

# A High-Order Cumulant Extension of Quasi-Linkage Equilibrium

Kai S. Shimagaki,<sup>1, 2,\*</sup> Jorge Fernandez-de-Cossio-Diaz,<sup>3</sup> Mauro Pastore,<sup>4</sup>  
Rémi Monasson,<sup>5, 6, 7</sup> Simona Cocco,<sup>5, 6, 7</sup> and John P. Barton<sup>1, 2, †</sup>

<sup>1</sup>*Department of Physics and Astronomy, and*

<sup>2</sup>*Department of Computational and Systems Biology,  
University of Pittsburgh School of Medicine, USA.*

<sup>3</sup>*Institut de Physique Théorique, Université Paris-Saclay, CNRS, CEA, Gif-sur-Yvette, France.*

<sup>4</sup>*The Abdus Salam International Centre for Theoretical Physics (ICTP), Strada Costiera 11, 34151 Trieste, Italy*

<sup>5</sup>*Laboratoire de Physique de École Normale Supérieure, France*

<sup>6</sup>*PSL & CNRS UMR 8023, France*

<sup>7</sup>*Sorbonne Université, 7005 Paris, France*

A central question in evolutionary biology is how to quantitatively understand the dynamics of genetically diverse populations. Modeling the genotype distribution is challenging, as it ultimately requires tracking all correlations (or cumulants) among alleles at different loci. The quasi-linkage equilibrium (QLE) approximation simplifies this by assuming that correlations between alleles at different loci are weak – i.e., low linkage disequilibrium – allowing their dynamics to be modeled perturbatively. However, QLE breaks down under strong selection, significant epistatic interactions, or weak recombination. We extend the multilocus QLE framework to allow cumulants up to order  $K$  to evolve dynamically, while higher-order cumulants ( $> K$ ) are assumed to equilibrate rapidly. This extended QLE (exQLE) framework yields a general equation of motion for cumulants up to order  $K$ , which parallels the standard QLE dynamics (recovered when  $K = 1$ ). In this formulation, cumulant dynamics are driven by the gradient of average fitness, mediated by a geometrically interpretable matrix that stems from competition among genotypes. Our analysis shows that the exQLE with  $K = 2$  accurately captures cumulant dynamics even when the fitness function includes higher-order (e.g., third- or fourth-order) epistatic interactions, capabilities that standard QLE lacks. We also applied the exQLE framework to infer fitness parameters from temporal sequence data. Overall, exQLE provides a systematic and interpretable approximation scheme, leveraging analytical cumulant dynamics and reducing complexity by progressively truncating higher-order cumulants.

## INTRODUCTION

One of the central questions in evolutionary biology is to understand how populations evolve under natural selection and other evolutionary forces [1–4]. The aim is to model the dynamics of genotype distribution over time. A major challenge arises from the non-random association of alleles (e.g., nucleotides) at different loci (positions in a linear sequence), known as linkage disequilibrium (LD) [5]. In the absence of LD, a state known as linkage equilibrium, each allele evolves independently and its dynamics can be solved exactly [1]. However, this condition is overly restrictive. Indeed, natural selection can induce LD, which causes the dynamics of different alleles to become correlated. A general description of population evolution requires accounting for an exponentially increasing number of allele combinations across loci, which are coupled through LD [4, 6].

Generally, the time scale of change in LD is faster than the change in allele frequency. As a result, LD rapidly reaches equilibrium and is determined by allele frequency. This state, where LD values are small and stable while allele frequencies evolve, is known as quasi-linkage equilibrium (QLE) [4, 7]. In contrast to linkage equilibrium,

QLE can naturally occur under conditions of weak selection and/or weak epistasis, coupled with high recombination rates. This results in distinct time scales: the slow dynamics of individual allele frequency, governed by selection (with a time scale of  $1/s$ , where  $s$  is the selection strength), and the rapid decay of linkage disequilibrium (with a time scale of  $1/r$ , where  $r$  is the recombination rate). The slow dynamics are driven by gradients of average fitness with respect to cumulants [7–10]. Practically, QLE serves as a useful platform to investigate the collective evolution of alleles in multilocus systems, as it simplifies the mathematical structure, and reduces the dimensionality of genotype distribution [10].

QLE has primarily been examined within a two-locus, two-allele framework, initially proposed by Kimura [7] and further investigated by Nagylaki, Barton and colleagues [11, 12]. Barton and Turelli developed a framework for describing evolution in centered moments [13, 14], encompassing second and higher orders under arbitrary selection and recombination, which was later generalized by Kirkpatrick et al. [15]. Nagylaki et al. also focused on the evolution of multilocus systems, rigorously examining the sufficient conditions for convergence to equilibria or QLE manifolds using a small-epistasis perturbation theory [16, 17]. Recently, Neher and Shraiman further developed the QLE theory for multilocus systems, building on the foundational work by Barton and Turelli [13, 14]. This has made the QLE theory more conceptu-

\* Corresponding author; [kais@pitt.edu](mailto:kais@pitt.edu)

† Corresponding author; [jpbarton@pitt.edu](mailto:jpbarton@pitt.edu)

ally and analytically streamlined [10], elucidating that the cumulant dynamics are driven by the gradient of average fitenss. Recent work leveraged the analytically tractable QLE's cumulant dynamics to infer fitness parameters, an important challenge in evolutionary biology [18, 19].

Strong selection can at least transiently break down QLE. During a selective sweep, linked loci can exhibit positive LD as the selected loci drag other linked loci, known as genetic hitchhiking [20] (see ref. [21] for a broader discussion). In contrast, negative LD may arise between selected loci, known as the Hill-Robertson interference effect [22] (see also ref. [23]). These dynamics violate QLE, but equilibrium may re-establish after the sweep concludes. However, epistatic interactions between different loci can permanently violate the QLE assumptions [8, 17]. For example, if pairs of loci are selected due to epistatic interactions, then the pairwise LD or pairwise cumulants of alleles at different loci cannot relax to equilibria rapidly. As a result, the decay times of the pairwise cumulants become comparable to the timescales of individual allele frequencies.

Theoretical and computational studies offer insights into how epistasis arises [24, 25], the structure of epistasis [26–29], and how epistasis influences evolutionary dynamics [30, 31] (See also recent reviews [32, 33]). However, it remains challenging to understand how populations evolve in the presence of higher-order epistasis. Recent studies have revealed the prevalence and complex patterns of epistasis [34–40], including the presence of hierarchical higher-order epistasis in various contexts [41–48]. This underscores the need to extend QLE theory to accommodate a prevalent and broader range of epistasis.

Here, we propose an extended QLE (exQLE) framework that relaxes the assumptions that while individual allele frequencies evolve slowly, second and higher-order cumulants are small and converge rapidly to equilibria. Instead, we allow cumulants up to an arbitrary order  $K$  to evolve dynamically while assuming those of order  $K + 1$  and above remain small and rapidly reach equilibrium. We first derived a general expression for cumulant dynamics under an arbitrary order of cumulants and genotype distributions, showing that their evolution is driven by the gradient of the average fitness function with respect to the cumulants. This forms the basis of the exQLE formulation, allowing for systematic relaxation of the assumption that cumulants above the  $K = 1$  order are effectively in steady state.

As an example, we analyze the  $K = 2$  case, which is the simplest extension of QLE, demonstrating the dynamics of first- and second-order cumulants. These expressions describe how fitness parameters influence the cumulant dynamics and can be used to infer these fitness parameters. Interestingly, the expressions for the cumulant dynamics from exQLE match exactly with those derived by projecting genotype dynamics onto the space of cumulants in the diffusion approximation of the Wright-Fisher (WF) process [49, 50]. Furthermore, the exQLE frame-

work, whose cumulant dynamics are fully characterized by combinations of cumulants, readily provides a systematic approach to approximate the dynamics by progressively reducing cumulants in an order-by-order manner. This systematic approach reproduces the previously reported QLE-based epistasis inference method with a Gaussian closure (GC) scheme [51], along with alternative inference methods.

## EVOLUTION OF GENETIC TRAITS AND QUASI-LINKAGE EQUILIBRIUM THEORY

Here, we consider the evolution of a population of individuals, described by a probability distribution of genotypes,  $P(\mathbf{g}, t)$  with a binary genotype  $\mathbf{g} \in \{-1, +1\}^L$  of length  $L$ . The average of an arbitrary genetic trait across a temporal genetic distribution is defined as  $\langle G \rangle := \sum_{\mathbf{g}} P(\mathbf{g}, t)G(\mathbf{g})$ . For the sake of simplicity, the time dependency ( $t$ ) has been omitted. The principle governing population evolution is that fitter genotypes produce more offspring, leading to their increased frequency in the next generation. Let  $F(\mathbf{g})$  denotes a fitness function, mapping genotypes to Malthusian fitness [52, 53], then the genotype dynamics can then be expressed as  $P(g, t + \Delta t) = \frac{e^{\Delta t F(g)}}{\langle e^{\Delta t F(g)} \rangle} P(g, t)$  over a time period  $\Delta t$ . When the selection is small such that  $|\log(e^{\Delta t F} / \langle e^{\Delta t F} \rangle)| \ll 1$ , the expected genotype follows  $P(g, t + \Delta t) \simeq P(g, t) + \Delta t[F(g) - \langle F \rangle]P(g, t)$ . Considering mutation and recombination effects, which operate on individuals and introduce genetic variation, the dynamics of the genotype distribution is described by the following master equation:

$$\begin{aligned} \dot{P}(\mathbf{g}, t) &= [F(\mathbf{g}) - \langle F \rangle] P(\mathbf{g}, t) \\ &+ \mu \sum_{\mathbf{g}' : d(\mathbf{g}, \mathbf{g}')=1} [P(\mathbf{g}', t) - P(\mathbf{g}, t)] \\ &+ r \sum_{\mathbf{g}', \mathbf{g}''} R(\mathbf{g} | \mathbf{g}', \mathbf{g}'') P(\mathbf{g}', t) P(\mathbf{g}'', t) - r P(\mathbf{g}, t) \end{aligned} \quad (1)$$

where  $\mu$ ,  $r$  are, respectively, mutation and recombination rates,  $d(\mathbf{g}, \mathbf{g}')$  represents the Hamming distance, and  $R(\mathbf{g} | \mathbf{g}', \mathbf{g}'')$  is the probability that genotypes  $\mathbf{g}'$  and  $\mathbf{g}''$  produce genotype  $\mathbf{g}$  through recombination.

For simplicity, we temporarily ignore the contributions of mutation and recombination ( $\mu = 0$  and  $r = 0$ ), as their effects do not alter the following discussion and are explicitly detailed in refs. [10, 50]. Given this dynamical rule, the equation of motion for the average arbitrary trait or arbitrary function of  $\mathbf{g}$  denoted as  $G$ , is given as:  $\frac{d}{dt} \langle G \rangle = \text{Cov}(F, G)$ , which is known as Price's equation [54] or “second theorem” of natural selection [55], and can be viewed as a generalization of Fisher's “fundamental theorem” [2, 54]. Throughout this paper, we assume the fitness function is expressed as

$$F(\mathbf{g}) = \sum_i s_i g_i + \sum_{i < j} s_{ij} g_i g_j + \dots, \quad (2)$$

where  $s_i$  and  $s_{ij}$  are time-independent coefficients characterizing the effects of a single mutation at site  $i$  (selection coefficient) and a double mutation at site  $i$  and  $j$  (pairwise epistatic coefficient). To describe dynamics of cumulants below, we define the following cumulant generating function under an arbitrary genotype distribution,  $P(\mathbf{g}, t)$ , parameterized by  $\phi = (\phi_1, \dots, \phi_L)^\top$ , given as:

$$\Phi(\phi, t) := \log \left( \sum_{\mathbf{g}} P(\mathbf{g}, t) e^{\phi^\top \mathbf{g}} \right). \quad (3)$$

For notational convenience, we define  $\chi_i^\phi := \partial_{\phi_i} \Phi$ ,  $\chi_{ij}^\phi := \partial_{\phi_i} \partial_{\phi_j} \Phi$ , and subsequent orders, which results in the cumulants,  $\chi_i = \chi_i^\phi|_{\phi=0}$ ,  $\chi_{ij} = \chi_{ij}^\phi|_{\phi=0}$ , respectively. Additionally, we define  $\langle G \rangle_\phi := \sum_{\mathbf{g}} G(\mathbf{g}) P(\mathbf{g}, t) e^{\phi^\top \mathbf{g}} / \sum_{\mathbf{g}} P(\mathbf{g}, t) e^{\phi^\top \mathbf{g}}$ , which yields  $\langle G \rangle_\phi|_{\phi=0} = \langle G \rangle$ . Based on Price's equation and (3), along with the assumption that  $P(\mathbf{g}, t)$  does not induce strong interactions between sites and that the dynamics of higher-order cumulants are negligibly small, it is derived that the dynamics of first-order cumulants as  $\dot{\chi}_i \simeq \sum_j \chi_{ij} \partial_{\chi_j} \langle F \rangle$  (ref. [9]). The dynamics of arbitrary traits were expressed as a linear combination of first-order cumulant dynamics,  $\frac{d\langle G \rangle}{dt} = \sum_i \dot{\chi}_i \partial_{\chi_i} \langle G \rangle$ . In the QLE framework, we assume that higher-order cumulants change faster than first-order cumulants, allowing the former to quickly reach equilibrium while the later are still evolving ( $\dot{\chi}_{ij} = \dot{\chi}_{ijk} = \dots = 0$ ).

## EXTENSION OF QLE THEORY

We begin by expressing the equation of motion for cumulants under any arbitrary genotype distribution. We then consider a scenario where the  $K$ -th order cumulants evolve dynamically, while higher-order cumulants (those of order  $K+1$  and above) remain small and rapidly reach equilibria.

To express the general form of cumulant dynamics, let  $\mathcal{I}, \mathcal{J}, \mathcal{K}$  denote multi-indices over loci, e.g.,  $\mathcal{I} = (i_1, i_2, \dots)$  where  $i_1, i_2, \dots$  are indices of loci. Using the cumulant generating function and the relation  $\partial_t \Phi(\phi, t) = \langle F \rangle_\phi - \langle F \rangle$ , the dynamics of cumulants of arbitrary order can be expressed as

$$\begin{aligned} \dot{\chi}_{\mathcal{I}} &= \partial_t \partial_{\phi_{\mathcal{I}}} \Phi(\phi, t)|_{\phi=0} \\ &= \partial_{\phi_{\mathcal{I}}} \langle F \rangle_\phi|_{\phi=0}. \end{aligned} \quad (4)$$

By denoting moments  $\mu_{\mathcal{I}}^\phi := e^{-\Phi} \partial_{\phi_{\mathcal{I}}} e^\Phi$  with  $\mu_{\mathcal{I}} = \mu_{\mathcal{I}}^\phi|_{\phi=0}$ , the last expression can be further written as:

$$\begin{aligned} \partial_{\phi_{\mathcal{I}}} \langle F \rangle_\phi|_{\phi=0} &= \sum_{\mathcal{K}} \frac{\partial \langle F \rangle}{\partial \mu_{\mathcal{K}}} \frac{\partial \mu_{\mathcal{K}}^\phi}{\partial \phi_{\mathcal{I}}}|_{\phi=0} \\ &= \sum_{\mathcal{J}, \mathcal{K}} \frac{\partial \langle F \rangle}{\partial \chi_{\mathcal{J}}} \frac{\partial \chi_{\mathcal{J}}}{\partial \mu_{\mathcal{K}}} \frac{\partial \mu_{\mathcal{K}}^\phi}{\partial \phi_{\mathcal{I}}}|_{\phi=0}. \end{aligned} \quad (5)$$

In the first equality, we use the fact that  $\frac{\partial \langle F \rangle}{\partial \mu_{\mathcal{L}}}$  is independent of the statistical variables. From the first to the second line, we convert from moment-based to cumulant-based expressions.

Therefore, by denoting  $\chi$  and  $\nabla_\chi$  as  $\chi := ((\chi_i)_i, (\chi_{ij})_{i < j}, \dots)$  and  $\nabla_\chi := ((\partial_{\chi_i})_i, (\partial_{\chi_{ij}})_{i < j}, \dots)$ , the cumulant dynamics can be generally expressed as:

$$\begin{aligned} \dot{\chi} &= D(\chi) \nabla_\chi \langle F \rangle \\ D_{\mathcal{I}, \mathcal{J}} &:= \sum_{\mathcal{K}} \frac{\partial \chi_{\mathcal{J}}}{\partial \mu_{\mathcal{K}}} \frac{\partial \mu_{\mathcal{K}}^\phi}{\partial \phi_{\mathcal{I}}}|_{\phi=0} \end{aligned} \quad (6)$$

Explicit computations for  $D_{\mathcal{I}, \mathcal{J}}$  are provided in the Supplementary Information (SI). The matrix  $D(\chi)$  is symmetric and positive definite, resulting from genotype competition and acting as a diffusion matrix in stochastic processes (see also SI). Alternatively,  $D(\chi)$  can also be seen as a mobility matrix in fluid dynamics, as it links velocity (i.e., cumulant dynamics) to potential force (i.e., the gradient of average fitness) [10, 56]. (6) suggests that cumulants evolve along the gradient of average fitness through a matrix that serves as a geometric metric – a picture that was also drawn in the QLE theory [10] – within this generalized setting.

Based on (6), the extended QLE (exQLE) is derived by truncating the dynamics of cumulants up to order  $K$ . In this setting, each row of  $D(\chi)$ , corresponding to cumulants up to order  $K$ , generally involves higher-order cumulants. For instance, the dynamics of first-order cumulants depend on second-, third-, and higher-order terms. However, cumulant dynamics up to order  $K$  depend only on a finite set of cumulants, determined by the fitness function. For example, in an additive fitness model ( $K^* = 1$ ), the gradient of the first-order cumulant is constant, and those of higher-order cumulants vanish. Thus, only cumulants up to order  $K + K^* = 2$  contribute to  $D(\chi)$ ; higher-order terms ( $> 2$ ) have no effect on first-order dynamics. Therefore, the equality in (6) holds as long as  $D(\chi)$  includes cumulants up to order  $K + K^*$ , where  $K^*$  is the highest-order cumulant contributing to the fitness function. Similarly, if the goal is to understand the relationship between cumulant dynamics and the gradient of average fitness, and the fitness depends on cumulants up to order  $K^*$ , then considering cumulant dynamics up to order  $K = K^*$  is necessary (note: the  $D(\chi)$  matrix still needs to include cumulants up to order  $K + K^* = 2K^*$ ). As we will discuss in the next section, the exQLE with  $K = 2$  under a pairwise fitness function ( $K^* = 2$ ) yields exact results for this reason.

It is important to note that the expression in (6) is not self-contained, as the dynamics of cumulants up to order  $K$  depend on higher-order cumulants up to order  $K + K^*$ , unless  $\chi$  includes cumulants of all orders. As a result, it is limited in describing cumulant dynamics over long time periods. To use this equation for simulating cumulant dynamics, the higher-order cumulants (beyond order  $K$ ) must be specified externally (e.g., assumed to be zero or

random) or described by using lower-order cumulants to make the equation self-contained.

### EXTENDED QLE: THE $K = 2$ CASE

To illustrate the extension of QLE theory, we consider the  $K = 2$  case, where cumulants up to the second order can dynamically evolve, while higher-order cumulants are small and rapidly reach equilibria. We assume a fitness function that depends on cumulants up to order  $K^* > 2$  and an arbitrary genotype distribution. Under these conditions, the dynamics of  $\chi_i$  are given by:

$$\begin{aligned}\dot{\chi}_i &= \partial_{\phi_i} \langle F \rangle_{\phi} |_{\phi=0} \\ &\simeq \sum_k \chi_{ik} \partial_{\chi_k} \langle F \rangle + \sum_{k < l} \chi_{ikl} \partial_{\chi_{kl}} \langle F \rangle.\end{aligned}\quad (7)$$

Similarly, the dynamics of  $\chi_{ij}$  follow:

$$\begin{aligned}\dot{\chi}_{ij} &= \partial_{\phi_i} \partial_{\phi_j} \langle F \rangle_{\phi} |_{\phi=0} \\ &\simeq \sum_k \chi_{ijk} \partial_{\chi_k} \langle F \rangle + \sum_{k < l} (\chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk}) \partial_{\chi_{kl}} \langle F \rangle.\end{aligned}\quad (8)$$

Details of the derivation (8) are provided in SI. Note that  $\dot{\chi}$  depends linearly on fitness parameters, which simplifies the inference of them as we discuss below. If the average fitness function is characterized by cumulants up to second-order, the relations in (7) and (8) hold exactly. Under this extended QLE framework, the dynamics of any arbitrary trait  $G$  is given as:

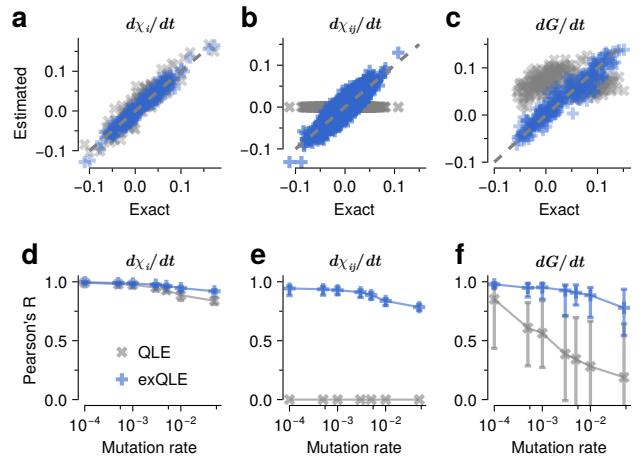
$$\frac{d\langle G \rangle}{dt} \simeq \sum_i \dot{\chi}_i \partial_{\chi_i} \langle G \rangle + \sum_{ij} \dot{\chi}_{ij} \partial_{\chi_{ij}} \langle G \rangle.$$

The dynamics effectively align with gradients of the average fitness surface and average traits. The dynamics of cumulants are characterized by a symmetric matrix  $D(\chi)$ , such that

$$\begin{aligned}\dot{\chi} &\simeq D(\chi) \nabla_{\chi} \langle F \rangle \\ D(\chi) &= \begin{pmatrix} \chi_{ik} & \chi_{ikl} \\ \chi_{ijk} & \chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} \end{pmatrix}.\end{aligned}\quad (9)$$

Here, the indices  $i, j, k, l$  spans all combinations: the top-left block  $\chi_{ik}$  spans rows and columns 1 to  $L$ ; the top-right spans rows 1 to  $L$ , columns  $L+1$  to  $L(L+1)/2$ ; and the bottom-right spans rows and columns  $L+1$  to  $L(L+1)/2$ . Although the interpretation of the gradient of average fitness and the mobility matrix is absent, equations equivalent to (7) and (8) can be found in ref. [57].

To examine how the exQLE framework incorporates the second-order cumulant dynamics and enhances the descriptive capacity of population evolution, we numerically compared the cumulant dynamics derived in (7) and (8) with the exact cumulant dynamics obtained from Price's equation. As we discussed earlier, the expressions



**FIG. 1. The accuracy of cumulant and trait dynamics by comparing estimated and exact values.** To examine the accuracy of the exQLE framework, we consider a fitness function that depends on interactions up to the fourth-order, providing a non-trivial test case. Genetic sequences were simulated using the Wright–Fisher (WF) process, incorporating mutation, recombination, and selection determined by this fitness function. Simulation conditions are detailed at the SI. This presents the dynamics of first-order cumulants (a), second-order cumulants (b), and a random trait defined below (c), comparing results from exact calculations based on Price's equation with those from the exQLE (with  $K = 2$ ) and QLE ( $K = 1$ ). Unlike the QLE case, exQLE closely match the exact dynamics for all cases. For the random trait, which is a pairwise function with parameters drawn from a normal distribution (see SI), exQLE can accurately capture the dynamics, while QLE cannot. To systematically assess accuracy, we performed 10 independent WF simulations and estimated the  $R$  values across multiple mutation rates (Fig. 1d–f). Overall, exQLE consistently outperformed QLE, maintaining higher  $R$  values across all mutation rates.

(7) and (8) are exact when  $K = K^* = 2$ , and the  $D(\chi)$  includes cumulants up to the 4-th order, corresponding to  $K + K^*$ . Therefore, as a non-trivial example, we considered a fitness function incorporating third- and fourth-order interactions between sites (Fig. 1). The first- and second-order cumulant dynamics derived from exQLE closely correlate with those obtained from the exact calculations, whereas the original QLE framework fails to accurately describe the cumulant dynamics (Fig. 1a–b). As an example of a more general trait, we evaluated the dynamics of a trait defined as a function of first- and second-order cumulants (its mathematical definition is provided in SI). The results demonstrate that trait dynamics derived from exQLE align well with the exact dynamics, whereas QLE, which excludes second-order cumulant dynamics, does not (Fig. 1c).

To assess the accuracy of exQLE in capturing cumulant and trait dynamics, we analyzed the correlation between the estimated  $\dot{\chi}_i$ ,  $\dot{\chi}_{ij}$  and random traits across multiple mutation rates. Increasing mutation rates lead to

a greater number of genotypes, effectively enhancing the presence of finite higher-order cumulants. The Pearson's  $R$  values comparing those from the exact results and estimated ones based on the exQLE framework achieve consistently higher values than those from the QLE method across various mutations (Fig. 1d-f).

Naive computation of dynamics under higher-order fitness is computationally challenging; therefore, we also propose an efficient method in SI.

## INFERRING FITNESS PARAMETERS

We now illustrate how to infer fitness parameters  $\mathbf{s} = ((s_i)_i, (s_{ij})_{i < j}, \dots)$  from cumulant dynamics. As discussed at the end of the section, the core idea parallels the derivation of the recently proposed marginal path-likelihood (MPL) method [49, 50]. A more mathematically explicit discussion is also provided in SI.

So far, our analysis has assumed deterministic dynamics. However, under finite population size  $N$ , stochasticity must be incorporated. To account for this effect, the cumulant dynamics can be represented as a Langevin equation:

$$\dot{\chi} = D(\chi) \nabla_\chi \langle F \rangle + \sqrt{D(\chi)/N} \boldsymbol{\eta}(t), \quad (10)$$

where  $\boldsymbol{\eta}(t)$  is white Gaussian noise, satisfying with  $\langle \boldsymbol{\eta} \rangle = \mathbf{0}$  and  $\langle \eta_J(t) \eta_K(t') \rangle = \delta_{J,K} \delta(t - t')$ . This form of noise arises under the WF process and has been used in the prior work [10, 49, 50].

Due to stochasticity  $\boldsymbol{\eta}(t)$ , the system can trace varied cumulant trajectories over time. The probability distribution of the cumulant trajectories over time,  $P((\chi(t_k))_{k=0}^K | \mathbf{s})$ , is obtained from the Langevin equation. The most probable parameter  $\hat{\mathbf{s}}$  maximizes this likelihood and equivalently, minimize the action, the time integral of the squared noise term (see SI). Therefore, the solution is given by:

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \max_{\mathbf{s}} P((\chi(t_k))_{k=0}^K | \mathbf{s}) \\ &= \left( \sum_{k=0}^K \Delta t_k \tilde{D}(\chi(t_k)) \right)^{-1} \sum_{k=0}^K \Delta \chi(t_k), \end{aligned} \quad (11)$$

where  $\Delta t_k := t_{k+1} - t_k$ ,  $\Delta \chi(t_k) := \chi(t_{k+1}) - \chi(t_k)$ , and the matrix  $\tilde{D}(\chi)$  is defined such that

$$\tilde{D}(\chi) \mathbf{s} := D(\chi) \nabla_\chi \langle F \rangle. \quad (12)$$

The simple closed-form inference arises from the linearity of the dynamics in the fitness parameters (see SI). This path-likelihood maximization framework provides a principled basis for inference, with statistical integrals arising naturally in the solution.

As an example, in the  $K = 2$  case, the maximum likelihood equation can be used to infer  $\mathbf{s}$  under WF process with mutation and recombination. By reintroducing the

expected cumulant changes due to mutation and recombination, as computed in prior work [10, 57], the solution becomes:

$$\begin{aligned} \hat{\mathbf{s}} &= \left( \sum_{k=0}^K \Delta t_k \tilde{D}(\chi(t_k)) \right)^{-1} \\ &\times \sum_{k=0}^K \left[ \begin{pmatrix} \Delta \chi_i(t_k) \\ \Delta \chi_{ij}(t_k) \end{pmatrix} + \Delta t_k \begin{pmatrix} 2\mu \chi_i(t_k) \\ (4\mu + rc_{ij}) \chi_{ij}(t_k) \end{pmatrix} \right], \end{aligned} \quad (13)$$

with

$$\tilde{D}(\chi) = \begin{pmatrix} \chi_{ik} & \chi_{ikl} + \chi_{ik}\chi_l + \chi_{il}\chi_k \\ \chi_{ijk} & \chi_{ijkl} + \chi_{ik}\chi_{jl} + \chi_{il}\chi_{jk} + \chi_{ijk}\chi_l + \chi_{ijl}\chi_k \end{pmatrix}, \quad (14)$$

which is derived from the definition of  $\tilde{D}$  in (12) and the matrix  $D(\chi)$  in (9).

This expression, (13) and (14), matches with the one derived from MPL upon transforming from the 0/1 basis and moment representation to the  $-/+$  basis and cumulants [49, 50, 58]. Besides mathematical conventions, exQLE and MPL differ in that exQLE employs a Langevin equation for cumulant dynamics for arbitrary order  $K$ , while MPL considers the WF process under the diffusion limit.

The cumulant-based approach allows for systematic approximations of the equation of motion by reducing the effects of higher-order cumulants in an order-by-order manner. For example, given the expression (13), it is relevant to consider the scenario where cumulants beyond the second order are negligible. In this case,  $D(\chi)$  matrix depends only on second-order cumulants, leading to

$$\begin{aligned} \dot{\chi}_i &\simeq \sum_k s_k \chi_{ik} + \sum_{k < l} s_{kl} \chi_{ik} \chi_{il} - 2\mu \chi_i \\ \dot{\chi}_{ij} &\simeq \sum_{k < l} s_{kl} (\chi_{il} \chi_{jk} + \chi_{jl} \chi_{ik}) - (4\mu + rc_{ij}) \chi_{ij}. \end{aligned} \quad (15)$$

When considering only up to second-order cumulants, the dependence of second-order cumulant dynamics on additive selection disappears, as also observed in the previous multi-locus QLE study [10]. The absence of additive fitness dependence in second-order cumulant dynamics, which ultimately enables a closed-form expression for epistasis [10, 57], arises because epistatic effects are coupled with third-order cumulants. By ignoring third-order cumulant dynamics, cumulant evolution becomes effectively independent of additive fitness effects. Suppose we assume steady-state conditions in  $\chi_{ij}$  and further approximate  $\sum_{k < l} s_{kl} \chi_{il} \chi_{jk} \simeq s_{ij} \chi_{ii} \chi_{jj}$ . In this case, we recover the identical solution as the QLE epistasis model with Gaussian closure (GC) scheme, which drops more than second-order cumulants [51, 59], given as:  $s_{ij} = (4\mu + rc_{ij}) \chi_{ij} / \chi_{ii} \chi_{jj}$ .

To evaluate how accurately the exQLE inference framework, represented by (13), and the exQLE with GC scheme (exQLEGc), represented by (15), can infer fitness parameters including selection and epistatic

coefficients, we performed 100 independent WF simulations under a pairwise fitness function across multiple mutation rates. This allows us to systematically assess the correlation between inferred coefficients and ground-truth fitness parameters. Fig. S1 also illustrates the results from the  $K^* = 4$  case.

Overall, exQLE and exQLEGc achieve similar inference accuracy (accuracy of the forward simulation using exQLEGc can be found in Fig. S2). For selection coefficients, the average Pearson's  $R$  values range from 0.74 to 0.87, with the lowest  $R$  value occurring at smaller mutation rates and the highest  $R$  values at the highest mutation rate ( $\mu = 0.05$ ) (Fig. 2a). For epistatic inference, mutation rates noticeably influence  $R$  values, with the lowest  $R$  value of 0.15 occurring at the lowest mutation rate. Higher mutation rates yielded higher  $R$  values, up to a maximum of 0.79 (Fig. 2b). Genetic diversity is measured by the average entropy across all sites, assuming that the sites are independent of one another, and varies with mutation rates in a similar manner to the Pearson's  $R$  values for inferred epistasis. This suggests that increased mutation rates diversify the population, effectively increasing the number of distinct genotypes, enhancing the precision of cumulant estimation, and resulting in higher  $R$  values for the epistasis inference (Fig. 2d). The inferred fitness values, obtained from inferred selection and epistatic coefficients as well as genetic sequences, also increase with mutation rates, though they exceed 0.81 with a maximum value of 0.97 at the highest mutation rate.

## DISCUSSION

Quantitatively understanding how fundamental evolutionary forces, such as selection, mutation, and recombination, shape the evolution of populations and phenotypic traits, is a central question in evolutionary biology [3, 60–62]. While considering multilocus effects is crucial for capturing the complexity of genetic evolution, and growing evidence shows that most phenotypic traits are governed by many alleles across different loci [34–48], many foundational results were based on single-locus or two-locus models [1, 7, 63–66].

Existing studies of genetic evolution in multilocus populations often assume that the non-random association between alleles in different loci, known as linkage disequilibrium (LD), rapidly vanishes and is negligibly small [6, 15, 63, 67]. Although these studies provide valuable insights into genetic evolution, they do not apply to collective evolution across different sites, which involves complex and rich phenomena [20–23, 68, 69].

Quasi-linkage equilibrium (QLE), where the LD is present but is weak and rapidly converges to the equilibrium state [4, 7, 8, 10, 12–16], allows for exploring collective allele evolution while simplifying mathematical structures. However, the existing QLE theories have limitations. Populations can reach the QLE phase only

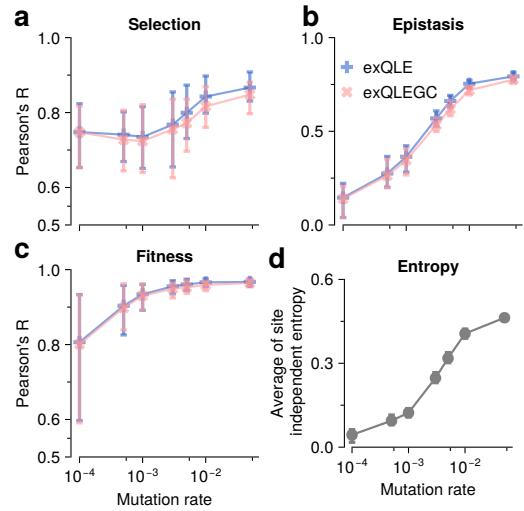


FIG. 2. Pearson's  $R$  values compare inferred and ground-truth fitness for selection coefficients (a), epistatic coefficients (b), and overall fitness (c), shown for  $K^* = 2$  (with  $K^* = 4$  in Fig. S2). The recombination rate was fixed at  $r = 3 \times 10^{-3}$  per generation per site, as recombination minimally affects inference accuracy, especially under exQLE. Inferred coefficients were obtained from temporal genetic sequences generated with the WF process (including recombination, mutation, and selection) based on the cumulant dynamics expressions (13) (or (15) for QLEGc). Inference of selection coefficients remains consistently accurate across mutation rates, whereas inference of epistatic coefficients strongly depends on mutation rates. This dependency reflects the effect of genetic diversity: higher mutation rates increase diversity, improving estimation of higher-order cumulants. Genetic diversity is quantified by the entropy profile, defined as the entropy of independent site frequency  $\nu_i = (1 + \chi_i)/2$ :  $S = -\frac{1}{L} \sum_{i=1}^L \langle \nu_i(t) \log(\nu_i(t)) \rangle_t$ , averaged every 50 generations between 1,500 and 2,000. The entropy profile closely parallels the  $R$ -value profile for epistatic inference, confirming the link between mutation-driven diversity and improved accuracy.

when the recombination rate is much larger than the selection [12], and epistasis is also smaller than the recombination rate [9], which significantly restricts the application of the QLE phase. In the inter-host evolution of SARS-CoV-2, selective pressures are likely strong, but the recombination rate is effectively zero, leading to significant LD [70]. Although the human immunodeficiency virus exhibits a higher recombination rate [71, 72], selective pressure can be significant, and often substantial LD persists over many years [49, 71, 73, 74]. Additionally, the recent progress in high-throughput deep mutational scanning to measure functional effects revealed the prevalence of widespread epistatic epistasis in viral and bacterial population in wild, and bacterial populations [27, 44, 75, 76]. Collectively, this evidence suggests that the QLE assumption violates conditions where LDs or higher-order cumulants of alleles across loci are unlikely to converge to equilibrium rapidly, and the timescales of

higher-order cumulants and individual allele frequencies are not well separated.

Here, we present an extension of QLE theory, exQLE, which generalizes QLE to allow cumulants up to any order  $K$  to evolve dynamically. To demonstrate this, we first expressed the cumulant dynamics for arbitrary orders and arbitrary genotype distributions (corresponding to (6)), providing a geometric interpretation of the dynamics and insights into when the expression can be exact. The resulting matrix  $D(\chi)$  servers as a geometric metric, and its general form is detailed in the Supplementary Information. The exQLE formulation naturally arises from (6) under conditions or genotype distributions where cumulants of order greater than  $K$  evolve rapidly toward equilibria, while the  $K$ -th order and lower cumulants remain dynamically changing. As an example of exQLE, we investigated the case  $K = 2$ , focusing on fitness functions whose averages depend on cumulants up to order  $K^* = 4$  (noting that the  $K^* = 2$  case is trivially exact). Under this condition, the  $K = 2$  exQLE framework can accurately demonstrate the cumulant dynamics, a capability that the standard QLE framework lacks. The derived cumulant dynamics also enable inference of fitness parameters by maximizing the likelihood of observed cumulant trajectories over evolutionary time, as in the marginal path-likelihood (MPL) method[49, 50, 58, 77].

Since the equations for cumulant dynamics are expressed as combinations of cumulants, their complexity can be systematically reduced by suppressing cumulants in a stratified, order-by-order manner. To achieve this, we employed an alternative Gaussian closure scheme, in which cumulants beyond the second order are assumed to be absent. The resulting novel family of inference methods accurately estimates selection and epistatic coefficients.

We could not fully explore the equilibrium distribution in the exQLE framework, in this work. Similar to the prior work [10], an explicit expression for the equilibrium distribution  $Q(\chi)$  could be derived from the forward Kolmogorov equation, which is characterized by  $K$ -th order cumulant dynamics, such that  $\partial_t Q(\chi, t) = \nabla_{\chi}^{\top} j(\chi, t)$ , where  $j(\chi, t)$  is the probability current. The equilibrium distribution can be obtained from the condition  $j(\chi) = \mathbf{0}$ . For example, when  $K = 2$ , the equilibrium distribution includes the interactions between first-order cumulants in an exponential function. Similar to the equilibrium distribution under QLE [10], the equilibrium distribution under exQLE also features an exponential form and an entropic term. Although, the general form of the equilibrium distribution remains conjectural, it likely depends on higher-order epistatic interactions between all possible combinations of cumulants up to order  $K$ , shaped by cumulants, mutation rates, and recombination rates, and appearing within the exponential function.

## ACKNOWLEDGMENTS

The work of K.S.S. and J.P.B. reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM138233. The work of M.P. was supported by the European Union (ERC, CHORAL, project number 101039794). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

# Supplementary Information

## STOCHASTIC PROCESSES IN POPULATION GENETICS

Following the convention of previous studies [10, 18], we represent genetic sequences as  $\mathbf{g} \in \{-1, +1\}^L$ , where  $L$  is the sequence length. This  $-/+$  encoding is commonly used in the physics literature. For a genotype  $a \in \{1, \dots, M\}$  with  $M = 2^L$ , and corresponding sequence  $\mathbf{g}^a$ , the fitness is defined as  $F^a = F(\mathbf{g}^a)$ . We consider the following general form for the fitness function:

$$F(\mathbf{g}) = \bar{F} + \sum_i s_i g_i + \sum_{i < j} s_{ij} g_i g_j + \dots, \quad (\text{S1})$$

where the fitness parameters  $s_i$  and  $s_{ij}$  represent selection coefficients and pairwise epistatic interactions, respectively. Interactions beyond pairwise, such as  $s_{ijk}, s_{ijkl}, \dots$ , can also occur, and we refer to these as higher-order epistatic interactions. For simplicity, we denote fitness parameters generically as  $s_e$ , where the index  $e$  may represent a single-site effect ( $i$ ), pairwise ( $i, j$ ), three-way ( $i, j, k$ ), four-way ( $i, j, k, l$ ), and so on.

To model the evolution of genetic sequences, we employed the Wright–Fisher (WF) process, a foundational stochastic model in population genetics that captures reproduction dynamics [1]. The WF process describes the evolution of a population with fixed size  $N$ . At time  $t_k$  (the parent generation), the population is represented by genotype counts  $(n_1(t_k), \dots, n_M(t_k))^\top = (n_a(t_k))_a^M$ , which give rise to the next generation at time  $t_{k+1}$ , denoted  $(n_a(t_{k+1}))_a^M$ .

Demographic noise, or genetic drift, arises from stochastic sampling and is on the order of  $\mathcal{O}(1/N)$ . It can be quantified through fluctuations in genotype frequencies, defined as  $\nu_a = n_a/N$ . Mathematically, the WF process is a discrete-time multinomial process. Additional evolutionary forces, including selection (with fitness  $F(\mathbf{g})$ ), mutation, and recombination, can be incorporated. Mutation and recombination occur at rates  $\mu$  and  $r$  per site per generation, respectively.

Let  $\boldsymbol{\nu} = \mathbf{n}/N$  denote the genotype frequency vector,  $p_a(\boldsymbol{\nu} \mid F, \mu, r)$  denote the probability that genotype  $a$  is selected, given the current genotype frequencies  $\boldsymbol{\nu}$ , fitness function  $F(\mathbf{g})$ , and evolutionary forces such as mutation and recombination (defined below). Then, the WF process can be expressed as:

$$p((\boldsymbol{\nu}(t_k))_{k=0}^K \mid F, \mu, r, N) = \prod_{k=0}^{K-1} p(\boldsymbol{\nu}(t_{k+1}) \mid \boldsymbol{\nu}(t_k); F, \mu, r, N), \quad (\text{S2})$$

where

$$p(\boldsymbol{\nu}(t_{k+1}) \mid \boldsymbol{\nu}(t_k); F, \mu, r, N) = N! \prod_a \frac{p_a(\boldsymbol{\nu}(t_k) \mid F, \mu, r)^{N\nu_a(t_{k+1})}}{[N\nu_a(t_{k+1})]!}. \quad (\text{S3})$$

Let  $y_a(\boldsymbol{\nu}; r)$  denote the probability that recombination events produce genotype  $a$ , which can be expressed as:

$$y_a(\boldsymbol{\nu}; r) = (1 - r)^{L-1} \nu_a + (1 - (1 - r)^{L-1}) \sum_{b,c} R_{a|b,c} \nu_b \nu_c. \quad (\text{S4})$$

The selection probability is then given by:

$$p_a(\boldsymbol{\nu} \mid F, \mu, r) = \frac{y_a(\boldsymbol{\nu}; r) F_a + \mu \sum_{b: d_{ab}=1} [y_b(\boldsymbol{\nu}; r) F_b - y_a(\boldsymbol{\nu}; r) F_a]}{\sum_b y_b(\boldsymbol{\nu}; r) F_b}. \quad (\text{S5})$$

In this expression,  $d_{ab}$  denotes the Hamming distance between genotypes  $a$  and  $b$ . When the mutation rate  $\mu$  is low, at most one mutation is expected per individual per generation. As a result, the contribution to the mutation flux in the numerator of (S5) comes only from genotype pairs that differ by a single mutation (i.e.,  $d_{ab} = 1$ ).

### Details of simulation conditions

To examine a non-trivial scenario, we considered a higher-order fitness function defined as

$$F(\mathbf{g}) = \bar{F} + \sum_i s_i g_i + \sum_{i < j} s_{ij} g_i g_j + \sum_{i < j < k} s_{ijk} g_i g_j g_k + \sum_{i < j < k < l} s_{ijkl} g_i g_j g_k g_l. \quad (\text{S6})$$

where the fitness coefficients  $s_e \in \{-0.03, 0, 0.03\}$  for indices  $e = i, (i, j), (i, j, k), (i, j, k, l)$ . The number of nonzero coefficients  $s_e$  was kept at  $\mathcal{O}(L)$  across orders one through four, with  $L$  denoting the sequence length. In Fig. 1, to assess the influence of higher-order cumulants on trait dynamics, we also considered random traits defined as

$$G^{\text{Rand}}(\mathbf{g}) = \sum_i a_i g_i + \sum_{i < j} a_{ij} g_i g_j, \quad (\text{S7})$$

where  $a_i \sim \mathcal{N}(0, 1/L)$  and  $a_{ij} \sim \mathcal{N}(0, 2/L(L-1))$  for all  $i, j$ . These coefficients were independently sampled.

The simulation conditions of **Fig. 1** are as follows. Fitness parameters were drawn from the set  $\{-0.03, 0, 0.03\}$ , while maintaining the number of nonzero parameters at  $\mathcal{O}(L)$  for each order, where  $L = 100$  is the sequence length. The recombination rate was fixed at  $r = 3 \times 10^{-3}$  per site per generation, as the results were robust across different recombination rates. In contrast, the simulation outcomes depend on the mutation rate, which we varied between  $10^{-4}$  and  $0.05$  per site per generation. The population size was set to  $N = 10^3$ . As we observed no significant variation in cumulant dynamics after  $10^3$  generations, we sampled genetic sequences every 200 generations between  $10^3$  and  $2 \times 10^3$  generations.

### COMPUTATION OF $D_{\mathcal{I}, \mathcal{J}}$ MATRIX

We now provide a more explicit expression for  $D_{\mathcal{I}, \mathcal{J}}$ . As we noted in the main text,  $\mathcal{I}, \mathcal{J}, \mathcal{K}$  are multi-indices over loci, and cumulants and moments of arbitrary order are defined as  $\chi_{\mathcal{I}}^\phi = \partial_{\phi_{\mathcal{I}}} \Phi$  and  $\mu_{\mathcal{I}}^\phi = e^{-\Phi} \partial_{\phi_{\mathcal{I}}} e^\Phi$ , respectively. The general form of cumulant dynamics is given as

$$\begin{aligned} \partial_{\phi_{\mathcal{I}}} \langle F \rangle_{\phi} |_{\phi=0} &= \sum_{\mathcal{J}} D_{\mathcal{I}, \mathcal{J}} \frac{\partial \langle F \rangle}{\partial \chi_{\mathcal{J}}} \\ D_{\mathcal{I}, \mathcal{J}} &= \sum_{\mathcal{K}} \frac{\partial \chi_{\mathcal{J}}}{\partial \mu_{\mathcal{K}}} \left. \frac{\partial \mu_{\mathcal{K}}^\phi}{\partial \phi_{\mathcal{I}}} \right|_{\phi=0}. \end{aligned} \quad (\text{S8})$$

We express moments  $\mu_{\mathcal{K}}$  in terms of cumulants since derivatives of cumulants with respect to  $\phi$  are more tractable. Specifically, moments can be written as:

$$\mu_{\mathcal{K}} = \sum_{\pi \in \mathcal{P}(\mathcal{K})} \prod_{B \in \pi} \chi_B. \quad (\text{S9})$$

where  $\mathcal{P}(\mathcal{K})$  denotes all partitions of the index set  $\mathcal{K}$ . For example, if  $\mathcal{K} = \{k_1\}$ , then  $\mathcal{P}(\mathcal{K}) = \{\{k_1\}\}$ . For  $\mathcal{K} = \{k_1, k_2\}$ ,  $\mathcal{P}(\mathcal{K}) = \{\{k_1, k_2\}, \{\{k_1\}, \{k_2\}\}\}$ , and so on.

Therefore,

$$\left. \frac{\partial \mu_{\mathcal{K}}^\phi}{\partial \phi_{\mathcal{I}}} \right|_{\phi=0} = \sum_{\pi \in \mathcal{P}(\mathcal{K})} \left[ \partial_{\phi_{\mathcal{I}}} \prod_{B \in \pi} \chi_B^\phi \right]_{\phi=0}. \quad (\text{S10})$$

To relate cumulants to moments, we use the Faà di Bruno formula:

$$\chi_{\mathcal{J}} = \sum_{\pi \in \mathcal{P}(\mathcal{J})} (-1)^{|\pi|-1} (|\pi|-1)! \prod_{B \in \pi} \mu_B, \quad (\text{S11})$$

from which, we obtain the derivative of cumulants with respect to moments:

$$\frac{\partial \chi_{\mathcal{J}}}{\partial \mu_{\mathcal{K}}} = \sum_{\pi \in \mathcal{P}(\mathcal{J})} (-1)^{|\pi|-1} (|\pi|-1)! \frac{\partial}{\partial \mu_{\mathcal{K}}} \prod_{B \in \pi} \mu_B. \quad (\text{S12})$$

Thus, combining (S12) and (S10), we obtain the explicit expression of  $D_{\mathcal{I}, \mathcal{J}}$  in (S8). Below, we demonstrate this for specific cases under a pairwise fitness function.

$K = 1$  case

Let us consider a simple  $\mathcal{I} = \{i\}$ ,  $\mathcal{J} = \{j\}$  case. The only partition for  $\mathcal{K} = \{j\}$  is  $\mathcal{P}(\mathcal{K}) = \{\{j\}\}$ , and we obtain:

$$\frac{\partial \chi_j}{\partial \mu_{\mathcal{K}}} = \delta_{\{j\}, \mathcal{K}},$$

where  $\delta_{\mathcal{K}, \mathcal{L}}$  returns 1 if  $\mathcal{K} = \mathcal{L}$ , otherwise returns 0. Thus, the  $D$  matrix reduces to:

$$D_{i,j} = \sum_{\mathcal{K}} \delta_{\{j\}, \mathcal{K}} \sum_{\pi \in \mathcal{P}(\mathcal{K})} \left[ \partial_{\phi_i} \prod_{B \in \pi} \chi_B^\phi \right]_{\phi=0} = \sum_{\mathcal{K}} \delta_{\{j\}, \mathcal{K}} \partial_{\phi_i} \chi_j^\phi |_{\phi=0} = \chi_{ij}.$$

where only  $\mathcal{K} = \{j\}$  contributes. Therefore, this result is consistent with the  $K = 1$  case.

$K = 2$  case

Since the expression for  $D_{ij}$  matches the  $K = 1$  case, we consider three additional cases:  $\mathcal{I} = \{i, j\}$ ,  $\mathcal{J} = \{k\}$ ;  $\mathcal{I} = \{i\}$ ,  $\mathcal{J} = \{k, l\}$ ; and  $\mathcal{I} = \{i, j\}$ ,  $\mathcal{J} = \{k, l\}$ .

For  $\mathcal{I} = \{i, j\}$ ,  $\mathcal{J} = \{k\}$ , we have seen that  $\frac{\partial \chi_k}{\partial \mu_{\mathcal{K}}} = \delta_{j,k}$ . from the example in  $K = 1$  case. Therefore, we have

$$D_{ij,k} = \sum_{\mathcal{K}} \delta_{\{k\}, \mathcal{K}} \left[ \partial_{\phi_i} \partial_{\phi_j} \chi_k^\phi \right]_{\phi=0} = \chi_{ijk}.$$

For  $\mathcal{I} = \{i\}$ ,  $\mathcal{J} = \{k, l\}$ , the cumulant-moment relation yields,

$$\frac{\partial \chi_{kl}}{\partial \mu_{\mathcal{K}}} = -\chi_l \delta_{\{k\}, \mathcal{K}} - \chi_k \delta_{\{l\}, \mathcal{K}} + \delta_{\{k, l\}, \mathcal{K}}.$$

Therefore,

$$\begin{aligned} D_{i,kl} &= \sum_{\mathcal{K}} (-\chi_l \delta_{\{k\}, \mathcal{K}} - \chi_k \delta_{\{l\}, \mathcal{K}} + \delta_{\{k, l\}, \mathcal{K}}) \left[ \partial_{\phi_i} \sum_{\pi \in \mathcal{P}(\mathcal{K})} \prod_{B \in \pi} \chi_B^\phi \right]_{\phi=0} \\ &= -\chi_l \chi_{ik} - \chi_k \chi_{il} + \left[ \partial_{\phi_i} (\chi_{kl}^\phi + \chi_k^\phi \chi_l^\phi) \right]_{\phi=0} \\ &= -\chi_l \chi_{ik} - \chi_k \chi_{il} + \chi_{ikl} + \chi_l \chi_{ik} + \chi_k \chi_{il} = \chi_{ikl}. \end{aligned}$$

For  $\mathcal{I} = \{i, j\}$ ,  $\mathcal{J} = \{k, l\}$  case,

$$\begin{aligned} D_{ij,kl} &= \sum_{\mathcal{K}} (-\chi_l \delta_{\{k\}, \mathcal{K}} - \chi_k \delta_{\{l\}, \mathcal{K}} + \delta_{\{k, l\}, \mathcal{K}}) \left[ \partial_{\phi_i} \partial_{\phi_j} \sum_{\pi \in \mathcal{P}(\mathcal{K})} \prod_{B \in \pi} \chi_B^\phi \right]_{\phi=0} \\ &= -\chi_l \chi_{ijk} - \chi_k \chi_{ijl} + \left[ \partial_{\phi_i} \partial_{\phi_j} (\chi_{kl}^\phi + \chi_k^\phi \chi_l^\phi) \right]_{\phi=0} \\ &= -\chi_l \chi_{ijk} - \chi_k \chi_{ijl} + \chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} + \chi_l \chi_{ijk} + \chi_k \chi_{ijl} \\ &= \chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk}. \end{aligned}$$

In summary, we recover the full  $D(\chi)$  matrix,

$$D(\chi) = \begin{pmatrix} D_{i,k} & D_{i,kl} \\ D_{ij,k} & D_{ij,kl} \end{pmatrix} = \begin{pmatrix} \chi_{ik} & \chi_{ikl} \\ \chi_{ijk} & \chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} \end{pmatrix},$$

consistent with the expression in (9).

**DERIVATION OF EXPLICIT EXPRESSIONS FOR FIRST- AND SECOND-ORDER CUMULANTS  
UNDER PAIRWISE FITNESS FUNCTION**

**Exact calculation**

Here, we derive explicit expressions for the first- and second-order cumulants' equations of motion under the pairwise fitness function and demonstrate that the results derived from the exQLE yield the exact results. Pairwise fitness is defined as:

$$\begin{aligned} F(g) &= \bar{F} + \sum_k s_k g_k + \sum_{k < l} s_{kl} g_k g_l \\ &\quad \bar{F} + \sum_k s_k g_k + \sum_{k < l} s_{kl} (g_k - \chi_k)(g_l - \chi_l) + \sum_{k < l} s_{kl} (g_k - \chi_k) \chi_l + \sum_{k < l} s_{kl} \chi_k (g_l - \chi_l) + \sum_{k < l} s_{kl} \chi_k \chi_l . \end{aligned} \quad (\text{S13})$$

Derivatives yield:

$$\begin{aligned} \partial_{\chi_i} \langle F \rangle &= s_i + \sum_{l; l > i} s_{il} \chi_l + \sum_{k; k < i} s_{ki} \chi_k , \\ \partial_{\chi_{ij}} \langle F \rangle &= s_{ij} . \end{aligned} \quad (\text{S14})$$

Subtracting,

$$F(g) - \langle F \rangle = \sum_k s_k (g_k - \chi_k) + \sum_{k < l} s_{kl} [(g_k - \chi_k)(g_l - \chi_l) + (g_k - \chi_k) \chi_l + \chi_k (g_l - \chi_l) - \chi_{kl}] \quad (\text{S15})$$

Taking expectations:

$$\begin{aligned} \langle F(g) - \langle F \rangle \rangle &= 0 , \\ \langle (g_i - \chi_i)(F(g) - \langle F \rangle) \rangle &= \sum_k s_k \chi_{ik} + \sum_{k < l} s_{kl} (\chi_{ikl} + \chi_{ik} \chi_l + \chi_k \chi_{il}) , \\ \langle (g_i - \chi_i)(g_j - \chi_j)(F(g) - \langle F \rangle) \rangle &= \sum_k s_k \chi_{ijk} + \sum_{k < l} s_{kl} (m_{ijkl} + \chi_{ijk} \chi_l + \chi_k \chi_{ijl} - \chi_{ij} \chi_{kl}) . \end{aligned} \quad (\text{S16})$$

where the fourth central moment is given by:

$$m_{ijkl} = \chi_{ijkl} + \chi_{ij} \chi_{kl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} . \quad (\text{S17})$$

Therefore, the exact equations of motion are:

$$\begin{aligned} \dot{\chi}_i &= \sum_k s_k \chi_{ik} + \sum_{k < l} s_{kl} (\chi_{ikl} + \chi_{ik} \chi_l + \chi_k \chi_{il}) , \\ \dot{\chi}_{ij} &= \sum_k s_k \chi_{ijk} + \sum_{k < l} s_{kl} (\chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} + \chi_{ijk} \chi_l + \chi_k \chi_{ijl}) . \end{aligned} \quad (\text{S18})$$

**exQLE calculation**

As the following calculations are valid for both the Gibbs distribution, and the cumulant distribution of any arbitrary distribution after taking the limit of  $\phi \rightarrow \mathbf{0}$ , we assume that the genotype distribution takes the form of a Gibbs distribution for simplicity. For further simplicity, we drop  $\phi$  from  $\chi_i^\phi$ ,  $\chi_{ij}^\phi$ , and also omit the operation of taking the limit as  $\phi \rightarrow \mathbf{0}$ . The equation of motion for the first-order cumulant (7) is straightforward to obtain, using the relationships  $\partial_{\phi_i} \chi_k = \chi_{ik}$  and  $\partial_{\phi_i} \chi_{kl} = \chi_{ikl}$ , which arise from the properties of the cumulant generating function or the normalization of the Gibbs distribution.

Here, we focus on the equation of motion for the second-order cumulant. The direct calculation of the exQLE for

the second-order cumulants yields (8), which is given as:

$$\begin{aligned}
\partial_{\phi_i} \partial_{\phi_j} \langle F \rangle &= \partial_{\phi_j} \left( \sum_k (\partial_{\phi_i} \chi_k) \partial_{\chi_k} \langle F \rangle + \sum_{k < l} (\partial_{\phi_i} \chi_{kl}) \partial_{\chi_{kl}} \langle F \rangle \right) \\
&= \sum_k (\partial_{\phi_j} \partial_{\phi_i} \chi_k) \partial_{\chi_k} \langle F \rangle + \sum_{k < l} (\partial_{\phi_j} \partial_{\phi_i} \chi_{kl}) \partial_{\chi_{kl}} \langle F \rangle + \sum_k (\partial_{\phi_i} \chi_k) \partial_{\phi_j} \partial_{\chi_k} \langle F \rangle + \sum_{k < l} (\partial_{\phi_i} \chi_{kl}) \partial_{\phi_j} \partial_{\chi_{kl}} \langle F \rangle^0 \\
&= \sum_k \chi_{ijk} \partial_{\chi_k} \langle F \rangle + \sum_{k < l} \chi_{ijkl} \partial_{\chi_{kl}} \langle F \rangle + \sum_k \chi_{ik} \left( \sum_l \chi_{jl} \partial_{\chi_l} \partial_{\chi_k} \langle F \rangle + \sum_{l < m} \chi_{jlm} \partial_{\chi_{lm}} \partial_{\chi_k} \langle F \rangle \right)^0 \\
&= \sum_k \chi_{ijk} \partial_{\chi_k} \langle F \rangle + \sum_{k < l} (\chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk}) \partial_{\chi_{kl}} \langle F \rangle.
\end{aligned} \tag{S19}$$

Here, we used the fact that the derivative of the average fitness beyond the second-order cumulant vanishes. Additionally we used:  $\partial_{\chi_k} \partial_{\chi_l} \langle F \rangle = \partial_{\chi_{kl}} \langle F \rangle$  and

$$\sum_{k,l} \chi_{ik} \chi_{jl} \partial_{\chi_{kl}} \langle F \rangle = \sum_{k < l} (\chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk}) \partial_{\chi_{kl}} \langle F \rangle. \tag{S20}$$

The equations of motion for exQLE, corresponding to (7) and (8), lead to:

$$\begin{aligned}
\dot{\chi}_i &= \sum_k \chi_{ik} \partial_{\chi_k} \langle F \rangle + \sum_{k < l} \chi_{ikl} \partial_{\chi_{kl}} \langle F \rangle \\
&= \sum_k \chi_{ik} \left( s_k + \sum_{l;l > k} s_{kl} \chi_l + \sum_{l;l < k} s_{lk} \chi_l \right) + \sum_{k < l} s_{kl} \chi_{ikl} \\
&= \sum_k \chi_{ik} s_k + \sum_{k < l} s_{kl} (\chi_{ik} \chi_l + \chi_{il} \chi_k + \chi_{ikl}),
\end{aligned} \tag{S21}$$

and

$$\begin{aligned}
\dot{\chi}_{ij} &= \sum_k \chi_{ijk} \partial_{\chi_k} \langle F \rangle + \sum_{k < l} (\chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk}) \partial_{\chi_{kl}} \langle F \rangle \\
&= \sum_k \chi_{ijk} \left( s_k + \sum_{l;l > k} s_{kl} \chi_l + \sum_{l;l < k} s_{lk} \chi_l \right) + \sum_{k < l} s_{kl} (\chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk}) \\
&= \sum_k s_k \chi_{ijk} + \sum_{k < l} s_{kl} (\chi_{ijk} \chi_l + \chi_{ijl} \chi_k + \chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk}).
\end{aligned} \tag{S22}$$

Here, where we used:

$$\sum_k \chi_{ik} \sum_{l;l < k} s_{lk} \chi_l = \sum_{k < l} s_{kl} \chi_{il} \chi_k, \tag{S23}$$

by simply renaming indices. The two sets of equations are identical.

Therefore, we directly confirmed that the equations of motion for the first and second cumulants, derived from the exact Price's equation and the exQLE equation, yield identical results in the case of pairwise fitness.

### PROOF OF POSITIVE SEMIDEFINITENESS OF $D$ MATRIX FOR $K \in \{1, 2\}$

For the  $K = 1$  case,  $D(\chi)$  is trivially positive semidefinite. For  $K = 2$  case, by defining  $\Delta_i := g_i - \chi_i$ , it can be expressed as

$$\begin{aligned} D(\chi) &= \begin{pmatrix} \chi_{ik} & \chi_{ikl} \\ \chi_{ijk} & \chi_{ijkl} + \chi_{ik}\chi_{jl} + \chi_{il}\chi_{jk} \end{pmatrix} \\ &= \left\langle \begin{pmatrix} \Delta_i\Delta_j & \Delta_i\Delta_k\Delta_l \\ \Delta_i\Delta_j\Delta_k & (\Delta_i\Delta_j - \chi_{ij})(\Delta_k\Delta_l - \chi_{kl}) \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} \Delta_i & \Delta_k \\ \Delta_i\Delta_j - \chi_{ij} & \Delta_k\Delta_l - \chi_{kl} \end{pmatrix}^\top \right\rangle \succeq 0. \end{aligned} \quad (\text{S24})$$

From the second line to the third line, we used the fact  $\langle \Delta_i\Delta_j\Delta_k \rangle = \langle (\Delta_i\Delta_j - \chi_{ij})\Delta_k \rangle$ . Therefore,  $D(\chi)$  is positive semidefinite.

### DERIVATION OF THE CUMULANT DYNAMICS FROM GENOTYPE DYNAMICS

We denote the genotype distribution  $P(\mathbf{g})$  for  $\mathbf{g} \in \{-1, 1\}^L$  as  $P(\mathbf{g}^a) \mapsto z_a$ , where each unique genotype is indexed by  $a \in \{1, \dots, 2^L\}$ . We also denote genotype-level selection coefficient as  $h_a$ , such that the fitness function satisfies  $F(\mathbf{g}^a) = \bar{F} + h_a$ .

The average fitness is then given by

$$\langle F \rangle = \bar{F} + \sum_a h_a z_a = \bar{F} + \sum_{\mathcal{J}} s_{\mathcal{J}} \mu_{\mathcal{J}}. \quad (\text{S25})$$

Here,  $\mu_{\mathcal{J}}$  denote a moment indexed by the set  $\mathcal{J}$  (which may include multiple indices), and given by

$$\mu_{\mathcal{J}} = \sum_a z_a \prod_{j \in \mathcal{J}} g_j^a. \quad (\text{S26})$$

Let define the matrix,

$$\mathcal{G}_{a,\mathcal{J}} = \prod_{j \in \mathcal{J}} g_j^a, \quad (\text{S27})$$

so that the moments and genotype distributions are related via  $\boldsymbol{\mu} = \mathcal{G}^\top \mathbf{z}$ .

Given Fisher's fundamental theorem, the genotype distribution evolves as

$$\dot{z}_a = \sum_{ab} C_{ab}(\mathbf{z}) h_b \quad (\text{S28})$$

where

$$C_{ab}(\mathbf{z}) = z_a \delta_{a,b} - z_a z_b, \quad (\text{S29})$$

is the covariance matrix of genotype frequencies. In the vector form, it can be expressed as  $\dot{\mathbf{z}} = C(\mathbf{z})\mathbf{h} = C(\mathbf{z})\nabla_{\mathbf{z}} \langle F \rangle$ .

Since  $\langle F \rangle$  depends on  $\mathbf{z}$  via the moments  $\boldsymbol{\mu}$ , we apply the chain rule:

$$\frac{\partial \langle F \rangle}{\partial z_a} = \sum_{\mathcal{J}} \frac{\partial \mu_{\mathcal{J}}}{\partial z_a} \frac{\partial \langle F \rangle}{\partial \mu_{\mathcal{J}}} = \sum_{\mathcal{J}} \mathcal{G}_{a,\mathcal{J}} \frac{\partial \langle F \rangle}{\partial \mu_{\mathcal{J}}}, \quad (\text{S30})$$

which can be expressed in vector form  $\nabla_{\mathbf{z}} \langle F \rangle = \mathcal{G} \nabla_{\boldsymbol{\mu}} \langle F \rangle$ .

Substituting into the dynamics:

$$\dot{\mathbf{z}} = C(\mathbf{z}) \mathcal{G} \nabla_{\boldsymbol{\mu}} \langle F \rangle. \quad (\text{S31})$$

Using the relation  $\boldsymbol{\mu} = \mathcal{G}^\top \mathbf{z}$ , the moment dynamics become

$$\dot{\boldsymbol{\mu}} = \mathcal{G}^\top C(\mathbf{z}) \mathcal{G} \nabla_{\boldsymbol{\mu}} \langle F \rangle, \quad (\text{S32})$$

which can be further transformed by using the chain rule:

$$\dot{\boldsymbol{\mu}} = \mathcal{G}^\top C(\boldsymbol{z}) \mathcal{G} (\nabla_{\boldsymbol{\mu}} \boldsymbol{\chi}^\top) \nabla_{\boldsymbol{\chi}} \langle F \rangle. \quad (\text{S33})$$

Transforming from moments to cumulants using the Jacobian matrix:  $d\boldsymbol{\mu} = (\nabla_{\boldsymbol{\chi}} \boldsymbol{\mu}^\top)^\top d\boldsymbol{\chi}$ . Substituting this into the dynamics gives:  $\dot{\boldsymbol{\mu}} = (\nabla_{\boldsymbol{\chi}} \boldsymbol{\mu}^\top)^\top \dot{\boldsymbol{\chi}}$ .

By using the relation,

$$((\nabla_{\boldsymbol{\chi}} \boldsymbol{\mu}^\top)^\top)^{-1} = ((\nabla_{\boldsymbol{\chi}} \boldsymbol{\mu}^\top)^{-1})^\top = (\nabla_{\boldsymbol{\mu}} \boldsymbol{\chi}^\top)^\top,$$

we obtain the cumulant dynamics:

$$\dot{\boldsymbol{\chi}} = D(\boldsymbol{\chi}) \nabla_{\boldsymbol{\chi}} \langle F \rangle, \quad (\text{S34})$$

where

$$D(\boldsymbol{\chi}) = (\nabla_{\boldsymbol{\mu}} \boldsymbol{\chi}^\top)^\top \mathcal{G}^\top C(\boldsymbol{z}) \mathcal{G} (\nabla_{\boldsymbol{\mu}} \boldsymbol{\chi}^\top). \quad (\text{S35})$$

This is the matrix  $D(\boldsymbol{\chi})$  appearing in (6) of the main text. Its origin lies in  $C(\boldsymbol{z})$ , which arises from competition between genotypes and also serves as the covariance matrix in the diffusion process. Importantly, since  $C(\boldsymbol{z})$  is symmetric and positive semidefinite, the matrix  $D(\boldsymbol{\chi})$  inherits these properties, it is symmetric and positive semidefinite as well.

## DERIVATION OF THE EQUATION TO INFER FITNESS VALUES

We now derive the equation used to infer fitness parameters from cumulant dynamics.

The cumulant dynamics described in (9) have been deterministic, assuming an infinitely large population. For a finite population of size  $N$ , however, stochastic effects must be considered. In this case, the dynamics become a Langevin equation:

$$\dot{\boldsymbol{\chi}} = D(\boldsymbol{\chi}) \nabla_{\boldsymbol{\chi}} \langle F \rangle + \sqrt{D(\boldsymbol{\chi}(t))/N} \boldsymbol{\eta}(t). \quad (\text{S36})$$

where  $\boldsymbol{\eta}(t)$  is a noise vector satisfying  $\langle \boldsymbol{\eta} \rangle = \mathbf{0}$  and  $\langle \boldsymbol{\eta}_{\mathcal{J}}(t) \boldsymbol{\eta}_{\mathcal{K}}(t') \rangle = \delta_{\mathcal{J}, \mathcal{K}} \delta(t - t')$ .

This Langevin equation is equivalent to the following Fokker–Planck equation [78]:

$$\partial_t P(\boldsymbol{\chi}, t) = -\nabla_{\boldsymbol{\chi}}^\top D(\boldsymbol{\chi}) \nabla_{\boldsymbol{\chi}} \langle F \rangle P(\boldsymbol{\chi}, t) + \frac{N}{2} \text{Tr} (\nabla_{\boldsymbol{\chi}} \nabla_{\boldsymbol{\chi}}^\top D(\boldsymbol{\chi})) P(\boldsymbol{\chi}, t). \quad (\text{S37})$$

The Fokker–Planck equation can be rewritten as a probability density over entire cumulant trajectories. For numerical implementation, we discretize time at points  $t_k$  for  $k \in \{0, 1, \dots, K+1\}$ , and define  $\Delta t_k := t_{k+1} - t_k$  and  $\Delta \boldsymbol{\chi}(t_k) := \boldsymbol{\chi}(t_{k+1}) - \boldsymbol{\chi}(t_k)$ .

The probability of a cumulant trajectory is then expressed as

$$P((\boldsymbol{\chi}(t_k))_{k=0}^{K+1}) \propto e^{-N\mathcal{S}((\boldsymbol{\chi}(t_k))_{k=0}^{K+1})}, \quad (\text{S38})$$

where  $\mathcal{S}$  is given by

$$\begin{aligned} \mathcal{S}((\boldsymbol{\chi}(t_k))_{k=0}^{K+1}) &= \sum_{k=0}^K \frac{1}{2\Delta t_k} \left[ \Delta \boldsymbol{\chi}(t_k) - \Delta t_k D(\boldsymbol{\chi}(t_k)) \nabla_{\boldsymbol{\chi}} \langle F \rangle \right]^\top \\ &\quad \times D^{-1}(\boldsymbol{\chi}(t_k)) \left[ \Delta \boldsymbol{\chi}(t_k) - \Delta t_k D(\boldsymbol{\chi}(t_k)) \nabla_{\boldsymbol{\chi}} \langle F \rangle \right]. \end{aligned} \quad (\text{S39})$$

Since  $\mathcal{S}$  is quadratic in  $\nabla_{\boldsymbol{\chi}} \langle F \rangle$ , the maximum likelihood estimate of the fitness parameters can be obtained analytically by solving the following linear equation:

$$\sum_{k=0}^K \Delta \boldsymbol{\chi}(t_k) = \sum_{k=0}^K \Delta t_k D(\boldsymbol{\chi}(t_k)) \nabla_{\boldsymbol{\chi}} \langle F \rangle. \quad (\text{S40})$$

By solving the above maximum likelihood equation, which is linear in  $\boldsymbol{s}$ , we can obtain  $\boldsymbol{s}$ .

## EFFICIENT COMPUTATION OF THE FORWARD EXQLE SIMULATION UNDER HIGHER-ORDER FITNESS

To obtain cumulant dynamics using either the exQLE framework or the exQLE with a Gaussian closure scheme, we must evaluate the products of the diffusion matrix, which involve third and fourth-order cumulants and fitness parameters. However, a direct computation of the diffusion matrix and its product with fitness parameters is computationally expensive. Estimating these values across multiple time points, on the order of 100 different time points, requires significant computational time. Additionally, incorporating a fitness function that depends on four-way interactions further increases computational complexity. To improve computational efficiency, we outline an optimized approach for computing cumulant dynamics.

Denote the gradient of the average fitness, which serves as the effective fitness parameter vector consisting of selection and epistatic coefficients, as  $\begin{pmatrix} \hat{s}_k \\ \hat{s}_{kl} \end{pmatrix} := \begin{pmatrix} \partial_{\chi_k} \\ \partial_{\chi_{kl}} \end{pmatrix} \langle F \rangle$ . For the fitness function with four-way interactions, let  $\mu_i, \mu_{ij}, \dots$  denote mutation frequencies, and let  $\langle F \rangle = \bar{F} + \sum_i s_i \mu_i + \sum_{i < j} s_{ij} \mu_{ij} + \sum_{i < j < k} s_{ijk} \mu_{ijk} + \sum_{i < j < k < l} s_{ijkl} \mu_{ijkl}$  represent the effective fitness parameters as

$$\begin{pmatrix} \hat{s}_k \\ \hat{s}_{kl} \end{pmatrix} = \begin{pmatrix} s_k + \sum_{i < k} s_{ik} \mu_i + \sum_{i < j; i < k} s_{ijk} \mu_{ij} + \sum_{i < j < l; i < k} s_{ijkl} \mu_{ijl} \\ s_{kl} + \sum_{i; i < k < l} s_{ikl} \mu_i + \sum_{i < j; i < k < l} s_{ijkl} \mu_{ij} \end{pmatrix}. \quad (\text{S41})$$

These products between fitness parameters and moments can be computed as the sum of matrix-vector products,  $s_{ijk} \mu_{ij} = \sum_{i < j < k} \langle s_{ijk} g_i g_j \rangle$  and  $s_{ijkl} \mu_{ijl} = \sum_{i < j < l < k} \langle s_{ijkl} g_i g_j g_l \rangle$ .

Let  $\Delta_i := g_i - \chi_i$ , then the cumulants can be obtained by  $\chi_i = \langle \Delta_i \rangle, \chi_{ij} = \langle \Delta_i \Delta_j \rangle, \chi_{ijk} = \langle \Delta_i \Delta_j \Delta_k \rangle, \chi_{ijkl} = \langle \Delta_i \Delta_j \Delta_k \Delta_l \rangle - (\chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} + \chi_{ij} \chi_{kl})$ . Additionally, define  $\chi^{(2)}$  as the second-order cumulant in matrix form and  $\hat{S}$  as the effective epistasis matrix, where  $\hat{s}_{kl}$  occupies the  $k$ -th row and  $l$ -th column. Thus, the cumulant dynamics can be expressed in the following computationally more efficient form:

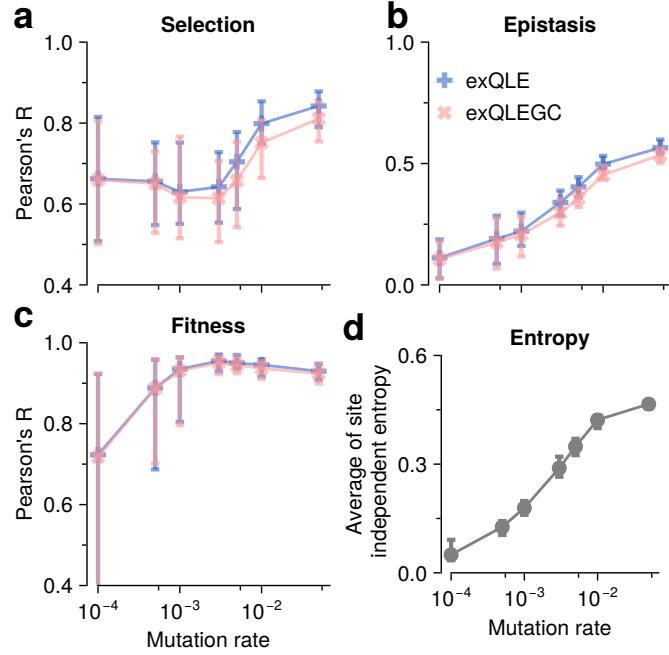
$$\begin{pmatrix} \dot{\chi}_i \\ \dot{\chi}_{ij} \end{pmatrix} = \begin{pmatrix} \chi_{ik} & \chi_{ikl} \\ \chi_{ijk} & \chi_{ijkl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} \end{pmatrix} \begin{pmatrix} \hat{s}_k \\ \hat{s}_{kl} \end{pmatrix} = \begin{pmatrix} \langle \delta(\Delta^\top \hat{s}) \rangle + \langle \Delta(\Delta^\top \hat{S} \Delta) \rangle / 2 \\ \langle \Delta \Delta^\top (\Delta^\top \hat{s}) \rangle + \langle \Delta \Delta^\top \Delta^\top \hat{S} \Delta \rangle - \chi^{(2)} \text{Sum}(\chi^{(2)} \odot \hat{S}) \end{pmatrix}. \quad (\text{S42})$$

The last expression in (S42) is efficient because this computation never explicitly requires obtaining matrices or matrix products with more than  $\mathcal{O}(L^2)$  elements.

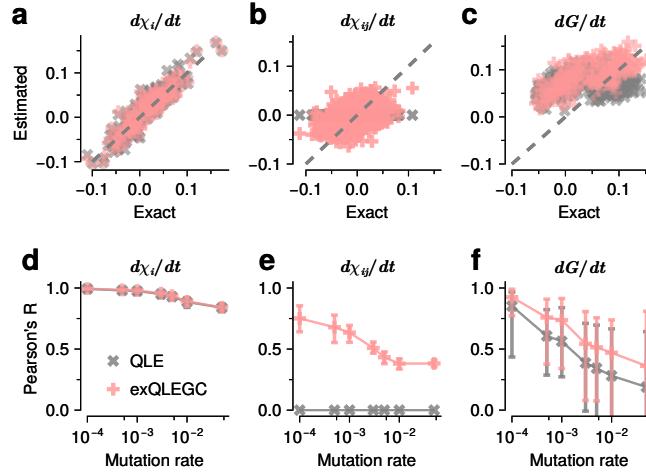
For the Gaussian closure scheme, the cumulant dynamics can be obtained in the same manner. The dynamics under the Gaussian closure scheme are given by:

$$\begin{aligned} \begin{pmatrix} \dot{\chi}_i \\ \dot{\chi}_{ij} \end{pmatrix} &= \begin{pmatrix} \chi_{ik} & \chi_{ikl}(\delta_{ik} + \delta_{jk}) \\ \chi_{ijk}(\delta_{ik} + \delta_{jk}) & \chi_{ijkl}\delta_{ik}\delta_{jl} + \chi_{ik}\chi_{jl} + \chi_{il}\chi_{jk} \end{pmatrix} \begin{pmatrix} \hat{s}_k \\ \hat{s}_{kl} \end{pmatrix} \\ &= \begin{pmatrix} \chi_{ik} & \chi_{ikl}(\delta_{ik} + \delta_{jk}) \\ \chi_{ijk}(\delta_{ik} + \delta_{jk}) & [\langle \Delta_i \Delta_j \Delta_i \Delta_j \rangle - \chi_{ii} \chi_{jj} - 2\chi_{ij}^2] \delta_{ik} \delta_{jl} + \chi_{ik} \chi_{jl} + \chi_{il} \chi_{jk} \end{pmatrix} \begin{pmatrix} \hat{s}_k \\ \hat{s}_{kl} \end{pmatrix} \\ &= \begin{pmatrix} \langle \Delta(\Delta^\top \hat{s}) \rangle + \left\langle \Delta \odot \left( \sum_k \left\{ (\Delta \Delta^\top \odot \hat{S})_{ik} + (\Delta \Delta^\top \odot \hat{S})_{ki} \right\} \right)_{i=1}^L \right\rangle \\ \langle \Delta \Delta^\top \odot (\mathbf{1}(\Delta \odot \hat{s})^\top + (\Delta \odot \hat{s}) \mathbf{1}^\top) \rangle + \text{diag} \left( \left( (\langle \Delta_i^2 \Delta_j^2 \rangle - \chi_{ii} \chi_{jj} - 2\chi_{ij}^2) s_{ij} \right)_{i < j} \right) + \chi^{(2)} \hat{S} \chi^{(2)} \end{pmatrix}, \end{aligned} \quad (\text{S43})$$

where  $\mathbf{1}$  is a vector of length  $L$  consisting entirely of ones,  $\odot$  denotes the elementwise product, and the notation  $(a_i)_i^L = (a_1, \dots, a_L)^\top$  is used.



**FIG. S1. Accuracy of inferred fitness parameters under higher-order fitness function.** This corresponds to **Fig. 2** in the main text, but here the underlying selective pressure is determined by a higher-order fitness function rather than a pairwise one, that is,  $K^* = 4$ , where  $K^*$  denotes the highest order of cumulants in the averaged fitness function (as defined in the main text). The functionality and model parameters are the same as those used in the simulation for **Fig. 1**. The inference approach remains pairwise ( $K = 2$ ), aiming to infer additive (selection) and pairwise (epistatic) fitness parameters. Despite the increased complexity of the true fitness landscape, the inferred parameters show high accuracy, as measured by Pearson's  $R$  values, for selection, epistasis, and overall fitness. The dependency of accuracy on mutation rate remains qualitatively similar to that observed under the pairwise fitness setting.



**FIG. S2. Dynamics of cumulants and traits using exQLE and Gaussian closure scheme.** This corresponds to **Fig. 1** in the main text but utilizes the exQLE framework with a Gaussian closure (GC) scheme, which suppresses all cumulants beyond second order. The GC scheme is efficient for inferring fitness parameters and is not limited to inference problems; it is also applicable to forward processes. Although the accuracy of the dynamics for cumulants (b) is not as high as that in the exQLE case without GC, the estimated values show a reasonable correlation with those from the exact calculations. Pearson's  $R$  values for additive selection (a), pairwise epistasis (b), and random traits (c) for exQLE with GC are 0.93, 0.43, and 0.65. For comparison, the exQLE values without GC are 0.97, 0.89, and 0.95, respectively.

## REFERENCES

---

- [1] S. Wright, *Genetics* **16**, 97 (1931).
- [2] R. A. Fisher, *The genetical theory of natural selection: a complete variorum edition* (Oxford University Press, 1999).
- [3] B. Walsh and M. Lynch, *Evolution and selection of quantitative traits* (Oxford University Press, 2018).
- [4] J. F. Crow, *An introduction to population genetics theory* (Scientific Publishers, 2017).
- [5] R. C. Lewontin and K.-i. Kojima, *Evolution* , 458 (1960).
- [6] N. H. Barton and M. Turelli, *Annual review of genetics* **23**, 337 (1989).
- [7] M. Kimura, *Genetics* **52**, 875 (1965).
- [8] N. H. Barton, *Genetics Research* **65**, 123 (1995).
- [9] R. A. Neher and B. I. Shraiman, *Proceedings of the National Academy of Sciences* **106**, 6866 (2009).
- [10] R. A. Neher and B. I. Shraiman, *Reviews of Modern Physics* **83**, 1283 (2011).
- [11] T. Nagylaki, *Proceedings of the National Academy of Sciences* **71**, 526 (1974).
- [12] N. H. Barton and S. P. Otto, *Genetics* **169**, 2353 (2005).
- [13] N. H. Barton and M. Turelli, *Genetics* **127**, 229 (1991).
- [14] M. Turelli and N. H. Barton, *Genetics* **138**, 913 (1994).
- [15] M. Kirkpatrick, T. Johnson, and N. Barton, *Genetics* **161**, 1727 (2002).
- [16] T. Nagylaki, *Genetics* **134**, 627 (1993).
- [17] T. Nagylaki, J. Hofbauer, and P. Brunovský, *Journal of mathematical biology* **38**, 103 (1999).
- [18] H.-L. Zeng and E. Aurell, *Physical Review E* **101**, 052409 (2020).
- [19] H.-L. Zeng, V. Dichio, E. Rodriguez Horta, K. Thorell, and E. Aurell, *Proceedings of the National Academy of Sciences* **117**, 31519 (2020).
- [20] J. M. Smith and J. Haigh, *Genetics Research* **23**, 23 (1974).
- [21] N. H. Barton, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **355**, 1553 (2000).
- [22] W. G. Hill and A. Robertson, *Genetics Research* **8**, 269 (1966).
- [23] J. Felsenstein, *Genetics* **78**, 737 (1974).
- [24] K. Husain and A. Murugan, *Molecular Biology and Evolution* **37**, 2865 (2020).
- [25] S. Kryazhimskiy, *Elife* **10**, e60200 (2021).
- [26] F. J. Poelwijk, V. Krishna, and R. Ranganathan, *PLoS computational biology* **12**, e1004771 (2016).
- [27] J. Otwinowski, D. M. McCandlish, and J. B. Plotkin, *Proceedings of the National Academy of Sciences* **115**, E7550 (2018).
- [28] J. Zhou, M. S. Wong, W.-C. Chen, A. R. Krainer, J. B. Kinney, and D. M. McCandlish, *Proceedings of the National Academy of Sciences* **119**, e2204233119 (2022).
- [29] K. Rijal, C. M. Holmes, S. Petti, G. Reddy, M. M. Desai, and P. Mehta, *arXiv preprint arXiv:2504.10388* (2025).
- [30] D. M. Lyons, Z. Zou, H. Xu, and J. Zhang, *Nature Ecology & Evolution* **4**, 1685 (2020).
- [31] G. Reddy and M. M. Desai, *Elife* **10**, e64740 (2021).
- [32] M. S. Johnson, G. Reddy, and M. M. Desai, *BMC biology* **21**, 120 (2023).
- [33] J. Diaz-Colunga, A. Skwara, K. Gowda, R. Diaz-Uriarte, M. Tikhonov, D. Bajic, and A. Sanchez, *Philosophical Transactions of the Royal Society B* **378**, 20220053 (2023).
- [34] H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx, *Science* **332**, 1190 (2011).
- [35] A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, *Science* **332**, 1193 (2011).
- [36] S. Kryazhimskiy, D. P. Rice, E. R. Jerison, and M. M. Desai, *Science* **344**, 1519 (2014).
- [37] C. A. Olson, N. C. Wu, and R. Sun, *Current biology* **24**, 2643 (2014).
- [38] T. N. Starr, J. M. Flynn, P. Mishra, D. N. Bolon, and J. W. Thornton, *Proceedings of the National Academy of Sciences* **115**, 4453 (2018).
- [39] Y. Park, B. P. Metzger, and J. W. Thornton, *Science* **376**, 823 (2022).
- [40] C. W. Bakerlee, A. N. Nguyen Ba, Y. Shulgina, J. I. Rojas Echenique, and M. M. Desai, *Science* **376**, 630 (2022).
- [41] Z. R. Sailer and M. J. Harms, *PLoS computational biology* **13**, e1005541 (2017).
- [42] F. J. Poelwijk, M. Socolich, and R. Ranganathan, *Nature communications* **10**, 4213 (2019).
- [43] A. M. Phillips, K. R. Lawrence, A. Moulana, T. Dupic, J. Chang, M. S. Johnson, I. Cvijovic, T. Mora, A. M. Walczak, and M. M. Desai, *Elife* **10**, e71393 (2021).
- [44] A. Moulana, T. Dupic, A. M. Phillips, J. Chang, A. A. Roffler, A. J. Greaney, T. N. Starr, J. D. Bloom, and M. M. Desai, *Elife* **12**, e83442 (2023).
- [45] A. M. Phillips, D. P. Maurer, C. Brooks, T. Dupic, A. G. Schmidt, and M. M. Desai, *Elife* **12**, e83628 (2023).
- [46] G. Yang, D. W. Anderson, F. Baier, E. Dohmen, N. Hong, P. D. Carr, S. C. L. Kamerlin, C. J. Jackson, E. Bornberg-Bauer, and N. Tokuriki, *Nature Chemical Biology* **15**, 1120 (2019).
- [47] K. Buda, C. M. Miton, and N. Tokuriki, *Nature Communications* **14**, 8508 (2023).
- [48] S. Schulz, T. J. Tan, N. C. Wu, and S. Wang, *Proceedings of the National Academy of Sciences* **122**, e2413884122 (2025).
- [49] M. S. Sohail, R. H. Louie, M. R. McKay, and J. P. Barton, *Nature biotechnology* **39**, 472 (2021).
- [50] M. S. Sohail, R. H. Louie, Z. Hong, J. P. Barton, and M. R. McKay, *Molecular biology and evolution* **39**, msac199 (2022).

- [51] E. Mauri, S. Cocco, and R. Monasson, *Europhysics Letters* **132**, 56001 (2021).
- [52] J. D. Murray, *Mathematical biology: I. An introduction*, Vol. 17 (Springer Science & Business Media, 2007).
- [53] W. J. Ewens, *Mathematical population genetics: theoretical introduction*, Vol. 1 (Springer, 2004).
- [54] G. R. Price, *Annals of human genetics* **36**, 129 (1972).
- [55] A. Robertson, *Animal Science* **8**, 95 (1966).
- [56] L. Durlofsky, J. F. Brady, and G. Bossis, *Journal of fluid mechanics* **180**, 21 (1987).
- [57] V. Dichio, H.-L. Zeng, and E. Aurell, *Reports on Progress in Physics* **86**, 052601 (2023).
- [58] K. S. Shimagaki and J. P. Barton, *Genetics* , iyaf118 (2025).
- [59] H.-L. Zeng, E. Mauri, V. Dichio, S. Cocco, R. Monasson, and E. Aurell, *Journal of Statistical Mechanics: Theory and Experiment* **2021**, 083501 (2021).
- [60] R. Lande and S. J. Arnold, *Evolution* , 1210 (1983).
- [61] A. Nourmohammad, T. Held, and M. Lässig, *Current opinion in genetics & development* **23**, 684 (2013).
- [62] A. Nourmohammad, J. Rambeau, T. Held, V. Kovacova, J. Berg, and M. Lässig, *Cell reports* **20**, 1385 (2017).
- [63] S. Wright, *Proceedings of the National Academy of Sciences* **23**, 307 (1937).
- [64] J. F. Crow and M. Kimura, *The American Naturalist* **99**, 439 (1965).
- [65] M. Kimura, *Journal of Applied Probability* **1**, 177 (1964).
- [66] T. Ohta and M. Kimura, *Genetics* **63**, 229 (1969).
- [67] M. Lynch, B. Walsh, *et al.*, *Genetics and analysis of quantitative traits*, Vol. 1 (Sinauer Sunderland, MA, 1998).
- [68] B. H. Good, A. M. Walczak, R. A. Neher, and M. M. Desai, *PLoS genetics* **10**, e1004222 (2014).
- [69] B. H. Good, *Genetics* **220**, iyac004 (2022).
- [70] B. Lee, A. A. Quadeer, M. S. Sohail, E. Finney, S. F. Ahmed, M. R. McKay, and J. P. Barton, *Nature Communications* **16**, 441 (2025).
- [71] F. Zanini, J. Brodin, L. Thebo, C. Lanz, G. Bratt, J. Albert, and R. A. Neher, *eLife* **4**, e11282 (2015).
- [72] E. V. Romero and A. F. Feder, *Molecular Biology and Evolution* **41**, msad260 (2024).
- [73] K. S. Shimagaki, R. M. Lynch, and J. P. Barton, *eLife* **14** (2025).
- [74] K. Shimagaki and J. P. Barton, *Physical Review E* **107**, 024116 (2023).
- [75] T. N. Starr, A. J. Greaney, S. K. Hilton, D. Ellis, K. H. Crawford, A. S. Dingens, M. J. Navarro, J. E. Bowen, M. A. Tortorici, A. C. Walls, *et al.*, *cell* **182**, 1295 (2020).
- [76] A. Moulana, T. Dupic, A. M. Phillips, J. Chang, S. Nieves, A. A. Roffler, A. J. Greaney, T. N. Starr, J. D. Bloom, and M. M. Desai, *Nature communications* **13**, 7011 (2022).
- [77] Y. Gao and J. P. Barton, *Proceedings of the National Academy of Sciences* **122**, e2405379122 (2025).
- [78] H. Risken, in *The Fokker-Planck equation: methods of solution and applications* (Springer, 1989) pp. 63–95.