A Coevolutionary Community Ecology

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Bioinformatics and Computational Biology
in the
College of Graduate Studies
University of Idaho
by
Robert Michael Week

Major Professor: Scott L. Nuismer, Ph.D.
Committee Members: Luke J. Harmon, Ph.D.; Paul A. Hohenlohe, Ph.D.;
Stephen M. Krone, Ph.D.
Department Administrator: David C. Tank, Ph.D.

August 2020

Authorization to Submit Dissertation

This dissertation of Robert Michael Week, submitted for the degree of Doctor of Philosophy with a Major in Bioinformatics and Computational Biology and titled "A Coevolutionary Community Ecology," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor:	Scott L. Nuismer, Ph.D.	Date
Committee Members:	Luke J. Harmon, Ph.D.	Date
	Paul A. Hohenlohe, Ph.D.	Date
	Stephen M. Krone, Ph.D.	Date
Department Chair:	David C. Tank, Ph.D.	Date

Abstract

Species in the wild exhibit an immense diversity of forms, functions and abundances. To understand this diversity we must consider the processes that have shaped it along with the biological details these processes emerge from. For example, species are often found interacting with other species in an ecological community and these interspecific interactions can lead to a wide variety of ecological and evolutionary processes. The frequency of interspecific interactions across a set of species is determined in part by the abundance of each species. More abundant species tend to interact more often. However, the fluctuations in species abundance is itself determined in part by interspecific interactions. The combination of these two causal pathways creates a feedback loop between the patterns of interspecific interactions among species and the patterns of species abundance. Although classical community ecology has largely focused on studying patterns of diversity through the lense of feedbacks between interspecific interactions and species abundances, much of this work has ignored an important aspect of species. In particular, species are not monolithic entities, but are comprised of diverse sets of individuals. Furthermore, these individuals are characterized by suites of behavioural and morphological traits that mediate the outcomes of interactions with other individuals, including lifetime reproductive output (i.e., fitness). When particular trait values are associated with increased fitness, the species experiences selection for those trait values. If those trait values are also heritable, so that offspring traits resemble parental traits, they will tend to increase in frequency. Hence, interactions among individuals provide building blocks from which both ecological and evolutionary processes emerge. In turn, the evolution of traits mediating interactions can modify the patterns and outcomes of interactions, generating feedbacks between the ecological and evolutionary processes emerging from individual interactions. Thus, to understand the patterns and dynamics of species and the ecological communities they compose, we must consider a holistic framework that treats the evolutionary and ecological responses of species to their interactions as the products of a common underlying phenomenon; interactions among individuals. In this dissertation I introduce a novel mathematical framework rigorously derived from biological first principles that accomplishes this goal. Using this framework to investigate communities of species competing along a resource gradient, I find a positive correlation between the strength of selection and the strength of competition across interacting species pairs. I then apply this framework to study the consequences of coevolutionary races for the stability of mutualistic interactions, demonstrating that the tendency for mutualisms to disintegrate into parasitisms depends on the phenotypic interface of the interaction. Finally, I introduce a new statistical method to measure the strength of coevolution between pairs of species in the wild. Applying this method to two well studied cases, I find support for the role of coevolution in both cases and calculate first ever estimates for the strength of coevolution in the wild. Overall, this work contributes to a coevolutionary theory of community ecology by developing rigorous mathematical tools that embrace the interwoven nature of ecological and evolutionary processes, deriving novel results on the consequences of coevolution in shaping interspecific interactions, and introducing a new statistical method to infer the coevolutionary process from empirical patterns. Future research based on this work may lead to a deeper synthetic understanding of the diverse and dynamical nature of life along with methods to forecast the responses of wild populations to our changing world.

Acknowledgements

I thank the multitude of organisms, organizations and abiotic processes that have provided support, have shaped who I am today, and have inspired me to always keep learning and always keep loving. In particular, I thank my adviser, Scott L. Nuismer, for the opportunity to pursue graduate studies at the University of Idaho, the Institute for Bioinformatics and Evolutionary Studies (IBEST) for encouragement, recognition and financial support and the National Science Foundation (NSF) for financial support. In particular, I would like to thank Dr. Paul Joyce. Paul co-authored a NSF grant with Scott that provided the initial financial support for this work and has become a source of inspiration in my studies of mathematical genetics. I deeply regret the loss of this brilliant mind. Members of the Nuismer lab circa 2015, including Ailene MacPherson, CJ Jenkins, Anahì Espìndola and E.T. (Emily Thornquist), provided important support and encouragement as I made the difficult transition from mathematics to biology. Current Nuismer lab members, including Beth Tuschhoff, Courtney Schreiner, Breanna Sipley, Tanner Varrelman, Andrew Basinski, Nathan Layman and Anna Sjodin, provided important support and encouragement as I rushed to finish this dissertation. I appreciate all the conversations with fellow Bioinformatics and Computational Biology students, Amanda Stahlke, Megan Ruffley and Austin Patton in particular, which have substantially contributed to my development as an evolutionary ecologist. I am also thankful for the community of ecologists and evolutionary biologists of the Palouse. It has been a pleasure to interact with these fantastic scientists and wonderful people. The mathematics and engineering communities at the University of Idaho also provided invaluable support starting in my undergraduate education. In particular, I thank Professors Jennifer Johnson-Leung, Somantika Datta and Hirotachi Abo for their excellence in teaching and insightful discussions. Finally, I thank the mathematics department of Clark Community College, where this journey began.

Dedication

To my parents for their enduring support, to my friends on the other side of the mountain, to the lizards basking in the sun and to the number twenty three.

Table of Contents

List of Tables

List of Figures

Chapter 1: A White Noise Approach to Evolutionary Ecology

Abstract

Although the evolutionary response to random genetic drift is classically modelled as a sampling process for populations with fixed abundance, the abundances of populations in the wild fluctuate over time. Furthermore, since wild populations exhibit demographic stochasticity, it is reasonable to consider the evolutionary response to demographic stochasticity and its relation to random genetic drift. Here we close this gap in the context of quantitative genetics by deriving the dynamics of the distribution of a quantitative character and the abundance of a biological population from a stochastic partial differential equation driven by space-time white noise. In the process we develop a useful set of heuristics to operationalize the powerful, but abstract theory of white noise and measure-valued stochastic processes. This approach allows us to compute the full implications of demographic stochasticity on phenotypic distributions and abundances of populations. We demonstrate the utility of our approach by deriving a quantitative genetic model of diffuse coevolution mediated by exploitative competition for a continuum of resources. In addition to trait and abundance distributions, this model predicts interaction networks parameterized by rates of interactions, competition coefficients, and selection gradients. Analyzing the relationship between selection gradients and competition coefficients reveals independence between linear selection gradients and competition coefficients. In contrast, absolute values of linear selection gradients and quadratic selection gradients tend to be positively correlated with competition coefficients. That is, competing species that strongly effect each others abundance tend to also impose selection on one another. This approach contributes to the development of a synthetic theory of evolutionary ecology by formalizing first principle derivations of stochastic models that underlie rigorous investigations of the relationship between feedbacks of biological processes and the patterns of diversity they produce.

1.1 Introduction

Current mathematical approaches to synthesize the dynamics of abundance and evolution in populations have capitalized on the fact that biological fitness plays a key role in determining both sets of dynamics. In particular, while covariance of fitness and genotype is the basis of evolution by natural selection, the mean fitness across all individuals in a population determines the growth, stasis or decline of abundance. Although this connection has been established in the contexts of population genetics (Crow and Kimura, 1970; Roughgarden, 1979), evolutionary game theory (Hofbauer and Sigmund, 1998; Lion, 2018; Nowak, 2006), quantitative genetics (Doebeli, 1996; Lande, 1982; Lion, 2018) and a unifying framework for these three distinct approaches to evolutionary theory (Champagnat et al., 2006), there remains a gap in incorporating the intrinsically random nature of abundance into the evolution of continuous traits. Specifically, in theoretical quantitative genetics the derivation of a populations response to random genetic drift is derived in discrete time under the assumption of constant effective population size using arguments based on properties of random samples (Lande, 1976). Though this approach conveniently mimics the formalism provided by the Wright-Fisher model

of population genetics, real population sizes fluctuate over time. Furthermore, since these fluctuations are themselves stochastic, it seems natural to derive expressions for the evolutionary response to demographic stochasticity and consider how the results relate to characterizations of random genetic drift. This can be done in continuous time for population genetic models without too much technical overhead, assuming a finite number of alleles (Gomulkiewicz et al., 2017; Lande et al., 2009; Parsons et al., 2010). However, for populations with a continuum of types, such as a quantitative trait, finding a formal approach to derive the evolutionary response to demographic stochasticity has remained a vexing mathematical challenge. In this paper we close this gap by combining the calculus of white noise with results on rescaled limits of measure-valued branching processes (MVBP) and stochastic partial differential equations (SPDE).

Our goal in this chapter has two components: 1) Establish a novel synthetic approach to theoretical evolutionary ecology that provides a formal connection between demographic stochasticity and random genetic drift in the context of quantitative traits. 2) Communicate some useful properties of space-time white noise, MVBP and SPDE to a wide audience of mathematical evolutionary ecologists. With this goal in mind we will not provide a rigorous treatment of any of these mathematically rich topics. Instead, we introduce a set of heuristics that only require the basic concepts of Riemann integration, partial differentiation and some exposure to Brownian motion and stochastic ordinary differential equations (SDE). A concise introduction to SDE and Brownian motion has been provided by Evans (2014).

Since MVBP are abstract mathematical objects and their associated literature tends to be written with an unfortunate amount of mathematical detail, their study is quite demanding. Hence, the use of MVBP in mainstream theoretical evolutionary ecology has been limited. However, they provide natural models of biological populations by capturing various mechanistic details. In particular, MVBP generalize classical birth-death processes, such as the Galton-Watson process (Dawson, 1993; Kimmel and Axelrod, 2015), to model populations of discrete individuals that carry some value in a given type-space. Selection can then be modelled by associating these values with average reproductive output and mutation can be incorporated using a model that determines the distribution of offspring values given their parental value. For population genetic models the type-space is the discrete set of possible alleles individuals can carry. In quantitative genetic models tracking the evolution of d-dimensional phenotypes, this type-space is typically set to the Euclidean space $\mathbb{R}^d$. By starting with branching processes we can implement mechanistic models of biological fitness that account for the phenotype of the focal individual along with the phenotypes and number of all other individuals in a population or community. By taking a rescaled limit, we can then use these detailed individual-based models to derive population-level models tracking the dynamics of population abundance and phenotypic distribution driven by selection, mutation and demographic stochasticity. Hence, rescaled limits of MVBP provide a means to derive mathematically tractable, yet biologically mechanistic models of eco-evolutionary dynamics.

For univariate traits (i.e., $d = 1$) Konno and Shiga (1988), Reimers (1989), Li (1998) and Champagnat et al. (2006) have shown rescaled limits for a large class of MVBP converge to solutions of SPDE. Although cases in which $d \geq 1$ can be treated using the so-called martingale problem formulation

(Dawson, 1993), the SPDE formulation provides a more intuitive description of the biological processes involved. We therefore focus on the case $d = 1$ here. This allows us to introduce a concrete set of heuristics for deriving SDE tracking the dynamics of abundance, phenotypic mean and phenotypic variance to a wide audience of mathematical evolutionary ecologists. Following our approach to simplify notation and develop heuristics for calculations, future work can use the martingale formulation to extend the results presented here for $d > 1$ and even for infinite-dimensional traits (Dawson, 1993; Stinchcombe et al., 2012). Rigorous introductions to SPDE and rescaled limits of MVBP have been respectively provided by Da Prato and Zabczyk (2014) and Etheridge (2000).

In this paper we begin in §1.2 by introducing the basic framework of our approach. We first outline the essential ideas behind deriving evolutionary dynamics from abundance dynamics using a deterministic partial differential equation (PDE). Following this, we review rescaled limits of MVBP, their associated SPDE and introduce an approach to derive SDE tracking the dynamics of abundance, phenotypic mean and phenotypic variance. This approach requires performing calculations with respect to space-time white noise processes and we provide heuristics for doing so in Appendix A.10. We then discuss consequences of the derived SDE for general phenotypic distributions and simplify their expressions by assuming normally distributed phenotypes. For added biological relevance, we incorporate models of inheritance and development following classical quantitative genetics. To demonstrate how our framework can be used to formulate a synthetic theory of evolutionary ecology, in §1.3 we derive a model of diffuse coevolution for a set of S species competing along a resource continuum. The basic approach follows classical niche theory to develop biological fitness as a function of niche parameters and niche locations of other individuals in the community. We then use this model to derive formula for selection gradients and competition coefficients. Finally, we investigate the relationship between selection gradients and competition coefficients using an analytical approximation.

1.2 The framework

At the core of our approach is a model of stochastic abundance dynamics for a structured population in continuous time and phenotypic space. From this stochastic equation we derive a system of SDE for the dynamics of total abundance, mean trait and additive genetic variance of a population. In particular, our approach develops a quantitative genetic theory of evolutionary ecology. A popular alternative to quantitative genetics is the theory of adaptive dynamics (Dieckmann and Law, 1996; Metz et al., 1996). As demonstrated by Page and Nowak (2002) and Champagnat et al. (2006), the canonical equation of adaptive dynamics can be derived from the replicator-mutator equation, which in turn can be derived from models of abundance dynamics, revealing a synthesis of mathematical approaches to theoretical evolutionary ecology. In this section we review derivations of the replicator-mutator equation and trait dynamics from abundance dynamics in the deterministic case. We then extend these formula along with related results to the case of random reproductive output (i.e., demographic stochasticity).

1.2.1 Deterministic dynamics

Finite number of types We start by considering the dynamics of an asexually reproducing population in a homogeneous environment. For simplicity, we first assume individuals are haploid and carry one of K alleles each with a different fitness expressed as growth rate before introducing a model involving a quantitative trait. Under these assumptions, the derivation of the evolution of allele frequencies due to natural selection can be derived from expressions of exponential growth. This, and a few related approaches, have been provided by Crow and Kimura (1970). Mutation can be included using a matrix of transition rates. Specifically, denoting ν_i the abundance of individuals with allele i, m_i the growth rate of allele i (called the Malthusian parameter in Crow and Kimura, 1970), μ_{ij} the mutation rate from allele i to allele j and assuming selection and mutation are decoupled (Bürger, 2000), we have

$$\frac{d\nu_i}{dt} = m_i\nu_i + \sum_{j=1}^{K}(\mu_{ji}\nu_j - \mu_{ij}\nu_i). \tag{1.1}$$

Starting from this model, we get the total abundance of the population as $N = \sum_i \nu_i$, the frequency of allele i as $p_i = \nu_i/N$ and the mean Malthusian fitness of the population as $\bar{m} = \sum_i p_i m_i$. Note we have used the abbreviation $\sum_i = \sum_{i=1}^{K}$ to simplify inline notation. Observing $\sum_{ij}\mu_{ji}\nu_j = \sum_{ij}\mu_{ij}\nu_i$, we use linearity of differentiation to derive the dynamics of abundance dN/dt as

$$\frac{dN}{dt} = \sum_{i=1}^{K} m_i\nu_i + \sum_{i,j=1}^{K}(\mu_{ij}\nu_j - \mu_{ji}\nu_i) = \bar{m}N. \tag{1.2}$$

To derive the dynamics of the allele frequencies $p_1, \ldots, p_K$, we use the quotient rule of elementary calculus to find

$$\frac{dp_i}{dt} = (m_i - \bar{m})p_i + \sum_{j=1}^{K}(\mu_{ji}p_j - \mu_{ij}p_i). \tag{1.3}$$

Two important observations of these equations include: (i) Mean Malthusian fitness $\bar{m}$ is equivalent to the population growth rate and thus determines the abundance dynamics of the entire population. (ii) Selection for allele i occurs when $m_i > \bar{m}$ and selection against allele i occurs when $m_i < \bar{m}$. Hence, as mentioned in the introduction, fitness plays a key role in determining both abundance dynamics and evolution.

Equation (1.3) is known in the field of evolutionary game theory as a replicator-mutator equation (Nowak, 2006). Instead of being explicitly focused on alleles, the replicator-mutator equation describes the fluctuations of relative abundances of various *types* in a population in terms of replication and annihilation rates of each type and hence can be used to model dynamical systems outside of evolutionary biology (Nowak, 2006).

Continuum of types Inspired by equations (1.1)-(1.3), we derive an analog of the replicator-mutator equation for a continuum of types (that is, for a quantitative trait). In particular, we model a continuously reproducing population with trait values $x \in \mathbb{R}$ and an abundance density $\nu(x,t)$ that represents

the amount of individuals in the population with trait value x at time t. Hence, the abundance density satisfies $N(t) = \int \nu(x,t)dx$ and $p(x,t) = \nu(x,t)/N(t)$ is the relative density of trait x which we also refer to as the phenotypic distribution. Note we have used the abbreviation $\int = \int_{-\infty}^{+\infty}$ to simplify inline notation.

In analogy with the growth rates m_i for equation (1.1) we write $m(\nu, x)$ as the growth rate associated with trait value x which depends on the abundance density ν. We assume mutation is captured by diffusion with coefficient $\frac{\mu}{2}$. Hence, we model the demographic dynamics of a population and the dynamics of a quantitative character simultaneously by the PDE

$$\frac{\partial}{\partial t}\nu(x,t) = m(\nu,x)\nu(x,t) + \frac{\mu}{2}\frac{\partial^2}{\partial x^2}\nu(x,t). \tag{1.4}$$

Equation (1.4) qualifies both as a semilinear evolution equation and also a scalar reaction-diffusion equation. Although the general theory of such equations is quite rich, it is also quite difficult (Evans, 2010; Zheng, 2004). Hence, to stay within the realms of analytical tractability and biological plausibility, we require a set of technical assumptions.

Firstly, our expression for the growth rate $m(\nu, x)$ is an abbreviation for $m((K\nu)(x,t), x)$ where K is an operator that accounts for nonlocal effects, such as resource competition, on growth rates (Champagnat et al., 2006; Volpert, 2014). In particular, we are concerned with operators K taking the form

$$(K\nu)(x,t) = \int_{-\infty}^{+\infty} \kappa(x-y)\nu(y,t)dy, \tag{1.5}$$

where κ is a non-negative bounded function.

Secondly, to ensure existence and uniqueness of solutions to (1.4), we further assume $m(\nu, x)$ is bounded above by some real number R and satisfies a continuity condition. In particular, for two abundance densities $\nu_1(x), \nu_2(x)$ with total abundances $N_1 = \int \nu_1(x)dx$, $N_2 = \int \nu_2(x)dx$, we assume that, for every positive number $M > 0$, there exists a constant $L_M > 0$ depending on M such that when $N_1, N_2 \leq M$, then

$$\int_{-\infty}^{+\infty} |m(\nu_1,x)\nu_1(x) - m(\nu_2,x)\nu_2(x)|\, dx \leq L_M \int_{-\infty}^{+\infty} |\nu_1(x) - \nu_2(x)|dx. \tag{1.6}$$

Existence and uniqueness of solutions to (1.4) also require an initial abundance density that is twice continuously differentiable and integrable. That is, we require the second partial derivative $\frac{\partial^2}{\partial x^2}\nu(x,0)$ to be continuous in x and $N(0) = \int |\nu(x,0)|dx < +\infty$.

These conditions satisfy some technical mathematical requirements, but to satisfy biological plausibility we also require the initial condition to be non-negative with finite trait mean and variance. That is, we assume $\nu(x,0) \geq 0$ for all $x \in \mathbb{R}$ and

$$-\infty < \bar{x}(0) = \int_{-\infty}^{+\infty} xp(x,0)dx < +\infty, \tag{1.7a}$$

$$\sigma^2(0) = \int_{-\infty}^{+\infty} (x - \bar{x}(0))^2 p(x,0)dx < +\infty, \tag{1.7b}$$

where $\bar{x}(t)$ and $\sigma^2(t)$ are respectively the phenotypic mean and variance at time $t \geq 0$. In Appendix A.1 we combine the above assumptions on growth rate, mutation and initial condition to prove that solutions to (1.4) satisfy $N(t), |\bar{x}(t)|, \sigma^2(t) < +\infty$ for all $t \geq 0$.

Equation (1.4) can be seen as an analog of equation (1.1) for a continuum of types. By assuming mutation acts via diffusion, the effect of mutation causes the abundance density $\nu(x,t)$ to flatten out across phenotypic space. In fact, if the growth rate is constant across x, then this model of mutation will cause $\nu(x,t)$ to converge to a flat line in x as $t \to \infty$. Interpreting the trait value x as location in geographic space, equation (1.4) becomes a well-studied model of spatially distributed population dynamics (Cantrell and Cosner, 2004).

Although clearly an idealized representation of biological reality, this model is sufficiently general to capture a large class of dynamics including density dependent growth and frequency dependent selection. As an example, logistic growth combined with stabilizing selection can be captured using the growth rate

$$m(\nu, x) = R - \frac{a}{2}(\theta - x)^2 - c\int_{-\infty}^{+\infty} \nu(y,t)dy = R - \frac{a}{2}(\theta - x)^2 - cN(t), \tag{1.8}$$

where $a > 0$ the is strength of abiotic stabilizing selection around the phenotypic optimum θ, $c > 0$ is the strength of intraspecific competition and we refer to R as the innate growth rate (see §1.3.3 below). In the language of population ecology, $r = R - \frac{a}{2}(\theta - x)^2$ is the intrinsic growth rate of the population (Chesson, 2000). We have set $\kappa(x - y) = 1$ so that competitive interactions cause the same reduction in fitness regardless of trait value.

This exemplary fitness function has a few convenient properties. First, the effect of competition induces a local carrying capacity on the population, leading to a finite equilibrium abundance over bounded subsets of phenotypic (or geographic) space. Second, abiotic selection prevents the abundance density from diffusing too far from the abiotic optimum. In particular, when $R > \frac{1}{2}\sqrt{a\mu} > 0$, $\bar{x}(0)$ is finite, $\sigma^2(0)$ is non-negative and finite and $N(0)$ is positive and finite, this leads to a unique stable equilibrium given by

$$\hat{N} = \tfrac{1}{c}(R - \tfrac{1}{2}\sqrt{a\mu}), \tag{1.9a}$$

$$\hat{\bar{x}} = \theta, \tag{1.9b}$$

$$\hat{\sigma}^2 = \sqrt{\frac{\mu}{a}}. \tag{1.9c}$$

We demonstrate this result in Appendix A.2. The equilibrial phenotypic variance predicted by this model coincides with a classic quantitative genetic result predicted by modelling the combined effects of Gaussian stabilizing selection and the Gaussian allelic model of mutation (Bürger, 2000; Johnson and Barton, 2005; Lande, 1975; Walsh and Lynch, 2018).

To derive a replicator-mutator equation from equation (1.4), we employ the chain rule from calculus. Writing $\bar{m}(t) = \int m(\nu, x)p(x,t)dx$ for the mean fitness, we have

$$\frac{d}{dt}N(t) = \frac{d}{dt}\int_{-\infty}^{+\infty} \nu(x,t)dx = \int_{-\infty}^{+\infty} \frac{\partial}{\partial t}\nu(x,t)dx$$
$$= \int_{-\infty}^{+\infty} m(\nu,x)\nu(x,t)dx + \int_{-\infty}^{+\infty} \frac{\mu}{2}\frac{\partial^2}{\partial x^2}\nu(x,t)dx$$
$$= N(t)\int_{-\infty}^{+\infty} m(\nu,x)p(x,t)dx = \bar{m}(t)N(t). \quad (1.10)$$

Using our assumptions on mutation and fitness, we show in Appendix A.1 $\nu(x,t)$ is twice differentiable with respect to x and $N(t) < +\infty$ for all $t \geq 0$. This implies that we are justified in swapping the order of differentiation and integration and the result $\int \frac{\partial^2}{\partial x^2}\nu(x,t)dx = 0$ can be derived from the fundamental theorem of calculus. Biological reasoning agrees with this latter result since mutation neither creates nor destroys individuals, but merely changes their type from their parental type. Taking the same approach, we derive the dynamics of the phenotypic distribution $p(x,t)$ in response to selection and mutation as

$$\frac{\partial}{\partial t}p(x,t) = \frac{\partial}{\partial t}\frac{\nu(x,t)}{N(t)} = \frac{1}{N^2(t)}\left(N(t)\frac{\partial}{\partial t}\nu(x,t) - \nu(x,t)\frac{d}{dt}N(t)\right)$$
$$= \frac{1}{N(t)}\Big(m(\nu,x)\nu(x,t) + \frac{\mu}{2}\frac{\partial^2}{\partial x^2}\nu(x,t) - \bar{m}(t)\nu(x,t)\Big)$$
$$= \Big(m(\nu,x) - \bar{m}(t)\Big)p(x,t) + \frac{\mu}{2}\frac{\partial^2}{\partial x^2}p(x,t). \quad (1.11)$$

This result closely resembles Kimura's continuum-of-alleles model (Kimura, 1965). The primary difference being that our model utilizes diffusion instead of convolution with an arbitrary mutation kernel. However, our model of mutation can be derived as an approximation to Kimura's model, which has been referred to as the Gaussian allelic approximation in reference to the distribution of mutational effects on trait values at each locus in a genome (Bürger, 1986, 2000; Johnson and Barton, 2005; Lande, 1975), the infinitesimal genetics approximation in reference to modelling continuous traits as being encoded by an infinite number of loci each having infinitesimal effect (Barton et al., 2017; Fisher, 1919) and the Gaussian descendants approximation in reference to offspring trait values being normally distributed around their parental values (Bulmer, 1971; Turelli, 2017).

To distinguish this model from previous models of phenotypic evolution we refer to PDE (1.4) from which (1.11) was derived as the Deterministic Asexual Gaussian allelic model with Abundance dynamics (abbreviated DAGA). Later, we will extend this model to include the effects of demographic stochasticity, which we refer to as the Stochastic Asexual Gaussian allelic model with Abundance dynamics (abbreviated SAGA).

Evolutionary dynamics We now apply DAGA to derive the dynamics of mean trait $\bar{x}$ and phenotypic variance σ^2. Both of these dynamics are expressible in terms of covariances with fitness. For an abundance distribution $\nu(x)$ and associated phenotypic distribution $p(x)$, the covariance of fitness and

phenotype across the population is defined as

$$\mathrm{Cov}_t\big(m(\nu,x),x\big) = \int_{-\infty}^{+\infty} \big(m(\nu,x) - \bar{m}(t)\big)\big(x - \bar{x}(t)\big)p(x,t)dx. \tag{1.12}$$

Hence, the dynamics of the mean trait $\bar{x}(t)$ can be derived as

$$\begin{aligned}\frac{d}{dt}\bar{x}(t) = \frac{d}{dt}\int_{-\infty}^{+\infty} xp(x,t)dx &= \int_{-\infty}^{+\infty} x\frac{\partial}{\partial t}p(x,t)dx \\ &= \int_{-\infty}^{+\infty} x\big(m(\nu,x) - \bar{m}(t)\big)p(x,t) + x\frac{\mu}{2}\frac{\partial^2}{\partial x^2}p(x,t)dx = \mathrm{Cov}_t\big(m(\nu,x),x\big).\end{aligned} \tag{1.13}$$

Equation (1.13) is a continuous time analog of the well known Robertson-Price equation without transmission bias (Frank, 2012; Lion, 2018; Price, 1970; Queller, 2017; Robertson, 1966). Whether or not the covariance of fitness and phenotype creates change in $\bar{x}$ to maximize mean fitness $\bar{m}$ depends on the degree to which selection is frequency dependent (Lande, 1976). Since this change is driven by a covariance with respect to phenotypic diversity, the response in mean trait to selection is mediated by the phenotypic variance. In particular, when $\sigma^2 = 0$, $\bar{x}$ will not respond to selection. The result $\int x\frac{\partial^2}{\partial x^2}p(x,t)dx = 0$ can be found by applying integration by parts.

Following the approach taken to calculate the evolution of $\bar{x}$, we find the response of phenotypic variation to this model of selection and mutation is

$$\frac{d}{dt}\sigma^2(t) = \mathrm{Cov}_t\Big(m(\nu,x),(x-\bar{x})^2\Big) + \mu. \tag{1.14}$$

For the sake of space we relegate the derivation of $d\sigma^2/dt$ to Appendix A.3. In the absence of mutation equation (1.14) mirrors the result derived by Lion (2018) for discrete phenotypes. From a statistical perspective, if we think of $(x-\bar{x})^2$ as a square error, then in analogy to the dynamics of the mean trait, we see that the response in σ^2 to selection can be expressed as a covariance of fitness and square error, which is defined in analogy to $\mathrm{Cov}_t(m(\nu,x),x)$. Just as for the evolution of $\bar{x}(t)$, this covariance also creates change in σ^2 that can either increase or decrease mean fitness $\bar{m}$, depending on whether or not selection is frequency dependent. The effect of selection on phenotypic variance can be positive or negative depending on whether selection is stabilizing or disruptive.

Extending DAGA to stochastic dynamics In Appendix A.12, we extend these results to include the effects of demographic stochasticity. The idea is to add an appropriate noise term to DAGA. Hence, we wish to study *stochastic* partial differential equations (SPDE) that provide natural generalizations of DAGA. Fortunately, rigorous first principle derivations of such SPDE have been provided by Li (1998) and Champagnat et al. (2006). The noise terms driving these SPDE are space-time white noise processes, which are random processes uncorrelated in both space and time. In Appendix A.10, we provide a set of heuristics for performing calculations with respect to space-time white noise, including methods to derive SDE from SPDE in analogy to our derivations of ordinary differential equations (ODE) from PDE above. Since our aim is to present this material to a wide audience of mathematical evolutionary ecologists, our treatment of space-time white noise and stochastic integration deviates

from standard definitions to remove the need for a detailed technical treatment. However, in Appendix A.11, we show our heuristics are consistent with the rigorous infinite-dimensional stochastic calculus presented in Da Prato and Zabczyk (2014). Using our simplified approach, the reader will only need some elementary probability and an intuitive understanding of SDE, including Brownian motion, in addition to the notions of Riemann integration and partial differentiation already employed.

To understand how SPDE can be derived from biological first principles, we provide in the following subsection an informal discussion of measure-valued branching processes (MVBP) (which treat populations as sets of discrete individuals) and their diffusion limits (which treat populations as a mass of "infinitesimal" individuals). We start by introducing a MVBP that models populations of individuals reproducing and passing away independently of each other and their trait value. We then discuss a so-called superprocess (Etheridge, 2000) derived from a diffusion limit of this MVBP and, in the case of univariate traits, an associated SPDE. Following this, we consider SPDE of superprocesses derived from MVBP accounting for biological fitness functions that depend on the trait values of interacting individuals.

Under the simplifying assumptions inherited from our treatment of deterministic dynamics and the additional assumption that the magnitude of demographic stochasticity is independent of trait values, we obtain as a special case a relatively simple expression for an SPDE that generalizes DAGA. The simplicity of our special case allows us to use properties of space-time white noise processes to derive a set of SDE that generalize equations (1.10), (1.13) and (1.14) to include the effects of demographic stochasticity. Classical expressions for the effects of random genetic drift on the evolution of mean traits are obtained as a further special case.

1.2.2 From branching processes to SPDE

In real populations individuals are born and potentially reproduce before they ultimately die. These three events provide the basic ingredients of a branching process. Mathematical investigations of such processes have a relatively deep history (Kendall, 1966). The most simple branching process, known as the Galton-Watson process, describes the number of individuals alive at a given time $t \geq 0$ as a non-negative random integer (Kimmel and Axelrod, 2015). In these models individuals give rise to random numbers of offspring which leads to a biological process known in theoretical ecology as demographic stochasticity (Gotelli, 2001). By increasing the rate at which individuals reproduce and decreasing their individual contributions to population size, Feller (1951) introduced a formal method to approximate branching processes with diffusion processes that are continuous in time and in state (i.e., population size is approximated as a continuously varying quantity). Since diffusion processes possess greater analytical tractability than branching processes, Feller's method, known as the diffusion limit, has acquired immense popularity particularly in the field of mathematical population genetics (Ewens, 2004).

For over the past half of a century a great deal of accomplishments have been achieved in formalizing the diffusion limits of structured branching processes that describe populations of individuals occurring in some continuous space (Barton and Etheridge, 2019; Bertoin and Le Gall, 2003; Dawson, 1975, 1978; Etheridge, 2008; Li, 1998; Méléard and Roelly, 1992, 1993; Perkins, 1992, 1995; Watanabe,

1968). This space can represent geographic space or, relevant to our context, phenotypic space. Mathematically, these processes can be formalized using MVBP (Dawson, 1993). In what follows we describe a particularly important MVBP known as branching Brownian motion (BBM). This process has been very useful in the study of SPDE due to its simplifying assumption that individuals do not interact. However, this assumption imposes an unfortunate restriction by precluding the modelling of ecological interactions. We therefore follow our discussion of BBM with a review of a few important results on spatially structured branching processes that account for interactions.

Branching Brownian motion A BBM tracks individuals navigating d-dimensional Euclidean space that reproduce and senesce between exponentially distributed intervals. Unlike other stochastic processes that take values in $\mathbb{R}^d$, BBM takes values in the set of *non-negative finite measures* over $\mathbb{R}^d$. An excellent introduction to the theory of measures has been provided by Axler (2019). Intuitively, one can think of a finite measure as a function that maps subsets of $\mathbb{R}^d$ to real numbers. In particular, denoting X a BBM, for a measurable subset $D \subset \mathbb{R}^d$ and a time $t \geq 0$, $X(D,t)$ returns the (random) integral number of individuals alive within the region D at time t. By concentrating on so-called *measurable* subsets (defined in Axler, 2019), we avoid pathological technical issues that would otherwise prevent the use of stochastic integration needed later. For brevity, we write $X(t)$ as shorthand for $X(D,t)$ evaluated at an arbitrary measurable subset D. To keep our focus on the biological application, we refer to a measure describing the state of a population, such as $X(t)$, as a population distribution.

It will be useful to interpret the BBM X as a sum of point masses. A point mass at $y \in \mathbb{R}^d$ can be represented by the Dirac measure δ_y which returns $\delta_y(D) = 1$ if $y \in D$ and $\delta_y(D) = 0$ if $y \notin D$. Hence, one can think of the measure δ_y as a probability measure for a degenerate random variable (i.e., a random variable that has no variance). Then, supposing at time t there are $n(t) \in \{0, 1, 2, \dots\}$ individuals with (random) trait values $x_1(t), \dots, x_{n(t)}(t)$ and denoting $\delta_{x_i(t)}$ the point mass located at $x_i(t)$, we can express $X(t)$ as

$$X(t) = \sum_{i=1}^{n(t)} \delta_{x_i(t)}. \tag{1.15}$$

The BBM has three main components:

1) **Branching rate:** In our formulation of BBM we assume lifetimes of individuals are exponentially distributed with death rate $\lambda > 0$ and reproduction occurs simultaneously with death. Biologically, this implies individuals are semelparous. An alternative formulation treats birth and death events separately to model iteroparity. However, under the appropriate rescaling, both approaches converge to the same diffusion limit (Champagnat et al., 2006; Li, 1998). We therefore choose the former approach for the sake of simplicity.

2) **Reproductive law:** When a birth event occurs we assume a random (possibly zero) number of offspring are produced. The distribution of offspring left is called the reproductive law or branching mechanism. Traditional biological notation utilizes W to denote the average lifetime reproductive output of an individual (i.e., fitness). However, since W is used here to denote either Brownian

motions or Brownian sheets, we adopt $\mathfrak{W}$ to denote fitness. The variance of reproductive output around this mean is denoted by V. The case of $\mathfrak{W} = 1$ is referred to as the critical condition. Since the process tracking the total number of individuals, calculated for time t as $X(\mathbb{R}^d, t)$, is a Galton-Watson branching process, it is well known the critical condition implies extinction in finite time with probability one, given the initial total number of individuals is finite (i.e., $X(\mathbb{R}^d, 0) < +\infty$) (Athreya and Ney, 1972).

3) **Spatial movement:** Each offspring is born at the current location of their parent. Immediately after birth they move around space according to d-dimensional Brownian motion with diffusion coefficient $\sqrt{\mu}$. In our context we interpret spatial movement as mutation so that the location where an individual dies represents the value of its expressed trait. Then an individual born at location $x \in \mathbb{R}^d$ that lives for $\tau > 0$ units of time will have a normally distributed trait centered on x with covariance matrix equal to $\tau\mu$ times the $d \times d$ identity matrix. Hence, offspring inherit normally distributed traits centered on their parental trait.

Although BBM provides a very flexible framework for simulating biological populations, its analytical tractability is limited due to the detailed description at the individual level. Evolutionary ecologists are often concerned with population level details, such as abundance, mean trait and variance around this mean trait. To extract tractable expressions that describe the evolution of such population level details while retaining key biological ingredients such as demographic stochasticity, we take a diffusion limit of BBM. There are several ways to do this, but we summarize a simple approach.

First we rescale the *mass* of individuals by N_0/n_0 for positive integers n_0 and some fixed positive continously valued number $N_0 > 0$. In particular, this means the point mass $\delta_{x_i(t)}$ used to represent the i-th individual of the population is replaced by $\frac{N_0}{n_0}\delta_{x_i(t)}$ for each $i = 1, \ldots, n(t)$. Hence, denoting $X^{(n_0)}(t)$ the rescaled version of the population distribution $X(t)$, we have

$$X^{(n_0)}(t) = \frac{N_0}{n_0} \sum_{i=1}^{n(t)} \delta_{x_i(t)}. \tag{1.16}$$

We also rescale the branching rate by $\lambda \to n_0$ and fitness by $\mathfrak{W} \to \mathfrak{W}^{1/n_0}$ and consider the limit as $n_0 \to \infty$. Note that the initial rescaled measures satisfy $X^{(n_0)}(\mathbb{R}^d, 0) = N_0$ for each $n_0 = 1, 2, \ldots$. If the sequence of rescaled initial population distributions $X^{(1)}(0), X^{(2)}(0), \ldots$ converges to a population distribution $\mathfrak{X}(0)$ with finite total abundance such that $\mathfrak{X}(\mathbb{R}^d, 0) = N_0$, the limiting process $\mathfrak{X}$ defined by $\mathfrak{X}(D, t) = \lim_{n_0 \to \infty} X^{(n_0)}(D, t)$ is a superprocess known as a super-Brownian motion (Etheridge, 2000; Watanabe, 1968). Technically, super-Brownian motion refers to the special case of $\mathfrak{W} = 1$ and $\mathfrak{X}(\mathbb{R}^d, 0) = 1$. However, we also refer to the superprocess limit $\mathfrak{X}$ as a super-Brownian motion for any $\mathfrak{W} > 0$ and $\mathfrak{X}(\mathbb{R}^d, 0) \geq 0$.

Instead of returning the integral number of individuals alive in a region of space, a super-Brownian motion returns the mass of the population concentrated in a region of space. Since we have rescaled individual mass by N_0/n_0 and took the limit $n_0 \to \infty$, individuals are no longer discrete units. Instead, the particle view of the population gets replaced by a blob spread across d-dimensional space. In

particular, the value of $\mathfrak{X}(D,t)$ is a continuously varying non-negative random variable for any $t \geq 0$ and $D \subset \mathbb{R}^d$ with an initial condition $\mathfrak{X}(0)$ that satisfies $\mathfrak{X}(\mathbb{R}^d, 0) = N_0$. The initial population distribution $\mathfrak{X}(0)$ can be chosen by specifying a sequence of initial rescaled population distributions $X^{(n_0)}(0)$ that converge to $\mathfrak{X}(0)$. Similar to the case of branching processes, when the initial total abundance is finite ($\mathfrak{X}(\mathbb{R}^d, 0) < +\infty$), we have finite total abundances ($\mathfrak{X}(\mathbb{R}^d, t) < +\infty$) for all $t > 0$ (Etheridge, 2000). For the sake of biological plausibility and analytical tractability, we always assume finite initial abundance.

Unfortunately, just as with cream cheese spread across too much toast, the blob perspective of the population may exhibit some complicated spatial discontinuities that make precise results difficult to obtain. However, for spatial dimension $d = 1$, it turns out that $\mathfrak{X}(t)$ is absolutely continuous with respect to the Lebesgue measure for each $t \geq 0$ (Konno and Shiga, 1988; Reimers, 1989). This means that we can write $\mathfrak{X}(D,t) = \int_D \nu(x,t)dx$ for some density process $\nu(x,t)$ that is continuous in both x and t and satisfies the SPDE

$$\frac{\partial}{\partial t}\nu(x,t) = (\ln \mathfrak{W})\nu(x,t) + \frac{\mu}{2}\frac{\partial^2}{\partial x^2}\nu(x,t) + \sqrt{V\nu(x,t)}\dot{W}(x,t), \tag{1.17}$$

where $\dot{W}(x,t)$ is a space-time white noise process as defined in Appendix A.10.

Since $\nu(x,t)$ is not generally differentiable in x, the spatial derivative in expression (1.17) is taken in the *weak* sense (Evans, 2010; Walsh, 1986). In particular, denoting $C_b^2(\mathbb{R})$ the set of bounded and twice continuously differentiable functions on $\mathbb{R}$, this means the spatial derivative in equation (1.17) is defined indirectly such that

$$\int_{\mathbb{R}} f(x)\frac{\partial^2}{\partial x^2}\nu(x,t)dx := \int_{\mathbb{R}} \nu(x,t)\frac{\partial^2}{\partial x^2}f(x)dx, \tag{1.18}$$

for every $f \in C_b^2(\mathbb{R})$. If $\nu(x,t)$ happens to be twice differentiable with respect to x, this definition coincides with integration by parts. Hence, to rigorously interpret SPDE (1.17), we rewrite it as

$$\begin{aligned}\int_{\mathbb{R}} \nu(x,t)f(x)dx - \int_{\mathbb{R}} \nu(x,0)f(x)dx \\ = \int_0^t \int_{\mathbb{R}} \nu(x,s)\left(\ln \mathfrak{W} f(x) + \frac{\mu}{2}\frac{\partial^2}{\partial x^2}f(x)\right)dsdx \\ + \sqrt{V}\int_0^t \int_{\mathbb{R}} f(x)\sqrt{\nu(x,s)}\dot{W}(x,s)dxds, \ \forall\ f \in C_b^2(\mathbb{R}).\end{aligned} \tag{1.19}$$

This expression is referred to as the *weak* solution of (1.17) (Walsh, 1986). Note that since $\nu(x,t)$ is the density of a finite measure, it is integrable for each $t \geq 0$. Thus, since $f \in C_b^2(\mathbb{R})$, it is bounded and hence for some $M > 0$, $|f(x)| \leq M$ for every $x \in \mathbb{R}$. This implies that $\int_0^t \int_{\mathbb{R}} |f(x)|^2\nu(x,s)dxds$ is finite with probability one for each $t \geq 0$. As mentioned above, $\nu(x,t)$ is also continuous with respect to t. In combination, these results enable us to employ the heuristics developed in Appendix A.10 to understand and evaluate the white-noise integral on the right-hand side of equation (1.19). In particular, evaluating equation (1.19) for the constant function $f(x) \equiv 1$ returns the total mass

process of the super-Brownian motion, which we refer to as the total abundance $N(t)$ in our biological application.

A convergence theorem for the diffusion limit of a generalization of BBM was established by Watanabe (1968). Dawson (1975) suggested that, for spatial dimension $d = 1$, this diffusion limit should admit a density process that satisfies a SPDE. Konno and Shiga (1988) and Reimers (1989) independently proved Dawson's suggestion was indeed correct. The diffusion limit of this more general branching process (in arbitrary spatial dimension) is referred to as a Dawson-Watanabe superprocess (Etheridge, 2000). Conditioning a Dawson-Watanabe superprocess to have constant mass returns a Fleming-Viot process (Etheridge and March, 1991; Perkins, 1991) which has been popular in studies of spatial population genetics. In particular, an extension of the Fleming-Viot process, which Etheridge (2008) has dubbed the Λ-Fleming-Viot process, was introduced by Bertoin and Le Gall (2003). Etheridge (2008) used the Λ-Fleming-Viot process to resolve a long-held technical challenge in modelling isolation by distance (Barton and Etheridge, 2019; Barton et al., 2013; Felsenstein, 1975).

Although this provides an impressive list of accomplishments, the Dawson-Watanabe superprocess falls short of our needs. In particular this process assumes individuals do not interact and thus precludes its ability to model nonlocal effects on the fitness of individuals, such as those produced via competitive interactions. Fortunately, this concern has been addressed by Champagnat et al. (2006); Evans and Perkins (1994); Li (1998); Méléard and Roelly (1992, 1993) and others, leading to constructions of superprocesses that account for interactions among individuals. In the next subsection we summarize relevant results in this area and introduce the SPDE that provides the basis for our approach to theoretical evolutionary ecology.

Interacting superprocesses Above we reviewed the diffusion limit of an especially tractable measure-valued branching process, the branching Brownian motion (BBM). However, we found the simplicity of this process restricts us from modelling the effects due to interactions between individuals. Here, we discuss superprocesses that account for such effects. The existence of diffusion limits for a class of MVBP allowing for individual interactions has been treated by Méléard and Roelly (1992, 1993). The interactions can manifest as dependencies of the spatial movement or reproductive law of individuals on their position x and the state of the whole population described by $X(t)$.

An important result of Méléard and Roelly (1992, 1993) is a theorem that provides sufficient conditions to construct a sequence of rescaled measure-valued branching processes that converge to a generalization of the Dawson-Watanabe superprocess that includes interactions. The rescaling is analogous to that described above for non-interacting Dawson-Watanabe superprocesses, but now the mean and variance in reproductive output $\mathfrak{W}(X(t), x)$ and $V(X(t), x)$, branching rate $\lambda(X(t), x)$ and mutation rate $\mu(X(t), x)$ are allowed to depend on the whole population $X(t)$ and individual location x. Since these parameters are now functions of $X(t)$ and x, we need to extend the diffusion limit presented for the non-interacting processes.

Although Méléard and Roelly (1992, 1993) provide a very general strategy for taking diffusion limits of interacting MVBP, we focus on a tractable, yet flexible approach. For simplicity we assume spatial dimension $d = 1$ and branching rate λ, mutation rate μ and variance in reproductive output V

are constant with respect to $X(t)$ and x so that the effects of interactions only manifest in the fitness $\mathfrak{W}(X(t), x)$ of individuals. When it exists, we denote by $\nu(x,t)$ the density of the limiting superprocess. We again rescale the branching rate by $\lambda \to n_0$ and individual mass by N_0/n_0. Then, the Malthusian fitness $m(\nu, x)$ can be calculated as

$$m(\nu, x) = \lim_{n_0 \to \infty} n_0 \left(\mathfrak{W}^{1/n_0}(X(t), x) - 1 \right). \tag{1.20}$$

In Figure 1.1 we demonstrate this rescaling in discrete time for a population experiencing stabilizing selection and logistic growth. In particular, we employed a rescaled fitness such that the limit (1.20) converges to the Malthusian fitness given by equation (1.8) in §1.2.1. Since time is discretized, the process we simulate is formally a branching random walk. For further details on our simulation see Appendix A.6.

Li (1998) built directly off of the construction of Méléard and Roelly (1992, 1993) to study properties of interacting superprocesses and, by assuming individual spatial movement occurs independently of location x and the entire population $X(t)$, showed the evolution of associated density processes can be described by a SPDE. Assuming the interactions manifest only in the reproductive law and that spatial movement follows Brownian motion with mutation rate μ independent of both $X(t)$ and x and the growth rate $m(\nu, x)$ is bounded both above and below, Li's (1998) result implies the interacting superprocess on one dimensional trait space has a density $\nu(x,t)$ which is non-negative, integrable, continuous in time and space and satisfies the SPDE

$$\frac{\partial}{\partial t}\nu(x,t) = m(\nu, x)\nu(x,t) + \frac{\mu}{2}\frac{\partial^2}{\partial x^2}\nu(x,t) + \sqrt{V\nu(x,t)}\dot{W}(x,t). \tag{1.21}$$

As mentioned in §1.2.1, we refer to SPDE (1.21) as the Stochastic Asexual Gaussian allelic model with Abundance dynamics (abbreviated SAGA).

By assuming growth rates are only bounded above, Evans and Perkins (1994) proved a result that demonstrates the existence and uniqueness for multiple interacting superprocesses. The result is known as a bivariate Girsanov transform and formally demonstrates existence for a pair of interacting superprocesses engaged in resource competition following our treatment provided in Appendix §A.7. By lumping the pair of interacting superprocesses into a single superprocess such that individuals are now represented by a discrete trait indicating which species they belong to in addition to their trait value, we can consider interactions with yet another superprocess and in this way extend the bivariate Girsanov transform to a multivariate Girsanov transform which then establishes existence of S superprocesses engaged in resource competition. Although we are unaware of conditions under which these superprocesses admit density processes that satisfy SPDE, the derivations of SDE from SPDE using weak solutions given below can be thought of as shorthand for deriving SDE from superprocesses by extending the dual space to include the test functions $f(x) = 1, x, x^2$ when possible. We therefore continue our treatment from the SPDE perspective in what follows.

Deriving SDE from SPDE Assuming solutions to SAGA are well defined, we can calculate the total mass process $N(t)$ using the weak solution (see eqn. (1.19)) of SPDE (1.21) with $f(x) \equiv 1$. This implies

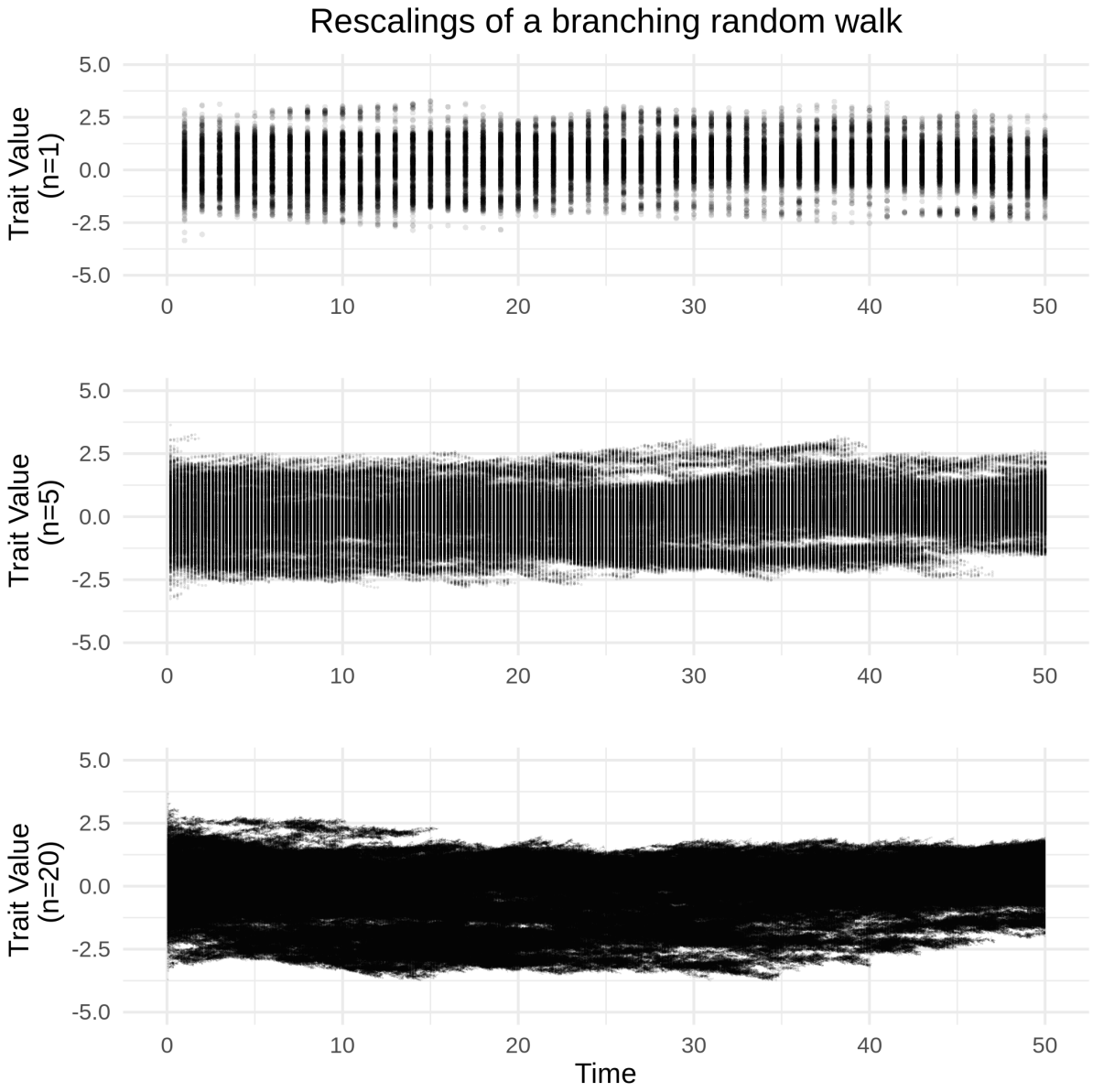

Figure 1.1: Rescaled sample paths of a branching random walk under stabilizing selection and logistic growth. The top plot displays a sample path without scaling ($n = 1$), the middle plot shows a sample path rescaled by $n = 5$ and the bottom plot shows a sample path rescaled by $n = 20$. For each of these plots we have set the innate growth rate to $R = 1.0$, phenotypic optimum to $\theta = 0$, strength of abiotic stabilizing selection to $a = 0.01$, sensitivity to competition to $c = 2 \times 10^{-3}$ and mutation rate to $\mu = 1 \times 10^{-3}$.

$$N(t) - N(0) = \int_0^t \int_{\mathbb{R}} \nu(x,s) \left(m(\nu,x) \cdot 1 + \frac{\mu}{2} \frac{\partial^2}{\partial x^2} 1 \right) + 1 \sqrt{V\nu(x,s)} \dot{W}(x,s) ds dx$$
$$= \int_0^t \bar{m}(s) N(s) dt + \sqrt{V} \int_0^t \sqrt{N(s)} d\hat{\mathbf{W}}_s \left(\sqrt{\nu(x,s)} \right), \quad (1.22)$$

where the population growth rate is calculated as

$$\bar{m}(t) = \frac{1}{N(t)} \int_{\mathbb{R}} m(\nu,x) \nu(x,t) dx, \quad (1.23)$$

and $\hat{\mathbf{W}}_s \left(\sqrt{\nu(x,s)} \right)$ is a standard Brownian motion (see Appendix A.10) given by

$$\int_0^t d\hat{\mathbf{W}}_s \left(\sqrt{\nu(x,s)} \right) = \int_0^t \int_{\mathbb{R}} \frac{\sqrt{\nu(x,s)}}{\sqrt{\int_{\mathbb{R}} \nu(x,s) dx}} \dot{W}(x,s) dx ds. \quad (1.24)$$

Setting $W_1(t) = \hat{\mathbf{W}}_t(\sqrt{\nu(x,t)})$, we can use traditional stochastic differential notation to write

$$dN = \bar{m} N dt + \sqrt{VN} dW_1. \quad (1.25)$$

To find the associated SDE for $\bar{x}(t)$ and $\sigma^2(t)$, we want to repeat the same approach for $f(x) = x, x^2$ and apply Itô's lemma. However, for these cases $f \notin C_b^2(\mathbb{R})$ since f will not be bounded. But, if we can show $\int_{\mathbb{R}} (|x| + x^2 + x^4) \nu(x,t) dx < +\infty$ for all $t > 0$ given this condition is satisfied by $\nu(x,0)$, then we can apply the weak solution of (1.21) to derive SDE for $\bar{x}(t)$ and $\sigma^2(t)$. To illustrate, let us suppose this is the case. Setting $\tilde{x}(t) = \int_{\mathbb{R}} x\nu(x,t) dx$, we have

$$\tilde{x}(t) = \tilde{x}(0) + \int_0^t \int_{\mathbb{R}} \nu(x,s) m(\nu,x) x + x \sqrt{V\nu(x,s)} \dot{W}(x,s) dx ds. \quad (1.26)$$

Similarly, setting $\tilde{\tilde{\sigma}}^2(t) = \int_{\mathbb{R}} x^2 \nu(x,t) dx$, we have

$$\tilde{\tilde{\sigma}}^2(t) = \tilde{\tilde{\sigma}}^2(0) + \int_0^t \int_{\mathbb{R}} \nu(x,s) \left(m(\nu,x) x^2 + \mu \right) + x^2 \sqrt{V\nu(x,s)} \dot{W}(x,s) dx ds. \quad (1.27)$$

Since $\bar{x}(t) = \tilde{x}(t)/N(t)$ and $\sigma^2(t) = \tilde{\tilde{\sigma}}^2(t)/N(t) - \bar{x}^2(t)$, we can use Itô's lemma to derive SDE for $\bar{x}(t)$ and $\sigma^2(t)$, which we perform in Appendix A.12. We make no attempt in finding sufficient conditions to ensure $\int_{\mathbb{R}} (|x| + x^2 + x^4) \nu(x,t) dx < +\infty$ and hence make no general assertions about the existence or uniqueness of $\bar{x}(t)$ or $\sigma^2(t)$. Regardless, we will later assume $\nu(x,t)$ can be approximated by a Gaussian curve in x for all $t \geq 0$. This assumption implies $\int_{\mathbb{R}} |x|^n \nu(x,t) dx < +\infty$ for all $n \in \{1, 2, \dots\}$ and for all $t \geq 0$ and hence guarantees the existence of $\bar{x}(t)$ and $\sigma^2(t)$ for all $t > 0$.

1.2.3 Equations of evolutionary and demographic dynamics

Here we return to our biological application of measure-valued branching processes and white noise calculus. In particular, we introduce SDE modeling the dynamics of total abundance $N(t)$, mean trait $\bar{x}(t)$ and trait variance $\sigma^2(t)$. In Appendix A.12 we derive these SDE from equation (1.21) of the main

text, which we refer to as the Stochastic Asexual Gaussian allelic model with Abundance dynamics (SAGA). SAGA is a SPDE model tracking the dynamics of an abundance density $\nu(x,t)$ driven by a Malthusian growth rate $m(\nu,x)$, diffusive mutation at the rate $\mu \geq 0$ and demographic stochasticity modulated by the variance in reproductive output $V \geq 0$. We begin by considering a general form of these expressions that hold for a wide range of phenotypic distributions. However, the generality of these expressions come at the cost of pragmatic applicability. Then, to transform these equations into useful tools for deriving stochastic models of biological populations, we consider a particular case by assuming normally distributed trait values. Following this simplification we incorporate a model of imperfect inheritance based on classical quantitative genetic theory. Under this model of inheritance we derive a SDE that tracks the stochastic dynamics of additive genetic variance $G(t)$ (the component of trait variance $\sigma^2(t)$ explained by additive allelic effects) in response to mutation, selection and demographic stochasticity. The expressions we present in this section provide a general set of tools for deriving stochastic eco-evolutionary dynamics and sets the stage for deriving our model of diffuse coevolution presented in §1.3.

General expressions As mentioned above, we begin with general expressions describing the stochastic evolution of abundance, mean trait and trait variance. In particular, these expressions hold for abundance densities $\nu(x,t)$ that are continuous (but not necessarily differentiable) in x and t such that $\int (x^2 + x^4)\nu(x,t)dx < +\infty$. The second condition implies the total abundance $N(t)$ is finite and the mass of the population is not spread out too far in phenotypic space (e.g., finite trait variance). Although future work is needed to understand the restrictions on growth rate $m(\nu,x)$ that rigorously justify the following expressions, our simulations suggest that quadratic stabilizing selection combined with an upper bound on $m(\nu,x)$ is sufficient. Assuming these conditions are met, we find the following system of SDE;

$$dN(t) = \bar{m}(t)N(t)dt + \sqrt{VN(t)}dW_1(t), \tag{1.28a}$$

$$d\bar{x}(t) = \mathrm{Cov}_t\Big(x, m(\nu,x)\Big)dt + \sqrt{V\frac{\sigma^2(t)}{N(t)}}dW_2(t), \tag{1.28b}$$

$$d\sigma^2(t) = \mathrm{Cov}_t\Big((x-\bar{x}(t))^2, m(\nu,x)\Big)dt + \left(\mu - V\frac{\sigma^2(t)}{N(t)}\right)dt + \sqrt{V\frac{\overline{(x-\bar{x}(t))^4} - \sigma^4(t)}{N(t)}}dW_3(t), \tag{1.28c}$$

where W_1, W_2 and W_3 are standard Brownian motions and barred expressions such as $\bar{m}(t)$ and $\overline{(x-\bar{x}(t))^4}$ are averaged quantities across phenotypic space with respect to $p(x,t) = \nu(x,t)/N(t)$. Dividing by dt one can interpret equations (1.28) as if they are ordinary differential equations, but this is not technically rigorous since Brownian motion is nowhere differentiable with respect to time. In Appendix A.12 we show that in general $W_1(t)$ is independent of both $W_2(t)$ and $W_3(t)$, but $W_2(t)$ and $W_3(t)$ may covary depending on the shape of $p(x,t)$.

Many known results follow directly from expressions (1.28). Firstly, assuming no variance in reproductive output so that $V = 0$ recovers the deterministic dynamics derived in §1.2.1. Alternatively, one can take $N(t) \to \infty$ to recover the deterministic dynamics for $\bar{x}(t)$ and $\sigma^2(t)$. Characteristically, we note the effect of demographic stochasticity on abundance grows with $\sqrt{N(t)}$. Hence, dividing by $N(t)$, we find the effects of demographic stochasticity on the per-capita growth rate diminish with increased abundance. Relating the response to demographic stochasticity derived here to the effect of random genetic drift derived in classic quantitative genetic theory, if we set $\sigma^2(t) = \sigma^2$ and $N(t) = N$ constant with respect to time, then integrating the stochastic term in equation (1.28b) over a single unit of time returns a normally distributed random variable with mean zero and variance equal to $V\sigma^2/N$. In particular, assuming perfect inheritance, when reproductive variance is unity ($V = 1$) this random variable coincides with the effect of random genetic drift on the change in mean trait over a single generation derived using sampling arguments (Lande, 1976). There is also an interesting connection with classical population genetics. A fundamental result from early population genetic theory is the expected reduction in diversity due to the chance loss of alleles in finite populations (Fisher, 1923; Wright, 1931). This expected reduction in diversity due to random genetic drift is captured by the third term in the deterministic component of expression (1.28c), particularly $-V\sigma^2(t)/N(t)$. The component of SDE (1.28c) describing random fluctuations in $\sigma^2(t)$ is more complicated and is proportional to the root of the difference between the centralized fourth moment of $p(x,t)$ and $\sigma^4(t)$.

These expressions can be used to investigate the dynamics of the mean and variance for a very general set of phenotypic distributions. However, in the next subsection we simplify these expressions by assuming normally distributed trait values, known as the Gaussian population assumption (Turelli 2017). Under this assumption we guarantee the existence of $\bar{x}(t)$ and $\sigma^2(t)$ for all t such that $N(t) > 0$. Furthermore, in Appendix A.12 we show that under the Gaussian case W_1, W_2 and W_3 are independent. Hence, although the Gaussian population assumption is very restrictive as a model of phenotypic diversity and, except for very special cases of Malthusian growth rates, is not formally justified, its exceedingly convenient properties make it an important initial approximation.

Particular results assuming a Gaussian phenotypic distribution By assuming normally distributed trait values, expressions (1.28) transform into efficient tools for deriving the dynamics of populations given a fitness function $m(\nu, x)$. Gaussian phenotypic distributions can be formally obtained through Gaussian, exponential or weak selection approximations together with a simplified model of mutation, genotype-phenotype mapping and asexual reproduction or random mating (Bürger, 2000; Lande, 1980; Turelli, 1984, 1986, 2017). Hence, given appropriate assumptions on selection, mutation and reproduction, the abundance density $\nu(x,t)$ can be approximated as a Gaussian curve in x when the ratio V/N is small (i.e., when the variance in reproductive output is much smaller than the population size). As with any diffusion approximation, this requires a sufficiently large abundance to accurately reflect the dynamics of populations. Therefore, models developed in this framework are not suitable for studies involving very small population sizes. Allowing for these restrictions, we assume

$$\nu(x,t) = \frac{N(t)}{\sqrt{2\pi\sigma^2(t)}} \exp\left(-\frac{\left(x - \bar{x}(t)\right)^2}{2\sigma^2(t)}\right). \tag{1.29}$$

Under this assumption we find in Appendix A.4 covariances with fitness can be written in terms of fitness gradients. In particular,

$$\mathrm{Cov}(x, m) = \sigma^2 \left(\frac{\partial \bar{m}}{\partial \bar{x}} - \overline{\frac{\partial m}{\partial \bar{x}}} \right), \tag{1.30a}$$

$$\mathrm{Cov}\Big((x - \bar{x})^2, m \Big) = 2\sigma^4 \left(\frac{\partial \bar{m}}{\partial \sigma^2} - \overline{\frac{\partial m}{\partial \sigma^2}} \right) \tag{1.30b}$$

and $\overline{(x - \bar{x})^4} = 3\sigma^4$. These results imply trait dynamics can be rewritten as

$$d\bar{x} = \sigma^2 \left(\frac{\partial \bar{m}}{\partial \bar{x}} - \overline{\frac{\partial m}{\partial \bar{x}}} \right) dt + \sqrt{V \frac{\sigma^2}{N}} dW_2, \tag{1.31a}$$

$$d\sigma^2 = 2\sigma^4 \left(\frac{\partial \bar{m}}{\partial \sigma^2} - \overline{\frac{\partial m}{\partial \sigma^2}} \right) dt + \left(\mu - V \frac{\sigma^2}{N} \right) dt + \sigma^2 \sqrt{\frac{2V}{N}} dW_3. \tag{1.31b}$$

These equations allow us to derive the response in trait mean and variance by taking derivatives of fitness, a much more straightforward operation than calculating a covariance for general phenotypic distributions. Note that in the above expressions, the partial derivatives of $\bar{m}$ represent frequency independent selection and the averaged partial derivatives of m represent frequency dependent selection. This relationship has already been pointed out by Lande (1976) for the evolution of the mean trait in discrete time, but here we see an analogous relationship holds in continuous time and also for the evolution of trait variance.

In what follows we generalize this result to the case when traits are imperfectly inherited. In this case, the phenotypic variance σ^2 is replaced by a genetic variance G. This genetic variance represents the component of σ^2 explained by additive effects among genetic loci encoding for the focal phenotype (Bulmer, 1971; Roughgarden, 1979; Walsh and Lynch, 2018). It is therefore fitting that G is referred to as the additive genetic variance.

Inheritance To model imperfect heritability we consider the relationship between expressed phenotypes $x \in \mathbb{R}$ and associated genetic values $g \in \mathbb{R}$ known as *breeding values*. The breeding value (called genotypic value in Bulmer, 1971; Walsh and Lynch, 2018) of an individual is the sum of additive effects of the alleles carried by the individual on its expressed trait. Hence, if the trait is encoded by L loci and the additive effect at locus l is a_l, then $g = \sum_{l=1}^{L} a_l$. The additive genetic variance G is just the variance of breeding values in a population (Bulmer, 1971; Walsh and Lynch, 2018). Following Lande (1975), we assume a mutation at locus l occurs with probability M and replaces the additive effect a_l with $a_l + \kappa_l$ where κ_l is normally distributed with a mean of zero and variance μ/M. Hence, we adopt the Gaussian allelic model of mutation. Next, we implement an infinitesimal approximation by assuming breeding values are determined by an infinite number of loci. Although very general infinitesimal approximations have been provided by Barton et al. (2017), for the sake of simplicity we employ a less technical approach. In particular, we rescale the mutational effects κ_l by $1/\sqrt{L}$ and take the limit $L \to \infty$. Then, denoting g' the breeding value of an offspring produced by a parent with breeding

value g and I_l the indicator variable determining whether or not a mutation occurs at locus l, we have

$$g' = g + \lim_{L\to\infty} \frac{1}{\sqrt{L}} \sum_{l=1}^{L} I_l \kappa_l. \tag{1.32}$$

This limit implies that g' has expected value g and variance μ. Thus, our assumptions yield the Gaussian descendants approximation coined by Turelli (2017). For a detailed treatment of breeding values, additive genetic variances and more general genetic architectures see Walsh and Lynch (2018).

Development Our treatment of the relationship between breeding values and expressed traits follows classical quantitative genetic assumptions such as those used by Bulmer (1971) to investigate the effect of selection on genetic variation. In particular, we ignore epistatic interactions so that effects at different loci combine additively. Since our treatment assumes haploid asexuals, there are no contributions of dominance or inbreeding depression to phenotypic variance. We assume expressed traits for given individuals are normally distributed around their breeding values with a fixed variance η. Hence, phenotypic variance decomposes as $\sigma^2 = G + \eta$. The variance η is referred to as developmental noise (Walsh and Lynch, 2018). For a fixed breeding value g, we denote the probability density of a randomly drawn expressed trait x by $\psi(x, g)$ so that

$$\psi(x, g) = \frac{1}{\sqrt{2\pi\eta}} \exp\left(-\frac{(x-g)^2}{2\eta}\right). \tag{1.33}$$

Selection On Breeding Values To include the relationship between breeding values and expressed traits in our framework, we write $\rho(g, t)$ as the abundance density of breeding values at time t so that

$$\int_{-\infty}^{+\infty} \rho(g, t) dg = \int_{-\infty}^{+\infty} \nu(x, t) dx = N(t). \tag{1.34}$$

We switch our focus from directly modelling the evolution of $\nu(x, t)$ to modelling the evolution of $\rho(g, t)$. Once $\rho(g, t)$ is determined, we can compute $\nu(x, t)$ via

$$\nu(x, t) = \int_{-\infty}^{+\infty} \rho(g, t)\psi(x, g) dg. \tag{1.35}$$

However, since selection acts on expressed phenotypes, we use the assumed relationship between breeding values and expressed traits to calculate the fitness of breeding values. To motivate the approach taken, consider the problem of inferring the breeding value of an individual given its expressed trait x. Denote $\mathfrak{g}$ a random variable representing the unknown breeding value. Under the above model of development we know x is a random sample from a normal distribution with mean $\mathfrak{g}$ and variance η. Maximizing likelihood suggests x is our best guess for $\mathfrak{g}$, but the actual value of $\mathfrak{g}$ is normally distributed around x with the variance η. Hence, for fixed x, we obtain $\psi(x, g)$ as the probability density of $\mathfrak{g}$. Thus, the mean fitness of a breeding value g across all individuals carrying g can be written as

$$m^*(\rho, g) = \int_{-\infty}^{+\infty} m(\nu, x)\psi(x, g) dx. \tag{1.36}$$

This is similar to the approach taken by Kimura and Crow (1978) to calculate the overall effects

of selection for expressed characters onto the changes in the distribution of alleles encoding those characters. However, instead of focusing on the frequencies of alleles at particular loci, our results focus on the densities of breeding values. With the relationship between $m(\nu, x)$ and $m^*(\rho, g)$ established, we define the evolution of $\rho(g,t)$ by the SPDE

$$\dot{\rho}(g,t) = m^*(\rho,g)\rho(g,t) + \frac{\mu}{2}\frac{\partial^2}{\partial^2 g}\rho(g,t) + \sqrt{V\rho(g,t)}\dot{W}(g,t). \tag{1.37}$$

Equation (1.37) is a stochastic generalization of DAGA, the deterministic PDE (1.4) from §1.2.1. However, equation (1.37) describes the evolution of the distribution of breeding values instead of expressed characters. Regardless, whether modelling expressed characters or breeding values, we refer to SPDE of the form (1.37) as Stochastic Asexual Gaussian allelic models with Abundance dynamics (abbreviated SAGA). In §1.2.2 we reviewed the origins of equation (1.37) and in Appendix A.10 we develop some heuristics to perform calculations with respect to the space-time white noise term $\dot{W}$.

Evolution Assuming $\rho(g,t)$ is Gaussian implies its mode coincides with $\bar{x}$. Furthermore, since $\sigma^2 = G + \eta$, we can use equation (1.36) and the chain rule from calculus (see Appendix A.5) to find

$$\frac{\partial \bar{m}}{\partial G} = \frac{\partial \bar{m}}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial G} = \frac{\partial \bar{m}}{\partial \sigma^2}, \tag{1.38a}$$

$$\overline{\frac{\partial m}{\partial G}} = \overline{\frac{\partial m}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial G}} = \overline{\frac{\partial m}{\partial \sigma^2}}. \tag{1.38b}$$

Thus, equations (1.31) become

$$d\bar{x} = G\left(\frac{\partial \bar{m}}{\partial \bar{x}} - \overline{\frac{\partial m}{\partial \bar{x}}}\right)dt + \sqrt{V\frac{G}{N}}dW_2, \tag{1.39a}$$

$$dG = 2G^2\left(\frac{\partial \bar{m}}{\partial G} - \overline{\frac{\partial m}{\partial G}}\right)dt + \left(\mu - V\frac{G}{N}\right)dt + G\sqrt{\frac{2V}{N}}dW_3. \tag{1.39b}$$

From expressions (1.39) we see that, under this model of inheritance, focusing on additive genetic variance G instead of the variance in expressed traits σ^2 makes no structural changes to the basic equations describing the dynamics of populations. Instead we see the role played by the variance of expressed traits is now being played by the additive genetic variance. In the next section, we make use of these expressions to develop a model of diffuse coevolution in a guild of S species competing along a resource continuum.

1.3 A model of diffuse coevolution

In this section we demonstrate the use of our framework by developing a model of diffuse coevolution across a guild of S species whose interactions are mediated by resource competition along a single niche axis. Because our approach treats abundance dynamics and evolutionary dynamics simultaneously, this model allows us to investigate the relationship between selection gradients and competition coefficients, which we carry out in what follows.

1.3.1 Formulation

The dynamics of phenotypic distributions and abundances have been derived above and so the only task remaining is the formulation of a fitness function. Our approach mirrors closely the theory developed by Levins (1968); MacArthur and Levins (1967) and MacArthur (1969, 1970, 1972). The most significant difference, aside from allowing evolution to occur, is the treatment of resource quality, which we replace with a model of abiotic stabilizing selection. A derivation is provided in Appendix A.7.

Abiotic selection and competition For species i we inherit the above notation for trait value, distribution, average, variance, abundance, etc., except with an i in the subscript. Real world examples of niche axes include the size of seeds consumed by competing finch species and the date of activity in a season for pollinators competing for floral resources. For mathematical convenience, we model the axis of resources by the real line $\mathbb{R}$. The value of a resource along this axis is denoted by the symbol ζ. For an individual in species i, we assume the resource utilization curve u_i can be written as

$$u_i(\zeta, x_i) = \frac{U_i}{\sqrt{2\pi w_i}} \exp\left(-\frac{(x_i - \zeta)^2}{2w_i}\right). \tag{1.40}$$

We further assume the niche center x_i is normally distributed among individuals in species i, but the niche breadth w_i and total niche utilization U_i are constant across individuals in species i and therefore cannot evolve. Among all the resources available (ie., among all possible real numbers), we suppose species i benefits most from resources with value $\theta_i \in \mathbb{R}$. In the absence of competition, we further suppose individuals leave on average Q_i offspring when their utilization curve is concentrated at θ_i (that is, when $x_i = \theta_i$ and $w_i = 0$). We assume the benefits for individuals of species i derived from resources with value $\zeta \in \mathbb{R}$ decreases as $(\zeta - \theta_i)^2$ increases at a rate $A_i > 0$. In particular, we assume abiotic stabilizing selection along the resource axis can be modelled by the curve

$$e_i(\zeta) = Q_i \exp\left(-\frac{A_i}{2}(\theta_i - \zeta)^2\right). \tag{1.41}$$

The effect of abiotic stabilizing selection on the fitness for an individual of species i with niche location x_i is then given by

$$\int_{-\infty}^{+\infty} e_i(\zeta) u_i(\zeta, x_i) d\zeta = \frac{Q_i U_i}{\sqrt{A_i w_i + 1}} \exp\left(-\frac{A_i}{2(A_i w_i + 1)}(\theta_i - x_i)^2\right). \tag{1.42}$$

To determine the potential for competition between individuals with niche locations x_i and x_j, belonging to species i and j respectively, we compute the niche overlap

$$\mathcal{O}_{ij}(x_i - x_j) = \int_{-\infty}^{+\infty} u_i(\zeta, x_i) u_j(\zeta, x_j) d\zeta = \frac{U_i U_j}{\sqrt{2\pi(w_i + w_j)}} \exp\left(-\frac{(x_i - x_j)^2}{2(w_i + w_j)}\right). \tag{1.43}$$

To map the degree of niche overlap to fitness, we assume competition between individuals with niche locations x_i and x_j decreases the expected reproductive output for the individual in species i at the rate $c_i \mathcal{O}_{ij}(x_i - x_j)$ for some $c_i > 0$. We refer to c_i as the strength of competition for species i.

The term $c_i\mathcal{O}_{ij}(x_i - x_j)$ coincides with a special case of a term used to capture competition in Dawson's geostochastic logistic model, an SPDE model developed to study the combined effects of demographic stochasticity, spatial dispersion and locally finite carrying capacity (Dawson, 1978). Recalling the constraints on fitness discussed in §1.2.1, we see this model of competition also coincides with $\kappa(x_i - x_j) = \mathcal{O}_{ij}(x_i - x_j)$, where κ was introduced in §1.2.1 to capture nonlocal effects of the abundance density $\nu(x,t)$ on the fitness of individuals (Champagnat et al., 2006; Volpert, 2014).

The fitness function Assuming the effects due to competitive interactions and abiotic stabilizing selection on the expected reproductive output of individuals accumulates multiplicatively, we derive in Appendix A.7 an expression for the expected reproductive output of individuals in each. Applying a series of diffusion limits, we then find the following expressions for the Malthusian growth rate associated with trait value x for species i along with the population growth rate of species i:

$$m_i(x) = R_i - \frac{a_i}{2}(\theta_i - x)^2 - c_i \sum_{j=1}^{S} N_j U_i U_j \sqrt{\frac{\tilde{b}_{ij}}{2\pi}} e^{-\frac{\tilde{b}_{ij}}{2}(x-\bar{x}_j)^2}, \tag{1.44a}$$

$$\bar{m}_i = R_i - \frac{a_i}{2}\Big((\theta_i - \bar{x}_i)^2 + G_i + \eta_i\Big) - c_i \sum_{j=1}^{S} N_j U_i U_j \sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i-\bar{x}_j)^2}, \tag{1.44b}$$

where a_i is the strength of abiotic stabilizing selection on species i. The variables $\tilde{b}_{ij}, b_{ij}$ determine the sensitivity of competitive effects on species i to differences in trait centers between species i and j. We refer to R_i as the innate growth rate of species i to distinguish it from the intrinsic growth rate commonly referred to in the field of population ecology. These are composite parameters given by the following expressions:

$$R_i = \ln\left(\frac{Q_i U_i}{\sqrt{1 + A_i w_i}}\right), \tag{1.45a}$$

$$a_i = \frac{A_i}{1 + A_i w_i}, \tag{1.45b}$$

$$\tilde{b}_{ij}(t) = \big(w_i + w_j + \eta_j + G_j(t)\big)^{-1}, \tag{1.45c}$$

$$b_{ij}(t) = b_{ji}(t) = \big(w_i + w_j + \eta_i + \eta_j + G_i(t) + G_j(t)\big)^{-1}. \tag{1.45d}$$

1.3.2 The model

In Appendix A.7 we combine equations (1.28a), (1.39) and (1.44) to find

$$dN_i = \left\{R_i - \frac{a_i}{2}\Big((\theta_i - \bar{x}_i)^2 + G_i + \eta_i\Big) - c_i \sum_{j=1}^{S} N_j U_i U_j \sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i-\bar{x}_j)^2}\right\} N_i dt + \sqrt{V_i N_i}\, dW_1, \tag{1.46a}$$

$$d\bar{x}_i = \left\{a_i G_i(\theta_i - \bar{x}_i) - c_i G_i\Big(\sum_{j=1}^{S} N_j U_i U_j b_{ij}(\bar{x}_j - \bar{x}_i)\sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i-\bar{x}_j)^2}\Big)\right\} dt + \sqrt{V_i \frac{G_i}{N_i}}\, dW_2, \tag{1.46b}$$

$$dG_i = \Bigg\{ c_i {G_i}^2 \Bigg(\sum_{j=1}^{S} N_j U_i U_j b_{ij} \left(1 - b_{ij}(\bar{x}_i - \bar{x}_j)^2\right) \sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i - \bar{x}_j)^2} + N_i U_i^2 b_{ii} \sqrt{\frac{b_{ii}}{2\pi}} + \mu_i - a_i {G_i}^2 - V_i \frac{G_i}{N_i} \Bigg) \Bigg\} dt + G_i \sqrt{\frac{2V_i}{N_i}} dW_3. \quad (1.46c)$$

Model behavior Despite the convoluted appearance of system (1.46), there are some nice features that reflect biological reasoning. For example, the dynamics of abundance are just a generalization of Lotka-Volterra dynamics. In particular, the effect of competition with species j on the fitness of species i grows linearly with N_j. However, as biotic selection pushes $\bar{x}_i$ away from $\bar{x}_j$, the effect of competition with species j on the fitness of species i rapidly diminishes due to the Gaussian weights capturing a reduction in niche overlap. These Gaussian weights have been usefully employed to capture interaction preference in recent investigations of coevolution in mutualistic networks (de Andreazzi et al., 2019; Guimarães et al., 2017; Medeiros et al., 2018). The divergence of $\bar{x}_i$ and $\bar{x}_j$ due to competition is referred to in the community ecology literature as character displacement (Brown and Wilson, 1956). We also see that the fitness of species i drops quadratically with the difference between $\bar{x}_i$ and the abiotic optimum θ_i. Hence, abiotic selection acts to pull $\bar{x}_i$ towards θ_i.

The response in mean trait $\bar{x}_i$ to natural selection is proportional to the amount of heritable variation in the population, represented by the additive genetic variance G_i. However, we have that G_i is itself a dynamic quantity. Under our model, abiotic stabilizing selection erodes away heritable variation at a rate that is independent of both N_i and $\bar{x}_i$. The effect of competition on G_i is a bit more complicated. When $b_{ij}(\bar{x}_i - \bar{x}_j)^2 < 1$, competition with species j acts as diversifying selection which tends to increase the amount of heritable variation. However, when $b_{ij}(\bar{x}_i - \bar{x}_j)^2 > 1$, competition with species j acts as directional selection and reduces G_i. In the following subsections we demonstrate the behavior of system (1.46) by plotting numerical solutions and investigate implications for the relationship between the strength of ecological interactions and selection.

Community dynamics For the sake of illustration we numerically integrated system (1.46) for a richness of $S = 100$ species. We assumed homogeneous model parameters across species in the community as summarized by Table 1.1. We repeated numerical integration under the two scenarios of weak and strong competition. For the first scenario of weak competition we set $c = 1.0 \times 10^{-7}$ and for the second scenario of strong competition we set $c = 5.0 \times 10^{-6}$. With these two sets of model parameters, we simulated our model for 1000.0 units of time. For both scenarios, we initialized the trait means to $\bar{x}_i = 0.0$, additive genetic variances to $G_i = 10.0$ and abundances to $N_i = 1000.0$ for each $i = 1, \ldots, S$.

Temporal dynamics for each scenario are provided in Figure 1.2. This figure suggests weaker competition leads to smoother dynamics and a higher degree of organization within the community. Considering expression (1.46a) we note that, all else equal, relaxed competition allows for larger growth rates which promote greater abundances. From (1.46a) we also note that the per-capita effects on demographic stochasticity diminish with abundance. To see this, divide both sides by N_i.

Inspecting expressions (1.46b) and (1.46c), we see that larger abundances also erode the effects of demographic stochasticity on the evolution of mean trait and additive genetic variance. These effects

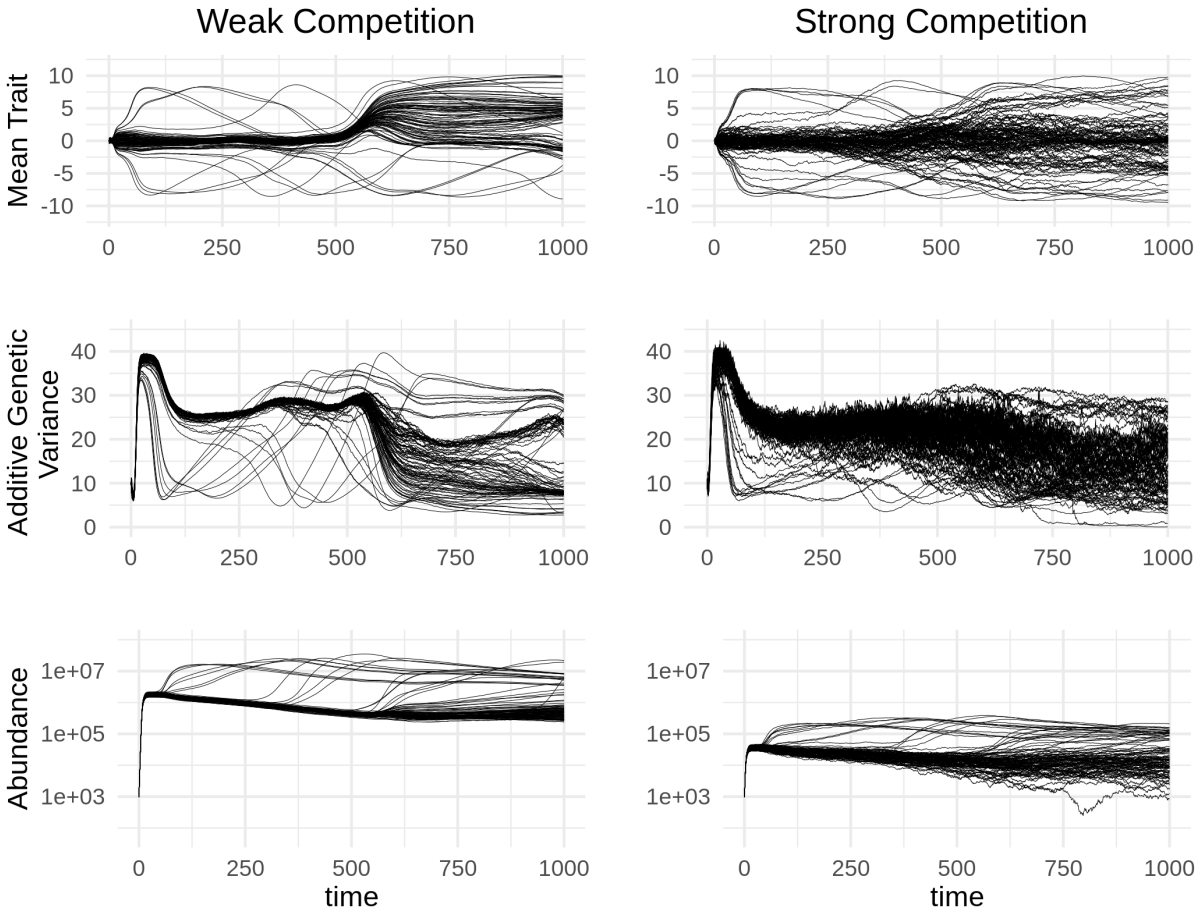

Figure 1.2: Temporal dynamics of mean trait (top), additive genetic variance (middle) and abundance (bottom) for the scenario of weak competition (left) and strong competition (right).

Table 1.1: Values of model parameters used for numerical integration.

Parameter	Description	Value
S	species richness	100
R	innate growth rate, see §1.3.3	1.0
θ	abiotic optimum	0.0
a	strength of abiotic selection	0.01
c	sensitivity to competition	$\{1.0 \times 10^{-7}, 5.0 \times 10^{-6}\}$
w	niche breadth	0.1
U	total niche use	1.0
η	developmental noise	1.0
μ	mutation rate	1.0×10^{-7}
V	variance of reproductive output	5.0

were already noted in §1.2.3, and thus are not a consequence of our model of coevolution per-se, but we revisit them here since Figure 1.2 demonstrates the importance of demographic stochasticity in structuring ecological communities even when populations are very large. Hence, contrary to the common assumption that stochastic effects can be ignored for large populations, we find that minute asymmetries generated by demographic stochasticity remain significant drivers of community structure. In particular, although we initialized each species with identical state variables and model parameters, we found an enormous amount of asymmetry in both the evolutionary and abundance dynamics and even some peculiar synchronized shifts. Although future work may show these bizarre features always dissipate after the system has been given sufficient time to evolve, we see demographic stochasticity has pronounced effects on communities experiencing non-equilibrium dynamics.

Although Figure 1.2 displays interesting patterns in the dynamics of abundance and trait evolution, we are interested in developing a quantitative understanding of the relationship between abundance dynamics and trait evolution. In the following subsection we take a step in this direction by approximating correlations between competition coefficients and components of selection gradients induced by interspecific interactions.

1.3.3 The relation between the strength of ecological interactions and selection

Here we investigate the relationship between competition coefficients, which measure the effect of ecological interactions on abundance dynamics, with selection gradients, which measure the magnitude and direction of selection on mean trait and trait variance. We start by considering the expressions of absolute competition coefficients implied by equations (1.46). However, it turns out absolute competition coefficients display some unfortunate behaviour with respect to our model. We therefore introduce a slightly modified form of absolute competition coefficients. We then provide formula for the components of linear and quadratic selection coefficients corresponding to the effects of interspecific interactions. Lastly, we use a high richness (large S) approximation to obtain analytical approximations of the means, variances and covariances between competition coefficients and selection gradients across the community. Associated calculations are provided in Appendix A.9.

Competition coefficients Relating our treatment of resource competition to theoretical community ecology, the absolute competition coefficient $\tilde{\alpha}_{ij}$, which measures the effect of species j on the growth rate of species i (sensu Chesson, 2000), becomes a dynamical quantity that can be written as

$$\begin{aligned}\tilde{\alpha}_{ij}(t) = \frac{c_i}{r_i(t)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_i(x,t) p_j(y,t) \mathcal{O}_{ij}(x,y) dx dy \\ = \frac{c_i U_i U_j}{r_i(t)} \sqrt{\frac{b_{ij}(t)}{2\pi}} \exp\left(-\frac{b_{ij}(t)}{2} \left(\bar{x}_i(t) - \bar{x}_j(t) \right)^2 \right), \quad (1.47)\end{aligned}$$

where

$$r_i(t) = R_i - \frac{a_i}{2}\Big((\bar{x}_i(t) - \theta_i)^2 + G_i(t) + \eta_i \Big), \quad (1.48)$$

is the intrinsic growth rate of species i. Then, $dN_i(t)$ can be expressed as

$$dN_i(t) = r_i(t)\left(1 - \sum_{j=1}^{S} \tilde{\alpha}_{ij}(t)N_j(t)\right)N_i(t)dt + \sqrt{V_i N_i(t)}dW_1(t). \tag{1.49}$$

Following our model, the classically defined absolute competition coefficient for species i is parameterized with the intrinsic growth rate of species i appearing in the denominator. In turn, these intrinsic growth rates depend on the balance between the innate growth rate R_i and the effect of abiotic stabilizing selection. However, this balance further depends on mean trait and additive genetic variance, which evolve freely. This leads to the potential for the signage of r_i to switch between positive and negative which implies the potential for infinite absolute competition coefficients. Furthermore, we see these competition coefficients are influenced by abiotic stabilizing selection instead of solely capturing the effects of inter/intraspecific interactions. Hence, we find it necessary to introduce a modification of the absolute competition coefficient $\tilde{\alpha}_{ij}$ that avoids these caveats. In particular, we define

$$\alpha_{ij} = r_i\tilde{\alpha}_{ij} = c_i U_i U_j \sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}\left(\bar{x}_i - \bar{x}_j\right)^2}. \tag{1.50}$$

We call α_{ij} the *specific* competition coefficient mediating the effects of species j on the growth rate of species i. Under this parameterization, the abundance dynamics of species i is now expressed as

$$dN_i = \left(r_i - \sum_{j=1}^{S} \alpha_{ij}N_j\right)N_i dt + \sqrt{V_i N_i}dW_1. \tag{1.51}$$

Selection gradients Linear and quadratic selection gradients have been defined by Lande and Arnold (1983). While the linear selection gradient β measures the effect of selection on mean trait evolution, the stabilizing selection gradient γ measures the effect of selection on additive genetic or phenotypic variance. Since these quantities are classically defined with respect to discrete-time models of trait evolution, we provide the analogous definitions for continuous-time models in Appendix A.8. Following our model of diffuse coevolution, we then show these selection gradients can be additively partitioned into components due to interactions with each species and abiotic stabilizing selection. In particular, we find the components of linear and quadratic selection gradients for species i induced by species j are given respectively by

$$\beta_{ij} = c_i U_i U_j N_j b_{ij}(\bar{x}_i - \bar{x}_j)\sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i - \bar{x}_j)^2}, \tag{1.52a}$$

$$\gamma_{ij} = c_i U_i U_j N_j b_{ij}\left(1 - b_{ij}(\bar{x}_i - \bar{x}_j)^2\right)\sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i - \bar{x}_j)^2},\ i \neq j, \tag{1.52b}$$

$$\gamma_{ii} = 2c_i N_i U_i^2 b_{ii}\sqrt{\frac{b_{ii}}{2\pi}},\ i = j. \tag{1.52c}$$

With these expressions, the dynamics of mean trait and additive genetic variance simplify to

$$d\bar{x}_i = G_i\left(a_i(\theta_i - \bar{x}_i) + \sum_{j=1}^{S}\beta_{ij}\right)dt + \sqrt{V_i\frac{G_i}{N_i}}dW_2, \tag{1.53a}$$

$$dG_i = \left\{G_i{}^2\left(-a_i + \sum_{j=1}^{S}\gamma_{ij}\right) + \mu_i - V_i\frac{G_i}{N_i}\right\}dt + G_i\sqrt{\frac{2V_i}{N_i}}dW_3. \tag{1.53b}$$

HIGH RICHNESS APPROXIMATION We now make use of the expressions derived for competition coefficients and selection gradients to investigate their relationship. As a first pass, we assume the niche-breadths w_i and intraspecific variances σ_i^2 are equivalent across species so that the sensitivity parameters $b_{ij} = 1/(w_i + w_j + \sigma_i^2 + \sigma_j^2) = b$ are constant across interacting pairs of species. We also assume abundances N_i, niche-use parameters U_i, strengths of competition c_i and mean traits $\bar{x}_i$ are distributed independently of each other with respective means and variances denoted by $\bar{N}$, V_N, $\bar{U}$, V_U, $\bar{c}$, V_c, $\bar{\bar{x}}$, $V_{\bar{x}}$. We further assume that richness S is large and the distribution of mean trait values is approximately normal.

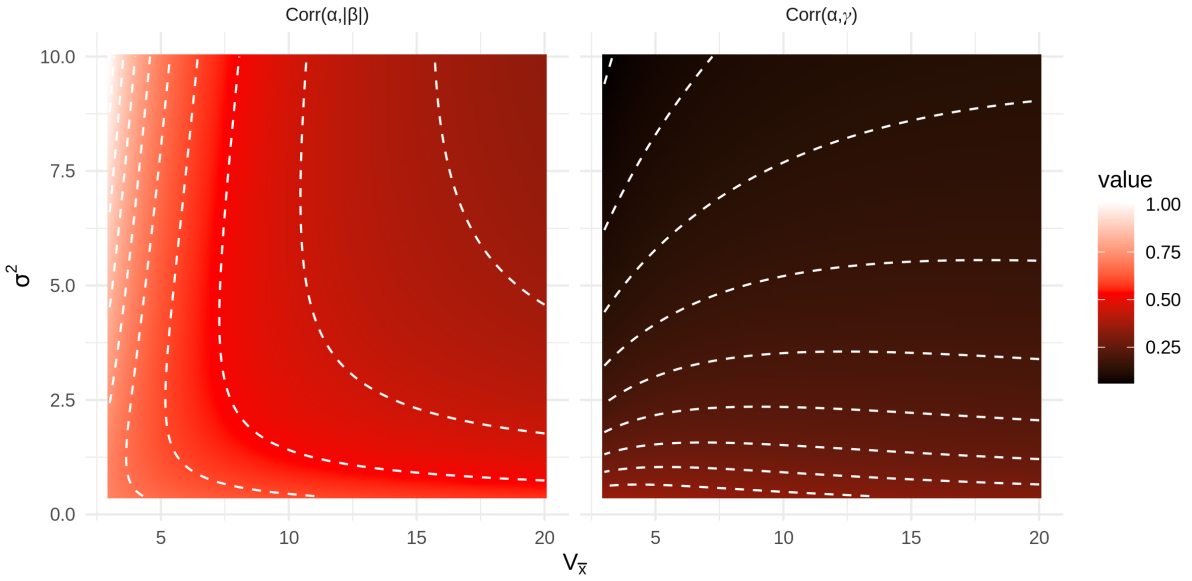

Figure 1.3: Heatmaps of the correlation between the magnitude of linear selection gradients and competition coefficients (left) and between stabilizing selection gradients and competition coefficients (right) as functions of community-wide variance of mean trait values $V_{\bar{x}}$ and intraspecific trait variances σ^2. In both plots we set $w = 1.0$, $\bar{c} = 1.0 \times 10^{-7}$, $V_c = 0.0$, $\bar{U} = 1.0$, $V_U = 0.0$, $\bar{N} = 1.0 \times 10^5$, and $V_N = 100.0$.

Under these assumptions we obtained analytical approximations for the correlations between specific competition coefficients α_{ij} and selection gradients β_{ij}, γ_{ij}. These calculations are provided in Appendix A.9. In particular, we found linear selection gradients are not associated with competition coefficients ($\text{Corr}(\alpha, \beta) \approx 0$). However, we did find a non-trivial relationship between the magnitudes of linear selection gradients and competition coefficients ($\text{Corr}(\alpha, |\beta|) \neq 0$) and also between quadratic

selection gradients and competition coefficients ($\text{Corr}(\alpha, \gamma) \neq 0$). Their expressions can be found in Appendix A.9.

To understand if associations between competition coefficients and selection gradients tend to be positive or negative, we visualized these relationships in Figure 1.3. We fixed $w, \bar{c}, V_c, \bar{U}, V_U, \bar{N}$ and V_N and allowed the amounts of intraspecific trait variance σ^2 and interspecific trait variance $V_{\bar{x}}$ to vary. We found, for biologically realistic areas of parameter space, absolute values of linear selection gradients and quadratic selection gradients tend to be positively associated with competition coefficients. Hence, if we know of competing species that strongly effect each others abundances then we can guess they also impose directional and diversifying selection on one another. However, based on this information alone, we cannot guess at the direction of selection.

1.4 Conclusion

We have introduced a novel approach to derive eco-evolutionary models using the calculus of white noise and diffusion limits of measure-valued branching processes (MVBP) and coined SAGA, a SPDE model of phenotypic evolution that accounts for demographic stochasticity. From SAGA we derived SDE that track the dynamics of abundance, mean trait and additive genetic variance. Observing the expressions of these SDE, we find the effects of demographic stochasticity on the evolution of mean trait and additive genetic variance characterize the effects of random genetic drift. Although Lande (1976) has previously characterized the effects of random genetic drift on mean trait evolution in quantitative genetic models, the approach taken assumed constant effective population size and discrete non-overlapping generations. In contrast, our approach shows random genetic drift is a result of demographic stochasticity for continuously reproducing populations with fluctuating abundances.

To illustrate the relevance of our approach to studies of evolutionary ecology, we combined our SDE with classical competition theory to derive a model of diffuse coevolution. We then used this model to investigate the relationship between standardized selection gradients and competition coefficients. We found absolute values of linear selection gradients and raw values of quadratic selection gradients are positively related with competition coefficients. In the process, we derived expressions for competition coefficients and components of selection gradients due to pairwise interactions as functions of niche-use parameters (niche breadth, total use and mean and variance of niche location), strength of competitive interactions and abundance.

Although the framework outlined here holds great potential for developing a synthetic theory of coevolving ecological communities, there are two technical gaps in the mathematical foundations of our approach. Firstly, we did not provide any formal conditions under which trait means and variances remain finite for finite time. However, a result due to Evans and Perkins (1994) shows that the diffusion limit for a pair of interacting MVBP following our simple niche-based treatment of competition exist when growth rates, as functions of trait values and abundances, are bounded above. This result can be easily extended to finite sets of competing species and therefore formally establishes the existence of abundances as diffusion processes. Further work is needed to determine the conditions under which trait means and variances exist as diffusion processes. The models studied

here provide likely sufficient conditions. In particular, since diffusive mutation does not lead to "heavy-tailed" phenotypic distributions, we expect the mean trait and trait variance to remain finite so long as total abundance is positive, given finite initial values for trait mean and variance. That is, since we have not included any processes that would cause blow-up either in mean trait or trait variance, we expect solutions of the SDE (1.28) to exist for all finite time t such that $N(t) > 0$ when $|\bar{x}(0)|, \sigma^2(0) < +\infty$. This assumption appears especially well-founded under quadratic stabilizing selection. Since fitness indefinitely decreases as individual trait value becomes indefinitely large (see equation (1.44)), the diversifying effects of mutation and competition will eventually be overwhelmed by stabilizing selection. Hence quadratic stabilizing selection prevents the abundance densities of populations from venturing indefinitely far from their phenotypic optima.

Secondly, although SDE derived under the assumption of normally distributed phenotypes provide particularly useful formula by replacing covariances between phenotype and fitness with fitness gradients, this assumption is mathematically rigorous only under deterministic dynamics and when the growth rate is a linear or concave-down quadratic function of trait value. However, following our derivation based on classical competition theory, we found the associated growth rate is highly non-linear. While this extreme non-linearity is mathematically inconvenient, it also captures important biological details and thus allows for a more realistic model of community dynamics. In spite of this inconsistency in our model formulation, we found resulting dynamics under the assumption of normally distributed trait values retained well-founded biological intuition. Furthermore, previous work in the field of theoretical quantitative genetics has demonstrated the assumption of normally distributed trait values is robust to fitness functions that select for non-normal trait distributions when inheritance is given a more realistic treatment and when populations reproduce sexually (Barton et al., 2017; Turelli and Barton, 1994). Hence, future work is needed to extend our approach to account for sexual reproduction, more realistic models of inheritance and to investigate the community-level consequences of non-normally distributed trait values.

Overall, this work demonstrates that connecting contemporary theoretical approaches of evolutionary ecology with some fundamental results in the theory of measure-valued branching processes and their diffusion limits allows for the development of a rigorous, yet flexible approach to synthesizing the dynamics of abundance and distribution of quantitative characters. In particular, equations (1.28a) and (1.39) provide a fundamental set of equations for deriving stochastic eco-evolutionary models involving quantitative traits. However, these equations require an expression for growth rates associated with each trait value. Conveniently, equation (1.20) in Appendix §1.2.2 provides a means to derive such growth rates from individual based models. Taken together, these results provide a means to derive analytically tractable dynamics from mechanistic formulations of fitness as a function of phenotype. The derivation of our model of diffuse coevolution, located in Appendix §A.7, demonstrates how to derive eco-evolutionary models involving a set of interacting species from biological first principles. Hence, this work provides a novel set of mathematical tools and a tutorial for their use in theoretical studies of evolutionary ecology and therefore paves the way for future work that provides a holistic theoretical treatment of coevolving ecological communities.

Chapter 2: Coevolutionary Arms Races and the Conditions for the Maintenance of Mutualisms

Abstract

Empirical evidence suggests coevolutionary arms races between flowering plants and their pollinators commonly occur in wild populations. In extreme cases, trait escalation can result in evolutionary switching from mutualism to parasitism. However, theoretical approaches to study coevolution typically assume fixed types of ecological interactions and ignore the evolution of absolute fitness. Here we introduce a novel approach to track the evolution of absolute fitness as a framework to determine when escalatory coevolution results in a switch from mutualism to parasitism. We apply our approach to two mechanisms mediating selection as a function of phenotype. Our results demonstrate interactions mediated by a "bigger-is-better" mechanism evolve towards parasitism. In contrast, generalizing the classical trait-matching mechanism so that fitness of each species is optimized when trait values mismatch by a particular amount, we find theoretical support for indefinite trait exaggeration that preserves mutualistic interactions. Building on our results, we discuss a path towards the development of statistical methods to project when mutualisms are at risk of evolutionary disintegration. Moving beyond pairwise interactions, we consider the ramifications of coevolution in a South African pollination network for the evolution of parasitism. Future work extending our approach beyond pairwise interactions can lead to a framework for understanding the evolution of parasitism in mutualistic networks and further insights into the structure and dynamic nature of ecological communities in general.

2.1 Introduction

Coevolution between pairs of species has long been considered an important driver of phenotypic exaggeration (Anderson and Johnson, 2007, 2009; Anderson et al., 2010; Benkman et al., 2003; Brodie et al., 2005; Darwin, 1862; Muchhala and Thomson, 2010; Nuismer, 2017; Nuismer and Week, 2019; Pauw et al., 2009; Thompson, 2014; Toju and Sota, 2006; Wallace, 1867; Week and Nuismer, 2019). In particular, coevolutionary theory predicts antagonistic victim-exploiter interactions, such as host-parasite and predator-prey interactions, often result in coevolutionary arms races (Gavrilets, 1997; Gavrilets and Hastings, 1998; Nuismer et al., 2007). If unchecked by external sources of stabilizing selection, the interacting species are predicted to evolve ever greater defensive and offensive trait values. However, this dynamic is not unique to antagonistic interactions. Indeed, several cases of apparent trait escalation involving mutualistic interactions have been documented (Anderson and Johnson, 2007, 2009; Pauw et al., 2009). Furthermore, the dichotomy between mutualism and antagonism is, to some degree, artificial. For example, if one mutualistic partner outpaces the other in the evolutionary race, the carefully balanced mutual benefits may disintegrate leading to the evolutionary switching from mutualism to parasitism Jones et al. (2015); Pauw et al. (2009). Here we study this evolutionary disintegration of mutualisms under two mechanisms mediating fitness as a function of phenotype. Our goal is to determine whether these interaction mechanisms preserve or destroy mutualistic interactions.

Although coevolutionary models of victim-exploiter interactions often predict indefinite trait esca-

lation, most coevolutionary models of mutualistic interactions rarely, if ever, predict coevolutionary arms races and sustained escalation of traits (Nuismer, 2017). Instead, coevolution driven by a mutualism is often modelled with a trait-matching mechanism that promotes evolutionary convergence of trait values (Kiester et al., 1984; Nuismer, 2017). While this dichotomy between antagonistic and mutualistic interactions is generally accepted, striking counter-examples exist in a range of empirically well-studied mutualistic interactions characterized by extremely exaggerated traits (Anderson and Johnson, 2007, 2009; Muchhala and Thomson, 2010; Pauw et al., 2009). For instance, in the interaction between the long-proboscid fly *Prosoeca ganglbaueri* and the flowering plant *Zaluzianskya microsiphon*, Anderson and Johnson (2007) found extreme trait exaggeration including proboscis lengths of up to 50mm and floral tube depths of up to 55mm. Furthermore, these authors discovered significant spatial correlations in mean proboscis length and floral tube depth, suggesting a role for coevolution in explaining patterns of phenotypic exaggeration. Similarly, Pauw et al. (2009) found significant spatial correlations between another long-proboscid fly, *Moegistorhynchus longirostris*, and a long-tubed flower *Lapeirousia anceps*. Using a clever experimental design, Pauw et al. (2009) also demonstrated evidence for reciprocal directional selection caused by this interaction for longer floral tubes and pollinator proboscises.

The biological reasoning behind mutualsitic trait exaggeration dates back at least to Darwin (1862) and Wallace (1867). In particular, it was reasoned that pollinators benefit from having slightly longer mouth-parts than nectar-tube depths in order to retrieve their nectar rewards, thus inducing selection for longer mouth-parts. Simultaneously, flowers benefit from having nectar-tubes slightly longer than pollinator mouth-parts to ensure pollen transfer, thus inducing selection for deeper nectar-tubes. This "bigger-is-better" mechanism of selection has been captured mathematically by the so-called trait-differences mechanism Nuismer (2017); Nuismer et al. (2007). In particular, given ample genetic variation and lack of selective forces that oppose trait elongation, the verbal arguments of Darwin and Wallace agree with mathematical analysis of the trait-differences mechanism in predicting a coevolutionary arms race for ever longer mouth-parts and ever deeper nectar-tubes. In fact, it was this line of reasoning that lead Darwin to famously predict the existence of a large moth pollinator while examining the exaggerated nectar spur of the orchid *Angraecum sesquipedale* (Darwin, 1862; Wasserthal, 1997).

The above examples suggest coevolutionary trait exaggeration commonly occurs in plant-pollinator mutualisms. However, if trait escalation of one partner outpaces that of the other, the interaction may cease to be beneficial for the second partner. That is, the mutualism may collapse into a parasitism (Jones et al., 2015). For instance, if the fly mouth-part evolves to exceed the nectar-tube depth by an extreme amount, the fly may be better described as a nectar thief rather than a pollinator (Inouye, 1980; Pauw et al., 2009). This motivates the following two questions;

1. Can sustained arms races within mutualistic interactions continue indefinitely without disintegrating into antagonisms?

2. If so, which mechanisms mediating fitness as a function of phenotype favor the stability of mutualism in the face of ongoing arms races for ever increasing trait values?

Here we answer these questions by analyzing two simple models of mutualistic arms races. In one

model, the arms race is driven by a "bigger-is-better" (ie., trait-differences) mechanism where fitness increases indefinitely with trait size. In the other model, the arms race is driven by an "offset-matching" mechanism where fitness is a unimodal function peaking at some value larger than the trait of the interacting partner. By tracking the absolute fitness of interacting species, our results demonstrate unchecked mutualistic arms races driven by a weak selection approximation of trait-differences are doomed to a parasitic fate. In contrast, offset-matching (with weak or strong selection) can preserve mutualistic interactions in spite of ever increasing trait values.

2.2 Methods

Our analysis focuses on two interaction mechanisms determining fitness of individuals as functions of their phenotype and the phenotype of an encountered individual; the trait-differences mechanism (Nuismer et al., 2007) and the offset-matching mechanism (Week and Nuismer, 2019). We therefore begin the description of our methods with a brief review of trait-differences and offset-matching. We then describe our approach to modelling coevolutionary dynamics and present population growth rates associated with each interaction mechanism. We show these growth rates can be additively decomposed into a component representing the overall effect on fitness due to the interaction and a component representing effects due to sources outside of the interaction. By tracking the component describing the overall effect of the interaction on absolute fitness we develop a method to model the evolutionary switching of interaction types and the transition from mutualism to parasitism in particular. In the results we combine our models of coevolutionary dynamics with our approach to model the evolutionary switching of interaction types to determine when trait-differences and offset-matching promote an evolutionary switch from mutualism to antagonism.

2.2.1 Trait-differences

The trait-differences mechanism assumes the component of fitness due to a biotic interaction changes monotonically with the trait value of the focal individual. Nuismer et al. (2007) used trait-differences to derive quantitative genetic models of coevolutionary trait escalation driven by antagonistic interactions. The coevolutionary behavior implied by the trait-differences mechanism exhibits two characteristic features. First, the mean trait will diverge towards ever larger values. Second, in the limit of weak selection, the rate of trait evolution for one species does not depend on the trait value of the other leading to independent evolutionary trajectories.

The trait differences mechanism has frequently been modeled using a logistic curve describing the probability of a successful interaction given the trait values of the interacting individuals (Nuismer, 2017; Nuismer and Week, 2019; Nuismer et al., 2007; Toju and Sota, 2006). Under this model, the probability of successful interaction grows asymptotically towards one as this difference in traits increases towards positive infinity and decreases towards zero as this difference decreases towards negative infinity. Theoretical studies of coevolutionary trait-escalation are traditionally based on a logistic curve fitness function and make use of weak selection approximations to obtain analytically tractable models. However, the same "bigger-is-better" mechanism can also be modelled with an exponential curve. Al-

though not as biologically realistic, the latter option simplifies analysis and yields the same dynamics as the logistic curve in the limit of weak selection (Nuismer et al., 2007; Week and Nuismer, 2019, also see Appendix B). We therefore make use of the exponential curve instead of the logistic curve in our derivations. In particular, denoting x the trait value of an individual of species X, y the trait value of an individual of species Y, and assuming each individual engages in a single interaction, individual fitness can be captured by

$$W_X(x,y) \propto e^{B_X(x-y)}, \tag{2.1a}$$

$$W_Y(y,x) \propto e^{B_Y(y-x)}, \tag{2.1b}$$

where B_X and B_Y determine the sensitivity of fitness to trait values and hence mediate the strengths of selection on the respective species. We therefore refer to these parameters as the strengths of biotic selection. This interaction mechanism is summarized graphically in Figure 2.1.

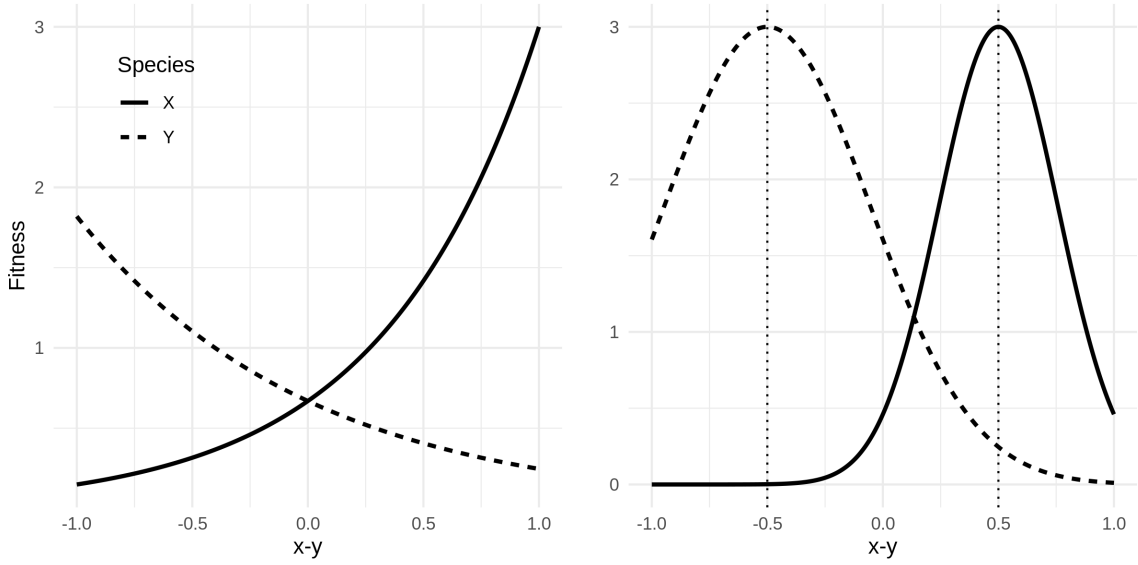

Figure 2.1: Fitness curves under trait-differences (left) and offset-matching (right) as functions of the difference in individual traits $x - y$. Solid lines represent the fitness of individuals in species X. Dashed lines represent the fitness of individuals in species Y. Vertical dotted lines mark the location of offset fitness optima. In the case shown we have set $\delta = 0.5$. In both plots we have set $B_Y < B_X$ so that species X experiences stronger selection than species Y.

2.2.2 Offset-matching

The offset-matching mechanism generalizes the trait-matching mechanism frequently used to model coevolutionary interactions (Kiester et al., 1984; Week and Nuismer, 2019). In particular, offset-matching relaxes the key assumption that trait values from interacting pairs must be perfectly matched to optimize fitness for both species. Instead, fitness for each species is optimized when their trait value differs from the other by a particular amount, which we refer to as the optimal offset. By assuming fitness

is maximized at a positive optimal offset for both species, we obtain a mechanism similar to that described by Darwin (1862) and Wallace (1867) reviewed in the introduction. That is, each species benefits from having a slightly larger trait, but suffers reduced fitness beyond a certain threshold. This model of fitness was introduced by Week and Nuismer (2019) to simultaneously explain trait exaggeration and correlations of mean traits across multiple populations, enabling the use of a simple maximum likelihood approach to measure the strength of coevolution in the wild. By setting the optimal offset to zero, the trait-matching model generally used to model mutualistic coevolution is recovered. Figure 2.1 compares fitness curves under trait-differences and offset-matching.

Following the notation introduced above, we denote by x the trait value of an individual in species X and y a trait value in species Y. Assuming each individual participates in a single interaction, individual fitness under the offset-matching model can be described by

$$W_X(x,y) \propto e^{-\frac{B_X}{2}(x+\delta-y)^2}, \tag{2.2a}$$

$$W_Y(y,x) \propto e^{-\frac{B_Y}{2}(y+\delta-x)^2}, \tag{2.2b}$$

where δ is the optimal offset and B_X, B_Y again determine the strength of biotic selection on the respective species. A formal connection with the trait-differences model can be established using a simultaneous weak selection and large optimal offset approximation. In particular, substituting $B_X, B_Y := \varepsilon^3$ and $\delta := 1/\varepsilon$, the trait-differences mechanism can be obtained using a second order Taylor expansion around $\varepsilon \approx 0$. We illustrate this calculation in Appendix B.

2.2.3 Modelling approach

Beginning with the above models of individual fitness, we follow Chapter 1 and formally derive the continuous-time growth rates for interacting populations using diffusion limits (see Appendix B). In particular, Chapter 1 introduced a framework to rigorously derive models of evolutionary and ecological dynamics from structured branching processes. These branching processes are used as individual-based models of populations evolving in phenotypic state space. By rescaling time, reproductive output and "mass" of individuals and taking a large population size limit, the derivations in Chapter 1 provide deterministic and stochastic models that extend classical quantitative genetic models to account for abundance dynamics and demographic stochasticity in continuous time. Specifically, the Stochastic Asexual Gaussian allelic model with Abundance dynamics (abbreviated SAGA), introduced in Chapter 1, provides a set of stochastic differential equations that determines the dynamics of mean trait, additive genetic variance and population abundance in response to mutation, demographic stochasticity, random genetic drift and selection. Although these equations can accommodate a wide array of phenotypic distributions, they express phenotypic responses to selection in terms of covariances between growth rates and trait values which are generally difficult to calculate. We therefore assume normally distributed phenotypes to formally replace these covariances with gradients of growth rates which are typically much easier to calculate. Although our focus here is neither on stochastic nor abundance dynamics, this framework provides a means to track absolute fitness through time. In the following paragraph, we summarize our approach to diffusion limits and a series of assumptions used

to isolate the effect of interaction mechanism and mean trait evolution on the evolutionary switching of ecological relationships.

To focus our models of coevolutionary dynamics on trait evolution and the effects of interaction mechanisms, we assume each individual of species X interacts with a single individual of species Y and each individual of species Y interacts with a single individual of species X. Under this condition, selection for species X does not depend on the density of species Y and vice versa, allowing us to control for the effects of eco-evolutionary feedbacks. We provide more detail on this assumption in Appendix B where we develop our approach to diffusion limits. Our diffusion limits lead to a set of stochastic differential equations tracking the dynamics of abundance, mean trait and trait variance. To control for the effects of demographic stochasticity and random genetic drift, we assume infinitely large abundances. These assumptions reduce each of our models to a set of ordinary differential equations describing the evolution of trait means, phenotypic and additive genetic variances. We further simplify our models by assuming conditions that fix phenotypic and additive genetic variances. In particular, our model of trait-differences implies phenotypic and additive genetic variances are fixed when mutation is absent. Hence, for the trait-differences model we set mutation rates to zero. In contrast, our model of offset-matching implies these variances evolve to unique positive stable equilibria given by mutation-selection balance. Hence, for the offset-matching model, we initiate phenotypic and additive genetic variance to these stable equilibria. Details are provided in Appendix B. The final models obtained each consist of a pair of ordinary differential equations tracking the coevolution of mean traits.

Aside from our simplifying assumptions of fixed phenotypic and additive genetic variances and infinite population abundances, we further assume the lack of external selective pressures such as abiotic stabilizing selection. Although this allows for indefinite trait escalation, which is clearly not biologically feasible, our goal is to focus on evolutionary and ecological outcomes driven solely by the ecological interaction. Future models using our framework could extend our investigation by confronting the combined effects of various selective agents, non-equilibrium additive genetic variances, eco-evolutionary feedbacks and demographic stochasticity, but such ambitions remain outside the scope of our inquiry.

To formalize our models mathematically, we denote by $\bar{m}_X$ and $\bar{m}_Y$ the continuous-time growth rates for species X and Y respectively. We assume trait values x and y are normally distributed with means $\bar{x}, \bar{y}$ and variances σ_X^2, σ_Y^2. We denote by G_X, G_Y the additive genetic variances, r_X, r_Y the intrinsic growth rates and e_X, e_Y the intrinsic effects of interactions on growth rates (ie., the effects on growth rates due to the interaction when $B_X = 0$ or $B_Y = 0$) for species X and Y respectively. This notation is summarized in Table 2.1.

Using $\mathcal{D}$ to denote results under trait-differences and $\mathcal{O}$ for offset-matching, we find the following pairs of growth rates;

$$\mathcal{D} \begin{cases} \bar{m}_X = r_X + e_X + B_X(\bar{x} - \bar{y}), \\ \bar{m}_Y = r_Y + e_Y + B_Y(\bar{y} - \bar{x}), \end{cases} \tag{2.3a}$$

Table 2.1: Summary of notation.

Parameter	Description
x, y	Individual trait values
$\bar{x}, \bar{y}$	Mean traits
σ_X^2, σ_Y^2	Traits variances
G_X, G_Y	Additive genetic variances
$\bar{m}_X, \bar{m}_Y$	Population growth rates
r_X, r_Y	Intrinsic growth rates
I_X, I_Y	Overall interaction effects
e_X, e_Y	Intrinsic interaction effects
B_X, B_Y	Strengths of biotic selection
δ	Optimal offset

$$\mathcal{O}\begin{cases}\bar{m}_X = r_X + e_X - \frac{B_X}{2}(\bar{y}+\delta-\bar{x})^2 - \frac{B_X}{2}(\sigma_X^2+\sigma_Y^2),\\ \bar{m}_Y = r_Y + e_Y - \frac{B_Y}{2}(\bar{x}+\delta-\bar{y})^2 - \frac{B_Y}{2}(\sigma_X^2+\sigma_Y^2).\end{cases} \tag{2.3b}$$

In Appendix B we show these growth rates do not exhibit frequency dependent selection. Hence, following our approach to derive deterministic dynamics, we calculate the evolution of mean traits via

$$\frac{d\bar{x}}{dt} = G_X \frac{\partial \bar{m}_X}{\partial \bar{x}}, \tag{2.4a}$$

$$\frac{d\bar{y}}{dt} = G_Y \frac{\partial \bar{m}_Y}{\partial \bar{y}}. \tag{2.4b}$$

This yields the following two sets of mean trait dynamics;

$$\mathcal{D}\begin{cases}\frac{d\bar{x}}{dt} = G_X B_X,\\ \frac{d\bar{y}}{dt} = G_Y B_Y,\end{cases} \tag{2.5a}$$

$$\mathcal{O}\begin{cases}\frac{d\bar{x}}{dt} = G_X B_X(\bar{y}+\delta-\bar{x}),\\ \frac{d\bar{y}}{dt} = G_Y B_Y(\bar{x}+\delta-\bar{y}).\end{cases} \tag{2.5b}$$

Under our assumption of fixed phenotypic variances, the population growth rates presented in equations (2.3) change only through mean trait evolution. In particular, we find growth rates for each species can be additively partitioned as $\bar{m}_X = r_X + I_X$ and $\bar{m}_Y = r_Y + I_Y$, where I_X, I_Y represent the components due to the interspecific interaction and the intrinsic growth rates r_X, r_Y represent the components due to everything else. Hence, I_X and I_Y determine the type of interaction between species X and Y. Specifically, if $I_X, I_Y > 0$, the interaction is a mutualism and if either $I_X < 0$ or $I_Y < 0$, the interaction is an antagonism. Thus, by tracking the signs of I_X and I_Y as mean traits evolve, we can track the evolutionary switching of ecological interaction types. In particular, if an interaction mechanism promotes the evolutionary switching from a mutualism to a parasitism, then I_X, I_Y begin with positive values and eventually one of either I_X or I_Y becomes negative due to the evolution of $\bar{x}$ and $\bar{y}$. Following equations (2.3), trait-differences and offset-matching respectively yield

$$\mathcal{D}\begin{cases} I_X = e_X + B_X(\bar{x} - \bar{y}), \\ I_Y = e_Y + B_Y(\bar{x} - \bar{y}), \end{cases} \tag{2.6a}$$

$$\mathcal{O}\begin{cases} I_X = e_X - \frac{B_X}{2}\left[(\bar{y} + \delta - \bar{x})^2 + \sigma_X^2 + \sigma_Y^2\right], \\ I_Y = e_Y - \frac{B_Y}{2}\left[(\bar{x} + \delta - \bar{y})^2 + \sigma_X^2 + \sigma_Y^2\right]. \end{cases} \tag{2.6b}$$

In the following section we combine the coevolutionary models presented above with this criteria to determine when an interaction mechanism promotes the evolutionary switching from mutualism to parasitism.

2.3 Results

2.3.1 Trait-differences promotes switching to parasitism

We begin by exploring the conditions under which sustained trait exaggeration can occur under the model of trait-differences without the interaction dissolving into antagonism. Specifically, because the trait differences model yields sustained trait exaggeration under positive selection strengths (ie., $B_X, B_Y > 0$), we study how the ecological nature of the interaction evolves under this regime. Combining equations (2.5a) and (2.6a), we find the overall effects of the interaction on species growth rates evolve via

$$\frac{dI_X}{dt} = B_X(G_X B_X - G_Y B_Y), \tag{2.7a}$$

$$\frac{dI_Y}{dt} = B_Y(G_Y B_Y - G_X B_X). \tag{2.7b}$$

Hence, unless the products of additive genetic variance and strength of biotic selection is perfectly balanced between the two species so that $G_X B_X = G_Y B_Y$, the effects on species growth rates will indefinitely evolve in opposite directions. This symmetry condition is unlikely to hold in nature implying one of the species will evolve to become a parasite of the other. Thus, no matter how large the intrinsic effects e_X and e_Y are, mutualisms mediated by a trait-differences mechanism tend towards parasitism. The left panel of Figure 2.2 displays this steady disintegration of mutualism by tracking the evolution of overall interaction effects I_X, I_Y. For the particular parameters chosen, species Y becomes the host and species X becomes the parasite. The right panel of Figure 2.2 displays the rate of transition from mutualism to parasitism as a function of $G_X B_X - G_Y B_Y$. When $G_X B_X < G_Y B_Y$, species Y becomes the parasite and when $G_X B_X > G_Y B_Y$, species X becomes the parasite.

There are two important caveats to this result: 1) the lack of stabilizing selection and 2) the assumed independence of ecological and evolutionary dynamics. Both may have important consequences for our predictions which we work through in detail in the discussion.

2.3.2 Offset-matching stabilizes mutualistic interactions

Following the logic of Darwin (1862) and Wallace (1867), we assume fitness for each individual is increased by having a trait value that exceeds that of its partner. However, we also assume a limit

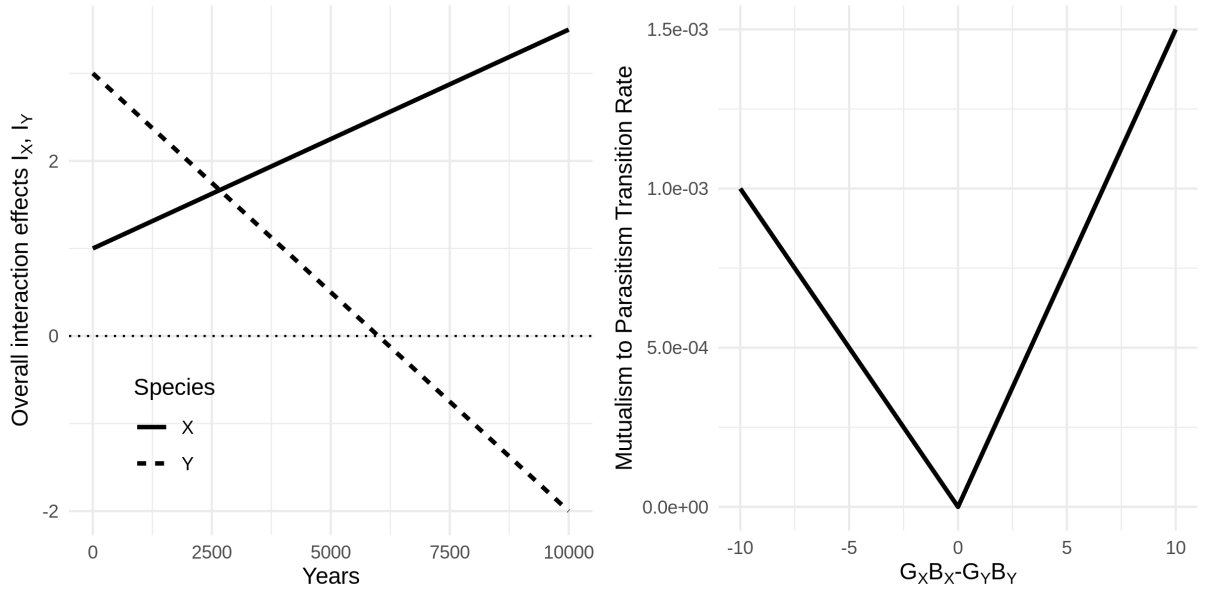

Figure 2.2: (Left) A time-series plot illustrating the trajectories of overall interaction effects I_X, I_Y. The solid line represents the overall interaction effect on species X and the dashed line represents the overall interaction effect on species Y. The horizontal dotted line marks the threshold determining interaction type. When I_X or I_Y are below this line, the interaction is an antagonism. Hence, for the particular parameters chosen, this figure shows species X evolves to become a parasite of species Y. (Right) Rate of transition from parasitism to mutualism as a function of $G_XB_X - G_YB_Y$ under the assumption $B_X < B_Y$. On the left side of the plane, species X becomes a parasite of species Y. On the right side of the plane this relationship is reversed.

exists, such that fitness is maximized when an individual's trait value exceeds that of its partner by some specific amount. Hence, we assu me a positive optimal offset $\delta > 0$ and positive selection strengths $B_X, B_Y > 0$. Under these assumptions equations (2.5b) yields ever increasing trait values. Combining equations (2.5b) and (2.6b), we find the effects on species growth rates evolve via

$$\frac{dI_X}{dt} = -B_X(\delta - \Delta)\frac{d\Delta}{dt}, \tag{2.8a}$$

$$\frac{dI_Y}{dt} = -B_Y(\delta + \Delta)\frac{d\Delta}{dt}, \tag{2.8b}$$

where $\Delta = \bar{x} - \bar{y}$. In turn, the difference in mean traits Δ evolves via

$$\frac{d\Delta}{dt} = (G_XB_X + G_YB_Y)\left(\frac{G_XB_X - G_YB_Y}{G_XB_X + G_YB_Y}\delta - \Delta\right). \tag{2.9}$$

Under our assumption of positive selection strengths $B_X, B_Y > 0$, consistent with escalatory trait evolution, equation (2.9) implies that, although the mean traits $\bar{x}, \bar{y}$ evolve indefinitely greater values, their difference Δ will always evolve to the stable equilibrium

$$\hat{\Delta} = \frac{G_XB_X - G_YB_Y}{G_XB_X + G_YB_Y}\delta. \tag{2.10}$$

In spite of ever escalating mean trait values $\bar{x}, \bar{y}$, when Δ is at its equilibrium, the components of species growth rates I_X, I_Y have also reached a stable equilibrium given by

$$\hat{I}_X = e_X - \frac{B_X}{2}\left[\left(\frac{2G_Y B_Y \delta}{G_X B_X + G_Y B_Y}\right)^2 + (\sigma_X^2 + \sigma_Y^2)\right], \tag{2.11a}$$

$$\hat{I}_Y = e_Y - \frac{B_Y}{2}\left[\left(\frac{2G_X B_X \delta}{G_X B_X + G_Y B_Y}\right)^2 + (\sigma_X^2 + \sigma_Y^2)\right]. \tag{2.11b}$$

Hence, trait escalation under the offset-matching model promotes the stabilization of interaction effects in the absence of all other evolutionary forces. This means if the intrinsic effects e_X and e_Y are large enough, the interaction will remain a mutualism even as trait values escalate indefinitely toward larger values. Furthermore, this result implies novel interactions that begin as a parasitism may evolve towards mutualism when mediated by an offset-matching mechanism. Figure 2.3 illustrates these results by tracking the evolution of overall interaction effects I_X, I_Y for two scenarios.

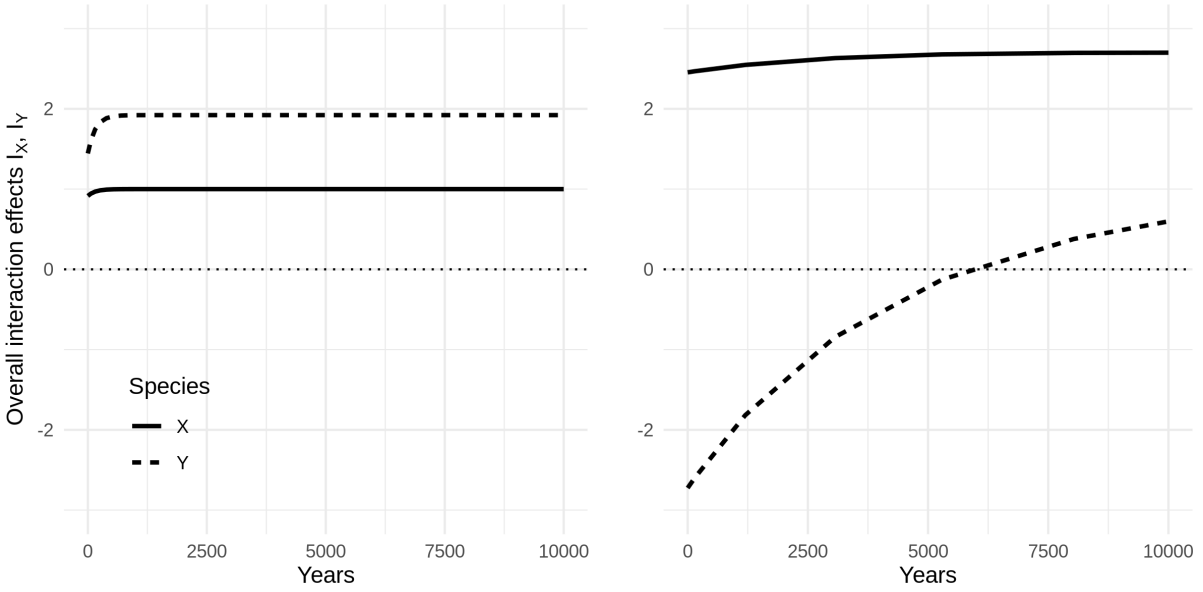

Figure 2.3: Time-series plots of overall interaction effects I_X, I_Y under two scenarios. Solid lines represent the overall interaction effect on species X and dashed lines represent the overall interaction effect on species Y. The horizontal dotted lines mark the threshold determining interaction type. When I_X or I_Y are below this line, the interaction is an antagonism. The left plot illustrates a mutualism that is preserved in spite of indefinitely escalating trait values. The right plot illustrates a novel parasitism that experiences an evolutionary switch to mutualism.

Inspection of equations (2.11) reveal the minimal intrinsic benefits e_X, e_Y needed to preserve mutualism increase logistically with the selection strength of the partner species. For example, the minimal value of e_X required to maintain benefits for species X does not increase indefinitely with B_Y. In contrast, we see these minimal quantities increase indefinitely with the selection strengths of their focal species, the optimal offset δ, and phenotypic variances σ_X^2, σ_Y^2. Since additive genetic variance com-

prises the component of expressed trait variance explained by additive allelic effects, we have $G_X \leq \sigma_X^2$ and $G_Y \leq \sigma_Y^2$. Hence, the minimal intrinsic benefits required to maintain mutualisms also increase indefinitely with additive genetic variances. To summarize, mutualisms mediated by an offset-matching mechanism are more likely will dissolve into parasitisms when fitness is maximized for large offsets, when either species maintains a sufficiently large trait variance it induces a phenotypic load, or when either species experiences strong selection.

2.4 Discussion

We have shown that mutualistic interactions can be preserved in the face of indefinite coevolutionary trait escalation. However, this result depends on the mechanism mediating interspecific interactions. In particular, we found interactions mediated by a weak selection approximation of the trait-differences mechanism, which has been traditionally employed to model trait escalation in coevolving antagonisms (Nuismer et al., 2007; Toju and Sota, 2006), encourages the dissolution of mutualism into parasitism. In contrast, employing the recently introduced offset-matching mechanism (Week and Nuismer, 2019), which generalizes the classical trait-matching mechanism (Kiester et al., 1984), we found mutualistic interactions are preserved when the intrinsic benefits of the interaction are sufficiently large (eqn. 2.11).

Our analyses are particularly relevant for the study of pairwise plant-pollinator interactions. Indeed, the primary motivation of this work is the hypothesis that exaggerated floral tubes and pollinator proboscises observed in the wild are explained by pairwise coevolutionary races. Plant-pollinator interactions provide classical examples of mutualisms in which the plant provides a nutritional or metabolic resource to the pollinator in trade for movement of pollen. Under the trait-differences mechanism, our results imply the difference between average proboscis length and average floral tube depth will increase indefinitely, leading to one of two outcomes; 1) Floral tube depth will eventually exceed pollinator proboscis length such that that the pollinator is essentially tricked into transferring pollen without access to any reward. 2) Pollinator proboscis length will eventually exceed floral tube depth such that the pollinator drains the flower of nectar without transferring pollen. In both of these cases one species is eventually exploited while the other continues to profit from the interaction. Quantitatively, this implies the overall interaction effect on absolute fitness becomes negative for one of the species and hence results in a host-parasite relationship.

Alternatively, results based on the offset-matching mechanism imply this parasitic fate is not universal among mutualistic pairs engaged in a coevolutionary arms race. Returning to the plant-pollinator example above, the offset-matching mechanism implies the difference in average proboscis length and average floral tube depth will converge to a stable equilibrium proportional to the optimal offset, even though these mean traits will themselves continue to escalate indefinitely. In this case there are three conditions required to maintain the mutualism; 1) The optimal offset must be sufficiently small, 2) biotic selection cannot be too strong and 3) phenotypic variances cannot be too large. If all three of these conditions are satisfied then the phenotypic distribution of each species remains in a range that benefits the other species, even though these ranges are dynamic.

Although these analyses shed light on the outcomes of pairwise interactions, plant-pollinator interactions based on a trade between pollen movement and food tend to be generalized (Vázquez and Aizen, 2004; Waser et al., 1996). To understand the dynamics of evolutionary interaction switching in the community context, the analytical methods developed here can be extended to models of coevolving mutualistic networks that account for multispecific interactions (eg., Medeiros et al., 2018). In the case that a subset of the community engages in a multispecific coevolutionary arms race, the species left behind are at risk of becoming parasitized. For example, in the fly-flower pollination system investigated by Pauw et al. (2009), the long-proboscid fly *M. longirostris* is known to visit at least 20 long-tubed flower species (Manning and Goldblatt, 1997). This plant guild includes *L. anceps*, which is the most abundant and widespread member (Pauw et al., 2009). The exaggerated nectar tube of *L. anceps* exhibits strong spatial correlations with the proboscis length of *M. longirostris*. By interfacing these patterns of trait exaggeration and spatial correlation with coevolutionary theory, Week and Nuismer (2019) were able to provide quantitative evidence for a coevolutionary arms race between the two species. However, in the course of this arms race it is likely *M. longirostris* imposed and received selection pressures from other members of the long-tubed plant guild, leading to a compartment of species engaged in a coevolutionary arms race (Pauw et al., 2009). As a consequence, mutualistic interactions between *M. longirostris* and flower species outside of this compartment are at risk of disintegrating into parasitisms.

In particular, *M. longirostris* has been observed to frequently visit flowers of *Babiana thunbergii*, which is usually pollinated by Malachite sunbirds and has a much shorter and wider floral morphology. Pauw et al. (2009) noted that *M. longirostris* rarely contacts the reproductive organs of *B. thunbergii* while draining its nectar and hence acts as a parasitic nectar thief. In the case that *B. thunbergii* originally profited from visits of *M. longirostris*, before the evolution of extreme proboscis lengths, external sources of selection, such as those due to sunbirds, may have countered selection for floral elongation. In this case, the disintegration of mutualism follows a different path from what we have described above. Instead of direct pairwise interactions leading to an intimate race for exaggerated trait values, indirect coevolutionary effects due to interactions with other community members impose inconsistent patterns of selection for each member of the focal pair. A recent theoretical result supports this view by suggesting such indirect effects are as important for shaping coevolutionary dynamics in ecological networks as direct effects (Guimarães et al., 2017).

Alternatively, the association between *B. thunbergii* and *M. longirostris* may be novel and thus have no coevolutionary history. If so, it is possible that future evolution induced by selection pressure from *M. longirostris* will cause the short and wide floral morphology of *B. thunbergii*, currently adapted for sunbirds, to become thinner and longer as it adapts to the fly. Thus, it is reasonable to ask whether the novel parasitism will evolve into a mutualism or whether patterns of selection induced by other community members will prevent *B. thunbergii* from building defenses to nectar robbers such as *M. longirostris*. Extending the analytical approach developed here to the context of mutualistic networks will undoubtedly shed light on this curious situation.

For the sake of analytical tractability and clarity of results, we have kept our analysis as simple as possible, while recognizing our assumptions are unlikely to be satisfied literally in real biological sys-

tems. For instance, ignoring stabilizing selection implies the absence of well-established physiological and anatomical constraints that prevent indefinite runaway coevolution in most systems. Similarly, we have ignored the possible depletion of heritable variation, the influence of abundance dynamics, and, in the case of trait-differences, considered only weak coevolution. Although these assumptions come at the cost of biological realism, they have allowed us to understand and illuminate the tendency of trait-differences and offset-matching to motivate or resist evolutionary transitions from mutualism to parasitism. Hence, our results should be interpreted in much the same way as experiments conducted in controlled, but biologically unrealistic environments. To predict and understand when wild populations of coevolving mutualists are doomed to an inevitable parasitic fate, however, we must confront these complications directly and with more complex and biologically realistic models.

Out of the large set of processes we have controlled for, perhaps the most important are eco-evolutionary feedbacks. Future models using our framework can incorporate the effects of eco- evolutionary feedbacks by replacing the assumption that each individual interacts with a single individual of their partner species with the assumption that individual interaction frequencies depend on abundances. As an example, assuming each individual of species X interacts once with each individual of species Y and vice versa yields intrinsic benefits e_X, e_Y and strengths of biotic selection B_X, B_Y that depend linearly on the abundance of the partner species. Hence, this assumption leads to generalizations of classical mass-action models of population dynamics, such as the Lotka-Volterra predator-prey model (Lotka, 1925; Volterra, 1926), that allow for the evolutionary switching of interaction type in tandem with the fluctuations of abundance. In this case, the overall interaction effects can be expressed as $I_X = \zeta_X N_Y$, $I_Y = \zeta_Y N_X$, where ζ_X, ζ_Y represent the interaction coefficients corresponding to species X and Y respectively. To formally remove demographic stochasticity while considering finite abundances, one can set the variance in reproductive output of individuals equal to zero (see Appendix B). This returns a set of ordinary differential equations modelling the dynamics of abundance, mean trait and additive genetic variance for each species. In principle, the same approach developed here to analyze evolutionary interaction switching can be directly extended to this more complicated scenario. However, without doing this analysis we can still draw some basic conclusions. For example, in the case that X evolves to parasitize species Y, the growth rate of species Y will diminish. If the intrinsic growth rate r_Y is not sufficiently large, this will result in decreased abundance of species Y which, in turn, results in decreased selection pressure on species X. In the extreme case, as the abundance of species Y falls towards zero, the absolute fitness of species X will be approximately equal to the intrinsic growth rate r_X. Hence, although intrinsic growth rates played no role in the conditions maintaining mutualisms based on our simple coevolutionary models, we anticipate intrinsic growth rates to play a fundamental role in these conditions when accounting for eco-evolutionary feedbacks.

Our results imply that empirical studies aimed at inferring or projecting evolutionary interaction switching will need to determine which interaction mechanisms are at play (eg., trait-differences, trait-matching or offset-matching) along with quantitative estimates of model parameters. For approaches building on models introduced here, these parameters include the strengths of biotic selection B_X, B_Y, additive genetic variances G_X, G_Y, intrinsic benefits e_X, e_Y and, in the case of offset-matching, phenotypic variances σ_X^2, σ_Y^2 and the optimal offset δ. Furthermore, these studies will need to confront

the effects of eco-evolutionary feedbacks and evolutionary processes occurring outside of the interaction, such as abiotic stabilizing selection, which will require the estimation of associated parameters. Although further development is needed to fully parameterize the models studied here, the maximum-likelihood and Bayesian methods of coevolutionary inference introduced by Week and Nuismer (2019) and Nuismer and Week (2019) provide a useful starting point. Extending these methods to disentangle absolute fitness into intrinsic growth rates r_X, r_Y, intrinsic benefits e_X, e_Y and the combined effects of phenotypic distributions and biotic selection would move us closer to the ultimate goal of forecasting the evolutionary stability of species interactions and the potential future dissolution of mutualistic interactions.

2.5 Conclusion

We introduced a novel approach to model the evolutionary switching of interaction types by tracking the evolution of absolute fitness. Applying our approach to two models of coevolutionary trait escalation inspired by plant-pollinator interactions revealed mutualisms mediated by a "bigger-is-better" (ie., trait-differences) mechanism inevitably dissolve into parasitism. In contrast, our results show that mutualisms mediated by an offset-matching mechanism, a generalization of classical trait-matching, are preserved when the intrinsic benefits of the interaction are large enough. Our results are based on minimal models of mean trait coevolution determined by different interaction mechanisms and hence ignore the effects of external evolutionary processes and eco-evolutionary feedbacks. Predicting when wild populations of mutualists are prone to disintegration via coevolutionary arms races will likely require more complex models that account for a variety of processes such as abiotic stabilizing selection, gene-flow, random genetic drift and multispecific interactions. Building on recently established methods of coevolutionary inference, we discussed a path towards developing methods to project the evolutionary switching of interaction types. Future work applying the analytical approach developed here to models that account for realistic sets of evolutionary processes can enrich our understanding of the dynamical nature of ecological relationships observed in the wild and produce novel statistical tools to forecast the evolutionary stability of ecological relationships.

Chapter 3: The Measurement of Coevolution in the Wild[1]

Abstract

Coevolution has long been thought to drive the exaggeration of traits, promote major evolutionary transitions such as the evolution of sexual reproduction, and influence epidemiological dynamics. Despite coevolution's long suspected importance, we have yet to develop a quantitative understanding of its strength and prevalence because we lack generally applicable statistical methods that yield numerical estimates for coevolution's strength and significance in the wild. Here we develop a novel method that derives maximum likelihood estimates for the strength of direct pairwise coevolution by coupling a well established coevolutionary model to spatially structured phenotypic data. Applying our method to two well-studied interactions reveals evidence for coevolution in both systems. Broad application of this approach has the potential to further resolve long-standing evolutionary debates such as the role species interactions play in the evolution of sexual reproduction and the organization of ecological communities.

3.1 Introduction

Our current understanding of coevolution's importance rests upon methods that fall into two general classes: those that are broadly applicable but yield only qualitative evidence for coevolution and those that produce quantitative estimates for the strength of coevolution but can be applied only in a narrow range of systems. For example, one popular approach for inferring coevolution relies on measuring the spatial correlation between traits of interacting species and using significant interspecific correlations as evidence of a coevolutionary process (Berenbaum et al., 1986; Hanifin et al., 2008; Pauw et al., 2009; Toju, 2008). Strengths of this approach include the relative ease of collecting the relevant data and its broad applicability to a wide range of species interactions. The critical weakness of this approach, however, is that significant interspecific correlations are neither necessary nor sufficient for demonstrating coevolution (Janzen, 1980; Nuismer et al., 2010). Similarly, time-shift experiments have been broadly implemented in systems where experimental evolution is a tractable approach, but do not yield quantitative estimates of the strength of coevolution (Blanquart and Gandon, 2013; Gaba and Ebert, 2009; Koskella, 2014). In contrast, more quantitative approaches such as selective source analysis, a method that additively partitions selection gradients into independent components of selection (Ridenhour, 2005), require the collection of extensive trait and fitness data from interacting species and thus have proven difficult to employ in all but a few specialized systems (Brodie III and Ridenhour, 2003; Burkhardt et al., 2012; Nuismer and Ridenhour, 2008). As a consequence of these trade-offs in existing approaches, rigorous quantitative estimates of the strength of coevolution in natural populations are extremely scarce.

A promising alternative to existing approaches is the development of model-based inference methods that use easily collected phenotypic data to estimate the significance of well established coevolutionary models and hence to test for the significance of coevolution. In particular, coevolutionary

[1]This chapter was previously published as: Week, B., Nuismer, S.L. 2019. The Measurement of Coevolution in the Wild. Ecology Letters 22(4):717–725. DOI: 10.1111/ele.13231.

models now exist that predict the statistical distribution of traits across multiple populations for a pair of interacting species that evolve in response to random genetic drift, abiotic selection, and coevolution (Nuismer et al., 2010). Crucially, these models predict that the distribution of local population trait means in the interacting species across a metapopulation will approach a bivariate normal distribution entirely described by five statistical moments: the average value of the key trait in each species among populations, the variance of the key trait in each species among populations, and the spatial association (covariance) between the key traits in each species. The phenotypic data necessary to calculate these statistical moments can be visualized as a two-dimensional scatter plot. Where each axis measures the mean trait value for one of the species. Hence, each point in the scatter plot corresponds to a pair of mean traits of the two interacting species within a given population.

Because the models predict a bivariate normal distribution of traits, calculating the likelihood of observing any particular set of trait values in a pair of interacting species is straightforward. With the five statistical moments that describe the bivariate normal distribution, we can infer up to five model parameters. The five parameters our method infers includes strengths of reciprocal selection caused by the focal interaction (the strengths of biotic selection B_1, B_2), the strengths of selection due to any other source (the strengths of "abiotic" selection A_1, A_2), and the optimal offset between trait values that optimize biotic fitness (δ). The parameters quantifying selection (B_i and A_i) are proportional to the selection gradients due to the biotic and abiotic components of selection in each population (see Appendix C.1.3). By maximizing the resulting likelihood with respect to these key parameters, our method can be used to rigorously test for the presence of coevolution. Specifically, for a coevolutionary hypothesis to be supported, reciprocal selection must be demonstrated (Janzen, 1980; Thompson, 1994). In our maximum likelihood framework, this long-standing and widely accepted definition of coevolution corresponds to demonstrating that both strengths of biotic selection are significantly non-zero. By performing likelihood ratio tests, support for the coevolutionary hypothesis can be compared relative to support for the null hypotheses of unilateral evolution where $B_1 = 0$ or $B_2 = 0$ (also referred to as tracking, see Figure 3.1). Due to the nested structure of these models, the likelihood of coevolution and the likelihoods of the null models can be directly compared via likelihood ratio tests. Figure 3.1 shows that each p-value p_1 and p_2 must be less than the significance threshold α (we use $\alpha = 0.05$) to support a coevolutionary hypothesis. Rejecting either null hypothesis of unilateral evolution automatically implies the rejection of evolution completely absent of biotic selection ($B_1 = B_2 = 0$) since the likelihood of the this third null model will always be less than the likelihoods of tracking.

While showing both B_1 and B_2 are non-zero is necessary for demonstrating the significance of pairwise coevolution, the strength of coevolution can most easily be quantified as the geometric mean of the absolute value of the two biotic selection strengths: $\mathfrak{C} \equiv \sqrt{|B_1 B_2|}$. If either strength of biotic selection is zero, and hence coevolution is absent, then $\mathfrak{C} = 0$ as desired and if $|B_1| = |B_2|$, then $\mathfrak{C} = |B_1| = |B_2|$. However, our metric $\mathfrak{C}$ fails to capture a sense of balance in the forces of biotic selection. We therefore propose an accompanying measure based on Shannon entropy that takes this into account. Setting $b_i = |B_i|/(|B_1| + |B_2|)$ we define the balance of coevolutionary selection as

$$\mathfrak{B} \equiv \frac{(b_1 \ln b_1 + b_2 \ln b_2)}{\ln(1/2)}. \tag{3.1}$$

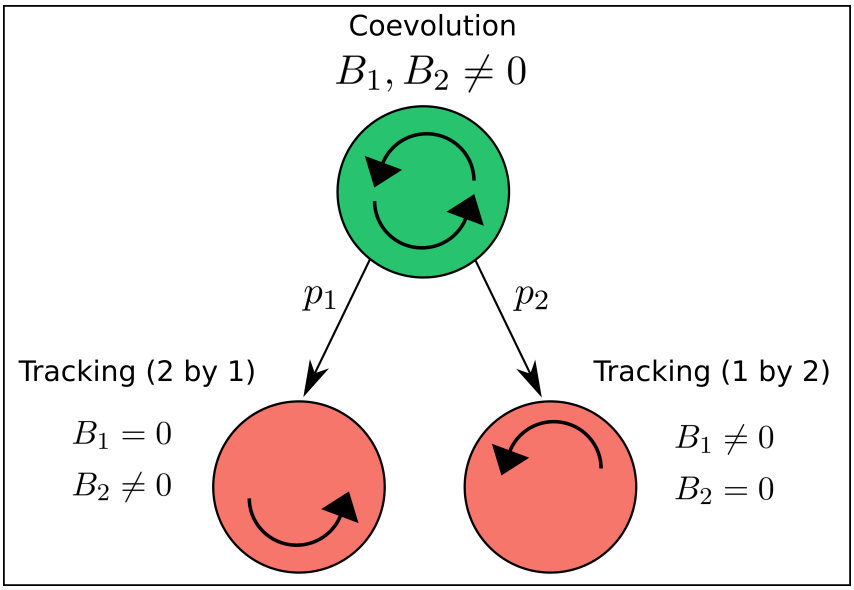

Figure 3.1: The network structure of hypotheses that can be distinguished using our approach. Nodes represent the three relevant hypotheses for coevolutionary inference. Edges represent comparisons labeled by their p-values. The upper node (in green) represents the coevolutionary hypothesis in which both strengths of selection induced by the interaction are non-zero. The pink colored nodes represent the hypotheses of unilateral evolution, or tracking, where one species experiences biotic selection, but the other does not. By ruling out tracking this approach automatically rejects evolution completely absent of biotic selection.

Standardizing by $\ln(1/2)$ makes $0 \leq \mathfrak{B} \leq 1$ with $\mathfrak{B} = 1$ representing perfect balance and $\mathfrak{B} = 0$ representing unilateral evolution. Though the strength and balance of coevolution can be subjectively inferred upon inspection of the biotic selection strengths, these two metrics provide a way to quantitatively compare these aspects of coevolution across systems.

3.2 Materials and methods

3.2.1 The coevolutionary model

To model the coevolutionary process, we begin by considering a local population level model of pairwise coevolution. This model assumes fitness is a function of the environment, the trait of the focal individual and the trait of the individual being encountered. In particular, we assume species i has an optimal phenotype θ_i that maximizes fitness in the absence of the interaction (the abiotic phenotypic optimum). We define A_i to be the strength of abiotic selection on species i so that the abiotic component of fitness ($W_{A,i}$), as a function of the trait value z_i, is proportional to

$$W_{A,i} \propto \exp\left(-\frac{A_i}{2}(\theta_i - z_i)^2\right). \tag{3.2}$$

Likewise, beginning from first principles, we derive the biotic component of fitness for an individual of species i. We assume that biotic fitness is maximized when the trait value of the focal individual z_i

is offset from the trait value being encountered z_j by an ideal amount δ. We refer to δ as the "optimal offset". A simple example of an optimal offset comes from considering the interaction between long-tubed flowers and the long-proboscid flies that visit them. The biotic component of fitness for the fly is maximized when its proboscis is slightly longer than the nectar tube depth of the flower, allowing the fly to easily extract its nectar reward. The difference between tube depth and proboscis length that maximizes the flies biotic fitness component is the optimal offset for the fly. Note how this differs from a "bigger is better" situation commonly referred to for the explanation of coevolutionary arms races. Under the optimal offset model, fitness is a unimodal function and therefore does not increase indefinitely with larger (or lesser) trait values. A more general model would allow different δ's for each species, but since our method can only infer up to five parameters we make the parsimonious assumption that both species have the same optimal offset. Defining B_i to be the strength of biotic selection on species i, the biotic component of fitness ($W_{B,i}$) is proportional to

$$W_{B,i} \propto \exp\left(-\frac{B_i}{2}(\bar{z}_j + \delta_i - z_i)^2\right) \tag{3.3}$$

when biotic selection is weak ($|B_i| \ll 1$). Here $\bar{z}_j$ is the within population average phenotype of species j. Net fitness is given by the product of the abiotic and biotic components of fitness. Since the amount by which fitness is proportional to these values is irrelevant for evolutionary dynamics, we leave them out here. Detailed derivations are provided in Appendix C.1. As noted above our method infers values for B_1, B_2, A_1, A_2 and δ and can thus accommodate most coevolutionary scenarios including escalation ($\delta \neq 0$) and matching ($\delta = 0,\ B_1, B_2 > 0$).

With a functional form of fitness in hand, we employed theoretical quantitative genetics to formally derive the local population model of mean trait dynamics for the two species. From this local model we derived the dynamics of the distribution of pairs of mean traits across the metapopulation. Since our model predicts the metapopulation distribution of mean-trait-pairs will converge to a bivariate normal (a proof is given in Appendix C.1.6), we are justified in tracking only the first five moments of the metapopulation distribution. These are the metapopulation mean traits of each species (μ_1 and μ_2), the metapopulation variance of local mean traits for each species (V_1 and V_2) and the metapopulation covariance of local mean traits for the two species (C). For species i we denote the additive genetic variance by G_i and the local effective population size by n_i. Results derived in Appendix C.1 demonstrate that the five moments change according to the following recursions:

$$\Delta\mu_1 = G_1\{B_1\delta + B_1(\mu_2 - \mu_1) + A_1(\theta_1 - \mu_1)\} \tag{3.4a}$$

$$\Delta\mu_2 = G_2\{B_2\delta + B_2(\mu_1 - \mu_2) + A_2(\theta_2 - \mu_2)\} \tag{3.4b}$$

$$\Delta V_1 = -2A_1G_1V_1 + 2B_2G_2(C - V_1) + G_1/n_1 \tag{3.4c}$$

$$\Delta V_2 = -2A_2G_2V_2 + 2B_1G_1(C - V_2) + G_2/n_2 \tag{3.4d}$$

$$\Delta C = B_2G_2(V_1 - C) + B_1G_1(V_2 - C) - (A_1G_1 + A_2G_2)C. \tag{3.4e}$$

3.2.2 Parameter estimation

After solving for the equilibrium expressions of the first five moments from equations (4), we use maximum likelihood to estimate the selection strengths (A_1, A_2, B_1 and B_2) and the optimal offset (δ). However, to do so requires more than estimates of mean trait pairs from multiple populations. Background parameters of the model also need to be estimated. These include the effective population sizes n_1, n_2, the optimal phenotypes favored by abiotic stabilizing selection θ_1, θ_2 and the additive genetic variances G_1, G_2.

We show in Appendix C.1.4 that if n_i has been estimated from multiple locations, these can be included by using their harmonic mean as the effective population size in our model. Likewise, if G_i has been estimated from multiple populations, these can be included by using their arithmetic mean as the effective additive genetic variance for our model. Finally, the model used in this manuscript assumes the abiotic optimum is constant across space. In the associated Mathematica notebook, we expand the model to formally account for variable θ_i. The results of this notebook demonstrate that the two models are equivalent when variation in θ_i is small and therefore implies that the average abiotic optimum across space works as the effective abiotic optimum needed to perform inference. This notebook also implies that our method is readily adaptable for the inclusion of spatially varying optima as such data become available.

The likelihood is a routine calculation in terms of the first five moments which are in turn functions of model parameters $(n_1, n_2, \theta_1, \theta_2, G_1, G_2, \delta)$ and selection strengths (A_1, A_2, B_1, B_2). In Appendix C.2 we show how to invert these expressions to obtain analytic solutions for the maximum likelihood estimates of selection strengths. Full expressions are provided in the associated Mathematica notebook. Although our focus is on finding point estimates for the strengths of biotic selection, coevolution and coevolutionary balance, we also estimated uncertainty due to error caused by sampling from the metapopulation. To do so we calculated 95% confidence intervals for each selection strength.

3.2.3 Estimating significance

Denoting the likelihood of the coevolutionary model by L_c and the likelihood of null model i (for which $B_i = 0$) by L_i, we compute the log-likelihood difference statistic by

$$\Lambda_i = 2(\ln L_c - \ln L_i). \tag{3.5}$$

Denote by $F_j(x)$ the distribution function of a χ^2 random variable with degrees of freedom j. Wilk's theorem implies the distribution of Λ_i is approximately a χ^2 (Wilks, 1938). Since in each null model we fix just one parameter, the degrees of freedom is one for both tests. Thus, the p-value associated with testing against null hypothesis i (written p_i) has the following approximation

$$p_i \approx 1 - F_1(\Lambda_i). \tag{3.6}$$

If both p_1 and $p_2 < 0.05$ for a given study system then our method asserts significant evidence for coevolution exists in this system. We provide a tutorial for implementing our approach using the

statistical programming language *R* at the following url:

`https://bobweek.github.io/measuring_coevolution.html`

3.2.4 Evaluation of performance

Before applying our maximum likelihood methodology to specific study systems, we evaluated its performance when challenged with simulated data. We assessed the type-1 error rate and statistical power of our method across a range of biotic selection strengths and metapopulation sample sizes. These analyses were performed by simulating data under the model with randomly drawn model parameters. Distributions used for each background parameter are reported in Table 3.1. For error rates as functions of biotic selection strengths, sample sizes were drawn at random from a Poisson distribution with a mean of 20. Draws were repeated until a sample size of at least three was obtained. For type-1 error rates as functions of unilateral selection we chose one biotic strength to be zero and set the other to the strength of unilateral selection. For type-2 error rates as functions of the strength of coevolution $\mathfrak{C}$, we drew one biotic selection strength from a uniform distribution on the interval $(\mathfrak{C}/10, 10\mathfrak{C})$ and set the other such that their geometric mean equates to $\mathfrak{C}$. When calculating type-2 error rates as functions of sample size, strengths of biotic selection were drawn independently from a uniform distribution on (0,0.01). A similar approach was taken for calculating type-1 error rate as a function of sample size, except one or both of the biotic selection strengths were set to zero at random. If either strength of biotic selection was set to zero in the simulation and reported significantly non-zero by our method, a false positive was accumulated. Likewise, if both strengths of biotic selection were set to some non-zero number and our method failed to detect coevolution, then a type-2 error was accumulated. This scheme was repeated 10,000 times for each estimated error rate.

Table 3.1: Distributions of background parameters used for generating error rates and regression analyses.

Parameter(s)	Description	Distribution
A_i	Strength of abiotic selection	Uniform(0,0.01)
δ	Optimal offset	Exp(0.1)
θ_i	Abiotic optima	Normal(0,10)
G_i	Additive genetic variance	Exp(1)
n_i	Effective population size	Exp(0.01)

Alongside our analyses of error rates, we investigated our methods ability to accurately infer the strength of coevolution using simulated data. For each replicate, we simulated phenotypic data using the coevolutionary model with known selection strengths and background parameters drawn from the same set of distributions as those used for the error rates as functions of sample size analysis. We then estimated the strength of coevolution as defined above using our maximum likelihood approach and compared it against its actual value via linear regression. Each regression was performed across a range of sample sizes (Figure 3.2). We also extended this analysis using more general simulations that relax key assumptions such as the absence of gene-flow and normality of data in Appendix C.3.

Numerical analyses of our methods performance were done using the statistical programming language *R*. The scripts are publicly available at the following Github repository:

`https://github.com/bobweek/measuring.coevolution`

3.2.5 Measuring coevolution in the wild

We next applied our maximum likelihood approach to two well-studied species interactions where previous work implicated coevolution as a cause of trait exaggeration and spatial variability (Pauw et al., 2009; Toju, 2011): the mutualism between the long tongued fly *Moegistorhynchus longirostris* and a plant it pollinates *Lapeirousia anceps* as well as the antagonism between the camellia plant *Camellia japonica* and its seed predator, the weevil *Curculio camelliae*. In both cases, the interactions are thought to depend largely on a single key trait in each species (fly proboscis and plant floral tube lengths or weevil rostrum length and camellia pericarp thickness). This is a crucial detail as the models upon which our method is based assume interactions are mediated by a single trait in each species. Phenotypic data for these systems have been collected from several populations, providing a sample of pairs of mean trait values, the core data required by our method. In addition to the essential phenotypic data, previous work in both systems provided valuable additional information that allowed us to estimate the key background parameters required by our method: the likely trait optima in the absence of the interaction (the "abiotic" optima), the effective population sizes for each species (assumed fixed over time and space), and the effective additive genetic variances for each species (also assumed to be fixed over time and space).

The long proboscid fly, *M. longirostris*, resides in lowland habitats near the coast of South Africa and pollinates a guild of at least 20 plant species (Manning and Goldblatt, 1997). Among these species, the most widespread is *L. anceps*, a long tubed perennial whose distribution extends outside the range of *M. longirostris* (Pauw et al., 2009). We were able to estimate the likely optimal tube and proboscis lengths for these species in the absence of this particular interaction. Using the phenotypic data published by Pauw et al. (2009), we inferred this parameter for the flower as the average mean tube length of two populations not visited by the fly. Estimating the abiotic optima for the fly was more challenging because we were unable to identify fly populations where the plant did not co-occur. However, there are data available for the proboscis lengths in three sister species of *M. Longirostris* (41.0 mm for *M. braunsi*, 11.5 mm for *M. brevirostris*, and 32.0 mm for *M. perplexus*) (Bequaert, 1935). Since these sister species do not interact with *L. anceps* (Barraclough and Slotow, 2010), their traits represent potential evolutionary trajectories that could have been taken by *M. longirostris* in the absence of its interaction with *L. anceps*. Given that none of the three sister species underwent a similar arms race with some other flower (which appears likely based on their relatively modest proboscis lengths), we therefore take these values as rough approximations of the actual abiotic optimal phenotype for *M. longirostris*. Hence, we estimated selection strengths and significance when the abiotic optimum was set equal to each of the three trait values and the average of all three. The result presented in the main text correspond to the average of all three sister species, but we present the results for all four abiotic optima in Appendix C.4. Effective population sizes have not been estimated for either species. We therefore

relied on the biologically plausible census sizes of 1000 for *L. anceps* and 100 for *M. longirostris*, as suggested by B. Anderson (personal communications). Since heritabilities for neither of these traits have been estimated, we relied on within population phenotypic variances as a rough proxy for the additive genetic variances in this system.

We complement our analysis of this plant pollinator mutualism with an analysis of the antagonistic interaction between *C. camelliae* and *C. japonica* (Toju and Sota, 2005). Female weevils bore holes into the woody pericarps of the camellia to oviposit. Inside the fruit, weevil larvae feed on the seeds of the camellia up until the fourth instar, at which time they exit the fruit and overwinter (Toju and Sota, 2005). These two species co-occur across Japan, although camellia populations where the weevil is absent also exist (Toju and Sota, 2005). We were able to establish point estimates of each background parameter using data from previously published work (Toju et al., 2011a,b) and the fact that male weevil rostrum lengths could be used as a proxy for the abiotic optimum of the female weevils since males do not interact with the camellia. Hence, our method does not inherently require estimates of abiotic optima to come from populations where the interaction is absent. However, using male traits as a surrogate for the abiotic optimum assumes that male and female trait values are either genetically uncorrelated or have reached equilibrium. The abiotic optimum for the pericarp thickness of the camellia was inferred by averaging pericarp thicknesses across populations where weevils are absent. Heritability of pericarp thickness has been estimated directly (Toju et al., 2011a) and can be at least crudely inferred for weevil rostrum length via estimates of related species (Toju and Sota, 2009). We used the average of these values for each species multiplied by the average within population phenotypic variances to estimate additive genetic variances in this system.

To assess the biological significance of the strengths of coevolution inferred, we compared the distribution of trait values we would expect in the presence vs absence of coevolution. This was accomplished by setting both B_1 and B_2 equal to zero and maximizing the resulting restricted likelihood function with the remaining free parameters (A_1, A_2 and δ). Using a multivariate generalization of effect size (see Appendix C.4.3), we summarize with a single number the effect of coevolution in each system.

3.3 Results

3.3.1 Evaluation of performance

Regressions of randomly drawn strengths of coevolution onto those inferred by our method were heteroskedastic with variation proportional to the strength of coevolution (Bartlett's test: p-value $< 2.22e-16$). To rectify this we used weighted least squares. For each point in the regression we set its weight equal to the inverse of its Euclidean distance to the origin. Analysis of regression results demonstrates that at low sample sizes our method tends to overestimate the strength of coevolution, but this bias rapidly diminishes with sample size (see Figure 3.2).

False positive rates are greatly exaggerated for small sample sizes (e.g., < 5), modestly inflated for sample sizes between $5-10$, but approach their set value (0.05) for sample sizes > 10 (Figure 3.2). This behavior is attributable to two factors. First, statistical artifacts accumulate in sample moments for

small sample sizes. For example, the correlation of a sample of size two will always be ± 1. Second, the distribution of our p-values may significantly diverge from a Chi-square distribution at small sample sizes (Wilks, 1938). We therefore suggest this method only be used for sample sizes of at least five. Another important caveat, however, is that as biotic selection becomes increasingly imbalanced under the null scenario when one strength is zero and the other set to some non-zero number, the false positive rate increases monotonically (see Figure 3.2). Hence, our method can be tricked by extreme unilateral selection.

Power to detect coevolution is reasonably high at low sample sizes (≈ 0.9) and increases monotonically with sample size. As a function of the strength of coevolution, power is initially negligible but increases quickly and monotonically.

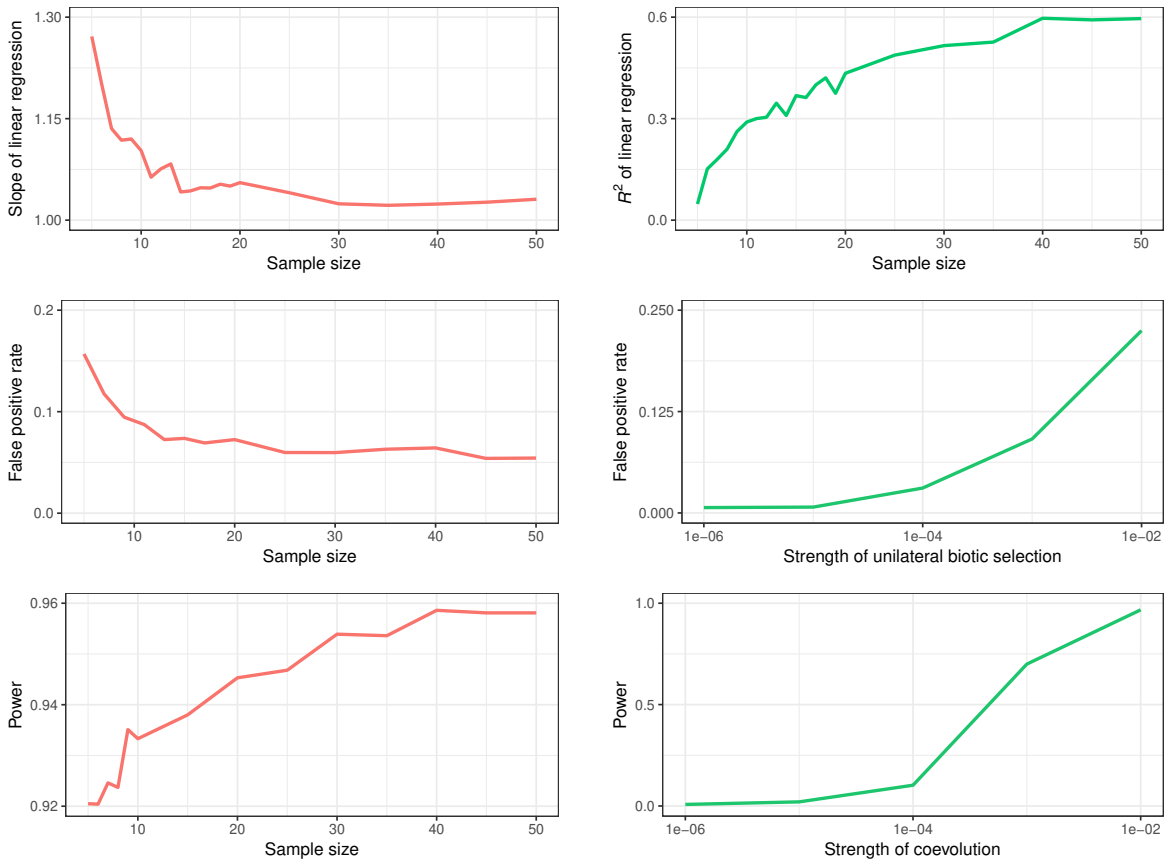

Figure 3.2: Top row: Performance of parameter estimation as a function of sample size. The left-hand panel shows the slope of the regressions converging near one as sample size increases. The right-hand plot shows the percent variance explained (R^2) increasing with sample size. Lower two rows: Error rates as functions of sample sizes and selection strengths. The left-hand column shows the type-1 and type-2 error rates as functions of sample size. The right-hand column shows type-1 error as a function of the strength of tracking (ie, unilateral selection where the species being tracked does not experience biotic selection) and power as a function of the strength of coevolution.

3.3.2 Measuring coevolution in the wild

We found that the biotic selection strengths B_1 and B_2 acting on *M. longirostris* and *L. anceps* both differ significantly from zero (Table 3.2). Thus, our analysis supports the hypothesis of pairwise coevolution in this system. Likewise, both B_1, the strength of biotic selection on the weevil, and B_2, biotic selection on the camellia plant, significantly differed from zero. Hence, we also found evidence for pairwise coevolution between the seed-eating weevil *C. camelliae* and its host plant *C. japonica*. For numerical estimates of biotic selection strengths, p-values, and the strength and balance of coevolution, see Table 3.2. Cross-system comparison of biotic selection strengths is visualized in Figure 3.3.

Table 3.2: Biotic and abiotic selection strengths, optimal offsets, p-values, and strengths of coevolution and coevolutionary balance for each system. CW refers to the camellia-weevil system and FF refers to the fly-flower system. Units of selection strengths are all inverse square phenotypic units (mm^{-2} in this case). Optimal offsets (δ) are in phenotypic units (mm). The p-values and balances of coevolutionary selection are unitless.

	CW	FF
B_1	7.17e-04	6.40e-05
B_2	5.00e-06	1.84e-06
A_1	2.59e-04	7.04e-06
A_2	8.05e-06	3.13e-06
δ	4.51	14.2
p_1	<2.22e-16	<2.22e-16
p_2	<2.22e-16	1.19e-07
$\mathfrak{C}$	5.99e-05	1.08e-05
$\mathfrak{B}$	5.97e-02	1.84e-01

In addition to providing information on the magnitude and significance of coevolution, we quantified the extent of trait exaggeration produced by coevolution by comparing the equilibrium phenotypic distribution we would expect with and without the levels of coevolution we estimated (Figure 3.4). This comparison reveals that although the numerical estimates of coevolutionary selection appear superficially small, for the camellia-weevil interaction coevolution results in a 111% increase in the mean rostrum length of the camellia weevil and a 66.0% increase in the pericarp thickness of the camellia fruit (Figure 3.4). For the fly-flower system coevolution appears to have caused a 134% increase in proboscis length and a 34.5% increase in floral tube depth compared to equilibrium estimates for these values we predict when coevolution is absent. Using a multivariate analog of effect size we calculated the effect of coevolution in each system. We found an effect size of 7.55 for the fly-flower system and an effect size of 3.07 for the camellia-weevil interaction.

3.4 Discussion

Our results demonstrate that coupling existing coevolutionary models with a maximum likelihood approach allows the strength of coevolutionary selection to be estimated using routinely collected phenotypic data. Regression analysis shows that with sufficient sample sizes we can obtain accurate

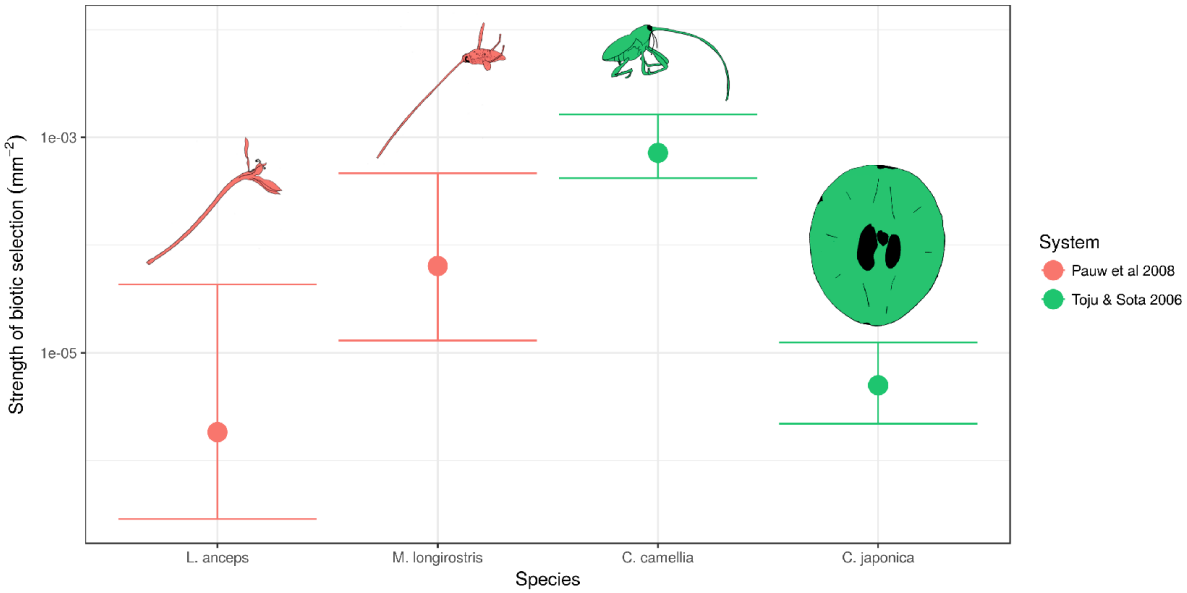

Figure 3.3: The estimated strength of biotic selection for the *M. longirostris-L. anceps* interaction (pink) and the *C. japonica-C. camelliae* interaction (green). Units for each strength are in mm^{-2}, the inverse of the square of the phenotypic units. 95% confidence intervals are shown around each estimate. Each selection strength was found to be statistically significant and hence coevolution was detected in both systems.

estimates of the strength and significance of coevolution. Furthermore, our method is robust to modest amounts of gene flow and weakly non-normal data (Appendix C.3).

Applying our method to two textbook examples of pairwise coevolution, we find strong evidence for significant coevolution in both systems. This qualitative result is complemented by quantitative estimates of the strength of coevolution in the wild. By applying this method to various systems, it will be possible to obtain an empirical distribution of the strength of coevolution in nature. After the appropriate transformation (analogous to standardizing selection gradients with respect to phenotypic distributions) such data will allow for a meta-analysis akin to (Kingsolver et al., 2001; Siepielski et al., 2009, 2013) which would provide a yardstick allowing us to further understand the biological significance of our numerical results.

In spite of the various merits of our method, there are serious limitations that must be confronted empirically. Most notable is the necessity of providing estimates of abiotic optima. Since these parameters are seldomly estimated for natural populations, we are restricted in our analysis here to two data sets in which sufficient information was provided. In particular, phenotypic measurements in populations that do not partake in the interaction (due to geographical isolation or sexual dimorphism) provide reasonable estimates of the abiotic optima, though other means of estimating these parameters exist as demonstrated above.

Alongside the empirical work necessary for estimating background parameters of our model, our results suggest that increasing the number of populations used in studies of trait matching would also substantially improve opportunities for coevolutionary inference. Specifically, we suggest sample sizes

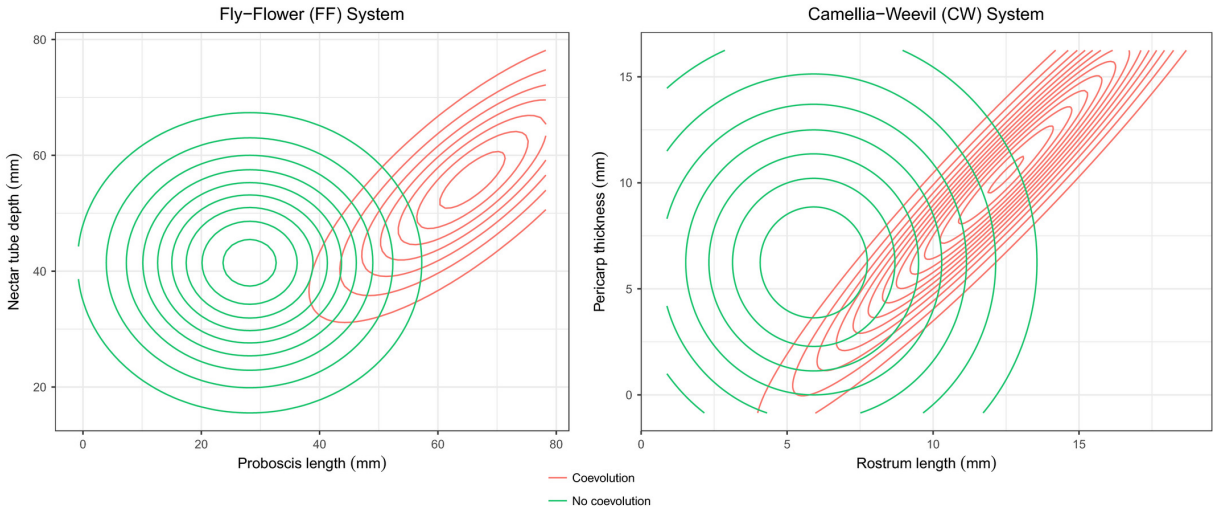

Figure 3.4: The effect of coevolution on the trait distributions predicted by our model. The point in the center of each contour represents the mean traits of the species involved. The green contours represent data predicted without coevolution and the pink contours represent the observed data.

of at least five and ideally more than twenty to avoid type-1 errors. Taken together, these considerations outline a reasonably tractable set of sufficient conditions empirical data-sets must meet in order to utilize our method.

Theoretical limitations of our approach stem from its grounding in classic quantitative genetics and include the assumptions of fixed additive genetic variance and weak selection. Although we do not assume strict equilibrium for each component population, we do assume that the system as a whole has reached approximate statistical equilibrium so that the means, variances and spatial covariance have become relatively constant with respect to time. This implies that pairs of species for which this method is ideal have been interacting for a sufficiently long period of time. In reality, however, empirical systems may be far enough from equilibrium that a significant contemporary trend in the five moments describing their distribution should be accounted for. Lastly, our method assumes the key traits mediating the interaction are univariate which may not be ubiquitous across coevolving systems. Future work that generalizes our approach to multivariate traits, strong selection and non-equilibrium (ie, time-series data) will result in a more broadly applicable method.

By providing a methodology that does not rely on extensive and system specific experimental manipulation, our approach greatly expands the range of systems for which the strength of coevolutionary selection can be estimated, paving the road for a more quantitative and critical assessment of coevolution's importance in natural systems. To add substance to this claim we provide three examples. First, with finer spatial resolution in phenotypic data this method can be applied to the same pair of species across different partitions of their range to infer the strength of selection mosaics argued to be central to the coevolutionary process by the Geographic Mosaic Theory of Coevolution (Thompson, 2005). Second, previous investigations have resulted in mixed views on the significance of pairwise coevolution in shaping various aspects of ecological communities including inter- and intraspecific diversity, demographic stability, network structure and ecosystem function (Althoff et al., 2014; Iwao and Rausher,

1997; Nuismer et al., 2013; Roughgarden, 1979; Yamamura et al., 2001). By applying our method to each pairwise interaction in a set of interacting species, the distribution of pairwise coevolution can be inferred within a community to provide empirical insight into the degree to which coevolution molds the previously mentioned properties of ecological communities. Third, theoretical studies suggest that only very strong coevolution favors the evolution of sexual reproduction (Agrawal, 2006; Lively, 2010; Otto and Nuismer, 2004). Our method could inform this hypothesis by determining the strength of coevolution in specific systems where the evolution of sex has been attributed to interspecific interactions. Hence, when coupled with data from a broad range of empirical systems, this method and its future iterations hold the potential to settle long standing debates involving the importance of species interactions and coevolution in the evolution of various phenomena including phenotypic diversity, sexual reproduction, community structure, and epidemiological dynamics (Anderson and May, 1982; Hamilton, 1980; McPeek, 2017; Yoder and Nuismer, 2010).

Literature Cited

A. F. Agrawal. Similarity selection and the evolution of sex: revisiting the red queen. *PLoS biology*, 4 (8):e265, 2006.

D. M. Althoff, K. A. Segraves, and M. T. Johnson. Testing for coevolutionary diversification: linking pattern with process. *Trends in Ecology & Evolution*, 29(2):82–89, 2014.

B. Anderson and S. D. Johnson. The geographical mosaic of coevolution in a plant–pollinator mutualism. *Evolution*, 62(1):220–225, Oct. 2007.

B. Anderson and S. D. Johnson. Geographical covariation and local convergence of flower depth in a guild of fly-pollinated plants. *New Phytologist*, 182(2):533–540, Feb. 2009.

B. Anderson, J. S. Terblanche, and A. G. Ellis. Predictable patterns of trait mismatches between interacting plants and insects. *BMC Evolutionary Biology*, 10(1):204, 2010.

R. M. Anderson and R. May. Coevolution of hosts and parasites. *Parasitology*, 85(2):411–426, 1982.

K. B. Athreya and P. E. Ney. *Branching Processes*. Springer Berlin Heidelberg, 1972.

S. Axler. *Measure, Integration & Real Analysis*. Graduate Texts in Mathematics. Springer International Publishing, 2019.

D. Barraclough and R. Slotow. The south african keystone pollinator moegistorhynchus longirostris (wiedemann, 1819)(diptera: Nemestrinidae): notes on biology, biogeography and proboscis length variation. *African Invertebrates*, 51(2):397–403, 2010.

N. H. Barton and A. M. Etheridge. Mathematical models in population genetics, 2019.

N. H. Barton, A. M. Etheridge, and A. Véber. Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01002, 2013.

N. H. Barton, A. M. Etheridge, and A. Véber. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, 118:50–73, Dec. 2017.

P. Beerli. Comparison of bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, 22(3):341–345, 2005.

C. W. Benkman, T. L. Parchman, A. Favis, and A. M. Siepielski. Reciprocal selection causes a coevolutionary arms race between crossbills and lodgepole pine. *The American Naturalist*, 162(2):182–194, Aug. 2003.

J. Bequaert. Notes on the genus moegistorhynchus and description of a new african species of nycterimyia (diptera, nemestrinidae). *Annals of the Transvaal Museum*, 15(4):491–502, 1935.

M. Berenbaum, A. Zangerl, and J. Nitao. Constraints on chemical coevolution: wild parsnips and the parsnip webworm. *Evolution*, pages 1215–1228, 1986.

J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. *Probability Theory and Related Fields*, 126(2):261–288, June 2003.

F. Blanquart and S. Gandon. Time-shift experiments and patterns of adaptation across time and space. *Ecology letters*, 16(1):31–38, 2013.

E. D. Brodie, C. R. Feldman, C. T. Hanifin, J. E. Motychak, D. G. Mulcahy, B. L. Williams, and E. D. Brodie. Parallel arms races between garter snakes and newts involving tetrodotoxin as the phenotypic interface of coevolution. *Journal of Chemical Ecology*, 31(2):343–356, Feb. 2005.

E. Brodie III and B. Ridenhour. Reciprocal selection at the phenotypic interface of coevolution. *Integrative and Comparative Biology*, 43(3):408–418, 2003.

W. L. Brown and E. O. Wilson. Character displacement. *Systematic Zoology*, 5(2):49, June 1956.

W. Bryc. *The normal distribution: characterizations with applications*, volume 100. Springer Science & Business Media, 1995.

M. G. Bulmer. The effect of selection on genetic variability. *The American Naturalist*, 105(943):201–211, May 1971.

R. Bürger. On the maintenance of genetic variation: global analysis of kimura's continuum-of-alleles model. *Journal of Mathematical Biology*, 24(3):341–351, 1986.

R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, 2000.

A. Burkhardt, B. Ridenhour, L. Delph, and G. Bernasconi. The contribution of a pollinating seed predator to selection on silene latifolia females. *Journal of evolutionary biology*, 25(3):461–472, 2012.

R. S. Cantrell and C. Cosner. *Spatial Ecology via Reaction-Diffusion Equations*. Wiley, Jan. 2004.

N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. *Theoretical Population Biology*, 69(3):297–321, May 2006.

P. Chesson. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31(1):343–366, Nov. 2000.

J. F. Crow and M. Kimura. *An Introduction to Population Genetics Theory*. The Blackburn Press, 1970.

G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, 2014.

C. Darwin. *On the Various Contrivances by which British and Foreign Orchids are Fertilised by Insects: And on the Good Effects of Intercrossing*. John Murray, 1862.

D. A. Dawson. Stochastic evolution equations and related measure processes. *Journal of Multivariate Analysis*, 5(1):1–52, 1975.

D. A. Dawson. Geostochastic calculus. *Canadian Journal of Statistics*, 6(2):143–168, 1978.

D. A. Dawson. Measure-valued markov processes. In *École d'été de Probabilités de Saint-Flour XXI-1991*, pages 1–260. Springer, 1993.

C. S. de Andreazzi, J. Astegiano, and P. R. Guimarães. Coevolution by different functional mechanisms modulates the structure and dynamics of antagonistic and mutualistic networks. *Oikos*, 129(2):224–237, Nov. 2019.

U. Dieckmann and R. Law. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of Mathematical Biology*, 34(5-6):579–612, May 1996.

M. Doebeli. Quantitative genetics and population dynamics. *Evolution*, 50(2):532–546, Apr. 1996.

M. Eigen and P. Schuster. *The Hypercycle: A Principle of Natural Self-Organization*. Springer Berlin Heidelberg, 1979.

M. Eigen, J. McCaskill, and P. Schuster. Molecular quasi-species. *The Journal of Physical Chemistry*, 92 (24):6881–6891, Dec. 1988.

A. M. Etheridge. *An Introduction to Superprocesses*. American Mathematical Society, aug 2000.

A. M. Etheridge. Drift, draft and structure: some mathematical models of evolution. In *Stochastic Models in Biological Sciences*. Institute of Mathematics Polish Academy of Sciences, 2008.

A. M. Etheridge and P. A. March. A note on superprocesses. *Probability Theory and Related Fields*, 89(2): 141–147, June 1991.

L. C. Evans. *Partial Differential Equations: Second Edition*. American Mathematical Society, 2010.

L. C. Evans. *An Introduction to Stochastic Differential Equations*. American Mathematical Society, 2014.

S. N. Evans and E. A. Perkins. Measure-valued branching diffusions with singular interactions. *Canadian Journal of Mathematics*, 46(1):120–168, Feb. 1994.

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

S. J. Farlow. *Partial Differential Equations for Scientists and Engineers*. Dover, sep 1993.

W. Feller. Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 227–246. University of California Press, 1951.

J. Felsenstein. A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist*, 109(967):359–368, 1975.

R. A. Fisher. XV.—the correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.

R. A. Fisher. XXI.-on the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, 42:321–341, 1923.

S. A. Frank. Natural selection. IV. the price equation$*$. *Journal of Evolutionary Biology*, $25(6) : 1002 - -1019, 2012$.

S. Gaba and D. Ebert. Time-shift experiments as a tool to study antagonistic coevolution. *Trends in Ecology & Evolution*, 24(4):226–232, 2009.

S. Gavrilets. Coevolutionary chase in exploiter–victim systems with polygenic characters. *Journal of Theoretical Biology*, 186(4):527–534, June 1997.

S. Gavrilets and A. Hastings. Coevolutionary chase in two-species systems with applications to mimicry. *Journal of Theoretical Biology*, 191(4):415–427, Apr. 1998.

R. Gomulkiewicz, S. M. Krone, and C. H. Remien. Evolution and the duration of a doomed population. *Evolutionary Applications*, 10(5):471–484, Mar. 2017.

N. Gotelli. *A Primer of Ecology*. Oxford University Press, 2001.

P. R. Guimarães, M. M. Pires, P. Jordano, J. Bascompte, and J. N. Thompson. Indirect effects drive coevolution in mutualistic networks. *Nature*, 550(7677):511–514, Oct. 2017.

W. D. Hamilton. Sex versus non-sex versus parasite. *Oikos*, pages 282–290, 1980.

C. T. Hanifin, E. D. Brodie Jr, and E. D. Brodie III. Phenotypic mismatches reveal escape from arms-race coevolution. *PLoS biology*, 6(3):e60, 2008.

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

D. W. Inouye. The terminology of floral larceny. *Ecology*, 61(5):1251–1253, Oct. 1980.

K. Iwao and M. D. Rausher. Evolution of plant resistance to multiple herbivores: quantifying diffuse coevolution. *The American Naturalist*, 149(2):316–335, 1997.

D. H. Janzen. When is it coevolution. *Evolution*, 34(3):611–612, 1980.

T. Johnson and N. H. Barton. Theoretical models of selection and mutation on quantitative traits. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1411–1425, July 2005.

E. I. Jones, M. E. Afkhami, E. Akçay, J. L. Bronstein, R. Bshary, M. E. Frederickson, K. D. Heath, J. D. Hoeksema, J. H. Ness, M. S. Pankey, S. S. Porter, J. L. Sachs, K. Scharnagl, and M. L. Friesen. Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecology Letters*, 18(11):1270–1284, Sept. 2015.

D. G. Kendall. Branching processes since 1873. *Journal of the London Mathematical Society*, s1-41(1): 385–406, 1966.

A. R. Kiester, R. Lande, and D. W. Schemske. Models of coevolution and speciation in plants and their pollinators. *The American Naturalist*, 124(2):220–243, Aug. 1984.

M. Kimmel and D. E. Axelrod. *Branching Processes in Biology*. Springer New York, 2015.

M. Kimura. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences*, 54(3):731–736, 1965.

M. Kimura and J. F. Crow. Effect of overall phenotypic selection on genetic change at individual loci. *Proceedings of the National Academy of Sciences*, 75(12):6168–6171, 1978.

J. G. Kingsolver, H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. Hill, A. Hoang, P. Gibert, and P. Beerli. The strength of phenotypic selection in natural populations. *The American Naturalist*, 157(3):245–261, 2001.

A. Kolmogorov and S. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Number v. 1. Dover, 1999.

N. Konno and T. Shiga. Stochastic partial differential equations for some measure-valued diffusions. *Probability Theory and Related Fields*, 79(2):201–225, 1988.

B. Koskella. Bacteria-phage interactions across time and space: merging local adaptation and time-shift experiments to understand phage evolution. *The American Naturalist*, 184(S1):S9–S21, 2014.

N. V. Krylov and B. L. Rozovskii. Stochastic evolution equations. *Journal of Soviet Mathematics*, 16(4): 1233–1277, 1981.

R. Lande. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical Research*, 26(3):221–235, Dec. 1975.

R. Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30(2):314–334, 1976.

R. Lande. The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. *Genetics*, 94(1):203–215, 1980.

R. Lande. A quantitative genetic theory of life history evolution. *Ecology*, 63(3):607–615, June 1982.

R. Lande and S. J. Arnold. The measurement of selection on correlated characters. *Evolution*, 37(6): 1210–1226, 1983.

R. Lande, S. Engen, and B.-E. Sæther. An evolutionary maximum principle for density-dependent population dynamics in a fluctuating environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1523):1511–1518, June 2009.

R. Levins. *Evolution in Changing Environments: Some Theoretical Explorations. (MPB-2) (Monographs in Population Biology)*. Princeton University Press, 1968.

Z.-H. Li. Absolute continuity of measure branching processes with interaction. *Chinese Journal of Applied Probability and Statistics*, 14:231–242, 1998.

S. Lion. Theoretical approaches in evolutionary ecology: Environmental feedback as a unifying perspective. *The American Naturalist*, 191(1):21–44, 2018.

C. M. Lively. A review of red queen models for the persistence of obligate sexual reproduction. *Journal of Heredity*, 101(suppl_1):S13–S20, 2010.

A. Lotka. *Elements of Physical Biology*. Williams & Wilkins, 1925.

E. Lukacs. *Characteristic functions*. Griffin, 1970.

R. H. MacArthur. Species Packing, and what Competition Minimizes. *Proceedings of the National Academy of Sciences*, 64(4):1369–1371, 1969.

R. H. MacArthur. Species packing and competitive equilibrium for many species. *Theoretical Population Biology*, 1(1):1–11, 1970.

R. H. MacArthur. *Geographical Ecology*. Princeton University Press, 1972.

R. H. MacArthur and R. Levins. The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, 101(921):377–385, 1967.

J. C. Manning and P. Goldblatt. The *Moegistorhynchus longirostris* (diptera: Nemestrinidae) pollination guild: long-tubed flowers and a specialized long-proboscid fly pollination system in southern africa. *Plant Systematics and Evolution*, 206(1-4):51–69, 1997.

M. A. McPeek. The ecological dynamics of natural selection: traits and the coevolution of community structure. *The American Naturalist*, 189(5):E91–E117, 2017.

L. P. Medeiros, G. Garcia, J. N. Thompson, and P. R. Guimarães. The geographic mosaic of coevolution in mutualistic networks. *Proceedings of the National Academy of Sciences*, 115(47):12017–12022, Nov. 2018.

M. Méléard and S. Roelly. Interacting branching measure processes. *Stochastic Partial Differential Equations and Applications (G. Da Prato and L. Tubaro, eds.)*, pages 246–256, 1992.

M. Méléard and S. Roelly. Interacting measure branching processes. some bounds for the support. *Stochastics and Stochastic Reports*, 44(1-2):103–121, Sept. 1993.

J. A. Metz, S. A. Geritz, G. Meszéna, F. J. Jacobs, and J. S. Van Heerwaarden. Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. 1996.

N. Muchhala and J. D. Thomson. Fur versus feathers: Pollen delivery by bats and hummingbirds and consequences for pollen production. *The American Naturalist*, 175(6):717–726, June 2010.

M. A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press, 2006.

S. L. Nuismer. *Introduction to Coevolutionary Theory*. W. H. Freeman, 2017.

S. L. Nuismer and B. J. Ridenhour. The contribution of parasitism to selection on floral traits in heuchera grossulariifolia. *Journal of evolutionary biology*, 21(4):958–965, 2008.

S. L. Nuismer and B. Week. Approximate bayesian estimation of coevolutionary arms races. *PLOS Computational Biology*, 15(4):e1006988, 2019.

S. L. Nuismer, B. J. Ridenhour, and B. P. Oswald. Antagonistic coevolution mediated by phenotypic differenes between quantitative traits. *Evolution*, 61(8):1823–1834, Aug. 2007.

S. L. Nuismer, R. Gomulkiewicz, and B. J. Ridenhour. When is correlation coevolution? *The American Naturalist*, 175(5):525–537, 2010.

S. L. Nuismer, P. Jordano, and J. Bascompte. Coevolution and the architecture of mutualistic networks. *Evolution*, 67(2):338–354, 2013.

S. P. Otto and S. L. Nuismer. Species interactions and the evolution of sex. *Science*, 304(5673):1018–1020, 2004.

K. M. Page and M. A. Nowak. Unifying evolutionary dynamics. *Journal of Theoretical Biology*, 219(1): 93–98, 2002.

T. L. Parsons, C. Quince, and J. B. Plotkin. Some consequences of demographic stochasticity in population genetics. *Genetics*, 185(4):1345–1354, May 2010.

A. Pauw, J. Stofberg, and R. J. Waterman. Flies and flowers in darwin's race. *Evolution*, 63(1):268–279, Jan. 2009.

E. A. Perkins. Conditional dawson-watanabe processes and fleming-viot processes. In *Seminar on Stochastic Processes, 1991*, pages 143–156. Birkhäuser Boston, 1991.

E. A. Perkins. Measure-valued branching diffusions with spatial interactions. *Probability Theory and Related Fields*, 94(2):189–245, June 1992.

E. A. Perkins. *On the Martingale Problem for Interactive Measure-Valued Branching Diffusions*. Amer Mathematical Society, may 1995.

G. R. Price. Selection and covariance. *Nature*, 227(5257):520–521, 1970.

D. C. Queller. Fundamental theorems of evolution. *The American Naturalist*, 189(4):345–353, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`.

M. Reimers. One dimensional stochastic partial differential equations and the branching measure diffusion. *Probability Theory and Related Fields*, 81(3):319–340, 1989.

B. J. Ridenhour. Identification of selective sources: partitioning selection based on interactions. *The American Naturalist*, 166(1):12–25, 2005.

A. Robertson. A mathematical model of the culling process in dairy cattle. *Animal Science*, 8(1):95–108, 1966.

J. Roughgarden. *Theory of population genetics and evolutionary ecology: An introduction*. Macmillan, 1979.

P. Schuster and K. Sigmund. Replicator dynamics. *Journal of Theoretical Biology*, 100(3):533–538, 1983.

A. M. Siepielski, J. D. DiBattista, and S. M. Carlson. It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology letters*, 12(11):1261–1276, 2009.

A. M. Siepielski, K. M. Gotanda, M. B. Morrissey, S. E. Diamond, J. D. DiBattista, and S. M. Carlson. The spatial patterns of directional phenotypic selection. *Ecology letters*, 16(11):1382–1392, 2013.

J. R. Stinchcombe, Function-valued Traits Working Group, and M. Kirkpatrick. Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. *Trends in Ecology & Evolution*, 27(11):637–647, Nov. 2012.

P. D. Taylor and L. B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1-2):145–156, 1978.

J. N. Thompson. *The coevolutionary process*. University of Chicago Press, 1994.

J. N. Thompson. *The geographic mosaic of coevolution*. University of Chicago Press, 2005.

J. N. Thompson. *Interaction and Coevolution*. Interaction and Coevolution. University of Chicago Press, 2014.

H. Toju. Fine-scale local adaptation of weevil mouthpart length and camellia pericarp thickness: altitudinal gradient of a putative arms race. *Evolution*, 62(5):1086–1102, 2008.

H. Toju. Weevils and camellias in a darwin's race: model system for the study of eco-evolutionary interactions between species. *Ecological research*, 26(2):239–251, 2011.

H. Toju and T. Sota. Imbalance of predator and prey armament: geographic clines in phenotypic interface and natural selection. *The American Naturalist*, 167(1):105–117, 2005.

H. Toju and T. Sota. Adaptive divergence of scaling relationships mediates the arms race between a weevil and its host plant. *Biology Letters*, 2(4):539–542, July 2006.

H. Toju and T. Sota. Do arms races punctuate evolutionary stasis? unified insights from phylogeny, phylogeography and microevolutionary processes. *Molecular ecology*, 18(18):3940–3954, 2009.

H. Toju, H. Abe, S. Ueno, Y. Miyazawa, F. Taniguchi, T. Sota, and T. Yahara. Climatic gradients of arms race coevolution. *The American Naturalist*, 177(5):562–573, 2011a.

H. Toju, S. Ueno, F. Taniguchi, and T. Sota. Metapopulation structure of a seed–predator weevil and its host plant in arms race coevolution. *Evolution*, 65(6):1707–1722, 2011b.

M. Turelli. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology*, 25(2):138–193, 1984.

M. Turelli. Gaussian versus non-gaussian genetic analyses of polygenic mutation-selection balance. In *Evolutionary Processes and Theory*, pages 607–628. Academic Press, 1986.

M. Turelli. Commentary: Fisher's infinitesimal model: A story for the ages. *Theoretical Population Biology*, 118:46–49, Dec. 2017.

M. Turelli and N. H. Barton. Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics*, 138(3):913–941, 1994.

D. P. Vázquez and M. A. Aizen. Asymmetric specialization: A pervasive feature of plant–pollinator interactions. *Ecology*, 85(5):1251–1257, May 2004.

V. Volpert. *Elliptic Partial Differential Equations*. Springer Basel, 2014.

V. Volterra. Fluctuations in the abundance of a species considered mathematically1. *Nature*, 118(2972): 558–560, Oct. 1926.

A. R. Wallace. Creation by law. *Quarterly Journal of Science*, 4(16):470–488, 1867.

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

J. B. Walsh. An introduction to stochastic partial differential equations. In *Lecture Notes in Mathematics*, pages 265–439. Springer Berlin Heidelberg, 1986.

N. M. Waser, L. Chittka, M. V. Price, N. M. Williams, and J. Ollerton. Generalization in pollination systems, and why it matters. *Ecology*, 77(4):1043–1060, June 1996.

L. T. Wasserthal. The pollinators of the malagasy star OrchidsAngraecum sesquipedale, a. sororiumandA. compactumand the evolution of extremely long spurs by pollinator shift. *Botanica Acta*, 110 (5):343–359, Oct. 1997.

S. Watanabe. A limit theorem of branching processes and continuous state branching processes. *Journal of Mathematics of Kyoto University*, 8(1):141–167, 1968.

B. Week and S. L. Nuismer. The measurement of coevolution in the wild. *Ecology Letters*, 22(4):717–725, 2019.

S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931. ISSN 0016-6731.

N. Yamamura, S. Yachi, and M. Higashi. An ecosystem organization model explaining diversity at an ecosystem level: Coevolution of primary producer and decomposer. *Ecological Research*, 16(5): 975–982, 2001.

J. B. Yoder and S. L. Nuismer. When does coevolution promote diversification? *The American Naturalist*, 176(6):802–817, 2010.

S. Zheng. *Nonlinear evolution equations*. Chapman & Hall/CRC Press, Boca Raton, Fla, 2004.

Appendix A: Supplementary Material for *A White Noise Approach to Evolutionary Ecology*

Throughout this supplement, we set use dot notation for time derivatives so that $\dot{f}(x,t) = \frac{\partial}{\partial t} f(x,t)$ and denote by $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ the Laplace operator on $\mathbb{R}^d$.

A.1 Sufficient conditions for finite moments under DAGA

In this section we investigate the conditions under which the trait mean $\bar{x}(t)$, trait variance $\sigma^2(t)$ and abundance $N(t)$ remain finite for finite time $t \geq 0$ when they evolve according to DAGA, the deterministic PDE (1.4) of the main text.

Recall the growth rate expression $m(\nu, x)$ is shorthand for the more accurate expression $m((K\nu)(x,t), x)$ where $(K\nu)(x,t) = \int_{\mathbb{R}} \kappa(x-y)\nu(y,t)dy$ for some non-negative and bounded function κ. Hence, $K\nu$ is a non-negative function whenever ν is a non-negative function. In particular, this implies m is actually a bivariate function of two real numbers $h \geq 0$ and $x \in \mathbb{R}$. Following the main text, we assume the existence of $R \in \mathbb{R}$ such that $m(h,x) \leq R$ across all $h \geq 0$ and $x \in \mathbb{R}$. We also assume a twice continuously differentiable and integrable initial condition $u(x)$ that satisfies

$$0 < \int_{\mathbb{R}} (|x| + x^2) u(x) dx < +\infty. \tag{A.1}$$

In particular, this implies finite initial moments $N(0), |\bar{x}(0)|, \sigma^2(0) < +\infty$ and positive initial abundance and trait variance $0 < N(0), \sigma^2(0)$. Following DAGA, we consider the Cauchy problem

$$\begin{cases} \dot{\nu}(x,t) = m(\nu,x)\nu(x,t) + \frac{\mu}{2}\Delta\nu(x,t) & t > 0 \\ \nu(x,0) = u(x) & t = 0. \end{cases} \tag{A.2}$$

We assume the operator F defined by $\nu(x,t) \to m(\nu,x)\nu(x,t)$ is locally Lipschitz continuous, corresponding to equation (1.6) of the main text. To be specific, we define the domain of the Laplacian as $D(\Delta) = C^2(\mathbb{R}) \cap L^1(\mathbb{R})$ with the norm $\|v\| = \int_{\mathbb{R}} v(x)dx$ and define F as an operator on $D(\Delta)$. That is, F maps between functions that are integrable and twice continuously differentiable. Then Theorem 2.5.6 of Zheng (2004) implies for some maximal $T > 0$, the Cauchy problem (A.2) admits a unique classical solution $\nu(x,t)$ for $t \in [0,T)$. This implies the solution $\nu(x,t)$ is continuously differentiable with respect to t and twice continuously differentiable with respect to x for all $t \in [0,T)$. Furthermore, Theorem 2.5.6 of Zheng (2004) implies either $T = +\infty$ and $N(t) < +\infty$ for all $t > 0$ or $T < +\infty$ and $\lim_{t \uparrow T} N(t) = +\infty$. The latter case corresponds to the notion of *blow-up*.

In this section we show that our assumption $m(h,x) \leq R$ for all $h \geq 0$ and $x \in \mathbb{R}$ implies $T = +\infty$ and $N(t) < +\infty$ for all $t > 0$. Replacing m with it's upper bound $R \in \mathbb{R}$, PDE (A.2) reduces to a simple parabolic equation that can be solved using elementary techniques (Farlow, 1993). In particular, when $m(h,x) \equiv R = 0$ we denote the solution to (A.2) by $\nu_0(x,t)$. Then, denoting

$$\Phi(x,t) = \frac{\exp\left(-x^2/2\mu t\right)}{\sqrt{2\pi\mu t}}, \tag{A.3}$$

we have

$$\nu_0(x,t) = \int_{\mathbb{R}} \Phi(x-y,t)u(y)dy. \tag{A.4}$$

In the more general case, when $m(\nu,x) \equiv R \in \mathbb{R}$, equation (A.2) has the solution $\nu_R(x,t) = e^{Rt}\nu_0(x,t)$. Hence, $\nu_R(x,t) \geq 0$ for all $x \in \mathbb{R}$ and $\int_{\mathbb{R}} \nu_R(x,t)dx = e^{Rt}N(0) < +\infty$ for all $t \geq 0$. Furthermore, denoting

$$N_R(t) = \int_{\mathbb{R}} \nu_R(x,t)dx, \tag{A.5a}$$

$$p_R(x,t) = \nu_R(x,t)/N_R(t), \tag{A.5b}$$

$$\bar{x}_R(t) = \int_{\mathbb{R}} x p_R(x,t)dx, \tag{A.5c}$$

$$\sigma_R^2(t) = \int_{\mathbb{R}} (x-\bar{x}_R(t))^2 p_R(x,t)dx, \tag{A.5d}$$

we have

$$\bar{x}_R(t) = \int_{\mathbb{R}} x \int_{\mathbb{R}} \Phi(x-y,t)p_R(y,0)dydx = \int_{\mathbb{R}} y p_R(y,0)dy = \bar{x}(0), \tag{A.6}$$

$$\sigma_R^2(t) = \int_{\mathbb{R}} (x-\bar{x}_R(t))^2 \int_{\mathbb{R}} \Phi(x-y,t)p_R(y,0)dydx = \int_{\mathbb{R}} \Big((y-\bar{x}(0))^2 + \mu t\Big) p_R(y,0)dy = \sigma^2(0) + \mu t. \tag{A.7}$$

Hence, $|\bar{x}_R(t)|, \sigma_R^2(t) < +\infty$ for all $t \geq 0$. For the sake of contradiction, suppose there exists $x \in \mathbb{R}$ and $t > 0$ such that $\nu(x,t) > \nu_R(x,t)$. Then

$$\nu(x,t) - u(x) = \int_0^t m(\nu,x)\nu(x,s) + \frac{\mu}{2}\Delta\nu(x,s)ds > \int_0^t R\nu_R(x,s) + \frac{\mu}{2}\Delta\nu_R(x,s)ds = \nu_R(x,t) - u(x) \tag{A.8}$$

which implies there exists $h \geq 0$ and $x \in \mathbb{R}$ such that $m(h,x) > R$. But this contradicts our assumption $m(h,x) \leq R$ for all $h \geq 0$ and $x \in \mathbb{R}$. So we have $\nu(x,t) \leq \nu_R(x,t)$ for each $x \in \mathbb{R}$ and $t \geq 0$. This implies that, for all $t > 0$, $N(t) = \int_{\mathbb{R}} \nu(x,t)dx < +\infty$ and

$$0 \leq \int_{\mathbb{R}} x^2\nu(x,t)dx \leq \int_{\mathbb{R}} x^2\nu_R(x,t)dx < +\infty. \tag{A.9}$$

Furthermore, since $\nu(x,t)$ is a classical solution of Cauchy problem (A.2) and since we assumed $N(0) > 0$, we conclude $N(t) > 0$ for all finite $t > 0$. Hence, for each $t > 0$,

$$0 \leq \sigma^2(t) + \bar{x}^2(t) = \frac{1}{N(t)} \int_{\mathbb{R}} x^2\nu(x,t)dx < +\infty. \tag{A.10}$$

A.2 Equilibrium moments for a population experiencing logistic growth and stabilizing selection under DAGA

Here we show, under DAGA, the population moments $N, \bar{x}$ and σ^2 evolve to the following stable equilibrium

$$\hat{N} = \tfrac{1}{c}(R - \tfrac{1}{2}\sqrt{a\mu}), \tag{A.11a}$$

$$\hat{\bar{x}} = \theta, \tag{A.11b}$$

$$\hat{\sigma}^2 = \sqrt{\tfrac{\mu}{a}}, \tag{A.11c}$$

given the initial condition $N(0) > 0$ and growth rate

$$m(\nu, x) = R - \frac{a}{2}(\theta - x)^2 - c\int_{\mathbb{R}} \nu(y,t)dy = R - \frac{a}{2}(\theta - x)^2 - cN(t) \tag{A.12}$$

that satisfies $\theta \in \mathbb{R}$, $a, c, \mu > 0$ and $R > \frac{1}{2}\sqrt{\mu a}$. Following equation (A.12), mean fitness becomes

$$\bar{m}(t) = R - \frac{a}{2}\Big[(\theta - \bar{x}(t))^2 + \sigma^2(t)\Big] - cN(t), \tag{A.13}$$

and the ODE for $N(t)$ becomes

$$\frac{d}{dt}N(t) = \Big\{R - \frac{a}{2}\Big[(\theta - \bar{x}(t))^2 + \sigma^2(t)\Big] - cN(t)\Big\}N(t). \tag{A.14}$$

Solving for equilibrium total abundance $\hat{N}$ amounts to setting $\frac{d}{dt}N(t) = 0$ and solving for $N(t)$. Ignoring the equilibrium $N(t) = 0$, this reduces to solving $\bar{m}(t) = 0$ for $N(t)$, which, assuming finite equilibrial $\hat{\bar{x}}$ and $\hat{\sigma}^2$, returns

$$\hat{N} = \frac{1}{c}\left\{R - \frac{a}{2}\left[(\theta - \hat{\bar{x}})^2 - \hat{\sigma}^2\right]\right\}. \tag{A.15}$$

Unfortunately, deriving ODE for $\bar{x}(t)$ and $\sigma^2(t)$ leads to expressions involving higher moments and finding ODE for these higher moments will lead to expressions involving yet even higher moments. To avoid this infinite regression, we find the equilibrium abundance density $\hat{v}(x)$ by solving $\frac{\partial}{\partial t}\nu(x,t) = 0$ for $\nu(x,t)$. This implies the following ordinary differential equation

$$\frac{d^2}{dx^2}\hat{v}(x) = \left(\frac{2c}{\mu}\hat{N} + \frac{a}{\mu}(\theta - x)^2 - \frac{2R}{\mu}\right)\hat{v}(x) \tag{A.16}$$

which has the solution

$$\hat{v}(x) = \frac{\hat{N}}{\sqrt{2\pi}}\left(\frac{a}{\mu}\right)^{\frac{1}{4}} \exp\left(-\sqrt{\frac{a}{\mu}}\frac{(\theta - x)^2}{2}\right). \tag{A.17}$$

From this expression we infer $\hat{\bar{x}} = \theta$ and $\hat{\sigma}^2 = \sqrt{\frac{\mu}{a}}$. Hence $\hat{N} = \frac{1}{c}\left(R - \frac{1}{2}\sqrt{a\mu}\right)$. To show this equilibrium is stable, we use linear stability analysis. Since $\hat{v}(x)$ is Gaussian, we do not run into the

same issue with higher moments as above. Furthermore, following equations (1.31) of the main text, ODE for $\bar{x}(t)$ and $\sigma^2(t)$ can now be expressed as

$$\frac{d}{dt}\bar{x}(t) = \sigma^2(t)\left(\frac{\partial \bar{m}(t)}{\partial \bar{x}(t)} - \overline{\frac{\partial m(t)}{\partial \bar{x}(t)}}\right) = a\sigma^2(t)(\theta - \bar{x}(t)), \tag{A.18a}$$

$$\frac{d}{dt}\sigma^2(t) = 2\sigma^4(t)\left(\frac{\partial \bar{m}(t)}{\partial \sigma^2(t)} - \overline{\frac{\partial m(t)}{\partial \sigma^2(t)}}\right) + \mu = \mu - a\sigma^4(t). \tag{A.18b}$$

These expressions confirm our findings that $\hat{\bar{x}} = \theta$ and $\hat{\sigma}^2 = \sqrt{\frac{\mu}{a}}$. Furthermore, calculating

$$\frac{\partial}{\partial \sigma^2(t)}\frac{d}{dt}\sigma^2(t) = -2a\sigma^2(t) \tag{A.19}$$

and evaluating at $\sigma^2(t) = \hat{\sigma}^2$ demonstrates the equilibrium phenotypic variance is stable when $a, \mu > 0$. Hence, calculating

$$\frac{\partial}{\partial \bar{x}(t)}\frac{d}{dt}\bar{x}(t) = -a\sigma^2(t) \tag{A.20}$$

and evaluating at $\sigma^2(t) = \hat{\sigma}^2$ and $\bar{x}(t) = \hat{\bar{x}}$ demonstrates the equilibrium phenotypic mean is stable when $a, \mu > 0$. Finally, calculating

$$\frac{\partial}{\partial N(t)}\frac{d}{dt}N(t) = R - \frac{a}{2}\Big[(\theta - \bar{x}(t))^2 + \sigma^2(t)\Big] - 2cN(t) \tag{A.21}$$

and evaluating at $\sigma^2(t) = \hat{\sigma}^2$, $\bar{x}(t) = \hat{\bar{x}}$ and $N(t) = \hat{N}$ demonstrates the equilibrium total abundance is stable when $a, c, \mu > 0$, and $R > \frac{1}{2}\sqrt{a\mu}$.

A.3 Dynamics of σ^2 under DAGA

In this section we derive the dynamics of the phenotypic variance σ^2 under DAGA. Recall $\nu(x,t)$ is the abundance density in phenotypic space, $N(t) = \int_{\mathbb{R}} \nu(x,t)dx$ is the total abundance and $p(x,t) = \nu(x,t)/N(t)$ is the phenotypic distribution. Picking up from the main text §1.2.1, we have

$$\begin{aligned}
\frac{d}{dt}\sigma^2(t) = \frac{d}{dt}\int_{\mathbb{R}}(x-\bar{x}(t))^2 p(x,t)dx &= \int_{\mathbb{R}} 2(x-\bar{x}(t))\frac{d}{dt}\bar{x}(t) + (x-\bar{x}(t))^2\frac{\partial}{\partial t}p(x,t)dx \\
&= \int_{\mathbb{R}}(x-\bar{x}(t))^2\left((m(\nu,x) - \bar{m}(t))p(x,t) + \frac{\mu}{2}\frac{\partial^2}{\partial x^2}p(x,t)\right)dx \\
= \int_{\mathbb{R}}\Big((x-\bar{x}(t))^2 - \sigma^2(t) + \sigma^2(t)\Big)(m(\nu,x) - \bar{m}(t))p(x,t) &+ (x-\bar{x}(t))^2\frac{\mu}{2}\frac{\partial^2}{\partial x^2}p(x,t)dx \\
&= \mathrm{Cov}_t\Big((x-\bar{x}(t))^2, m(\nu,x)\Big) + \frac{\mu}{2}\int_{\mathbb{R}}(x-\bar{x}(t))^2\frac{\partial^2}{\partial x^2}p(x,t)dx,
\end{aligned} \tag{A.22}$$

where $\mathrm{Cov}_t\Big(f(x), g(x)\Big) = \int_{\mathbb{R}}\big(f(x) - \bar{f}\big)\big(g(x) - \bar{g}\big)p(x,t)dx$ is the covariance between f and g across the phenotypic distribution $p(x,t)$. Applying integration by parts twice yields

$$\int_{\mathbb{R}} (x - \bar{x}(t))^2 \frac{\partial^2}{\partial x^2} p(x,t) dx = 2. \tag{A.23}$$

Hence,

$$\frac{d}{dt}\sigma^2(t) = \mathrm{Cov}_t\Big((x - \bar{x}(t))^2, m(\nu, x)\Big) + \mu. \tag{A.24}$$

A.4 Simplifying fitness covariances when traits are normally distributed

Here we show when traits are normally distributed the evolutionary dynamics of $\bar{x}$ and σ^2 can be expressed in terms of growth rate gradients. In particular, we set

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left(-\frac{(x - \bar{x}(t))^2}{2\sigma^2(t)}\right) \tag{A.25}$$

and calculate

$$\begin{aligned}
\sigma^2\left(\frac{\partial \bar{m}}{\partial \bar{x}} - \overline{\frac{\partial m}{\partial \bar{x}}}\right) &= \sigma^2\left(\frac{\partial}{\partial \bar{x}}\int_{\mathbb{R}} m(\nu,x)p(x,t)dx - \int_{\mathbb{R}} p(x,t)\frac{\partial}{\partial \bar{x}} m(\nu,x)dx\right) \\
&= \sigma^2 \int_{\mathbb{R}} m(\nu,x)\frac{\partial}{\partial \bar{x}} p(x,t)dx = \sigma^2 \int_{\mathbb{R}} \frac{x - \bar{x}(t)}{\sigma^2} m(\nu,x)p(x,t)dx \\
&= \int_{\mathbb{R}} (x - \bar{x})(m(\nu,x) - \bar{m})p(x,t)dx = \mathrm{Cov}_t(m,x),
\end{aligned} \tag{A.26}$$

$$\begin{aligned}
2\sigma^4\left(\frac{\partial \bar{m}}{\partial \sigma^2} - \overline{\frac{\partial m}{\partial \sigma^2}}\right) &= 2\sigma^4\left(\frac{\partial}{\partial \sigma^2}\int_{\mathbb{R}} m(\nu,x)p(x,t)dx - \int_{\mathbb{R}} p(x,t)\frac{\partial}{\partial \sigma^2} m(\nu,x)dx\right) \\
&= 2\sigma^4 \int_{\mathbb{R}} \frac{(x - \bar{x})^2 - \sigma^2}{2\sigma^4} m(\nu,x)p(x,t)dx = \int_{\mathbb{R}} \Big((x - \bar{x})^2 - \sigma^2\Big)\big(m(\nu,x) - \bar{m}\big)p(x,t)dx \\
&= \mathrm{Cov}_t\Big(m, (x - \bar{x})^2\Big).
\end{aligned} \tag{A.27}$$

A.5 Relating fitness of expressed traits to fitness of breeding values

In the main text we extended our models of trait evolution to the case of imperfect inheritance. In this case, the expressed trait x of an individual is normally distributed around the individuals breeding value g. We denote by $\psi(x,g)$ the density of this normal distribution and by η its variance. We then apply either DAGA or SAGA to track the dynamics of the abundance density across breeding values, denoted $\rho(g,t)$, instead of tracking the dynamics of the the abundance density across expressed traits $\nu(x,t)$. We denote by G the variance of breeding values, traditionally referred to as the additive genetic variance, and by $m^*(\rho,g)$ the continuous time growth rate of breeding value g. To obtain $\nu(x,t)$ from $\rho(g,t)$, we calculate

$$\nu(x,t) = \int_{\mathbb{R}} \rho(g,t)\psi(x,g)dg. \tag{A.28}$$

Denoting by σ^2 the variance of expressed traits and $m(\nu,x)$ the continuous time growth rate of expressed trait x, equation (A.28) implies $\sigma^2 = G + \eta$ and

$$m^*(\rho,g) = \int_{\mathbb{R}} m(\nu,x)\psi(x,g)dx. \tag{A.29}$$

Furthermore, equation (A.28) implies expressed traits are normally distributed whenever the breeding values are. In this case both $\nu(x,t)$ and $\rho(g,t)$ share a common mean $\bar{x}$. Hence, the averaged fitness gradients $\overline{\frac{\partial m^*}{\partial \bar{x}}}$, $\overline{\frac{\partial m^*}{\partial G}}$ can be expressed as

$$\begin{aligned}\overline{\frac{\partial m^*}{\partial \bar{x}}} = \int_{\mathbb{R}} \frac{\rho(g,t)}{N(t)} \frac{\partial}{\partial \bar{x}} \left(\int_{\mathbb{R}} m(\nu,x)\psi(x,g)dx \right) dg = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{\rho(g,t)}{N(t)} \psi(x,g)dg \right) \frac{\partial}{\partial \bar{x}} m(\nu,x)dx \\ = \int_{\mathbb{R}} p(x,t) \frac{\partial}{\partial \bar{x}} m(\nu,x)dx = \overline{\frac{\partial m}{\partial \bar{x}}},\end{aligned} \tag{A.30}$$

$$\begin{aligned}\overline{\frac{\partial m^*}{\partial G}} = \int_{\mathbb{R}} \frac{\rho(g,t)}{N(t)} \frac{\partial}{\partial G} \left(\int_{\mathbb{R}} m(\nu,x)\psi(x,g)dx \right) dg = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{\rho(g,t)}{N(t)} \psi(x,g)dg \right) \frac{\partial}{\partial G} m(\nu,x)dx = \\ \int_{\mathbb{R}} p(x,t) \frac{\partial m}{\partial \sigma^2} \frac{\partial \sigma^2}{\partial G} dx = \overline{\frac{\partial m}{\partial \sigma^2}}.\end{aligned} \tag{A.31}$$

A.6 Simulating the rescaled process

Here we provide a detailed description of the branching random walk and how we have chosen to rescale it. To reduce the potential for extinction and to keep the population density concentrated near a particular trait value, we focus on the case of logistic growth and stabilizing selection as described in equation (A.12). In particular, we focus on a growth rate which, as a function of trait value $x \in \mathbb{R}$ and a measure-valued process X, can be written as

$$m(X(t),x) = R - \frac{a}{2}(\theta - x)^2 - cX(\mathbb{R},t). \tag{A.32}$$

We have implemented this simulation in the programming language Julia. A copy can be found at the url:

`https://github.com/bobweek/branching.brownian.motion.and.spde`

A.6.1 Description of simulation

We begin by describing the branching random walk before introducing our scheme to rescale it. Our branching random walk follows closely the description of branching Brownian motion in the main

text. However, we replace exponentially distributed lifetimes with deterministic unit time steps for easier implementation. Hence, we restrict time to $t = 0, 1, 2, \ldots$, and so on. Furthermore, we allow individual fitness to depend on both trait value and the state of the entire population. At time t we write $\{x_1(t), \ldots, x_{n(t)}(t)\}$ as the set of trait values across all $n(t)$ individuals alive in the population. Since our simulation follows discrete individuals instead of continuous distributions of trait values, we can write

$$X(t) = \sum_{i=1}^{n(t)} \delta_{x_i(t)}, \tag{A.33}$$

where $\delta_{x_i(t)}$ denotes the point mass located at $x_i(t)$. For simplicity we assume perfect heritability. At each iteration we draw, for each individual, a random number of offspring from a Negative-Binomial distribution. We use the Negative-Binomial distribution so that we can fix the variance in reproductive output while allowing the mean reproductive output to change. In particular, this coincides with our treatment of diffusion limits of interacting measure-valued processes in the main text.

Recall the Negative-Binomial distribution models the number of failed Bernoulli trials that occur before a given number of successful trials. Denoting q the probability of success for each trial and s the number of successes, the mean and variance are given respectively by

$$\mathfrak{W} = \frac{s(1-q)}{q}, \; V = \frac{s(1-q)}{q^2}. \tag{A.34}$$

This imposes the restriction $V > \mathfrak{W}$. Requiring the ith individual to have mean number offspring $\mathfrak{W}(X, x_i)$ and variance equal to V, the parameters of the associated Negative-Binomial distribution become

$$q(X, x_i) = \frac{\mathfrak{W}(X, x_i)}{V}, \; s(X, x_i) = \frac{\mathfrak{W}^2(X, x_i)}{V - \mathfrak{W}(X, x_i)}. \tag{A.35}$$

For each offspring produced by the individual with trait value x_i, we assign independently drawn trait values normally distributed around x_i with variance μ. This summarizes the basic structure of our simulation. To impose selection and density dependent growth rates, we set

$$\mathfrak{W}(X, x_i) = \exp\left(R - \frac{a}{2}(\theta - x_i)^2 - cX(\mathbb{R}, t)\right), \tag{A.36}$$

where $X(\mathbb{R}, t) = n(t)$.

A.6.2 Rescaling

To rescale the branching random walk by a positive integer n_0, we replace individual mass with $\frac{N_0}{n_0}$ for some fixed continuously varying number $N_0 > 0$, write the rescaled population distribution as $X^{(n_0)}$, rescale generation time by $1/n_0$ (which implies mutational variance is rescaled by $1/n_0$) and expected reproductive output by

$$\mathfrak{W}(X^{(n_0)}, x_i) \to \left(\mathfrak{W}(X^{(n_0)}, x_i)\right)^{1/n_0} = \exp\left(\frac{R}{n_0} - \frac{a}{2n_0}(\theta - x_i)^2 - \frac{c}{n_0^2}X^{(n_0)}(\mathbb{R}, \cdot)\right)$$
$$= \exp\left(\frac{R}{n_0} - \frac{a}{2n_0}(\theta - x_i)^2 - \frac{cn}{n_0^2}\right). \quad \text{(A.37)}$$

When it exists, we denote by $\mathfrak{X} = \lim_{n_0\to\infty} X^{(n_0)}$, the limiting process of the sequence of rescaled processes $X^{(1)}, X^{(2)}, \ldots$. Then, as desired, we find

$$\lim_{n_0\to\infty} n_0\left(\left(\mathfrak{W}(X^{(n_0)}(t), x)\right)^{1/n_0} - 1\right) = R - \frac{a}{2}(\theta - x)^2 - c\mathfrak{X}(\mathbb{R}, t). \quad \text{(A.38)}$$

Since the limiting growth rate is bounded above by R, the bivariate Girsanov transform given by Evans and Perkins (1994) can be used to demonstrate existence and uniqueness of $\mathfrak{X}$ (see also §7.5 of Etheridge, 2000).

A.7 Derivation of diffuse coevolution model

In this section we provide a derivation of our model of diffuse coevolution driven by resource competition. Since most of the work in this derivation has been completed in Appendix A.12, we focus here on deriving the Malthusian fitness of each species as a function of trait values and abundance densities of across all species in the community. We then use this fitness function to calculate selection gradients.

A.7.1 Individual fitness

We begin with discrete populations of individuals. In particular, we begin by assuming population size n_i is an integer for each species $i = 1, \ldots, S$ before passing to the large population size limit. We assume the competitive effects on fitness for each individual accumulates multiplicatively. For species i, the magnitudes of these negative effects increase with the degree of niche-overlap, mediated by the sensitivity $c_i > 0$.

We model niche space using the real line $\mathbb{R}$ and represent locations along this gradient with the symbol ζ. We assume individuals of species i sample the niche gradient following a probability distribution with density $u_i(\zeta, x)$, x being the average niche location sampled or niche center. In particular, we assume individuals sample their environment following a normal distribution so that

$$u_i(\zeta, x) = \frac{U_i}{\sqrt{2\pi w_i}} e^{-\frac{(\zeta - x)^2}{2w_i}}, \quad \text{(A.39)}$$

where U_i represents total niche use (since $U_i = \int u_i(\zeta, x)d\zeta$) and w_i represents niche breadth (the width of the bell curve u_i). Following the main text, we define the niche-overlap between individuals of species i and j with trait values x_i and x_j respectively as

$$\mathcal{O}_{ij}(x_i - x_j) = \int_{\mathbb{R}} u_i(\zeta, x_i) u_j(\zeta, x_j) d\zeta = \frac{U_i U_j}{\sqrt{2\pi(w_i + w_j)}} e^{-\frac{(x_i - x_j)^2}{2(w_i + w_j)}}. \tag{A.40}$$

Denote by x_{ij} the niche center of the j-th individual belonging to species i. The set of niche centers across all individuals in the community is written $\mathcal{C} = \{x_{ij}\}$. We denote by $\mathfrak{B}_{ij}$ a function that maps $\mathcal{C}$ to the cumulative effect of all competitive interactions on the fitness of the j-th individual in species i. Since individuals do not compete with themselves the net multiplicative effects on fitness of both interspecific and intraspecific competition on the j-th individual in species i can be summarized by

$$\mathfrak{B}_{ij}(\mathcal{C}) = \exp\left(-c_i \sum_{l \neq j} \mathcal{O}_{ii}(x_{ij} - x_{il}) - c_i \sum_{k \neq i} \sum_{l=1}^{n_k} \mathcal{O}_{ik}(x_{ij} - x_{kl}) \right). \tag{A.41}$$

To capture abiotic stabilizing selection we assume resources are normally distributed along the niche gradient. We also assume the concentration of resources is proportional to expected reproductive output. Combining these assumptions, we denote by $e_i(\zeta)$ the fitness benefits for individuals sampling at niche location ζ so that

$$e_i(\zeta) = Q_i e^{-\frac{A_i}{2}(\theta_i - \zeta)^2}, \tag{A.42}$$

where Q_i is the maximum expected reproductive output in the absence of competitive interactions, θ_i is the phenotypic optimum (location along niche axis of most abundant resources) and $A_i > 0$ determines the strength of abiotic stabilizing selection (the sharpness of the resource distribution). Then, we calculate the effect of mismatch between resource use and resource distribution on the fitness of individuals in species i with niche center x as

$$\mathfrak{A}_i(x) = \int_{\mathbb{R}} e_i(\zeta) u_i(\zeta, x) d\zeta = \frac{Q_i U_i}{\sqrt{1 + A_i w_i}} e^{-\frac{A_i}{1 + A_i w_i}(\theta_i - x)^2}. \tag{A.43}$$

Writing $\mathfrak{W}_{ij}(\mathcal{C})$ as the average number of offspring left by the j-th individual of species i, we have

$$\begin{aligned} \mathfrak{W}_{ij}(\mathcal{C}) &= \mathfrak{A}_i(x_{ij}) \mathfrak{B}_{ij}(\mathcal{C}) \\ &= \frac{Q_i U_i}{\sqrt{1 + A_i w_i}} \exp\left(-\frac{A_i}{1 + A_i w_i}(\theta_i - x)^2 - c_i \sum_{l \neq j} \mathcal{O}_{ii}(x_{ij} - x_{il}) - c_i \sum_{k \neq i} \sum_{l=1}^{n_k} \mathcal{O}_{ik}(x_{ij} - x_{kl}) \right). \end{aligned} \tag{A.44}$$

A.7.2 The diffusion limit

To make sense of the diffusion limit, we recall the components of the interacting measure-valued branching process discussed in the main text §1.2.2: (1) the branching rate λ, (2) the mean $\mathfrak{W}(X(t), x)$ and variance V of reproductive output and (3) spatial movement given by Brownian motion with diffusion parameter $\sqrt{\mu}$. For integers $n = 1, 2, \ldots$, we rescale the branching rate by $\lambda \to n$ and fitness by $\mathfrak{W}_{ij}(\mathcal{C}) \to \mathfrak{W}_{ij}^{1/n}(\mathcal{C})$. We rescale the mass of individuals in species i by $N_i(0)/n$ for a fixed positive continuously valued number $N_i(0) > 0$. In the diffusion limit, we take $n \to \infty$. We assume the

sequence of initial measures for species i, $X_i^{(1)}(0), X_i^{(2)}(0), \ldots$, converges to a limiting measure $\mathfrak{X}_i(0)$ that admits a density $\nu_i(x,0)$ (i.e., $\mathfrak{X}_i(D,0) = \int_D \nu_i(x,0)dx$) such that $\int_{\mathbb{R}}(|x| + x^2 + x^4)\nu_i(x,0)dx < +\infty$ and $\mathfrak{X}_i(\mathbb{R},0) = \int_{\mathbb{R}} \nu_i(x,0)dx = N_i(0) < +\infty$. Hence, rescaled fitness becomes

$$\mathfrak{W}_{ij}^{1/n}(\mathcal{C}) = \mathfrak{A}_i(x_{ij})^{1/n} \exp\left(-\frac{c_i}{n}\frac{N_i(0)}{n}\sum_{l\neq j}\mathcal{O}_{ii}(x_{ij}-x_{il}) - \frac{c_i}{n}\sum_{k\neq i}\frac{N_k(0)}{n}\sum_{l=1}^{n}\mathcal{O}_{1k}(x_{ij}-x_{kl})\right). \tag{A.45}$$

For large n, we have the approximation

$$\mathfrak{W}_{ij}^{1/n}(\mathcal{C}) \approx \mathfrak{A}_i(x_{ij})^{1/n}\left(1 - \frac{c_i}{n}\frac{N_i(0)}{n}\sum_{l\neq j}\mathcal{O}_{ii}(x_{ij}-x_{il}) - \frac{c_i}{n}\sum_{k\neq i}\frac{N_k(0)}{n}\sum_{l=1}^{n}\mathcal{O}_{ik}(x_{ij}-x_{kl})\right). \tag{A.46}$$

Then, writing $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_S)$, the Malthusian growth of an individual with trait value x_{ij} is

$$\begin{aligned}
m_i(\boldsymbol{\nu}, x_{ij}) &= \lim_{n\to\infty} n\left(\mathfrak{W}_{ij}^{1/n}(\mathcal{C}) - 1\right) \\
&= \lim_{n\to\infty} n\left(\mathfrak{A}_i(x_{ij})^{1/n} - 1\right) - c_i\mathfrak{A}_i(x_{ij})^{1/n}\left(\frac{N_i(0)}{n}\sum_{l\neq j}\mathcal{O}_{ii}(x_{ij}-x_{il}) + \sum_{k\neq i}\frac{N_k(0)}{n}\sum_{l=1}^{n}\mathcal{O}_{ik}(x_{ij}-x_{kl})\right) \\
&= \ln\mathfrak{A}_i(x_{ij}) - c_i\left(\int_{\mathbb{R}}\mathcal{O}_{ii}(x_{ij}-y)\nu_i(y,t)dy + \sum_{k\neq i}\int_{\mathbb{R}}\mathcal{O}_{ik}(x_{ij}-y)\nu_k(y,t)dy\right) \\
&= \ln\mathfrak{A}_i(x_{ij}) - c_i\left(\sum_{k=1}^{S}\int_{\mathbb{R}}\mathcal{O}_{ik}(x_{ij}-y)\nu_k(y,t)dy\right).
\end{aligned} \tag{A.47}$$

The resulting expression can be used to compute the Malthusian growth rate for species i associated with any trait value $x \in \mathbb{R}$, which we write as $m_i(\boldsymbol{\nu}, x)$.

A.7.3 Computing fitness gradients

We compute the average niche overlap of an individual in species i with niche location x across all individuals in species j as

$$\bar{\mathcal{O}}_{ij}(x,t) = \frac{\int_{\mathbb{R}}\mathcal{O}_{ij}(x-y)\nu_j(y,t)dy}{\int_{\mathbb{R}}\nu_j(y,t)dy} = \frac{1}{N_j(t)}\int_{\mathbb{R}}\mathcal{O}_{ij}(x-y)\nu_j(y,t)dy. \tag{A.48}$$

Following our assumption that individuals of species i sample their environment via a normal distribution with density $u_i(\zeta)$, we further assume normally distributed phenotypes for each of the S species. In this case $\bar{\mathcal{O}}_{ij}(x,t)$ simplifies to

$$\bar{\mathcal{O}}_{ij}(x,t) = \frac{1}{N_j(t)} \int_{\mathbb{R}} \mathcal{O}_{ij}(x-y)\nu_j(y,t)dy$$
$$= \frac{U_i U_j}{\sqrt{2\pi(w_i + w_j + {\sigma_j}^2(t))}} \exp\left(-\frac{(x-\bar{x}_j(t))^2}{2(w_i + w_j + {\sigma_j}^2(t))}\right), \quad \text{(A.49)}$$

where $\sigma_i^2(t)$ is the variance of niche-centers in species i at time t. Adopting the model of imperfect inheritance formulated in the main text we recall the expressed trait of an individual x_i is normally distributed around its breeding value g_i with variance η_i. We call η_i the variance of environmental deviation and G_i, which is the variance of breeding values, the additive genetic variance for species i. Under this model of inheritance variance of expressed traits decomposes as $\sigma_i^2(t) = G_i(t) + \eta_i$.

To simplify notation we set

$$R_i = \ln\left(\frac{Q_i U_i}{\sqrt{1 + A_i w_i}}\right), \quad \text{(A.50a)}$$
$$a_i = \frac{A_i}{1 + A_i w_i}, \quad \text{(A.50b)}$$
$$\tilde{b}_{ij}(t) = \frac{1}{w_i + w_j + \sigma_j^2(t)}, \quad \text{(A.50c)}$$

where R_i is the innate growth rate, a_i is the strength of abiotic stabilizing selection and $\tilde{b}_{ij}$ is an intermediate variable mediating the sensitivity of fitness of individuals in species i to interactions with species j. With this notation, the Malthusian fitness $m_i(\boldsymbol{\nu}, x)$ can be expressed as

$$m_i(\boldsymbol{\nu}, x) = R_i - \frac{a_i}{2}(x-\theta_i)^2 - c_i \sum_{j=1}^{S} N_j U_i U_j \sqrt{\frac{\tilde{b}_{ij}}{2\pi}} \exp\left(-\frac{\tilde{b}_{ij}}{2}(x-\bar{x}_j)^2\right). \quad \text{(A.51)}$$

For the remainder of the derivation we suppress notation indicating dependency on $\boldsymbol{\nu}$ and x. From (A.51) we calculate

$$\frac{\partial m_i}{\partial \bar{x}_i} = c_i N_i U_i^2 \tilde{b}_{ii}(x-\bar{x}_i)\sqrt{\frac{\tilde{b}_{ii}}{2\pi}} \exp\left(-\frac{\tilde{b}_{ii}}{2}(x-\bar{x}_i)^2\right), \quad \text{(A.52)}$$

$$\frac{\partial m_i}{\partial G_i} = \frac{c_i N_i U_i^2}{2}\left(\frac{(x-\bar{x}_i)^2 - G_i - \eta_i - 2w_i}{(G_i + \eta_i + 2w_i)^2}\right)\sqrt{\frac{\tilde{b}_{ii}}{2\pi}} \exp\left(-\frac{\tilde{b}_{ii}}{2}(x-\bar{x}_i)^2\right)$$
$$= \frac{c_i N_i U_i^2 \tilde{b}_{ii}^2}{2}\left((x-\bar{x}_i)^2 - {\sigma_i}^2 - 2w_i\right)\sqrt{\frac{\tilde{b}_{ii}}{2\pi}} \exp\left(-\frac{\tilde{b}_{ii}}{2}(x-\bar{x}_i)^2\right). \quad \text{(A.53)}$$

Note that

$$\sqrt{\frac{\tilde{b}_{ii}}{2\pi}}\exp\left(-\frac{\tilde{b}_{ii}}{2}(x-\bar{x}_i)^2\right)\sqrt{\frac{1}{2\pi{\sigma_i}^2}}\exp\left(-\frac{(x-\bar{x}_i)^2}{2{\sigma_i}^2}\right)$$
$$=\sqrt{\frac{1}{2\pi({\sigma_i}^2+1/\tilde{b}_{ii})}}\sqrt{\frac{{\sigma_i}^2+1/\tilde{b}_{ii}}{2\pi{\sigma_i}^2/\tilde{b}_{ii}}}\exp\left(-\frac{{\sigma_i}^2+1/\tilde{b}_{ii}}{2{\sigma_i}^2/\tilde{b}_{ii}}(x-\bar{x}_i)^2\right)$$
$$=\sqrt{\frac{1}{4\pi({\sigma_i}^2+w_i)}}\sqrt{\frac{2({\sigma_i}^2+w_i)}{2\pi{\sigma_i}^2({\sigma_i}^2+2w_i)}}\exp\left(-\frac{{\sigma_i}^2({\sigma_i}^2+2w_i)}{4({\sigma_i}^2+w_i)}(x-\bar{x}_i)^2\right). \quad \text{(A.54)}$$

Hence, gradients of the Malthusian fitness m_i averaged across the phenotypic distribution p_i become

$$\overline{\frac{\partial m_i}{\partial \bar{x}_i}}=0, \quad \text{(A.55)}$$

$$\overline{\frac{\partial m_i}{\partial G_i}}=\frac{c_i N_i U_i^2}{2({\sigma_i}^2+2w_i)^2}\left(\frac{({\sigma_i}^2+2w_i){\sigma_i}^2}{2(w_i+{\sigma_i}^2)}-{\sigma_i}^2-2w_i\right)\sqrt{\frac{b_{ii}}{2\pi}}$$
$$=\frac{c_i N_i U_i^2}{2({\sigma_i}^2+2w_i)}\left(\frac{{\sigma_i}^2}{2({\sigma_i}^2+w_i)}-1\right)\sqrt{\frac{b_{ii}}{2\pi}}=-\frac{c_i N_i U_i^2 b_{ii}}{2}\sqrt{\frac{b_{ii}}{2\pi}}, \quad \text{(A.56)}$$

where

$$b_{ij}=\frac{1}{w_i+w_j+{\sigma_i}^2+{\sigma_j}^2}. \quad \text{(A.57)}$$

The growth rate for species i is

$$\bar{m}_i=R_i-\frac{a_i}{2}\left((\bar{x}_i-\theta_i)^2+G_i+\eta_i\right)-c_i\sum_{j=1}^{S}N_j U_i U_j\sqrt{\frac{b_{ij}}{2\pi}}\exp\left(-\frac{b_{ij}}{2}(\bar{x}_i-\bar{x}_j)^2\right). \quad \text{(A.58)}$$

Thus, we find the following growth rate gradients

$$\frac{\partial\bar{m}_i}{\partial\bar{x}_i}=a_i(\theta_i-\bar{x}_i)-c_i\sum_{j=1}^{S}N_j U_i U_j b_{ij}(\bar{x}_j-\bar{x}_i)\sqrt{\frac{b_{ij}}{2\pi}}\exp\left(-\frac{b_{ij}}{2}(\bar{x}_i-\bar{x}_j)^2\right), \quad \text{(A.59)}$$

$$\frac{\partial\bar{m}_i}{\partial G_i}=-\frac{a_i}{2}+\frac{c_i}{2}\sum_{j=1}^{S}N_j U_i U_j b_{ij}\left(1-b_{ij}(\bar{x}_i-\bar{x}_j)^2\right)\sqrt{\frac{b_{ij}}{2\pi}}\exp\left(-\frac{b_{ij}}{2}(\bar{x}_i-\bar{x}_j)^2\right). \quad \text{(A.60)}$$

In particular, we will see

$$\left(\frac{\partial\bar{m}_i}{\partial G_i}-\overline{\frac{\partial m_i}{\partial G_i}}\right)=-\frac{a_i}{2}+\frac{c_i}{2}\left(N_i U_i^2 b_{ii}\sqrt{\frac{b_{ii}}{2\pi}}+\sum_{j=1}^{S}N_j U_i U_j b_{ij}\left(1-b_{ij}(\bar{x}_i-\bar{x}_j)^2\right)\sqrt{\frac{b_{ij}}{2\pi}}e^{-\frac{b_{ij}}{2}(\bar{x}_i-\bar{x}_j)^2}\right). \quad \text{(A.61)}$$

Applying equations (1.28a), (1.39a) and (1.39b) of the main text recovers system (1.46) of the main text.

A.8 Selection gradients

Here we derive expressions for selection gradients under our model of diffuse coevolution driven by resource competition.

A.8.1 Definition

Our definition of selection gradients differs slightly from traditional definitions. In particular, Lande and Arnold (1983) express the linear selection gradient β in general as

$$\beta = \frac{1}{\sigma^2}\mathrm{Cov}_t(\mathfrak{W}, x). \tag{A.62}$$

When expressed traits are normally distributed, Lande (1976) has shown this simplifies to

$$\beta = \left(\frac{\partial \ln \bar{\mathfrak{W}}}{\partial \bar{x}} - \overline{\frac{\partial \ln \mathfrak{W}}{\partial \bar{x}}}\right), \tag{A.63}$$

where $\bar{\mathfrak{W}}$ is individual fitness averaged across expressed trait values and $\overline{\frac{\partial \ln \mathfrak{W}}{\partial \bar{x}}}$ represents frequency dependent selection. This is convenient for discrete time models of mean trait evolution where the change in mean trait between generations is captured by

$$\Delta \bar{x} = \frac{G}{\sigma^2}\mathrm{Cov}_t(\mathfrak{W}, x) = G\beta. \tag{A.64}$$

However, in our case, we model mean trait evolution in continuous time via

$$\frac{d\bar{x}}{dt} = \frac{G}{\sigma^2}\mathrm{Cov}_t(m, x), \tag{A.65}$$

which in the case of normally distributed expressed traits simplifies to

$$\frac{d\bar{x}}{dt} = G\left(\frac{\partial \bar{m}}{\partial \bar{x}} - \overline{\frac{\partial m}{\partial \bar{x}}}\right). \tag{A.66}$$

Hence, we define the linear selection gradient β as

$$\beta := \frac{1}{\sigma^2}\mathrm{Cov}_t(m, x) \tag{A.67}$$

which, under normally distributed expressed traits, simplifies to

$$\beta = \left(\frac{\partial \bar{m}}{\partial \bar{x}} - \overline{\frac{\partial m}{\partial \bar{x}}}\right). \tag{A.68}$$

Similarly, the quadratic selection gradient γ is expressed in Lande and Arnold (1983) as

$$\gamma = \frac{1}{\sigma^4}\mathrm{Cov}_t(\mathfrak{W}, (x-\bar{x})^2). \tag{A.69}$$

Then, in analogy to our definition of β, we define γ by

$$\gamma := \frac{1}{\sigma^4}\mathrm{Cov}_t(m, (x-\bar{x})^2). \tag{A.70}$$

Following results derived in Appendix A.4, the case of normally distributed expressed traits simplifies γ to

$$\gamma = 2\left(\frac{\partial \bar{m}}{\partial G} - \overline{\frac{\partial m}{\partial G}}\right). \tag{A.71}$$

A.8.2 Selection gradients under abiotic stabilizing selection and resource competition

Combining our definitions of selection gradients with the results found in Appendix A.7, our model of diffuse coevolution yields, for species i,

$$\beta_i = a_i(\theta_i - \bar{x}_i) - c_i \sum_{j=1}^{S} N_j U_i U_j b_{ij}(\bar{x}_j - \bar{x}_i)\sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_j - \bar{x}_i)^2}, \tag{A.72a}$$

$$\gamma_i = -a_i + c_i\left(N_i U_i^2 b_{ii}\sqrt{\frac{b_{ii}}{2\pi}} + \sum_{j=1}^{S} N_j U_i U_j b_{ij}\left(1 - b_{ij}(\bar{x}_i - \bar{x}_j)^2\right)\sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i - \bar{x}_j)^2}\right). \tag{A.72b}$$

Note these selection gradients can be additively partitioned as $\beta_i = \beta_i^{(a)} + \sum_{j=1}^{S}\beta_{ij}$ and $\gamma_i = \gamma_i^{(a)} + \sum_{j=1}^{S}\gamma_{ij}$ where $\beta_i^{(a)}, \gamma_i^{(a)}$ denote the components due to abiotic stabilizing selection and β_{ij}, γ_{ij} denote the components due to interactions with species j. In particular, we find $\beta_i^{(a)} = a_i(\theta_i - \bar{x}_i)$, $\gamma_i^{(a)} = -a_i$ and

$$\beta_{ij} = c_i N_j U_i U_j b_{ij}(\bar{x}_i - \bar{x}_j)\sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i - \bar{x}_j)^2}, \tag{A.73a}$$

$$\gamma_{ij} = c_i N_j U_i U_j b_{ij}\left(1 - b_{ij}(\bar{x}_i - \bar{x}_j)^2\right)\sqrt{\frac{b_{ij}}{2\pi}} e^{-\frac{b_{ij}}{2}(\bar{x}_i - \bar{x}_j)^2},\ i \neq j \tag{A.73b}$$

$$\gamma_{ii} = 2c_i N_i U_i^2 b_{ii}\sqrt{\frac{b_{ii}}{2\pi}},\ i = j. \tag{A.73c}$$

A.9 The relation between competition coefficients and selection gradients

Here we derive covariances between competition coefficients and selection gradients following the model of diffuse coevolution derived in Appendix A.7. We assume the community is very rich (i.e., the number of species S is very large) and that the distribution of mean traits is approximately normal and independent of the distribution of abundance. We denote by $\bar{\bar{x}}$, $V_{\bar{x}}$ the community-wide mean

and variance of species mean traits and by $\bar{N}, V_N$ the community-wide mean and variance of species abundances. For simplicity we assume constant species trait variances and niche breadths so that $\sigma_i^2 = \sigma^2$ and $w_i = w$ for some $\sigma^2, w > 0$. Thus $b_{ij} = b = 1/(2\sigma^2 + 2w)$ for each $i, j = 1, \ldots, S$. Under these conditions, we can express competition coefficients, linear selection gradients and quadratic selection gradients respectively as

$$\alpha_{ij} = c_i U_i U_j \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2}(\bar{x}_i - \bar{x}_j)^2}, \tag{A.74a}$$

$$\beta_{ij} = c_i U_i U_j N_i b(\bar{x}_i - \bar{x}_j) \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2}(\bar{x}_i - \bar{x}_j)^2}, \tag{A.74b}$$

$$\gamma_{ij} = c_i U_i U_j N_i b(1 - b(\bar{x}_i - \bar{x}_j)^2) \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2}(\bar{x}_i - \bar{x}_j)^2}, \ i \neq j. \tag{A.74c}$$

To compute statistical distributions of these quantities we draw i and j independently from the set $\{1, \ldots, S\}$ each with probability $1/S$. Then the event $i = j$ occurs with probability $1/S^2$. We suppose S is large enough that we can safely ignore the event $i = j$.

Under our model of diffuse coevolution, the competition coefficients and selection gradients can be written in terms of the difference $D_{ij} = \bar{x}_i - \bar{x}_j$. By our assumption that i and j are drawn independently and that $\bar{x}_i$, $\bar{x}_j$ approximately follow a normal distribution with mean $\bar{\bar{x}}$ and variance $V_{\bar{x}}$, we see the distribution of D_{ij} is approximated by a normal distribution with mean zero and variance $2V_{\bar{x}}$.

We suppose the strengths of competition c_i and niche-use parameters U_i are distributed independently of mean traits, abundances and each other. We write $\bar{c}$, $\bar{U}$ and V_c, V_U as the mean and variance of these parameters respectively.

A.9.1 Means and variances of competition coefficients and selection gradients

Combining the above assumptions and notation, we can approximate the expectations of competition coefficients and selection gradients via

$$\bar{\alpha} = \frac{1}{S^2} \sum_{i,j=1}^{S} \alpha_{ij} \approx \bar{c}\bar{U}^2 \int_{\mathbb{R}} \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2}D^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD = \bar{c}\bar{U}^2 \sqrt{\frac{b}{2\pi(2V_{\bar{x}}b + 1)}}, \tag{A.75a}$$

$$\bar{\beta} = \frac{1}{S^2} \sum_{i,j=1}^{S} \beta_{ij} \approx \bar{c}\bar{U}^2 \bar{N} b \int_{\mathbb{R}} D \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2}D^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD = 0, \tag{A.75b}$$

$$\begin{aligned} \bar{\gamma} = \frac{1}{S^2} \sum_{i,j=1}^{S} \gamma_{ij} &\approx \bar{c}\bar{U}^2 \bar{N} b \int_{\mathbb{R}} (1 - bD^2) \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2}D^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD \\ &= \bar{c}\bar{U}^2 \bar{N} b \sqrt{\frac{b}{2\pi(2V_{\bar{x}}b + 1)}} \left(\frac{1}{2V_{\bar{x}}b + 1} \right) = \frac{\bar{\alpha}\bar{N}b}{2V_{\bar{x}}b + 1}. \end{aligned} \tag{A.75c}$$

Similarly, their variances can be approximated as

$$\begin{aligned}\mathrm{Var}(\alpha) = \overline{\alpha^2} - \bar{\alpha}^2 &\approx (V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 \int_{\mathbb{R}} \frac{b}{2\pi} e^{-bD^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD - \bar{\alpha}^2 \\ &= \frac{(V_c + \bar{c}^2)(V_U + \bar{U}^2)^2}{2} \sqrt{\frac{b}{\pi}} \sqrt{\frac{2b}{2\pi(4V_{\bar{x}}b + 1)}} - \bar{\alpha}^2 \\ &= \frac{(V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 b}{2\pi\sqrt{4V_{\bar{x}}b + 1}} - \bar{\alpha}^2, \quad \text{(A.76a)}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(\beta) = \overline{\beta^2} - \bar{\beta}^2 &\approx (V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 (V_N + \bar{N}^2) b \int_{\mathbb{R}} D^2 \frac{b}{2\pi} e^{-bD^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD \\ &= (V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 (V_N + \bar{N}^2) b \sqrt{\frac{b}{\pi}} \sqrt{\frac{2b}{2\pi(4V_{\bar{x}}b + 1)}} \left(\frac{V_{\bar{x}}}{4V_{\bar{x}}b + 1}\right) \\ &= \frac{(V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 (V_N + \bar{N}^2) b^2 V_{\bar{x}}}{\pi(4V_{\bar{x}}b + 1)^{3/2}}, \quad \text{(A.76b)}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(\gamma) = \overline{\gamma^2} - \bar{\gamma}^2 &\approx (V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 (V_N + \bar{N}^2) b \int_{\mathbb{R}} (1 - bD^2)^2 \frac{b}{2\pi} e^{-bD^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD - \bar{\gamma}^2 \\ = (V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 (V_N + \bar{N}^2) b \sqrt{\frac{b}{\pi}} \sqrt{\frac{2b}{2\pi(4V_{\bar{x}}b + 1)}} &\left(1 - 2b\left(\frac{V_{\bar{x}}}{4V_{\bar{x}}b + 1}\right) + 3b^2\left(\frac{V_{\bar{x}}}{4V_{\bar{x}}b + 1}\right)^2\right) - \bar{\gamma}^2 \\ = \frac{(V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 (V_N + \bar{N}^2) b^2}{\pi\sqrt{4V_{\bar{x}}b + 1}} &\left(1 - 2b\left(\frac{V_{\bar{x}}}{4V_{\bar{x}}b + 1}\right) + 3b^2\left(\frac{V_{\bar{x}}}{4V_{\bar{x}}b + 1}\right)^2\right) - \bar{\gamma}^2. \quad \text{(A.76c)}\end{aligned}$$

To accomplish these calculations, we used the fact that if

$$f_1(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(\mu_1 - x)^2}{2\sigma_1^2}\right), \; f_2(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(\mu_2 - x)^2}{2\sigma_2^2}\right), \quad \text{(A.77)}$$

then

$$f_1(x) f_2(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{(\mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left(-\frac{(\tilde{\mu} - x)^2}{2\tilde{\sigma}^2}\right), \quad \text{(A.78)}$$

where

$$\tilde{\mu} = \frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}, \; \tilde{\sigma}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \quad \text{(A.79)}$$

A.9.2 Mean and variance of absolute values of linear selection gradients

Since our above assumptions imply a certain degree of symmetry across the community, we find the average linear selection gradient is zero. Then, to extract information about the total quantity of linear selection occurring in the community, we consider the absolute values of linear selection gradients. Following the above assumptions we can express $|\beta_{ij}|$ as

$$|\beta_{ij}| = c_i U_i U_j N_i b |D_{ij}| \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2} D_{ij}^2}. \tag{A.80}$$

The mean of $|\beta_{ij}|$ can then be approximated as

$$\begin{aligned} \overline{|\beta|} &\approx \bar{c}\bar{U}^2 \bar{N} b \int_{\mathbb{R}} |D| \sqrt{\frac{b}{2\pi}} e^{-\frac{b}{2} D^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD \\ &= \bar{c}\bar{U}^2 \bar{N} b \sqrt{\frac{b}{2\pi(2V_{\bar{x}} b + 1)}} \int_{\mathbb{R}} |D| \sqrt{\frac{b}{2\pi(2V_{\bar{x}} b + 1)}} e^{-\frac{2V_{\bar{x}} b + 1}{2b} D^2} dD. \end{aligned} \tag{A.81}$$

Computing the integral on the RHS is equivalent to computing the mean of the absolute value of a normally distributed random variable with mean zero and variance $\frac{b}{2V_{\bar{x}} b + 1}$. It is well known that the absolute value $|Z|$ of a normally distributed random variable Z, itself taking mean zero and variance V_Z, has mean $\overline{|Z|} = \sqrt{2V_Z/\pi}$. Hence, we can use this information to compute

$$\overline{|\beta|} \approx \bar{c}\bar{U}^2 \bar{N} b \sqrt{\frac{b}{2\pi(2V_{\bar{x}} b + 1)}} \sqrt{\frac{2b}{\pi(V_{\bar{x}} b + 1)}} = \frac{\bar{c}\bar{U}^2 \bar{N} b^2}{\pi(2V_{\bar{x}} b + 1)}. \tag{A.82}$$

The variance $\mathrm{Var}(|\beta|)$ is a bit easier to calculate. In particular, we can approximate the variance of absolute values of β via

$$\mathrm{Var}(|\beta|) = \overline{|\beta|^2} - \overline{|\beta|}^2 = \overline{\beta^2} - \overline{|\beta|^2} \approx \mathrm{Var}(\beta) - \overline{|\beta|^2}, \tag{A.83}$$

where we have capitalized on the result $\bar{\beta} \approx 0$.

A.9.3 Correlations between competition coefficients and selection gradients

Following the above assumptions and notation, the covariance of competition coefficients α_{ij} and linear selection gradients β_{ij} can be approximated as

$$\mathrm{Cov}(\alpha, \beta) = \overline{\alpha\beta} - \bar{\alpha}\bar{\beta} \approx (V_c + \bar{c}^2)(V_U + \bar{U}^2)^2 \bar{N} b \int_{\mathbb{R}} D \frac{b}{2\pi} e^{-bD^2} \frac{1}{\sqrt{4\pi V_{\bar{x}}}} e^{-\frac{D^2}{4V_{\bar{x}}}} dD = 0. \tag{A.84}$$

Again, this result follows from our assumptions on the distribution of model parameters across the community. Instead, to extract information about the covariance between competition coefficients and the magnitude of linear selection, we compute $\mathrm{Cov}(\alpha, |\beta|)$. This quantity can be approximated by

$$\begin{aligned}\mathrm{Cov}(\alpha,|\beta|) = \overline{\alpha|\beta|} - \bar{\alpha}\overline{|\beta|} &\approx (V_c+\bar{c}^2)(V_U+\bar{U}^2)^2\bar{N}b\int_{\mathbb{R}}|D|\frac{b}{2\pi}e^{-bD^2}\frac{1}{\sqrt{4\pi V_{\bar{x}}}}e^{-\frac{D^2}{4V_{\bar{x}}}}dD - \bar{\alpha}\overline{|\beta|} \\ &= \frac{(V_c+\bar{c}^2)(V_U+\bar{U}^2)^2\bar{N}b^2}{\pi(4V_{\bar{x}}b+1)}\sqrt{\frac{V_{\bar{x}}}{\pi}} - \bar{\alpha}\overline{|\beta|}, \quad \text{(A.85)}\end{aligned}$$

where we have again made use of the properties of absolute values of normally distributed random variables.

The covariance between competition coefficients and quadratic selection gradients can be approximated by

$$\begin{aligned}\mathrm{Cov}(\alpha,\gamma) = \overline{\alpha\gamma} - \bar{\alpha}\bar{\gamma} &\approx (V_c+\bar{c}^2)(V_U+\bar{U}^2)^2\bar{N}b\int_{\mathbb{R}}(1-bD^2)\frac{b}{2\pi}e^{-bD^2}\frac{1}{\sqrt{4\pi V_{\bar{x}}}}e^{-\frac{D^2}{4V_{\bar{x}}}}dD - \bar{\alpha}\bar{\gamma} \\ &= \frac{1}{2}(V_c+\bar{c}^2)(V_U+\bar{U}^2)^2\bar{N}b\sqrt{\frac{b}{\pi}}\sqrt{\frac{2b}{2\pi(4V_{\bar{x}}b+1)}}\left(\frac{2V_{\bar{x}}b+1}{4V_{\bar{x}}b+1}\right) - \bar{\alpha}\bar{\gamma} \\ &= \frac{(V_c+\bar{c}^2)(V_U+\bar{U}^2)^2\bar{N}b^2(2V_{\bar{x}}b+1)}{2\pi(4V_{\bar{x}}b+1)^{3/2}} - \bar{\alpha}\bar{\gamma}. \quad \text{(A.86)}\end{aligned}$$

Finally, these approximations can be used to approximate the correlations between competition coefficients and selection gradients via

$$\mathrm{Corr}(\alpha,|\beta|) = \frac{\mathrm{Cov}(\alpha,|\beta|)}{\sqrt{\mathrm{Var}(\alpha)\mathrm{Var}(|\beta|)}}, \quad \text{(A.87a)}$$

$$\mathrm{Corr}(\alpha,\gamma) = \frac{\mathrm{Cov}(\alpha,\gamma)}{\sqrt{\mathrm{Var}(\alpha)\mathrm{Var}(\gamma)}}. \quad \text{(A.87b)}$$

A.10 Heuristics for the White Noise Calculus

Before introducing heuristics to operationalize white noise calculus, we begin with a brief conceptual introduction. One can think of white noise as the static seen on old television sets or infinitely detailed random dust spread across both time and space. From a more mathematical, yet still informal perspective, white noise can be thought of as a stochastic process. That is, we can picture white noise as a collection of random variables indexed by time and possibly space. In relation to Brownian motion, denoted by W, white noise can be interpreted of as the derivative of Brownian motion with respect to time, denoted $\dot{W}$. Since Brownian motion can be thought to take infinitesimally small Gaussian distributed jumps at each time point, this leads to the conceptualization of white noise as a collection of independent Gaussian distributed random variables. Figure A.1 illustrates realizations of this conceptualized white noise in one (left) and two (right) dimensions.

However, it turns out that realizations of white noise do not exist as functions in the classical sense. Indeed, since Brownian motion is nowhere differentiable with respect to time, white noise cannot be formally understood as a time derivative. Thus our notation $\dot{W}$ is only meant to aid intuition and not

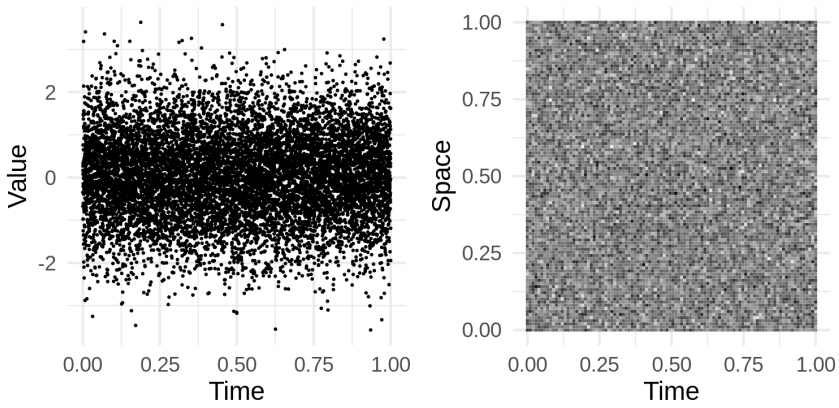

Figure A.1: Approximations of sample paths of temporal white noise (left) and space-time white noise (right) with brightness scaled to value.

be taken as formal. A formal understanding is possible by considering white noise as a generalized process that acts on classically defined stochastic processes to return other classically defined stochastic processes (Da Prato and Zabczyk, 2014; Krylov and Rozovskii, 1981).

Although the treatments of SPDE provided by Krylov and Rozovskii (1981) and Da Prato and Zabczyk (2014) extend the theory of SDE to formally treat SPDE in a general and elegant fashion, they require the navigation of an enormous amount of technical definitions and detailed proofs. To extract some particularly useful results from this theory relevant to our goal of synthesizing the stochastic dynamics of biological populations, we provide a streamlined approach by capitalizing on the solid ground these authors have established. For instance, instead of rigorously proving properties of white noise, we simply define them to be so, taking comfort in the fact that the technical details have been worked out elsewhere. In Appendix A.11 we show how our informal treatment is related to the rigorous treatment provided by Da Prato and Zabczyk (2014).

Before diving in, we shed a bit of light on the idea of a generalized process. A generalized process is analogous to a generalized function, such as the Dirac delta distribution δ_y, which places a point mass at position y. Often one sees δ_y defined as a function satisfying the properties $\delta_y(x) = 0$ for $x \neq y$ and $\int \delta_y(x)dx = 1$. However, since there is no function that satisfies these properties, we refer to δ_y as a generalized function. An alternative definition of δ_y considers its action on classically defined functions f. In particular, $\delta_y(f) = f(y)$, which can be heuristically written as $\int f(x)\delta_y(x)dx = f(y)$. Similarly, other generalized functions can be defined by their action on classically defined functions. Then, just as a generalized function operates on classical functions to return a value, a generalized process acts on a set of processes to return another stochastic process. For a brief primer on the theory of generalized functions, see the addendum to chapter 3 of Kolmogorov and Fomin (1999).

A.10.1 Definition and basic properties of space-time white noise

Throughout this section, we write $\int_{\mathbb{R}} f(x)dx$ for the integral of f over the whole real line and similarly $\int_D f(x)dx$ for the integral of f over $D \subset \mathbb{R}$. We define $\mathfrak{N}_2$ as the set of stochastic processes

$f(x,t)$ that are continuous in t and satisfy $\mathbb{E}\left(\int_0^t \int_{\mathbb{R}} |f(x,s)|^2 dxds\right) < +\infty$ for each $t \geq 0$. The operator $\mathbb{E}$ denotes expectation with respect to the underlying probability space. For each $t \geq 0$ we set

$$\|f\|_t = \sqrt{\mathbb{E}\left(\int_0^t \int_{\mathbb{R}} |f(x,s)|^2 dxds\right)}, \tag{A.88}$$

and make use of the convention $f = g$ if $\|f - g\|_t = 0$ for all $t \geq 0$. Following the main text, since the abundance density process $\nu(x,t)$ satisfying SAGA is continuous in t and integrable with respect to x for each $t \geq 0$, it also satisfies $\sqrt{\nu} \in \mathfrak{N}_2$. This enables us to utilize the heuristics developed in this section for the derivation of SDE describing the stochastic dynamics of $N(t)$, $\bar{x}(t)$ and $\sigma^2(t)$. To begin developing these heuristics, we introduce a generalized process that captures the essence of space-time white noise in a mathematically tractable format.

We define a generalized stochastic process $\mathbf{W}$ that maps processes $f \in \mathfrak{N}_2$ to real-valued stochastic processes indexed by time $t \geq 0$, but not by space. To evaluate $\mathbf{W}$ for a process $f \in \mathfrak{N}_2$ and some time $t \geq 0$ we write $\mathbf{W}_t(f)$. Specifically, for any $f, g \in \mathfrak{N}_2$, we define $\mathbf{W}(f)$ and $\mathbf{W}(g)$ to be Gaussian processes satisfying, for any $t, t_1, t_2 \geq 0$,

$$\mathbb{E}(\mathbf{W}_t(f)) = \mathbb{E}(\mathbf{W}_t(g)) = 0, \tag{A.89a}$$

$$\mathbb{C}(\mathbf{W}_{t_1}(f), \mathbf{W}_{t_2}(g)) = \mathbb{E}\left(\int_0^{t_1 \wedge t_2} \int_{\mathbb{R}} f(x,s) g(x,s) dxds\right), \tag{A.89b}$$

where $t_1 \wedge t_2 = \min(t_1, t_2)$ and $\mathbb{C}$ denotes covariance with respect to the underlying probability space. In particular, denoting $\mathbb{V}$ the variance operator with respect to the underlying probability space, we have $\mathbb{V}(\mathbf{W}_t(f)) = \|f\|_t^2$ for all $t \geq 0$ and $f \in \mathfrak{N}_2$.

The operators $\mathbb{E}$ and $\mathbb{C}$ are to be distinguished from expectations and covariances with respect to phenotypic diversity such as $\bar{x}(t)$ and $\mathrm{Cov}_t(m,x)$. In particular, since we model phenotypic diversity as a random process, the phenotypic moments $\bar{x}(t)$ and $\mathrm{Cov}_t(m,x)$ are random variables and $\mathbb{E}(\bar{x}(t))$, $\mathbb{E}(\mathrm{Cov}_t(m,x))$ denote the expectations of these random variables with respect to the underlying probability space.

Since Gaussian processes are characterized by their expectations and covariances and since we assume the $\mathfrak{N}_2$ processes are continuous in time, the processes $\mathbf{W}(f)$ and $\mathbf{W}(g)$ are well defined. As an example, if $f \in \mathfrak{N}_2$ is independent of time, then $\mathbf{W}(f)$ is a Brownian motion with variance at time $t \geq 0$ equal to $\|f\|_t^2 = t\, \mathbb{E}(\int_{\mathbb{R}} f^2(x,0) dx)$. With the generalized process $\mathbf{W}$ defined, we define the space-time white noise $\dot{W}(x,t)$ implicitly via the stochastic integral

$$\text{“} \int_0^t \int_{\mathbb{R}} f(x,s) \dot{W}(x,s) dxds\text{”} = \mathbf{W}_t(f), \ \forall\ f \in \mathfrak{N}_2,\ t \geq 0. \tag{A.90}$$

We place quotations in the above expression to emphasize its informal nature and that it should not be confused with classical Riemann integration. Following this definition of white noise, we compute its value by sampling it using $\mathfrak{N}_2$ processes. For example, integrating white noise over a region $D \times [0,t]$, with $t > 0$ and D a bounded subset of $\mathbb{R}$, is equivalent to evaluating $\mathbf{W}_t(I_D)$ for the deterministic process

$$I_D(x,t) = \begin{cases} 0, & x \notin D \\ 1, & x \in D \end{cases}. \tag{A.91}$$

Since

$$\|I_D\|_t^2 = \mathbb{E}\left(\int_0^t \int_{\mathbb{R}} I_D^2(x,s)dxds\right) = t\int_D dx = t|D| < +\infty, \tag{A.92}$$

where $|D|$ denotes the length of D, we have $I_D \in \mathfrak{N}_2$. Thus, using equations (A.89a) and (A.89b) and adopting the informal notation introduced in equation (A.90), we can write the following

$$\mathbb{E}\left(\int_0^t \int_D \dot{W}(x,s)dxds\right) = 0, \tag{A.93a}$$

$$\mathbb{V}\left(\int_0^t \int_D \dot{W}(x,s)dxds\right) = t|D|. \tag{A.93b}$$

Using this informal notation, equations (A.89a) and (A.89b) can be rewritten as

$$\mathbb{E}\left(\int_0^t \int_{\mathbb{R}} f(x,s)\dot{W}(x,s)dxds\right) = 0, \tag{A.94a}$$

$$\mathbb{C}\left(\int_0^{t_1} \int_{\mathbb{R}} f(x,s)\dot{W}(x,s)dxds, \int_0^{t_2} \int_{\mathbb{R}} g(x,s)\dot{W}(x,s)dxds\right) = \int_0^{t_1\wedge t_2} \int_{\mathbb{R}} f(x,s)g(x,s)dxds. \tag{A.94b}$$

To relate these formula to the common notation used for SDE, we write

$$\hat{f}(x,t) = \frac{f(x,t)}{\sqrt{\int_{\mathbb{R}} f^2(y,t)dy}} \text{ and } d\hat{\mathbf{W}}_t(f) = \left(\int_{\mathbb{R}} \hat{f}(x,t)\dot{W}(x,t)dx\right)dt \tag{A.95}$$

so that

$$\int_0^t d\hat{\mathbf{W}}_s(f) = \int_0^t \int_{\mathbb{R}} \frac{f(x,s)}{\sqrt{\int_{\mathbb{R}} f^2(s,y)dy}}\dot{W}(x,s)dxds. \tag{A.96}$$

This implies

$$\mathbb{E}\left(\int_0^t d\hat{\mathbf{W}}_s(f)\right) = 0,\ \mathbb{C}\left(\int_0^{t_1} d\hat{\mathbf{W}}_s(f), \int_0^{t_2} d\hat{\mathbf{W}}_s(f)\right) = t_1 \wedge t_2 \tag{A.97}$$

and in particular, as a function of t, $\int_0^t d\hat{\mathbf{W}}_s(f)$ is a standard Brownian motion for any $f \in \mathfrak{N}_2$. Hence, $d\hat{\mathbf{W}}_t(f)$ is analogous to the traditional shorthand used to denote stochastic differentials. Thus, equation (A.94b) effectively extends Itô's multiplication table to:

The extension of Itô's multiplication table and properties of white noise outlined in this subsection provide a useful set of tools for working with SPDE. In Appendix A.12 we employ these tools to derive SDE that track the dynamics of abundance, mean trait and phenotypic variance of a population from

Table A.1: An extension of Itô's multiplication table.

$\times$	$d\hat{\mathbf{W}}_t(f)$	$d\hat{\mathbf{W}}_t(g)$	dt
$d\hat{\mathbf{W}}_t(f)$	dt	$\left(\int_{\mathbb{R}} \hat{f}(x,t)\hat{g}(x,t)dx\right) dt$	0
$d\hat{\mathbf{W}}_t(g)$	$\left(\int_{\mathbb{R}} \hat{f}(x,t)\hat{g}(x,t)dx\right) dt$	dt	0
dt	0	0	0

a particular SPDE. In the following subsection, we review how this particular SPDE naturally arises as the diffusion limit of a measure-valued branching process (MVBP).

A.11 Comparing the white noise heuristics to Da Prato and Zabczyk (2014)

Our approach in the main text is inspired by the treatment provided in §4.2 of Da Prato and Zabczyk (2014). Here the authors develop a stochastic integral of operator-valued processes. In particular, they consider processes indexed by time $t \geq 0$ valued as Hilbert-Schmidt operators $\Phi(t)$ and define the norm

$$\|\Phi\|_t = \sqrt{\mathbb{E}\left(\int_0^t \mathrm{Tr}[\Phi(s)\Phi^*(s)]ds\right)},\ t \geq 0. \tag{A.98}$$

In our case we only consider the so-called multiplication operators. That is, processes that consist of operators $\Phi(t)$ having the form $\Phi(t)g(x) = \varphi(x,t)g(x)$ such that $\varphi(\cdot,t) \in L^2(\mathbb{R})$ a.s. for each $t \geq 0$. In this case $\Phi(t) = \Phi^*(t)$ and

$$\|\Phi\|_t = \|\varphi\|_t = \sqrt{\mathbb{E}\left(\int_0^t \int_{\mathbb{R}} \varphi^2(x,s)dxds\right)},\ t \geq 0. \tag{A.99}$$

Da Prato and Zabczyk (2014) form the space $\mathfrak{N}_W^2(0,T)$ of Hilbert-Schmidt operator-valued predictable processes $\Phi(t)$ that satisfy $\|\Phi\|_T < +\infty$ for some $T > 0$. This corresponds to our more specialized space $\mathfrak{N}_2$ that consists of $L^2(\mathbb{R})$ valued processes $\varphi(x,t)$ such that $\|\varphi\|_t < +\infty$ for all $t \geq 0$. In their treatment, $W(t)$ plays a similar role to our generalized process $\mathbf{W}_t$. For $\Phi \in \mathfrak{N}_W^2(0,T)$, they denote the stochastic integral for $t \in [0,T]$ by $\Phi \cdot W(t)$. Hence, for $\Phi(t)g(x) = \varphi(x,t)g(x)$ as above, $\mathbf{W}_t(\varphi) = \Phi \cdot W(t)$. The authors then prove the following:

Proposition 4.28 *Assume that* $\Phi_1, \Phi_2 \in \mathfrak{N}_W^2(0,T)$. *Then*

$$\mathbb{E}(\Phi_i \cdot W(t)) = 0,\ \ \mathbb{E}(\|\Phi_i \cdot W(t)\|^2) < +\infty,\ \ \forall\, t \in [0,T].$$

Corollory 4.29 *Under the same assumptions as Proposition 4.28,*

$$\mathbb{C}(\Phi_1 \cdot W(t), \Phi_2 \cdot W(s)) = \mathbb{E}\left(\int_0^{t\wedge s} \mathrm{Tr}[\Phi_2(r)\Phi_1^*(r)]dr\right), \quad \forall\, t,s \in [0,T].$$

Simplifying these expressions for the multiplication operators described above returns equations (A.94a) and (A.94b) above.

A.12 Derivation of SDE for $\bar{x}$ and σ^2

Here we derive the stochastic dynamics of $\bar{x}$ and σ^2 under SAGA by combining weak solutions of SPDE, an extension Itô's multiplication table summarized in Table 1 of the main text and Itô's quotient rule. This calculation requires the abundance density $\nu(x,t)$ to have finite first, second and fourth moments. Hence, we assume

$$\int_{\mathbb{R}} \nu(x,t)(|x| + x^2 + x^4)dx < +\infty. \tag{A.100}$$

Following §1.2.2 of the main text, we set

$$\tilde{x}(t) = \int_{\mathbb{R}} x\nu(x,t)dx, \quad \tilde{\tilde{\sigma}}^2(t) = \int_{\mathbb{R}} x^2\nu(x,t)dx. \tag{A.101}$$

Applying the weak solution of SAGA we obtain diffusion processes defined by

$$N(t) = N(0) + \int_0^t \int_{\mathbb{R}} \nu(x,s)m(\nu,x) + x\sqrt{V\nu(x,s)}\dot{W}(x,s)dxds, \tag{A.102a}$$

$$\tilde{x}(t) = \tilde{x}(0) + \int_0^t \int_{\mathbb{R}} \nu(x,s)m(\nu,x)x + x\sqrt{V\nu(x,s)}\dot{W}(x,s)dxds, \tag{A.102b}$$

$$\tilde{\tilde{\sigma}}^2(t) = \tilde{\tilde{\sigma}}^2(0) + \int_0^t \int_{\mathbb{R}} \nu(x,s)\left(m(\nu,x)x^2 + \mu\right) + x^2\sqrt{V\nu(x,s)}\dot{W}(x,s)dxds. \tag{A.102c}$$

In the following two sections we use Itô's quotient rule to derive expressions for the evolution of $\bar{x} = \tilde{x}/N$ and $\sigma^2 = \tilde{\tilde{\sigma}}^2 - \bar{x}^2$. Following these two sections we investigate stochastic dependencies between the processes N, $\bar{x}$ and σ^2.

A.12.1 Derivation for trait mean

We make use of the notation

$$\begin{cases} \|N\|_2 = \sqrt{V\int_{\mathbb{R}} \nu(x,t)dx} = \sqrt{VN} \\ \\ \|\tilde{x}\|_2 = \sqrt{V\int_{\mathbb{R}} x^2\nu(x,t)dx} \\ \\ \langle \tilde{x}, N\rangle = V\int_{\mathbb{R}} x\nu(x,t)dx = \bar{x}VN. \end{cases} \tag{A.103}$$

Rewriting formula (A.102b) as an SDE provides

$$d\tilde{x} = \left(\overline{xm}N + \frac{\mu}{2}\int_{\mathbb{R}} x\Delta\nu(x,t)dx\right)dt + \|\tilde{x}\|_2 d\tilde{W}_2, \tag{A.104}$$

where,

$$d\tilde{W}_2 = d\hat{\mathbf{W}}_t(\sqrt{Vx^2\nu}) = \frac{1}{\|\tilde{x}\|_2}\int_{\mathbb{R}} x\sqrt{V\nu(x,t)}\dot{W}(x,t)dxdt. \tag{A.105}$$

Using Itô's quotient rule on $\bar{x} = \tilde{x}/N$, we obtain

$$d\bar{x} = d\left(\frac{\tilde{x}}{N}\right) = \frac{\tilde{x}}{N}\left(\frac{d\tilde{x}}{\tilde{x}} - \frac{dN}{N} - \frac{d\tilde{x}}{\tilde{x}}\frac{dN}{N} + \left(\frac{dN}{N}\right)^2\right) = \frac{d\tilde{x}}{N} - \bar{x}\frac{dN}{N} - \frac{d\tilde{x}}{N}\frac{dN}{N} + \bar{x}\left(\frac{dN}{N}\right)^2. \tag{A.106}$$

From Table A.1 we have $d\tilde{x}dN = \langle\tilde{x}, N\rangle$ and $dN^2 = \|N\|_2^2$. Hence,

$$\begin{aligned} d\bar{x} &= \overline{xm}dt + \frac{\|\tilde{x}\|_2}{N}d\tilde{W}_2 - \bar{x}\left(\bar{m}dt + \sqrt{\frac{V}{N}}dW_1\right) - \frac{\langle\tilde{x}, N\rangle}{N^2}dt + \bar{x}\frac{\|N\|_2^2}{N^2}dt \\ &= (\overline{xm} - \bar{x}\bar{m})dt + \frac{\|\tilde{x}\|_2}{N}d\tilde{W}_2 - \bar{x}\sqrt{\frac{V}{N}}dW_1 - V\frac{\bar{x}}{N}dt + V\frac{\bar{x}}{N}dt \\ &= \mathrm{Cov}_t(x,m) + \frac{\|\tilde{x}\|_2}{N}d\tilde{W}_2 - \bar{x}\sqrt{\frac{V}{N}}dW_1. \end{aligned} \tag{A.107}$$

Note that

$$\begin{aligned} \frac{\|\tilde{x}\|_2}{N}d\tilde{W}_2 - \bar{x}\sqrt{\frac{V}{N}}dW_1 &= \frac{1}{N}\int_{\mathbb{R}} x\sqrt{V\nu(x,t)}\dot{W}(x,t)dx - \frac{\bar{x}}{N}\int_{\mathbb{R}}\sqrt{V\nu(x,t)}\dot{W}(x,t)dx \\ &= \int_{\mathbb{R}}\frac{x-\bar{x}}{N}\sqrt{V\nu(x,t)}\dot{W}(x,t)dx \end{aligned} \tag{A.108}$$

and

$$\mathbb{V}\left(\int_{\mathbb{R}}\frac{x-\bar{x}}{N}\sqrt{V\nu(x,t)}\dot{W}(x,t)dx\right) = \frac{V}{N}\int_{\mathbb{R}}(x-\bar{x})^2 p(x,t)dx = V\frac{\sigma^2}{N}. \tag{A.109}$$

Hence, by setting

$$dW_2 = \frac{\int_{\mathbb{R}}\frac{(x-\bar{x})}{N}\sqrt{V\nu(x,t)}\dot{W}(x,t)dx}{\sqrt{V\sigma^2/N}} \tag{A.110}$$

we can write

$$d\bar{x} = \mathrm{Cov}_t(x,m)dt + \sqrt{V\frac{\sigma^2}{N}}dW_2. \tag{A.111}$$

A.12.2 Derivation for trait variance

We make use of the notation

$$\begin{cases} \|\tilde{\bar{\sigma}}^2\|_2 = \sqrt{V \int_{\mathbb{R}} x^4 \nu(x,t) dx} \\ \\ \langle \tilde{\bar{\sigma}}^2, N \rangle = V \int_{\mathbb{R}} x^2 \nu(x,t) dx = \overline{x^2} V N. \end{cases} \tag{A.112}$$

Applying formula (A.102c) provides

$$d\tilde{\bar{\sigma}}^2 = \left(\overline{x^2 m} N + \mu N \right) dt + \|\tilde{\bar{\sigma}}^2\|_2 d\tilde{W}_3 \tag{A.113}$$

where

$$d\tilde{W}_3 = d\hat{\mathbf{W}}_t(\sqrt{V x^4 \nu}) = \frac{1}{\|\tilde{\bar{\sigma}}^2\|_2} \int_{\mathbb{R}} x^2 \sqrt{V \nu(x,t)} \dot{W}(x,t) dx. \tag{A.114}$$

Using Itô's quotient rule on $\overline{x^2} = \tilde{\bar{\sigma}}^2 / N$, we obtain

$$d\overline{x^2} = d\left(\frac{\tilde{\bar{\sigma}}^2}{N} \right) = \frac{\tilde{\bar{\sigma}}^2}{N} \left(\frac{d\tilde{\bar{\sigma}}^2}{\tilde{\bar{\sigma}}^2} - \frac{dN}{N} - \frac{d\tilde{\bar{\sigma}}^2}{\tilde{\bar{\sigma}}^2} \frac{dN}{N} + \left(\frac{dN}{N} \right)^2 \right) = \frac{d\tilde{\bar{\sigma}}^2}{N} - \overline{x^2} \frac{dN}{N} - \frac{d\tilde{\bar{\sigma}}^2}{N} \frac{dN}{N} + \overline{x^2} \left(\frac{dN}{N} \right)^2. \tag{A.115}$$

Table A.1 implies $d\tilde{W}_3 dW_1 = \langle \tilde{\bar{\sigma}}^2, N \rangle$ and hence

$$\begin{aligned} d\overline{x^2} &= \left(\overline{x^2 m} + \mu \right) dt + \frac{\|\tilde{\bar{\sigma}}^2\|_2}{N} d\tilde{W}_3 - \overline{x^2} \left(\bar{m} dt + \sqrt{\frac{V}{N}} dW_1 \right) - \frac{\langle \tilde{\bar{\sigma}}^2, N \rangle}{N^2} dt + \overline{x^2} \frac{\|N\|_2^2}{N^2} dt \\ \\ &= \left(\overline{x^2 m} - \overline{x^2} \bar{m} dt + \mu \right) dt + \frac{\|\tilde{\bar{\sigma}}^2\|_2}{N} d\tilde{W}_3 - \overline{x^2} \sqrt{\frac{V}{N}} dW_1 - \overline{x^2} \frac{V}{N} dt + \overline{x^2} \frac{V}{N} dt \\ \\ &= \left(\mathrm{Cov}_t\left(x^2, m \right) + \mu \right) dt + \frac{\|\tilde{\bar{\sigma}}^2\|_2}{N} d\tilde{W}_3 - \overline{x^2} \sqrt{\frac{V}{N}} dW_1. \end{aligned} \tag{A.116}$$

Setting $F(y,z) = y - z^2$, use Itô's formula on $\sigma^2 = F(\overline{x^2}, \bar{x}) = \overline{x^2} - \bar{x}^2$ to obtain:

$$d\sigma^2 = d\overline{x^2} - 2\bar{x}d\bar{x} - (d\bar{x})^2 = \left(\mathrm{Cov}_t\left(x^2, m\right) + \mu\right)dt + \frac{\|\tilde{\tilde{\sigma}}^2\|_2}{N}d\tilde{W}_3 - \overline{x^2}\sqrt{\frac{V}{N}}dW_1$$

$$- 2\bar{x}\left(\mathrm{Cov}_t(x,m) + \mu dt + \sqrt{\frac{V\sigma^2}{N}}dW_2\right) - \left(\mathrm{Cov}_t(x,m)dt + \mu dt + \sqrt{\frac{V\sigma^2}{N}}dW_2\right)^2$$

$$= \left(\mathrm{Cov}_t\left(x^2 - 2\bar{x}x, m\right) + \mu\right)dt + \frac{\|\tilde{\tilde{\sigma}}^2\|_2}{N}d\tilde{W}_3 - \overline{x^2}\sqrt{\frac{V}{N}}dW_1 - 2\bar{x}\sqrt{\frac{V\sigma^2}{N}}dW_2 - \left(\frac{V\sigma^2}{N}\right)dt$$

$$= \left(\mathrm{Cov}_t\left(x - \bar{x})^2, m\right) + \mu - \frac{V\sigma^2}{N}\right)dt + \frac{\|\tilde{\tilde{\sigma}}^2\|_2}{N}d\tilde{W}_3 - \overline{x^2}\sqrt{\frac{V}{N}}dW_1 - 2\bar{x}\sqrt{\frac{V\sigma^2}{N}}dW_2. \quad \text{(A.117)}$$

In light of

$$\frac{\|\tilde{\tilde{\sigma}}^2\|_2}{N}d\tilde{W}_3 - \overline{x^2}\sqrt{\frac{V}{N}}dW_1 - 2\bar{x}\sqrt{\frac{V\sigma^2}{N}}dW_2 = \frac{1}{N}\int_{\mathbb{R}}\left(x^2 - \tilde{\sigma}^2 - 2\bar{x}(x-\bar{x})\right)\sqrt{V\nu(x,t)}\dot{W}(x,t)dx$$
$$= \frac{1}{N}\int_{\mathbb{R}}\left((x-\bar{x})^2 - \sigma^2\right)\sqrt{V\nu(x,t)}\dot{W}(x,t)dx \quad \text{(A.118)}$$

and

$$\frac{1}{N}\int_{\mathbb{R}}\left(\left((x-\bar{x})^2 - \sigma^2\right)\sqrt{V\nu(x,s)}\right)^2 dx = \frac{V}{N}\left(\int_{\mathbb{R}}((x-\bar{x})^4 - 2(x-\bar{x})^2\sigma^2 + \sigma^4)p(x,t)dx\right)$$
$$= \frac{V}{N}\left(\overline{(x-\bar{x})^4} - \sigma^4\right) \quad \text{(A.119)}$$

we set

$$dW_3 = \frac{\int_{\mathbb{R}}\left((x-\bar{x})^2 - \sigma^2\right)\sqrt{V\nu(x,t)}\dot{W}(x,t)dx}{V\left(\overline{(x-\bar{x})^4} - \sigma^4\right)} \quad \text{(A.120)}$$

so that

$$d\sigma^2 = \mathrm{Cov}_t\left((x-\bar{x})^2, m\right)dt + \left(\mu - V\frac{\sigma^2}{N}\right)dt + \sqrt{V\frac{\overline{(x-\bar{x})^4} - \sigma^4}{N}}dW_3. \quad \text{(A.121)}$$

A.12.3 Stochastic dependencies between N, $\bar{x}$ and σ^2

Table A.1 implies

$$dW_1 dW_2 = \frac{\int_{\mathbb{R}} (x-\bar{x})\nu(x,t)dx}{\sqrt{N\sigma^2}}dt = 0, \tag{A.122a}$$

$$dW_1 dW_3 = \frac{\int_{\mathbb{R}} \left((x-\bar{x})^2-\sigma^2\right)\nu(x,t)dx}{\sqrt{\overline{(x-\bar{x})^4}-\sigma^4}}dt = 0, \tag{A.122b}$$

$$dW_2 dW_3 = \frac{\int_{\mathbb{R}} (x-\bar{x})\left((x-\bar{x})^2-\sigma^2\right)p(x,t)dx}{\sqrt{\sigma^2\left(\overline{(x-\bar{x})^4}-\sigma^4\right)}}dt = \frac{N\overline{(x-\bar{x})^3}}{\sqrt{\sigma^2\left(\overline{(x-\bar{x})^4}-\sigma^4\right)}}dt. \tag{A.122c}$$

Hence, stochastic fluctuations in the evolution of abundance N are independent of stochastic fluctuations in the evolution of $\bar{x}$ and σ^2. However, the stochastic fluctuations in the evolutions of $\bar{x}$ and σ^2 may be correlated. This is not the case when p is a Gaussian curve as equation (A.122c) would then imply $dW_2 dW_3 = 0$.

Appendix B: Supplementary Material for *Coevolutionary Arms Races and the Conditions for the Maintenance of Mutualisms*

B.1 Derivation of evolutionary dynamics from individual-based models

Our approach to deriving models of evolutionary dynamics follows three main steps. First we introduce the individual fitness functions for the trait-differences and offset-matching interaction mechanisms. Second, we derive the continuous-time growth rates of trait values using diffusion limits. Averaging these growth rates over phenotypic distributions returns population growth rates. Third, following Chapter 1, we apply SAGA, a continuous time stochastic model of trait dynamics that generalizes the well-known breeders equation (Lande, 1976) by accounting for demographic dynamics and arbitrary trait distributions. Assuming normally distributed traits, we use SAGA to derive mean trait dynamics in terms of growth rate gradients. We then assume infinitely large abundances to remove stochastic effects.

B.1.1 Individual fitness

Trait-matching For simplicity, we begin by formulating fitness under trait-matching. Letting x denote the trait value for the individual of species X and y for species Y, we can model the effect of trait-matching on fitness for individuals of species X by the concave down quadratic polynomial $-\frac{B_X}{2}(x-y)^2$. Since we assume both species interact via the same mechanism, the formula determining their dynamics will be of the same form. Hence, we focus on species X without losing any generality in our conclusions. We call the parameter $B_X \geq 0$ the strength of biotic selection (introduced in Week and Nuismer, 2019). Denote by $R_X > 0$ the number offspring produced by individuals in species X in the absence of ecological interactions and $E_X > 0$ the multiplicative effect on R_X due to the interspecific interaction when $B_X = 0$. We call E_X the intrinsic effect of the interspecific interaction. Suppose each individual of species X interacts with a single individual of species Y with trait value denoted by y. Then, under trait-matching, we model the number of offspring produced when $B_X > 0$ by

$$W_X(x,y) = R_X E_X e^{-\frac{B_X}{2}(x-y)^2}. \tag{B.1}$$

Note if $E_X e^{-\frac{B_X}{2}(x-y)^2} < 1$, the interaction is antagonistic and if $E_X e^{-\frac{B_X}{2}(x-y)^2} > 1$, species X benefits from the interaction. Following the same approach for species Y, we have

$$W_Y(y,x) = R_Y E_Y e^{-\frac{B_Y}{2}(x-y)^2}. \tag{B.2}$$

Trait-differences Following the approach taken to formulate individual fitness for the trait-matching mechanism, we replace the concave quadratic polynomial with the linear term $B_X(x-y)$, yielding

$$W_X(x,y) = R_X E_X e^{B_X(x-y)}, \tag{B.3}$$

$$W_Y(y,x) = R_Y E_Y e^{B_Y(y-x)}. \tag{B.4}$$

Under this model, the interaction is antagonistic if $E_X e^{B_X(x-y)} < 1$ or $E_Y e^{B_Y(y-x)} < 1$, and is a mutualism if $E_X e^{B_X(x-y)}, E_Y e^{B_Y(y-x)} > 1$.

Offset-matching Offset-matching is just a slight modification of trait-matching, summarized by

$$W_X(x,y) = R_X E_X e^{-\frac{B_X}{2}(y+\delta_X-x)^2}, \tag{B.5}$$

$$W_Y(y,x) = R_Y E_Y e^{-\frac{B_Y}{2}(x+\delta_Y-y)^2}. \tag{B.6}$$

Here we have assumed the optimal offset δ_X of species X is distinct from the optimal offset δ_Y of species Y. This seems more likely to occur in nature rather than identical offsets between the two species. However, we can always reduce such asymmetry via a change of variables. In particular, setting $\delta = \frac{\delta_X+\delta_Y}{2}$, $x' = x + \frac{\delta_Y}{2}$ and $y' = y + \frac{\delta_X}{2}$, we find

$$W_X(x,y) = R_X E_X e^{-\frac{B_X}{2}(y'-\frac{\delta_X}{2}+\delta_X-x'+\frac{\delta_Y}{2})^2}, \tag{B.7}$$

$$W_Y(y,x) = R_Y E_Y e^{-\frac{B_Y}{2}(x'-\frac{\delta}{2}+\delta_Y-y'+\frac{\delta_X}{2})^2}. \tag{B.8}$$

which simplifies to

$$W_X(x,y) = R_X E_X e^{-\frac{B_X}{2}(y'+\delta-x')^2}, \tag{B.9}$$

$$W_Y(y,x) = R_Y E_Y e^{-\frac{B_Y}{2}(x'+\delta-y')^2}. \tag{B.10}$$

Thus, from hereon we assume the common optimal offset δ for both species.

B.1.2 Deriving growth rates

To derive the dynamical equations of the main text starting with individual fitness we begin by taking a diffusion limit. Our approach follows a special case of the framework developed in Chapter 1, which we now summarize.

Outline of diffusion limits In the special case treated here, a diffusion limit implies that we begin with a population at the initial time $t = 0$ of n discrete individuals and rescale their "mass" by $\frac{N_0}{n}$ for some continuously positive number $N_0 > 0$. We then consider the limit $n \to \infty$. The total mass of the population, which is interpreted as the initial abundance $N(0)$, is computed as the sum of individual masses and hence remains equal to

$$N(0) = \sum_{i=1}^{n} \frac{N_0}{n} = N_0 \tag{B.11}$$

as we send $n \to \infty$. Once the diffusion limit is taken, we arrive at a model for the dynamics of the abundance $N(t)$. When taken appropriately, this diffusion limit results in a continuous-time, continuous-state stochastic process where the magnitude of stochasticity is modulated by the variance

in reproductive output V.

Starting with individual-based models of discrete individuals allows us to formulate mechanistic models of fitness and, by taking their diffusion limits, we formally derive models at the population level. In particular, these diffusion limits track the dynamics of abundance and phenotypic distribution instead of tracking the dynamics of a discrete set of individuals. Results found in Chapter 1 demonstrates these dynamics can be summarized by a set of stochastic differential equation tracking the dynamics of abundance and phenotypic moments. Under the assumption of normally distributed phenotypes, the population can be completely described by abundance $N(t)$, mean trait $\bar{x}(t)$ and trait variance $\sigma^2(t)$. Since the dynamics of abundance and phenotypic moments are interwoven via the expected reproductive output as a function of phenotype (ie., individual fitness) $W(x)$, the crux component for deriving the dynamics of $N(t)$, $\bar{x}(t)$ and $\sigma^2(t)$ is calculating a rescaled limit of $W(x)$. Although there are many ways to compute this limit, a useful approach is to set

$$m(x) = \lim_{n\to\infty} n\left(W(x)^{1/n} - 1\right). \tag{B.12}$$

When fitness is constant with respect to trait value (ie., $W(x) \equiv W$), then this limit converges to $m(x) = \ln W$ for each x. When this limit converges in general, $m(x)$ is the Malthusian (ie., continuous time) growth rate associated with trait value x. The average of $m(x)$ across all trait values in the population, denoted $\bar{m}$, is the growth rate of the population as a whole. Chapter 1 revealed, under the assumption of normally distributed phenotypes, the dynamics of mean trait can be derived by taking partial derivatives of $\bar{m}$ and $m(x)$ with respect to $\bar{x}$. While the partial derivative $\frac{\partial \bar{m}}{\partial \bar{x}}$ represents the effect of frequency independent selection on mean trait evolution, the partial derivative $\frac{\partial m(x)}{\partial \bar{x}}$ represents the effect of frequency dependent selection on mean trait evolution. In Chapter 1 this observation was extended by noting, again under normally distributed phenotypes, the action of selection on the evolution of phenotypic variance can be summarized by the partial derivatives $\frac{\partial \bar{m}}{\partial \sigma^2}$, $\frac{\partial m(x)}{\partial \sigma^2}$ which again represent frequency independent and frequency dependent selection respectively. The resulting stochastic differential equations describing the dynamics of $N(t)$ and the evolution of $\bar{x}(t)$ and $\sigma^2(t)$ are provided below in the section titled *Deriving evolution equations*.

Diffusion limits for trait-differences and offset-matching Following the main text, we isolate the effect of interaction mechanism by assuming each individual of species X interacts with a single individual of species Y and each individual of species Y interacts with a single individual of species X. To formally justify this assumption we require identical abundances for each species ($n_X = n_Y = n$). Although this will clearly never hold in the wild, this assumption acts as a mere stepping stone in our derivation as eventually we take $n_X, n_Y, n \to \infty$. In particular, for any continuously positive numbers $N_X, N_Y > 0$, we rescale individual mass by N_X/n_X for individuals in species X and by N_Y/n_Y for individuals in species Y. Hence, in the diffusion limit, the initial abundances for species X and Y are given by

$$\lim_{n_X\to\infty} \sum_{i=1}^{n_X} \frac{N_X}{n_X} = N_X \tag{B.13a}$$

$$\lim_{n_Y\to\infty}\sum_{i=1}^{n_Y}\frac{N_Y}{n_Y}=N_Y. \tag{B.13b}$$

In particular, since this holds for any N_X, N_Y, we can choose $N_X \neq N_Y$ in spite of setting $n_X = n_Y = n$. To picture why this is so, consider a specific case such as $n = 4$. Since we have exactly n individuals in each species, we can pair each individual of species X with a unique individual of species Y, forming n distinct pairs of interacting individuals. However, individuals in species X are weighted by N_X/n and individuals in species Y are weighted by N_Y/n and, since we can freely choose N_X and N_Y, these weights can be different. Hence, the "total mass" of species X at time $t = 0$, which is interpreted as the initial abundance of X, is N_X for any n. By the same argument, the initial abundance of species Y is N_Y for any n.

Following Chapter 1, we derive the growth rates for traits x and y respectively as

$$m_X(x,y)=\lim_{n\to\infty} n\left(W_X(x,y)^{1/n}-1\right), \tag{B.14a}$$

$$m_Y(y,x)=\lim_{n\to\infty} n\left(W_Y(y,x)^{1/n}-1\right). \tag{B.14b}$$

In the following, we apply these limits to derive the continuous time growth rates for species X and Y under trait-differences and offset-matching.

Trait-differences For the trait-differences mechanism, equation (B.14a) becomes

$$m_X(x,y)=\lim_{n\to\infty} n\left(\Big(W_X(x,y)\Big)^{1/n}-1\right)=\lim_{n\to\infty} n\left((R_XE_X)^{1/n}\exp\left(\frac{B_X}{n}(x-y)\right)-1\right). \tag{B.15}$$

For large n we have

$$\begin{aligned} &n\left((R_XE_X)^{1/n}\exp\left(\frac{B_X}{n}(x-y)\right)-1\right)\\ &\qquad\approx n\left((R_XE_X)^{1/n}\left(1+\frac{B_X}{n}(x-y)\right)-1\right)\\ &\qquad= n\big((R_XE_X)^{1/n}-1\big)+(R_XE_X)^{1/n}B_X(x-y). \end{aligned} \tag{B.16}$$

Since $n\big((R_XE_X)^{1/n}-1\big)\to\ln(R_XE_X)$ and $(R_XE_X)^{1/n}\to 1$, we find

$$m_X(x,y)=r_X+e_X+B_X(x-y), \tag{B.17}$$

where $r_X=\ln R_X$ and $e_X=\ln E_X$. Applying the same limit to equation (B.14b) returns

$$m_Y(y,x)=r_Y+e_Y+B_Y(y-x). \tag{B.18}$$

Recall that we assumed individuals interact at random with each other. This implies the growth rates $m_X(x,\cdot)$ and $m_Y(y,\cdot)$ are random variables with means

$$E_{m_X}(x) = r_X + e_X + B_X(x - \bar{y}), \tag{B.19a}$$

$$E_{m_Y}(y) = r_Y + e_Y + B_Y(y - \bar{x}), \tag{B.19b}$$

and variances

$$V_{m_X}(x) = B_X^2\sigma_Y^2, \tag{B.20a}$$

$$V_{m_Y}(y) = B_Y^2\sigma_X^2. \tag{B.20b}$$

However, by formally following our assumption that individuals interact with a single individual of the other species, averaging $m_X(x,y)$ across species X requires us to simultaneously average $m_X(x,y)$ across species Y as well and vice versa for average $m_Y(y,x)$ across species Y. Hence, we find the population growth rates

$$\bar{m}_X = r_X + e_X + B_X(\bar{x} - \bar{y}), \tag{B.21}$$

$$\bar{m}_Y = r_Y + e_Y + B_Y(\bar{y} - \bar{x}), \tag{B.22}$$

where $\bar{x}, \bar{y}$ are the mean traits for species X and Y respectively.

Offset-matching For the offset-matching mechanism, equation (B.14a) becomes

$$\begin{aligned} m_X(x,y) &= \lim_{n\to\infty} n\left(\Big(W_X(x,y)\Big)^{1/n} - 1\right) \\ &= \lim_{n\to\infty} n\left((R_XE_X)^{1/n}\exp\left(-\frac{B_X}{2n}(y+\delta-x)^2\right) - 1\right). \end{aligned} \tag{B.23}$$

For large n we have

$$\begin{aligned} &n\left((R_XE_X)^{1/n}\exp\left(-\frac{B_X}{2n}(y+\delta-x)^2\right) - 1\right) \\ &\approx n\left((R_XE_X)^{1/n}\left(1 - \frac{B_X}{2n}(y+\delta-x)^2\right) - 1\right) \\ &= n\big((R_XE_X)^{1/n} - 1\big) - (R_XE_X)^{1/n}\frac{B_X}{2}(y+\delta-x)^2. \end{aligned} \tag{B.24}$$

Hence,

$$m_X(x,y) = r_X + e_X - \frac{B_X}{2}(\delta + y - x)^2. \tag{B.25}$$

Applying the same limit to equation (B.14b) returns

$$m_Y(y,x) = r_Y + e_Y - \frac{B_Y}{2}(x + \delta - y)^2. \tag{B.26}$$

Then, following the same argument made in deriving growth rates under trait-differences, we find the population growth rates

$$\bar{m}_X = r_X + e_X - \frac{B_X}{2}(\bar{y} + \delta - \bar{x})^2 - \frac{B_X}{2}(\sigma_X^2 + \sigma_Y^2), \tag{B.27}$$

$$\bar{m}_Y = r_Y + e_Y - \frac{B_Y}{2}(\bar{x} + \delta - \bar{y})^2 - \frac{B_Y}{2}(\sigma_X^2 + \sigma_Y^2), \tag{B.28}$$

where σ_X^2, σ_Y^2 are the trait variances for species X and Y respectively.

B.1.3 Deriving evolution equations

With the growth rates for each mechanism derived, we apply a general formula for the dynamics of abundances N_X, N_Y, mean traits $\bar{x}, \bar{y}$ and additive genetic variances G_X, G_Y following the model SAGA introduced in Chapter 1. We also assume a simple model of inheritance where expressed traits are normally distributed around genotypic values (ie., breeding values). This implies $\sigma_X^2 = \eta_X + G_X$ and $\sigma_X^2 = \eta_X + G_X$ where η_X, η_Y capture developmental noise (Walsh and Lynch, 2018). In particular, assuming normally distributed trait values, normally distributed offspring breeding values, and Gaussian alleles, the dynamics of abundances, mean traits and additive genetic variances are given by the stochastic differential equations;

$$\frac{dN_X}{dt} = \bar{m}_X N_X + \sqrt{V_X N_X}\frac{dW_{N_X}}{dt}, \tag{B.29a}$$

$$\frac{dN_Y}{dt} = \bar{m}_Y N_Y + \sqrt{V_Y N_Y}\frac{dW_{N_Y}}{dt}, \tag{B.29b}$$

$$\frac{d\bar{x}}{dt} = G_X\left(\frac{\partial \bar{m}_X}{\partial \bar{x}} - \overline{\frac{\partial m_X}{\partial \bar{x}}}\right) + \sqrt{V_X\frac{G_X}{N_X}}\frac{dW_{\bar{x}}}{dt}, \tag{B.29c}$$

$$\frac{d\bar{y}}{dt} = G_Y\left(\frac{\partial \bar{m}_Y}{\partial \bar{y}} - \overline{\frac{\partial m_Y}{\partial \bar{y}}}\right) + \sqrt{V_Y\frac{G_Y}{N_Y}}\frac{dW_{\bar{y}}}{dt}, \tag{B.29d}$$

$$\frac{dG_X}{dt} = \left[2G_X^2\left(\frac{\partial \bar{m}_X}{\partial G_X} - \overline{\frac{\partial m_X}{\partial G_X}}\right) + \mu_X - V_X\frac{G_X}{N_X}\right] + G_X\sqrt{\frac{2V_X}{N_X}}\frac{dW_{G_X}}{dt}, \tag{B.29e}$$

$$\frac{dG_Y}{dt} = \left[2G_Y^2\left(\frac{\partial \bar{m}_Y}{\partial G_Y} - \overline{\frac{\partial m_Y}{\partial G_Y}}\right) + \mu_Y - V_Y\frac{G_Y}{N_Y}\right] + G_Y\sqrt{\frac{2V_Y}{N_Y}}\frac{dW_{G_Y}}{dt}. \tag{B.29f}$$

Here $V_X, V_Y \geq 0$ are variances in reproductive output, $\mu_X, \mu_Y \geq 0$ are rates of mutation, terms of the form $\frac{dW_Q}{dt}$ represent independent white noise processes driving the stochastic component of

variable Q and the terms $\overline{\frac{\partial m_X}{\partial Q}}$ and $\overline{\frac{\partial m_Y}{\partial Q}}$ are averages of the partial derivatives $\frac{\partial m_X}{\partial Q}$ and $\frac{\partial m_Y}{\partial Q}$ with respect trait distributions of species X and Y respectively. As noted above, these partial derivatives capture frequency dependent selection. With respect to both trait-differences and offset-matching mechanisms, we find

$$\frac{\partial m_X}{\partial \bar{x}} = \frac{\partial m_Y}{\partial \bar{y}} = \frac{\partial m_X}{\partial G_X} = \frac{\partial m_Y}{\partial G_Y} = 0. \tag{B.30}$$

Hence, frequency dependent selection is absent in our models of trait evolution. In contrast, frequency independent selection is summarized by;

$$\mathcal{D}\begin{cases} \frac{\partial \bar{m}_X}{\partial \bar{x}} = B_X, \\ \frac{\partial \bar{m}_Y}{\partial \bar{y}} = B_Y, \\ \frac{\partial \bar{m}_X}{\partial G_X} = 0, \\ \frac{\partial \bar{m}_Y}{\partial G_Y} = 0. \end{cases} \tag{B.31}$$

$$\mathcal{O}\begin{cases} \frac{\partial \bar{m}_X}{\partial \bar{x}} = B_X(\bar{y} + \delta - \bar{x}), \\ \frac{\partial \bar{m}_Y}{\partial \bar{y}} = B_Y(\bar{x} + \delta - \bar{y}), \\ \frac{\partial \bar{m}_X}{\partial G_X} = B_X/2, \\ \frac{\partial \bar{m}_Y}{\partial G_Y} = B_Y/2. \end{cases} \tag{B.32}$$

Taking the limits $N_X, N_Y \to \infty$ and incorporating the growth rate gradients above, equations (B.29) return the ordinary differential equations;

$$\mathcal{D}\begin{cases} \frac{d\bar{x}}{dt} = G_X B_X, \\ \frac{d\bar{y}}{dt} = G_Y B_Y, \\ \frac{dG_X}{dt} = \mu_X, \\ \frac{dG_Y}{dt} = \mu_Y, \end{cases} \tag{B.33}$$

$$\mathcal{O}\begin{cases} \frac{d\bar{x}}{dt} = G_X B_X(\bar{y} + \delta - \bar{x}), \\ \frac{d\bar{y}}{dt} = G_Y B_Y(\bar{x} + \delta - \bar{y}), \\ \frac{dG_X}{dt} = \mu_X - G_X^2 B_X, \\ \frac{dG_Y}{dt} = \mu_Y - G_Y^2 B_Y. \end{cases} \tag{B.34}$$

Then, to further simplify our models, for trait-differences we set $\mu_X = \mu_Y = 0$ and for offset-matching we set $G_X = \sqrt{\mu_X / B_X}$ and $G_Y = \sqrt{\mu_Y / B_Y}$ so that additive genetic variances remain fixed. This reduces each of our models to a pair of ordinary differential equations describing mean trait dynamics.

B.2 Weak selection approximations

Here we show how various interaction mechanisms can be related through weak selection and/or large offset approximations.

B.2.1 Exponential trait-differences approximates logistic trait-differences under weak selection

Under the logistic trait-differences mechanism traditionally employed (Nuismer et. al., 2006; Nuismer 2017), fitness for an individual of species X that engages in a single interaction is captured by

$$W_X(x,y) = \frac{R_X E_X}{1 + e^{2B_X(y-x)}}, \tag{B.35}$$

where, as above, R_X and E_X denote the intrinsic fitness and the intrinsic effect on fitness due to an interaction and the variable B_X is the strength of biotic selection. We take a weak selection approximation by Taylor expanding around $B_X \approx 0$. To leading order this returns

$$W_X(x,y) \approx \frac{R_X E_X}{2}\left(1 + B_X(x-y)\right). \tag{B.36}$$

Following the approach taken above, we calculate the continuous-time growth rate of individuals of species X with trait value x via $m_X(x,y) = \lim_{n\to\infty} n((W_X(x,y))^{1/n} - 1)$. This provides

$$m_X(x,y) = r_X + e_X + \ln\left(1 + B_X(x-y)\right), \tag{B.37}$$

where $r_X = \ln R_X$ and $e_X = \ln E_X - \ln 2$. Following through again with our weak selection approximation, we have

$$m_X(x,y) \approx r_X + e_X + B_X(x-y). \tag{B.38}$$

This approximation coincides with the growth rate we found using the exponential version of the trait-differences mechanism above. Hence, an exponential fitness curve provides an approximation for the logistic fitness curve when selection is weak.

B.2.2 Offset-matching approximates trait-differences under weak selection and a large offset

As mentioned in the main text, the offset-matching model is formally connected to the trait-differences model via a combined weak selection, large offset approximation. Recall individual fitness for species X under offset-matching is given by

$$W_X(x,y) = R_X E_X \exp\left(-\frac{B_X}{2}(y + \delta - x)^2\right), \tag{B.39}$$

Under the substitution $B_X := \varepsilon^3$ and $\delta := 1/\varepsilon$, equation (B.39) simplifies to

$$W_X(x,y) = R_X E_X \exp\left(-\varepsilon^3\frac{(y-x)^2}{2} - \varepsilon^2(y-x) + \frac{\varepsilon}{2}\right), \tag{B.40}$$

Hence, a second order Taylor expansion around $\varepsilon \approx 0$ leads to

$$W_X(x,y) \approx R_X E_X\left(1 + \varepsilon^2(x-y) + \frac{\varepsilon}{2}\right). \tag{B.41}$$

This approximation of fitness has the same form as the approximation of the logistic trait-differences mechanism derived above. Thus, the offset-matching mechanism produces an approximation of trait-differences for weak selection and large offset.

Appendix C: Supplementary Material for *The Measurement of Coevolution in the Wild*

Sections C.1 and C.2 derive the mathematical and statistical foundations for our approach respectively. Section C.3 evaluates the performance of our approach. Section C.4 describes the data and data analysis.

C.1 Derivation of the coevolutionary model

C.1.1 Selection

We assume the biotic effects on fitness act on the fecundity of an individual while the abiotic effects act on the viability of an individual with $W_{B,i}$ as the expected reproductive output of an individual of species i given that it survives to reproduce and $W_{A,i}$ being the probability of surviving until reproduction for an individual of species i. Specifically, assuming that abiotic selection is Gaussian with a constant phenotypic optimum θ_i, tolerance α_i and maximum $0 < p_i \leq 1$, we define the probability of a member of species i with trait z_i surviving abiotic selection by

$$W_{A,i}(z_i) \equiv p_i \exp\left(-\frac{(\theta_i - z_i)^2}{2\alpha_i}\right). \tag{C.1}$$

To model fecundity selection based on the interactions between species we assume the interactions are mediated by what we call an offset matching mechanism. In this model biotic fitness is maximized for an individual when its trait is offset from the trait of the individual it is interacting with by some constant δ. A simple example of an optimal offset comes from considering the interaction between long-tubed flowers and the long-proboscid flies that visit them. The biotic component of fitness for the fly is maximized when its proboscis is slightly longer than the nectar tube depth of the flower, allowing the fly to easily extract its nectar reward. The difference between tube depth and proboscis length that maximizes the flies biotic fitness component is the optimal offset for the fly. Note how this differs from a "bigger is better" situation since fitness with an optimal offset is unimodal and therefore does not increase indefinitely with larger (or lesser) trait values.

To capture this we assume biotic selection is also Gaussian but with γ_i as the tolerance and $z_j + \delta_i$ as the phenotypic optimum where z_j is the trait value of the individual being encountered and δ is the optimal offset in trait value that maximizes the biotic component of fitness. By assuming there is an upper bound on the potential number of offspring (which we will denote by c_i), we define the proportion of this upper bound achieved by an individual of species i given that it has encountered an individual of species j by

$$s_i(z_i, z_j) \equiv \exp\left(-\frac{(z_j + \delta - z_i)^2}{2\gamma_i}\right). \tag{C.2}$$

Next, we assume individuals encounter each other at random (with respect to their trait values) and that traits are normally distributed with densities denoted by $\phi_1(z_1)$ and $\phi_2(z_2)$. We denote the means of these distributions by $\bar{z}_1, \bar{z}_2$ and variances by σ_1^2, σ_2^2. Then for an individual in species i with trait z_i,

the expected proportion of the maximal reproductive output is defined as

$$b_i(z_i) \equiv \int_{\mathbb{R}} \phi_j(z_i) s(z_i, z_j) dz_j \tag{C.3}$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(\bar{z}_j - z_j)^2}{2\sigma_j^2}\right) \exp\left(-\frac{(z_j + \delta - z_i)^2}{2\gamma_i}\right) dz_j$$

$$= \sqrt{\frac{\gamma_i}{\gamma_i + \sigma_j^2}} \exp\left(-\frac{(\bar{z}_j + \delta - z_i)^2}{2(\gamma_i + \sigma_j^2)}\right).$$

Proving this statement is a matter of elucidating the algebra of Gaussian functions. For example, if $f_1(x)$ and $f_2(x)$ are Gaussian functions of x then $f_1(x)f_2(x)$ will also be a Gaussian function of x. Similarly, if $f_1(x)$ and $f_2(x)$ are the densities of two normal distributions then, $f_1(x)f_2(x)$ will be proportional to a density of another normal distribution. Specifically, we have

$$\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(\mu_1 - x)^2}{2\sigma_1^2}\right) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(\mu_2 - x)^2}{2\sigma_2^2}\right) \tag{C.4}$$

$$= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{(\mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left(-\frac{(\tilde{\mu} - x)^2}{2\tilde{\sigma}^2}\right)$$

where

$$\tilde{\mu} = \frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_2^2 + \sigma_1^2},\ \tilde{\sigma}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

To demonstrate this equality rigorously requires a lengthy calculation involving the Fourier transform and can be found elsewhere. We therefore omit it here. This result can be used to evaluate equation (C.3) and will be employed a great deal throughout the rest of the derivation.

Putting the pieces together, biotic fitness for an individual of species i with trait z_i can then be written as

$$W_{B,i}(z_i) = c_i b_i(z_i) = c_i \sqrt{\frac{\gamma_i}{\gamma_i + \sigma_j^2}} \exp\left(-\frac{(\bar{z}_j + \delta - z_i)^2}{2(\gamma_i + \sigma_j^2)}\right). \tag{C.5}$$

Hence the expected reproductive output of an individual with trait z_i in species i is expressed as

$$W_i(z_i) = W_{B,i}(z_i) W_{A,i}(z_i) = c_i p_i \sqrt{\frac{\gamma_i}{\gamma_i + \sigma_j^2}} \exp\left(-\frac{(\bar{z}_j + \delta - z_i)^2}{2(\gamma_i + \sigma_j^2)}\right) \exp\left(-\frac{(\theta_i - z_i)^2}{2\alpha_i}\right). \tag{C.6}$$

C.1.2 Phenotypic response to selection

Denoting the population mean fitness by $\overline{W}_i$ and the phenotypic distribution of species i by $\phi_i(z_i)$, we have

$$\overline{W}_i = \int_{\mathbb{R}} W_i(z_i) \phi_i(z_i) dz_i. \tag{C.7}$$

Using equation (C.6) from above we have

$$\overline{W}_i = K_i \exp\left(-\frac{\alpha_i(\bar{z}_j + \delta - \bar{z}_i)^2 + (\gamma_i + \sigma_j^2)(\theta_i - \bar{z}_i)^2}{2(\gamma_i\alpha_i + \gamma_i\sigma_j^2 + \gamma_i\sigma_i^2 + \alpha_i\sigma_j^2 + \alpha_i\sigma_i^2 + \sigma_i^2\sigma_j^2)}\right), \tag{C.8}$$

where K_i is independent of both $\bar{z}_i$ and $\bar{z}_j$. Once again, we have capitalized on two iterations of (C.4) to obtain (C.8). We can now employ our weak selection approximations by assuming both selective tolerances α_i and γ_i are sufficiently large so that $\gamma_i + \sigma_j^2 + \sigma_i^2 \approx \gamma_i$ and $\alpha_i + \sigma_j^2 + \sigma_i^2 \approx \alpha_i$. Mean fitness is then approximated as

$$\overline{W}_i \approx K_i \exp\left(-\frac{(\bar{z}_j + \delta - \bar{z}_i)^2}{2\gamma_i} - \frac{(\theta_i - \bar{z}_i)^2}{2\alpha_i}\right). \tag{C.9}$$

Hence, the breeders equation (sensu Lande, 1976) returns

$$\Delta\bar{z}_i = G_i\frac{\partial \ln \overline{W}_i}{\partial \bar{z}_i} \approx G_i\left(\gamma_i^{-1}(\bar{z}_j + \delta - \bar{z}_i) + \alpha^{-1}(\theta_i - \bar{z}_i)\right) \tag{C.10}$$

where G_i is the additive genetic variance for the trait in species i. For simplicity, we assume that the G_i are constant with respect to time for both species. We define the strength of abiotic selection as $A_i \equiv \alpha_i^{-1}$. Similarly we define the strength of biotic selection by $B_i \equiv \gamma_i^{-1}$. Hence, the response of the population mean phenotype to selection and reproduction is approximated by

$$\Delta\bar{z}_i \approx G_i\left(B_i\delta + B_i(\bar{z}_j - \bar{z}_i) + A_i(\theta_i - \bar{z}_i)\right). \tag{C.11}$$

C.1.3 Relating selection strengths to selection gradients

A selection gradient, commonly denoted by β, provides an alternative measure of selection. It occurs in the breeders equation given by $\Delta\bar{z} = G\beta$ (Lande and Arnold, 1983). Assuming that independent sources of selection can be decomposed (Ridenhour, 2005) so that $\beta = \beta_B + \beta_A$ represents the decomposition of selection into "biotic" and "abiotic" sources then, in terms of the offset matching model, we have

$$\beta_{B,i} = B_i\delta + B_i(\bar{z}_j - \bar{z}_i),\ \beta_{A,i} = A_i(\theta_i - \bar{z}_i). \tag{C.12}$$

C.1.4 Drift

Drift is modeled as the response to randomly sampling the offspring population. Denoting $Z_1, \ldots, Z_{n_i}$ as the breeding values of a random sample of individuals drawn from species i where n_i is the effective population size, we can calculate the random variable $\tilde{z}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} Z_k$. Assuming these breeding values are drawn independently of one another and that their distribution is normal with mean $\bar{z}_i$ and variance $h_i^2\sigma_i^2 = G_i$ (h_i^2 denotes heritability), we have $\tilde{z}_i$ is distributed normally with mean $\bar{z}_i$ and variance G_i/n_i. We denote the response to drift in species i by $\xi_i \equiv \tilde{z}_i - \bar{z}_i$.

In the case that n_i is different across N populations we must alter our treatment of drift. We denote by n_{ij} the effective population size of species i in population j and Z_{ijk} the trait of individual k in this

population. At first it may seem that replacing ξ_i with the average $\frac{1}{N}\sum_j \xi_{ij}$ would be appropriate. However this average reduces in variance as N grows. The effect of drift on local populations should not attenuate as sample size increases. We therefore instead take motivation from the Lindeberg-Lévy central limit theorem and define $\tilde{\xi}_i \equiv \frac{1}{\sqrt{N}}\sum_j \xi_{ij}$ as our "effective" response to drift. Under this definition we have

$$\mathrm{Var}(\tilde{\xi}_i) = \frac{1}{N}\sum_j \sum_k \mathrm{Var}(Z_{ikj}) = \frac{1}{N}\sum_j \frac{G_i}{n_{ij}} = G_i \frac{\sum_j n_{ij}^{-1}}{N}. \tag{C.13}$$

Hence, if there are multiple estimates of effective population size available from different locations (such as for our camellia-weevil example) we recommend using the harmonic mean of these values. Likewise, to calculate the effective additive genetic variance for our model when the additive genetic variance is known to be different across space we repeat the above approach to obtain

$$\mathrm{Var}(\tilde{\xi}_i) = \frac{1}{N}\sum_j \sum_k \mathrm{Var}(Z_{ikj}) = \frac{1}{N}\sum_j \frac{G_{ij}}{n_i} = \frac{1}{n_i}\frac{\sum_j G_{ij}}{N}. \tag{C.14}$$

Thus, we recommend using the arithmetic average of additive genetic variances when they vary across space.

Beyond the change in mean phenotype, we ignore the effects of drift on variance or any other property of the population. Then the total change in the mean trait of the population for species i after selection, reproduction and drift is

$$\Delta \bar{z}_i = G_i \left(B_i \delta + B_i(\bar{z}_j - \bar{z}_i) + A_i(\theta_i - \bar{z}_i) \right) + \xi_i. \tag{C.15}$$

C.1.5 Metapopulation dynamics

We denote by the random vector $(\bar{Z}_1(t), \bar{Z}_2(t))$ the pair of trait means for each species within a given (but unspecified) population at time step t. Motivated by the result for local mean trait dynamics, we define the next generation mean trait pair by

$$\bar{Z}_i(t+1) \equiv \bar{Z}_i(t) + G_i \left[B_i \delta + B_i(\bar{Z}_j(t) - \bar{Z}_i(t)) + A_i(\theta_i - \bar{Z}_i(t)) \right] + \xi_i(t). \tag{C.16}$$

This defines a two dimensional random process. We describe the statistical behavior of this process by focusing on its distribution. In general, to fully describe the distribution of a random process we need to write down all of the statistical moments at each time step. Of course this is not feasible, but, as shown in Appendix C.1.6 below, for our model we only need to calculate the first five moments. The first two are the average values of the two local mean traits (μ_1, μ_2) across the entire metapopulation. The third and fourth moments are the variances of the local mean traits among the metapopulation (V_1, V_2), which can take on positive values in our model solely due to drift. Finally, the fifth moment is the spatial covariance of the two mean traits (C) which can be non-zero in our model due to either $B_1 \neq 0$, $B_2 \neq 0$ or both $B_1, B_2 \neq 0$. Then to construct the metapopulation model we identify $\mu_i(t) \equiv \mathbb{E}\bar{Z}_i(t)$, $V_i(t) \equiv \mathbb{E}[(\mu_i(t) - \bar{Z}_i(t))^2]$, and $C(t) \equiv \mathbb{E}[(\mu_1(t) - \bar{Z}_1(t))(\mu_2(t) - \bar{Z}_2(t))]$. The simplicity of the local dynamics makes it a straightforward calculation to show that this definition coincides with the

recursions

$$\mu_i' = \mu_i + G_i\left[B_i\delta + B_i(\mu_j - \mu_i) + A_i(\theta_i - \mu_i)\right] \tag{C.17a}$$

$$V_i' = (1 - A_iG_i)^2V_i + 2(1 - A_iG_i)B_jG_j(C - V_i) + B_i^2G_i^2(V_i + V_j - 2C) + G_i/n_i \tag{C.17b}$$

$$C' = (1 - A_1G_1)(1 - A_2G_2)C + (1 - A_1G_1)B_2G_2(V_1 - C) + (1 - A_2G_2)B_1G_1(V_2 - C) \tag{C.17c}$$

$$+B_1B_2G_1G_2(2C - V_1 - V_2).$$

Taking this approach makes the implicit assumption of infinitely many localities. This is due to the fact that the moments are calculated are over all possible sample paths. Each sample path is associated with some theoretical location. However, in the real world metapopulations are comprised of just a finite number of localities. Then, in view of our model, they provide just a subset of all the possibilities. The true metapopulation moments of those empirical systems would then be looked at as sample moments with respect to the theoretical metapopulation moments predicted by our model. These observations dove-tail nicely with the fact that true empirical metapopulation moments exhibit some stochastic behavior while the metapopulation moments of our model evolve deterministically.

Finally, we implement simplifications of the expressions (C.17) by assuming each B_i and A_i are of some small order $\epsilon \ll 1$ and neglect all terms of order ϵ^2 and higher.

C.1.6 Proof of normality

Here we demonstrate that as long as selection is weak, the distribution describing population mean phenotypes across the metapopulation converges to a bivariate normal distribution regardless of the initial distribution of the metapopulation. Our model is 2-dimensional, but the proof is easily obtained for the more general d-dimensional situation. Using $\|x\| = \sum_{i=1}^{d}|x_i|$ for $x \in \mathbb{R}^d$ and $\|U\| = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } U\}$ for matrices, we have the following:

Claim: If $X(t)$ is a d-dimensional random process defined by $X(t+1) = UX(t) + W(t)$ where U is a $d \times d$ nonsingular matrix with real entries, $\|U\| < 1$ and the sequence $W(t)$ is comprised of iid d-dimensional multivariate normal variables, then $X(t)$ will converge in distribution to a multivariate normal random variable regardless of the distribution of $X(0)$.

Proof: For clarity we move the time index t to the subscript. The solution is immediately obtained as $X_t = U^tX_0 + \sum_{k=0}^{t-1}U^kW_k$. The hypothesis implies $\lim_{t\to\infty}U^t = 0$, the $d \times d$ zero matrix. So, we have $X^* \equiv \lim_{t\to\infty}X_t = \sum_{k=0}^{\infty}U^kW_k$. Note that $\|U^k\| \le \|U\|^k$ for each $k = 0, 1, \ldots$, that $\mathbb{E}\|W_i\| = \mathbb{E}\|W_j\| < \infty$ for each $i, j = 0, 1, \ldots$, and that $\sum_{k=0}^{\infty}\|U\|^k = (1 - \|U\|)^{-1} < \infty$. Hence, using Tonelli's Theorem to reverse the order of the expectation and summation, we have

$$\mathbb{E}\|X^*\| = \mathbb{E}\Big\|\sum_{k=0}^{\infty}U^kW_k\Big\| \le \mathbb{E}\left(\sum_{k=0}^{\infty}\|U^k\|\|W_k\|\right)$$

$$= \mathbb{E}\|W_0\|\sum_{k=0}^{\infty}\|U^k\| \le \mathbb{E}\|W_0\|\sum_{k=0}^{\infty}\|U\|^k < \infty.$$

This demonstrates that the mean vector of X^* has finite magnitude. Denote by $\mathrm{Var}(X)$ the variance covariance matrix of a d-dimensional random vector X. Then, by the iid assumption on the W_t we have

$$\|\mathrm{Var}(X^*)\| = \Big\| \sum_{k=0}^{\infty} \mathrm{Var}(U^k W_k) \Big\| = \Big\| \sum_{k=0}^{\infty} U^k \mathrm{Var}(W_0)(U^k)^\top \Big\| \leq \sum_{k=0}^{\infty} \|U^k\| \|\mathrm{Var}(W_0)\| \|U^k\| = \star.$$

By hypothesis $\|U\| < 1$ which implies $\|U^k\|^2 \leq \|U\|^k$ for each $k = 1, 2, \ldots$. Hence,

$$\star = \|\mathrm{Var}(W_0)\| \sum_{k=0}^{\infty} \|U^k\|^2 \leq \|\mathrm{Var}(W_0)\| \sum_{k=0}^{\infty} \|U\|^k < \infty.$$

This shows that $\mathrm{Var}(X^*)$ has finite entries. The higher cumulants of X^* will be zero since the W_t contribute nothing more than second order cumulants, which is characteristic of the multivariate normal distribution (Bryc, 1995; Lukacs, 1970). Hence, X^* is a d-dimensional multivariate normal random variable. □

In the notation used in the above proof, our model of coevolution corresponds to

$$U = \begin{pmatrix} 1 - G_1(B_1 + A_1) & G_1 B_1 \\ G_2 B_2 & 1 - G_2(B_2 + A_2) \end{pmatrix}. \tag{C.18}$$

The eigenvalues λ of the above matrix satisfy

$$\begin{aligned} |\lambda| \leq \frac{1}{2} |2 - G_1(A_1 + B_1) - G_2(A_2 + B_2)| + \\ \frac{1}{2} \sqrt{4 G_1 G_2 B_1 B_2 + [G_1(A_1 + B_1) - G_2(A_2 + B_2)]^2} \end{aligned} \tag{C.19}$$

By construction of our model we assume $A_i > 0$. If we futher impose $0 < G_i(B_i + A_i) < 1$ it can be demonstrated with basic calculus that the right hand side of the above inequality is restricted to the half-open unit interval $[0, 1)$. Hence, these conditions are sufficient for our model to satisfy the above claim. This allows us to focus on just the first and second order moments without losing any information predicted by the model.

C.2 Maximum likelihood

In this section we describe our use of maximum likelihood for estimating model parameters defining the strength of coevolution.

C.2.1 Estimating selection parameters and hypothesis testing

To connect our model to data, we assume that the metapopulation dynamics have reached equilibrium in distribution. This means that the distribution of the random variables representing the mean traits of each species within a given population is not changing. This assumption allows us to write the

means and variances as static functions of the model parameters. By assuming the equilibrium distribution of the metapopulation is bivariate normal, the likelihood of the data (the mean traits of each species in each population; $D = \{D_i\}_{i=1}^N \equiv \{(\bar{z}_1, \bar{z}_2)_i^\top\}_{i=1}^N$) given the two metapopulation mean trait values ($\mu = (\mu_1, \mu_2)^\top \equiv \mathbb{E}(\bar{Z}_1, \bar{Z}_2)^\top$), the two metapopulation trait variances ($V_1 \equiv \mathrm{Var}(\bar{Z}_1),\ V_2 \equiv \mathrm{Var}(\bar{Z}_2)$), and the metapopulation spatial covariance of the two traits ($C \equiv \mathrm{Cov}(\bar{Z}_1, \bar{Z}_2)$), ie, the first five moments, can be expressed as

$$L(D|\mu, \Sigma) = \prod_{i=1}^{N} \frac{\exp\left(-\frac{1}{2}(D_i - \mu)^\top \Sigma^{-1}(D_i - \mu)\right)}{2\pi\sqrt{|\Sigma|}},\ \Sigma = \begin{pmatrix} V_1 & C \\ C & V_2 \end{pmatrix}. \tag{C.20}$$

Using equations (C.17) we can easily derive the following equilibrium solutions of the first five moments

$$\mu_i = \frac{A_i A_j \theta_i + A_i B_j \theta_i + A_j B_i (\theta_j + \delta) + 2 B_i B_j \delta}{A_i A_j + A_i B_j + A_j B_i} \tag{C.21a}$$

$$V_i = \frac{B_i C + \frac{1}{2n_i}}{A_i + B_i} \tag{C.21b}$$

$$C = \frac{B_1(A_1 + B_1)G_1 n_1 + B_2(A_2 + B_2)G_2 n_2}{2(A_1 A_2 + A_1 B_2 + A_2 B_1)((A_1 + B_1)G_1 + (A_2 + B_2)G_2) n_1 n_2}. \tag{C.21c}$$

These expressions can in turn be used to solve for the two strengths of abiotic selection, the two strengths of biotic selection and the optimal offset. Because these expressions are quite cumbersome, however, we refer the reader to the Mathematica notebook associated with this paper for their full expressions. However, for the sake of developing intuition and for later use, we show the expressions of just the selection strengths as functions of other model parameters and statistical moments

$$A_i = \frac{\mu_j + \delta - \mu_i}{2n_i(V_i(\mu_j + \delta - \theta_i) - C(\mu_i - \theta_i))} \tag{C.22a}$$

$$B_i = \frac{\mu_i - \theta_i}{2n_i(V_i(\mu_j + \delta - \theta_i) - C(\mu_i - \theta_i))}. \tag{C.22b}$$

Since the sample moments of a multivariate normal distribution are the maximum likelihood estimates of its underlying moments and our solutions for the selection strengths and optimal offset correspond precisely with the sample moments, our solutions also maximize likelihood of our model given the data. That is, they maximize the function

$$L(D|A_1, A_2, B_1, B_2) \equiv L(D|\mu(A_1, A_2, B_1, B_2), \Sigma(A_1, A_2, B_1, B_2)). \tag{C.23}$$

Restricting the selection parameters B_i to match our null hypotheses of unilateral evolution ($B_1 = 0$ or $B_2 = 0$) results in a restricted likelihood (specifically, $L(D|A_1, A_2, 0, B_2)$ or $L(D|A_1, A_2, B_1, 0)$), one whose value is always less than that of $L(D|A_1, A_2, B_1, B_2)$. Then to calculate the significance of B_1, for

example, we compute the probability that the statistic

$$\Lambda_1 = 2\ln\left(\frac{L(D|A_1, A_2, B_1, B_2)}{L(D|A_1, A_2, 0, B_2)}\right) \tag{C.24}$$

lands within a given percentile of its distribution under the assumption of the null hypothesis $B_1 = 0$. We adopt the standard 95th percentile as our threshold for determining significance (hence, $\alpha = 0.05$). According to Wilk's Theorem, the distribution of this statistic is approximated by a Chi-square (Wilks, 1938). We therefore use this distribution to approximate our p-values.

We now use the results of this section to derive expressions for the restricted maximum likelihood moments under the two null conditions just described. Note that according to equation (C.22b), when $B_i = 0$ we must have $\mu_i = \theta_i$ for otherwise some parameter or moment must be infinite. This leaves one parameter fixed and one statistical moment fixed. The remaining four parameters can be solved for in terms of the remaining four statistical moments. Without loss of generality, we consider the situation when $B_1 = 0$. This results in the following solutions for the remaining strengths of selection and optimal offset

$$A_1 = \frac{1}{2n_1V_1} \tag{C.25a}$$

$$A_2 = \frac{1}{C^2 - V_1V_2}\left(\frac{C - V}{2n_2} + \frac{CG_1(V_2 - C)}{2n_1G_2V_1}\right) \tag{C.25b}$$

$$B_2 = \frac{1}{C^2 - V_1V_2}\left(-\frac{C}{2n_2} - \frac{CG_1V_2}{2n_1G_2V_1}\right) \tag{C.25c}$$

$$\delta = (\theta_2 - \theta_1) + (\mu_2 - \theta_1)\frac{G_2n_1V_1^2 + G_1n_2C^2}{G_2n_1V_1C + G_1n_2V_2C} \tag{C.25d}$$

That the remaining four equations for the equilibrium of statistical moments are invertible with respect to the remaining selection strengths and optimal offset indicates that the only moment whose value changes under the restricted maximum likelihood when $B_1 = 0$ is μ_1. Likewise, when we fix $B_2 = 0$ the only change in the resulting maximum likelihood moments is $\mu_2 = \theta_2$. Hence, formulating the likelihood ratios used to estimate the significance of the B_i are relatively straightforward.

C.3 Inference under broken assumptions

In this appendix we evaluate the performance of our method when some key assumptions are broken. Specifically, we extend the analyses described in the main text to inference when gene-flow is present, when data are significantly non-normal and when there is error in estimating the abiotic optima.

C.3.1 Inference for systems with gene-flow

Although our method formally assumes the absence of gene-flow among populations, we anticipate its usefulness persists in the face of weak gene-flow. To determine the validity of this assertion we

extend our model to include gene-flow as captured by an island model. Under this model the effect of gene-flow is to reduce the metapopulation variances (and hence covariance, but not correlation). As both strengths of selection also act to reduce variance across the metapopulation, we expect the inclusion of gene-flow to negatively bias our estimates of biotic selection.

In terms of the local model, the inclusion of gene-flow is approximated by an additional linear term. Denoting the strength of gene-flow by m_i, we have

$$\Delta\bar{z}_i = G_i(A_i(\theta_i - \bar{z}_i) + B_i\delta + B_i(\bar{z}_j - \bar{z}_i)) + m_i(\mu_i - \bar{z}_i) + \xi_i. \tag{C.26}$$

Assuming weak migration along with weak selection, we obtain

$$\Delta\mu_i = G_i\left[B_i\delta + B_i(\mu_j - \mu_i) + A_i(\theta_i - \mu_i)\right] \tag{C.27a}$$

$$\Delta V_i = -2(A_iG_i + m_i)V_i + 2B_jG_j(C - V_i) + G_i/n_i \tag{C.27b}$$

$$\Delta C = B_2G_2(V_1 - C) + B_1G_1(V_2 - C) - (A_1G_1 + A_2G_2 + m_1 + m_2)C. \tag{C.27c}$$

We can then solve for the selection strengths at equilibrium to arrive at

$$A_i = \left(\frac{1}{2n_i} - \frac{m_iV_i}{G_i}\right)\left(\frac{\mu_j + \delta - \mu_i}{V_i(\mu_j + \delta - \theta_i) - C(\mu_i - \theta_i)}\right) \tag{C.28a}$$

$$B_i = \left(\frac{1}{2n_i} - \frac{m_iV_i}{G_i}\right)\left(\frac{\mu_i - \theta_i}{V_i(\mu_j + \delta - \theta_i) - C(\mu_i - \theta_i)}\right). \tag{C.28b}$$

From these expressions we see that if the additive genetic variance within populations is much greater than the variance in mean phenotype among populations, gene-flow can safely be ignored. Of course, this will not likely be the case. Hence, we repeat our analysis of method performance described in the main text for data simulated with gene-flow present. In Figure C.1 of this appendix we display error rates for detecting coevolution and regression statistics for inferring the quantity of coevolution as functions of the amount of gene-flow present. These results demonstrate that low rates of gene-flow have negligible effects on our methods ability to infer coevolution, but that more extreme rates of gene-flow (e.g. 0.01) significantly reduces our methods precision in inferring the strength of coevolution as demonstrated by a reduced R^2.

C.3.2 Inference using non-normal data

The requirement that phenotypic data be bivariate normal will likely restrict the applicability of this method. However, if this method can be shown to be robust to non-normal data, the range of potentially coevolving systems for which this method can be used will expand greatly. We are therefore motivated to investigate the effect of non-normal data on inference. To understand this effect we repeat the methods discussed in the main text for simulated non-normal data sets.

To achieve this we repeatedly simulated data under our model until the resulting data set failed to be normal. We used the Shapiro-Wilks test with a threshold of normality τ to determine whether or not

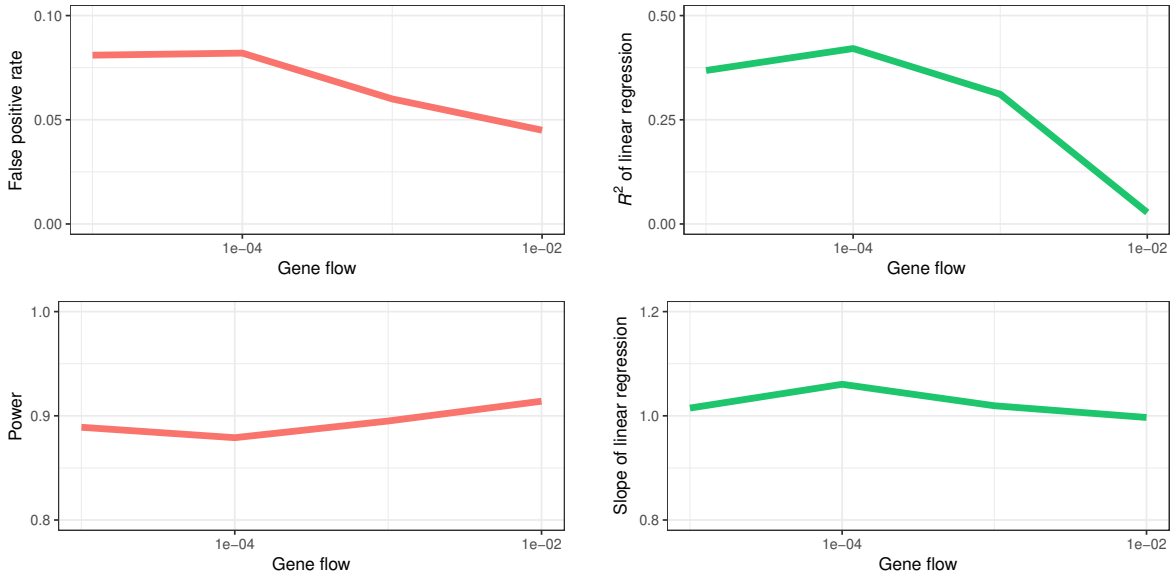

Figure C.1: Error rates and regression statistics as functions of the strength of gene-flow.

the data are normal (R Core Team, 2016). That is, if the p-value returned by the Shapiro-Wilks test is less than τ, we conclude the data to be non-normal. Using only those data sets which were determined to be significantly non-normal, we proceeded to perform inference as described above. We performed regression analysis on our estimates of biotic selection strengths and analysed false positive rates and power across a range of threshold values. Results are displayed in Figure C.2 of this appendix. From this figure we see the effect of non-normality on inference is negligible, even for highly non-normal data.

C.3.3 Inference with measurement error

We analyzed the effect of measurement error in the background parameters has on our ability to infer coevolution. In doing so we focused on error in measuring abiotic optima. We assume the estimates of abiotic optima are normally distributed and centered on the ”true” optima. Using the same approach for performance analysis as described in the main text, we examine regression statistics and error rates as functions of the standard deviation in the estimated abiotic optima. Results are displayed in Figure C.3 of this appendix. These results demonstrate that our method of coevolutionary inference is insensitive to measurement error in the abiotic optima.

C.4 Analysis of data

Here we describe the data and methods used for the analysis of coevolution. At its most basic level, our approach requires estimates of the mean phenotypes for both species in at least two populations where they are known to interact. This data is used to calculate the metapopulation means, variances and covariance upon which our maximum likelihood estimation rests. In addition to this core data,

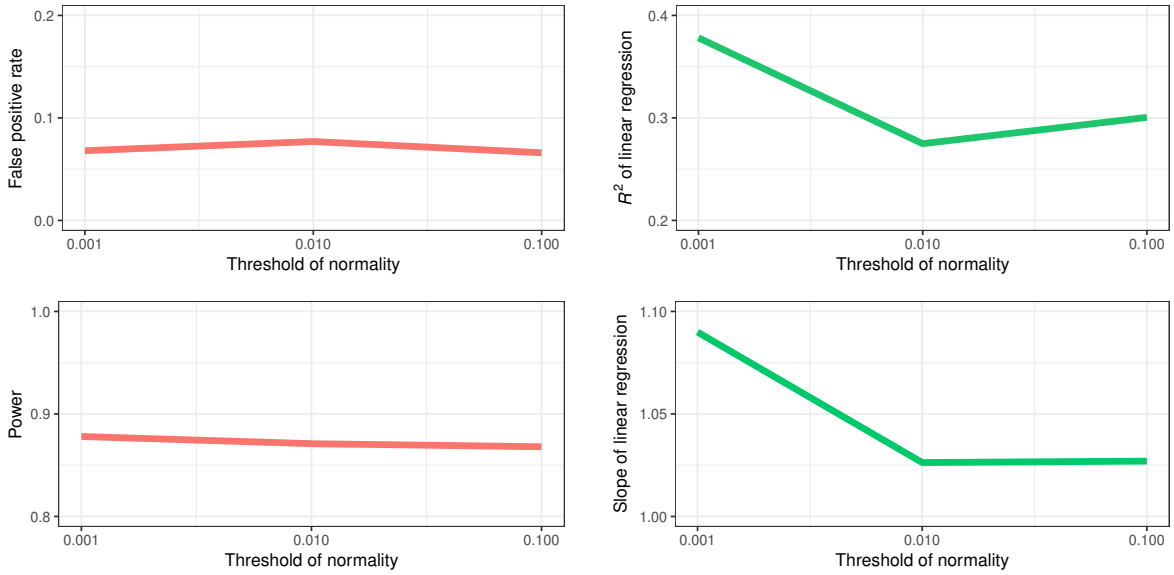

Figure C.2: Error rates and regression statistics as functions of τ, the threshold for which data is determined to be non-normal. The smaller τ is the more non-normal the data is according to the Shapiro-Wilks test.

our approach requires estimates for several key background parameters: the effective local population sizes for each species n_i, the optimal traits in the absence of the interaction θ_i (the abiotic optima) and the additive genetic variances G_i. Model parameters that are inferred by our method are the strengths of abiotic selection A_i, the strengths of biotic selection B_i, and the optimal offset δ.

For many systems, it will be possible to estimate effective population sizes using molecular techniques (e.g. Beerli, 2005). If these are known for each population, or for a subset of populations, the harmonic mean of these should be used for the parameter n_i. This is justified in Appendix C.1.4. In the absence of effective population size estimates, one can instead infer the composite parameters $A_i n_i$ and $B_i n_i$. As shown in equation (C.22b) the selection strengths are directly proportional to the effective population sizes. Hence, the suggested composite parameters provide reasonable proxies for the strengths of selection.

The abiotic optima are particularly important for inference. If these parameters are close to the biotic optima, it becomes difficult to assess the relative importance of biotic selection as its signal becomes partially masked by the patterns resulting from abiotic selection. In our analyses, three distinct approaches were used to estimate these parameters: (1) We used the trait means of populations isolated from the interacting species as an estimate for θ_i. However, if there is significant gene flow between these populations and populations where the interaction is taking place, the resulting estimate for θ_i may be biased towards the biotic optimum, which would weaken the coevolutionary signal and make our approach conservative. Also, this approach assumes the form and strength of abiotic selection is equal in locations where the interaction does and does not occur. (2) We also used the mean phenotypes of members of a species that do not partake in the interaction at any point in their life history to estimate θ_i. For example, in the camellia-weevil interaction, the males do not oviposit and hence do

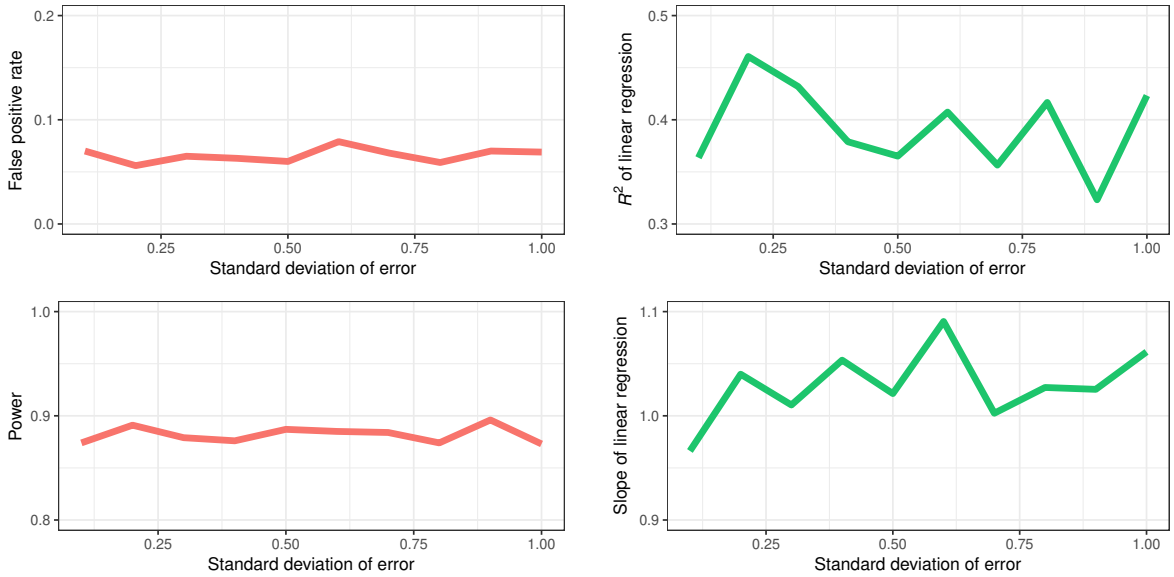

Figure C.3: Error rates and regression statistics as functions of the amount of measurement error in abiotic optima.

not interact with camellia by boring holes in its pericarp at any point in their life histories. Hence, we assume their rostrum lengths provide decent estimates of the abiotic optima. This approach also assumes approximately equivalent abiotic selection surfaces for both individuals within a population that do and do not partake in the interaction. (3) Lastly, we estimated the abiotic optimum for the fly *M. longirostris* through data on closely related species. We reason that each sibling species represents an alternative evolutionary trajectory, especially since they do not coincide with the *M. longirostris* pollination guild (Barraclough and Slotow, 2010). Hence, we may crudely estimate the abiotic optimal phenotype as the trait values of these sibling species.

In the following sections, we describe how we applied this general approach to interactions between the flower *Lapeirousia anceps* and its fly pollinator *Moegistorhynchus longirostris* as well as the camellia fruit *Camellia japonica* and its weevil seed-predator *Curculio camelliae*. In both interactions, values subscripted with the number 1 indicate the insect species and values subscripted with the number 2 the plant species. All computations were performed in the statistical programming language *R* (R Core Team, 2016). Data and scripts used in this analysis will be made available at:

`https://github.com/bobweek/measuring.coevolution`

C.4.1 The interaction between *L. anceps* and *M. longirostris*

Phenotypic data was obtained from previously published work (Pauw et al., 2009) estimating the population mean proboscis length and population mean tube length for eight populations in South Africa where this interaction occurs. These data are summarized in Table C.1.

The abiotic optimum trait value for *L. anceps* was assumed to be equal to the average of the mean

tube lengths found in Red Hill and Uitkomsberge, the only two populations recorded which are not visited by *M. longirostris* (Pauw et al., 2009). This average came to be 41.45 mm. For the fly, we estimated the abiotic optimum in four different ways. For the first three we set the abiotic optimum equal to the recorded trait-value of one of the sister species (found in Bequaert, 1935): (1) *M. braunsi*, (2) *M. brevirostris* and (3) *M. perplexus*. Only for *M. braunsi* were separate measurements recorded for each sex. We therefore use the average of these two lengths as the species wide mean. (4) Lastly, we estimated the abiotic optimum as the average of the mean traits for the three sibling species. These estimates are summarized in Table C.2. We implemented our method for measuring coevolution separately for each of these four different parameterizations. Table C.2 shows the results under each of these parameterizations do not substantially differ.

Table C.1: Moments used for analysis in the fly-flower interaction.

Moments	Sample Moments
μ_1	65.83 mm
μ_2	55.74 mm
V_1	206.4 mm^2
V_2	163.6 mm^2
C	143.4 mm^2

Table C.2: Results under different abiotic optima for *M. longirostris*.

Parameter	*M. brevirostris*	*M. perplexus*	*M. braunsi*	Average
θ_1	11.5 mm	32.0 mm	41.0 mm	28.17 mm
B_1	6.409660e-05 mm^{-2}	6.394011e-05 mm^{-2}	6.386160e-05 mm^{-2}	6.397159e-05 mm^{-2}
B_2	1.718896e-06 mm^{-2}	1.867071e-06 mm^{-2}	1.941416e-06 mm^{-2}	1.837269e-06 mm^{-2}
A_1	7.025746e-06 mm^{-2}	7.073487e-06 mm^{-2}	7.097440e-06 mm^{-2}	7.063885e-06 mm^{-2}
A_2	3.140721e-06 mm^{-2}	3.122435e-06 mm^{-2}	3.113260e-06 mm^{-2}	3.126113e-06 mm^{-2}
δ	16.03591 mm	13.82320 mm	12.84025 mm	14.23957 mm
p_1	<2.22e-16	<2.22e-16	2.22e-16	<2.22e-16
p_2	1.19e-07	1.19e-07	1.19e-07	1.19e-07
$\mathfrak{C}$	1.049645e-05 mm^{-2}	1.092615e-05 mm^{-2}	1.113472e-05 mm^{-2}	1.084126e-05 mm^{-2}
$\mathfrak{B}$	0.1745277	0.1861598	0.1918956	0.1838421

C.4.2 The interaction between *C. camelliae* and *C. japonica*

Phenotypic data was obtained from previously published work (Toju et al., 2011a) estimating the population mean rostrum length and population mean pericarp thickness for weevil and camellia populations throughout Japan. We follow the approach of Toju et al. (2011a) by focusing on female weevils as male weevils do not oviposit and hence are not involved with the interaction. Of the phenotypic data available we included only those populations for which each species was sampled. This made for 25 populations (see the csv for more details). Table C.3 summarizes the metapopulation means, variances and covariance of these data.

Effective population sizes were taken as the harmonic means of those reported in Toju et al. (2011b). We assumed the male weevil rostrum length to be the abiotic optimum of the weevil species. Means for the rostrum length of male weevils are available from 11 different populations (Nagaoka, Kyoto, Jurinji, Taiji, Muroto, Ashizuri, Reihoku, Takahama, Yahazu, Shitoko and Hyanyama). The optimal pericarp thickness for the camellia was estimated as the average of mean pericarp thicknesses from 17 populations with no recorded weevil populations (Miyatsuka, Hatake, Mukai, Izuhara, Kamitsuki, Izu, Igatani, AkoN2, Tsubota, Jinoshima, Kishiku, Tomie, Kuwanoura, Ichinoura, Nakakoshiki, Kashima and Teuchi). These values were taken from Toju et al. (2011a). The resulting background parameters can be found in Table C.4.

Table C.3: Moments used for analysis in the camellia-weevil interaction.

Moments	Sample Moments
μ_1	12.5 mm
μ_2	10.4 mm
V_1	14.2 mm^2
V_2	28.8 mm^2
C	19.3 mm^2

Table C.4: Parameter values used for analysis in the camellia-weevil interaction.

Parameters	*C. camelliae*	*C. japonica*
θ_i	5.91 mm	6.26 mm
n_i	3.08e+04	1.79e+03

C.4.3 Effect sizes

To estimate the effect of coevolution on the bivariate distribution of mean-trait-pairs, we used our model to predict the parameters of the metapopulation distribution in the absence of coevolution. Referring to the equlibrium expressions C.21 in Appendix likelihood, we can calculate the maximum likelihood solutions for A_1 and A_2 when $B_1 = B_2 = 0$. Under this condition we find $\mu_i = \theta_i$, $C = 0$ and

$$V_i = \frac{1}{2n_i A_i}.$$

Thus, in the absence of coevolution both V_1 and V_2 remain equal to their observed values. The remaining three moments, however, must shift from their observed values.

To calculate the magnitude by which these moments shift, we calculate a metric for the effect size. In general, the effect size of a treatment is typically defined as the difference between sample means with and without the treatment divided by the pooled standard deviation of the samples. A natural multivariate extension of such a metric is the Mahalonobis distance. Setting μ_c, Σ_c and μ_0, Σ_0 equal to the observed and predicted moments with and without coevolution respectively, we use the following expression as our effect size

$$\mathcal{E} \equiv \sqrt{(\mu_c - \mu_0)^\top \Sigma^{-1} (\mu_c - \mu_0)}$$

where Σ is the pooled covariance matrix, which in our case simplifies to

$$\Sigma = \frac{\Sigma_c + \Sigma_0}{2}.$$