, p.39). While our population game does not necessarily reduce to a two-player symmetric game in mixed strategies (it can under certain cases), it is an infinite game and thus computational issues are likely to be harder. This does not mean that the uninvadable coalition can never be computed. If the invasion fitness has a special structure, it can be solved in certain cases. For instance, in "succinct games" (Papadimitriou, 2007), the analysis becomes more tractable, as seen in congestion games. In biological terms, this corresponds to models where competition for resources depends on some aggregate effect, such as foraging models with ideal free distribution (e.g., Křivan et al., 2008). The conclusion is twofold. First, evolutionary biologists should be aware of the computational complexity of evaluating a uninvadable coalition, and that results from algorithmic game theory may be useful (e.g., Papadimitriou, 2007; Conitzer and Sandholm, 2008; Papadimitriou, 2014; Narahari, 2014, Theorem 2.1). Second, identifying conditions necessary for the emergence of adaptive polymorphism without computing the long term equilibrium is a useful first-step to understanding polymorphism.

Box 2. Intensive versus extensive variables and explicit ecological models

When regulating factors are extensive variables they represent dynamical state variables that have reached some equilibrium (or more generally some attractor). For instance, when the regulating factors $e_1(y)$ and $e_2(y)$ in invasion fitness eq. (20) are extensive variables, they can be written to satisfy the following system of equations:

$$\begin{aligned} R_1(z_1(y), z_2(y), e_1(y), e_2(y)) &= 0 \\ R_2(z_1(y), z_2(y), e_1(y), e_2(y)) &= 0. \end{aligned} \tag{2.A}$$

Here, R_1 and R_2 are mappings that characterise the equilibrium of the regulating variables, which is assumed to be hyperbolically stable (see also Kisdi and Geritz, 2016, eq. 8). Since individuals interact with their environment via their production factor, the equilibrium values of both regulating factors $e_1(y)$ and $e_2(y)$ depend on both production factors $z_1(y)$ and $z_2(y)$. This could describe how competition for food resources $e_1(y)$ and breeding sites $e_2(y)$ is impacted by individuals' efforts in foraging $z_1(y)$ or reproduction $z_2(y)$.
An explicit example of an ecological map of the form eq. (2.A) and how extensive variables can be expressed in terms of intensive variables can be obtained as follows. Let us consider a population of iteroparous individuals that survive from one time step to the next with probability $z_1(x)$ and produce offspring according to a Beverton-Holt model of reproduction $z_2(x)/[1 + \gamma e_1(y) + \eta e_2(y)]$, where $z_2(x)$ is the baseline fecundity in the absence of density-dependence, $e_1(y)$ is the density of the focal species, $e_2(y)$ is the density of some other species, and γ and η are parameter tuning the importance of density-dependence. Invasion fitness of the focal species then takes the form

$$\rho(x, y) = z_1(x) + \frac{z_2(x)}{1 + \gamma e_1(y) + \eta e_2(y)}, \tag{2.B}$$

which is a Leslie-Gower model (Leslie and Gower, 1958, 1960); the discrete analogue of the continuous time Lotka-Voltera model (Leslie and Gower, 1960). Since both $e_1(y)$ and $e_2(y)$ are extensive variables, we need maps describing population dynamics at equilibrium in these two species: $e_1(y) = \rho(y, y)e_1(y)$ and $e_2(y) = w_2(y)e_2(y)$ where $w_2(y)$ is the fitness of an individual from the second species, which may depend on all quantities $z_1(y), z_2(y), e_1(y)$ and $e_2(y)$. Hence, in order to arrive at a system of the form of eq. (2.A), we can use

$$R_1(z_1(y), z_2(y), e_1(y), e_2(y)) = \rho(y, y) - 1 = z_1(y) + \frac{z_2(y)}{1 + \gamma e_1(y) + \eta e_2(y)} - 1 = 0 \tag{2.C}$$

for the focal species, while for the second species we can use

$$R_2(z_1(y), z_2(y), e_1(y), e_2(y)) = w_2(y) - 1. \tag{2.D}$$

These allow us to solve for the population dynamic equilibrium of the focal species regardless of the functional form of eq. (2.D). Indeed, by solving eq. (2.C) for $e_2(y)$ and by substituting into eq. (2.B) produces

$$\rho(x, y) = z_1(x) + (1 - z_1(y))\frac{z_2(x)}{z_2(y)}. \tag{2.E}$$

Thus, invasion fitness can be expressed in terms of mutant and resident production factors only, thus in terms of intensive variables. Interestingly, invasion fitness (2.E) can be seen as abiding to a lottery model for iteroparous reproduction with the following life-cycle. (a) A large but fixed number of adults produce a large number of juveniles and then each adult either survives or dies independently. (b) Density-dependent competition occurs among juveniles for the breeding spots vacated by the death of adults so that the population is regulated back to the fixed adult size. Such life-cycles are standard in population genetics. Assuming further $z_1(x) = 0$ gives the Wright-Fisher model (Ewens, 2004; Hartl and Clark, 2007). These connections showcase how standard models of population genetics do incorporate ecological feedbacks, even when these feedbacks are not explicitly specified (as well established, e.g., Crow and Kimura, 1970; Nagylaki, 1992).

Box 3. Class structured populations

We consider here heterogeneous populations where individuals can belong to a finite number of classes, where $\mathcal{C}$ denotes the set of possible classes (e.g., genetic background, condition, age, environmental state). Assuming that this heterogeneity occurs either only within-generations (e.g. stage) or only between-generations – in which case heterogeneity is independently and identically distributed (e.g. temporal fluctuations in climate) – invasion fitness is given by

$$\rho(x,y) = \begin{cases} \sum_{c'\in\mathcal{C}}\sum_{c\in\mathcal{C}} w_{c'c}(x,y)q_c(x,y) & \text{if only within generation heterogeneity} \quad (3.A) \\ \prod_{c\in\mathcal{C}} w_c(x,y)^{q_c(y)} & \text{if only between generation heterogeneity,} \quad (3.B) \end{cases}$$

where as before mutant individuals express trait value x while the resident population has trait y. In eq. (3.A), $w_{c'c}(x,y)$ is the expected number of offspring of class c' (possibly including the surviving self) produced by a mutant in class c, and $q_c(x,y)$ is the probability that a mutant is in class c (see Appendix E.1 or Lehmann et al., 2016, Box 3). In eq. (3.B) $w_c(x,y)$ is the expected number of offspring produced by a mutant in class or environmental state c and $q_c(y)$ is the probability of occurrence of environmental state $c \in \mathcal{C}$, which depends only on resident traits (as in standard invasion analysis, e.g., Ferrière and Gatto, 1995). Note that owing to between-generation heterogeneity, there is only type of offspring so $w_c(x,y)$ depends only on the environmental state c in which the individual reproduces. We show in Appendix E.3 how these class-specific quantities can be expressed in terms of regulation and production factors such that invasion fitness in both cases of heterogeneity can be conceptualised as a mapping from fitness factors to geometric growth ratio (i.e. as in eq. 18).

Using eqs. (3.A)-(3.B), the selection gradient can be expressed as

$$S(y) = \begin{cases} \sum_{c'\in\mathcal{C}}\sum_{c\in\mathcal{C}} v_{c'}(y)\left.\frac{\partial w_{c'c}(x,y)}{\partial x}\right|_{x=y} q_c(y,y) & \text{if only within generation heterogeneity} \quad (3.C) \\ \sum_{c\in\mathcal{C}} \left.\frac{\partial \log(w_c(x,y))}{\partial x}\right|_{x=y} q_c(y) & \text{if only between generation heterogeneity,} \quad (3.D) \end{cases}$$

where $v_c(y)$ is the reproductive value of an individual in class c in a (monomorphic) resident population y, satisfying $v_c(y) = \sum_{c'\in\mathcal{C}} v_{c'}(y)w_{c'c}(y,y)$ while $q_{c'}(y,y) = \sum_{c\in\mathcal{C}} w_{c'c}(y,y)q_c(y,y)$ (eq. 3.C is a standard result, e.g., Taylor and Frank, 1996; Rousset, 2004; Lehmann et al., 2016 and eq. 3.D is derived in Appendix E.2). The singular trait value x^* thus reflects a trade-off of fitness effects among different classes, which are appropriately weighted according to how often an individual is in class c (via $q_c(y,y)$ or $q_c(y)$) and offspring reproductive value under within generation heterogeneity (via $v_{c'}(y)$ in eq. 3.C).

The interaction coefficient, meanwhile, can be written as,

$$I(x^*) = \sum_{c'\in\mathcal{C}}\sum_{c\in\mathcal{C}} v_{c'}(x^*)q_c(x^*,x^*)\left[\frac{\partial^2 w_{c'c}(x,y)}{\partial x\partial y} + \frac{\partial w_{c'c}(x,y)}{\partial x}\left(\frac{\partial \log(v_{c'}(y)q_c(x,y))}{\partial y} - \frac{\partial \log(q_c(x,y))}{\partial x}\right)\right]_{x=y=x^*} \quad (3.E)$$

for within generation heterogeneity, and

$$I(x^*) = \sum_{c\in\mathcal{C}} q_c(x^*)\left[\frac{\partial^2 \log(w_c(x,y))}{\partial x\partial y} + \frac{\partial \log(w_c(x,y))}{\partial x}\frac{\partial \log(q_c(y))}{\partial y}\right]_{x=y=x^*} \quad (3.F)$$

for between generation heterogeneity (Appendix E for derivation). The first term of eqs. (3.E) and (3.F) shows that NTDS can simply be due to negative interaction effects on class specific fitness components (i.e. via $\partial^2 w_{c'c}(x,y)/(\partial x\partial y)$ and $\partial^2 \log(w_c(x,y))/(\partial x\partial y)$). More interestingly, the rest of eqs. (3.E) and (3.F) highlight how population heterogeneity opens new pathways for NTDS. These can be understood by assuming that trait expression increases individual fitness in a given class ($\partial w_{c'c}(x,y)/\partial x > 0$ or $\partial \log(w_c(x,y)/\partial x > 0$). An increases in the resident population trait value then tends to have negative effects on mutant fitness when such an increase reduces: (i) the probability that mutants belong to the class in which their fitness is enhanced (via $\partial \log(q_c(x,y))\partial y < 0$ in eq. 3.E or $\partial \log(q_c(y))\partial y < 0$ in eq. 3.F); and/or (ii) the reproductive value of offspring produced in that class under within generation heterogeneity (via $\partial \log(v_{c'}(y))\partial y < 0$ in eq. 3.E). We discuss these pathways more generally in section 6.3.

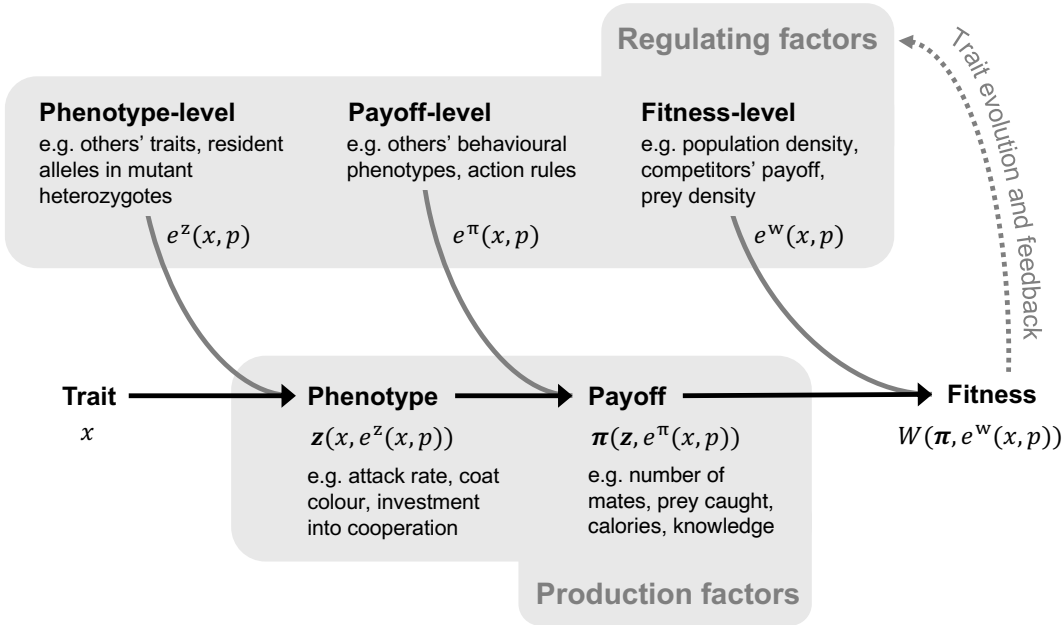

Figure 1: **Production and regulating fitness factors.** Diagram showing the construction of fitness, from genetic trait value x through production factors: phenotype and payoff. These various layers and their regulating factors offer different opportunities for population feedback as trait evolution leads to changes in regulating factors. Section 3.2 for details.

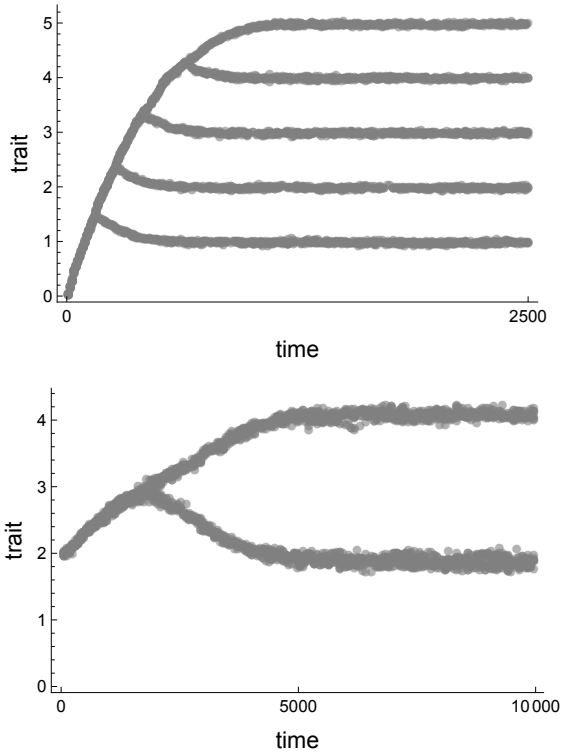

Figure 2: **Illustration of evolutionary branching and long term evolution under individual-based simulations for within generation heterogeneity in resource exploitation.** The two panels illustrate trait evolution under the generalization to 5 habitats of the resource specialisation fitness model (26) with production and regulating factors through morph x given by $z_i(x) = e_i(x) = e^{-(x-o_i)^2/(2\sigma^2)}$, where the optima are taken to be $o_i = i$, and habitat frequency is uniform (see eq. A-32 for the invasion fitness function of this model). Reproduction follows a Wright-Fisher process (lottery model) with $N = 5000$ individuals. Mutations occur with probability 0.05 and effects size are sampled from a Normal distribution with mean 0 and standard deviation 0.01. Top panel is for the case where $\sigma^2 = 0.04$ so that niche overlap is small. This results in a saturated uninvadable coalition: there are five morphs at the evolutionary equilibrium, each specializing on one habitat and with trait value corresponding to the habitat specific optimum. Evolution thus exhausts trait-dependent selection, whereby eq. (17) holds. Lower panel is for the case $\sigma^2 = 1$ where niche overlap is large. Then, the long-term polymorphism is unsaturated resulting in the coexistence of only two morphs in the uninvadable coalition, since with more morphs and large fitness overlap, the fitness of the different morphs is no longer sufficiently differentiated in how they respond to frequency-dependent selection.

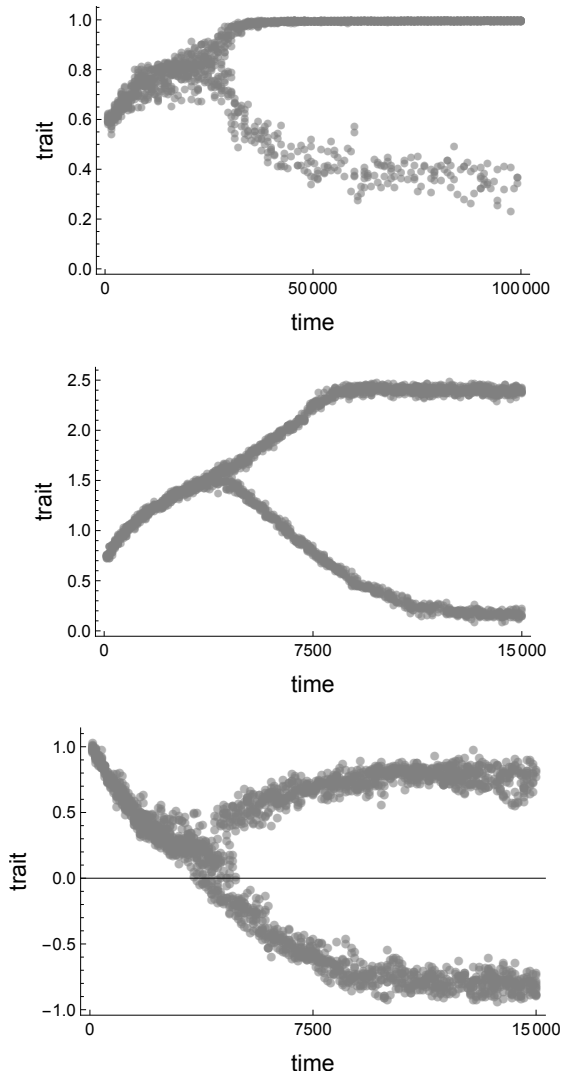

Figure 3: **Illustration of evolutionary branching and long term evolution under individual-based simulations for a sample of novel polymorphisms**. Top panel is for the fitness model eq. (29), which leads to the coexistence of an immortal morph investing all resources into survival and another one expressing a mixed strategy (the trait x is here investment into survival, with fraction $1-x$ invested into fecundity, see Appendix B.2.2 for the corresponding formal analysis). Middle panel is for payoff model eq. (43), which leads to the coexisting of one morph investing a lot into contests ("Hawk") and another one investing moderately ("Dove") (the trait x here is investment into contest, see Appendix C.2.1 for the corresponding formal analysis). Bottom panel is for the payoff model eqs. (46)–(48), which leads to the coexisting of optimists and pessimists (the trait x here is a self esteem bias, with $x > 0$ describing optimism and $x < 0$ pessimism, see Appendix C.2.2 for the corresponding formal analysis). In each case, reproduction follows a Wright-Fisher process (lottery model) with $N = 10000$ individuals. Mutations occur with probability 0.015 and effect sizes are sampled from a Normal distribution with mean 0 and standard deviation 0.01 (trait values are capped between zero and one in the top panel).

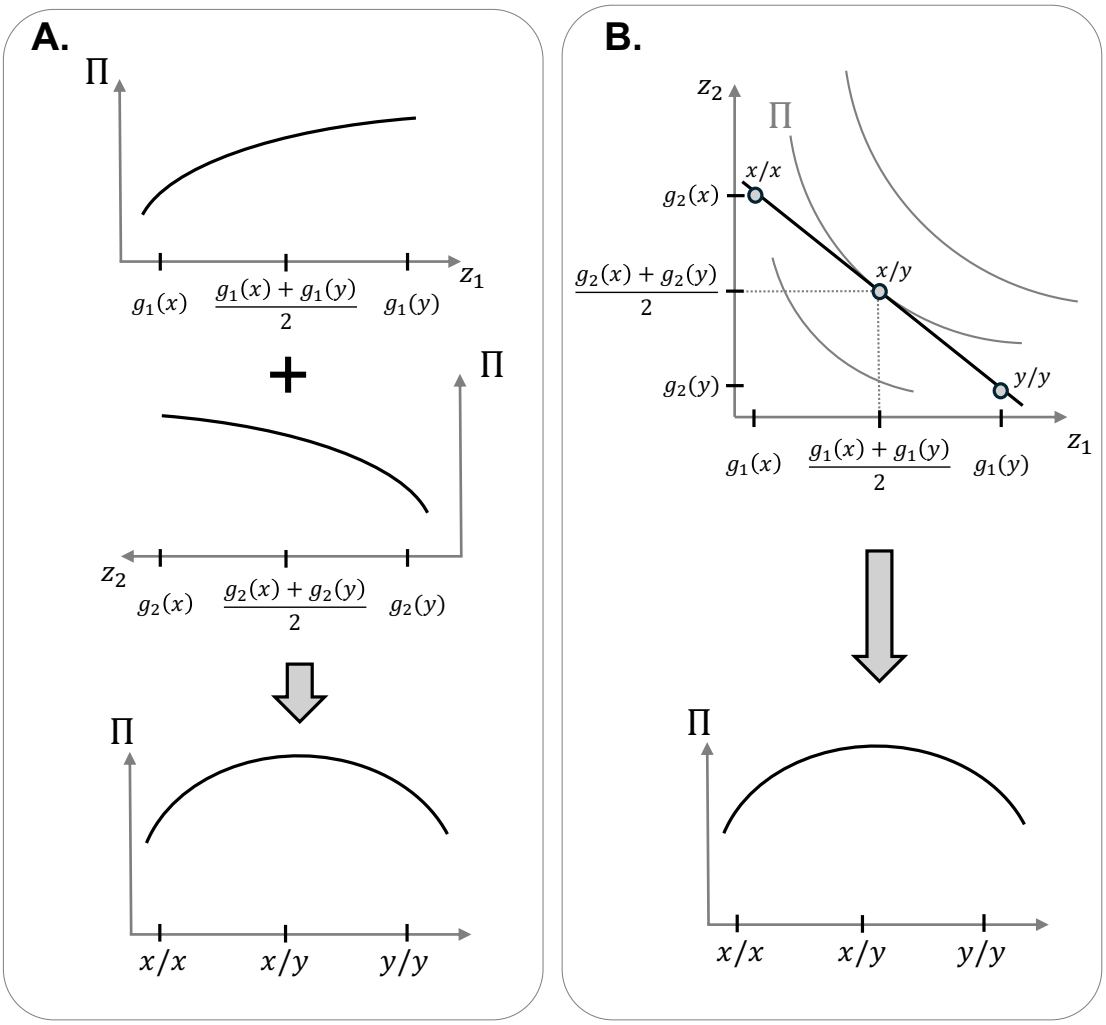

Figure 4: **A graphical illustration of overdominance origin under pleiotropic constraints. A**. Overdominance can arise when the payoff function $\Pi(z_1, z_2)$ decelerates with both phenotypes increasing (first two terms of eq. 56). The top and middle graphs show the payoff $\Pi(z_1, z_2)$ increasing in a decelerating manner with both z_1 (top) and z_2 (middle). When the payoff increases with both phenotypes, z_1 and z_2 must change in opposite directions near a singular trait value (as implied by eq. 55). The direction of the x-axes are reversed in the top and middle graphs to ensure genotype alignment: x/x homozygotes on the left, x/y heterozygotes in the centre, and y/y homozygotes on the right. Summing the effects across both phenotypes to compute the average payoff for each genotype leads to overdominance, with heterozygotes (x/y) achieving the highest payoff (captured by the two first terms in eq. 56). **B**. Overdominance can also emerge when phenotypes have complementary effects on payoff (last term of eq. 56). The top graph illustrates this situation when the payoff increases with both phenotypes, with payoff isoclines shown in grey. According to eq. (55), z_1 and z_2 change in opposite directions near a singular trait value, moving along tangents to a payoff contour (black lines, following the genotypes shown on graph). When phenotypes act complementarily on payoff, such changes along contour tangents lead to overdominance (compare black and grey curves and see bottom).

Appendix A Coexistence

Appendix A.1 Coexistence between dimorphism

Under gradual evolution, evolutionary convergence towards an invadable monomorphism x^* results in the population subsequently spiting into distinct morphs leading to their coexistence in a protected polymorphism (Eshel et al., 1997; Geritz et al., 1998). This coexistence can be understood by considering two morphs, say $x_1 = y + \delta \in \mathcal{X}$ and $x_2 = y - \delta\epsilon \in \mathcal{X}$, that are weakly diverged around y and where $|\delta| \ll 1$ and $\epsilon > 0$ assumed to be of order $O(1)$. These morphs will coexist if they can mutually invade one another, that is when $\rho(x_1, x_2) > 1$ and $\rho(x_2, x_1) > 1$ (e.g., Hofbauer and Sigmund, 1998; Turelli et al., 2001). Second order Taylor expanding $\rho(x_1, x_2)$ and $\rho(x_2, x_1)$ in δ around $\delta = 0$, and using the condition $\rho(y, y) = 1$ differentiated twice with respect to y, we get

$$\begin{aligned}\rho(x_1, x_2) &= 1 - \delta(1+\epsilon)\left[\epsilon\delta I(y) - \left(S(y) + \frac{\delta(1-\epsilon)}{2}H(y)\right)\right] + O(\delta^3)\\ \rho(x_2, x_1) &= 1 - \delta(1+\epsilon)\left[\delta I(y) + \left(S(y) + \frac{\delta(1-\epsilon)}{2}H(y)\right)\right] + O(\delta^3),\end{aligned} \tag{A-1}$$

where $S(y)$, $H(y)$ and $I(y)$ are given in eq. (3). The term $(S(y) + \delta(1-\epsilon)H(y)/2)$ has opposite effects on the invasion fitness of each morph: increasing fitness of x_2 and diminishing that x_1. Any mechanism that makes this term closer to zero can therefore be seen as an equalizing mechanism in the jargon of coexistence theory (Chesson, 2000b). These reduce inherent fitness differences among morphs thus favouring coexistence. From that perspective, directional selection leading to a singular trait value x^* such that $S(x^*) = 0$ can be seen as an equalizing mechanism. The term $I(y)$, in contrast, influences the invasion fitness of each morph in the same way. Any mechanism that makes $I(y)$ more negative thus acts as a stabilizing mechanism, favouring mutual invasion. When $I(y) > 0$, mutual invasion is not possible, i.e. $I(y) < 0$ is a necessary condition for mutual invasion. Additionally, when $\epsilon = 0$ (such that the two morphs are symmetrically distributed around y) and $y = x^*$ is singular, then $I(y) < 0$ also becomes a sufficient condition for mutual invasion.

Appendix A.2 Stability of oligomorphism

Appendix A.2.1 Hyperbolic stability of an uninvadable coalition

We here discuss the condition for the hyperbolic stability of an uninvadable coalition. For this, first recall eq. (7) and write it as

$$p_i = \tilde{\rho}(y_i, \mathbf{y}, \mathbf{p})p_i \qquad \text{for all } y_i \in \mathcal{P}(\mathbf{p}), \tag{A-2}$$

where the right-hand side describes the map of the dynamical system for change in morph frequencies. Expressing the map in vector form, taking its the derivative with respect to $\mathbf{p}$, and using to the fact that $\tilde{\rho}(y_i, \mathbf{y}, \mathbf{p}) = 1$ at equilibrium, we obtain that the Jacobian of the map of the dynamical system at $\mathbf{p}$ can be written as

$$\mathbf{I} + \mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}), \tag{A-3}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{D}(\mathbf{p})$ is the diagonal matrix with elements $\mathbf{p}$, and

$$\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = \begin{pmatrix} \dfrac{\partial \tilde{\rho}(y_1, \mathbf{y}, \mathbf{p})}{\partial p_1} & \cdots & \dfrac{\partial \tilde{\rho}(y_1, \mathbf{y}, \mathbf{p})}{\partial p_n} \\ \vdots & & \vdots \\ \dfrac{\partial \tilde{\rho}(y_n, \mathbf{y}, \mathbf{p})}{\partial p_1} & \cdots & \dfrac{\partial \tilde{\rho}(y_n, \mathbf{y}, \mathbf{p})}{\partial p_n} \end{pmatrix}. \tag{A-4}$$

Recall that the fixed point $\mathbf{p}$ is locally asymptotically stable if all eigenvalues of eq. (A-3) lie strictly inside the unit circle (i.e. $|\lambda| < 1$ for any eigenvalue λ of $\mathbf{I} + \mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$ where $|\cdot|$ denotes the modulus of a possibly complex eigenvalue). If any of the eigenvalues lie outside the unit circle, i.e., $|\lambda| > 1$, then the equilibrium point is unstable (Stuart and Humphries, 1996, Thm.1.3.7). Hence, a necessary condition for local asymptotic stability is that all the eigenvalues of eq. (A-3) belong to the unit circle. A fixed point $\mathbf{p}$ is said to be hyperbolic if none of the eigenvalues of the Jacobian of the dynamical system lie on the unit circle (Stuart and Humphries, 1996, Def.1.3.6). Therefore, a necessary and sufficient condition for a fixed point $\mathbf{p}$ to be hyperbolically stable is that any eigenvalue of eq. (A-3) lie strictly inside the unit circle.

Because any eigenvalue of eq. (A-3) is equal to 1 plus the eigenvalue of $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$, it follows that the fixed point $\mathbf{p}$ is hyperbolically stable if (and only if) the modulus of the eigenvalues of $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$ is smaller than zero [i.e. $|\lambda| < 0$ for any eigenvalue λ of $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$]. Because $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$ is a matrix with real entries, then its characteristic polynomial has real coefficients, so its complex eigenvalues (if any) come in conjugate pairs (Strang, 2006, p. 318). Hence, if $|\lambda| \neq 0$ for any eigenvalue λ of $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$, we have det $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$, which follows from the observation that the matrix determinant is equal to the product of all matrix eigenvalues and that the product of any conjugate pair involving eigenvalue λ is equal to $(|\lambda|)^2$. Then, owing to the facts (i) det $\mathbf{D}(\mathbf{p}) \neq 0$ because the diagonal elements are non zero and (ii) det $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$=det $\mathbf{D}(\mathbf{p})$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$, the condition $\mathbf{D}(\mathbf{p})\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$ implies det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$. In other words, a necessary condition for the hyperbolic stability of an equilibrium $\mathbf{p}$ in the interior of the state space is that det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$.

Appendix A.2.2 Convergence stability of an uninvadable coalition

We here explain the conditions for convergence stability and derive eq. (14) of the main text. Under the assumptions detailed in section 2.1 and the assumption that $\mathbf{p}(\mathbf{y})$ satisfying eq. (7) is hyperbolically stable, the selection coefficient $S_i(\mathbf{y}) = \partial\rho(x, \mathbf{y})/\partial x|_{x=y_i}$ tells us whether a mutant $y_i + \delta$ arising from morph y_i in population state $(y_1, y_{i-1}, y_i, y_{i+1}, ..., y_n)$ goes extinct or substitutes the ancestral type—leading to a new

polymorphic population state $(y_1, y_{i-1}, y_i + \delta, y_{i+1}, ..., y_n)$ (Cai and Geritz, 2020; Priklopil and Lehmann, 2020, 2021). Specifically, the rate of change of morph y_i in the population is of the form

$$\frac{\mathrm{d}y_i(t)}{\mathrm{d}t} = v_i(\mathbf{y}(t))S_i(\mathbf{y}(t)), \tag{A-5}$$

where $v_i(\mathbf{y}(t)) > 0$ is some measure of genetic variance in the subpopulation of morph i. Eq. (A-5) can typically take two types of interpretations, either as a rate of trait change under a trait substitution sequence or as the rate of mean trait change in a quantitative genetics model (e.g., Leimar, 2009; Lehmann and Rousset, 2014 for a review). More often than not, the variances $v_i(\mathbf{y}(t))$ differ among different morphs (due to e.g. different subpopulation sizes of morphs or selection, Sasaki and Dieckmann, 2011).

For an equilibrium $\mathbf{x}^*$ satisfying $S_i(\mathbf{x}^*) = 0$ for all $i \in \{1, 2, ...n\}$ to be a local attractor of the evolutionary dynamics (be a locally asymptotically stable fixed point of eq. A-5) requires that the Jacobian matrix of the flow of the dynamical system eq. (A-5), i.e. that the matrix

$$\mathbf{V}\mathbf{J}^*_{\mathbf{y}}(\mathbf{s}), \tag{A-6}$$

where $\mathbf{V}$ is a diagonal matrix with (i, i)-element $v_i(\mathbf{y}(t) > 0$ and $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ is given by eq. (13), has all eigenvalues with negative real parts. Then, in force of the the results in Appendix A.2.3, we have that a necessary condition for convergence stability is that all eigenvalues of $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$ have negative real parts. Sufficient conditions for convergence stability include (i) that $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ is negative definite and (ii) that $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ is sign-similar matrix (see Appendix A.2.3 for a characterization).

We now show that $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ can be decomposed according to eqs. (14)-(16) of the main text. For this, let us first substitute eq. (8) into the selection gradient on morph i (eq. 9) and write it as

$$S_i(\mathbf{y}) = \left.\frac{\partial \tilde{\rho}(x, \mathbf{y}, \mathbf{p}(\mathbf{y}))}{\partial x}\right|_{x=y_i} = \tilde{S}_i(\mathbf{y}, \mathbf{p}(\mathbf{y})), \tag{A-7}$$

where $\tilde{S}_i(\mathbf{y}, \mathbf{p}(\mathbf{y}))$ is written to emphasize the dependence of the selection gradient on morph frequencies. From there, we have that

$$\frac{\partial S_i(\mathbf{y})}{\partial y_j} = \frac{\partial \tilde{S}_i(\mathbf{y}, \mathbf{p})}{\partial y_j} + \sum_{k=1}^{n} \frac{\partial \tilde{S}_i(\mathbf{y}, \mathbf{p})}{\partial p_k} \frac{\partial p_k}{\partial y_j}. \tag{A-8}$$

Collecting all (i, j) elements $\partial S_i(\mathbf{y})/\partial y_j$ into the matrix $\mathbf{J}_{\mathbf{y}}(\mathbf{s})$, the above can be written in matrix form as

$$\mathbf{J}_{\mathbf{y}}(\mathbf{s}) = \mathbf{J}_{\mathbf{y}}(\tilde{\mathbf{s}}) + \mathbf{J}_{\mathbf{p}}(\tilde{\mathbf{s}})\mathbf{J}_{\mathbf{y}}(\mathbf{p}), \tag{A-9}$$

where $\tilde{\mathbf{s}} = (\tilde{S}_1(\mathbf{y}, \mathbf{p}(\mathbf{y})), \ldots, \tilde{S}_n(\mathbf{y}, \mathbf{p}(\mathbf{y})))$ and recall that $\mathbf{J}_{\mathbf{z}}(\mathbf{g})$ denotes the derivative of the vector valued function $\mathbf{g}$ with respect to vector $\mathbf{z}$.

The first matrix on the right-hand side of eq. (A-9) can be written as

$$\mathbf{J}_{\mathbf{y}}(\tilde{\mathbf{s}}) = \mathbf{H} + \boldsymbol{I}_{\mathbf{y}}, \tag{A-10}$$

where $\mathbf{H}$ is a diagonal matrix with (i,i)-element $H_i(\mathbf{y}) = \partial^2\tilde{\rho}(x_i,\mathbf{y},\mathbf{p})/\partial x_i^2|_{\mathbf{x}=\mathbf{y},\mathbf{p}=\mathbf{p}(\mathbf{y})}$ and

$$\boldsymbol{I}_{\mathbf{y}} = \begin{pmatrix} \dfrac{\partial^2\tilde{\rho}(x_1,\mathbf{y},\mathbf{p})}{\partial x_1\partial y_1} & \cdots & \dfrac{\partial^2\tilde{\rho}(x_1,\mathbf{y},\mathbf{p})}{\partial x_1\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2\tilde{\rho}(x_n,\mathbf{y},\mathbf{p})}{\partial x_n\partial y_1} & \cdots & \dfrac{\partial^2\tilde{\rho}(x_n,\mathbf{y},\mathbf{p})}{\partial x_n\partial y_n} \end{pmatrix}_{\substack{\mathbf{x}=\mathbf{y}\\ \mathbf{p}=\mathbf{p}(\mathbf{y})}} \tag{A-11}$$

is a matrix capturing trait-dependent selection holding morph frequencies constant. The second matrix on the right-hand side of eq. (A-9), meanwhile, can be written as

$$\mathbf{J}_{\mathbf{p}}(\tilde{\mathbf{s}}) = \boldsymbol{I}_{\mathbf{p}} = \begin{pmatrix} \dfrac{\partial^2\tilde{\rho}(x_1,\mathbf{y},\mathbf{p})}{\partial x_1\partial p_1} & \cdots & \dfrac{\partial^2\tilde{\rho}(x_1,\mathbf{y},\mathbf{p})}{\partial x_1\partial p_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2\tilde{\rho}(x_n,\mathbf{y},\mathbf{p})}{\partial x_n\partial p_1} & \cdots & \dfrac{\partial^2\tilde{\rho}(x_n,\mathbf{y},\mathbf{p})}{\partial x_n\partial p_n} \end{pmatrix}_{\substack{\mathbf{x}=\mathbf{y}\\ \mathbf{p}=\mathbf{p}(\mathbf{y})}}, \tag{A-12}$$

which captures frequency-dependent selection holding traits constant.

Finally, the third matrix on the right-hand side of eq. (A-9), $\mathbf{J}_{\mathbf{y}}(\mathbf{p})$, can be interpreted as a conversion factor such the the two terms of eq. (A-9) are in the same unit. To see this, let us recall eq. (7) and focus on a polymorphic population satisfying

$$\tilde{\rho}(y_i,\mathbf{y},\mathbf{p}) = 1 \qquad \text{for all } y_i \in \mathcal{P}(\mathbf{p}). \tag{A-13}$$

Upon implicit differentiation of eq. (A-13) and re-arrangements, one then gets

$$\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})\mathbf{J}_{\mathbf{y}}(\mathbf{p}) = -\mathbf{J}^*_{\mathbf{y}}(\tilde{\boldsymbol{\rho}}), \tag{A-14}$$

where

$$\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = \begin{pmatrix} \dfrac{\partial\tilde{\rho}(y_1,\mathbf{y},\mathbf{p})}{\partial p_1} & \cdots & \dfrac{\partial\tilde{\rho}(y_1,\mathbf{y},\mathbf{p})}{\partial p_n} \\ \vdots & & \vdots \\ \dfrac{\partial\tilde{\rho}(y_n,\mathbf{y},\mathbf{p})}{\partial p_1} & \cdots & \dfrac{\partial\tilde{\rho}(y_n,\mathbf{y},\mathbf{p})}{\partial p_n} \end{pmatrix}_{\mathbf{p}=\mathbf{p}(\mathbf{y})}, \tag{A-15}$$

$$\mathbf{J}_{\mathbf{y}}(\mathbf{p}) = \begin{pmatrix} \dfrac{\partial p_1(\mathbf{y})}{\partial y_1} & \cdots & \dfrac{\partial p_1(\mathbf{y})}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial p_n(\mathbf{y})}{\partial y_1} & \cdots & \dfrac{\partial p_n(\mathbf{y})}{\partial y_n} \end{pmatrix}, \tag{A-16}$$

and

$$\mathbf{J}_{\mathbf{y}}(\tilde{\boldsymbol{\rho}}) = \begin{pmatrix} \frac{\partial \tilde{\rho}(y_1, \mathbf{y}, \mathbf{p})}{\partial y_1} & \cdots & \frac{\partial \tilde{\rho}(y_1, \mathbf{y}, \mathbf{p})}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial \tilde{\rho}(y_n, \mathbf{y}, \mathbf{p})}{\partial y_1} & \cdots & \frac{\partial \tilde{\rho}(y_n, \mathbf{y}, \mathbf{p})}{\partial y_n} \end{pmatrix}_{\mathbf{p}=\mathbf{p}(\mathbf{y})} . \tag{A-17}$$

Given that $\det \mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$ and thus $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})$ is invertible, eq. (A-14) can be re-arranged as

$$\mathbf{J}_{\mathbf{y}}(\mathbf{p}) = -\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})^{-1}\mathbf{J}_{\mathbf{y}}(\tilde{\boldsymbol{\rho}}). \tag{A-18}$$

Substituting eq. (A-18) and eq. (A-10) into eq. (A-9) finally gives $\mathbf{J}_{\mathbf{y}}(\mathbf{s}) = \mathbf{H} + \boldsymbol{I}$, where $\boldsymbol{I} = \boldsymbol{I}_{\mathbf{y}} - \boldsymbol{I}_{\mathbf{p}}\mathbf{M}$ and $\mathbf{M} = \mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}})^{-1}\mathbf{J}_{\mathbf{y}}(\tilde{\boldsymbol{\rho}})$, as required.

Appendix A.2.3 D-stable matrices

As seen in Appendix A.2.1 and Appendix A.2.2, the asymptotic stability of morph frequency dynamics eq. (A-2) and the convergence stability of trait dynamics eq. (A-5) both depend on the eigenvalues of a matrix of the form $\mathbf{DJ}$ with $\mathbf{D}$ being a diagonal positive matrix, with diagonal elements $d_i > 0$ for each i, and $\mathbf{J}$ being some real valued Jacobian matrix. Much research has been devoted to understand the relationship between the stability properties of dynamical systems given by a matrix $\mathbf{J}$ and those given by $\mathbf{DJ}$ (Quirk and Ruppert, 1965; Kushel, 2019). All relevant results about this relationship, however, concern the case where $\mathbf{DJ}$ is the Jacobian of the flow of a continuous time dynamical system and not the map of discrete time system (see Kushel, 2019 for treating continuous and discrete time systems). These results therefore apply to eq. (A-6) , but not to eq. (A-3), unless we replace fitness by growth rate. Therefore we now focus on the continuous time case.

Recall that the fixed point of a continuous time dynamical system is locally asymptotically stable if all eigenvalues of its Jacobian $\mathbf{DJ}$ have negative real parts. If any of the eigenvalues has a positive real part, then the fixed point is unstable (Stuart and Humphries, 1996, Thm.2.3.5). A matrix is said to be stable if all its eigenvalue have negative real parts. A matrix $\mathbf{J}$ is further said to be $\mathbf{D}$-stable if $\mathbf{DJ}$ is stable for all diagonal matrices $\mathbf{D}$, with positive entries on the diagonal.

There are few necessary conditions for $\mathbf{D}$-stability and the most relevant for our purpose is that $\mathbf{J}$ itself is stable (Quirk and Ruppert, 1965, p. 314). There are, however, numerous sufficiency conditions for $\mathbf{D}$-stability. Perhaps the two most relevant ones in the present context are the two following reported in Quirk and Ruppert (1965). First, it is sufficient that $\mathbf{J}$ is negative-definite, meaning that all the eigenvalues of its symmetric part $(\mathbf{J} + \mathbf{J}^{\mathrm{T}})/2$ are negative (where T denotes transposition; note that the eigenvalues are necessarily real here since $(\mathbf{J} + \mathbf{J}^{\mathrm{T}})/2$ is symmetric). Negative-definiteness then requires that the diagonal elements of $\mathbf{J}$ are all non-positive, $j_{ii} \leq 0$, where j_{ij} denotes the (i, j)-entry of $\mathbf{J}$. The second sufficiency condition for $\mathbf{D}$-stability is that $\mathbf{J}$ is a sign-similar matrix (and used in theoretical ecology, e.g. May, 1973),

which means that the following conditions holds:

$$\text{(i)} \quad j_{ii} \leq 0 \text{ for all } i \text{ and } j_{ii} < 0 \text{ for some } i. \tag{A-19}$$

$$\text{(ii)} \quad j_{ij} j_{ji} \leq 0 \text{ for all } i \neq j. \tag{A-20}$$

$$\text{(iii)} \quad j_{ij} j_{jk} \cdots j_{ri} = 0 \text{ for any 3 or more indices } i \neq j \neq k \cdots \neq r. \tag{A-21}$$

$$\text{(iv)} \quad \det \mathbf{J} \neq 0. \tag{A-22}$$

Appendix A.2.4 Equilibrium number of morphs and vanishing determinant sequence

We here discuss conditions for determining the candidate number of morphs to coexist at a long-term evolutionary oligomorphic equilibrium. For this, let us first specify invasion fitness according to production and regulation factors in a population with n morphs with distribution $\mathbf{p}$ and support $\mathcal{P}(\mathbf{p})$. From eq. (18), the invasion fitness $\tilde{\rho}(x_i, \mathbf{y}, \mathbf{p})$ of type x_i in such a population can be expressed as

$$\tilde{\rho}(x_i, \mathbf{y}, \mathbf{p}) = W\big(\boldsymbol{\pi}(\mathbf{z}(x_i, \mathbf{e}^{\mathrm{z}}(x_i, \mathbf{p})), \mathbf{e}^{\pi}(x_i, \mathbf{p})), \mathbf{e}^{\mathrm{w}}(x_i, \mathbf{p})\big) = \tilde{W}(x_i, \mathbf{e}(x_i, \mathbf{p})) \tag{A-23}$$

for some real valued function $\tilde{W}$ of the trait of morph i that is written as a function of

$$\mathbf{e}(y_i, \mathbf{p}) = (\mathbf{e}^{\mathrm{z}}(y_i, \mathbf{p}), \mathbf{e}^{\pi}(y_i, \mathbf{p}), \mathbf{e}^{\mathrm{w}}(y_i, \mathbf{p})) = \{e_k(y_i, \mathbf{p})\}_{k=1}^{n_{\mathrm{e}}}, \tag{A-24}$$

which is the collection of the $n_{\mathrm{e}} = n_{\mathrm{z}} + n_{\pi} + n_{\mathrm{w}}$ regulating factors appearing in eq. (18), arranged in some lexicographic order. As usual, the dependence on $\mathbf{p}$ means a functional dependence on the distribution $\mathbf{p}$ and its support $\mathcal{P}(\mathbf{p})$.

Plugging eq. (A-23) into eq. (A-15) allows us to consider the Jacobian matrix of invasion fitness with respect to the type distribution $\mathbf{p}$ in a population with n segregating morphs as,

$$\mathbf{J}_{\mathbf{p}}^{(n)}(\tilde{\boldsymbol{\rho}}) = \begin{pmatrix} \sum_{k=1}^{n_{\mathrm{e}}} \dfrac{\partial \tilde{W}(x_1, \mathbf{e})}{\partial e_k} \left.\dfrac{\partial e_k}{\partial p_1}\right|_{\mathbf{e}=\mathbf{e}(x_1,\mathbf{p})} & \cdots & \sum_{k=1}^{n_{\mathrm{e}}} \dfrac{\partial \tilde{W}(x_1, \mathbf{e})}{\partial e_k} \left.\dfrac{\partial e_k}{\partial p_n}\right|_{\mathbf{e}=\mathbf{e}(x_1,\mathbf{p})} \\ \vdots & & \vdots \\ \sum_{k=1}^{n_{\mathrm{e}}} \dfrac{\partial \tilde{W}(x_n, \mathbf{e})}{\partial e_k} \left.\dfrac{\partial e_k}{\partial p_1}\right|_{\mathbf{e}=\mathbf{e}(x_n,\mathbf{p})} & \cdots & \sum_{k=1}^{n_{\mathrm{e}}} \dfrac{\partial \tilde{W}(x_n, \mathbf{e})}{\partial e_k} \left.\dfrac{\partial e_k}{\partial p_n}\right|_{\mathbf{e}=\mathbf{e}(x_n,\mathbf{p})} \end{pmatrix}_{\mathbf{p}=\mathbf{p}(\mathbf{y})}. \tag{A-25}$$

We can then define a sequence of Jacobian matrices $\mathbf{J}_{\mathbf{p}}^{(2)}(\tilde{\boldsymbol{\rho}}), \mathbf{J}_{\mathbf{p}}^{(3)}(\tilde{\boldsymbol{\rho}}), \mathbf{J}_{\mathbf{p}}^{(4)}(\tilde{\boldsymbol{\rho}})$... for an increasing number of morphs. Suppose that we are told that $\det \mathbf{J}_{\mathbf{p}}^{(n)}(\tilde{\boldsymbol{\rho}}) = 0$. What are the conditions on $\mathbf{J}_{\mathbf{p}}^{(n)}(\tilde{\boldsymbol{\rho}})$ such that $\det \mathbf{J}_{\mathbf{p}}^{(n)}(\tilde{\boldsymbol{\rho}}) = 0 \Rightarrow \det \mathbf{J}_{\mathbf{p}}^{(n+1)}(\tilde{\boldsymbol{\rho}}) = 0$? If $\operatorname{rank} \mathbf{J}_{\mathbf{p}}^{(n)}(\tilde{\boldsymbol{\rho}}) < n$, then unless the added row and column to form $\mathbf{J}_{\mathbf{p}}^{(n+1)}(\tilde{\boldsymbol{\rho}})$ restore full rank, $\operatorname{rank} \mathbf{J}_{\mathbf{p}}^{(n+1)}(\tilde{\boldsymbol{\rho}}) < n+1 \Rightarrow \det \mathbf{J}_{\mathbf{p}}^{(n+1)}(\tilde{\boldsymbol{\rho}}) = 0$. According to the Schur complement determinant formula (Horn and Johnson, 2012, p. 26, eq. 0.8.5.10), the added row and column to form $\mathbf{J}_{\mathbf{p}}^{(n+1)}(\tilde{\boldsymbol{\rho}})$ need to be carefully chosen to restore full rank. Restoration of full rank is thus unlikely to obtain

in our biological context,especially when $\mathbf{e}(x_i, \mathbf{p}) = \mathbf{e}(\mathbf{p})$ for all i. In this case, if we were to add a column and a row to matrix eq. (A-25), then the new matrix $\mathbf{J}_{\mathbf{p}}^{(n+1)}(\boldsymbol{\rho})$ has entries that have the same structure as in $\mathbf{J}_{\mathbf{p}}^{(n)}(\boldsymbol{\rho})$ and thus likely to have columns or rows that are linearly dependent.

From these considerations suggesting that restoration of full rank is unlikely (though not impossible), a way to determine the candidate number n^* of coexisting morph at a long term equilibrium is to compute the determinant of the sequence of matrices $\mathbf{J}_{\mathbf{p}}^{(2)}(\tilde{\boldsymbol{\rho}}), \mathbf{J}_{\mathbf{p}}^{(3)}(\tilde{\boldsymbol{\rho}}), ...$ until the first n such that $\det \mathbf{J}_{\mathbf{p}}^{(n)}(\tilde{\boldsymbol{\rho}}) = 0$, whereby $n^* = n - 1$. Additionally, it is useful to observe that when $\mathbf{e}(x_i, \mathbf{p}) = \mathbf{e}(\mathbf{p})$ for all i, the competitive exclusion principle (e.g., Levin, 1970; Meszéna et al., 2006) also tells us that the maximum number of morph that can coexist is

$$n^* \leq n_{\mathrm{z}} + n_{\pi} + n_{\mathrm{w}}, \tag{A-26}$$

where $n_{\mathrm{z}} + n_{\pi} + n_{\mathrm{w}}$ (the total number of regulating factors) is the maximal number of independent regulating factors.

Finally, we observe that the entries of $\mathbf{J}_{\mathbf{p}}^{(n)}(\tilde{\boldsymbol{\rho}})$ given in eq. (A-25) are conceptually equivalent to eq. (47) of Meszéna et al. (2006), so that their limiting similarity principle (Meszéna et al., 2006, p. 78) applies to our model. Here, this principle entails that for coexistence to occur, the different morphs need to differ in both their dependence on regulating variables (differences in $\partial \tilde{W}(x_i, \mathbf{e})/\partial e_k$) and in how their frequencies affect these variables (differences in $\partial e_k/\partial p_n|_{\mathbf{e}=\mathbf{e}(x_i,\mathbf{p})}$). Based on this connection with the so-called limiting similarity principle, we take these conditions to mean that the fitnesses of different morphs need to be "qualitatively dissimilar" in how they respond to intraspecific frequency-dependent feedback for them to coexist.

Appendix B Ecologically mediated interactions

Appendix B.1 Interaction coefficient

We derive here eqs. (23)–(24) of the main text . We first substitute eq. (20) into eq. (3), which after factoring out common terms gives

$$I(x^*) = \left[\frac{\mathrm{d}z_1(y)}{\mathrm{d}y} \frac{\mathrm{d}e_1(y)}{\mathrm{d}y} \left(\frac{\partial^2 W}{\partial z_1 \partial e_1} + \left(\frac{\mathrm{d}z_2(y)}{\mathrm{d}y} / \frac{\mathrm{d}z_1(y)}{\mathrm{d}y} \right) \frac{\partial^2 W}{\partial z_2 \partial e_1} \right) \right.$$
$$\left. + \frac{\mathrm{d}z_1(y)}{\mathrm{d}y} \frac{\mathrm{d}e_2(y)}{\mathrm{d}y} \left(\frac{\partial^2 W}{\partial z_1 \partial e_2} + \left(\frac{\mathrm{d}z_2(y)}{\mathrm{d}y} / \frac{\mathrm{d}z_1(y)}{\mathrm{d}y} \right) \frac{\partial^2 W}{\partial z_2 \partial e_2} \right) \right]_{x=y=x^*}. \tag{A-27}$$

Then, using eqs. (21)–(22) to substitute $\left(\frac{\mathrm{d}z_2(y)}{\mathrm{d}y} / \frac{\mathrm{d}z_1(y)}{\mathrm{d}y} \right)\Big|_{x=y=x^*} = -\left(\partial W/\partial z_1\right)/(\partial W/\partial z_2)|_{x=y=x^*}$ and $\left(\frac{\mathrm{d}e_2(y)}{\mathrm{d}y} / \frac{\mathrm{d}e_1(y)}{\mathrm{d}y} \right)\Big|_{x=y=x^*} = -\left(\partial W/\partial e_1\right)/(\partial W/\partial e_2)|_{x=y=x^*}$ produces the desired result.

Appendix B.2 Within generation heterogeneity

Appendix B.2.1 Resource specialization under panmixia

Substituting eq. (26) and $z_i(x) = e^{-(x-o_i)^2/(2\sigma^2)}$ into eqs. (3)–(4), we obtain that there is a unique singular strategy

$$x^* = \rho o_1 + (1-\rho)o_2, \tag{A-28}$$

and that

$$\begin{aligned} J(x^*) &= -1/\sigma^2 \\ H(x^*) &= (1/\sigma^4)\left((1-\rho)\rho(o_1-o_2)^2 - \sigma^2\right). \end{aligned} \tag{A-29}$$

Hence, x^* is always convergence stable (as $\sigma^2 > 0$) and evolutionary branching occurs when $H(x^*) > 0$, i.e. when the variance $(1-\rho)\rho(o_1-o_2)^2$ in environmental states is larger than σ^2, which can be though as a measure of spread of the production factor. For instance, this obtains when $o_1 = 1$, $o_2 = -1$, $\rho = 1/2$ and $\sigma^2 = 0.625$, which we use for illustration.

Given that evolutionary branching occurs, let us now consider evolution in an oligomorphic population and write $\mathbf{p} = (p_1, p_2, ..., p_n)$ for the probability mass function that has support given by the finite vector of morphs $\mathcal{P}(\mathbf{p}) = \mathbf{y} = (y_1, y_2, ..., y_n)$. Then, generalizing eq. (26) to the polymorphic case, the invasion fitness of type x in population $\mathcal{P}(\mathbf{p})$ is

$$\rho(x, \mathbf{p}) = \tilde{\rho}(x, \mathcal{P}(\mathbf{p}), \mathbf{p}) = q\frac{z_1(x)}{\sum_{i=1}^n e_1(y_i)p_i} + (1-q)\frac{z_2(x)}{\sum_{i=1}^n e_2(y_i)p_i}. \tag{A-30}$$

Using a symbolic computer algebra system such as Mathematica (Wolfram Research, 2016), it can be checked that for $n = 2$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$, while for $n = 3$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = 0$. There is thus a maximum of two morphs in this model, which is in line with the fact that there are only two regulating factors. We can therefore assume a dimorphism. Using eq. (7) (or eq. A-2), we obtain that the equilibrium frequency of morph one in a dimorphic population is

$$p_1(y_1, y_2) = \frac{1-q}{1-e^{(2o_1-y_1-y_2)(y_1-y_2)/(2\sigma^2)}} + \frac{q}{1-e^{(2o_2-y_1-y_2)(y_1-y_2)/(2\sigma^2)}}. \tag{A-31}$$

Applying eq. (9) to eq. (A-30) with eq. (A-31), we can obtain the selection gradients $S_1(\mathbf{y})$ and $S_2(\mathbf{y})$ on morphs y_1 and y_2. However, we were not able to find analytical expressions for the singular coalition $\mathbf{x}^* = (x_1^*, x_2^*)$. Nevertheless, $S_1(\mathbf{x}^*) = 0$ and $S_2(\mathbf{x}^*) = 0$ can be solved numerically and given the parameter values $o_1 = 1$, $o_2 = -1$, $\rho = 1/2$ and $\sigma^2 = 0.625$, there are two singular coalitions $(x_1^* = 0.987633, x_2^* = -0.987633)$ and $(x_1^* = -0.987633, x_2^* = 0.987633)$ with $p_1 = 1/2$. These are symmetric around zero. They are close to but not equal to the optimal values in each patch which are $o_1 = 1$ and $o_2 = -1$. This dimorphism is locally

uninvadable and strongly convergence stable. It can also be checked that at this dimorphism $\boldsymbol{I}^*_{\mathbf{y}} = \boldsymbol{I}^*_{\mathbf{p}}\mathbf{M}^*$, whereby the conditions of uninvadability and (strong) convergence stability are equivalent, i.e. $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s}) = \mathbf{H}^*$. This makes the dimorphism saturated.

It is straightforward to extent the above model to an arbitrary number n_{e} of environments, in which case the main text invasion fitness (26) for a monomorphic population becomes

$$\rho(x, \mathbf{p}) = \sum_{j=1}^{n_{\mathrm{e}}} q_j \frac{z_j(x)}{\sum_{i=1}^{n} e_j(y_i)p_i}, \tag{A-32}$$

where q_j is the frequency of the jth environmental state. When $z_i(z) = e_i(z) = e^{-(z-o_i)^2/(2\sigma^2)}$, this is a special case of the model of (Svardal et al., 2015) who analysed the conditions for evolutionary branching in a monomorphic resident population and found that there is a unique singular coalition $x^* = \bar{o} = \sum_i^{n_{\mathrm{e}}} q_i o_i$, the condition for convergence stability thereoff is $-1/\sigma^2 < 0$, and the condition for disruptive selection is $\sum_i q_i(o_i - \bar{o})^2 > \sigma^2$ (obtained from eq. 13 and eq. 15 of Svardal et al., 2015 under the present modelling assumptions). Assuming $q_j = 1/n_{\mathrm{e}}$, $o_i = i$, and $n_{\mathrm{e}} = 5$, Fig. 2 displays a situation where the evolutionary dynamics leads to uninvadable coalition that is saturated (5 coexisting morphs when $\sigma^2 = 0.024$) and an unsaturated uninvadable coalition (two coexisting morphs when $\sigma^2 = 1$). Note that the singular trait value corresponding to this this model is $x^* = 3$, which is attained in the second panel of Fig. 2, but not the first panel. The reason is that we kept population size small ($N = 2500$) for computational reasons in this model and that this induces branching to occur before the singularity is actually reached.

Appendix B.2.2 Life-history under limited dispersal

Substituting fitness model eq. (29) along $z_1(x) = x^{1/2}$ and $z_2(x) = (1-x)^{\beta}$ into eqs. (3)–(4) reveals a unique singular strategy

$$x^* = \frac{1}{(1 - 4\beta + 2\beta m)^2}, \tag{A-33}$$

and

$$\begin{aligned} J(x^*) &= -\frac{m\left(1 + x^* + 2\sqrt{x^*} - 2x^*\beta(2-m)\right)}{4\left(\sqrt{x^*}+1\right)^2 (x^*)^{3/2}} \\ H(x^*) &= \frac{m\Big(x^*(1+4\beta(1-m)) - (x^*)^{3/2}\left(4\beta(1+\beta m(2m-3))-1\right) - \sqrt{x^*} - 1\Big)}{4\left(1-\sqrt{x^*}\right)\left(\sqrt{x^*}+1\right)^2 (x^*)^{3/2}}. \end{aligned} \tag{A-34}$$

It can be checked that there is a range of parameters values in m and β such that $J(x^*) < 0$ and $H(x^*) > 0$, whereby dimorphism is favored by gradual evolution at x^*. This is for instance the case when $m = 0.5$ and $\beta = 0.7$.

Given branching occurs, let us now consider evolution in a polymorphic population with trait values $\mathcal{P}(\mathbf{p}) =$

$\mathbf{y} = (y_1, y_2, ..., y_n)$. Generalizing eq. (29) to the polymorphic case, the invasion fitness of type x in population $\mathcal{P}(\mathbf{p})$ is then given by

$$\rho(x, \mathbf{p}) = z_1(x) + (1 - z_1(x)) \frac{(1-m)z_2(x)}{(1-m)z_2(x) + m\sum_{i=1}^{n} e_2(y_i)p_i} + m\sum_{i=1}^{n} p_i \frac{(1 - e_1(y_i))z_2(x)}{\sum_{i=1}^{n} e_2(y_i)p_i}. \tag{A-35}$$

It can be checked that for $n = 2$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$, while for $n = 3$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = 0$, which is in line with the fact that there are only two regulating factors in this model. Focusing on a dimorphism and applying eq. (7), the equilibrium frequency $p_1(y_1, y_2)$ of morph 1 can be computed analytically, but the expression is complicated and not illuminating so we refrain from giving it here.

Applying eq. (9), and evaluating the selection gradients $S_1(\mathbf{y})$ and $S_2(\mathbf{y})$ on morph y_1 and y_2 using eq. (A-35), we were not able to find analytical expressions for the singular coalition $\mathbf{x}^* = (x_1^*, x_2^*)$. But numerical analysis of the selection gradients shows that one morph is favored to take the value of one, say $x_1^* = 1$, whereby all resources are channelled into survival. At this point, the selection gradient on the second morphs vanishes, which results in a line of neutral equilibria for the second morph. However, in the evolutionary process with mutations, the exact value of one is never reached, whereby selection becomes stabilizing on the second morph resulting in the coexistence of one (close to) immortal morph and another investing a fraction of resources into reproduction (see Fig. 3).

Appendix B.3 Between generation heterogeneity

Substituting fitness model eq. (32) with $z_i(x) = e^{-(x-o_i)^2/(2\sigma^2)}$ into eqs. (3)–(4) shows that there is a unique singular strategy given by eq. (A-28), and that

$$\begin{aligned} J(x^*) &= -2(1-s)/\sigma^2 \\ H(x^*) &= \frac{2(1-s)}{\sigma^4}\left(2s(1-\rho)\rho(o_1 - o_2)^2 - \sigma^2\right). \end{aligned} \tag{A-36}$$

This shows that x^* is convergence stable for $s \in [0, 1)$ and that evolutionary branching can occur under a certain range of parameter values. The larger s is, the more likely polymorphism is to emerge.

Given that branching occurs, let us now consider evolution in a polymorphic population with trait values $\mathcal{P}(\mathbf{p}) = \mathbf{y} = (y_1, y_2, ..., y_n)$. Generalizing eq. (32) to the polymorphic case, the invasion fitness of type x in population $\mathcal{P}(\mathbf{p})$ is given by

$$\rho(x, \mathbf{p}) = \left(s + (1-s)\frac{z_1(x)}{\sum_{i=1}^{n} e_1(y_i)p_i}\right)^q \left(s + (1-s)\frac{z_2(x)}{\sum_{i=1}^{n} e_2(y_i)p_i}\right)^{(1-q)}. \tag{A-37}$$

It can be checked that for $n = 2$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$, while for $n = 3$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = 0$. Focusing again on a dimorphism, we did not find any relevant analytical expression concerning the gradual evolution of such a

polymorphism. But it can be checked numerically using the selection gradients $S_1(\mathbf{y})$ and $S_2(\mathbf{y})$ on morphs y_1 and y_2 derived from eq. (A-37) that, as for within generation heterogeneity, there are two symmetric singular coalitions. For instance, when $s = 0.5$, $o_1 = 1$, $o_2 = -1$, $\rho = 1/2$ and $\sigma^2 = 0.625$ (same parameters as in section Appendix B.2.1, except that survival is non-zero), the singular coalitions $(x_1^* = 0.999785, x_2^* = -0.999785)$ and $(x_1^* = -0.999785, x_2^* = 0.999785)$ with $p_1 = 1/2$ are locally uninvadable, strongly convergence stable, and the resulting evolutionary stable dimorphism is saturated (i.e. $\boldsymbol{I}_{\mathbf{y}}^* = \boldsymbol{I}_{\mathbf{p}}^* \mathbf{M}^*$).

Appendix C Socially mediated interactions

Appendix C.1 Payoff based gradients

We here derive condition (37). For this, first substitute eq. (35) into the selection gradient of eq. (3) and use the chain rule to see that it can be written in terms of the payoff gradient as

$$S(y) = k(y) \left. \frac{\partial \pi(x,y)}{\partial x} \right|_{x=y}, \tag{A-38}$$

where $k(y) = \partial W(\pi, e(y))/\partial \pi|_{\pi=\pi(y,y)} > 0$ by assumption. Then, any singular strategy x^* satisfying $S(x^*)$ satisfies

$$\left. \frac{\partial \pi(x,y)}{\partial x} \right|_{x=y=x^*} = 0. \tag{A-39}$$

Using this equality after having substituted eq. (35) into the second-order selection coefficients of eqs. (3)–(4), we find that

$$\begin{aligned} H(x^*) &= k(x^*) \left. \frac{\partial^2 \pi(x,y)}{\partial x^2} \right|_{x=y=x^*} \\ I(x^*) &= k(x^*) \left. \frac{\partial^2 \pi(x,y)}{\partial x \partial y} \right|_{x=y=x^*} \\ J(x^*) &= k(x^*) \left[\frac{\partial^2 \pi(x,y)}{\partial x^2} + \frac{\partial^2 \pi(x,y)}{\partial x \partial y} \right]_{x=y=x^*}. \end{aligned} \tag{A-40}$$

Hence, the Hessian, interaction and Jacobian coefficients can all directly be computed from the payoff function. The necessary condition $I(x^*) = k(x^*) I_\pi(y) < 0$ for protected polymorphism can then be reduced to ineq. (37), since $k(x^*) > 0$.

Appendix C.2 Examples: contests and preference evolution

Appendix C.2.1 Contest evolution

Substituting payoff model eq. (43) into eq. (A-40) and using $a(x) = e^{\omega x}$ along $C(a(x)) = x^2$, we get a unique singular point

$$x^* = \frac{(N-1)v\omega}{2N^2} \tag{A-41}$$

and

$$\begin{aligned} J(x^*) &\propto -2 \\ H(x^*) &\propto \frac{(N-1)(N-2)v\omega^2}{N^3} - 2. \end{aligned} \tag{A-42}$$

Thus, when $N \leq 2$, the singular trait value is always convergence stable and uninvadable. When $N > 2$, one can find parameter values such that $J(x^*) < 0$ and $H(x^*) > 0$. Thus, there is a range of parameter values where a dimorphism can be favored by gradual evolution at x^*.

Given that branching occurs, let us consider evolution in a polymorphic population $\mathcal{P}(\mathbf{p}) = \mathbf{y} = (y_1, y_2, ..., y_n)$. Assuming that invasions fitness in a polymorphic population is $\rho(x, y) = \Pi(a(x), b(y))/\Pi(a(y), b(y))$ with payoff given by eq. (43), the invasion fitness in a polymorphic population of type x is

$$\tilde{\rho}(x, \mathcal{P}(\mathbf{p}), \mathbf{p}) = \frac{\left(\frac{a(x)}{a(x)+(N-1)\sum_{i=1}^{n} b(y_i)p_i} v\right) - C(a(x))}{\sum_{j=1}^{n} p_j \left[\left(\frac{a(y_j)}{a(y_j)+(N-1)\sum_{i=1}^{n} b(y_i)p_i} v\right) - C(a(y_j))\right]}. \tag{A-43}$$

It can be checked that for $n = 2$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$, while for $n = 3$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = 0$. Hence, only a dimorphism should be stable in this model. The singular coalition(s) can then be evaluated numerically given parameter values for N, ω, and v and checked for convergences stability and locally uninvadability. For instance, when $N = 5$, $\omega = 2$, and $v = 10$, we find that there is unique (strongly) convergence stable and uninvadable coalition at $(x_1^* = 0.16, x_2^* = 2.4)$, which is saturated and confirmed by individual-based simulations (see Fig. 3).

Appendix C.2.2 Preference evolution: dispositions

Here, we analyze the model of dispositions of section 5.3.2 but extend the payoff of this model to

$$\pi(x, y) = \Pi(a(x, y), a(y, x)) - \eta x^{\nu}, \tag{A-44}$$

where $a(x, y)$ is given by eq. (48) and the term ηx^{ν} represents a cost of overconfidence, which increases with x and is controlled by the parameters $\eta \geq 0$ and $\nu > 1$. This payoff implements the main text model eqs. (45)–

(46) but adds a cost to dispositions. This prevents the open ended evolution of exaggerated self-perception. Note that the right hand side of eq. (A-44) is more general than the right hand side of eq. (38). As a result the first-order condition for model eq. (A-44) is not given directly by eq. (39). Nevertheless the synergy coefficient corresponding to eq. (A-44) is equivalent to that obtained from model eqs. (45)–(46).

Substituting payoff model eq. (A-44) along equilibrium actions (48) into eq. (36) yields a unique feasible singular strategy x^*, which can be expressed as

$$x^* = \frac{\alpha\beta^2}{4\gamma^2 + 2\beta\gamma - \beta^2} + \eta\, h(\alpha, \beta, \gamma, \eta) \tag{A-45}$$

for some complicated function $h(\alpha, \beta, \gamma, \eta)$ of the parameters. When $\eta = 0$ and $\gamma = 1$, x^* reduces to the level of disposition found by Heifetz et al. (2007a, p. 267). The Jacobian and the Hessian coefficients at this singular point are obtained from eq. (A-40) to read as

$$\begin{aligned} J(x^*) &\propto -\frac{\beta^2 - 2\gamma^2 - 2\gamma(\beta+\gamma)}{(\beta - 2\gamma)(\beta + 2\gamma)^2} - 12\eta\,(x^*)^2 \\ H(x^*) &\propto \frac{4\gamma\left(\beta^2 - 2\gamma^2\right)}{\left(\beta^2 - 4\gamma^2\right)^2} - 12\eta\,(x^*)^2. \end{aligned} \tag{A-46}$$

Straightforward analysis of these equations reveals that there is a range of parameter values such that $J(x^*) < 0$ and $H(x^*) > 0$ and so evolutionary branching will occur. For instance, this is the case when $\alpha = 0.2$, $\beta = 1.6$, $\gamma = 1$, and $\eta = 0.45$.

Given that branching occurs, let us consider evolution in a polymorphic population $\mathcal{P}(\mathbf{p}) = \mathbf{y} = (y_1, y_2, ..., y_n)$. Assuming a Wright-Fisher process, the invasion fitness of type x in a polymorphic population is

$$\tilde{\rho}(x, \mathcal{P}(\mathbf{p}), \mathbf{p}) = \frac{\sum_{i=1}^{n} \left[(\alpha - \beta a(y_i, x) - \gamma a(x, y_i))a(x, y_i) - \eta x^{\nu}\right] p_i}{\sum_{j=1}^{n} p_j \left[\sum_{i=1}^{n} \left[(\alpha - \beta a(y_i, y_j) - \gamma a(y_j, y_i))a(y_j, y_i) - \eta y_j^{\nu}\right] p_i\right]}. \tag{A-47}$$

It can be checked that for $n = 2$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) \neq 0$, while for $n = 3$ det $\mathbf{J}_{\mathbf{p}}(\tilde{\boldsymbol{\rho}}) = 0$, whereby only a dimorphism should be stable in this model.

By solving $\tilde{\rho}(y_1, \mathcal{P}(\mathbf{p}), \mathbf{p}) = \tilde{\rho}(y_2, \mathcal{P}(\mathbf{p}), \mathbf{p})$ for p_1 by setting $p_2 = 1 - p_1$, we find a unique complicated analytical expression for $p_1(y_1, y_2)$. Nevertheless, this expression allows us to compute the selection gradients $S_1(\mathbf{y})$ and $S_2(\mathbf{y})$ analytically. By solving $S_1(\mathbf{x}^*) = 0$ and $S_2(\mathbf{x}^*) = 0$ for $\mathbf{x}^* = (x_1^*, x_2^*)$, we find that

$$x_1^* = -\frac{\sqrt{\gamma}\sqrt{\beta^2 - 2\gamma^2}}{\sqrt{\eta}\,(\beta^2 - 4\gamma^2)} \qquad \text{and} \qquad x_2^* = \frac{\sqrt{\gamma}\sqrt{\beta^2 - 2\gamma^2}}{\sqrt{\eta}\,(\beta^2 - 4\gamma^2)}, \tag{A-48}$$

which leads to

$$p_1 = \frac{\frac{\alpha\beta\sqrt{\eta}\left(\beta^2 - 4\gamma^2\right)}{\sqrt{\gamma}\sqrt{\beta^2 - 2\gamma^2}} - \frac{2\alpha\sqrt{\gamma}\sqrt{\eta}\left(\beta^2 - 4\gamma^2\right)}{\sqrt{\beta^2 - 2\gamma^2}} + \beta}{2\beta}. \tag{A-49}$$

Computing the Hessian for each morph (eq. 4), we find that

$$H_1(\mathbf{x}^*) = H_2(\mathbf{x}^*) = -\frac{8\gamma\eta\left(\beta^2-4\gamma^2\right)^2\left(\beta^2-2\gamma^2\right)}{\gamma\left(2\gamma\left(-3\beta^2\gamma^2+\beta^4+2\gamma^4\right)-\alpha^2\eta(\beta-2\gamma)^4(\beta+2\gamma)^2\right)}. \tag{A-50}$$

Computing the Jacobian matrix $\mathbf{J}^*_{\mathbf{y}}(\mathbf{s})$ (eq. 14) shows that both eigenvalues take the same value as the Hessian coefficients eq. (A-50). Hence, whenever the coalition is locally uninvadable, it is also strongly convergence stable and the polymorphism is saturated, i.e. $\boldsymbol{I}^*_{\mathbf{y}} = \boldsymbol{I}^*_{\mathbf{p}}\mathbf{M}^*$. This is the case for instance with the aforementioned parameter values ($\alpha = 0.2$, $\beta = 1.6$, $\gamma = 1$, and $\eta = 0.45$), which leads to a symmetric uninvadable coalitions $[(x_1^* = 0.77, x_2^* = -0.77)$ and $(x_1^* = -0.77, x_2^* = 0.77)]$, which can be confirmed by individual-based simulations (see Fig. 3).

Appendix C.2.3 Preference evolution: other-regarding motivations

Here, we consider an alternative model of preference, in which the evolving trait x measures an individual's other-regard. We assume that the utility functions for a pair of interacting individual, one mutant with action a and trait x and a resident with action b and trait y, are given by

$$\begin{aligned} u_x(a,b) &= \Pi(a,b) + x\Pi(b,a) \\ u_y(b,a) &= \Pi(b,a) + y\Pi(a,b). \end{aligned} \tag{A-51}$$

Each utility is the sum of own material payoff to the actor plus the payoff to the interaction partner and thus represents an other-regarding preference. The weight of other-regard is the evolving traits. To close the model, we postulate the linear-quadratic payoff eq. (46) of the main text. The necessary first-conditions for a Nash equilibrium under this game are that $\partial u_x(a,b)/\partial a = 0$ and $\partial u_y(b,a)/\partial b = 0$, which can be expressed as

$$\begin{aligned} a &= \frac{\alpha}{2\gamma} - b\frac{\beta(1+x)}{2\gamma} \\ b &= \frac{\alpha}{2\gamma} - a\frac{\beta(1+y)}{2\gamma}. \end{aligned} \tag{A-52}$$

Solving for the equilibrium actions $a = a(x,y)$ and $b = a(y,x)$ yields

$$a(x,y) = \frac{\alpha(\beta-2\gamma+\beta x)}{\beta^2(x+1)(y+1)-4\gamma^2}. \tag{A-53}$$

This is indeed a Nash equilibrium action since the solution is unique and satisfies the second-order condition $\partial^2 u_x(a,b)/\partial a^2 = -2\gamma \leq 0$ and $\partial^2 u_y(b,a)/\partial b^2 = -2\gamma \leq 0$.

For this model, the corresponding singular strategy, response coefficient and action interaction coefficients

are

$$x^* = \lambda(x^*) = -\frac{\beta}{\beta + 2\gamma} \quad \text{and} \quad A(x^*) = \frac{8\gamma^2(\beta+\gamma)}{\alpha(\beta+2\gamma)^2}. \tag{A-54}$$

Since $\beta > 0$ and $\gamma > 0$, the singular strategy is always convergence stable and $\partial^2\Pi(a,b)/(\partial a \partial b) < 0$, $\partial\Pi(a,b)/(\partial b) < 0$, $[\partial^2\Pi(a,b)/\partial a^2 + \partial^2\Pi(a,b)/\partial b^2] = -2\gamma < 0$, and $\partial\Pi/\partial a + \partial\Pi/\partial b = -\alpha\beta/2\gamma < 0$. This entails that the two component of eq. (41) result in negative trait-dependent selection:

$$K_\pi(x^*) = -\frac{2\beta^2(\beta+\gamma)}{(\beta+2\gamma)^2} \qquad K_\lambda(x^*) = -\frac{4\beta\gamma(\beta+\gamma)}{(\beta+2\gamma)^2} \tag{A-55}$$

Polymorphism in other-regard can thus evolve and it can be checked that there is a range of parameter values where evolutionary branching occurs, leading to the open-ended evolution of a positive $x > 0$ and negative $x < 0$ other-regarding morph unless a cost is added to the expression of the trait.

Appendix D Genetically mediated interactions

Appendix D.1 No effect of diploidy on the instability of monomorphism

We here show eq. (51). For this, first substitute eq. (50) into the selection gradient of eq. (3) and use the chain rule to show that it can be written as

$$S(y) = \tilde{k}(y) \left.\frac{\partial W(z(x,y), z(y,y))}{\partial z(x,y)}\right|_{x=y=x^*}, \tag{A-56}$$

where $\tilde{k}(y) = \partial z(x,y)/\partial x|_{x=y} \neq 0$ by assumption. Thus, any singular strategy x^* satisfying $S(x^*) =$ has to satisfy

$$\left.\frac{\partial W(z(x,y), z(y,y))}{\partial z(x,y)}\right|_{x=y=x^*} = 0. \tag{A-57}$$

Using this equality along $\partial z(x,y)/\partial x|_{x=y} = \partial z(x,y)/\partial y|_{x=y}$ by substituting eq. (50) into the second-order coefficients of eqs. (3)–(4), we find that

$$J(x^*) = \tilde{k}(x^*)^2 \left[\frac{\partial^2 W(z(x,y), z(y,y))}{\partial z(x,y)^2} + \frac{\partial^2 W(z(x,y), z(y,y))}{\partial z(x,y)\partial z(y,y)}\right]_{x=y=x^*} \tag{A-58a}$$

$$H(x^*) = \tilde{k}(x^*)^2 \left.\frac{\partial^2 W(z(x,y), z(y,y))}{\partial z(x,y)^2}\right|_{x=y=x^*}. \tag{A-58b}$$

Hence, for $J(x^*) < 0$ and $H(x^*) > 0$ to hold, requires that

$$\left.\frac{\partial^2 W(z(x,y), z(y,y))}{\partial z(x,y)\partial z(y,y)}\right|_{x=y=x^*} < 0, \tag{A-59}$$

i.e. there is NTDS at the phenotypic level.

Because in eqs. (A-60)–(A-59) all derivatives are phenotypic and these phenotypes are evaluated at $z(x^*, x^*) = z^*$, the allelic value x^* is in fact not needed to perform the local evolutionary analysis, which can be carried out by focusing only on the phenotype value z^* solving

$$\left.\frac{\partial W(z,\bar{z})}{\partial z}\right|_{z=\bar{z}=z^*} = 0, \tag{A-60}$$

and this leads to condition (51) of the main text.

Appendix D.2 Average heterozygote advantage

Using $\Pi(z_1, z_2) = z_1^q z_2^{1-q}$ in eq. (53) when substituting eq. (54) along $g_i(x) = e^{-\omega(x-o_i)^2}$ and eq. (52) into eqs. (3)–(4) shows that there is a unique singular strategy

$$x^* = \frac{qu_1}{qu_1 + (1-q)u_2} \tag{A-61}$$

and that

$$\begin{aligned} J(x^*) &- \frac{(q(u_1-u_2)+u_2)^3}{2q(1-q)u_1u_2} \\ H(x^*) &\propto \frac{q\left((u_1-u_2)+u_2\right)^2\left(2qu_1+u_2\left(2-2q-u_1\right)\right)}{2q(1-q)u_1u_2}. \end{aligned} \tag{A-62}$$

Thereby, we find that provided $u_i > 2$ and $u_j/(2u_i) > (1-q)/(u_i - 2q)$ (for $i \neq j$), evolutionary branching will occur.

Appendix E Class structured populations

Appendix E.1 Within generation heterogeneity

Here, we show eq. (3.E) in Box 3. To that end, we first note that eq. (3.A) of that box can be written as

$$\rho(x,y) = 1 + \frac{L(x,y)}{V(x,y)}, \tag{A-63}$$

with

$$L(x,y) = \sum_{c'\in\mathcal{C}}\sum_{c\in\mathcal{C}} v_{c'}(y)\left[w_{c'c}(x,y) - w_{c'c}(y,y)\right] q_c(x,y) \tag{A-64}$$

$$V(x,y) = \sum_{c'\in\mathcal{C}}\sum_{c\in\mathcal{C}} v_{c'}(y) w_{c'c}(y,y) q_c(x,y), \tag{A-65}$$

where $v_c(y)$ is the (neutral) reproductive value of an individual in class c in a resident population y defined to satisfy the recursion $v_c(y) = \sum_{c' \in \mathcal{C}} v_{c'}(y) w_{c'c}(y, y)$ (see Ohtsuki et al., 2020 for a proof). For completeness, note that one also has $q_{c'}(x, y)\rho(x, y) = \sum_{c \in \mathcal{C}} w_{c'c}(x, y) q_c(x, y)$, whereby $\rho(x, y)$ is the leading eigenvalue of the matrix with elements $w_{c'c}(x, y)$ that has leading right eigenvector with elements $q_c(x, y)$ [and summing the equation over all $c' \in \mathcal{C}$ produces eq. 3.A, see Lehmann et al., 2016; Ohtsuki et al., 2020 for more details on invasion fitness derivation and the interpretations of the weights $q_c(x, y)$].

Owing to the fact that $L(y, y) = 0$ and $V(y, y) = 1$ (Ohtsuki et al., 2020), we have from eq. (A-63) that

$$\frac{\partial^2 \rho(x, y)}{\partial x \partial y} = \frac{\partial^2 L(x, y)}{\partial x \partial y} - \frac{\partial L(x, y)}{\partial x} \frac{\partial V(x, y)}{\partial y} - \frac{\partial V(x, y)}{\partial x} \frac{\partial L(x, y)}{\partial y}. \tag{A-66}$$

Now note that $S(y) = \partial L(x, y)/\partial x$ and that by total differentiating $\rho(y, y)$ with respect to y one obtains that $\partial L(x, y)/\partial x = -\partial L(x, y)/\partial y$. Since at a singular point $S(x^*) = 0$, we have

$$I(x^*) = \left.\frac{\partial^2 \rho(x, y)}{\partial x \partial y}\right|_{x=y=x^*} = \left.\frac{\partial^2 L(x, y)}{\partial x \partial y}\right|_{x=y=x^*}. \tag{A-67}$$

Using eq. (A-64) for $L(x, y)$ we thus obtain

$$\begin{aligned} I(x^*) = \sum_{c' \in \mathcal{C}} \sum_{c \in \mathcal{C}} \Bigg[\left(v_{c'}(y) \frac{\partial^2 w_{c'c}(x, y)}{\partial x \partial y} + \frac{\mathrm{d} v_{c'}(y)}{\mathrm{d} y} \frac{\partial w_{c'c}(x, y)}{\partial x} \right) q_c(y, y) \\ + v_{c'}(y) \frac{\partial w_{c'c}(x, y)}{\partial x} \left(\frac{\partial q_c(x, y)}{\partial y} - \frac{\partial q_c(x, y)}{\partial x} \right) \Bigg]_{x=y=x^*}. \end{aligned} \tag{A-68}$$

Finally, by factoring out $q_c(y, y)$ and $v_{c'}(y)$ and using the property of the logarithm function, we obtain eq. (3.E).

Appendix E.2 Between generation heterogeneity

We here show eq. (3.D) and eq. (3.F) of Box 3. For this, it is useful to express invasion fitness as $\rho(x, y) = e^{a(x,y)}$ where $a(x, y)$ is the (stochastic) growth (or invasion exponent) of the mutant x in a resident y population (e.g., Tuljapurkar, 1990; Ferrière and Gatto, 1995). Owing to the fact that $a(y, y) = 0$, we obtain by direct differentiation

$$S(y) = \left.\frac{\partial a(x, y)}{\partial x}\right|_{x=y} \tag{A-69}$$

$$I(y) = S(y)^2 + \left.\frac{\partial^2 a(x, y)}{\partial x \partial y}\right|_{x=y} \tag{A-70}$$

and at a singularity $I(x^*) = \partial^2 a(x, y)/\partial x \partial y\big|_{x=y=x^*}$ since $S(x^*) = 0$. Thus, we can compute all derivative of invasion fitness directly from all corresponding derivatives of the growth rate, which facilitates calculations in what follows.

Substituting eq. (3.B) into $a(x,y) = \log(\rho(x,y))$, we get upon rearrangment

$$a(x,y) = \sum_{c\in\mathcal{C}} \log(w_c(x,y))q_c(y), \tag{A-71}$$

which is the familiar expression for the mutant growth rate under between generation (e.g., Tuljapurkar, 1990; Ferrière and Gatto, 1995). Differentiating, we get

$$\left.\frac{\partial a(x,y)}{\partial x}\right|_{x=y} = \sum_{c\in\mathcal{C}} \frac{1}{w_c(y,y)} \left.\frac{\partial w_c(x,y)}{\partial x}\right|_{x=y} q_c(y) \tag{A-72}$$

and

$$\left.\frac{\partial^2 a(x,y)}{\partial x \partial y}\right|_{x=y} = \sum_{c\in\mathcal{C}} \left[\frac{1}{w_c(y,y)} \frac{\partial^2 w_c(x,y)}{\partial x \partial y} q_c(y) - \frac{1}{w_c(y,y)^2} \frac{\partial w_c(x,y)}{\partial x} \frac{\partial w_c(x,y)}{\partial y} q_c(y) + \frac{1}{w_c(y,y)} \frac{\partial w_c(x,y)}{\partial x} \frac{\partial q_c(y)}{\partial y}\right]_{x=y} \tag{A-73}$$

which, after re-arrangements by factoring out $q_c(y)$ and using the property of logarithm gives eq. (3.F).

Appendix E.3 Class-specific regulation and production factors

Following the same construction of fitness factors as in section 3.2.1, we can also express $w_c(x,y)$ and $q_c(x,y)$ in terms of fitness factors as follows by writing

$$w_{c',c}(x,y) = \tilde{w}(\boldsymbol{\pi}_c(\mathbf{z}_c(x, \mathbf{e}^{\mathrm{z}}_c(x,y)), \mathbf{e}^{\pi}_c(x,y)), \mathbf{e}^{\mathrm{w}}_c(x,y)), \tag{A-74}$$

where we have indexed by c the production and regulating factors by the class of the bearer of this fitness, since in class structured population they will be class specific as both the phenotypes expressed by an individual and its experienced environment may depend on the class in which the individual resides. The right hand side of eq. (A-74) can also be applied to fitness $w_c(x,y)$ for temporal heterogeneity.

The distribution $q_c(x,y)$ represents the likelihood of being in a certain state. Since the occurrence of any state can depend on all factors affecting reproduction and survival across all states, we generally have

$$q_c(x,y) = \tilde{q}(\boldsymbol{\pi}(\mathbf{z}(x, \mathbf{e}^{\mathrm{z}}(x,y)), \mathbf{e}^{\pi}(x,y)), \mathbf{e}^{\mathrm{w}}(x,y)), \tag{A-75}$$

for some function $\tilde{q}$ mapping the collection of all fitness factors to probabilities of state occurrence, and where a vector $\mathbf{v}$ without subscript (where $\mathbf{v}$ could be $\boldsymbol{\pi}$, $\mathbf{z}$, $\mathbf{e}^{\mathrm{z}}$, $\mathbf{e}^{\pi}$, or $\mathbf{e}^{\mathrm{w}}$) concatenates the vectors with subscript, i.e. $\mathbf{v} = \{\mathbf{v}_c\}_{c\in\mathcal{C}}$. By substituting eqs. (A-74)–(A-75) into eq. (3.A), we can express invasion fitness as in eq. (18), i.e., as a mapping from all fitness factors to geometric growth ratio (and the same logic applies to $w_c(x,y)$ and $q_c(y)$ under temporal heterogeneity).

References

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

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

Akzcay, E. and J. V. Cleve. 2014. Pathways to social evolution: reciprocity, relatedness, and synergy. *Evolution* 68:2245–58.

Alger, I. 2023. Evolutionarily stable preferences. *Proceedings of the Royal Society of London Series B-Biological Sciences* 378.

Alger, I. and L. Lehmann. 2023. Evolution of semi-Kantian preferences in two-player assortative interactions with complete and incomplete Iiformation and plasticity. *Dynamic Games and Applications* 13:1288–1319.

Alipantris, C. D. and K. C. Border. 2006. Infinite Dimensional Analysis: a Hitchiker's Guide. Springer, Berlin, 3th edn.

Altenberg, L. and M. W. Feldman. 1987. Selection, generalized transmission and the evolution of modifier genes. I. The reduction principle. *Genetics* 117:559–572.

Altenberg, L., U. Liberman, and M. W. Feldman. 2017. Unified reduction principle for the evolution of mutation, migration, and recombination. *Proceedings of the National Academy of Sciences of the United States of America* 114:E2392–E2400.

Anderson, B. G., , S. Sojoudi, and M. Arcak. 2013. Evolutionary games on infinite strategy sets: convergence to Nash equilibria via dissipativity. arXiv:2312.08286v2.

André, J. B. and T. Day. 2007. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. *Journal of Theoretical Biology* 247:11–22.

Armstrong, R. A. and R. McGehee. 1980. Competitive exclusion. *The American Naturalist* 115:151–170.

Arnold, L., V. M. Gundlach, and L. Demetrius. 1994. Evolutionary formalism for products of positive random matrices. *The Annals of Applied Probability* 4:859–901.

Avila, P. and C. Mullon. 2023. Evolutionary game theory and the adaptive dynamics approach: adaptation where individuals interact. *Philosophical Transactions of the Royal Society B* 378:615255.

Barabas, G., R. D'Andrea, and S. Maccracken. 2008. Chesson's coexistence theory. *Ecological Monographs* 88:277–303.

Barton, N. H., D. Briggs, J. A. Eisen, G. D. B., and P. N. H. 2007. Evolution. Cold Spring Harbor Laboratory Press, New York.

Begon, M. and C. R. Townsend. 2021. Ecology: from Individuals to Ecosystems. Wiley, West Sussex, 5th edn.

Brännström, A., J. Johansson, and N. Von Festenberg. 2013. The hitchhiker's guide to adaptive dynamics. *Games* 4:304–328.

Cai, Y. and S. A. H. Geritz. 2020. Resident-invader dynamics of similar strategies in fluctuating environments. *Mathematical Biology* 81:907–959.

Cant, M. 2012. Suppression of social conflict and evolutionary transitions to cooperation. *American Naturalist* 179:293–301.

Case, T. J. 2000. An Illustrated Guide to Theoretical Ecology. Oxford University Press, Oxford.

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

Charnov, E. L. 1993. Life History Invariants. Oxford University Press, Oxford, UK.

Charnov, E. L. 1997. Trade-off-invariant rules for evolutionarily stable life histories. *Nature* 387:393–4.

Chesson, P. L. 2000a. General theory of competitive coexistence in spatially-varying environments. *Theoretical Population Biology* 58:211–237.

Chesson, P. L. 2000b. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* 31:343–66.

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

Conitzer, V. and T. Sandholm. 2008. New complexity results about Nash equilibria. *Games and Economic Behavior* 63:621–641.

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

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

de Mazancourt, C. and U. Dieckmann. 2004. Trade-off geometries and frequency-dependent selection. *The American Naturalist* 164:765–778.

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

Diekmann, O., M. Gyllenberg, and J. A. J. Metz. 2003. Steady-state analysis of structured population models. *Theoretical Population Biology* 63:309–338.

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

Eshel, I. 1983. Evolutionary and continuous stability. *Journal ot Theoretical Biology* 103:99–111.

Eshel, I. and M. W. Feldman. 1984. Initial increase of new mutants and some continuity properties of ESS in two-locus systems. *The American Naturalist* 124:631–640.

Eshel, I., U. Motro, and E. Sansone. 1997. Continuous stability and evolutionary convergence. *Journal of Theoretical Biology* 074:222–232.

Ewens, W. J. 2004. Mathematical Population Genetics. Springer-Verlag, New York.

Fairbairn, D. J. 2013. Odd couples: extraordinary differences between the sexes in the animal kingdom. Princeton University Press.

Ferrière, R. and M. Gatto. 1995. Lyapunov exponents and the mathematics of invasion in oscillatory or chaotic populations. *Theoretical Population Biology* 48:126–171.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Flintham, E., V. Savolainen, S. P. Otto, M. Reuter, and C. Mullon. 2024. The maintenance of genetic polymorphism underlying sexually antagonistic traits. *Evolution Letters* 9:259–272.

Fudenberg, D. and J. Tirole. 1991. Game Theory. MIT Press, Massachusetts.

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

Gigord, L. D., M. Macnair, and A. Smithson. 2001. Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid Dactylorhiza sambucina (L.) Soò. *Proceedings of the National Academy of Sciences* 98:6253–6255.

Gillespie, J. H. 1978. A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. *Theoretical population biology* 14:1–45.

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

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

Gross, M. R. 1996. Alternative reproductive strategies and tactics: diversity within sexes. *Trends in Ecology & Evolution* 11:92–98.

Grovers, J. P. 1997. Resource Competition. Chapman and Hall, New York.

Gyllenberg, M. and G. Meszéna. 2005. On the impossibility of coexistence of infinitely many strategies. *Mathematical Biology* 50:133–160.

Haigh, J. and J. Maynard Smith. 1972. Can there be more predators than prey? *Theoretical Population Biology* 3:290–299.

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

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

Hartl, D. and A. G. Clark. 2007. Principles of Population Genetics. Sinauer, Massachusetts, 4th edn.

Heifetz, A., C. Shannon, and Y. Spiegel. 2007a. The dynamic evolution of preferences. *Economic Theory* 32:251–286.

Heifetz, A., C. Shannon, and Y. Spiegel. 2007b. What to maximize if you must. *Journal of Economic Theory* 133:31 – 57.

Heino, M., J. A. J. Metz, and V. Kaitala. 1998. The enigma of frequency-dependent selection. *Trends in Ecology and Evolution* pp. 367–370.

Hingu, D., K. S. Mallikarjuna Rao, and A. J. Shaiju. 2008. Evolutionary stability of polymorphic population states in continuous games. *Dynamic Games and Applications* 8:141–156.

Hirshleifer, J. 1987. Conflict and rent-seeking success functions: ratio vs. difference models of relative success. *Public Choice* 63:101–112.

Hofbauer, J. and K. Sigmund. 1998. Evolutionary Games and Population Dynamics. Cambridge University Press.

Horn, R. A. and C. R. Johnson. 2012. Matrix Analysis. Cambridge University Press, Cambridge, UK, 2th

edn.

Houston, A. I. and J. M. McNamara. 1999. Models of Adaptive Behaviour: An Approach Based on State. Cambridge University Press, Cambridge, UK.

Ito, H. and U. Dieckmann. 2007. A new mechanism for recurrent adaptive radiations. *The American Naturalist* 170:E96–E111.

Ito, H. and A. Sasaki. 2016a. Evolutionary branching under multi-dimensional evolutionary constraints. *Journal of Theoretical Biology* 407:409–428.

Ito, H. and A. Sasaki. 2016b. Evolutionary branching under multi-dimensional evolutionary constraints. *Journal of Theoretical Biology* 407:409–428.

Jehle, A. G. and P. J. Reny. 2011. Advanced Microeconomics Theory. Prentice Hall, Essex.

Karlin, S. and J. McGregor. 1972. Polymorphisms for genetic and ecological systems with weak coupling. *Theoretical Population Biology* 3:210–38.

Kisdi, E. 2006. Trade-off geometries and the adaptive dynamics of two co-evolving species. *Evolutionary Ecology Research* 8:959–973.

Kisdi, E. 2015. Construction of multiple trade-offs to obtain arbitrary singularities of adaptive dynamics. *Journal of Mathematical Biology* 70:1093–1117.

Kisdi, E. and S. A. H. Geritz. 1999. Adaptive dynamics in allele space: evolution of genetic polymorphism by small mutations in a heterogeneous environment. *Evolution* 53:993–1008.

Kisdi, E. and S. A. H. Geritz. 2016. Adaptive dynamics of saturated polymorphisms. *Journal of Mathematical Biology* 72:1039–1079.

Kondepuli, D. and I. Prigogine. 2015. Modern Thermodynamics: From Heat Engines to Dissipative Structures. Wiley, New York, second edn.

Kushel, O. Y. 2019. Unifying matrix stability concepts with a view to applications. *SIAM Review* 61:643–729.

Křivan, V., R. Cressman, and C. Schneider. 2008. The ideal free distribution: A review and synthesis of the game-theoretic perspective. *Theoretical Population Biology* 73:403–425.

Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* 34:292–305.

Lehmann, L. 2022. Hamilton's rule, the evolution of behavior rules and the wizardry of control theory. *Journal of Theoretical Biology* 55.

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

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

Leimar, O. 2009. Multidimensional convergence stability. *Evolutionary Ecology Research* 11:191–208.

Leslie, P. H. and J. C. Gower. 1958. The properties of a stochastic model for two competing species. *Biometrika* 45:316–330.

Leslie, P. H. and J. C. Gower. 1960. The properties of a stochastic model for the predator-prey type of

interaction between two species. *Biometrika* 47:219–234.

Lessard, S. 1990. Evolutionary stability: one concept, several meanings. *Theoretical Population Biology* 37:159–170.

Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. *The American Naturalist* 87:331–333.

Levin, S. A. 1970. Community equilibria and stability, and an extension of the competitive exclusion principle. *The American Naturalist* pp. 413–423.

Lewontin, R. C. 1974. The Genetic Basis of Evolutionary Change. Columbia University Press.

Liberman, U. 1988. External stability and ESS: criteria for initial increase of a new mutant allele. *Journal of Mathematical Biology* 26:477–485.

Lion, S. 2018. Theoretical approaches in evolutionary ecology: environmental feedback as a unifying perspective. *American Naturalist* 191:21–44.

Lion, S., M. Boots, and A. Sasaki. 2022. Multimorph eco-evolutionary dynamics in structured populations in structured populations. *American Naturalist* 200:345–371.

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

Maire, M., M. Ackermann, and M. Doebeli. 2001. Evolutionary branching and the evolution of anisogamy. *Selection* 2:119–131.

Mas-Colell, A., M. D. Whinston, and J. R. Green. 1995. Microeconomic Theory. Oxford Unversity Press, Oxford.

May, R. M. 1973. Stability and Complexity in Model Ecosystems. Princeton University Press, New Jersey.

Maynard Smith, J. 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Maynard Smith, J. 1998. Evolutionary Genetics. Oxford University Press, Oxford, UK.

McNamara, J. M., C. E. Gasson, and A. I. Houston. 1999. Incorporating rules for responding into evolutionary games. *Nature* 401:368–71.

McNamara, J. M. and O. Leimar. 2020. Game Theory in Biology. Oxford University Press.

Meszéna, G., M. Gyllenberg, L. Pásztor, and J. A. J. Metz. 2006. Competitive exclusion and limiting similarity: a unified theory. *Theoretical Population Biology* 69:68–87.

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

Metz, J. A. J., S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs, and J. S. van Heerwaarden. 1996. Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. In van Strien, S. J. and S. M. Verduyn Lunel (eds.), *Stochastic and Spatial Structures of Dynamical Systems*, pp. 183–231. North-Holland, Amsterdam.

Metz, J. A. J., S. D. Mylius, and O. Diekmann. 2008. When does evolution optimize? *Evolutionary Ecology Research* 10:629–654.

Metz, J. A. J., R. M. Nisbet, and S. A. H. Geritz. 1992. How should we define fitness for general ecological scenarios? *Trends in Ecology and Evolution* 7:198–202.

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

Mode, C. J. 1971. Multitype branching processes: theory and applications. No. 34. American Elsevier Publishing Company, New York.

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

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

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

Mullon, C. and L. Lehmann. 2019. An evolutionary quantitative genetics model for phenotypic (co)variances under limited dispersal, with an application to socially synergistic traits. *Evolution* 73:1695–1728.

Nagylaki, T. 1992. Introduction to Population Genetics. Springer-Verlag, Heidelberg.

Narahari, Y. 2014. Game Theory and Mechanism Design. World Scientific, New Jersey.

Nurmi, T. and K. Parvinen. 2017. Evolution of specialization under non-equilibrium population dynamics. *Journal of Theoretical Biology* 321:63–77.

Ohtsuki, H., C. Rueffler, J. Y. Wakano, K. Parvinen, and L. Lehmann. 2020. The components of directional and disruptive selection in heterogeneous group-structured populations. *Journal of Theoretical Biology* 507:1–16.

Oliveira, R. F., M. Taborsky, and H. J. Brockmann. 2008. Alternative reproductive tactics: an integrative approach. In *Alternative reproductive tactics: An integrative approach*, pp. 1–22. Cambridge University Press.

Otto, S. P. and T. Day. 2007. A biologist's Guide to Mathematical Modeling in Ecology and Evolution. Princeton University Press, Princeton, NJ.

Papadimitriou, C. 2014. Algorithms, complexity, and the sciences. *Proceedings of the National Academy of Sciences of the United States of America* 111:15881–15887.

Papadimitriou, C. H. 2007. The complexity of finding Nash equilibrium. In Nisan, N., T. Roughgarden, E. Tardos, and V. V. (eds.), *Algorithmic Game Theory*, pp. 29–52. Cambridge University Press.

Pásztor, L., Z. Botta-Dukát, G. Magyar, T. Czárán, and G. Meszéna. 2016. Theory-based ecology: a Darwinian approach. Oxford University Press, Oxford.

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

Prigent, I. and C. Mullon. 2023. The molding of intraspecific trait variation by selection under ecological inheritance. *Evolution* 77:1–18.

Priklopil, T. and L. Lehmann. 2020. Invasion implies substitution in ecological communities with class-structured populations. *Theoretical Population Biology* 134:36–52.

Priklopil, T. and L. Lehmann. 2021. Metacommunities, fitness and gradual evolution. *Theoretical Population*

Biology 142:12–35.

Quirk, J. and R. Ruppert. 1965. Qualitative economics and the stability of equilibrium. *The Review of Economic Studies* 32:311–326.

Rand, D. A., H. B. Wilson, and J. M. McGlade. 1994. Dynamics and evolution: evolutionarily stable attractors, invasion exponents and phenotype dynamics. *Philosophical Transactions of the Royal Society B-Biological Sciences* 343:261–83.

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

Reeve, H. K. and P. W. Sherman. 1993. Adaptation and the goal of evolutionary research. *The Quarterly Review of Biology* 68:1–32.

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

Rousset, F. 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press, Princeton, NJ.

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

Sandholm, W. H. 2011. Population Games and Evolutionary Dynamics. MIT Press, Cambridge, MA.

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

Sasaki, A. and S. Ellner. 1995. The evolutionarily stable phenotype distribution in a random environment. *Evolution* 49:337–350.

Schmid, M., C. Rueffler, L. Lehmann, and C. Mullon. 2024. Resource variation within and between patches: where exploitation competition, local adaptation, and kin selection meet. *The American Naturalist* 203:E19–E34.

Schreiber, S. J. 2020. When do factors promoting genetic diversity also promote population persistence? A demographic perspective on Gillespie's SAS-CFF model. *Theoretical Population Biology* 133:141–149.

Sellis, D., B. J. Callahan, D. A. Petrov, and P. W. Messer. 2011. Heterozygote advantage as a natural consequence of adaptation in diploids. *Proceedings of the National Academy of Sciences* 108:20666–20671.

Siljestam, M. and C. Rueffler. 2024. Heterozygote advantage can explain the extraordinary diversity of immune genes. *Elife* pp. 1–35.

Sinervo, B. and C. M. Lively. 1996. The rock–paper–scissors game and the evolution of alternative male strategies. *Nature* 380:240–243.

Skaperdas, S. 1996. Contest success functions. *Economic Theory* 7:283–290.

Stephens, D. W. and J. R. Krebs. 1987. Foraging Theory. Princeton University Press, Princeton, NJ.

Strang, G. 2006. Linear Algebra and its Applications. Cengage Learning, Connecticut, fourth edn.

Stuart, A. M. and A. R. Humphries. 1996. Dynamical Systems and Numerical Analysis. Cambridge University Press, Cambridge, UK.

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

Sydsaeter, K., P. Hammond, A. Seierstad, and A. Strøm. 2008. Further Mathematics for Economic Analysis. Prentice Hall, Essex, 2th edn.

Tanny, D. 1981. On multitype branching processes in a random environment. *Advanced Apllied Probability* 13:464–497.

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

Taylor, P. D. and T. Day. 2004. Stability in negotiation games and the emergence of cooperation. *Proceedings of the Royal Society of London Series B-Biological Sciences* 271:669–674.

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

Tishkoff, S. A., R. Varkonyi, N. Cahinhinan, and et al. 2001. Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. *Science* 293:455–462.

Tuljapurkar, S. 1990. Population Dynamics in Variable Environments. Springer-Verlag, New York.

Turelli, M., D. W. Schemske, and P. Bierzychudek. 2001. Stable two-allele polymorphisms maintained by fluctuating fitnesses and seed banks: protecting the blues in Linanthus parryae. *Evolultion* 55:1283–98.

Van Cleve, J. 2023. Evolutionarily stable strategy analysis and its links to demography and genetics through invasion fitness. *Philosophical transactions of the Royal Society B* 378:20210496.

Van Dooren, T. J. 1999. The evolutionary ecology of dominance-recessivity. *Journal of theoretical Biology* 198:519–532.

Van Dooren, T. J., M. Durinx, and I. Demon. 2004. Sexual dimorphism or evolutionary branching? *Evolutionary Ecology Research* 6:857–871.

Vincent, T. L. and J. S. Brown. 2005. Evolutionary Game Theory, Natural Selection, and Darwinian Dynamics. Cambridge University Press, Cambridge.

Vincent, T. L., M. V. Van, and B. S. Goh. 1996. Ecological stability, evolutionary stability and the ESS maximum principle. *Evolutionary Ecology* 10,:567–591.

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

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

West-Eberhard, M. J. 2003. Developmental Plasticity and Evolution. Oxford University Press.

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

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

Yamamichi, M., A. D. Letten, and S. J. Schreiber. 2023. Eco-evolutionary maintenance of diversity in fluctuating environments. *Ecology Letters* 26:S152–S167.