

Cryptic selection forces and dynamic heritability in generalized phenotypic evolution

William Gilpin^{1*}, Marcus W. Feldman²

¹Department of Applied Physics, ²Department of Biology,

Stanford University, Stanford, CA

*wgilpin@stanford.edu

March 29, 2018

Abstract

Individuals with different phenotypes can have widely-varying responses to natural selection, yet many classical approaches to evolutionary dynamics emphasize how a population's average phenotype increases in fitness over time. However, recent experimental results have produced examples of populations that have multiple fitness peaks, or that experience frequency-dependence that affects the direction and strength of selection on certain individuals. Here, we extend classical fitness gradient formulations of natural selection in order to describe the dynamics of a phenotype distribution in terms of its moments—such as the mean, variance, skewness, etc. The number of governing equations in our model can be adjusted in order to capture different degrees of detail about the population. We compare our simplified model to direct Wright-Fisher simulations of evolution in several canonical fitness landscapes, we find

that our model provides a low-dimensional description of complex dynamics not typically explained by classical theory, such as cryptic selection forces due to selection on trait ranges, time-variation of the heritability, and nonlinear responses to stabilizing or disruptive selection due to asymmetric trait distributions. In addition to providing a framework for extending general understanding of common qualitative concepts in phenotypic evolution—such as fitness gradients, selection pressures, and heritability—our approach has practical importance for studying evolution in contexts in which genetic analysis is infeasible.

1 Introduction

The effects of evolutionary forces may be apparent in natural populations even when their underlying genetic consequences are not known. The size of river guppies increases when their natural predators are depleted;¹ the beaks of Darwin’s finches grow larger after droughts;² and mammals grow smaller in response to climate change.³ These and other natural and experimental studies demonstrate that rapid selection can produce noticeable changes in specific traits, underscoring the importance of considering phenotypic models of natural selection. These models are particularly relevant to studies in the field or of the fossil record where genetic analysis is unavailable or infeasible.⁴

A widely-used framework for such theories is the fitness landscape,⁵ an abstract function that describes the collective survival or reproductive benefits conferred by a given phenotype: an evolving population typically approaches a locally or globally maximum value in this space, subject to constraints on its rate of adaptation. But the underlying dynamics of this process may depend strongly on the context,⁶ and molecular techniques have only recently begun to shed light on the individual steps of adaptation and the intermediate phenotypes they produce.^{7, 8} Moreover, macroscale analyses have produced examples of non-monotone fitness functions with elaborate, multi-peaked topographies subject to strong frequency dependence,⁹ selection that acts on trait ranges in addition to average values,¹⁰ and selection

forces with nonlinear effects on metric traits.¹¹ These studies indicate the need for simple, analytical models that can provide heuristic insight into the complex dynamics of phenotypic adaptation. Of particular importance are cases in which traditional experimental metrics, such as the realized heritability or selection response, have complex time-dependence over long timescales,^{12, 13} rendering derived experimental quantities such as selection coefficients insufficient as large-scale descriptors of population dynamics.

Many classical approaches to phenotypic evolutionary theory readily describes the evolution of a bell-shaped and fixed-width trait distribution within a given fitness landscape, resulting in the widely-known result that the rate of adaptation of the mean trait is directly proportional to the gradient of the mean fitness.^{14–17} This depiction of evolution as a fitness gradient-climbing problem is particularly intuitive, and it mirrors earlier work on the adaptive landscape of individual genes.⁵ However, many phenotypes fail to satisfy these conditions,^{10, 18, 19} as a result of which many experimental and theoretical studies have noted strong limitations on the applicability of fitness gradient dynamics.^{20–23} Efforts to establish more general rules for the evolution of an arbitrary trait distribution typically reformulate the underlying mathematics in terms of agent-based rules or a stochastic transmission kernel;^{24–26} however, these alternative formulations are difficult to compare directly with classical fitness gradient dynamics, and they typically introduce new assumptions regarding the underlying genetic processes or functional forms of the trait distribution or fitness.

Here, we develop a general model of phenotypic evolution that seeks to avoid imposing functional constraints on the fitness or trait by using a moment series expansion of the fitness landscape. Our approach has its origin in classical approaches that describe evolution of the mean trait as a fitness gradient-climbing problem, but we add additional dynamical equations for the variance, skewness, kurtosis, and finer-scale statistical features of the trait distribution. By adding or removing dynamical equations that describe various moments of the trait distribution, the level of detail our model captures about the evolutionary dynamics may be tuned. Importantly, our model reduces to a classical fitness gradient model when

only the mean trait is allowed to vary. Our model explicitly relates the topology of the fitness landscape to the timescales (and thus dynamical relevance) of various phenomena through a series of coupling constants, which we compute analytically for several canonical fitness landscapes in a series of demonstrative examples. Using these examples, we show how even a simple generalization of classical fitness gradient dynamics can lead to a series of surprising effects in the evolutionary dynamics, including cryptic forces of selection that cause changes in the trait distribution even when the local fitness gradient is zero, as well as suppression of disruptive selection due to asymmetry and skewness in the trait distribution. We also show that our model can be re-formulated in terms of the coupled dynamics of the narrow-sense heritability and the mean fitness, and we show how these two quantities may jointly evolve under various conditions.

2 Model

Our model is based on a series expansion of the trait distribution in terms of its moments (such as the mean trait, trait variance, trait skewness, etc). Each moment has a separate dynamical equation, which couples its dynamics to those of the remaining moments. The strength and magnitude of the coupling of each moment’s dynamics to other moments depends on the specific fitness landscape, and can be summarized in terms of a finite set of “coupling coefficients” intrinsic to the landscape. Our basic approach is shown schematically in Figure S1, and below we describe how we derive dynamical equations for the mean trait \bar{z} , trait range σ , and mean fitness \bar{w} . We also describe below the assumptions of our approach, as well as the relationship between our approach and similar series-based approaches developed by others.

2.1 Trait mean dynamics in an arbitrary phenotype distribution

We follow the classical approach to deriving phenotypic evolution originally developed by Lande.¹⁵ We start with the breeder's equation, which relates the dynamics of the trait mean

95 \bar{z} to its change after a period of selection

$$\dot{\bar{z}} = \frac{V}{\sigma^2}(\bar{z}_w - \bar{z}), \quad (1)$$

where the mean trait value \bar{z} , mean fitness \bar{W} , and mean trait after selection \bar{z}_w are defined in terms of statistical averages over the entire population,

$$\bar{z} = \int z p(z) dz, \quad \bar{z}_w = \int z p(z) \frac{W(z)}{\bar{W}} dz, \quad \bar{W} = \int W(z) p(z) dz. \quad (2)$$

Here $p(z)$ is the frequency of the trait z in the population, $W(z)$ is an individual's fitness as a function of its trait. The prefactor V/σ^2 in Eq. 1 is equivalent to the narrow-sense heritability h^2 of the trait, with V representing the additive genetic variance. Eq. 1 incorporates the primary assumptions regarding the underlying genetics; namely, that there is no direct gene-environment interaction, and that there is a linear regression between the selection differential and the mean trait value over short timescales^{27, 28} (although the slope of this regression, h^2 , need not be constant). In many previous studies in which Eq. 1 appears, the phenotypic variance σ^2 is also assumed to be fixed, either due to logarithmic ranges in the values of metric traits that suppress the magnitude of fluctuations in trait variance, or due to the assumption of a fixed trait distribution shape that becomes a Gaussian distribution after an appropriate nonlinear transformation.^{15, 27} We will relax this requirement here.

Constancy of genetic variance V must be decided empirically for a given system under study; in general, V may change due to drift or direct gene-environment interactions.^{13, 28, 29} However, in the model of phenotypic evolution presented here, as in similar approaches based on the dynamics of single continuous traits,³⁰⁻³² V may be assumed to vary much more slowly

than the phenotypic variance σ^2 provided there is linkage equilibrium, weak selection, or a near-equilibrium trait distribution. Additionally, for cases in which V changes quickly, such as during periods of strong directional selection or from transient effects such as the accumulation of linkage disequilibrium, rapid compensatory effects (for example, recombination) can allow V to stabilize at a well-defined average value over long timescales.^{1, 12} Accordingly, low or stationary genetic variance has been reported in certain experimental systems that may satisfy these conditions.³³

Here, we use a standardized trait distribution $p(z) = p((z - \bar{z})/\sigma)$ and also parametrize the fitness as $W(z) = W(z - c\bar{z})$, where c is an arbitrary positive constant. The case $c = 0$ corresponds to a trait that directly affects survival independently of other individuals in the population (examples include metabolic efficiency, camouflage coloration, or immune system competency), while the case $c = 1$ corresponds to a trait that affects fitness of a given individual only relative to others in the population (examples include secondary sex characteristics or running speed relative to a herd). Under this assumption, taking the derivative of both sides of Eq. 2 with respect to \bar{z} yields a simple relationship between an individual's fitness landscape and the population mean fitness,

$$\frac{d\bar{W}}{d\bar{z}} = (1 - c) \int \frac{dp(z)}{d\bar{z}} W(z) dz. \quad (3)$$

See Appendix C for full derivation. We note that if $c = 1$, then $\frac{d\bar{W}}{d\bar{z}} = 0$; thus when the fitness landscape depends only on the difference between each individual trait and the trait mean, the mean fitness itself cannot depend on the mean trait.

Next, we assume that the trait distribution before selection $p(z)$ is separable into a Gaussian part $\mathcal{N}[\bar{z}, \sigma](z)$ and a dimensionless non-Gaussian component f , which without loss of generality we express in terms of the dimensionless standardized coordinate $\tilde{z} = (z - \bar{z})/\sigma$,

$$p(z) = \left(\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(z-\bar{z})^2}{2\sigma^2}} \right) f(z) = \mathcal{N}[\bar{z}, \sigma](z) f\left(\frac{z - \bar{z}}{\sigma}\right). \quad (4)$$

135 We emphasize that $f = 1$ corresponds to a purely Gaussian trait distribution, in which case our theory recreates standard evolutionary dynamics.^{15–17}

Inserting Eq. 4 and Eq. 3 into Eq. 1, we arrive at a dynamical equation for the mean trait of a non-Gaussian trait distribution,

$$\dot{\bar{z}} = \frac{V}{\bar{W}} \left(\frac{1}{1-c} \frac{d\bar{W}}{d\bar{z}} \right) + \frac{V}{\sigma \bar{W}} \int \mathcal{N}[\bar{z}, \sigma](z) f'(\bar{z}) W(z) dz. \quad (5)$$

Asymptotic analysis confirms that the first term vanishes when $c = 1$ due to a first-order zero
 140 in $\frac{d\bar{W}}{d\bar{z}}$ at $c = 1$ (see Appendix D). The first term in Eq. 5 corresponds to a classical fitness gradient dynamics model, and represents the complete dynamics if the trait distribution is Gaussian ($f' = 0$).^{17, 34} The second term determines how the higher-order moments of the trait distribution affect the evolutionary dynamics. Importantly, this new term depends explicitly on the values of higher-order moments at each timestep. Therefore, Eq. 5 can only
 145 be used to determine the dynamics if additional differential equations are specified for the trait variance, skewness, etc, which makes full determination of the dynamics a “moment closure” problem because the dynamics of the n^{th} moment will, in general, depend on the $(n+1)^{\text{th}}$ moment.^{35, 36} However, Eq. 5 can still be used to find the mean trait equilibrium and its stability as functions of the other moments.

150 We next expand the trait distribution in Eq. 4 as a Gram-Charlier A series,

$$p(z) = \mathcal{N}[\bar{z}, \sigma](z) \left(1 + \sum_{n=3}^{\infty} c_n He_n \left(\frac{z - \bar{z}}{\sigma} \right) \right), \quad (6)$$

where the expansion coefficients c_n are uniquely determined by the moments of the trait distribution (see Appendix B). Inserting the rightmost parenthetical term into Eq. 5 as f , we use the property of Hermite polynomials $He'_n = n He_{n-1}$ and the change of variables $\tilde{z} = (\bar{z} - z)/\sigma$ to find,

$$\dot{\bar{z}} = \frac{V}{\bar{W}} \left(\frac{1}{1-c} \frac{d\bar{W}}{d\bar{z}} \right) + \frac{V}{\sigma \bar{W}} \sum_{n=3}^{\infty} c_n n \mathcal{W}_{n-1}, \quad (7)$$

where the first-order term corresponds to classical phenotypic evolutionary theory. In the higher-order terms, the singly-indexed family of integrals \mathcal{W}_n is given by

$$\mathcal{W}_n \equiv \int \mathcal{N}[0, 1](\tilde{z}) He_n(\tilde{z}) W(\sigma \tilde{z} + \bar{z}) d\tilde{z}. \quad (8)$$

This equation represents a projection of the fitness landscape onto a basis of Hermite polynomials, with finer-scale features in the fitness landscape being represented by larger values of n in the series. However, if the fitness landscape is sufficiently smooth, there always exists some n above which the sequence \mathcal{W}_n continuously decreases, suggesting that a finite set of \mathcal{W}_n would be sufficient to describe many simple fitness landscapes.

We refer to the series \mathcal{W}_n as the “coupling coefficients” because the form of Eq. 7 suggests that the values of \mathcal{W}_n represent the degree of coupling between the fitness landscape and progressively larger cumulants of the trait distribution. Larger-scale features of the fitness landscape affect the dynamics over longer timescales, and thus appear in lower-order \mathcal{W}_n ; conversely, higher-order coupling coefficients provide increasingly precise information about the fitness landscape, and small-scale dynamical changes within it. Thus \mathcal{W}_n serve a similar function to the individual selection differentials described in previously-developed models of non-Gaussian breeding value distributions using multilocus genetic theory.^{23, 37}

Importantly, the coupling coefficients \mathcal{W}_n are parameters that only need to be computed once as long as the fitness landscape itself remains constant. This allows the integro-differential dynamics of an arbitrary trait distribution and fitness landscape to be expressed as a series of coupled ordinary differential equations for the time-varying moments (appearing in the individual c_n of the Gram-Charlier series), with the fitness landscape appearing only as a set of known coupling terms \mathcal{W}_n . For most fitness landscapes the individual \mathcal{W}_n are straightforward to compute analytically (using generating functions) or numerically, due to the orthogonality properties of Hermite integrals.³⁸ This allows the set of \mathcal{W}_n to be “pre-computed” for a given fitness landscape and then inserted as explicit terms into the

dynamical equation Eq. 7. The dynamics of the trait distribution subject to this fitness land-
 180 scape can then be determined without the need to perform any additional time-dependent
 integrals over the fitness landscape. If the fitness landscape were to change in time, any
 change arising from a process independent of the trait distribution (e.g. rainfall variation
 due to large scale weather patterns, or depletion of predators due to overfishing) would allow
 the time variation of each \mathcal{W}_n to be solved separately from the trait distribution dynamics.
 185 The resulting set of $\mathcal{W}_n(t)$ may then be inserted into Eq. 7 to yield the trait dynamics.

2.2 Trait variance dynamics

The method used above to derive the dynamics of the trait mean may be employed to derive
 corresponding dynamical equations for any moment or cumulant of the trait distribution.
 Here, we determine the dynamical equation for the trait standard deviation (and thus vari-
 190 ance) using a similar method to that above.

As with the mean trait, we assume that the trait variance $M \equiv \sigma^2$ has linear heritability
 with dynamics specified by

$$\frac{d}{dt}(\sigma^2) = h_\sigma^2(\sigma_w^2 - \sigma^2), \quad (9)$$

where h_σ^2 is the *variance heritability*, or the degree to which the phenotypic variance in
 one generation influences the phenotypic variance of the next generation. σ_w^2 is the trait
 195 variance after selection, which is defined in a manner analogous to \bar{z}_w . If the (unmodeled)
 recombination and mutation processes are sufficiently smooth, such a linear relationship and
 the value of h_σ^2 represents the lowest-order term in a suitable series expansion due to the
 summation properties of cumulants of random variables.^{39, 40} This relation based on early
 work on the infinitesimal model,²⁸ in which a given trait is continuous due to an effectively
 200 infinite number of individual loci contributing to it. We further discuss below the limitations
 of using a linear heritability equation.

If the trait distribution has the separable form Eq. 4, then it can be shown (see Appendix

E) that Eq. 9 simplifies to a dynamical equation of the form,

$$\dot{\sigma} = \frac{U}{\overline{W}} \left(\frac{d\overline{W}}{d\sigma} + \frac{1}{\sigma} \int \tilde{z} \mathcal{N}[0, 1](\tilde{z}) W(\sigma \tilde{z} + \bar{z}) f' d\tilde{z} \right),$$

where $U \equiv (\sigma h_\sigma)^2/2$. Inserting the Gram-Charlier series (Eq. 6) for $p(z)$ and exploiting the
 205 properties of the Hermite polynomials results in a final expression,

$$\dot{\sigma} = \frac{U}{\overline{W}} \frac{d\overline{W}}{d\sigma} + \frac{U}{\sigma \overline{W}} \sum_{n=3}^{\infty} c_n n \left((n-1) \mathcal{W}_{n-2} + \mathcal{W}_n \right). \quad (10)$$

As with the mean trait dynamics Eq. 7, the variance dynamics (and thus the width of the trait distribution) simplifies to a gradient of the mean fitness, plus an infinite summation over \mathcal{W}_n that takes into account increasingly fine-scale moments of the trait distribution. Using this construction technique, the dynamics of any arbitrary moment or cumulant of
 210 the trait distribution may be expressed as a dynamical equation with a similar form. The properties of Hermite polynomials guarantee that the fitness landscape is coupled to the dynamics of each moment solely through a fixed set of \mathcal{W}_n , further emphasizing how the moment-based formulation abets analytical work (through truncation of the infinite series) or numerical study (through pre-computation of an arbitrary number of terms in the series
 215 of \mathcal{W}_n).

2.3 Mean fitness dynamics

Based on the forms of Eq. 7 and Eq. 10, the dynamics of the mean fitness are given by

$$\dot{\overline{W}} = \frac{d\overline{W}}{d\bar{z}} \dot{\bar{z}} + \frac{d\overline{W}}{d\sigma} \dot{\sigma} + \sum_{i=3}^{\infty} \frac{d\overline{W}}{d\kappa_i} \kappa_i, \quad (11)$$

where κ_i are the cumulants of $p(z)$. Because all derivatives of \overline{W} may themselves be expressed in terms of the cumulants of the trait distribution $p(z)$, the dynamics of \overline{W} do not involve
 220 an additional dependent variable in the dynamical system. The form of $\dot{\overline{W}}$ suggests that,

even in the absence of additive genetic variance ($V = 0$) or in the case of a fixed mean trait $\dot{\bar{z}} = 0$, the mean fitness may continue to change due to the contribution of higher-order cumulants of the trait distribution to the dynamics. These higher-order effects are absent in the standard breeder’s equation and will be discussed further below.

2.4 Gradient dynamics with leading-order corrections

Here, we simplify equations Eq. 7 and Eq. 10 to account for only the leading-order effects of non-Gaussian moments in the trait distribution, resulting in a closed form for the dynamical equations.

The general framework used above for deriving $\dot{\bar{z}}$ and $\dot{\sigma}$ may, in principle, be used to derive dynamical equations for an arbitrary number of moments of the trait distribution. In these cases, a Gram-Charlier series with a non-Gaussian leading-kernel may be preferable.⁴¹ However, in the following sections we focus primarily on the dynamics of the first and second moments of the trait distribution ($\dot{\bar{z}}$ and $\dot{\sigma}$ derived above) because in many standard fitness models, selection acts directly on the mean and width of the trait distribution, but not necessarily the skewness and higher moments. As a result, the terms in the series \mathcal{W}_n decrease quickly in magnitude, causing the dynamics of \bar{z} and σ to be nearly uncoupled from the dynamics of higher moments. This is equivalent to assuming that higher-order cumulants of the trait distribution affect the dynamics solely as fixed parameter values in the series terms in Eq. 7 and Eq. 10. This restriction implies that higher moments of the trait distribution have zero effective heritability, and that natural selection, together with reproduction, mutation, and recombination, causes these moments to vary much more slowly than the mean and variance.^{36, 42} However, even if higher moments are allowed to vary continuously due to the action of natural selection, the equilibria found under our assumptions would remain the same—although their stability may be subject to additional constraints.

In order to illustrate potential applications of this formulation and to produce closed-form

results, in what follows we truncate at fourth order the infinite series c_n appearing in Eq. 7 and Eq. 10. This fourth-order closure is used in order to isolate the effects of asymmetry (c_3) and heavy-tailedness (c_4) on the dynamics of the trait distribution: neither effect can be described by the original breeder's equation due to its implicit assumption of a Gaussian trait distribution. Because of the ease with which individual \mathcal{W}_n may be computed, additional terms could easily be added to account for other effects; however, the contribution of these higher-order terms to the dynamics of lower-order moments is bounded due to the scaling properties of the Gram-Charlier series ($c_n \sim 1/\sigma^n$). Similar cumulant closure relations appear in moment-based models of ecological dynamics.^{35, 43, 44} In previous work by other authors,^{23, 37} the series terms in a discrete-time cumulant dynamical model were computed explicitly for the case of truncation selection, for which the short-time dynamics primarily depend on the leading moments. Additionally, in some discrete-time models of multilocus selection, the cumulant dynamical equations intrinsically contain a finite number of terms and cross terms.^{45, 46}

Together, these assumptions result in a simplified set of dynamical equations,

$$\dot{z} = \frac{V}{\bar{W}} \left(\frac{1}{1-c} \frac{d\bar{W}}{d\bar{z}} + \frac{1}{2\sigma} \left(\gamma \mathcal{W}_2 + \frac{1}{3}(k-3)\mathcal{W}_3 \right) \right) \quad (12)$$

$$\dot{\sigma} = \frac{U}{\bar{W}} \left(\frac{d\bar{W}}{d\sigma} + \frac{1}{2\sigma} \left(\gamma(\mathcal{W}_2 + \mathcal{W}_0) + \frac{1}{3}(k-3)(\mathcal{W}_3 + 2\mathcal{W}_1) \right) \right), \quad (13)$$

where γ and k are, respectively, the skewness and kurtosis of the trait distribution. While γ may take any value, the kurtosis is mathematically bounded from below by $k \geq \gamma^2 + 1$. Together, Eq. 12 and Eq. 13 may be considered a first-order “correction” to the classical fitness gradient dynamics equation, and they account for the leading-order effects of non-Gaussian features of the trait distribution. When the fitness landscape is centered ($c = 1$), the first terms vanish from both of these equations. In this case, if the trait distribution itself is purely Gaussian ($\gamma = 0, k = 3$), then the remaining terms vanish from the right hand sides of both equations and the trait distribution evolves under classical phenotypic evolutionary

theory. However, if the trait distribution is non-Gaussian but $c = 1$, then \bar{z} and σ will vary
 270 due entirely to the non-Gaussian components of the trait distribution.

3 Assumptions and related models

Our model is comparable to moment-series approaches previously used to study natural selection in phenotypic and genotypic systems under various assumptions;^{23, 36, 37, 46–48} we review these approaches and in greater detail in Appendix L. As in previous models,
 275 assume that the full dynamics of the trait distribution may be approximated by a finite series of ordinary differential equations, thus reducing a complex partial differential equation problem to a lower-dimensional moment evolution problem. Mathematically, the set of techniques upon which we base our analysis parallels those found in models of genetic processes under the infinitesimal model, in which a given continuous trait is assumed to depend on
 280 an arbitrary number of alleles—in particular, the use of a Gram-Charlier series as a starting point for cumulant iteration equations was pioneered in genetics by Zeng,⁴⁸ as well as Turelli and Barton.²³ Additionally, we note that several related works have focused on the distribution of fitness values,^{46, 49} including recent work producing the intriguing result that many fitness distributions asymptotically approach a fixed class of distributions.³⁶

285 Our assumption of a linear heritability for higher cumulants, Eq. 9, represents the primary assumption of our model regarding the underlying mechanisms of genetic inheritance in our system; it thus introduces the primary limitations of this purely phenotypic approach because it does not include an explicit mating mechanism. For extremely strong selection (leading to large changes in the trait distribution within one generation), our model may fail due
 290 to both the continuous time assumption and the presence of higher-order, terms in Eq. 9. The form of these terms depends on the underlying genetic process, and their general form has previously been found using multilocus theory.^{23, 45} These subleading terms affect the dynamics over long timescales, and may alter the stability criteria of equilibria.

Our continuous time phenotypic equations may be compared to previous work on cumu-
 295 lant dynamics that treat selection as a discrete-time repeated sampling process weighted by
 the fitness.⁵⁰ Thus we test our findings below against a simulated Wright-Fisher process,
 and we find general agreement (see Appendix N for details of this numerical work). We thus
 emphasize that our model is best applied to the study of short-term phenotypic evolutionary
 trends when genetic assays are unavailable or infeasible; however, over longer timescales in
 300 which the additive genetic variance parameter V varies, we expect high-order effects in her-
 itability to manifest. These limitations are consistent with the classical usages of “gradient
 dynamics” models (which our model essentially generalizes), which have found particular
 utility for the study of coupled ecological and evolutionary processes.^{15–17, 34}

4 Results

4.1 Cryptic forces of selection arise from non-Gaussian trait dis- 305 tributions

In classical phenotypic evolution, the first term in Eq. 12 is associated with the “force of
 selection” on the mean trait.⁵¹ However, the remaining terms in Eq. 12 and Eq. 13 show that
 the trait distribution can change even when this term is zero, allowing the trait distribution
 310 to evolve in the absence of any apparent force of selection. These cryptic selection forces can
 have significant effects on the overall dynamics of natural selection.

As a demonstration of this effect, Figure 1 shows the behavior of the system Eq. 12,
 Eq. 13 relative to a null model in which the trait distribution is always Gaussian (in which
 case all terms containing \mathcal{W}_n vanish in the dynamical equations). The figure illustrates two
 315 separate cryptic forces of selection: the excess selection force on the mean trait (left plots)
 and the excess selection force on the trait variance (right plots). For the Figure, we use a

simple fitness landscape consisting of exponential directional selection,^{40, 52, 53}

$$W = W_0 e^{sz}. \quad (14)$$

Such a landscape represents limiting case in several common contexts, including selection on metric traits (which frequently have logarithmic ranges),^{18, 21} evolution of biochemical reac-
 320 tions subject to microscale energetic constraints,^{54, 55} and cases in which fitness scales with mutation count.^{56, 57} The parameter s in Eq. 14 describes the relative strength of selection on the trait z . For this fitness landscape, the individual coupling coefficients are observed to obey the general relationship $\mathcal{W}_n \sim s^n$ (Appendix H); thus the broadest features of the fitness landscape (lowest order \mathcal{W}_n) contribute the most to directional selection.

325 The upper portion of the figure describes the relative magnitude of the excess force of selection as a function of the trait skewness γ and the selection strength s ; for simplicity, the kurtosis k is held fixed at its theoretical minimum $k = \gamma^2 + 1$. At the origin, the trait distribution is Gaussian and so the cryptic selection force is zero; however, as either the skewness γ or the selection strength s increase, the relative contributions of the non-
 330 Gaussian terms in Eq. 12, Eq. 13 increases. Red regions correspond to cases in which the cryptic selection force is positive (and thus assists directional selection), whereas blue regions correspond to cases in which the non-Gaussian contributions retard directional selection. The plot suggests that positive skewness (corresponding to a trait distribution with a long tail of large trait values) generally speeds the evolutionary dynamics of both the trait mean
 335 and trait variance, primarily due to the added contributions of extremal individuals. The opposite is true for negatively skewed populations in which most individual traits exceed the trait mean, due to outlier individuals producing offspring with lower fitness.

In the lower portions of the figure, trajectories of $\bar{z}(t)$ and $\sigma(t)$ are shown for a numerically-constructed non-Gaussian trait distribution (Appendix M) using (γ, s) values marked by open
 340 circles in the blue and red regions of the upper plots; for comparison, a trajectory consisting

of the “null” case of a Gaussian trait distribution is shown in gray. The dynamics of each distribution relative to the Gaussian case proceed as predicted by the magnitude of the cryptic force, with the primary advantage/detriment due to skewness occurring initially before the rate of evolution eventually stabilizes. Notably, directional selection causes a continued
 345 increase in the mean trait ($\bar{z} \rightarrow \infty$ as $t \rightarrow \infty$), but the trait distribution width σ stabilizes to a constant value that is proportional to the skewness. Thus the effect of non-Gaussian features in the trait distribution may manifest experimentally as a constant variance that is larger or smaller than that predicted under classical evolutionary theory.

In the Appendix, we compare these results to Wright-Fisher simulations of phenotypic
 350 evolution in a population initialized to the same starting values of the mean, variance, skewness, and kurtosis as was used in these equations, and we find general agreement. In the simulations, as in real populations, the values of the skewness and kurtosis drift over time due to accumulated sampling errors (this corresponds to a timescale over which the “fixed cumulant” assumption above is no longer valid). However we observe that trait means and
 355 variances tend to drift monotonically under exponential directional selection, and so even over long timescales the calculated direction of the cryptic force of selection correctly predicts the dynamics relative to the Gaussian distribution.

4.2 Transient evolutionary responses to stabilizing and disruptive selection

360 In addition to generating qualitatively different dynamics, non-Gaussian features of the trait distribution may affect the long-term duration and direction of natural selection. In order to illustrate this effect in a fitness landscape with a variable number of local maxima, we next consider a general fitness landscape described by an arbitrary polynomial,

$$W(z) = W_0 + \sum_{m=1}^{\infty} \alpha_m (z - c\bar{z})^m, \quad (15)$$

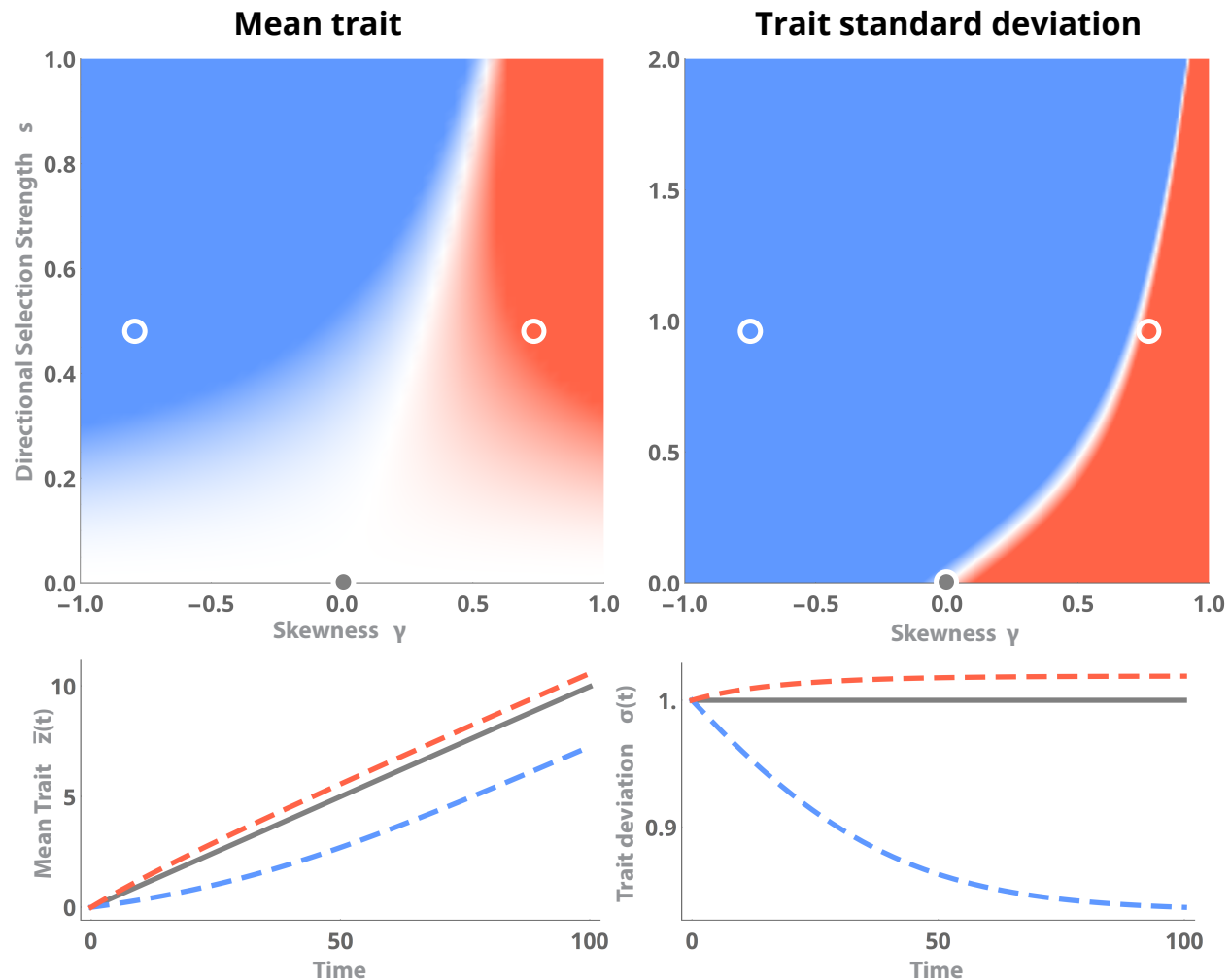


Figure 1 Cryptic forces of selection under directional selection. Top: The excess force of selection due to the non-Gaussian form of the trait distribution, for the mean trait dynamics (left) and trait standard deviation dynamics (right). Colored shading indicates the relative direction and magnitude of the cryptic terms in Eq. 12 and Eq. 13, with red (blue) indicating cryptic forces that accelerate (retard) the growth of each moment. Lower plots represent example dynamics of the mean and standard deviation for representative points on each color plot. White circles on upper plots indicate parameter values (γ, s) used for the trajectories in the lower plots: the grey trace indicates classical fitness gradient dynamics with an unperturbed Gaussian trait distribution, while red and blue traces indicate cases in which higher trait moments speed or hinder the evolutionary dynamics, respectively.

where the relative magnitudes of the various coefficients α_m determine the number of local
 365 maxima and minima of the landscape. In general, if the largest non-zero α_m is positive,
 then for large values the fitness landscape looks like a “U” ($\lim_{z \rightarrow \infty} W(|z|) = \infty$), and so the
 value of the mean trait \bar{z} eventually diverges: natural selection proceeds indefinitely in the
 system as the mean trait continuously increases. Sub-leading terms in the polynomial Eq. 15
 induce short term transients that may affect the dynamics only temporarily depending on
 370 the initial conditions. Globally, however, natural selection will proceed continuously in a U-
 shaped landscape, as has been reported experimentally.⁵⁸ Conversely, if the largest non-zero
 α_m is negative, then the fitness landscape looks like a hill at large z and so the the mean trait
 and trait variance will always eventually equilibrate at an intermediate stable solution—in
 which case natural selection proceeds transiently until this solution is reached.⁵⁹

375 We can investigate the manner in which skewness and kurtosis affect the timescale of
 natural selection by studying the specific case of a quartic fitness landscape under transient
 natural selection ($\alpha_m = 0$ for $m > 4$; $\alpha_4 < 0$). This landscape can have either one or two
 local maxima depending on the relative magnitudes of α_2 and α_4 in Eq. 15. Thus for a
 continuous, unbounded trait, a quartic fitness landscape represents the simplest landscape
 380 that can model both stabilizing (one maxima) or disruptive (two maxima) selection.³⁴

For each value of the skewness γ and kurtosis k , the equilibrium solutions of Eq. 12 and
 Eq. 13 are independent of the initial conditions and can be found analytically along with
 the Jacobian matrix describing their local stability. Figure 2 shows plots of the maximum
 eigenvalue of this Jacobian matrix associated with the intermediate phenotype, parametrized
 385 by the skewness γ and kurtosis k (the kurtosis lower bound $k > \gamma^2 + 1$ is indicated by a
 solid gray line). For cases in which the dynamical equations yield multiple solutions, the
 intermediate equilibrium phenotype is defined as the solution of Eq. 12 and Eq. 13 with \bar{z}
 closest to 0. In the blue regions on the plots, this intermediate phenotype is stable and so the
 mean trait approaches this point. In the red regions of the plots, the intermediate phenotype
 390 is unstable and so the mean trait instead approaches another, extremal equilibrium point.

The relative darkness of the plot colors represents the relative speed of the evolutionary dynamics: darker blue indicates that the intermediate phenotype is achieved relatively quickly, while darker red indicates that extremal phenotypes are reached quickly.

The points $\gamma = 0, k = 3$ on each plot correspond to the default case of a purely Gaussian trait distribution with a time-varying mean and standard deviation. For stabilizing selection this point resides within a blue region, consistent with the intuitive result that a Gaussian trait distribution evolves towards the location of the global maximum of the fitness landscape, and that the range of trait values decreases over time as more and more individuals in the population approach this optimum ($\sigma(t) \rightarrow 0$).⁵⁴ Indeed, for stabilizing selection, non-Gaussian features in the trait distribution only barely affect the rate of the dynamics, consistent with experimental results suggesting that traits under stabilizing selection universally attain intermediate optima.⁶⁰ Additionally, this suggests that if γ and k were themselves dynamical variables that changed either due to selection or mating effects, the mean fitness would nonetheless always reach a finite value dictated by the maximum of the fitness distribution.

In a disruptive landscape, the intermediate phenotype is disfavored for almost all trait distributions near the classical case of a Gaussian distribution at $\gamma = 0, k = 3$ (red regions in Figure 2). This is expected because the maxima of the disruptive landscape occur away from the intermediate phenotype at $z = 0$, and so the trait moves towards these dispersed values in the absence of additional destabilizing dynamics. However, when the trait distribution is strongly non-Gaussian, the intermediate phenotype near $z = 0$ regains stability, an effect that is particularly pronounced when the trait distribution is strongly asymmetric and flat (large γ , $k \approx \gamma^2 + 1$). High skewness and kurtosis correspond to the trait distribution containing a relatively large fraction of individuals with extremal phenotypes, which represent a large enough trait range that the distance between the two symmetric equilibria in the disrupting fitness landscape becomes relatively insignificant (the locations of these off-center equilibria depend on α_2/α_4 in Eq. 15). As a result, the overall average phenotype returns to the center

due to the leading-order effect of the negative α_4 term in Eq. 15.

In the Appendix, we compare the dynamics predicted by local stability analysis of Eq. 12
 420 and Eq. 13 with the dynamics of a population subject to Wright-Fisher dynamics with match-
 ing starting cumulants as used here. We find that the short-time dynamics of the Wright-
 Fisher process are directly analogous to those predicted by the local analysis, particularly
 because additional effects (such as variation in the genetic variance V , and fluctuations in the
 values of higher cumulants due to sampling drift) do not manifest over the infinitesimally-
 425 short timescales which local stability analysis applies.

Significant skewness and flatness in the trait distribution can obscure the effects of dis-
 ruptive selection and potentially serve as a mechanism to preserve or enhance phenotypic
 variance, even when selection itself acts on the trait variance—potentially implicating higher
 trait moments in speciation and ecological phenomena that depend on the phenotypic vari-
 430 ance.¹⁹ Although the dynamical equations would become substantially more complex if the
 skewness and kurtosis also responded to selection, results from other cumulant dynamical
 systems^{39, 40} suggest that genetic or mating processes that preserve an arbitrarily high or-
 der moment would result in similar “freezing” of the dynamics of lower-order cumulants.
 Thus, even if natural selection alters arbitrarily high moments of the trait distribution, if
 435 mutation or mating serve as a “source” that constrains an arbitrary moment of the trait
 distribution, then lower-order moments such as the phenotypic variance would also be pre-
 vented from reaching zero in certain regions of parameter space. This process represents a
 generalization of the concept of a “mutation-selection balance,” a concept typically invoked
 in order to justify holding the phenotypic variance fixed during selection.^{15, 36, 46} However,
 440 as with the traditional breeder’s equation, an additional equation specifically incorporating
 the underlying genetics would be required in order to justify this process biologically.⁴⁷

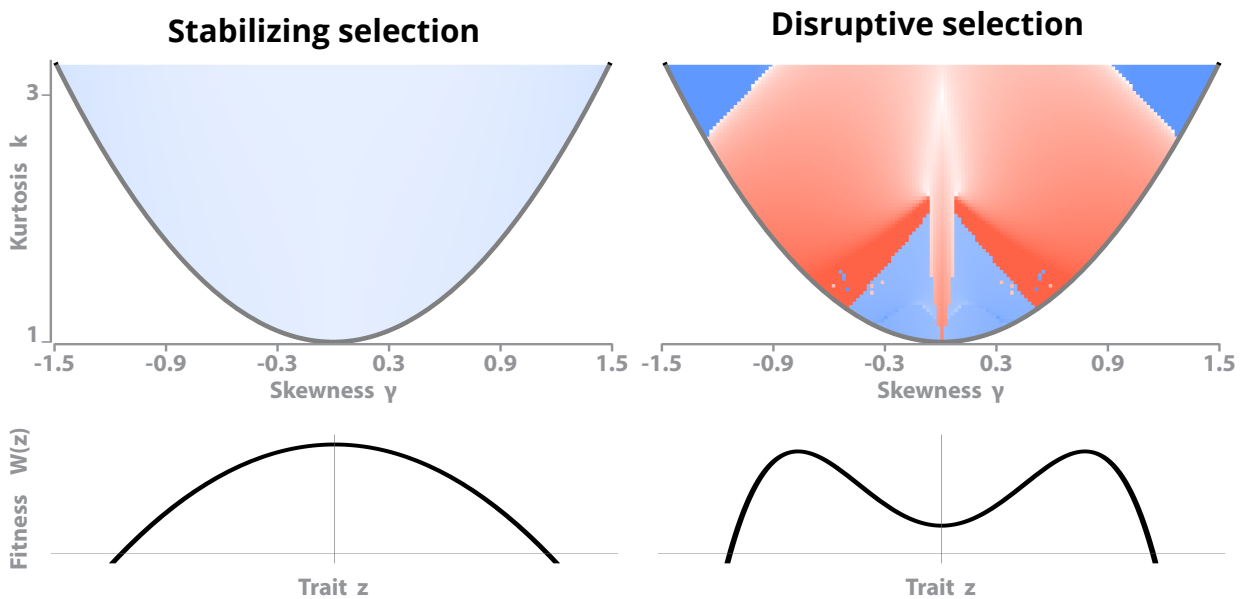


Figure 2 Evolutionary dynamics under stabilizing or disruptive selection. Colored shading represents the maximum eigenvalue associated with the Jacobian matrix of $\dot{\bar{z}}, \dot{\sigma}$ on a logarithmic scale, evaluated at the equilibrium nearest to $\bar{z} = 0$ (the intermediate phenotype) and parametrized by the skewness and kurtosis of the trait distribution (all other parameters are held constant). Negative maximum eigenvalues (blue regions) represent dynamics that eventually converge to the intermediate phenotype, positive maximum eigenvalues (red regions) represent dynamics that approach an extremal phenotype, and the intensity of shading indicates the instantaneous rate of the dynamics. The solid gray line indicates the lowest mathematically-valid value for the kurtosis, $k > \gamma^2 + 1$. Beneath each figure is a diagram of the fitness landscape $W(z)$ that produced it. For this figure, $c = 1$, $V = 1$, $U = 10$ in Eq. 12 and Eq. 13. Shading ranges from eigenvalue values of -2 (dark blue) to 2 (dark red).

4.3 Time-variation of the heritability

Experimental studies of phenotypic evolution in artificial selection regimes generally quantify genetic effects using the narrow sense heritability, $h(t)^2 \equiv V/\sigma(t)^2$, defined as the ratio of additive genetic variance V to overall phenotypic variance σ^2 . In large populations, h^2 changes slowly enough that it can be estimated without the need for explicit identification of a trait's genetic origin. However, recent experimental results have suggested that h^2 may change appreciably during short periods of strong selection, especially when the underlying genetics (and thus V) exhibit complex dynamics.²¹ In particular, rapid changes in the phenotypic variance σ^2 may underly this phenomenon over short timescales,^{19, 60} even when an insufficient number of generations have elapsed for V to change appreciably.

We note that by including changes in higher moments than the mean, our framework naturally describes changes in the observed heritability arising from the dynamics of trait variation. In order to study this effect, we re-parametrize our model in terms of quantities more readily measured experimentally: our two dynamical variables $\bar{z}(t)$ and $\sigma(t)$ may be replaced by the equivalent conjugate variables of heritability $h^2(t)$ and mean fitness $\bar{W}(t)$ (see Appendix G). Because the exact underlying trait values or distributions may not be accessible in certain experimental contexts, these variables are more descriptive of macroscopic population trends in population-level assays, such as long-term bacterial evolution experiments.⁶¹

In Figure 3, we show the short-time evolutionary dynamics of the heritability under a linear directional fitness landscape,

$$W(z) = \alpha_1(z - c\bar{z}), \quad (16)$$

for which the coupling coefficients are computed in Appendix K. Because any arbitrary, smooth fitness landscape may be approximated as linear in the neighborhood of the mean trait, this parametrization illustrates a general relationship between the heritability and mean

fitness dynamics over short timescales. Setting $\alpha_1 > 0$ results in positive directional selection ($\dot{\bar{W}} > 0$, $\dot{\bar{z}} > 0$) because regions of a fitness landscape with positive slope tend to drive a population towards a higher mean fitness. The solutions to Eq. 12, Eq. 13 corresponding to each timepoint are shown in Figure 3A, where they are parametrized in terms of the heritability as a function of mean fitness (which itself increases in time). In the figure, different traces correspond to various values of the skewness γ and kurtosis k .

In general, a Gaussian trait distribution ($\gamma = 0, k = 3$ in Eq. 12 and Eq. 13) always produces the default case of constant heritability ($\dot{h}^2 = 0$, gray dashed line in Figure 3A). An evolving population with positive skewness (blue traces) exhibits heritability that eventually decreases in time, primarily due to a high fraction of outlier individuals in the high-fitness tail of the trait distribution. These individuals produce higher-fitness offspring quickly enough that the overall trait range increases in time, leading to a corresponding decrease in the overall heritability—which is consistent with several field studies showing that increasing mean fitness also increases trait variation and lowers the observed heritability.^{60, 62} This growth of the variance due to tail effects is consistent with prior analytic solutions for the case of directional selection;^{37, 63} however we note that we hold the skewness and higher moments fixed in these calculations. Intriguingly, we find that high values of kurtosis retard this process at short timescales, producing some scenarios where the heritability appears to increase transiently, before eventually relaxing non-monotonically to zero over longer timescales (topmost blue trace). Such non-monotonicity in the heritability under directional selection would be an observable experimental signature of non-Gaussian dynamics.

Conversely, a population with negative skewness (red traces in Figure 3A) has a long tail of individuals with comparatively low fitness; these individuals serve to counteract the tendency of directional selection to increase trait ranges, and thus cause the trait variation to decrease in time—leading to an accompanying increase in the trait heritability. Because this effect is driven by the lower-fitness tail of the distribution, it may partly explain experimental results that have reported a disproportionately large contribution of rare phenotypes to

the heritability of certain deleterious traits.^{64, 65} The differential effect of left and right skewness is further apparent in the trait distributions at three representative timepoints in the dynamics shown in Figure 3B, where the apparent width of the trait distribution varies non-monotonically and causes continuous changes in the heritability observed in the upper portion of the figure. This substantial variation in the distribution’s width at various points in the dynamics may confound efforts to study experimentally the evolution of ecological niche width,⁶⁶ as non-Gaussian features in the trait distribution may cause transient contraction and expansion of the observed trait range, even in the absence of competition.

In the Appendix, we perform a comparable set of simulations using Wright-Fisher dynamics, and we find general agreement in the dynamics, including non-monotonicity in the dynamics at large values of k , as well as a qualitative shift from $h^2 \rightarrow 1$ to $h^2 \rightarrow 0$ when γ changes sign from negative to positive.

5 Discussion

We have presented a formulation of phenotypic evolution that seeks to iteratively relate the dynamics of an arbitrary trait distribution to an arbitrary fitness landscape. This simplified model explains several phenomena observed in numerical simulations of non-Gaussian trait distributions evolving in simple fitness landscapes. We have shown that skewness or asymmetry in the trait distribution can delay directional selection or prevent disruptive selection, and that the heritability—typically assumed to be constant over short timescales—can vary in time with complex, non-monotonic dynamics as the mean fitness changes.

Our results are consistent with empirical and theoretical analyses that formulate phenotypic evolution in terms of the probability distribution of fitness values $p(W)$ rather than in terms of an explicit trait distribution $p(z)$ and auxiliary fitness function $W(z)$.⁴⁹ However, the trait-based model here is more readily applied to arbitrarily complex fitness functions (particularly non-invertible and multimodal fitness landscapes),⁴⁶ and thus may assist the

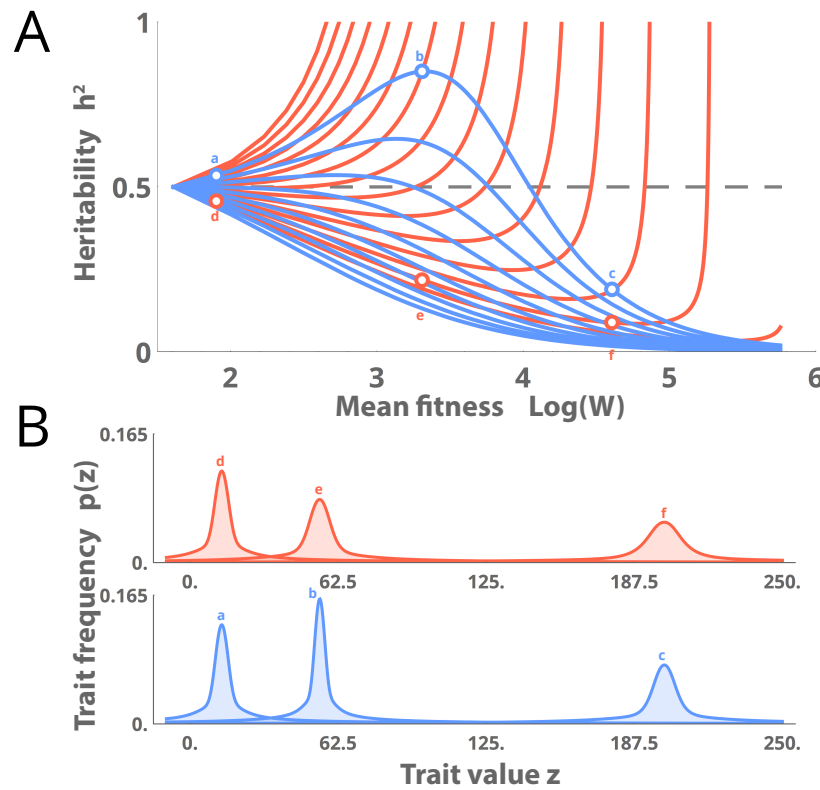


Figure 3 Time-dependent heritability under directional selection. (A) The heritability h^2 as a function of the mean fitness, which increases continuously over time in a directional fitness landscape. Colors represent cases in which the trait distribution skewness is $\gamma = -0.05$ (red) or $\gamma = 0.05$ (blue). Different traces of the same color correspond to increasing values of the kurtosis k in the range $k = \gamma^2 + 1$ to $k = \gamma^2 + 6$, with slower timescales (lower traces) corresponding to larger values of the kurtosis. The gray dashed line corresponds to evolutionary dynamics with constant heritability, which the dynamical equations recreate when $\gamma = 0, k = 3$. (B) The trait distributions for three representative timepoints (and thus values of \overline{W}) marked by open circles in (A). For this figure, $\alpha = V = 1, U = 0.01, c = 0.5$.

study of explicitly experimental quantities such as selection coefficients and realized heritability. For fitness distribution models, similar cumulant dynamical equations have been
 520 derived under distinct constraints and approximations, and for certain types of experiments (e.g., microbiological lineage assays) these formalisms may be preferable.^{36, 40, 46, 61}

Our explicitly phenotypic approach does not involve an explicit underlying genetic model; we assume a linear heritability relation and note that this approximation holds over the relatively short timescales and slow selection regimes observed in our simulations.²⁸ A more
 525 detailed model would include explicit information about the mechanics of inheritance, and how these parameters contribute to the breeder's equation and determine the parameters it contains. An important starting point for such work would be models of non-Gaussian evolution based on a multilocus genetic models,^{37, 47} which have shown that the dynamics of the phenotypic moments may vary appreciably depending on mating effects, transmission
 530 effects, and whether selection occurs before or after transmission²³

Other potential improvements include alternative series closure schemes to raw truncation that depend on the type of fitness landscape being evaluated, as well as additional cross terms in the dynamical equations that account for assortative mating effects. Additionally, we have assumed that the additive genetic variance remains fixed during the phenotypic dynamics;
 535 however over long timescales natural selection and mating may affect this value considerably. Coupling genotypes and phenotypes may also require the phenotypic portion of the model to be generalized for multivariate traits by defining a fully time-dependent phenotypic variance-covariance matrix, which may be particularly important for selection experiments in which the traits with the strongest selection responses are unknown a priori.^{13, 27}

Another limitation of our approach comes from the truncation approximations necessary
 540 to make the dynamical equations closed-form. Here, we have chosen canonical directional and minimal stabilizing/disruptive fitness landscapes, in order to illustrate the simplest non-trivial dynamics arising from our model and to justify our inclusion of only the skewness and kurtosis (which represent the leading-order departures from Gaussian trait distributions).

545 Additionally, we have retained only two of the dynamical equations in our model: one
for the mean trait and one for the trait variance. But for certain types of initial trait
distribution or fitness landscape, the leading order equations Eq. 12 and Eq. 13 may not
be sufficiently accurate, and instead the full dynamical equations Eq. 7, Eq. 10, and their
equivalents for higher cumulants may be necessary. The number of necessary dynamical
550 equations, and the number of terms to include in them, depends primarily on the form of
the coupling coefficients \mathcal{W}_n , which may be found as long as the fitness landscape is known.
A more detailed fitness landscape, such as one derived empirically, may require retention
of more terms in the series \mathcal{W}_n , because higher-order terms correspond to finer-scale fitness
landscape variations which affect dynamics over smaller length and timescales. However,
555 we note that because computing \mathcal{W}_n simply requires projecting a known fitness landscape
onto a polynomial basis, these coupling coefficients may easily be found numerically if the
parameter values for the initial trait distribution and fitness landscape are known.

6 Acknowledgements

W. G. thanks the NDSEG Fellowship program for support.

7 Competing interests

We declare no competing interests.

References

- [1] Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. Evaluation of the rate of evolution in natural populations of guppies (*poecilia reticulata*). *Science* **275**, 1934–1937 (1997).
- [2] Grant, P. R. & Grant, B. R. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution* **49**, 241–251 (1995).
- [3] Gingerich, P. D. Mammalian responses to climate change at the Paleocene-Eocene boundary: Polecat Bench record in the northern Bighorn Basin, Wyoming. *Special Papers-Geological Society of America* 463–478 (2003).
- [4] Gingerich, P. D. Paleontology and phylogeny: patterns of evolution at the species level in early tertiary mammals. *American Journal of Science* **276**, 1–28 (1976).
- [5] Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution **1**, 356–366 (1932).
- [6] Mustonen, V. & Lässig, M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends in Genetics* **25**, 111–119 (2009).
- [7] Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383 (2007).

- [8] Hartl, D. L. What can we learn from fitness landscapes? *Current Opinion in Microbiology* **21**, 51–57 (2014).
580
- [9] Martin, C. H. & Wainwright, P. C. Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science* **339**, 208–211 (2013).
- [10] Shen, B., Dong, L., Xiao, S. & Kowalewski, M. The avalon explosion: evolution of ediacara morphospace. *Science* **319**, 81–84 (2008).
- [11] Miles, D. B. The race goes to the swift: fitness consequences of variation in sprint
585 performance in juvenile lizards. *Evolutionary Ecology Research* **6**, 63–75 (2004).
- [12] Björklund, M., Husby, A. & Gustafsson, L. Rapid and unpredictable changes of the g-matrix in a natural bird population over 25 years. *Journal of Evolutionary Biology* **26**, 1–13 (2013).
- [13] Stepan, S. J., Phillips, P. C. & Houle, D. Comparative quantitative genetics: evolution
590 of the g matrix. *Trends in Ecology & Evolution* **17**, 320–327 (2002).
- [14] Arnold, S. J. Constraints on phenotypic evolution. *The American Naturalist* **140**, S85–S107 (1992).
- [15] Lande, R. Natural selection and random genetic drift in phenotypic evolution. *Evolution*
595 314–334 (1976).
- [16] Abrams, P. A. & Matsuda, H. Fitness minimization and dynamic instability as a consequence of predator–prey coevolution. *Evolutionary Ecology* **11**, 1–20 (1997).
- [17] Cortez, M. H. & Weitz, J. S. Coevolution can reverse predator–prey cycles. *Proceedings of the National Academy of Sciences* **111**, 7486–7491 (2014).
- [18] Karlin, S. *Non Gaussian Phenotypic Models of Quantitative Traits* (Morrison Institute for Population and Resource Studies at Stanford, 1987).
600

- [19] Bolnick, D. I. *et al.* Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution* **26**, 183–192 (2011).
- [20] Waxman, D. & Gavrillets, S. 20 questions on adaptive dynamics. *Journal of Evolutionary Biology* **18**, 1139–1154 (2005).
- [21] Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, 1998).
URL <https://books.google.com/books?id=UhCCQgAACAAJ>.
- [22] Barton, N. H. & Polechová, J. The limitations of adaptive dynamics as a model of evolution. *Journal of Evolutionary Biology* **18**, 1186–1190 (2005).
- [23] Turelli, M. & Barton, N. Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? *Genetics* **138**, 913–941 (1994).
- [24] Nowak, M. A. & Sigmund, K. Evolutionary dynamics of biological games. *Science* **303**, 793–799 (2004).
- [25] Sato, K. & Kaneko, K. Evolution equation of phenotype distribution: general formulation and application to error catastrophe. *Physical Review E* **75**, 061909 (2007).
- [26] Sasaki, A. & Ellner, S. The evolutionarily stable phenotype distribution in a random environment. *Evolution* **49**, 337–350 (1995).
- [27] Lande, R. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–416 (1979).
- [28] Bulmer, M. The effect of selection on genetic variability. *The American Naturalist* **105**, 201–211 (1971).
- [29] Turelli, M. Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**, 1342–1347 (1988).

[30] Bulmer, M. G. *et al.* *The mathematical theory of quantitative genetics*. (Clarendon Press., 1980).

[31] Falconer, D. S. *Introduction to Quantitative Genetics* (Pearson Education, 1981), second edn.

[32] Jones, A. G., Arnold, S. J. & Bürger, R. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**, 1747–1760 (2003).

[33] Garant, D., Hadfield, J. D., Kruuk, L. E. & Sheldon, B. C. Stability of genetic variance and covariance for reproductive characters in the face of climate change in a wild bird population. *Molecular Ecology* **17**, 179–188 (2008).

[34] Gilpin, W. & Feldman, M. W. A phase transition induces chaos in a predator-prey ecosystem with a dynamic fitness landscape. *PLoS Computational Biology* **13**, e1005644 (2017).

[35] Whittle, P. On the use of the normal approximation in the treatment of stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)* 268–281 (1957).

[36] Smerlak, M. & Youssef, A. Limiting fitness distributions in evolutionary dynamics. *Journal of Theoretical Biology* **416**, 68–80 (2017).

[37] Bürger, R. Moments, cumulants, and polygenic dynamics. *Journal of mathematical biology* **30**, 199–213 (1991).

[38] Babusci, D., Dattoli, G. & Quattromini, M. On integrals involving hermite polynomials. *Applied Mathematics Letters* **25**, 1157–1160 (2012).

[39] Rattray, M. & Shapiro, J. L. Cumulant dynamics of a population under multiplicative selection, mutation, and drift. *Theoretical Population Biology* **60**, 17–31 (2001).

- [40] Turelli, M. & Barton, N. Dynamics of polygenic characters under selection. *Theoretical Population Biology* **38**, 1–57 (1990).
- 650 [41] Berberan-Santos, M. N. Expressing a probability density function in terms of another pdf: A generalized gram-charlier expansion. *Journal of Mathematical Chemistry* **42**, 585–594 (2007).
- [42] Prügel-Bennett, A. & Shapiro, J. L. The dynamics of a genetic algorithm for simple random ising systems. *Physica D: Nonlinear Phenomena* **104**, 75–114 (1997).
- 655 [43] Matis, J. H. & Kiffe, T. R. Effects of immigration on some stochastic logistic models: a cumulant truncation analysis. *Theoretical Population Biology* **56**, 139–161 (1999).
- [44] Krishnarajah, I., Cook, A., Marion, G. & Gibson, G. Novel moment closure approximations in stochastic epidemics. *Bulletin of Mathematical Biology* **67**, 855–873 (2005).
- [45] Barton, N. & Turelli, M. Natural and sexual selection on many loci. *Genetics* **127**, 229–255 (1991).
- 660 [46] Neher, R. A. & Shraiman, B. I. Statistical genetics and evolution of quantitative traits. *Reviews of Modern Physics* **83**, 1283 (2011).
- [47] Barton, N. & Turelli, M. Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genetics Research* **49**, 157–173 (1987).
- 665 [48] Zeng, Z.-B. Genotypic distribution at the limits to natural and artificial selection with mutation. *Theoretical population biology* **32**, 90–113 (1987).
- [49] Good, B. H. & Desai, M. M. Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. *Theoretical Population Biology* **85**, 86–102 (2013).
- 670 [50] Rogers, A. & Prügel-Bennett, A. Evolving populations with overlapping generations. *Theoretical population biology* **57**, 121–129 (2000).

[51] Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).

[52] Lande, R. Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences* **78**, 3721–3725 (1981).

[53] Balagam, R., Singh, V., Sagi, A. R. & Dixit, N. M. Taking multiple infections of cells and recombination into account leads to small within-host effective-population-size estimates of hiv-1. *PLoS One* **6**, e14531 (2011).

[54] Tsimring, L. S., Levine, H. & Kessler, D. A. Rna virus evolution via a fitness-space model. *Physical Review Letters* **76**, 4440 (1996).

[55] Haldane, A., Manhart, M. & Morozov, A. V. Biophysical fitness landscapes for transcription factor binding sites. *PLoS Computational Biology* **10**, e1003683 (2014).

[56] Tripathi, K., Balagam, R., Vishnoi, N. K. & Dixit, N. M. Stochastic simulations suggest that hiv-1 survives close to its error threshold. *PLoS Computational Biology* **8**, e1002684 (2012).

[57] Saakian, D. B., Ghazaryan, M. H. & Hu, C.-K. Punctuated equilibrium and shock waves in molecular models of biological evolution. *Physical Review E* **90**, 022712 (2014).

[58] Kupriyanova, E. K. What do egg size distributions in marine invertebrates tell us about validity of fecundity-time models? *Marine Ecology* **35**, 249–253 (2014).

[59] Cortez, M. H. & Ellner, S. P. Understanding rapid evolution in predator-prey interactions using the theory of fast-slow dynamical systems. *The American Naturalist* **176**, E109–E127 (2010).

[60] Hoffmann, A. A. & Merilä, J. Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution* **14**, 96–101 (1999).

- 695 [61] Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* (2017).
- [62] Kruuk, L. E. *et al.* Heritability of fitness in a wild mammal population. *Proceedings of the National Academy of Sciences* **97**, 698–703 (2000).
- [63] Bürger, R. & Lande, R. On the distribution of the mean and variance of a quantitative
700 trait under mutation-selection-drift balance. *Genetics* **138**, 901–912 (1994).
- [64] Mancuso, N. *et al.* The contribution of rare variation to prostate cancer heritability. *Nature Genetics* **48**, 30 (2016).
- [65] Schork, N. J., Murray, S. S., Frazer, K. A. & Topol, E. J. Common vs. rare allele hypotheses for complex diseases. *Current Opinion in Genetics & Development* **19**,
705 212–219 (2009).
- [66] Bolnick, D. I. *et al.* Ecological release from interspecific competition leads to decoupled changes in population and individual niche width. *Proceedings of the Royal Society of London B: Biological Sciences* **277**, 1789–1797 (2010).
- [67] Slatkin, M. Selection and polygenic characters. *Proceedings of the National Academy of Sciences* **66**, 87–93 (1970).
710
- [68] Oster, G., Ipaktchi, A. & Rocklin, S. Phenotypic structure and bifurcation behavior of population models. *Theoretical Population Biology* **10**, 365–382 (1976).
- [69] Bürger, R. Predictions of the dynamics of a polygenic character under directional selection. *Journal of Theoretical Biology* **162**, 487–513 (1993).
- 715 [70] Prügel-Bennett, A. Modelling evolving populations. *Journal of Theoretical Biology* **185**, 81–95 (1997).

Contents

	1 Introduction	2
	2 Model	4
720	2.1 Trait mean dynamics in an arbitrary phenotype distribution	5
	2.2 Trait variance dynamics	9
	2.3 Mean fitness dynamics	10
	2.4 Gradient dynamics with leading-order corrections	11
	3 Assumptions and related models	13
725	4 Results	14
	4.1 Cryptic forces of selection arise from non-Gaussian trait distributions	14
	4.2 Transient evolutionary responses to stabilizing and disruptive selection . . .	16
	4.3 Time-variation of the heritability	22
	5 Discussion	24
730	6 Acknowledgements	28
	7 Competing interests	28
	8 Appendix	36
	A Supplementary Figures	36
	B Dynamical equations for cumulants and moments	36
735	C Simplification of trait dynamics when $W(z) = W(z - c\bar{z})$, $p(z) = p(z - \bar{z})$. .	39
	D Behavior of $\dot{\bar{z}}$ as $c \rightarrow 1$	40
	E Derivation of trait variance dynamics $\dot{\sigma}$	42
	F Dynamics of mean fitness	43
	G Formulation of dynamics in terms of heritability $h^2(t)$ and mean fitness $\bar{W}(t)$	45

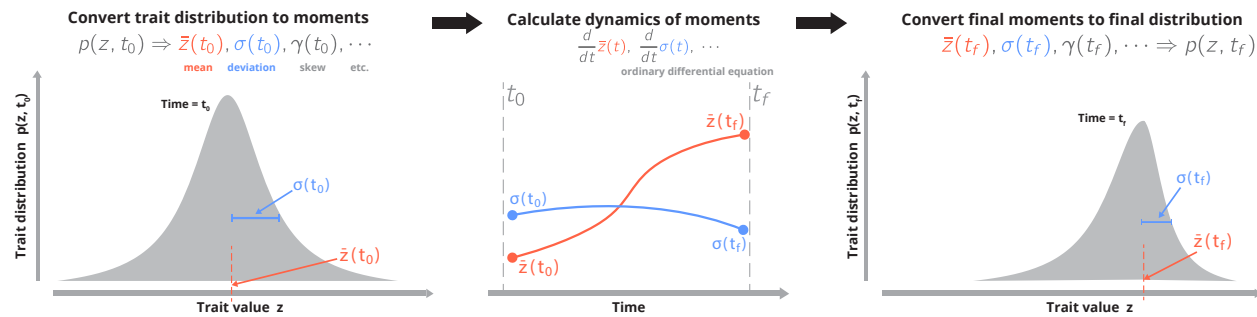


Figure S1 Overview of model approach. (Left) An initial distribution of traits in a population is parametrized into an arbitrary number of moments (the mean trait, trait range, skew, kurtosis, etc), which, collectively, describe the full state of the population at the initial time t_0 . (Middle) Using the breeder's equation, a known fitness landscape, and a series expansion of the trait distribution, a set of ordinary differential equations is derived that describes the time-evolution of each moment of the trait distribution, with the moment values in the initial trait distribution acting as initial conditions. (Right) The solutions of these differential equations at some later time may then be used to reconstruct the full trait distribution at that time.

740	H	\mathcal{W}_n for exponential directional selection	47
	I	Orthogonality of Hermite polynomials and shifted Hermite polynomials . . .	48
	J	Orthogonality of Hermite polynomials and regular polynomials	49
	K	\mathcal{W}_n for polynomial fitness landscapes	50
	L	Comparison with other models of cumulant dynamics	50
745	M	Numerical solution of cumulant dynamical equations	56
	N	Numerical simulations of selection on individuals	58

8 Appendix

A Supplementary Figures

B Dynamical equations for cumulants and moments

750 Standard derivations based on non-overlapping generations (discrete time) start with the breeder's equation, which supposes that the mean trait distribution in the next generation

depends linearly on the effect of selection on the current generation,

$$\bar{z}(t+1) = h^2 \bar{z}_w(t) + (1 - h^2) \bar{z}(t), \quad (\text{A1})$$

where the linear parameter h^2 represents the narrow-sense heritability of the trait z . Rearranging the terms in this equation and taking the limit of vanishingly small generation time
755 leads to the dynamical equation,

$$\dot{\bar{z}} = h^2 (\bar{z}_w(t) - \bar{z}(t))$$

where the mean reproductive rate has been used to make time dimensionless.

One way of deriving Eq. A1 involves expressing the full distribution $p(z, t+1)$ as an integral over the distribution at the previous timestep, $p(z, t)$. Various “survival kernels” and “mating kernels” are included within this integral, and they determine the contribution
760 of mating preference, fitness, and even migration to the form of the trait distribution.^{67, 68} In these models, the linear form of Eq. A1 originates from the convolution of multiple probability distributions, which in general will result in a new distribution with a mean given by a linear combination of the means of the original distributions.

In general, the convolution of two arbitrary distributions has a distribution with cumu-
765 lants given by the sum of the cumulants of the original variables’ distributions. This property is responsible for the linearity of Eq. A1, and its importance for the generation of cumulant dynamical equations for evolution has been noted by Bürger.³⁷ For this reason, the analogy to Eq. A1 for higher-order cumulants is given by,

$$\dot{\kappa}_i = h_i^2 (\kappa_{iw} - \kappa_i) \quad (\text{A2})$$

where h_i^2 represents a generalization of the heritability for higher-order cumulants ($h_1^2 = h^2$
770 is the standard narrow-sense heritability). This equation represents an extension of the

implicit assumption in the breeder's equation that the trait response depends linearly on the strength of selection: if the means respond linearly, then other cumulants must as well as long as the trait has a well-defined probability distribution in the current generation. Linear heritability equations for the variance in the form of Eq. A2 have previously been derived explicitly for the "infinitesimal" genetic model (in which a given continuous trait arises from the independent contributions of infinitely many distinct loci) by Bulmer.²⁸ Similar relations of this form are found by Turelli and Barton in both few-locus and infinitesimal genetic models (See Eq. 45 of Turelli and Barton 1994,²³ and Eq. 3.2a,b of Turelli and Barton 199040).

While the cumulants of a random variable have the advantage of being additive, the moments of a probability distribution often provide a more familiar description of the variable. Fortunately, the dynamical equations Eq. A2 may readily be converted into dynamical equations for the central moments (μ_i) using standard reference formulas; we note however that the first three cumulants are directly proportional to the first three central moments:

$$\kappa_1 = \bar{z}$$

$$\kappa_2 = \sigma^2$$

$$\kappa_3 = \mu_3$$

The fourth and higher cumulants have more complicated expressions given by Bell polynomials, $B_{n,k}$, of the various lower-order central moments,

$$\begin{aligned}\mu_n &= \sum_{k=1}^n B_{n,k}(0, \kappa_2, \dots, \kappa_{n-k+1}) \\ \kappa_n &= \sum_{k=1}^n (-1)^{k-1} (k-1)! B_{n,k}(0, \mu_2, \dots, \mu_{n-k+1})\end{aligned}\tag{A3}$$

These equations may be inserted directly into Eq. A2 to obtain dynamical equations in terms of central moments. Dynamical closure in this context thus represents an assumption that

cumulants above a certain order are not heritable: $h_i^2 = 0$ for some $i > 1$. This results in the infinite tower of coupled moment equations given by Eq. A2 becoming finite, because the explicit form of the n^{th} cumulant only depends on the n^{th} and smaller central moments.

785 The heritability requirements above suggest how each cumulant of an arbitrary trait distribution may have its own dynamical equation. Importantly, each of these dynamical equations may itself depend on an arbitrary number of cumulants of the trait distribution, depending on the current form of the distribution. This is because an arbitrary trait distribution may be expressed in terms of a Gram-Charlier A series,

$$p(z) = \mathcal{N}[\bar{z}, \sigma](z) \left(1 + \sum_{n=3}^{\infty} c_n H e_n \left(\frac{z - \bar{z}}{\sigma} \right) \right) \quad (\text{A4})$$

790 where each c_n is defined explicitly in terms of the moments or cumulants of $p(z)$,

$$c_n \equiv (1 - \delta_{n,2}) \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-\frac{1}{2})^j}{(j!(n-2j)!) \sqrt{\kappa_2}^{n-2j}} \mu_{n-2j}.$$

Here $\delta_{n,m}$ is the Kronecker delta function ($\delta_{n,m} = 1$ when $n = m$; $\delta_{n,m} = 0$ otherwise) and the central moments μ_j are given in terms of the cumulants by Eq. A3.

C Simplification of trait dynamics when $W(z) = W(z - c\bar{z})$, $p(z) = p(z - \bar{z})$

795 The mean fitness is given by

$$\bar{W} = \int W(z) p(z) dz, \quad (\text{A5})$$

Taking the derivative of both sides of this equation,

$$\frac{d\bar{W}}{d\bar{z}} = \int \frac{dp(z)}{d\bar{z}} W(z) dz + \int \frac{dW(z)}{d\bar{z}} p(z) dz.$$

We assume that $W(z) = W(z - c\bar{z})$,

$$\int p(z) \frac{dW(z)}{d\bar{z}} dz = -c \int p(z) \frac{dW(z)}{dz} dz.$$

Performing integration by parts,

$$\int p(z) \frac{dW(z)}{d\bar{z}} dz = -c \left((W(z)p(z)) \Big|_{-\infty}^{\infty} - \int \frac{dp(z)}{dz} W(z) dz \right).$$

The first term on the left hand side vanishes due the compactness of $p(z)$,

$$\int p(z) \frac{dW(z)}{d\bar{z}} dz = c \int \frac{dp(z)}{dz} W(z) dz.$$

800 If $p(z) = p(z - \bar{z})$, then,

$$\int p(z) \frac{dW(z)}{d\bar{z}} dz = -c \int \frac{dp(z)}{d\bar{z}} W(z) dz.$$

Inserting this expression into Eq. A5 results in a final equation,

$$\frac{d\bar{W}}{d\bar{z}} = (1 - c) \int \frac{dp(z)}{d\bar{z}} W(z) dz.$$

This result is intuitive for $c = 1$: if the fitness landscape depends only on the difference between each individual trait the mean trait, then the mean fitness itself cannot depend on the mean trait.

805 **D Behavior of $\dot{\bar{z}}$ as $c \rightarrow 1$**

We start with the definition of the mean fitness,

$$\bar{W}(\bar{z}) = \int p(z - \bar{z}) W(z - c\bar{z}) dz.$$

Taking a series expansion of this equation around $c = 1$,

$$\overline{W}(\bar{z}) = \int p(z - \bar{z})W(z - \bar{z}) dz - (1 - c) \int p(z - \bar{z}) \frac{d}{dc} \left(W(z - c\bar{z}) \right) \Big|_{c=1} dz + \mathcal{O}((1 - c)^2).$$

$$\overline{W}(\bar{z}) = \int p(z)W(z) dz + (1 - c)\bar{z} \int p(z) \frac{dW(z)}{dz} dz + \mathcal{O}((1 - c)^2).$$

We take the derivative of both sides with respect to \bar{z} ,

$$\frac{d\overline{W}(\bar{z})}{d\bar{z}} = 0 + (1 - c) \int p(z) \frac{dW(z)}{dz} dz + \mathcal{O}((1 - c)^2).$$

810 Next we multiply both sides by $1/(1 - c)$ and take the limit $c \rightarrow 1$. We note that all terms $\mathcal{O}((1 - c)^2)$ vanish in this limit,

$$\lim_{c \rightarrow 1} \frac{1}{1 - c} \frac{d\overline{W}(\bar{z})}{d\bar{z}} = \int p(z) \frac{dW(z)}{dz} dz = \int W(z) \frac{dp(z)}{dz} dz.$$

Importantly, this suggests that the dynamics of $\dot{\bar{z}}$ do not diverge as $1/(1 - c)$, because $\frac{d\overline{W}(\bar{z})}{d\bar{z}}$ has a leading-order zero, namely $(1 - c)$.

Rearranging the asymptotic equation results in the dynamics when $c = 1$,

$$\dot{\bar{z}} = \frac{V}{\sigma \overline{W}} \int \mathcal{N}[\bar{z}, \sigma](z) f'(\bar{z}) W(z) dz.$$

815 The behavior of the series in the vicinity of $c = 1$ can be found by including a higher-order term in the Taylor series found above,

$$\dot{\bar{z}} = (1 - c) \frac{V}{\overline{W}} \bar{z} \int W(z) \frac{d^2 p(z)}{dz^2} + \frac{V}{\overline{W}} \frac{1}{\sigma} \int \mathcal{N}[\bar{z}, \sigma](z) f'(\bar{z}) W(z) dz.$$

E Derivation of trait variance dynamics $\dot{\sigma}$

As with the trait mean, we assume that the trait variance $M \equiv \sigma^2$ has linear heritability with a dynamical equation given by

$$\dot{M} = h_{\sigma}^2(M_w - M),$$

820 where h_{σ}^2 is the *variance heritability*; it determines the degree to which the phenotypic variance in one generation influences the phenotypic variance of the next generation. We expand this equation as

$$\dot{M} = h_{\sigma}^2 \left(\int (z - \bar{z})^2 p(z) \frac{W(z)}{\bar{W}} dz - \int (z - \bar{z})^2 p(z) dz \right) \quad (\text{A6})$$

We use the substitution,

$$p(z) = \mathcal{N}[\bar{z}, \sigma](z) f\left(\frac{z - \bar{z}}{\sigma}\right)$$

Taking the derivative of both sides of this expression results in the relationship

$$\frac{dp(z)}{dM} = \left(\frac{(z - \bar{z})^2}{2M^2} - \frac{1}{2M} \right) p(z) - \frac{1}{2M^{3/2}} (z - \bar{z}) f'(\tilde{z}) \mathcal{N}[\bar{z}, \sigma](z)$$

825 Rearranging this equation results in the expression

$$p(z)(z - \bar{z})^2 \frac{W(z)}{\bar{W}} = \frac{W(z)}{\bar{W}} \left(2M^2 \frac{dp(z)}{dM} + Mp(z) + \sqrt{M}(z - \bar{z}) f'(\tilde{z}) \mathcal{N}[\bar{z}, \sigma](z) \right)$$

Substituting the right hand side of this expression into the integral in Eq. A6 results in the equation,

$$\dot{M} = \frac{h_{\sigma}^2}{\bar{W}} \left(2M^2 \int W(z) \frac{dp(z)}{dM} dz + \sqrt{M} \int W(z) \mathcal{N}[\bar{z}, \sigma](z) f'(\tilde{z}) (z - \bar{z}) dz \right)$$

Using the substitution $M = \sigma^2$, $dM = 2\sigma d\sigma$ results in a dynamical equation for the standard deviation of the trait distribution,

$$\dot{\sigma} = \frac{U}{\overline{W}} \left(\int W(z) \frac{dp(z)}{d\sigma} dz + \frac{1}{\sigma} \int \tilde{z} \mathcal{N}[0, 1](\tilde{z}) W(\sigma\tilde{z} + \bar{z}) f' d\tilde{z} \right),$$

830 where $U \equiv (\sigma h_\sigma)^2/2$. We note that in many standard real-world contexts and theoretical frameworks, the fitness W does not explicitly depend on σ , and so $dW/d\sigma = 0$, resulting in a simpler form for this dynamical equation,

$$\dot{\sigma} = \frac{U}{\overline{W}} \left(\frac{d\overline{W}}{d\sigma} + \frac{1}{\sigma} \int \tilde{z} \mathcal{N}[0, 1](\tilde{z}) W(\sigma\tilde{z} + \bar{z}) f' d\tilde{z} \right).$$

Inserting the Gram-Charlier series (Eq. 6) for $p(z)$ results in an infinite series of the form

$$\dot{\sigma} = \frac{U}{\overline{W}} \left(\frac{d\overline{W}}{d\sigma} + \frac{1}{\sigma} \sum_{n=3}^{\infty} c_n n \int \tilde{z} \mathcal{N}[0, 1](\tilde{z}) He_{n-1}(\tilde{z}) W(\sigma\tilde{z} + \bar{z}) dz \right).$$

Using the properties of Hermite polynomials, this equation simplifies to

$$\begin{aligned} \dot{\sigma} = \frac{U}{\overline{W}} \frac{d\overline{W}}{d\sigma} + \frac{U}{\sigma \overline{W}} \sum_{n=3}^{\infty} c_n n \left((n-1) \int \mathcal{N}[0, 1](\tilde{z}) He_{n-2}(\tilde{z}) W(\sigma\tilde{z} + \bar{z}) dz \right. \\ \left. + \int \mathcal{N}[0, 1](\tilde{z}) He_n(\tilde{z}) W(\sigma\tilde{z} + \bar{z}) dz \right). \end{aligned} \quad (\text{A7})$$

F Dynamics of mean fitness

835 As the trait distribution $p(z)$ changes in time, the mean fitness \overline{W} also changes. These dynamics may be expressed in terms of the trait distribution cumulants or moments by noting that the fitness landscape $W(z)$ may be expanded in terms of an infinite series around $z = 0$,

$$W(z) = W_0 + W_1 z + W_2 z^2 + W_3 z^3 + \dots$$

Thus the mean fitness has the form

$$\overline{W} = W_0 + W_1\bar{z} + W_2(\sigma^2 + \bar{z}^2) + W_3\langle z^3 \rangle + \dots \quad (\text{A8})$$

840 where the relationship $\sigma^2 = \langle z^2 \rangle - \bar{z}^2$ has been used. The remaining individual raw moments $\langle z^n \rangle$ may be expressed in terms of the variance σ^2 and remaining cumulants κ_n using the relation Eq. A3.

For a given fitness landscape, the individual cumulants (which may vary in time) may be substituted into the expression for \overline{W} given by Eq. A8. The resulting explicit expression for
845 \overline{W} may then be substituted into the cumulant dynamical equations in order to express the dynamics in a form where the cumulants are the only dynamical variables. Depending the fitness landscape being used, the resulting equations may have a particularly simple form.

While the number of time-dependent dynamical variables required to study the trait distribution is simply the number of moments or cumulants that vary in time, for some
850 fitness landscapes the dynamics become more tractable if \overline{W} is instead treated as its own dynamical variable, with dynamics given its total derivative with respect to the remaining dynamical variables,

$$\dot{\overline{W}} = \frac{d\overline{W}}{d\bar{z}}\dot{\bar{z}} + \frac{d\overline{W}}{d\sigma}\dot{\sigma} + \sum_{i=3}^{\infty} \frac{d\overline{W}}{d\kappa_i}\kappa_i, \quad (\text{A9})$$

where κ_i are the cumulants of $p(z)$. Because all derivatives of \overline{W} may themselves be expressed in terms of the cumulants of the trait distribution $p(z)$, the dynamics of \overline{W} do not represent
855 an additional dependent variable in the dynamical system. This form of $\dot{\overline{W}}$ suggests that, even in the absence of additive genetic variance ($V = 0$) or in case of a fixed mean trait $\dot{\bar{z}} = 0$, the mean fitness may continue to change due to higher-order cumulants of the trait distribution contributing to the dynamics. These higher-order effects are absent in the standard breeder's equation and will be discussed further below.

860 Finally, we note that under the specific closure assumptions described above and in the main text, the cumulants series κ_n and thus the Gram-Charlier coefficients of the trait

distribution c_n are assumed to truncate at finite order, and the only dynamical moments of the trait distribution are assumed to be the mean \bar{z} and width/deviation σ . The latter assumption either entails that higher moments of the trait distribution are not heritable due to specific features of recombination or mating dynamics (which are not explicitly modelled here), or—less restrictively—that the fitness landscape does not affect these higher moments. This allows Eq. A9 to have a closed form due to truncation of the infinite series,

$$\dot{\bar{W}} = \frac{d\bar{W}}{d\bar{z}}\dot{\bar{z}} + \frac{d\bar{W}}{d\sigma}\dot{\sigma}$$

For fitness landscapes in which the dynamics of \bar{W} cannot be solved explicitly, this expression may be used as an additional dynamical equation without greatly affecting numerical performance, due to its simple dependence on the other dynamical equations.

G Formulation of dynamics in terms of heritability $h^2(t)$ and mean fitness $\bar{W}(t)$

We work under the “closure” assumption of truncated Gram-Charlier moments and dynamical equations, as described above. The resulting two-variable dynamical system $\bar{z}(t), \sigma(t)$ may instead be expressed in terms of $h^2(t)$ and $\bar{W}(t)$, where

$$\begin{aligned} h^2(t) &\equiv \frac{V}{\sigma(t)^2} \\ \bar{W}(t) &\equiv \int p(z, t) W(z) dz \end{aligned} \tag{A10}$$

which have the dynamical equations

$$\begin{aligned} (h(t)^2) &= -\frac{2V}{\sigma(t)^3}\dot{\sigma}(t) \\ \dot{\bar{W}}(t) &= \frac{d\bar{W}}{d\bar{z}}\dot{\bar{z}} + \frac{d\bar{W}}{d\sigma}\dot{\sigma} \end{aligned} \tag{A11}$$

The exact form of the equivalence $\bar{z}(t), \sigma(t) \leftrightarrow h^2(t), \bar{W}(t)$ depends on the exact functional form of the fitness landscape $W(z)$, but the general approach consists of the following steps:

1. Derive an expression for the fitness in terms of the trait moments. The fitness landscape $W(z)$ is written as an infinite series in z ,

$$W(z) = W_0 + W_1 z + W_2 z^2 + W_3 z^3 + \dots$$

This series is substituted into Eq. A10, and then individual terms are averaged in order to generate a series representation of \bar{W} in terms of its raw moments $\langle z \rangle, \langle z^2 \rangle, \langle z^3 \rangle, \dots$ (as is also described in Appendix F). This series representation may then be converted into a sum over cumulants or moments using Eq. A3, and then the resulting series may be directly substituted into the dynamical equations for $(\bar{z}(t), \sigma(t))$. The resulting cumulant series may then be truncated in order to yield a closed form expression for \bar{W} . For clarity refer to this finite series expression as $\tilde{\bar{W}}$.

2. Solve for the mean trait. The closed-form expression $\bar{W} = \tilde{\bar{W}}$ may then be algebraically inverted in order to generate a solution for the mean trait \bar{z} in terms of σ and \bar{W} . Importantly, because the mean appears in the explicit definitions of each of the raw moments given by Eq. A3, this step requires inverting a high-degree polynomial. Practically, symbolic solutions for \bar{z} may be kept as placeholders until numerical values for individual parameters are chosen. We use the notation $\tilde{\bar{z}}$ to refer to the values of the mean trait obtained by inversion of $\tilde{\bar{W}}$.

3. Substitute into the dynamical equations. We perform the following substitutions (in order) into the right hand side of the dynamical equations Eq. A11.

1. $\dot{\bar{z}}$ and $\dot{\sigma}$ are replaced by their explicit expressions in terms of \mathcal{W}_n (as given in the main text)

2. Wherever it appears, \bar{z} is replaced by $\tilde{\bar{z}}$, effectively replacing \bar{z} with \bar{W} in the dynamical

system.

3. Wherever it appears, σ is replaced by $\sqrt{V/h^2}$, removing σ from the dynamical system.

For a specific functional form of $W(z)$ (and thus \mathcal{W}_n), these steps may yield simple analytical results (as occurs for low-order polynomial or Gaussian fitness landscapes) or complex but complete analytic results, which require a symbolic algebra computer program to compute accurately.

H \mathcal{W}_n for exponential directional selection

A fitness landscape for exponential directional selection is given by

$$W(z) = W_0 e^{s(z - c\bar{z})}.$$

We insert this equation into the definition of \mathcal{W}_n ,

$$\mathcal{W}_n \equiv \int \mathcal{N}[0, 1](\tilde{z}) H e_n(\tilde{z}) W(\sigma \tilde{z} + \bar{z}) d\tilde{z},$$

which results in a simple expression for \mathcal{W}_n associated with exponential directional selection

$$\mathcal{W}_n = \sqrt{2\pi} W_0 e^{-sc\bar{z}} e^{s^2/2} s^n$$

Inserting these terms into the truncated four-moment differential equations for $\dot{\bar{z}}$, $\dot{\sigma}$,

$$\dot{\bar{z}} = \frac{V}{\bar{W}} \left(\frac{1}{1-c} \frac{d\bar{W}}{d\bar{z}} + \frac{W_0}{3\sigma} \sqrt{\frac{\pi}{2}} e^{-sc\bar{z}} e^{s^2/2} s^2 (3\gamma + (k-3)s) \right) \quad (\text{A12})$$

$$\dot{\sigma} = \frac{U}{\bar{W}} \left(\frac{d\bar{W}}{d\sigma} + \frac{W_0}{3\sigma} \sqrt{\frac{\pi}{2}} e^{-sc\bar{z}} e^{s^2/2} (s(2-s^2)(k-3) + 3\gamma(1+s^2)) \right), \quad (\text{A13})$$

We note that similar equations are obtained by Bürger for a discrete-time model of cumulant dynamics under directional selection.⁶⁹

Under the truncated model, we define the cryptic force of selection simply as the additional terms in each of Eq. A12 and Eq. A13, which would not be present if the trait distribution were purely Gaussian:

$$F_{\bar{z}}^c \equiv \frac{W_0}{3\sigma} \sqrt{\frac{\pi}{2}} e^{-sc\bar{z}} e^{s^2/2} s^2 (3\gamma + (k-3)s)$$

$$F_{\sigma}^c \equiv \frac{W_0}{3\sigma} \sqrt{\frac{\pi}{2}} e^{-sc\bar{z}} e^{s^2/2} (s(2-s^2)(k-3) + 3\gamma(1+s^2))$$

Depending on the overall strength of selection, these terms may be comparable in magnitude to the leading order gradient terms in the dynamical equations Eq. A12, Eq. A13.

I Orthogonality of Hermite polynomials and shifted Hermite polynomials

We wish to solve the following integral

$$\mathcal{I}_{m,n} = \int e^{-z^2/2} He_n(z) He_m(az+b) dz.$$

We define a generating function,

$$\phi(s, t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \mathcal{I}_{m,n} \frac{s^m}{m!} \frac{t^n}{n!}$$

such that

$$\mathcal{I}_{m,n} = \left. \frac{d^m}{ds^m} \frac{d^n}{dt^n} \phi(s, t) \right|_{s \rightarrow 0, t \rightarrow 0}.$$

The generating function may thus be expressed as,

$$\begin{aligned}\phi(s, t) &= \int e^{-\frac{z^2}{2}} \left(\sum_{m=0}^{\infty} He_m(az + b) \frac{s^m}{m!} \right) \left(\sum_{n=0}^{\infty} He_n(z) \frac{t^n}{n!} \right) dz \\ \phi(s, t) &= \int e^{-\frac{z^2}{2}} \left(\exp(2(az + b)s - s^2) \right) \left(\exp(2zt - t^2) \right) dz \\ \phi(s, t) &= \exp \left(2bs + (2a^2 - 1)s^2 + 4ast + t^2 \right).\end{aligned}$$

The derivatives of this generating function have a regular pattern due to the polynomial argument of the exponential, resulting in an analytic solution for the integral

$$\mathcal{I}_{m,n} = \begin{cases} \frac{m!}{(m-n)!} a^n (1-a^2)^{(m-n)/2} He_{m-n} \left(\frac{b}{\sqrt{1-a^2}} \right) & n \leq m \\ 0 & n > m \end{cases}$$

J Orthogonality of Hermite polynomials and regular polynomials

Using the same method of characteristic functions used in Appendix I, it can be shown that

the solution to the integral

$$\mathcal{I}_{m,n} = \int e^{-z^2/2} He_n(z) (az + b)^m dz.$$

is

$$\mathcal{I}_{m,n} = \begin{cases} \frac{m!}{(m-n)!} (-1)^{(m-n)/2} a^m He_{m-n} \left(\frac{b}{\sqrt{-a^2}} \right) & n \leq m \\ 0 & n > m \end{cases}$$

As an example application of this integral, we consider a fitness landscape given by a sum over Hermite polynomials

$$W(z) = \sum_{m=0}^{\infty} \alpha_m He_m(z - c\bar{z}).$$

Inserting this expression into the definition of \mathcal{W}_n results in a series with fewer terms,

$$\mathcal{W}_n = \sigma^n \sum_{m=n}^{\infty} \alpha_m \frac{m!}{(m-n)!} (1-\sigma^2)^{(m-n)/2} He_{m-n} \left(\frac{(1-c)\bar{z}}{\sqrt{1-\sigma^2}} \right)$$

925 **K \mathcal{W}_n for polynomial fitness landscapes**

We consider a fitness landscapes of the form

$$W(z) = W_0 + \sum_{m=1}^{\infty} \alpha_m (z - c\bar{z})^m \quad (\text{A14})$$

In general, analytic solutions are known for integrals of the form³⁸

$$\int \mathcal{N}[\mu, \sigma](z) He_n(z) z^m dz,$$

as given in Appendix J. Using these results, we find that the coupling coefficients \mathcal{W}_m themselves have the form of an infinite series,

$$\mathcal{W}_n = \sqrt{2\pi} W_0 \delta_{n,0} + \sum_{m=0}^{\infty} \alpha_m \mathcal{W}_n^m,$$

930 where

$$\mathcal{W}_n^m = \mathcal{I}_{m,n} = \begin{cases} \sqrt{2\pi} & n = 0 \\ \sqrt{2\pi} \frac{m!}{(m-n)!} \sigma^m (-1)^{(m-n)/2} He_{m-n} \left(\frac{\bar{z}(1-c)}{\sqrt{1-\sigma^2}} \right) & 0 < n \leq m \\ 0 & n > m \end{cases}$$

L Comparison with other models of cumulant dynamics

Here, we describe the relationship between our approach and those taken in previous studies that have formulated natural selection in terms of moment or cumulant dynamics equations. We note that, to our knowledge, our specific model and derivation—inserting a Gram-

Charlier series into the breeder’s equation, and then iteratively finding terms—is distinct from other approaches previously taken in existing literature. Our work is primarily relevant for describing phenotypic evolution over intervals in which the additive genetic variance V remains constant, such as during the phenotypic Wright-Fisher simulations we describe in the main text. The more detailed models described below include specific underlying genetic models; extending our results to work within these frameworks may require adding additional terms to the cumulant-wise generalized breeder’s equation Eq. A2, such as those found by Turelli and Barton 1994 for a broader study of non-Gaussian evolutionary processes.^{23, 28} However, the techniques that we use (the Gram Charlier series and cumulant-wise expansions of the fitness landscape) have also appeared in other contexts in evolutionary theory,³⁶ such as the infinitesimal model in which a continuous genotype is associated with a continuous fitness landscape.^{30, 67} However, due to its origin in phenotypic theory, our approach entails certain distinct assumptions and regimes of applicability which we aim to outline below.

Zeng 1987 reports, to our knowledge, the earliest application of the Gram-Charlier series to a problem in population genetics.⁴⁸ Zeng expands the joint distribution of genotypes and phenotypes in terms of a multivariate Gram-Charlier series (Eq. 3 of the referenced paper), and then studies the discrete-time consequences of truncation selection on an initially bivariate Gaussian distribution. Similar results are presented using different techniques in Bulmer 1980.³⁰ Section 3.2.1 of Zeng provides the general form of cumulant iteration equations for the infinitesimal model under random mating, and Section 3.2.2 describes iteration equations under an explicit multilocus theory. For the multilocus case, his Eq. 46, 47 represent explicit dynamical equations for the mean and variance under truncation selection. The other results in the paper require explicit cross terms that depend on the fitness landscape to be calculated, a landscape-dependent calculation similar in principal to the \mathcal{W}_n coefficients we describe above.

Related to the notion of dynamical equations arising from a cumulant-based approxima-

tion of the fitness distribution, Barton and Turelli 1987 presents a continuous-time matrix equation for the temporal dynamics of individual moments.⁴⁷ The individual terms in the symmetric transfer matrix have a form in which the i^{th} term in a given row or column depends on the mean allelic effect of the $i + 1^{th}$ and lower moments. They then show show for certain assumptions (such as diallelic inheritance) this matrix equation becomes finite, allowing the dynamics to be analyzed in closed form. The authors successfully use this model to describe the well-known mutation-selection balance effect.

Turelli and Barton 1990 extend this model in order to include recombination and linkage effects.⁴⁰ The theoretical framework of their model focuses on the distribution of genotypes, each of which is assumed to be associated with a single phenotype—making this work comparable to the continuous trait approach we present in our work above. Within a multilocus framework that includes linkage, Turelli and Barton present discrete time iteration equations in terms of infinite summations over gradients of the mean fitness \bar{W} with respect to various phenotypic moments, each multiplied by its respective phenotypic moment. These recursions simplify under the assumption of many loci (the infinitesimal model), resulting in a linear “variance heritability” equation (Eq. 3.2b of the referenced paper) similar to the one found by Bulmer 1971.²⁸ While our model above encodes information about the fitness landscape in terms of a set of “coupling coefficients” \mathcal{W}_n , under the framework of Turelli and Barton 1990 information about the fitness landscape is encoded by computing the mean fitness gradients with respect to the moments—making these gradients the primary determinants of the tractability of a given landscape. We compare the mean trait evolution equation found by Turelli and Barton 1990 in the many-loci limit (Eq. 3.15 of the referenced paper) to the dynamical equations we derive for $\dot{\bar{z}}$ above. For a directional landscape, our results have slightly different coefficients (arising primarily from the continuous-time derivation and the form of our generalized breeder’s equation Eq. A2). However, the form of our solutions for $\dot{\bar{z}}, \dot{\sigma}$ qualitatively match Eq. 3. 23 of Turelli and Barton (when higher order selection gradients are added back into their expression); in particular, the lowest order term is pro-

portional to the additive genetic variance. The qualitative stability properties of Turelli and
 990 Barton's results for a Gaussian fitness landscape entails global convergence for a stabilizing
 landscape, similar to our results for a polynomial stabilizing landscape.

Bürger 1991 notes the relative tractability of using cumulants, as opposed to moments,
 in the dynamical equations, and introduces the notion of taking a series expansion in the
 fitness landscape (Eq. 3.3 of the referenced work) in order to provide iterative terms in the
 995 dynamical equations (Eq. 4.8a,b).³⁷ These equations have a similar form to Eq. 3.13 of
 Turelli and Barton 1990,⁴⁰ although the latter is expressed in terms of moments. Bürger
 notes that linear selection advantages in a standard discrete time fitness model correspond
 to a Malthusian fitness landscape; we emphasize that due to the continuous time dynamics
 we study in the model we present above, the equivalent to this landscape in our formalism is
 1000 simply a linear (and not exponential) function. Otherwise, the general dependence of terms
 in Bürger's directional selection equation on the strength of directional selection has a power
 series form similar to the one we derive above.

Bürger 1993 studies cumulant dynamics under both directional and stabilizing (Gaussian)
 fitness landscapes.⁶⁹ For the directional terms, the drift of an initially Gaussian distribution
 1005 away from Gaussianity due to selection and mutation has the form of a power series in the
 selection strength multiplied by a Gaussian (Eq. 16 of the referenced paper); we note that
 these results are similar to those obtained for the directional linear and exponential land-
 scapes above, with the linear directional landscape in continuous time being most analogous
 to linear directional selection in discrete time. Bürger also notes the timescale isolation
 1010 associated with the dynamics of different cumulants. In the model we present above, this
 timescale isolation is used as an *ad hoc* heuristic to justify truncating the series of differential
 equations, since a finite number of differential equations may describe the dynamics for short
 timescales. Bürger also presents numerical results via Monte Carlo simulations illustrating
 the dynamics of cumulants in different fitness landscapes, and finds that the skewness and
 1015 kurtosis generally remain bounded over a range of selection strengths. The coupling between

selection strength, skewness, and the time evolution of the variance in these results is consistent with Barton and Turelli 1987,⁴⁷ as well as the calculations we present above for a directional fitness landscape, which suggest a strong role of tail effects in determining the dynamical regime.

Barton and Turelli 1994 start from the breeder’s equation and develop a general non-Gaussian infinitesimal model.²³ Their formalism takes root in the concept of using convolutional integrals to construct discrete-time integral iteration equations, a method originally developed by Slatkin 1970.⁶⁷ Assuming a Gaussian transmission kernel centered on the midparent breeding value, Turelli and Barton note that cumulant iterations are a natural simplification for such integrals due to the linearity of cumulants under convolution—an approach that motivates our use of them as well above, and which is also discussed by Bulmer 1980.³⁰ However, as with the method that we present above, Turelli and Barton note that cumulants present issues when recombination is included, due to a multiplicative increase in the number of cross-cumulant terms in the dynamical equations (Eq. 13 of the referenced work). As in Turelli and Barton 1990, they note that the primary challenge in constructing cumulant iteration equations for a given fitness landscape consists of computing closed form expressions for the gradients of the mean fitness with respect to individual cumulants. Analogously to the “coupling constants” \mathcal{W}_n that we describe above, Turelli and Barton present an integral equation (Eq. 26 of the referenced paper) that allows individual terms of the fitness landscape to be included iteratively in the dynamics. They calculate these gradients analytically for the case of truncation selection, in which case the cumulant series readily truncates due to the leading order effects of truncation acting primarily on lower moments—resulting in dynamical equations complementary to those found by Zeng 1987.⁴⁸ Turelli and Barton also provide general cumulant iteration equations in terms of selection gradients for both an explicit multilocus model (Eq. 35-40 of the referenced paper) and for the limit of an infinitesimal model, in which an arbitrarily large number of alleles may contribute to the fitness (Eq. 42, 43). Because the formulation of our phenotypic evolution

model is mathematically similar to an infinitesimal genetic model, the dynamical equations Eq. 42, 43 of Turelli and Barton have a comparable form to the general dynamical equations for $\dot{\bar{z}}$ and $\dot{\sigma}$ we calculate. However we truncate our series based explicitly on the order of the cumulant, whereas the selection gradient approach may allow higher cumulants to appear due to residual terms in a given derivative of the mean fitness \bar{W} . Turelli and Barton find that their iteration equations reduce to a higher-order variant of the breeder's equation under the assumption of Gaussian breeding value distributions (Eq. 45). We note that we have assumed breeder's equation type construction for the underlying genetics of our phenotypic model; an extension of our approach with a specific underlying genetic process may require additional terms to be added to the breeder's equation.

Distinct from the approaches used above, Prügel-Bennet studies explicitly the natural selection process as an iterated random sampling process biased by the fitness, an approach similar in motivation to the phenotypic evolutionary model we derive above.^{42, 50, 70} Prügel-Bennet's formulation of the selection process is parametrized by a selection rate parameter, β , analogous to the inverse temperature of a thermodynamic ensemble. Analytic forms of the cumulant dynamical equations require taking the equivalent of a "high temperature limit," wherein the sampling exponential is linearized and thus extremal phenotypes are sampled less frequently. Our model presented above is similar to a continuous time, continuous trait formulation of this approach, with individual coupling coefficients \mathcal{W}_n analogous to the individual series terms computed by Prügel-Bennet for the case where some phenotypes have a linear fitness advantage.

Another cumulant-based approach for a discrete allele sampling process is provided by Rattraay and Shapiro, who study a case of multiplicative selection in which the number of beneficial alleles accrued by an individual confers that individual with a linear fitness advantage relative to the rest of the population.³⁹ This multiplicative fitness landscape in discrete time is equivalent to the linear directional fitness landscape we study above for continuous time. Rattraay and Shapiro exploit a useful property of multinomial sampling

processes: in general, the n^{th} cumulant of the distribution may be written as a summation over terms of order p_i^n associated with the set of alleles indexed by i (Eq. 7 of the referenced paper). By assuming that each allele follows an independent diffusion process within the population, and that the growth rate of an allele in the diffusion equation is associated with its mean (the lowest cumulant), Rattray and Shapiro derive a series of dynamical equations describing the evolution of moments within the population (Eq. 12), which are linear (and thus may be solved via matrix inversion) for the multiplicative fitness landscape case. We note that our approach presented above differs in construction from Rattray and Shapiro, with the most substantive differences being (1) the explicit multilocus evolutionary model, which allows the cumulant equations to have forms as summations, and (2) the use of a diffusion process for individual alleles. In the many-allele limit implicit to the infinitesimal model, these conditions would describe circumstances with weak selection or extremely high mutation rates.

M Numerical solution of cumulant dynamical equations

Efficient simulation of the dynamics of cumulants requires that moments and cumulants are not re-calculated at each time step, as integrals over the entire trait domain are generally computationally expensive to perform numerically. However, if high order cumulants are not heritable (and thus do not change in time), these values can be calculated in advance, thus truncating the number of quantities that must be calculate. Any distribution may be used to determine realistic constant values of these higher-order cumulants; in general, however, distributions are chosen that have cumulants that grow slowly in n —this is a requirement to safely truncate the Gram-Charlier series. Distributions used for the numerical work include the skew normal distribution, the exponentially-modified Gaussian distribution, and the uniform distribution.

The characteristic function for a given probability distribution is given by

$$\phi_Z(s) = \int_{-\infty}^{\infty} e^{isz} p(z) dz \quad (\text{A15})$$

1095 The cumulants are given by the derivatives of the logarithm of this distribution,

$$\kappa_n = \frac{1}{i^n} \left(\frac{d^n}{ds^n} \log(\phi_Z(s)) \right) \Big|_{s \rightarrow 0},$$

and they can generally be calculated to arbitrary order for a given distribution, abetting their use in numerical work. However, for certain distributions the series κ_n has a closed form. For example, for many of the examples simulated in this paper, the uniform distribution is used because it has the property $\kappa_n = B_n/n$, where B_n is the n^{th} Bernoulli number.

1100 As a result, simulations of $\bar{z}(t), \sigma(t)$ can be interpreted as the dynamics of a smoothed boxcar distribution with time-varying center and width. However, if other distributions with necessarily different cumulant series are used, the general dynamics will remain the same so long as the distributions and their associated cumulants result in a strictly decreasing series of magnitudes of Gram-Charlier coefficients $|c_n|$. For numerical simulations shown in the
1105 paper, equivalent simulations were performed with a uniform distribution, a skew normal distribution, and an exponentially-modified Gaussian distribution, in order to ensure that qualitative results regarding the dynamics of the mean and variance remained the same.

For cases in which the underlying trait distribution as a function of time is needed, the time-dependent series of cumulant values, $\kappa_i(t)$, is first computed using the dynamical
1110 equations presented in the main text. These are then used to compute a time-dependent characteristic function,

$$\phi_Z(s, t) = \exp \left(\sum_{r=1}^{\infty} \kappa_r(t) \frac{(is)^r}{r!} \right)$$

This expression derived by inserting a general Gram-Charlier series (Eq. A4) into Eq. A15. From $\phi_Z(s, t)$, the trait distribution may then be computed by numerically performing the

inverse integral of Eq. A15,

$$p(z, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isz} \phi_Z(s, t) ds$$

1115 Using this equation, the dynamics of a sufficiently large number of cumulants may be converted to those of the full trait distribution. This underscores how moment transform methods are reminiscent of traditional spectral methods for dealing with partial differential equations, which provide insight into the properties of time evolution operators by recasting them within the frequency domain: similarly to Fourier modes, the moments of a distribution constitute parametrized integral transforms across real space. This analogy that can be made
1120 more direct when the distribution has a well-defined characteristic function, as it does here.

N Numerical simulations of selection on individuals

Numerical simulations were performed using a variation of a standard Wright-Fisher process that accounts for selection on a continuous phenotype. For a given initial set of moments
1125 μ, σ^2, γ, k , a set of N_i phenotypes was initialized using Johnson’s S_U -distribution (numerical optimization was used to select distribution parameters that produced desired moments). From these values, an initial mean trait value \bar{z} was calculated, allowing the full set of z values to be assigned relative fitnesses $w = W(z, \bar{z})/\bar{W}$. At the next timestep, N_i new values were drawn (with replacement) from the initial set of z values, and these values were
1130 taken as the “offspring” generation of phenotypes. This process was repeated for subsequent timesteps, with each offspring generation serving as the parent phenotype distribution for the next generation.

Because here we are primarily interested in the behavior of average moments of distributions in large populations, a relatively large value of $N_i = 10^6$ was used in all simulations
1135 (effectively suppressing drift effects). However, smaller effective population sizes may be used in cases in which drift is important. In the simulations, the first four moments of the

trait distribution vary gradually over time (primarily due to drift, here represented by the accumulation of sampling errors). However, as expected for large sample sizes, for simple fitness landscapes at large N_i the higher moments varied less than the lower moments over the simulation timescales, leading to a natural hierarchy in the timescales of dynamical variation in moments.

Figures S2-S4 show the results of Wright-Fisher simulations for the various fitness landscapes and conditions explored in the main text:

Figure S2 shows the dynamics of the mean trait $\bar{z}(t)$ and trait range $\sigma(t)$ under an exponential fitness landscape. The dynamics match those found for the simplified model in Figure 1; positive skewness produces gradual widening of the trait distribution, which in turn allows natural selection to proceed more rapidly when compared to the null case of a purely Gaussian distribution. Conversely, the case with negative skewness lags behind the default rate of natural selection, as predicted, resulting in a gradual narrowing of the trait range. For the timescales explored in these simulations, both the skewness γ and the kurtosis k are observed to fluctuate by less than 1%, despite there being no *a priori* constraints on their dynamics in the simulations. At smaller population sizes these quantities vary more due to drift, leading to more pronounced short-timescale departure from the theoretical predictions.

Figure S3 shows the results of Wright-Fisher simulations in a quartic fitness landscape with all parameters equal to those in the disruptive selection panel shown in Figure 2. Due to constraints on the properties of Johnson's S_U -distribution, arbitrary values of γ , k cannot be probed over the full range of values shown in the Figure. Instead, we choose a line of values of γ and k that crosses through regions in which the intermediate phenotype equilibrium is active and inactive. We initialize each Wright-Fisher simulation at the each value of γ , k along this line, and set the initial mean and standard deviation to match the equilibrium solution of the cumulant model at that particular value of γ and k . We then simulate the Wright-Fisher dynamics for a short time interval, in order to yield an estimate for the velocity of the

mean and trait range $(\dot{z}, \dot{\sigma})$ in the vicinity of the equilibrium point. These velocity values
 1165 are difficult to directly compare to the eigenvalues yielded by the local stability analysis in
 Figure 2; instead, we convert these velocity values into an “angle of motion” towards or
 away from the equilibrium using the transformation $\theta_{sim} = \arctan(v_y/v_x)$. We then compare
 this angle to the direction of the eigenvector associated with the largest eigenvalue of the
 local stability matrix for the cumulant dynamics model, in order to assess whether the local
 1170 dynamics of the Wright-Fisher system match those predicted by the local stability of the
 simpler cumulant dynamical model. Figure S3 shows the resulting comparison; at large
 values of γ , the two cases nearly exactly coincide. This strong coincidence is expected due
 to the extremely short timescales over which local stability analysis predicts the dynamics,
 because drift in the values of the individual cumulants does not have time to substantially
 1175 affect the relative dynamics.

Figure S4 shows the result of numerical simulations of for identical parameter values to
 those used in Figure 3. To facilitate easier comparison, a panel from that Figure is reprinted
 adjacent to the numerical results. It is apparent that the numerical work shows the same
 general trends as the theoretical model, including differences in the growth of the heritability
 1180 depending on the sign of the skewness, and non-monotonicity in the dynamics that depends
 on the relative value of the kurtosis. For these simulations, the values of the skewness and
 kurtosis were observed to vary by less than 5% over the duration of the dynamics. Many
 of the qualitative differences in the dynamics (particularly at large values of the kurtosis k)
 occur due to the the accumulated effects of these moments gradually varying.

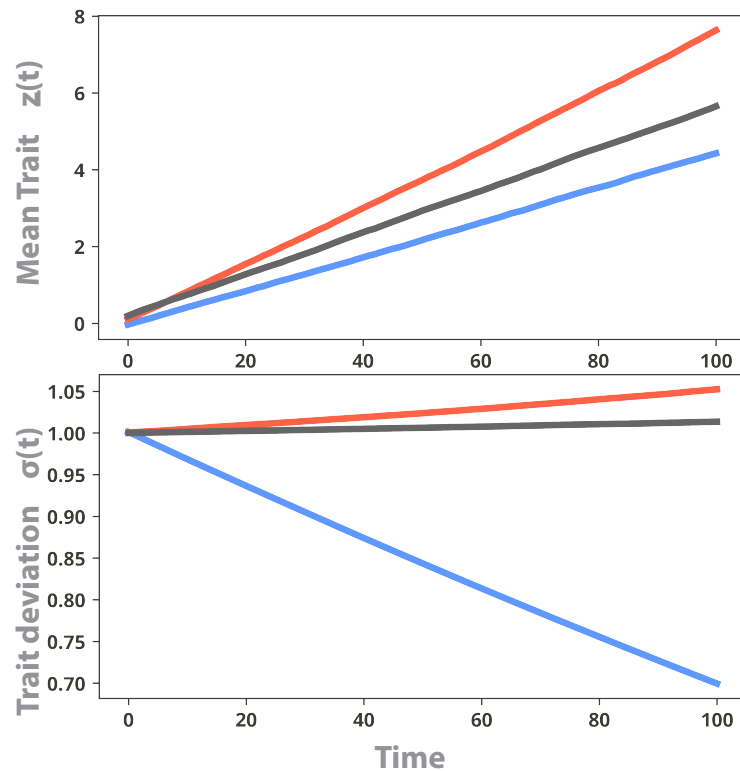


Figure S2 Wright-Fisher dynamics in exponential fitness landscape. The mean trait value and trait range for Wright-Fisher simulations in an exponential fitness landscape. Parameters are chosen to match those of Figure 1: gray curves correspond to an initially Gaussian population, blue curves correspond to initial distribution with negative skewness, and red curves correspond to initial distribution with positive skewness.

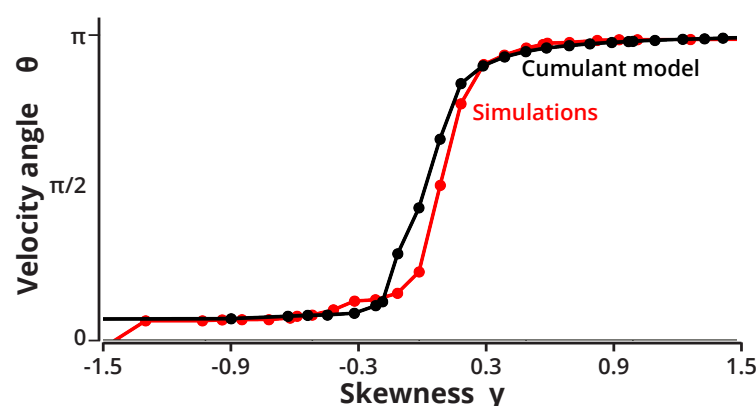


Figure S3 Wright-Fisher dynamics in quartic fitness landscape. A variety of γ values was first chosen, followed by a range of corresponding k values along the line $k = \gamma^2 + 2$. At each point, the cumulant dynamical equations under the four-moment model were solved and the direction of the principal eigenvector associated with the dynamics was computed (black traces). For each value of γ, k , corresponding values of \bar{z}, σ corresponding to the equilibrium point associated with the four-moment model were then used as starting distributions for Wright-Fisher dynamics simulations. From the short-time dynamics of the simulations, the direction of travel relative to the equilibrium point is calculated and shown in red on the plot.

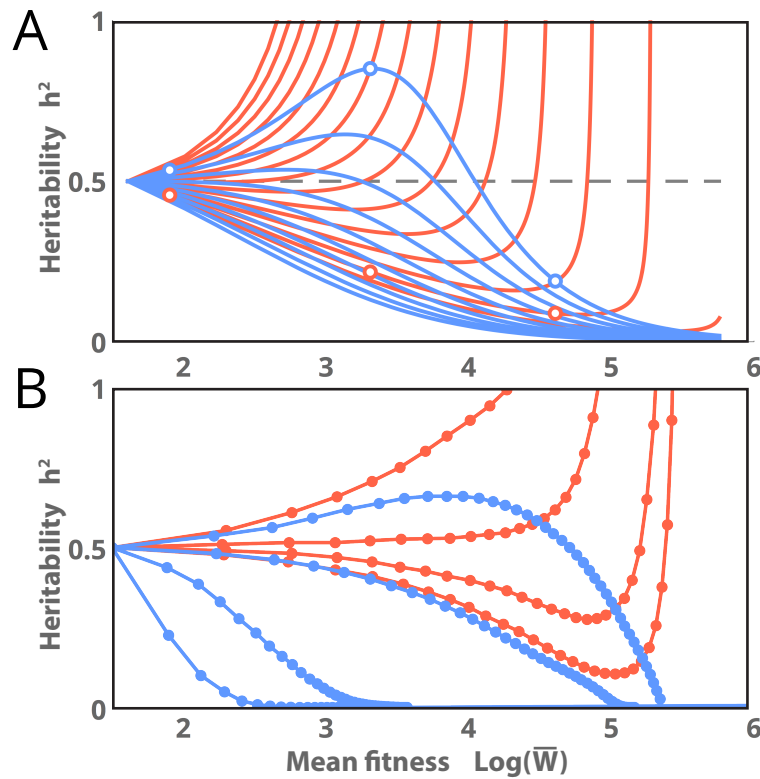


Figure S4 Comparison of analytic work with individual simulations. (A) Analytic results for heritability vs mean fitness under the four-moment model, with a linear fitness landscape. This panel is identical to Figure 3A of the main text. (B) Results of a numerical simulation of the Wright-Fisher model with 10^6 individuals in the same linear fitness landscape of (A); blue traces correspond to initial conditions with $\gamma = 0.05$ (blue), red traces correspond to initial conditions with $\gamma = -0.05$, and lower traces within each colored group correspond to linearly increasing initial values of the kurtosis k between $\gamma^2 + 1$ and $\gamma^2 + 6$. All parameter values are constant, with values as given in Figure 3