OXFORD GENETICS

A theory of oligogenic adaptation of a quantitative trait

Ilse Höllinger,[1,‡] Benjamin Wölfl,[1,2,3,‡] Joachim Hermisson[1,4,*]

[1]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
[2]Vienna Graduate School of Population Genetics, University of Vienna and Veterinary Medical University of Vienna, Vienna, Austria
[3]Vienna Doctoral School of Ecology and Evolution, University of Vienna, Vienna, Austria
[4]Max Perutz Labs, Vienna Biocenter Campus (VBC), Dr.-Bohr-Gasse 9, 1030 Vienna, Austria

*Corresponding author: Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria. Email: joachim.hermisson@univie.ac.at
‡These authors contributed equally to this work.

Abstract

Rapid phenotypic adaptation is widespread in nature, but the underlying genetic dynamics remain controversial. Whereas population genetics envisages sequential beneficial substitutions, quantitative genetics assumes a collective response through subtle shifts in allele frequencies. This dichotomy of a monogenic and a highly polygenic view of adaptation raises the question of a middle ground, as well as the factors controlling the transition. Here, we consider an additive quantitative trait with equal locus effects under Gaussian stabilizing selection that adapts to a new trait optimum after an environmental change. We present an analytical framework based on Yule branching processes to describe how phenotypic adaptation is achieved by collective changes in allele frequencies at the underlying loci. In particular, we derive an approximation for the joint allele-frequency distribution conditioned on the trait mean as a comprehensive descriptor of the adaptive architecture. Depending on the model parameters, this architecture reproduces the well-known patterns of sequential, monogenic sweeps, or of subtle, polygenic frequency shifts. Between these endpoints, we observe oligogenic architecture types that exhibit characteristic patterns of partial sweeps. We find that a single compound parameter, the population-scaled background mutation rate Θ_{bg}, is the most important predictor of the type of adaptation, while selection strength, the number of loci in the genetic basis, and linkage only play a minor role.

Keywords: oligogenic adaptation; polygenic adaptation; quantitative trait; population genetics; quantitative genetics

Introduction

Quantitative traits (QTs) are everywhere in nature, from body size and milk yield to melanism and life expectancy. This has made the study of their evolution a subject of interest for more than a century (Walsh and Lynch 2018; Sella and Barton 2019). Ever since the modern synthesis, it has been clear that even complex QTs are governed by a finite genetic basis (Fisher 1918). Nevertheless, the genomic changes underlying phenotypic adaptation have long been elusive. Instead, much effort has gone into theoretical exploration. Two schools of thought—population genetics and quantitative genetics—developed independent narratives for the genetics of adaptation, with notable differences.

On the one hand, population genetics is concerned with the dynamics of allele frequencies. Consequently, adaptation is understood as major changes in allele frequencies driven by selection. The archetypal scenario is a (hard) selective sweep, where an initially rare beneficial allele at a single locus quickly rises to fixation (Maynard-Smith and Haigh 1974; Kaplan et al. 1989; Barton 1998). More complex models examine sweep patterns in diversity and divergence data across multiple loci in the basis of a trait or along a biological pathway (e.g., Gouy et al. 2017), discuss linkage and epistasis, and the effect of recurrent sweeps on genome-wide diversity levels observable through genome scans (reviewed in Stephan 2019). At the level of phenotype or fitness, this results in a view

of adaptation via successive selective sweeps as discrete steps of a so-called adaptive walk (reviewed in Orr 2005).

On the other hand, quantitative genetics focuses on phenotype data. It thrives on being able to abstract from the underlying genetics, which only enter as summary statistics across loci, such as the genetic variance and higher moments of the trait distribution (Turelli and Barton 1990; Bürger 2000). Accordingly, the dynamics at single loci do not play a role. This view is taken to the extreme in the infinitesimal model, where phenotypic adaptation in completely decoupled from the changes in underlying allele frequencies (Barton et al. 2017; Fisher 1918). At the genotypic level, adaptation is viewed as tiny allele frequency shifts at a myriad of loci with small individual effects, but these dynamics are not explicitly described.

After decades of research with little exchange between communities, the emergence of data from genome-wide association studies (GWAS) led to a growing interest in an integrated view linking phenotypic adaptation with its genetic underpinnings. Pritchard and colleagues (Pritchard et al. 2010; Pritchard and Di Rienzo 2010) proposed the concept of "polygenic adaptation" as an alternative to the selective sweep model with the goal of characterizing—and, if possible, detecting—patterns of QT adaptation at the genomic level.

If adaptation can be either highly polygenic or involve only single loci, this raises the question of a middle ground, that is, whether there is room for an "oligogenic view of adaptation" (Bell 2009), with its own characteristic patterns. To address this question, a framework is needed that connects the opposite endpoints and also covers the parameter range in between. Several recent concept papers discuss such frameworks (Barghi *et al.* 2020; Láruson *et al.* 2020; Fagny and Austerlitz 2021).

It turns out that the decisive parameter for determining the type of adaptation is not simply the number of loci that underlies a trait. What matters is a measure of *redundancy* (Barghi *et al.* 2020; Láruson *et al.* 2020), which determines how many alternative genotypes that solve the adaptive task can be readily generated in a population (and therefore coexist), either from standing genetic variation (SGV) or new mutation. To make these notions precise, a model is needed.

Here, we consider a classical additive trait under Gaussian stabilizing selection that adapts to a new trait optimum after an environmental change. This is a classic scenario that has been studied many times before (e.g., Lande 1976; Jain and Stephan 2017b; Stetter *et al.* 2018; Thornton 2019; Hayward and Sella 2022). We ask how adaptive progress of the trait mean toward the new optimum breaks down into contributions of alleles at the underlying loci. Specifically, we develop an analytical theory to approximate the joint distribution of these allelic contributions across replicates.

Our method builds on a two-step approach first used in Höllinger *et al.* (2019) to describe adaptation of a binary polygenic trait (such as pesticide resistance). In the first step, a multi-type branching process is used to infer the joint frequency distribution of co-segregating alleles, as long as these frequencies are small and strongly affected by genetic drift, but hardly interact (both before and after the environmental change). The second step describes how this distribution is transformed when the frequencies grow larger. While genetic drift can be ignored in this phase, epistatic interactions begin to play a role. This transformation is straightforward for a binary trait, but it can be extended to a much larger class of models, including the additive QT, in particular.

In this manuscript, we describe this method for the case of a trait with arbitrarily many unlinked additive loci of equal effect. We derive the joint frequency distribution of all alleles that contribute to phenotypic adaptation at arbitrary, fixed values of the mean trait. The results show how the genotypic patterns change with an increasing phenotypic distance and provide a comprehensive classification of adaptive architectures, from single, consecutive sweeps to small polygenic shifts. In particular, we describe and discuss the characteristics of oligogenic architectures that lie between the better known monogenic or highly polygenic endpoints.

The remainder of the article is organized as follows: In the "Model and methods" section, we first introduce the model and the assumptions of the simulation methods (individual-based (IB) and Wright–Fisher). Then, both steps of the analytical approximation are explained, while all detailed derivations can be found in the separate Supplementary File S1. In the "Results" section, we compare the analytical predictions with simulation results and describe the most important patterns for both short- and long-distance adaptation. We also assess the effects of linkage between selected loci. The "Discussion" section examines key notions and highlights the scope and limits of the model. The Appendix shows complementary figures. Scripts and data are deposited on *Dryad* (DOI: https://doi.org/10.5061/dryad.573n5tbc9, Höllinger *et al.* 2023).

Model and methods
The model

We model a panmictic population of size N_e and follow the adaptation of an additive QT, Z, that is governed by L loci. Each locus $i \in \{1, 2, \ldots, L\}$ is biallelic with alleles a_i and A_i and equal effect sizes set to 0 (for a_i) and $\gamma > 0$ (for A_i). We focus on haploid genetics in the main text for simplicity, and present extensions to diploids in the Appendix (Section "Adaptation with linkage"). With $\eta_i \in \{0, 1\}$ indicating the allelic state, we thus have

$$Z = \sum_{1 \leq i \leq L} \eta_i \gamma. \tag{1}$$

Wrightian fitness is modeled by time-dependent Gaussian stabilizing selection towards a trait optimum $Z_{opt}(t)$,

$$W(Z) = \exp\left[-\frac{\sigma(t)}{2}(Z - Z_{opt}(t))^2\right], \tag{2}$$

where $\sigma(t) > 0$ (the inverse width of the fitness function) measures the selection strength. At time $t = 0$, a sudden environmental change occurs and the trait optimum jumps from the ancestral optimum, $Z_{opt}(t < 0) = Z_{opt}^0$, to a new optimum, $Z_{opt}(t \geq 0) = Z_{opt}^{new} > Z_{opt}^0$. As we will see, the selection strength $\sigma(t)$, which may be constant or variable, does not affect our results.

Denote the frequency of the A_i allele in the population as p_i. New mutations from a_i to A_i arise at rate μ_i per generation and back-mutations at rate ν_i. Loci may be linked, assuming a single linear chromosome with recombination rate r between neighboring loci. Prior to the environmental change, population variation segregates at mutation–selection–drift balance. If locus mutation rates are sufficiently small (and/or selection is sufficiently strong), the trait mean is $\bar{Z} = \sum_i p_i \gamma \approx Z_{opt}^0$. Moreover, allele frequencies, p_i, in the SGV follow a U-shaped distribution, such that, for an ancestral trait optimum of $Z_{opt}^0 = (L - d)\gamma$, there are $d \leq L$ loci almost fixed for the a_i allele ($p_i \approx 0$, "beneficial variation"), while $L - d$ loci are almost fixed for the A_i allele ($p_i \approx 1$, "deleterious variation") at $t = 0$. The case of no SGV is formally included in this framework as the limit of very strong purifying selection (very large $\sigma(t)$ for $t < 0$).

After the environmental change, the trait mean $\bar{Z}$ starts to move toward the new optimum. During this adaptation process, we take "snapshots" of the population, that is, we record all allele frequencies when the trait mean reaches $\bar{Z} = Z_{opt}^0 + c_Z \gamma \leq Z_{opt}^{new}$ (see Fig. 1). Here, c_Z measures adaptation at the level of the phenotype in units of mutational steps. The *joint distribution of allele frequencies* across evolutionary replicates at such phenotypic passage points serves as our measure of the adaptive architecture (following Höllinger *et al.* 2019; Barghi *et al.* 2020). An overview over the used nomenclature is provided in Table A1.

Individual-based simulations

We resort to IB simulations to compute the full adaptation dynamics in discrete time for a population of N_e haploids. Each generation, N_e mating pairs are chosen via stochastic acceptance (Lipowski and Lipowska 2012) according to their fitness Equation (2). For each pair, a binomially distributed number of crossover points are randomly placed on a single chromosome to construct the genotype of a single recombinant offspring individual: the probability of a recombination event between neighboring sites is r. This is followed by bidirectional mutation ($a_i \leftrightarrow A_i$), where the number of mutated sites in the entire population is Poisson
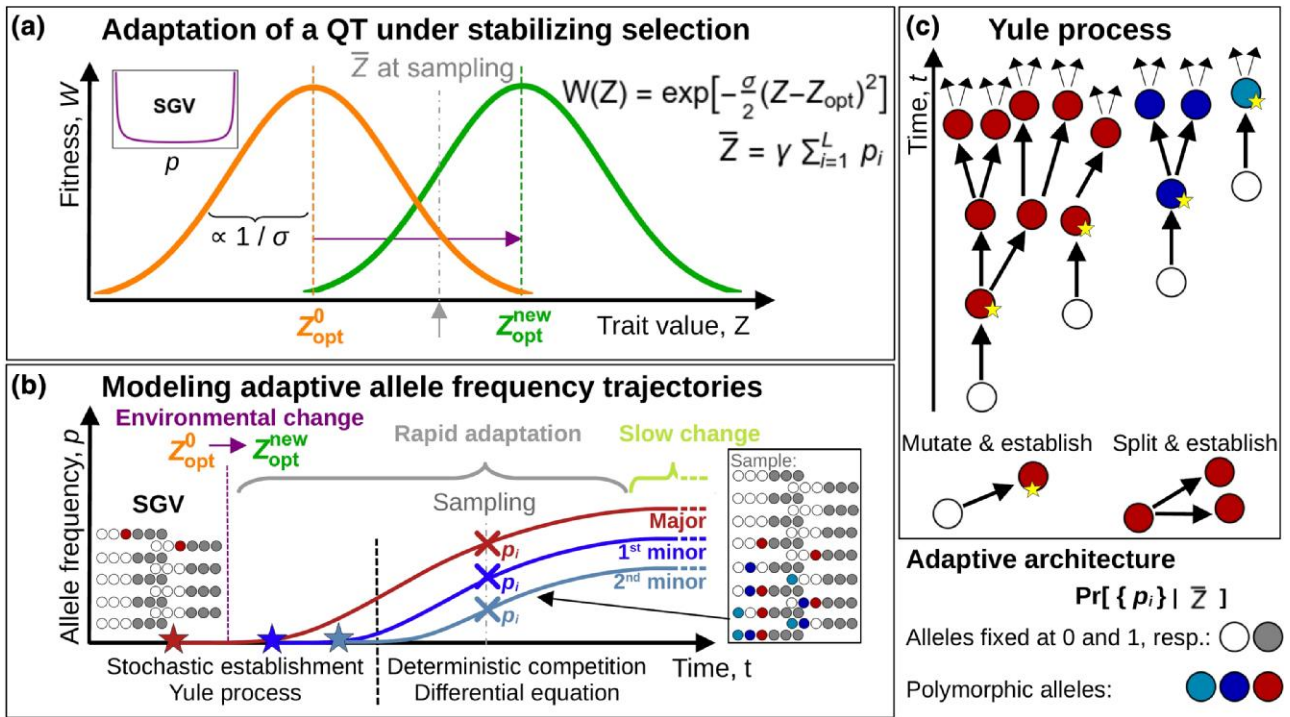
Fig. 1. Modeling approach. a) After an environmental change, the optimum of a trait Z under Gaussian stabilizing selection, $W(Z)$, shifts from Z_{opt}^0 to Z_{opt}^{new} (orange/left and green/right bell curves). The population adapts from mutation–selection–drift balance (SGV). Allele frequencies at all L loci underlying the trait are recorded once the trait mean, $\bar{Z}$, reaches a predefined value (gray arrow). b) We dissect the adaptive process into separate phases: During the early establishment phase, both prior and after the environmental change, mutant allele frequencies are typically small. Mutants evolve largely independently, but are strongly affected by the stochastic forces of new mutation and genetic drift. Once beneficial mutants have grown to higher frequency, stochasticity can be ignored, but competition and epistatic interaction become relevant. As long as the mean phenotype is not too close to the new optimum, $\bar{Z} \ll Z_{opt}^{new}$, directional selection prevails and leads to rapid adaptation. Close to the optimum, the dynamics slow down, and evolution under weak disruptive selection leads to a depletion of variation. (C) We use a multi-type Yule branching process to track the allele counts at different loci during the establishment phase. The process only follows "immortal" mutant lineages that escape loss due to drift. New immortal lines originate either via new mutation at all loci, or by birth events (splits) of existing lines.

distributed with parameter $N_e L \mu$. This completes the simulated life cycle.

All individuals are initialized with d randomly chosen loci carrying the allele a_i and $L - d$ loci with A_i in order to match the ancestral optimum, $\bar{Z} = Z_{opt}^0$. Then, the population equilibrates for $20N_e$ generations under constant stabilizing selection to reach mutation–selection–drift balance prior to the environmental change. At $t = 0$, the optimum jumps to Z_{opt}^{new}, and the population starts to adapt. Allele frequencies at all loci and the full phenotype distribution are recorded when $\bar{Z}$ first reaches a prescribed value, $\bar{Z} = Z_{opt}^0 + c_Z \gamma$. We usually consider a population of size $N_e = 1,000$ and summarize results for 10,000 replicates. Further details can be found in the *Computational Supplement* (uses *Mathematica* by Wolfram Research 2019) and the provided IB simulation script (in C++, Stroustrup 2013).

Linkage equilibrium simulations

For weak selection and/or strong recombination, we can ignore all linkage disequilibria (LD) and reduce the multi-locus dynamics to the dynamics of single-locus frequencies, p_i. Analytic expressions for the single-locus dynamics can be derived for a large class of models (see Supplementary File S1). For Gaussian selection, in particular, the single-locus equations are well known (Wright 1935a, 1935b; Barton 1986; de Vladar and Barton 2014; Jain and Stephan 2015, 2017b),

$$\dot{p}_i = p_i(1 - p_i)\sigma(t)\gamma_i$$
$$\cdot \left((Z_{opt}(t) - \bar{Z}) - \frac{\gamma_i}{2}(1 - 2p_i) \right) + \mu_i(1 - p_i) - \nu_i p_i. \qquad (3)$$

The first selection term, $Z_{opt}(t) - \bar{Z}$, describes directional selection toward $Z_{opt}(t)$. It dominates as long as the trait mean is sufficiently far from the optimum, which is typically the case in the early phase of rapid adaptation. The second term, $(\gamma_i/2)(1 - 2p_i)$ corresponds to disruptive selection and dominates the dynamics once the trait mean gets very close to the new optimum. In this later phase of adaptation, selection for reduced genetic variation drives alleles to the boundaries of the allele frequency range—either to loss or fixation.

We can assess these dynamics using efficient Wright–Fisher simulations that track loci separately in discrete time. Allele frequencies at a locus in the offspring generation are generated by forward and backward mutation with equal rates, $\Theta_i/2 = N_e\mu$, followed by binomial sampling to implement selection and drift. Let the sampling weight of the a_i allele be normalized to 1, then the weight of the A_i allele is given by its Wrightian (multiplicative) fitness, corresponding to the selection term in Equation (3),

$$\exp\left[\sigma(t)\gamma_i \left((Z_{opt}(t) - \bar{Z}) - \frac{\gamma_i}{2}(1 - 2p_i) \right) \right]. \qquad (4)$$

The dynamics at individual loci influence each other via the mean trait, $\bar{Z}$, which is recorded in every generation. Allele frequencies at all loci are reported whenever $\bar{Z}$ reaches a predefined value.

We usually simulate a population of $N_e = 10,000$ haploid individuals, with $3 \leq L \leq 10,000$, loci underlying the trait, and evaluate 10,000 replicates per parameter combination. Prior to the environmental change, we let the population equilibrate for at least $8N_e$ generations to build up SGV. Further details can be found in the *Computational Supplement* and the included LE simulation script (in C++).

Analytical approximations

To obtain an analytical approximation for the joint allele frequency distribution during rapid adaptation, we extend a framework developed for a binary trait in Höllinger et al. (2019) to the case of an additive QT. The approach separates the dynamical process into two phases: an initial stochastic *establishment phase* and a subsequent *deterministic phase* during which established mutants grow and compete (see Fig. 1b). Similar approaches have been used by Uecker and Hermisson (2011) and Martin and Lambert (2015) to describe the dynamics of a single locus and by Götsch and Bürger (2023) for independent loci.

We start with a population in mutation–selection–drift balance and the trait mean, $\bar{Z}$, close to its ancestral optimum, Z_{opt}^0. During the first, stochastic phase, mutant frequencies both before and directly after the environmental change are small, such that epistatic interactions can be neglected. Assuming linkage equilibrium (LE), the dynamics at individual loci (due to drift, mutation, and selection) are approximately independent and can be described by a branching process. Once the frequencies of beneficial alleles become sufficiently large, the phenotype adapts rapidly towards the new optimum, Z_{opt}^{new}. Since drift and new mutation can be ignored during this phase, the allele frequency dynamics are effectively deterministic. They are driven by directional selection, but are no longer independent, because alleles at different loci are coupled due to (fitness) epistasis. Mutants also compete for their relative contribution to adaptive change in the mean trait. The end of this phase is reached, when the rapid change of the mean population phenotype slows down in the vicinity of the new optimum and the dynamics of disruptive selection and drift take over.

Yule branching process

During the establishment phase, we describe the allele dynamics by a multi-type *Yule process* (see Fig. 1c). This process tracks the origin and spread of mutations that successfully establish within the population, that is, that escape stochastic loss and leave descendants in the population until observation. It thus corresponds to a coalescent genealogy, but is constructed forward in time, at multiple loci simultaneously. Details of the construction are given in the Supplementary File S1.

We denote the establishment probability of a new mutant copy (the probability that descendants still exist at the time of sampling) as $p_{est}(t)$. Such copies found new "immortal lineages" of the Yule tree. Since selection is time-dependent, also $p_{est}(t)$ depends on time. In particular, establishment of a later-beneficial A_i mutant is much less likely if it originates before the environmental change ($t < 0$), while the allele is still deleterious. Analytical approximations for $p_{est}(t)$ lead to complex expressions (Uecker and Hermisson 2011), but this is not needed here. Our

results simply exploit the fact that for small p_i the establishment probability of mutants with the same phenotypic effect is (approximately) identical for all loci.

There are two events that create new immortal lines of the Yule process: New successful mutation (before and after the environmental change) seeds novel Yule trees at rate

$$p_{mut,i}(t) \approx N_e \mu_i \cdot p_{est}(t) = \Theta_i \frac{p_{est}(t)}{2}$$

for locus i. Birth events lead to splits of existing immortal lines—branching of an existing Yule tree—at rate

$$p_{split}(t) \approx \frac{p_{est}(t)}{2}.$$

We start the Yule process at some time $t \leq 0$ before the first successful mutation has originated and stop it when a certain number of immortal lineages has been generated. At this point, we assess the distribution of immortal lineages across all loci. Since the rates of all events at loci are proportional to the same establishment probability, $p_{est}(t)$, we can drop this common factor if we are only interested in the sequence of events and not in their timing. (This is where our assumption of equal locus effects becomes important.) Mathematically, this corresponds to a time rescaling. On the new timescale, the process is time-homogeneous, with constant rates $p_{mut,i} = \Theta_i$ and $p_{split} = 1$. For this simple process, we find that the distribution of *ratios* of allele frequencies of beneficial A_i mutants at the end of the stochastic phase is given by an inverted Dirichlet distribution (see Höllinger et al. 2019, and the Supplementary File S1).

Deterministic phase

Once the allele frequencies are no longer small, and as long as the trait mean is not yet very close to the new optimum, the dynamics of beneficial A_i mutants are well described by the deterministic *directional selection model* (Jain and Stephan 2017b),

$$\dot{p}_i = \sigma(t)\gamma p_i(1 - p_i)(Z_{opt}(t) - \bar{Z}). \tag{5}$$

In terms of odds, $u_i := p_i/(1 - p_i)$, this reads

$$\dot{u}_i = \sigma(t)\gamma u_i(Z_{opt}(t) - \bar{Z})$$

such that odds ratios, u_i/u_j, remain constant under the dynamics,

$$\frac{\partial}{\partial t} \frac{u_i}{u_j} = \frac{\dot{u}_i u_j - u_i \dot{u}_j}{u_j^2} = 0. \tag{6}$$

For small allele frequencies, $u_i/u_j \approx p_i/p_j$. The distribution of odds ratios thus approximately follows the inverted Dirichlet distribution derived above from the Yule process. Due to Equation (6), the u_i maintain this distribution throughout the deterministic phase. To obtain a result for the joint allele frequency distribution, we need to transform the distribution of the odds u_i back to the frequencies p_i at the stopping condition for the mean trait $\bar{Z}$.

In a wide parameter region, when effects of rare a_i mutants (that change the trait mean in the opposite direction of phenotypic adaptation) can be ignored, the condition on frequencies p_i and

corresponding odds u_i of A_i mutants at the loci 1 to d (beneficial variation) at the sampling point reads

$$\frac{\bar{Z} - Z^0_{opt}}{\gamma} = \sum_{i=1}^{d} p_i = \sum_{i=1}^{d} \frac{u_i}{u_i + 1} =: c_Z, \qquad (7)$$

with $c_Z \in [0, d]$. We then obtain the joint distribution of the p_i with $p_1 = c_Z - \sum_{i=2}^{d} p_i$ and

$$P[\{p_i\}|c_Z]$$

$$= \frac{1}{B(\Theta)} \prod_{i=2}^{d} \frac{p_i^{\Theta_i - 1}}{(1 - p_i)^{\Theta_i + 1}} \left(1 + \frac{\sum_{i=2}^{d} p_i(1 - p_i)}{(c_Z - \sum_{k=2}^{d} p_k)(1 + \sum_{k=2}^{d} p_k - c_Z)}\right)$$

$$\cdot \left(1 + \frac{1 + \sum_{k=2}^{d} p_k - c_Z}{c_Z - \sum_{k=2}^{d} p_k} \sum_{i=2}^{d} \frac{p_i}{1 - p_i}\right)^{-\Theta_1} \qquad (8)$$

$$\cdot \left(\frac{c_Z - \sum_{k=2}^{d} p_k}{1 + \sum_{k=2}^{d} p_k - c_Z} + \sum_{i=2}^{d} \frac{p_i}{1 - p_i}\right)^{-\sum_{i=2}^{d} \Theta_i},$$

for $i = 2, \ldots, d$, where

$$B(\Theta) := \frac{\prod_{i=1}^{d} \Gamma(\Theta_i)}{\Gamma\left(\sum_{i=1}^{d} \Theta_i\right)}$$

is the multivariate *beta*-function and Γ is the *gamma*-function (see Supplementary File S1). Equation (8) is the analog of the joint distribution function for adaptation of a binary trait with only two phenotypes (Höllinger *et al.* 2019, Eq. 8). Note that the expression does not depend on any selection parameters (before or after the environmental change), but only on the mutation rates, Θ_i. For two loci, in particular, the marginal distribution at the second locus can be derived explicitly ($p_2 = p$),

$$P[p \,|\, c_Z] = \frac{\Gamma(\Theta_1 + \Theta_2)}{\Gamma(\Theta_1)\Gamma(\Theta_2)} \frac{p(1 - p) + (c_Z - p)(1 + p - c_Z)}{(c_Z - 2p(c_Z - p))^{\Theta_1 + \Theta_2}}$$

$$\cdot ((c_Z - p)(1 - p))^{\Theta_1 - 1}((1 + p - c_Z)p)^{\Theta_2 - 1}. \qquad (9)$$

For three or more loci, marginal distributions can only be obtained by $(d - 1)$-fold integration. The special functional form of the joint distribution (cf., Ghorbel 2009) and its connection to the gamma distribution allow for efficient numerical techniques and precise results even for highly polygenic traits and d of the order of 100 or 1,000. This is explained in detail in the *Computational Supplement*.

For high levels of SGV (large Θ_i, many loci) and/or small adaptation distances (small $\bar{Z} - Z^0_{opt}$ at sampling), the contribution of segregating minority a_i mutants contributing to deleterious variation cannot be neglected. It is still possible to derive (and numerically evaluate) an expression for the joint distribution of allele frequencies at all L loci underlying the trait for this case. Since the terms become increasingly complex, the results (and all derivations) are relegated to the Supplementary File S1. The associated, complete numerical solution procedure can be found in the *Computational Supplement*.

Results

Below, we compare our analytical results with comprehensive computer simulations and discuss characteristic features of the adaptive architecture of a QT across broad parameter ranges in terms of selection strength, mutation rates, number of loci in the genetic basis, linkage, presence/absence of SGV, and distance of the adaptive phenotype from its ancestral value.

Loci have equal strength in our model, and for simplicity we also assume equal locus mutation rates in the results part. This leaves two main sources of differences between loci: asymmetric initial conditions (some loci are nearly fixed for A_i alleles and others for a_i alleles before the environmental change) and the stochastic effects of mutation and genetic drift. We begin by explaining how we define and represent "adaptive architecture."

Adaptive architecture

Following Barghi *et al.* (2020), the adaptive architecture of a trait informs about the number and relative contributions of loci that collectively cause a given level of phenotypic adaptation. By definition, the joint allele frequency distribution at all loci in the genetic basis of the trait, taken across replicates at arbitrary points of phenotypic adaptation, provides an exhaustive description. However, for a trait with a genetic basis of more than a few loci, this is a complex high-dimensional quantity that needs to be projected onto one-dimensional marginal distributions for visualization.

Rather than simply marginalizing for a fixed locus, we follow Höllinger *et al.* (2019) and derive marginal distributions for groups of loci ordered by allele frequency. Therefore, we order all loci according to the frequency of the A_i allele in each evolutionary replicate. Since all locus effects are equal, this also corresponds to their relative contribution to the phenotype. We then construct univariate frequency distributions for loci with the same frequency rank across replicates (which may be different physical loci between replicates). For L loci in the genetic basis of the trait, this yields a set of L size-ordered marginal distributions to visualize the original L-dimensional joint distribution.

In designating frequency-ordered loci, we focus on the loci with A_i as minority allele (the "beneficial variation"), since these contribute the largest allele frequency changes. We therefore often discard the $(L - d)$ loci with the largest frequency of the A_i allele (assume them to be fixed) and call the locus with the $(L - d + 1)$-largest frequency the "major locus" of the adaptive process. Subsequent loci are called the "first," "second," "third," etc., minor loci, accordingly. This is consistent with Höllinger *et al.* (2019) and also with standard nomenclature. Indeed, although all locus effects are equal in our model, the major locus would typically yield the strongest signal in an association study of the adaptive trait (if there is any signal at all) because it has the largest frequency change.

Sampling in "pheno-time"

The adaptive architecture relates adaptive change at the level of the phenotype to changes in the underlying genotype. Accordingly, we record allele frequencies during the adaptive process not after a fixed time, but when a stopping condition $\bar{Z} = Z^0_{opt} + c_Z\gamma$ is reached for the mean phenotype (fixed *pheno-time*). For each sampling point, we calculate analytical predictions for the joint distribution of allele frequencies and compare them with numerical simulations.

In this manuscript, we focus entirely on the architecture of the early, rapid phase of adaptation, while the trait is still predominantly subject to directional selection. We thus consider sampling points before the trait mean reaches the new optimum, $\bar{Z} < Z^{new}_{opt}$. The architecture observed at these points is transient and is eventually replaced by a pattern of adaptive substitutions as stabilizing selection at the new optimum drives alleles to either loss or
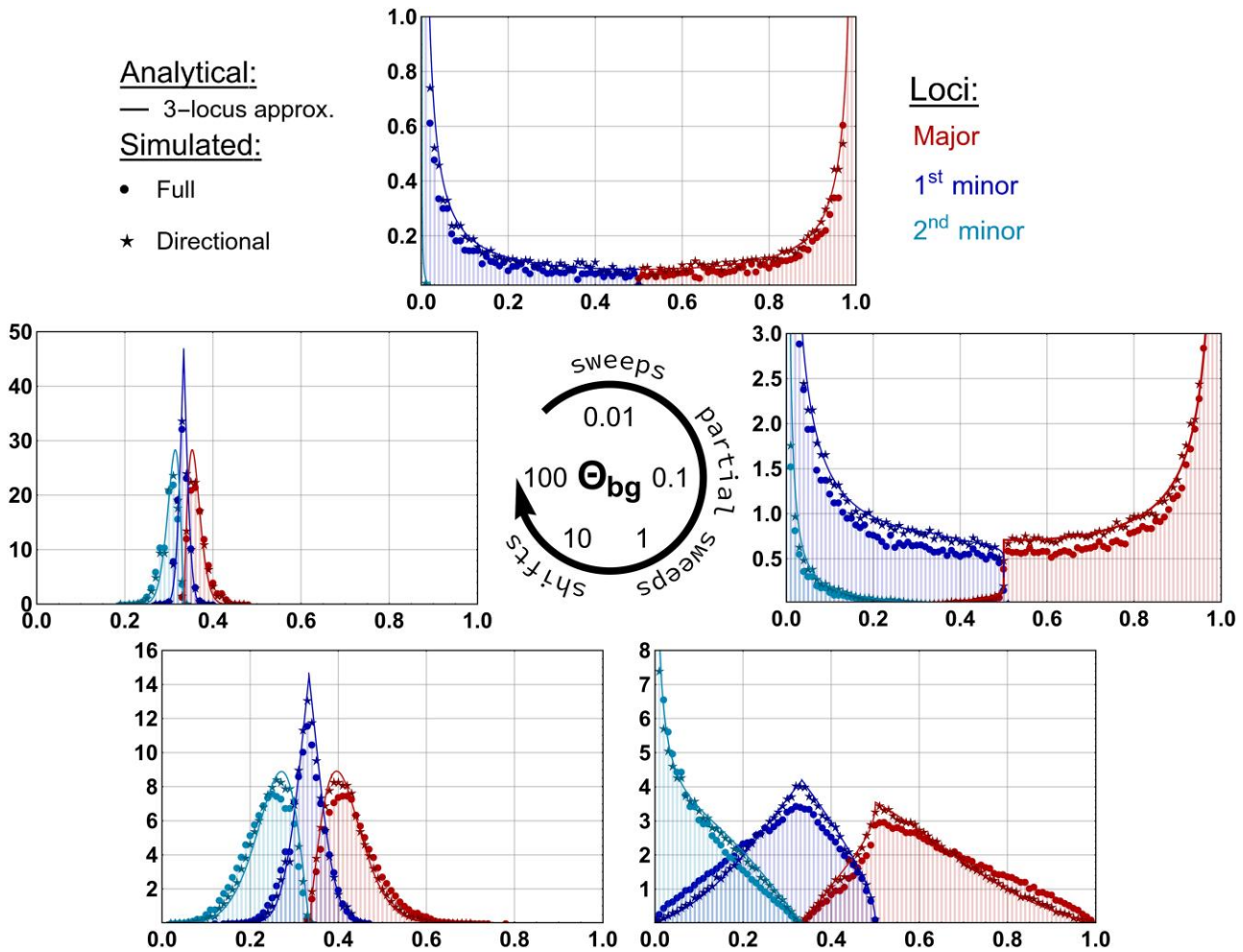
Fig. 2. Types of adaptive architecture. After an environmental change, an initially monomorphic trait adapts from $Z = 0$ towards a new optimum at $Z^{new}_{opt} = 3\gamma$, using new beneficial mutations at three loci. We observe the marginal distributions of ordered allele-frequency classes when phenotypic adaptation has proceeded by one mutational step, $\bar{Z}/\gamma = c_Z = 1$. Analytical predictions (lines) are compared to the numerical simulations of the directional selection model (asterisks) and the full model (dots). In each panel we display the distributions of (from left to right) the second minor, first minor and major locus. We find three distinct patterns of adaptive architectures: single selective sweeps for low Θ_{bg}, partial sweeps for intermediate Θ_{bg} and a collective, shift-like response of all three loci for high Θ_{bg}. $N_e = 10,000$, $N_e\sigma\gamma^2 = 100$, 100,000 replicates. Note the differences in the y-axes ranges.

fixation. In the following sections, we describe the key aspects of this transient architecture for scenarios of increasing complexity.

Architecture types and background mutation rate

We start with a particularly simple scenario that nevertheless highlights the key features of the basic architecture types. To this end, consider a trait with only three unlinked haploid loci that are all initially fixed for the a_i allele (no SGV). Selection prior to the environmental change is directional toward the lower bound of the phenotype range, $Z^0_{opt} = 0$. At $t = 0$, the trait optimum switches to the opposite end of the phenotype range, $Z^{new}_{opt} = 3\gamma$. Adaptation occurs from recurrent new mutation, with the same rate $\Theta = 2N_e\mu$ at all three loci. We record the frequencies of A_i alleles when the trait mean has increased by a single mutational step, $\bar{Z} = \gamma$ (i.e., $c_Z = 1$), which is two steps below the new optimum.

Figure 2 shows the marginal distributions of the major locus with the largest allele frequency at sampling (red), followed by the first and second minor loci (dark and light blue, respectively) for various values of the mutation rate, where $\Theta_{bg} = 2\Theta = 4N_e\mu$ measures the *background mutation rate* that is further discussed below. We show Wright- Fisher simulation results (assuming LE) for

two models: dots for the full model (Equations 3 and 4) and asterisks for the simplified directional selection model, Equation (5), without the disruptive selection term of the full model (the second term in Equation 4). The analytical predictions (solid lines) show a perfect fit for the directional selection model, for which they were derived. For the full model (dots) deviations become visible, as all marginal distributions are slightly pushed towards the boundaries of the frequency range by the disruptive selection term (with a relative strength of up to 1/4 of the strength of directional selection). In more detail, we observe the following:

- For very low mutation rates, $\Theta_{bg} \lesssim 0.1$, the adaptive change in phenotype is usually accomplished by a single locus. Adaptation thus occurs by a classic selective sweep at the first locus where an A_i allele appears and is picked up by selection. This locus is the major locus with a pronounced peak of its frequency distribution (red) at 1. Both minor loci hardly contribute to adaptation at all (blue distributions peak at 0). Note that this strong heterogeneity across loci is not visible in the marginal distribution of a single focal locus, since all

three loci can end up as the major locus with equal probability.

- For intermediate mutation rates, $0.1 \lesssim \Theta_{\mathrm{bg}} \lesssim 10$, the distributions of major and minor locus frequencies are still clearly distinct. However, minor loci now also contribute significantly to phenotypic adaptation, leading to two ($\Theta_{\mathrm{bg}} = 0.1$: major and first minor) or even three ($\Theta_{\mathrm{bg}} \geq 1$: all loci involved) partial sweeps.
- Finally, at high mutation rates, $\Theta_{\mathrm{bg}} \gtrsim 10$, the three marginal allele frequency distributions of major and minor loci gradually converge. Phenotypic adaptation occurs via the collective and homogeneous response at all three underlying loci (frequency shifts around $1/3$ at all three loci for a 3-locus trait).

The striking differences in adaptive architecture, from heterogeneous single sweeps to homogeneous collective shifts, reflect the decreasing role of stochasticity for the evolutionary process as Θ_{bg} increases. All characteristic features are captured by the analytical approximation of the joint distribution, Equation (8).

The background mutation rate Θ_{bg}

The type of architecture, between sweeps and shifts, is already decided early in the initial stochastic phase of the adaptive process. As it turns out, its degree of homogeneity is largely determined by a single composite parameter, the so-called background mutation rate Θ_{bg} (Höllinger *et al.* 2019). For a haploid model with equal locus mutation rates μ and $d \leq L$ loci carrying an a_i majority allele for $t < 0$, it reads

$$\Theta_{\mathrm{bg}} = 2N_e\mu(d - 1). \tag{10}$$

Θ_{bg} can be understood as a measure of redundancy (Barghi *et al.* 2020), because it measures the total mutational input of beneficial alleles with an equivalent effect to the allele at the major locus and are thus redundant options for the next adaptive step. Its role in the process can also be understood as follows. Consider the early stochastic phase of adaptation described by the Yule process. At some point in time, the first Yule tree is seeded by the first A_i mutation that escapes stochastic loss. Then $1/\Theta_{\mathrm{bg}}$ is the average waiting time for a second Yule tree to emerge at a different locus (measured on the timescale of the exponential growth of the first Yule tree, which has a split rate of 1). Θ_{bg} thus quantifies the expected head start of the front-runner allele over its competitors, and thus the heterogeneity of allelic contributions to the adaptive trait.

Effect of selection, trait basis size, and SGV

Equipped with these concepts, we can explore the adaptive architecture of an additive QT across a wider parameter space. In particular, we study the effects of the size, L, of the genetic basis and the selection strength, before and after the environmental change. We also include SGV as a source of adaptive variants.

Our basic setup is as follows: for a trait of size L (we use $L = 10$ and $L = 100$ in the figures), we place the ancestral optimum in the middle of the phenotype range, $Z_{\mathrm{opt}}^0 = L\gamma/2$, where we let the population equilibrate to mutation–selection–drift balance. LE is assumed. At time $t = 0$, the trait optimum jumps to a new value at a distance of three mutational steps, $Z_{\mathrm{opt}}^{\mathrm{new}} = Z_{\mathrm{opt}}^0 + 3\gamma$. As in the previous section, we record the adaptive architecture once the trait mean has increased by one mutational step ($\bar{Z} = Z_{\mathrm{opt}}^0 + \gamma$, that is, $c_Z = 1$). Adaptation over larger distances and effects of LD are discussed in the following sections. When comparing adaptive

architectures of traits with different size L of their genetic bases, we need to decide how model parameters are scaled. Following the insights of the previous section, we first make sure that the background mutation rate $\Theta_{\mathrm{bg}} = 2N_e\mu(L/2 - 1)$ is kept constant in all comparisons. We further keep the scaled selection strength $N_e\sigma\gamma^2$ prior to the environmental change constant, which leads to total expected levels of SGV that are (almost) independent of N_e and L (see the Supplementary File S1, Remark 3, for details).

Figure 3 summarizes the effects of three parameters on the adaptive architecture: the background mutation rate Θ_{bg} (between 0.01 and 100), the selection strength $N_e\sigma\gamma^2$ (10 for weak and 100 for strong) and the number of loci L (10 and 100). The simulation results are compared with analytical predictions. Most importantly, we see that neither selection strength nor trait size nor the origin of the mutations from SGV or new mutation has a qualitative effect on the adaptive architecture (compare panels in the same row of Fig. 3 and with the same Θ_{bg} in Fig. 2). As in the three-locus case, the background mutation rate emerges as the (only) crucial parameter for determining the architecture type or mode of adaptation: For $\Theta_{\mathrm{bg}} \lesssim 0.1$, we observe a single sweep at the major locus, while all other loci remain fixed at $p_i = 0$ or 1. At intermediate values, $0.1 \lesssim \Theta_{\mathrm{bg}} \lesssim 10$, adaptation proceeds via partial sweeps at a limited number of loci. The relative contribution of different (but identical) loci remains very heterogeneous (major and minor loci), highlighting a dominant role of genetic drift. For large $\Theta_{\mathrm{bg}} \geq 10$, very many loci contribute to adaptation. Individual contributions become more homogeneous, and thus approach $1/L$, leading to "subtle allele frequency shifts" for traits with a large genetic basis. For a detailed, quantitative assessment, we distinguish parameter regions where the purging of deleterious mutations does or does not play a role for phenotypic adaptation.

Adaptation dominated by beneficial variants

When a trait under stabilizing selection adapts to a shift in the optimum, there are two ways how selection on SGV can contribute: either by increasing the frequency of beneficial mutants (A_i), or by eliminating deleterious mutants (a_i) that change the phenotype in the opposite direction. In our example, with the original trait optimum in the middle of the phenotype range, half of the loci may carry rare A_i alleles and the other half rare a_i mutants for $t \leq 0$. After the environmental change, the former constitute the "beneficial variation" and the latter the "deleterious variation."

There is a large parameter range in which the maximal contribution of all rare deleterious variants to the change in $\bar{Z}$ (by eliminating all of them) is small relative to the total adaptive change in $\bar{Z}$. This is the case if mutation rates are small and/or stabilizing selection is strong, such that levels of SGV are low. In contrast, even single beneficial mutants can make a substantial contribution if they progress to fixation from a low starting frequency or if they originate as a new mutation after the environmental change.

In Fig. 3, there is hardly any deleterious variation segregating at sampling for weak selection and $\Theta_{\mathrm{bg}} \leq 0.1$, and for strong selection and $\Theta_{\mathrm{bg}} \leq 1$ (orange/yellow/green colors). We can thus assume that the corresponding loci are fixed for the A_i allele in the joint distribution and use Equation (8) to describe the joint variation at the beneficial loci. The analytical prediction is independent of selection parameters. Its good fit shows that the adaptive architecture, in this parameter region, neither depends on the strength of stabilizing selection prior to the environmental change (and, hence, on the amount of SGV)

nor on the strength of directional selection that drives the adaptation.

Contribution of deleterious variation

For high initial levels of SGV (high mutation rates and/or weak selection), deleterious variants are not yet fully purged at the sampling point (orange/yellow/green distributions in Fig. 3). Since the beneficial variation must compensate for the effect of the deleterious variation, this affects the distributions at all loci. In particular, this makes the adaptive architecture dependent on selection *before the environmental change*, because the strength of stabilizing selection affects the amount of SGV. As an example, consider the case $L = 10$ and $\Theta_{bg} = 1$ in Fig. 3. With weak selection, substantial amounts of SGV accumulate before the environmental change (shown as open squares), and some of the deleterious variation is not yet eliminated at the time of sampling. In contrast, with strong selection, the deleterious variation is completely lost. The distributions at the beneficial loci are similar in both cases, but not identical. The differences between weak and strong selection become larger for higher mutation rates, such as $\Theta_{bg} \geq 10$.

Our approximation can account for the contribution of both the beneficial and the deleterious variation to the adaptive architecture, see the *Mathematical* and *Computational* appendices for details. In short, the joint distribution of beneficial and deleterious minority alleles still follows a transformed Dirichlet distribution, as in Equation (8), but with an additional parameter κ to scale down the frequencies of the deleterious variants relative to the beneficial variants. This scaling factor depends on the total amount of SGV and thus also on the selection intensity before the environmental change. In the figure, we use this extended approximation in all panels with $\Theta_{bg} \geq 1$, where non-trivial contributions of deleterious variation can occur. For $\Theta_{bg} \leq 0.1$, the scaling factor becomes $\kappa \approx 0$ and the extended approximation reduces to Equation (8).

Effect of the number of loci L underlying the trait

Even if a trait is highly polygenic, phenotypic adaptation can be oligogenic if the number of loci that contribute to the adaptive change is much smaller. Indeed, our results show that the polymorphic part of the adaptive architecture is hardly affected by the size of the genetic basis at all, as long as L is sufficiently large (or the background mutation rate Θ_{bg} sufficiently small) such that typically fewer than L loci are polymorphic. In this case, differences in the number of loci L only lead to differences in the number of fixed loci at frequency 0 or 1 (compare $L = 10$ and $L = 100$ in Fig. 3 for $\Theta_{bg} \leq 0.1$ and approximately still for $\Theta_{bg} = 1$). This necessarily changes for larger background mutation rates once traits with a smaller genetic basis run out of further loci that could contribute to adaptation ($\Theta_{bg} \geq 10$ in Fig. 3).

We can use the approximate invariance of the adaptive architecture on L to describe the adaptive dynamics of a trait with a large genetic basis by a simpler model with few loci and rescaled locus mutation rates. As explained above, $1/\Theta_{bg}$ is the expected waiting time between the origin of the first and second beneficial mutation that contribute to phenotypic adaptation (the first and second Yule tree in our framework). Following Höllinger *et al.* (2019), we can refine this approach and match the waiting time between the first Yule tree and its jth follower to approximate the distribution of the jth minor locus. Details are given in the *Computational Supplement*.

In Fig. 3, we show how the marginal distributions of the loci with the largest contribution to the adaptive change for traits with $L = 10$ or 100 can be approximated by a 2-locus model (for $\Theta_{bg} = 0.01$) or a 3-locus model (for $\Theta_{bg} = 0.1$ and 1). The match is

excellent (with lines almost indistinguishable from the full approximation for $\Theta_{bg} \leq 0.1$) whenever the number of contributing loci does not exceed the size of the approximating model. For $\Theta_{bg} = 1$, where often more than three loci contribute, deviations appear. In particular, oligogenic adaptation beyond single sweeps is a collective phenomenon that cannot be reduced to a single-locus picture. Due to our conditioning on the phenotype (and also due to fitness epistasis), the shapes of the marginal distributions at different loci with segregating alleles are not independent.

Limits of the analytical approximation

The approximation produces a good fit of the simulation data, as long as the relevant alleles are confined to relatively small frequencies $\lesssim 10\%$ in the SGV. When the initial allele frequency distributions (open squares) extend to intermediate frequencies, two factors become important that are not included in the Yule formalism. First, disruptive selection (the term $\sim 2p - 1$ in Equation 3) adds a frequency-dependent component that increases the distance between the major and minor locus distributions in the SGV and, consequently, also in the adaptive architecture. This is clearly visible, for example, for $\Theta_{bg} = 10$ and weak selection in Fig. 3. Second, for large mutation rates, back-mutation becomes an important factor in shaping allele frequency distributions in the SGV. When strong mutation overwhelms selection, all allele frequencies are pushed toward the mutation equilibrium ($p_i = 0.5$ for equal forward and backward rates). This decreases rather than increases the distance between major and minor locus distributions. Sometimes the effects of mutation and disruptive selection can almost cancel (e.g., for $L = 10$, $\Theta_{bg} = 100$, weak selection in Fig. 3), but usually they do not, leading to deviations from the theoretical predictions. Note that high mutation rates *per se* do not compromise the approximation if the SGV is controlled by strong selection (for $N_e\sigma\gamma^2 = 100$ and $L = 100$, $\Theta_{bg} = 100$) or if adaptation occurs only from new mutation (cf., Fig. 2).

Adaptation dynamics across larger distances

So far, we have analyzed the types of adaptive architecture that emerge at a single point in the very early phase of phenotypic adaptation—a change of $\bar{Z}$ by one mutational step. We now complement this "single snapshot" with a dynamic approach to explore the changes in the adaptive architecture as the mean trait evolves across larger phenotypic distances toward a more distant optimum.

In Fig. 4, we track the adaptation dynamics of a haploid trait with 10 identical loci evolving from mutation–selection–drift balance around an ancestral optimum at $Z^0_{opt} = 2\gamma$ to a new optimum six mutation steps away, $Z^{new}_{opt} = 8\gamma$. We display marginals of the joint allele frequency distribution at five points throughout the rapid adaptive phase, when the trait mean $\bar{Z}$ reaches consecutive values one to five mutational steps away from the ancestral optimum (i.e., $\bar{Z} = 3\gamma, \ldots, 7\gamma$). As in the previous section, we examine five orders of magnitude for the background mutation rate, $\Theta_{bg} = 7 \cdot 2N_e\mu = 0.01, \ldots, 100$ (at the first mutational step, the major locus has 7 competing minor loci). We assume strong selection ($N_e\sigma\gamma^2 = 100$). Additional figures for traits with 10 and 100 loci are provided in the Appendix (Section "Adaptation with LE").

The first column in Fig. 4 (after a single mutational step) is analogous to Fig. 3 and shows different types of adaptive architectures, from a single sweep to collective shifts, as the background mutation rate increases. Further columns show that the type of the adaptive architecture is largely preserved at later observation points, while the color coded distributions of frequency-ordered loci move towards fixation at frequency $p = 1$. For small $\Theta_{bg} < 1$,
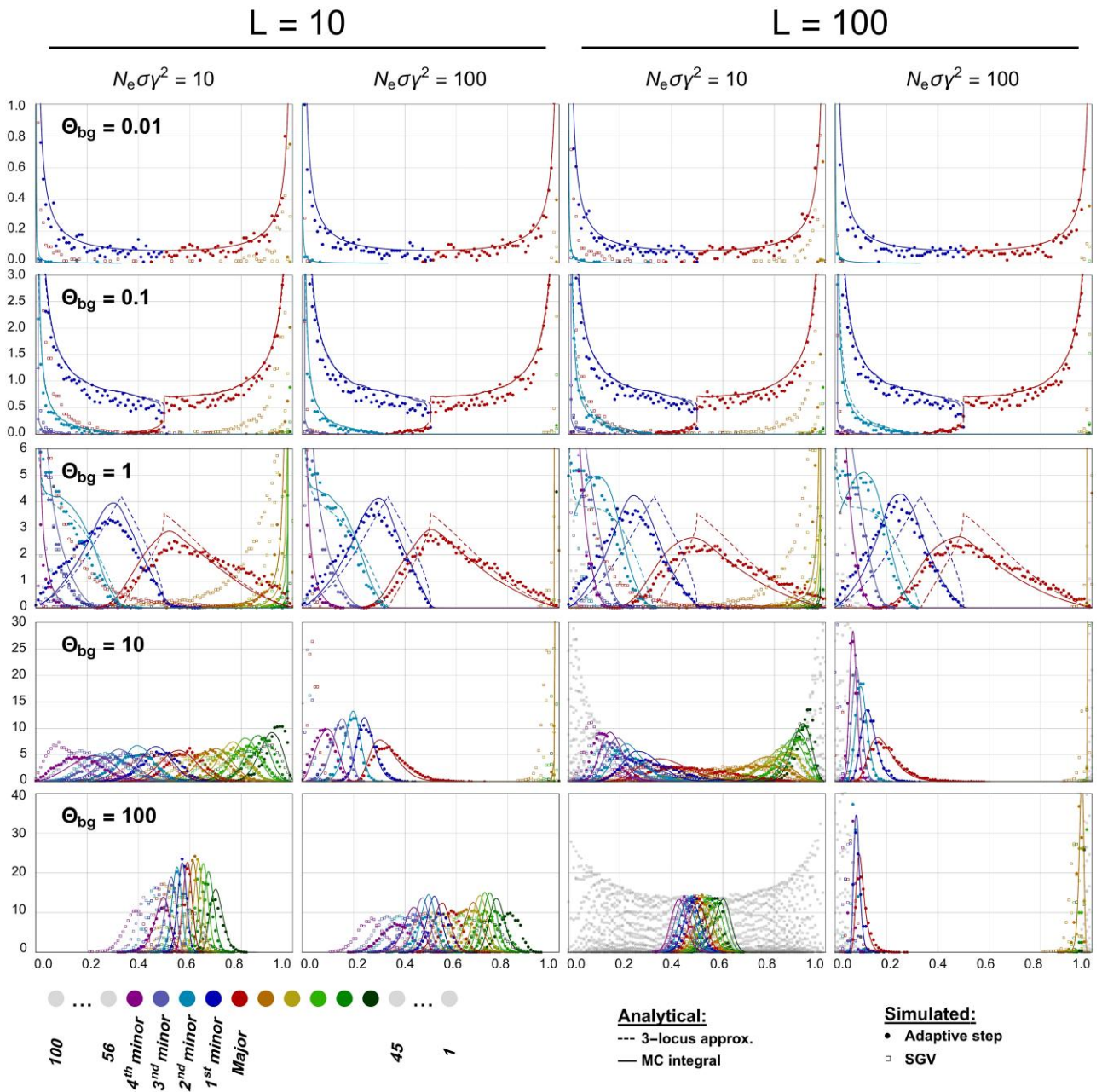
Fig. 3. Effect of selection strength and trait basis size on the adaptive architecture of a QT. The shape of the adaptive architecture, indicating selective sweeps ($\Theta_{bg} \leq 0.01$), partial sweeps ($0.1 \leq \Theta_{bg} < 10$), or shifts ($\Theta_{bg} \geq 10$), is independent of the selection strength and the number of loci in the trait basis. For a trait of size $L = 10$ or $L = 100$, the optimum shifts at time $t = 0$ from $Z^0_{opt} = L\gamma/2$ to $Z^{new}_{opt} = Z^0_{opt} + 3 \cdot \gamma$. Stabilizing selection is either weak ($N_e\sigma\gamma^2 = 10$) or strong ($N_e\sigma\gamma^2 = 100$). The allele frequency distributions of all loci ordered according to their frequency are obtained from simulations at $t = 0$ (SGV, open symbols) and after adaptation of a single mutational step, $\bar{Z} = c_Z = 1 \cdot \gamma$ (closed symbols). Lines show the analytical approximation, using either the full L-locus model (solid lines) or a 2- or 3-locus formalism where closed expressions can be derived (dashed lines), see the text for details. For $L = 100$, the distributions of the 10 "middle" loci (46 through 55), which correspond most closely to the 10-locus case, are shown in color. All other distributions are shown in gray. $N_e = 10{,}000$, 10,000 replicates. Note the differences in scaling of the y-axis.

we obtain a pattern typical of mutation-limited adaptation with successive and largely non-interfering selective sweeps. For large $\Theta_{bg} \geq 1$, we observe concerted movement of alleles at several (or all) loci simultaneously.

Limit shapes of adaptive architecture

From the figure panels of Fig. 4 (clearest for small mutation rates), it appears that the adaptive architecture, when observed at intervals of $k\gamma$ (full mutational steps), approaches a "quasi-stable" limit as phenotypic adaptation progresses. The shape of the joint distribution remains almost invariant, while the role of each locus within this joint distribution changes: a locus that contributes a small-frequency allele after the first step will contribute a larger frequency after the second and later steps, etc. In combination, these changes effectively lead to one fewer locus fixed at 0 and one more fixed at 1 with each step, but a constant pattern in the interior of the frequency space.

A stable limit shape requires that sufficiently many loci have traversed the entire frequency range from 0 to 1. This happens quickly in the sweep regime with only few segregating loci, but takes much longer for large Θ_{bg} when adaptation is achieved by small shifts at many loci. A limit is never reached if the distance
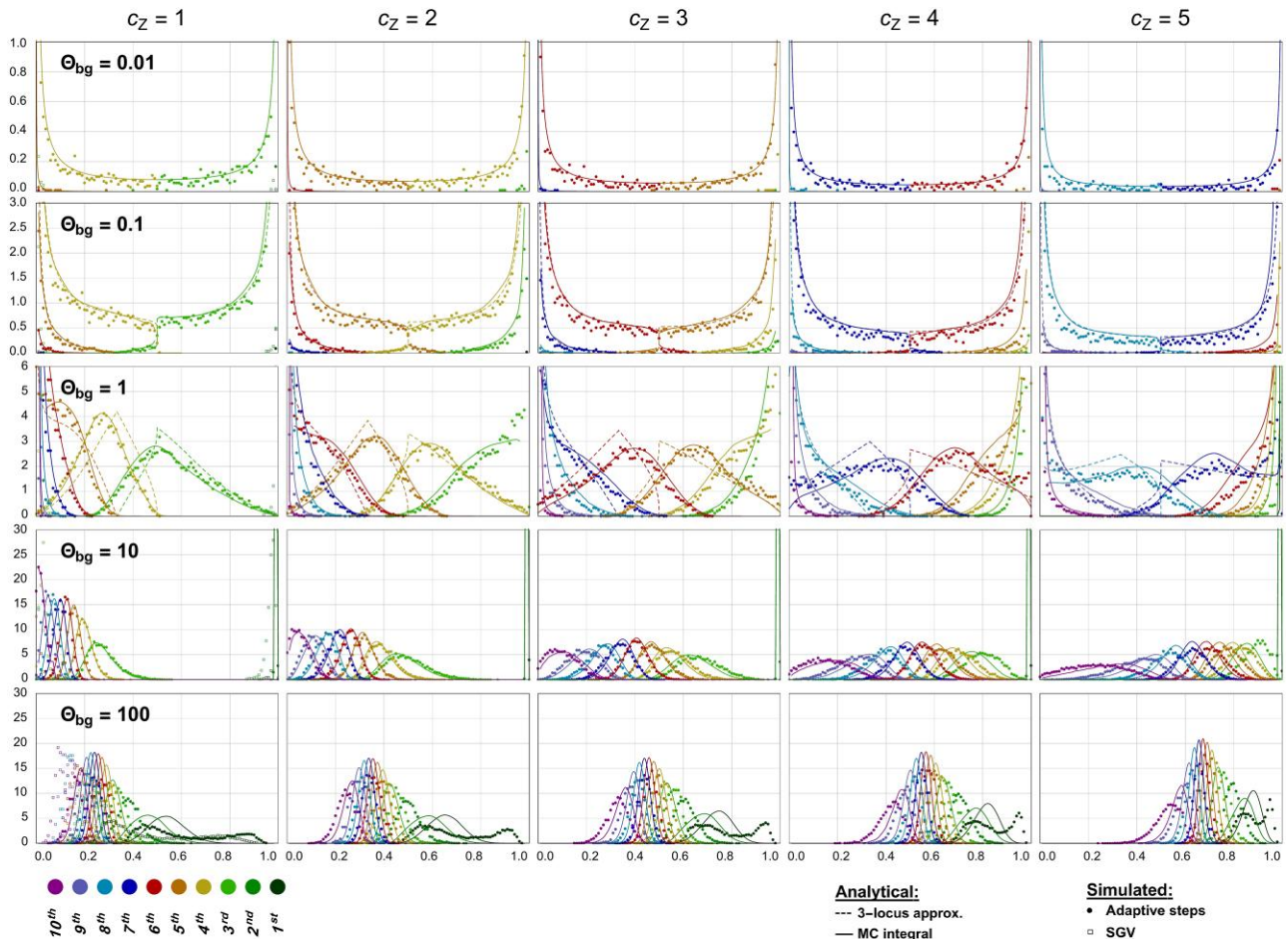
Fig. 4. Adaptation across larger distances. Successive snapshots of the adaptive architecture are shown over the course of adaptation of a QT with 10 loci. Populations evolve from mutation–selection–drift balance around the initial optimum $Z_{opt}^0 = 2\gamma$ toward a new optimum $Z_{opt}^{new} = 8\gamma$ after a change in the environment. Closed symbols show simulation results for the ordered marginal allele frequency distributions after a change in the trait mean $\bar{Z}$ by $c_Z = 1$ to $c_Z = 5$ mutational steps. The first column also shows the SGV distributions (open squares). Solid lines are analytical predictions for the full model, dashed lines (for $\Theta_{bg} \leq 0.1$) predictions from an adjusted 3-locus model (see main text). $N_e = 10,000$, $N_e\sigma\gamma^2 = 100$, 10,000 replicates. Note the differences in scaling of the y-axis.

to the new optimum is too small, or if the trait runs out of further loci that could start their adaptive course at $p = 0$ before the first loci have reached frequency $p = 1$. In our example of a trait with only 10 loci, this is already the case for $\Theta_{bg} = 10$. For traits with a larger genetic basis that adapt to a distant optimum, convergence can also be observed for larger background mutation rates (see the case $\Theta_{bg} = 10$ for a trait with $L = 100$ loci in the Appendix, Fig. A1, and also Fig. 5 for $L = 10,000$ below).

A more detailed look shows that the limit distribution in our model is only *quasi*-stable. Indeed, due to the dwindling supply of adaptive material, the expected waiting time between the origin of one beneficial mutation and the next, and thus the distance between the corresponding marginal distributions, increases with each step. As a consequence, the joint distribution gradually becomes more U-shaped (or sweep-like), with lower probability weights at intermediate frequencies, as phenotypic adaptation progresses. In Fig. 4, this is most clearly visible for $\Theta_{bg} = 0.1$ (see also the cases $\Theta_{bg} = 1$ and $\Theta_{bg} = 10$ in Fig. A1). Furthermore, see Fig. A2 in the Appendix for the dynamics of trait summary statistics for $L = 10$ and 100.

We can describe this effect as a reduction in the "effective background mutation rate" that occurs when a trait with finitely

many loci adapts over a longer distance. For the final adaptive step to the sampling point at $\bar{Z} = k\gamma$ in Fig. 4, there are $(10 - k)$ "redundant" loci that are not required to reach this value, but can still contribute to adaptation and compete with the first k loci. We thus have an effective background rate of $\Theta_{bg,eff} = \Theta_{bg}(10 - k)$ /7 in terms of the initial background mutation rate (for the first step, where $k = 3$). In our approximation of the 10-locus architecture by an effective 3-locus model (for $\Theta_{bg} \leq 1$ in Fig. 4), we match this effective rate $\Theta_{bg,eff}$ to obtain a good fit (for $\Theta_{bg} \leq 0.1$ the differences are hardly visible).

Highly polygenic traits

A model of a QT with a constant supply of new mutations (infinite loci or infinite alleles) would lead to a proper limit in the joint distribution without reduction in $\Theta_{bg,eff}$. Such "clean" results can also be observed for a trait with a finite, but very large genetic basis. Figure 5 shows results for adaptation over a large phenotypic distance (up to 800 mutational steps) of a trait with an even much larger genetic basis of $L = 10,000$ loci.

Panels a and b of Fig. 5 show the (scaled) trait variance $v_g = \sum_i p_i(1 - p_i)$, and the skew $\kappa_3(Z) = \sum_i p_i(1 - p_i)(1 - 2p_i)$, which are both summary statistics of the joint allele frequency
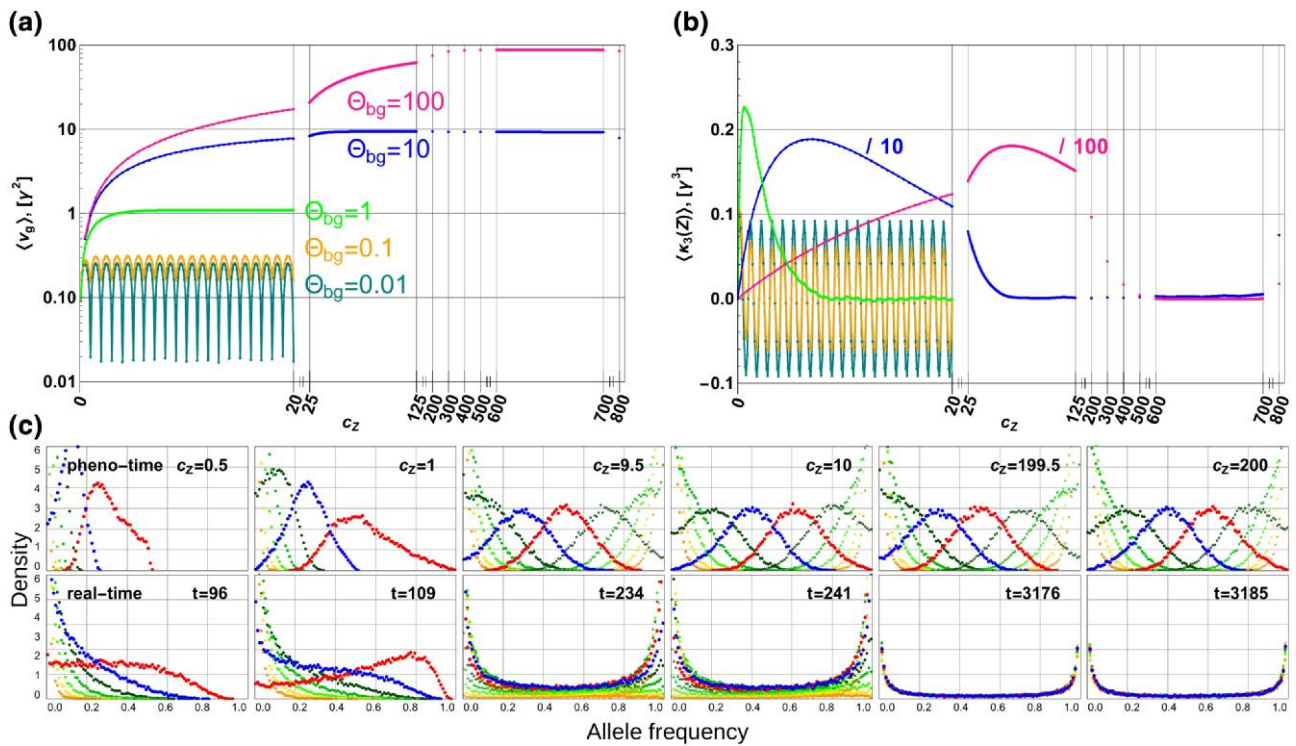
Fig. 5. Dynamics of adaptive architecture of a highly polygenic QT. Panels a) and b) show the expected genetic variance, $\langle v_g \rangle$, and skew, $\langle \kappa_3(Z) \rangle$, of the trait at sampling points in "pheno-time" for five values of $\Theta_{bg} = 0.01$. Panel c) shows the adaptive architecture (ordered marginal distributions) for $\Theta_{bg} = 1$, for sampling points in pheno-time (top row) and in real time (bottom). Stopping times are chosen such that the expected mean phenotype $\langle \bar{Z} \rangle$ across replicates in real time corresponds to the trait mean $\bar{Z}$ (in each replicate run) in the corresponding panel in pheno-time. Two central marginal distributions are shown in red (highest for the first two columns, later the one in the center for half steps or first above the center for full steps) and blue (next lower to red). To highlight the stability of the adaptive architecture at both full and half mutational steps in pheno-time, the colors are not attached to a fixed frequency rank (as in Fig. 4), but to the role of the locus in the joint distribution, that is, "red" is the largest locus in the first two panels, the 10th largest for panel 3 and 4, and the 200th largest for panels 5 and 6. Adaptation proceeds from a monomorphic starting state at $Z_{opt}^0 = 0$ by recurrent, new mutations towards $Z_{opt}^{new} = 800\gamma$. $N_e = 10{,}000$, $L = 10{,}000$, $N_e\sigma\gamma^2 = 1$, 35,000 replicates. Note that the skew for $\Theta_{bg} = 10$ and 100 in panel b) is downscaled as indicated: the inset comment "/10" indicates that presented data are 1/10th of the actual data.

distribution (the second and third cumulant). We evaluate these measures at multiple observation points for short-range (up to $c_Z = 20$ mutational steps), mid-range (up to $c_Z = 125$) and long-range (up to $c_Z = 800$) phenotypic adaptation. For short-range, in particular, this includes sampling points at non-integer intervals of mutational steps, showing that the joint distribution oscillates with period γ. Periodic fluctuations are pronounced for low Θ_{bg}, where single peaks in the marginal distributions (representing ongoing sweeps) traverse the frequency space, but become weak and almost disappear in the collective frequency-shifts regime.

While visual convergence to a stable limit period occurs within a few steps for small $\Theta_{bg} < 1$, the adaptation distance that is required rapidly increases (roughly proportional to the number of co-segregating alleles) for more polygenic architectures: 9–10 steps for $\Theta_{bg} = 1$, ≈ 70 for $\Theta_{bg} = 10$, and almost 600 for $\Theta_{bg} = 100$. Due to the large size, $L = 10{,}000$, of the genetic basis, the supply of further beneficial alleles does not change much, even over these distances. As a consequence, the slow change in the adaptive architecture that we have seen for $L = 10$ is no longer visible in the figure. In the Appendix (Section "Dynamics of Summary Statistics"), we show how the limit values for v_g relate to analytical approximations. Convergence of the skewness to (a period around) zero reflects the fact that the limit architecture (if averaged over a period) is approximately symmetric for all values of the background mutation rate.

As shown in panel C of Fig. 5, a stable shape of the adaptive architecture results only for sampling points at constant phenotype (i.e., in "pheno-time"), but not in real time. Indeed, stochastic effects on the waiting times between successive new mutations lead to a broadening of the marginal distributions, when sampling occurs at a fixed time after the environmental change. For larger times, the probability density piles up at 0 and 1, and all characteristic features of the joint distribution are eliminated.

Our analytical approximation describes not only the limit distribution, but also all dynamical changes in the adaptive architecture, as long as selection is primarily directional (Figs. 4 and A1). In particular, this includes a switch from using alleles from the SGV in the initial steps to primarily alleles that enter the population as new mutations in the later adaptive steps. This transition is seamless and occurs without any discontinuity in the shape of the joint distribution. Likewise, the shape of the limit architecture is not affected by the decrease in the strength of directional selection as the trait mean approaches the new optimum.

Once the trait mean closes in on the new phenotypic optimum, the shape of the joint allele frequency distribution eventually does change due to the action of disruptive selection against standing variation. In Figures 4 and 5, these changes become visible at the last observation point. When we follow adaptation further, selection against genetic variation pushes all frequencies toward the boundaries until all frequency distributions once again assume the typical U-shape of mutation–stabilizing–selection balance. Thus, the shapes of adaptive architecture described above are only transient, their characteristic differences eventually disappearing as the population equilibrates at the new optimum.
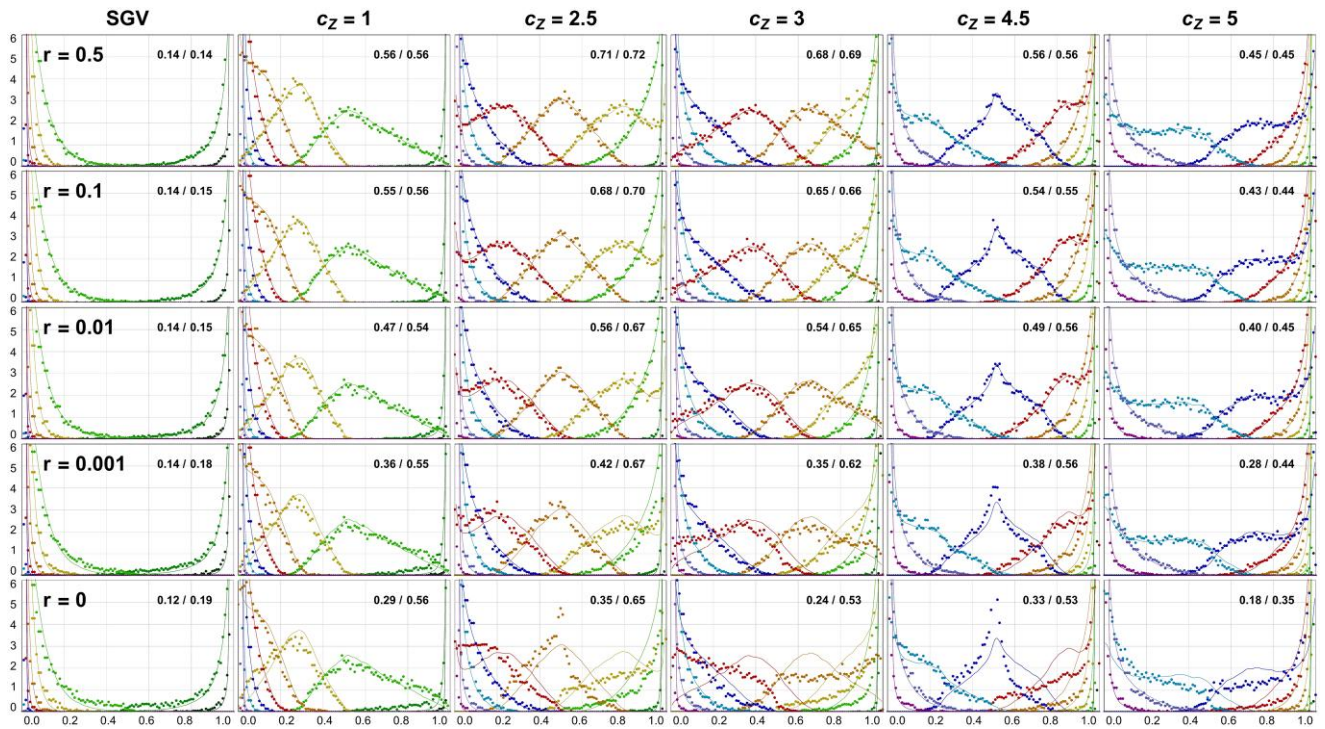
Fig. 6. Effect of linkage on the adaptive architecture of a haploid QT ($\Theta_{bg} = 1$). IB simulations with linkage (dots) are compared with LE results (lines). Numbers in the top right corner of each panel show the average genetic variance, $v_g = \text{Var}[\sum_i p_i]$, and genic variance, $v_g^{LE} = \sum_i \text{Var}[p_i]$ (in units of $[\gamma^2]$). While negative LD build up once recombination is weaker than selection ($v_g < v_g^{LE}$ for $r \lesssim 0.01 = \sigma\gamma^2$), larger deviations of the adaptive architecture only emerge for (almost) complete linkage. $N_e = 1,000$, $L = 10$, $N_e\sigma\gamma^2 = 10$, 10,000 replicates (IB), 125,000 replicates and very mild spline-smoothing for LE simulations.

Linkage and ploidy

Our analytical results and all simulations so far assume LE between all loci under selection. This assumption becomes unrealistic for highly polygenic traits when beneficial alleles co-segregate at many loci. In Fig. 6 and in the Appendix (Section "Adaptation with Linkage"), we present IB simulation results to address this issue. In addition, we analyze how our results apply to diploid genetics.

We use rescaled mutation and selection parameters for our IB simulations with $N_e = 1,000$ to match the LE simulations with $N_e = 10,000$ of the previous sections. In particular, the selection intensity of $\sigma\gamma^2 = 0.01$ prior to the environmental change (that we will use throughout) corresponds to the case of "weak" selection ($N_e\sigma\gamma^2 = 10$) in the previous figures. We assume that the loci underlying the trait are equally spaced on a single, linear chromosome, with recombination probability $r \leq 0.5$ between adjacent loci. Accordingly, the recombination probability between two loci at distance $\Delta\ell \cdot r$ Morgan reads $r_{\Delta\ell} = \sum_{k=0}^{\lfloor(\Delta\ell-1)/2\rfloor} \binom{\Delta\ell}{2k+1} r^{2k+1}(1-r)^{\Delta\ell-2k-1} \geq r$. For 10 loci and $r = 0.01$, in particular, $0.01 \leq r_{\Delta\ell} \leq r_9 \approx 0.08$. This is of the same order as the selection strength, which is $\sigma\gamma^2 = 0.01$ prior to the environmental change and $6 \cdot \sigma\gamma^2 = 0.06$ directly after the change. Recombination operates faster than selection (*loose linkage*) for $r \geq 0.1$, whereas linkage is *tight* for $r \leq 0.001$.

Figure 6 shows the effects of linkage on the adaptive architecture of a 10-locus trait with background mutation rate $\Theta_{bg} = 1$. At the environmental change, the trait optimum switches from $Z_{opt}^0 = 2\gamma$ to $Z_{opt}^{new} = 8\gamma$. Adaptation starts from SGV and is assessed at several values of $\bar{Z}$, representing both full and half mutational

steps. The top row represents the case of free linkage ($r = 0.5$). To validate our simulations and connect them to the previous sections, we also performed LE simulations for the same scenario, which are included in the figure as thin solid lines. For $r = 0.5$, we observe a perfect match for all sampling points, including at $\bar{Z} = 7\gamma$, close to the new optimum, where disruptive selection already has a strong effect. The same holds for $r = 0.1$ (loose linkage, second row). Even for $r = 0.01$ and $r = 0.001$ only minor deviations are visible, although linkage is strong and a comparison of the genetic variance, $v_g = \text{Var}[\sum_i p_i]$, and genic (or LE-)variance, $v_g^{LE} = \sum_i \text{Var}[p_i]$ (upper right corner of the figure panels) shows that negative LD build up. Negative LD is expected, both due to negative fitness epistasis and Hill–Robertson interference (Hill and Robertson 1966).

Larger deviations appear only for complete linkage ($r = 0$, bottom row), as a consequence of strong clonal interference and "mutational stacking" (Gerrish and Lenski 1998; Desai and Fisher 2007). Although the dynamics in this limit are entirely driven by the competition among haplotypes, and site frequencies are only a by-product, the qualitative features of the adaptive architecture (as measured by the frequency-ordered single-locus marginal distributions) remain surprisingly robust.

Results for other values of the background mutation rate Θ_{bg} are presented in the Appendix (Section "Adaptation with linkage"). Generally, linkage effects are very weak for small $\Theta_{bg} \ll 1$, because multi-locus polymorphism is rare. They also become weaker for (very) large $\Theta_{bg} \gtrsim L$, because recurrent new mutation at a locus reduces LD. Last, not least, we extend our results to diploid genetics. In the absence of dominance, we find perfect agreement (see Fig. A7) with the haploid predictions when the key model parameters (mutation rate, selection strength, effect sizes) are scaled

appropriately, see the Appendix (Section "Adaptation with linkage") for details.

Discussion

How genetically complex traits adapt to novel environments is a classical question of quantitative genetics. In the current study, we consider the standard model of quantitative genetics, an additive trait under Gaussian stabilizing selection that adapts to a shift in the trait optimum. However, we use a modeling approach (following Höllinger *et al.* 2019) rooted in population genetics. While we measure adaptive progress at the level of the phenotype, we set up an analytical framework to explicitly track the joint allele frequency dynamics at all loci of the genetic basis. The fundamental question is: How does a given level of phenotypic adaptation (change in the trait mean) translate into contributions (allele frequency changes) at the underlying loci? We call this collective genotypic pattern, conditional on the phenotype, the corresponding "adaptive architecture." Mathematically, it is described by the joint distribution of beneficial allele frequencies across all loci of the genetic basis. Depending on the model parameters, this distribution may describe large allele frequency changes at a few loci (i.e., sweeps) or small shifts at very many loci. Importantly, it sheds light on a large parameter range in between, where many, but not extremely many loci contribute to what we call *oligogenic adaptation* of a QT.

Adaptive architectures of a QT

When a QT adapts to a new optimum, there are two phases of adaptive evolution. First comes a (relatively) rapid phase driven by directional selection, during which the trait mean approaches the new optimum. This is followed by a (relatively much longer) equilibration phase of fine-tuning and allele sorting, in which selection against genetic variation and other, weaker evolutionary forces take over. Our model and results focus entirely on the first, rapid phase.

We assume a simple trait architecture with additive biallelic loci of all the same effect size. This assumption entails a permutation symmetry that greatly simplifies our mathematical analysis. While all loci are equivalent with respect to selection, they may differ due to the stochastic forces of mutation and genetic drift. The strength of directional selection is arbitrary and can (and usually does) depend on time and on the genetic background via epistasis for fitness. We assume that the trait adapts either through new mutations or from mutation–selection–drift balance. Under these conditions, the adaptive architecture (the joint distribution of allele frequencies at loci contributing to phenotypic adaptation) is fully described by our analytical results, with the following main features.

A single composite parameter, the population background mutation rate Θ_{bg} emerges as the critical factor to determine the characteristics of the adaptive regime, from a sweep-type architecture for small $\Theta_{bg} \ll 1$ to highly polygenic ones with small shifts for $\Theta_{bg} \gg 1$, and oligogenic patterns in between. Qualitatively, the background mutation rate can be understood as a measure of (expected) *segregating redundancy* (Láruson *et al.* 2020): alternative adaptive mutants that likely co-occur in the population for the given mutation rates. From the standpoint of a single adaptive allele, it measures the extent of competition it (likely) faces from equivalent alleles at other loci that originate and/or rise in frequency in the time interval during which the focal allele is on its way to fixation. There are three important observations. First, the adaptive regime is independent of the strength of directional selection acting on the trait. The intuitive reason for

this is that two effects of selection cancel: While the number of alternative alleles that establish per generation increases with selection (establishment probability $p_{est} \sim s$), faster growth shortens the time window for competing alleles to arise ($\sim 1/s$ in the early phase of exponential growth). This result is analogous to the probability of soft selective sweeps, where the effects of selection on p_{est} and on the window of opportunity in which further beneficial alleles at the same locus can arise also cancel (Hermisson and Pennings 2005; Pennings and Hermisson 2006). Second, it does not matter whether adaptation occurs from SGV in mutation–selection–drift balance or only from new mutation. In particular, for a given Θ_{bg}, adaptation from SGV is not "more polygenic." Indeed, the heterogeneity between allele frequencies of equivalent loci in the SGV is (maybe surprisingly) large and exactly the same as the one that results from recurrent new mutation. Third, the size L of the genetic basis of the trait has only an indirect effect on the type of adaptation, via Θ_{bg}. In particular, adaptation for a highly polygenic trait (large L, as would be observed in GWAS) can still be oligogenic or even sweep-like if beneficial mutation rates are low. In this case, the adaptive process can readily be described by a low-dimensional effective model with small L and appropriately matched Θ_{bg} (see "Results").

Adaptation dynamics in pheno-time

Our analytical method relies on a change of scale, on which the adaptive process is described. Instead of the usual time dynamics, we assess the adaptive architecture of a trait for fixed values of the corresponding trait mean, $\bar{Z}$, that is, we follow "pheno-time" instead of real time. From an empirical perspective, pheno-time is a natural choice because adaptive progress at the phenotypic level is often observable. Mathematically, the change of scale reveals a robust pattern that is not visible for replicate samples that are taken after a fixed time. Due to epistasis (the selection strength changes with the distance to the optimum) and genetic drift (replicates reach phenotype values at different times), both scales are not related by a simple transformation.

We observe a gradual buildup of adaptive architecture with the following steps. Any adaptive process starts with small frequency changes of alleles in the SGV: on average across replicates, the frequencies of beneficial variants slightly increase, while those of deleterious variants decrease. The adaptive architecture in this initial phase is simply a perturbed version of the joint allele frequency distribution in the balance of mutation, stabilizing selection, and drift. The phenotypic distance that is traversed in this initial phase depends on the amount of SGV and thus not only on mutation rates, but also on the strength of stabilizing selection. It can be large in the highly polygenic regime (large mutation/selection ratio, $\Theta_{bg}/(\sigma\gamma^2)$, in our model).

Subsequently, as frequency changes get larger, deleterious variants are gradually purged from the population. The non-trivial polymorphic part of the adaptive architecture then consists purely of beneficial variants, both from SGV and from new mutation. Henceforth, the shape of the architecture is governed solely by the background mutation rate and is independent of the selection strength and the proportion of alleles that originate from SGV (which could be zero). In pheno-time, its shape follows the beneficial alleles through frequency space and converges to a quasi-stable limit once these reach fixation. Alleles that exit the frequency space by fixation are replaced by alleles at other loci that enter by new mutation, as long as there is still a sufficient supply.

The quasi-stable limit shape reflects the type of phenotypic adaptation: isolated peaks in the size-ordered marginal

distributions represent successive sweeps, whereas broad distributions at many loci with strong overlaps indicate highly collective modes of adaptation. Limit architectures are quickly reached for small background mutation rates, but require adaptation across large phenotype distances in the highly polygenic case.

Finally, once adaptation progresses over distances on the order of the total phenotype range, the shape slowly changes in the direction of a less polymorphic type (smaller Θ_{bg}) due to the dwindling supply of new mutations. For a trait with finitely many loci, it is therefore only quasi-stable.

All these results assume that selection is directional, and the mean trait value has not yet reached the new optimum (the *rapid phase* in Hayward and Sella 2022) Once the new adaptive optimum is reached, selection at the locus level is no longer directional, but disruptive and generally much weaker. The slow process of allele sorting that follows, and the eventual return to equilibrium at the new optimum, are not captured by our formalism.

Theory of oligogenic adaptation

The adaptive architectures described by our framework interpolate between two regimes that provide a simplified description of the adaptive process. In the "monogenic" limit, adaptation can be conceptualized as an adaptive walk. Stochastic effects are important in this regime and lead to a heterogeneous response (complete sweeps at some loci, no response at others), but the individual steps of the walk are largely independent and can be described by single-locus theory (e.g., Kopp and Hermisson 2009b). At the other, polygenic, end of the scale, phenotypic adaptation is a collective response of alleles at very many loci. Genetic drift can be ignored, and deterministic theory can be used to describe the initial, rapid adaptive phase (e.g., Lande 1983). Loci may interact, but as long as the individual allele frequency shifts are tiny, epistatic interactions hardly play a role.

In the intermediate, oligogenic regime, phenotypic adaptation is achieved by simultaneous allele frequency shifts at several, but not very many loci. In contrast to the monogenic case, adaptation is collective, and polymorphic alleles at different loci interact due to fitness epistasis. Sweeps are often partial, rather than completed, as beneficial alleles only rise to intermediate frequencies. In contrast to the highly polygenic scenario, however, frequency changes are more than just a perturbation of the standing variation. Stochastic effects are important and lead to heterogeneous contributions of otherwise identical loci (major–minor locus structure due to a mix of larger and smaller shifts).

Mathematically, oligogenic adaptation is the most challenging regime. Previous models mostly rely on deterministic theory and/or simulation studies. An insightful analytical approach is the model by Jain and Stephan (2015, 2017a, 2017b). Building on work by de Vladar and Barton (2014), these authors study an additive QT and derive frequency trajectories of single alleles with arbitrary effects when adaptation occurs from deterministic mutation–selection balance. This initial condition favors small-effect alleles, which start from a high frequency of $p = 0.5$ (de Vladar and Barton 2014). As a consequence, the prevalence of large, sweep-like allele-frequency changes depends on the effect size, with a higher prevalence of sweeps if locus effects are large. These results extend to finite populations if and only if the population-scaled mutation rate per locus is large, $\Theta = 2N_e\mu > 1$, such that the distribution of weakly selected small-effect alleles in the SGV is unimodal (Devi and Jain 2023). In our model, with low Θ and/or stronger selection, alleles typically start from a low frequency. In this case, the type of adaptive

architecture (sweeps or shifts) does not depend on the effect size. Instead, the background mutation rate Θ_{bg} (which does not exist in a deterministic model) emerges as the decisive parameter. Note that our notion of a "sweep-type architecture" only refers to the size of the change in allele frequency and not to the speed: Slow frequency changes for weak-effect alleles hardly produce a discernible footprint in linked neutral variation (Thornton 2019).

The closest correspondence to our study is the analysis of a binary trait in Höllinger *et al.* (2019), where the Yule-framework has first been used. In their model, adaptation to the new optimum is achieved by a single mutational step at one of several loci underlying the trait. Many results, such as the key role of the background mutation rate Θ_{bg} and the independence of the adaptive architecture from selection strength, are equivalent in both models, demonstrating their generality. However, phenomena of multistep adaptation, such as a stable limit architecture or the effect of deleterious variation, can only be studied for a QT.

A related method, combining a Galton–Watson branching process with a deterministic logistic growth model, has recently been presented by Götsch and Bürger (2023) to study the adaptation of an additive QT under exponential directional selection from new mutation. Due to the absence of epistasis, the method can be set up in real time and allows for arbitrary locus effect sizes. Götsch and Bürger (2023) use the number of segregating alleles under selection to characterize the adaptive architecture. Once again, the total population-scaled rate of new mutations Θ_t (the infinite-loci counterpart of our parameter Θ_{bg}) emerges as the main determinant of the pattern of adaptation.

For highly polygenic traits, further analytical approaches become available. They describe the dynamics of individual loci in a "mean-field" quantitative background, which evolves according to simple deterministic dynamics. The classical approach dates back to Lande (1983) and has been used by Chevin and Hospital (2008), Chevin (2019) to describe adaptation of a major-effect allele ("QTL") in a highly polygenic background approximated by a normal distribution. With stabilizing selection, single loci are quickly outcompeted by the joint action of the background, preventing sweep-like changes. Sweeps only occur if the shift in trait optimum is large and large-effect alleles already contribute significantly to the initial SGV (Chevin and Hospital 2008; John and Stephan 2020; Stephan and John 2020; Devi and Jain 2023).

An elegant alternative approach was recently developed by Hayward and Sella (2022). Their method uses that, for an additive trait and assuming LE, selection at single loci depends on the genetic background only via the trait mean $\bar{Z}(t)$ (compare Equation 3). For a highly polygenic trait, $\bar{Z}(t)$ becomes approximately independent of the dynamics at single loci. One can then insert the deterministic solution for $\bar{Z}(t)$ into a single-locus diffusion equation to analyze the stochastic dynamics of individual alleles in the genetic basis.

A condition for Hayward and Sella (2022)'s method is that the shift in the trait optimum is small enough that "adaptation to the new optimum requires only a small average frequency change per segregating site." Under this assumption, all frequency changes during the rapid adaptive phase are first-order perturbations of SGV frequencies. In our model and the examples shown, these conditions are met only in the highly polygenic case of $\Theta_{bg} = 100$ and phenotypic adaptation of no more than a few ($\lesssim 10$) steps. In this case, adaptation to the new optimum is already complete with the first step of the adaptive architecture construction described above, before its characteristic features begin to show. Accordingly, the analysis by Hayward and Sella (2022) mainly describes the relative frequency change of alleles with

different effect size and the consequences for allele sorting during the equilibration phase—issues that are beyond the scope of our work.

Oligogenic and polygenic adaptation

The transition between oligogenic and polygenic adaptation is gradual, with no clear demarcation. Besides the background mutation rate, Θ_{bg} (which determines the number of co-segregating alleles), it also depends on the range of phenotypic adaptation. Indeed, if (say) 20 beneficial alleles from the SGV are picked up by selection, adaptation can appear as "polygenic," with small frequency shifts when the phenotype only changes by a single mutational step, but no longer at larger distances. At, say, 10 steps or more, frequency changes become large, interactions matter, and new mutations start to play a role.

A hallmark of the highly polygenic regime is that the genetic variance remains approximately constant, which is often used as a model assumption in quantitative genetic approaches (e.g., Lande 1983; Chevin and Hospital 2008; de Vladar and Barton 2014). For the infinitesimal model, a stable variance is a consequence of the even stronger assumption that selection does not change allele frequencies. As Hayward and Sella (2022) point out, it still holds for small, but non-zero frequency shifts, as long as the increase in variance due to the increase of aligned alleles (the beneficial variation) is offset by the decrease in variance due to the frequency decrease of opposing alleles (deleterious variation).

We can contrast this with characteristics of oligogenic adaptation. With a limited number of polymorphic alleles, allele frequency changes necessarily become larger. As a result, the effect of beneficial variation for changing genetic variance, as well as its contribution to the adaptive progress of the trait mean, outweighs the effect of the deleterious variation that is eventually eliminated from the population. Genetic variance typically increases during buildup of the adaptive architecture until the quasi-stable limit shape is reached. In this later phase, the genetic variance is once again approximately stable, either constant or oscillating. However, this is for different reasons than for highly polygenic adaptation over short distances: While the deleterious variation no longer plays a role, beneficial variants reach much higher frequencies. Depending on whether this frequency is smaller or larger than 0.5, further frequency increases can increase or decrease the genetic variance. In the limit, the contributions of different loci approximately cancel. The loss of variation due to fixation of alleles is compensated by recurrent new mutation, as previously described by Hill (1982a, 1982b) for adaptation of a quantitative trait under long-term truncation selection. A comprehensive discussion of both types of stable variance can be found in Götsch and Bürger (2023).

Biology of oligogenic adaptation

Throughout the history of quantitative genetics, evidence for highly polygenic adaptation has been gathered from various sources. This includes the classic observation that the response to artificial or natural selection on QTs is typically rapid and without major changes in genetic variance (reviewed in Sella and Barton 2019; Flatt 2020). Modern GWAS results largely supports the view of phenotypic adaptation proceeding via frequency changes at many loci with small individual effect (Shi et al. 2016). The contribution of genomic regions to the heritability is often roughly proportional to their length, which is consistent with an infinitesimal or "omnigenic" model (Boyle et al. 2017; Liu et al. 2019).

However, empirical evidence that goes beyond the infinitesimal model is equally widespread. In particular, major allele frequency changes driven by positive selection are frequently observed. This includes signals of sweeps and partial sweeps, but also polygenic footprints in quintessential QTs, such as human body size (Field et al. 2016). Such footprints do not exist with truly infinitesimal genetics.

A polygenic genetic architecture implies a high level of redundancy (Barghi et al. 2020; Láruson et al. 2020), reflecting the number of different ways how alleles can combine to produce the adaptive phenotype. As argued by Hayward and Sella (2022), it then "becomes uninteresting to focus on the particular subset of alleles" that was recruited—largely at random—to accomplish this task. However, data from replicated events of adaptive evolution often show a much higher level of parallelism than might be expected from the number of underlying loci and the segregating variation. Examples are summarized in Láruson et al. (2020), Barghi et al. (2020) and come from both natural evolution (e.g., Conte et al. 2015; Yeaman et al. 2018a) and from "Evolve and Resequence" experiments (e.g., Barghi et al. 2019).

Larger frequency changes and substantial parallelism are hallmarks of oligogenic adaptation: They show that the "segregating redundancy" that is available for adaptation is not unlimited. In models with equivalent loci (both for an additive and a binary trait, cf., Höllinger et al. 2019), the composite parameter $\Theta_{bg} = 2N_e\mu(d-1)$ is the appropriate measure of redundancy and determines the type of adaptation. Oligogenic characteristics are expected if (at least) one of the factors in Θ_{bg} is small. Empirically, they are highly variable: While $2N_e\mu$ measures the expected diversity per locus and strongly depends on the population size, d is the number of loci available to respond to a new selection pressure and depends on the size of the genetic basis. For example, melanism in flies (Bastide et al. 2016) or lipid traits in humans (Shi et al. 2016) have a much smaller genetic basis than classical size- or yield-traits.

However, it is important to distinguish the genetic basis of a trait from its "adaptive basis," that is, the loci that are sufficiently free of pleiotropic, epistatic, or developmental constraints to contribute to sustained adaptive change (Yeaman et al. 2018b; Yeaman 2022). Only the latter enter into d and, thus, affect Θ_{bg}. Pleiotropy is necessarily ubiquitous for omnigenic traits, where trans-acting alleles in peripheral pathways with minute effects on the focal trait are thought to contribute the bulk of heritable variation (Liu et al. 2019). If these alleles are strongly constrained due to their primary function, the basis of the adaptive architecture could be much smaller than the total genetic basis of these traits as seen in GWAS (Barghi et al. 2020; Láruson et al. 2020).

Scope and limits of our model

The most stringent limitation of the model concerns the assumptions on the trait genetics: alleles at all loci are additive, with equal effect sizes on the trait and on fitness. While our method describes the heterogeneity among frequencies of beneficial alleles due to mutation and drift (the stochastic variation within fitness classes), it does not capture heterogeneity due to selection differences (the deterministic variation between fitness classes). The only exception is the distinction of two classes with loci that harbor the beneficial and deleterious variation, respectively. Already this extension shows that the consequences of unequal effect sizes on the adaptive architecture are not only quantitative, but also qualitative. In particular, the background mutation rate is no longer the only relevant parameter, but both absolute and relative selection strength matter. Similarly, the presence (or not) of

standing variation and its distribution across different locus classes affect the results. While selection differences among loci increase the heterogeneity of the adaptive architecture, our results show, however, that observed patterns from a single replicate may not be easy to distinguish from a "null model" with equal loci. The extension to deleterious loci also shows how, in principle, multiple classes of loci can be included that differ either in their effect on the trait or in pleiotropic effects on fitness. Each additional class requires an additional model parameter and a separate fit to the starting configuration prior to the onset of directional selection. A general model with many classes can become complex, and the scope of this approach remains to be explored.

A second limitation refers to the selection regime. Our method relies on directional selection acting on all loci, as is typical during the early phase of phenotypic adaptation after a change of the optimum. There is no straightforward extension to other modes of selection, including disruptive selection (at the locus-level) that drives adaptive fine-tuning during the equilibration phase. The assumption is also important prior to the environmental change, where standing variation is maintained by a balance of mutation and negative directional selection. For stabilizing selection, this is approximately true if selection is strong enough to keep deleterious alleles at a low starting frequency, $p_i \lesssim 0.1$. It does not hold if much of the trait adaptation is due to common alleles that are already segregating at intermediate frequencies before adaptation, either because of very weak selection, or because of past episodes of directional selection, or because they are maintained by other forces, such as balancing selection or recurrent gene flow.

Despite these restrictions, the directional selection model encompasses a large set of scenarios. In particular, our results for the adaptive architecture after a sudden shift of the trait optimum can also be applied to the second standard model of phenotypic adaptation: They hold (without change) for a trait with a moving optimum (e.g., Lynch *et al.* 1991; Bürger 2000; Jones *et al.* 2004; Kopp and Hermisson 2009a, 2009b), as long as the gap between the optimum and the trait mean remains sufficiently large that overshooting can be ignored (the "mutation-limited regime" of Kopp and Hermisson 2009a, 2009b). In general, the method is much less restrictive with respect to the ecological assumptions, which determine the shape of the phenotypic fitness function and its change in time, than with respect to the genetic assumptions. This includes fitness epistasis of any order (see the Supplementary File S1 for details), as long as the simple additive trait genetics ensures that all loci are affected in the same way.

Our analytical predictions assume LE between all loci in the genetic basis of the trait, which is unrealistic, especially for traits with a highly polygenic basis. Comparison with IB simulations show, however, that the results for the joint allele-frequency distribution remain accurate even for strong linkage and long-distance adaptation. This is true even though we observe (as expected, Hill and Robertson 1966) a significant increase in negative LD when the average recombination rate between pairs of loci is smaller than the selection coefficient. Although the model assumption of LE is violated, this has surprisingly little effect on the distribution of allele frequencies. Only if linkage is complete do larger deviations in the direction of a more polygenic adaptive architecture occur. It would be interesting to relate our results to the predictions from models of clonal interference (Gerrish and Lenski 1998; Desai and Fisher 2007) in this limit. Relevant effects are also expected if selection effectively acts on linkage blocks, either because initial levels of LD are large (e.g., due to admixture) or when loci underlying the trait are densely distributed along the genome (cf., Sachdeva and Barton 2018; Sachdeva 2022). An important condition for the robustness of the LE results is the additive trait architecture. In particular, the mean trait $\bar{Z}$ that defines the pheno-time scale is invariant under changes in LD. This is different for the strongly epistatic binary trait architecture considered in Höllinger *et al.* (2019), where stronger effects of linkage have been observed.

Finally, all our results assume a single, panmictic population of constant size. Specifically, the population-scaled locus mutation rates $\Theta_i = 2N_e\mu_i$ should remain approximately constant over the time period described by the dynamics (before and after the environmental change). If adaptation occurs in a spatially extended population, different beneficial variants may dominate in different regions and subsequently mix through gene flow. For a single adaptive step and alleles that are mutually exclusive, the resulting spatial patterns have been analyzed (Ralph and Coop 2010; Paulose *et al.* 2019) and compared with empirical data (Feder *et al.* 2019). Furthermore, a trait optimum can vary in space and time (Polechová *et al.* 2009; Polechová 2018). All of these scenarios are expected to significantly affect patterns of adaptive architecture, both in the short term and in the long term, when gene flow drives allelic turnover (Yeaman 2015). A simpler situation in which our method could be applied is adaptation from migration–selection balance in a continent-island model. Extensions are needed to deal with strong LD due to multi-locus migration and potential barriers to gene flow. For this model, Sachdeva (2022) recently found that the effects of population-wide LD in migration–selection equilibrium can be accounted for by an adequately defined effective migration rate, providing a starting-point for a non-equilibrium analysis.

Data availability

The associated *Dryad* directory is accessible under the DOI https://doi.org/10.5061/dryad.573n5tbc9 (Höllinger *et al.* 2023). This repository contains comprehensive *Mathematica* Notebooks (*Computational* and *Figures Supplement*), simulation scripts (C++) and created (simulation) data used for plotting. All data and scripts to recreate the presented figures are available there.

Supplementary material are available at GENETICS online.

Conflicts of interest statement

The authors declare no conflict of interest.

Literature cited

Barghi N, Hermisson J, Schlötterer C. Polygenic adaptation: a unifying framework to understand positive selection. Nat Rev Genet. 2020;21:769–781. doi:10.1038/s41576-020-0250-z

Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, Dolezal M, Taus T, Kofler R, Schlötterer C. Genetic redundancy fuels polygenic adaptation in Drosophila. PLOS Biol. 2019;17:e3000128. doi:10.1371/journal.pbio.3000128

Barton NH. The maintenance of polygenic variation through a balance between mutation and stabilizing selection. Genet Res. 1986;47:209–216. doi:10.1017/S0016672300023156

Barton NH. The effect of hitch-hiking on neutral genealogies. Genet Res. 1998;72:123–133. doi:10.1017/S0016672398003462

Barton NH, Etheridge A, Véber A. The infinitesimal model: definition, derivation, and implications. Theor Popul Biol. 2017;118:50–73. doi:10.1016/j.tpb.2017.06.001

Bastide H, Lange JD, Lack JB, Yassin A, Pool JE. A variable genetic architecture of melanic evolution in Drosophila melanogaster. Genetics. 2016;204:1307–1319. doi:10.1534/genetics.116.192492

Bell G. The oligogenic view of adaptation. Cold Spring Harb Symp Quant Biol. 2009;74:139–144. doi:10.1101/sqb.2009.74.003

Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. Cell. 2017;169:1177–1186. doi:10.1016/j.cell.2017.05.038

Bürger R. The Mathematical Theory of Selection, Recombination, and Mutation. Chichester (UK): Wiley; 2000.

Chevin LM. Selective sweep at a QTL in a randomly fluctuating environment. Genetics. 2019;213:987–1005. doi:10.1534/genetics.119.302680

Chevin LM, Hospital F. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics. 2008;180:1645–1660. doi:10.1534/genetics.108.093351

Conte GL, Arnegard ME, Best J, Chan YF, Jones FC, Kingsley DM, Schluter D, Peichel CL. Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. Genetics. 2015;201:1189–1200. doi:10.1534/genetics.115.182550

Desai MM, Fisher DS. Beneficial mutation–selection balance and the effect of linkage on positive selection. Genetics. 2007;176:1759–1798. doi:10.1534/genetics.106.067678

Devi A, Jain K. Polygenic adaptation dynamics in large, finite populations. bioRxiv. 2023.

de Vladar HP, Barton NH. Stability and response of polygenic traits to stabilizing selection and mutation. Genetics. 2014;197:749–767. doi:10.1534/genetics.113.159111

Fagny M, Austerlitz F. Polygenic adaptation: integrating population genetics and gene regulatory networks. Trends Genet. 2021;37:631–638. doi:10.1016/j.tig.2021.03.005

Feder AF, Pennings PS, Hermisson J, Petrov DA. Evolutionary dynamics in structured populations under strong population genetic forces. G3. 2019;9:3395–3407. doi:10.1534/g3.119.400605

Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, Yengo L, Rocheleau G, Froguel P, McCarthy MI, et al. Detection of human adaptation during the past 2000 years. Science. 2016;354:760–764. doi:10.1126/science.aag0776

Fisher R. The correlation between relatives on the supposition of mendelian inheritance. Trans R Soc Edinb. 1918;52:339–433.

Flatt T. Life-history evolution and the genetics of fitness components in Drosophila melanogaster. Genetics. 2020;214:3–48. doi:10.1534/genetics.119.300160

Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. Genetica. 1998;102/103:127–144. doi:10.1023/A:1017067816551

Ghorbel M. On the inverted Dirichlet distribution. Commun Stat - Theory Methods. 2009;39:21–37. doi:10.1080/03610920802627062

Götsch H, Bürger R. Evolution of quantitative traits under directional selection: selective sweeps or small allele-frequency changes? bioRxiv. 2023.

Gouy A, Daub JT, Excoffier L. Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res. 2017;45:e149–e149. doi:10.1093/nar/gkx626

Hayward LK, Sella G. Polygenic adaptation after a sudden change in environment. eLife. 2022;11:e66697. doi:10.7554/eLife.66697

Hermisson J, Pennings PS. Soft sweeps. Genetics. 2005;169:2335–2352. doi:10.1534/genetics.104.036947

Hill WG. Predictions of response to artificial selection from new mutations. Genet Res. 1982a;40:255–278. doi:10.1017/S0016672300019145

Hill WG. Rates of change in quantitative traits from fixation of new mutations. Proc Natl Acad Sci USA. 1982b;79:142–145. doi:10.1073/pnas.79.1.142

Hill W, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8:269–294. doi:10.1017/S0016672300010156

Höllinger I, Pennings PS, Hermisson J. Polygenic adaptation: from sweeps to subtle frequency shifts. PLOS Genet. 2019;15:e1008035.

Höllinger I, Wölfl B, Hermisson J. A theory of oligogenic adaptation of a quantitative trait, Dryad, Dataset. https://doi.org/10.5061/dryad.573n5tbc9, 2023.

Jain K, Stephan W. Response of polygenic traits under stabilizing selection and mutation when loci have unequal effects. G3. 2015;5:1065–1074. doi:10.1534/g3.115.017970

Jain K, Stephan W. Modes of rapid polygenic adaptation. Mol Biol Evol. 2017a;34:3169–3175. doi:10.1093/molbev/msx240

Jain K, Stephan W. Rapid adaptation of a polygenic trait after a sudden environmental shift. Genetics. 2017b;206:389–406. doi:10.1534/genetics.116.196972

John S, Stephan W. Important role of genetic drift in rapid polygenic adaptation. Ecol Evol. 2020;10:1278–1287. doi:10.1002/ece3.5981

Jones AG, Arnold SJ, Burger R. Evolution and stability of the G-matrix on a landscape with a moving optimum. Evolution. 2004;58:1639–1654.

Kaplan NL, Hudson RR, Langley CH. The "hitchhiking effect" revisited. Genetics. 1989;123:887–899. doi:10.1093/genetics/123.4.887

Kopp M, Hermisson J. The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. Genetics. 2009a;182:233–249. doi:10.1534/genetics.108.099820

Kopp M, Hermisson J. The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. Genetics. 2009b;183:1453–1476. doi:10.1534/genetics.109.106195

Lande R. Natural selection and random genetic drift in phenotypic evolution. Evolution. 1976;30:314–334. doi:10.2307/2407703

Lande R. The response to selection on major and minor mutations affecting a metrical trait. Heredity. 1983;50:47–65. doi:10.1038/hdy.1983.6

Láruson ÁJ, Yeaman S, Lotterhos KE. The importance of genetic redundancy in evolution. Trends Ecol Evol. 2020;35:809–822.

Lipowski A, Lipowska D. Roulette-wheel selection via stochastic acceptance. Physica A. 2012;391:2193–2196. doi:10.1016/j.physa.2011.12.004

Liu X, Li YI, Pritchard JK. Trans effects on gene expression can drive omnigenic inheritance. Cell. 2019;177:1022–1034.e6. doi:10.1016/j.cell.2019.04.014

Lynch M, Gabriel W, Wood AM. Adaptive and demographic responses of plankton populations to environmental change. Limnol Oceanogr. 1991;36:1301–1312. doi:10.4319/lo.1991.36.7.1301

Martin G, Lambert A. A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. Theor Popul Biol. 2015;101:40–46. doi:10.1016/j.tpb.2015.01.004

Maynard-Smith J, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974;23:23–35.

Orr HA. The genetic theory of adaptation: a brief history. Nat Rev Genet. 2005;6:119–127. doi:10.1038/nrg1523

Paulose J, Hermisson J, Hallatschek O. Spatial soft sweeps: patterns of adaptation in populations with long-range dispersal. PLoS Genet. 2019;15:e1007936. doi:10.1371/journal.pgen.1007936

Pennings PS, Hermisson J. Soft sweeps II–molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol. 2006;23:1076–1084. doi:10.1093/molbev/msj117

Polechová J. Is the sky the limit? On the expansion threshold of a species' range. PLoS Biol. 2018;16:e2005372.

Polechová J, Barton N, Marion G. Species' range: adaptation in space and time. Am Nat. 2009. 174:E186–E204.

Pritchard JK, Di Rienzo A. Adaptation–not by sweeps alone. Nat Rev Genet. 2010;11:665–667. doi:10.1038/nrg2880

Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010;20:R208–R215. doi:10.1016/j.cub.2009.11.055

Ralph P, Coop G. Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics. 2010;186:647–668. doi:10.1534/genetics.110.119594

Sachdeva H. Reproductive isolation via polygenic local adaptation in sub-divided populations: effect of linkage disequilibria and drift. PLoS Genet. 2022;18:e1010297. doi:10.1371/journal.pgen.1010297

Sachdeva H, Barton NH. Introgression of a block of genome under infinitesimal selection. Genetics. 2018;209:1279–1303. doi:10.1534/genetics.118.301018

Sella G, Barton NH. Thinking about the evolution of complex traits in the era of genome-wide association studies. Annu Rev Genomics Hum Genet. 2019;20:461–493. doi:10.1146/annurev-genom-083115-022316

Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet. 2016;99:139–153. doi:10.1016/j.ajhg.2016.05.013

Stephan W. Selective sweeps. Genetics. 2019;211:5–13. doi:10.1534/genetics.118.301319

Stephan W, John S. Polygenic adaptation in a population of finite size. Entropy. 2020;22:907. doi:10.3390/e22080907

Stetter MG, Thornton K, Ross-Ibarra J. Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. PLoS Genet. 2018;14:1–24. doi:10.1371/journal.pgen.1007794

Stroustrup B. The C++ Programming Language. 4th ed. Boston (MA): Addison-Wesley Professional; 2013.

Thornton KR. Polygenic adaptation to an environmental shift: temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. Genetics. 2019;213:1513–1530. doi:10.1534/genetics.119.302662

Turelli M, Barton NH. Dynamics of polygenic characters under selection. Theor Popul Biol. 1990;38:1–57. doi:10.1016/0040-5809(90)90002-D

Uecker H, Hermisson J. On the fixation process of a beneficial mutation in a variable environment. Genetics. 2011;188:915–930. doi:10.1534/genetics.110.124297

Walsh B, Lynch M. Evolution and Selection of Quantitative Traits. Oxford (UK): Oxford University Press; 2018.

Wolfram Research I. Mathematica, Version 12.0. Champaign, IL, 2019.

Wright S. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. J Genet. 1935a;30:243–256. doi:10.1007/BF02982239

Wright S. Evolution in populations in approximate equilibrium. J Genet. 1935b;30:257–266. doi:10.1007/BF02982240

Yeaman S. Local adaptation by alleles of small effect. Am Nat. 2015; 186:S74–S89. doi:10.1086/682405

Yeaman S. Evolution of polygenic traits under global vs local adaptation. Genetics. 2022;220:134. doi:10.1093/genetics/iyab134

Yeaman S, Gerstein AC, Hodgins KA, Whitlock MC. Quantifying how constraints limit the diversity of viable routes to adaptation. PLoS Genet. 2018a;14:1–25. doi:10.1371/journal.pgen.1007717

Yeaman S, Gerstein AC, Hodgins KA, Whitlock MC. Quantifying how constraints limit the diversity of viable routes to adaptation. PLoS Genet. 2018b;14:e1007717. doi:10.1371/journal.pgen.1007717

Appendix

In this appendix, we present complementary figures to the figures in the main text with different choices of the parameter values. We also explain how our model extends to diploid genetics and present results for various levels of linkage in this case.

Adaptation with LE

Adaptive architecture for 100 loci

Figure A1 complements main text Fig. 4 (for a 10-locus trait) and shows the buildup of adaptive architecture for long-range adaptation of a trait with $L = 100$. Approximate convergence to a stable limit architecture can be observed for $\Theta_{bg} \leq 10$. For increasing distance of phenotypic adaptation, the results also show the broadening of the architecture (larger distances between successive size-ordered marginal distributions). As explained in the main text, this is due to the dwindling of adaptive material, corresponding to a reduction in the effective $\Theta_{bg,eff}$. The fit between LE simulation results and analytical curves is even better than in the main text Fig. 4. This is because the $\approx$10-fold smaller mutation rates *per locus* for $L = 100$ lead to a reduced effect of back mutation, which is ignored in the approximation.

Dynamics of summary statistics

Adding to Fig. 5 for a 10,000-locus trait in the main text, we show summary statistics of the adaptive architecture of QTs with 10 and 100 loci in Fig. A2. The adaptive scenarios correspond to the ones of Figs. 4 and A1, respectively. In addition, we also show a case of adaptation from only new mutation (*de novo*) for $L = 10$. Furthermore, we show how our results for pheno-time relate to the usual dynamics in real time.

The first row of Fig. A2 shows the scaled number of generations (on the *y*-axis, in units of $\Theta_{bg}\sigma\gamma^2$) required for the mean phenotype $\bar{Z}$ to reach passage points in units of mutational steps c_Z (*x*-axis). To derive a simple analytical prediction for this dependence, we use a result by Götsch and Bürger (2023) for an additive trait under constant directional selection and constant rates of beneficial mutation (assumption of an infinite-locus model). Translated to our model parameters, Götsch and Bürger (2023, Eq. 4.21) derive an expected change in the mean trait of

$$\Delta\bar{Z} \approx \Theta(L\gamma - \bar{Z})\sigma\gamma(Z^{new}_{opt} - \bar{Z})$$

for long-term adaptation. That is, to leading order in the selection coefficient, the average adaptive progress in the trait mean is simply given by the product of the selection strength and the population mutation rate. In our model, both the selection strength (which scales with the distance to the new optimum, $Z^{new}_{opt} - \bar{Z}$) and the rate of new beneficial mutations (which is proportional to the distance to the end of the phenotype range, $L\gamma - \bar{Z}$) are variable. This leads to a differential equation, which, in dimensionless

Table A1. Nomenclature. Overview of the main parameters used in the manuscript.

Symbol	Meaning	Formula
N_e	(effective) population size	
p_i	frequency of allele A_i at locus i	$0 \leq p_i \leq 1$
γ, μ_i	effect of allele A_i (all equal) and locus mutation rate (usually equal, $\mu_i = \mu$)	
L, d	total number L of loci in the genetic basis, d of which have $p_i(0) \leq 1/2$	
$Z, \bar{Z}$	trait value and its mean[a]	$\bar{Z} = \sum_{i=1}^{L} \gamma p_i$
$v_g, \kappa_3(Z)$	genetic variance and skew of Z	
Z_{opt}^0, Z_{opt}^{new}	initial trait optimum ($t < 0$), new trait optimum ($t \geq 0$)	$Z_{opt}^0 = d\gamma$
c_Z	stopping condition (sampling point) for $\bar{Z}$[b]	$\bar{Z} = Z_{opt}^0 + c_Z\gamma \leq Z_{opt}^{new}$
Θ_{bg}	population-scaled background mutation rate ($\mu = \mu_i$)	$\Theta_{bg} = 2N_e\mu(d-1)$
$W(Z), w(Z)$	Wrightian (discrete-time) and Malthusian (continuous-time, $\ln W(Z)$) fitness	$W(Z) = \exp\left[-\frac{\sigma}{2}(Z - Z_{opt})^2\right]$
$\sigma\gamma^2$	selection coefficient against minority-allele in SGV	$s_{eq} = \sigma\gamma^2 = s(t), \, t < 0$

[a]We denote means of x across individuals within a population as $\bar{x}$ and across evolutionary replicates as $\langle x \rangle$.
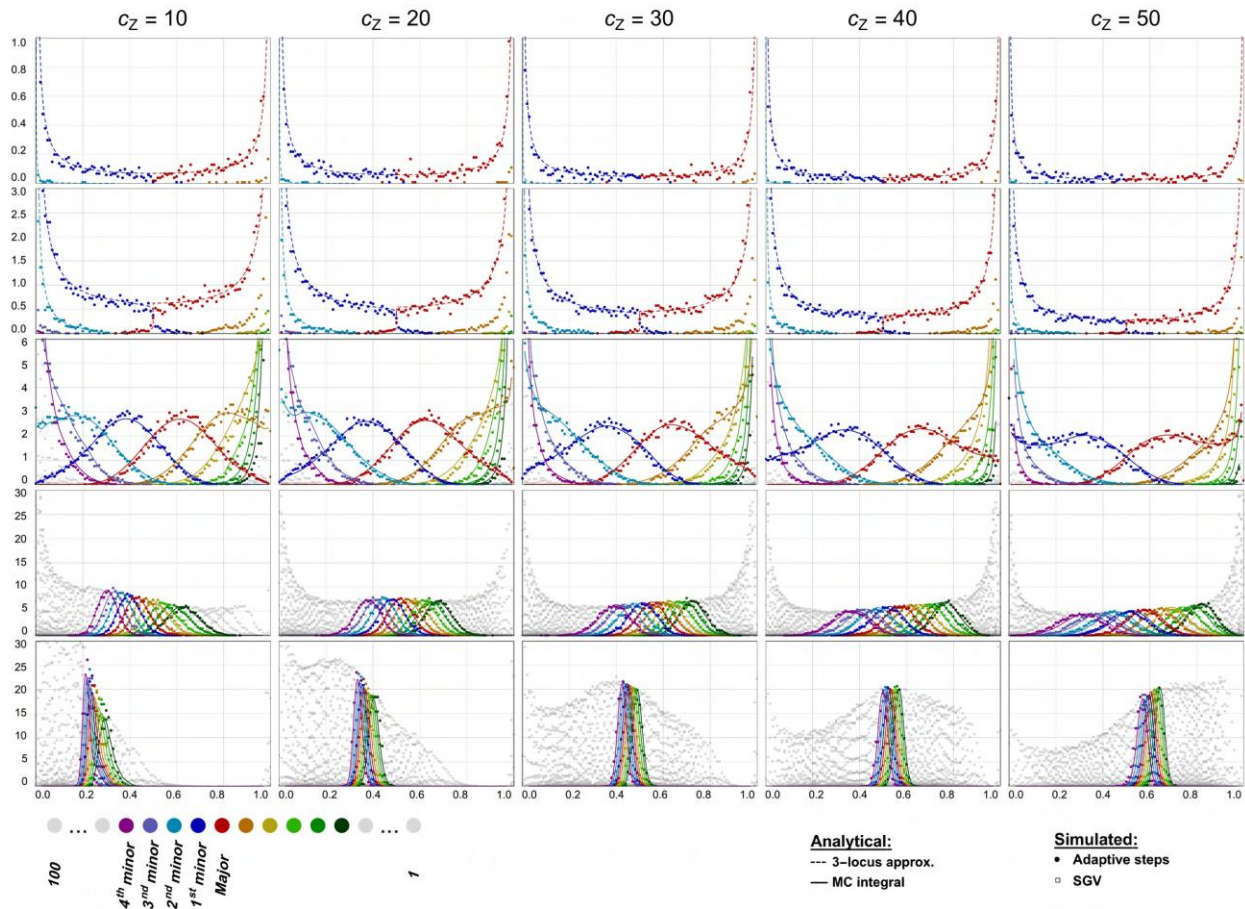[b]In simulations, we observe the joint allele frequency distribution at first-hitting time t, such that $\bar{Z}(t) \geq Z_{opt}^0 + c_Z\gamma$. Accordingly, t can vary between realizations.

Fig. A1. Dynamics of the adaptive process for a QT with 100 loci. Successive snapshots (in intervals of 10 mutational steps) of the ordered marginal allele-frequency distributions over the course of adaptation of a QT with 100 loci. Equilibrated populations evolve from an initial optimum $Z_{opt}^0 = 20\gamma$ toward a new optimum $Z_{opt}^{new} = 80\gamma$. LE simulation densities (points) are accurately predicted by the analytical model. Dashed curves (for $\Theta_{bg} = 0.01$ and 0.1) show analytical three-locus approximations accounting for the progressive reduction in the effective $\Theta_{bg,eff}$ with c_Z (see main text). Solid lines correspond to results of the 100-locus Monte-Carlo method, including deleterious variation (cf., the Supplementary File S1). The distributions of the loci with the ($Z_{opt}^0 + c_Z - 5$)th (dark green) to ($Z_{opt}^0 + c_Z + 4$)th (magenta) largest frequency are shown in bright colors. All others are shown in gray. We include a step-change in the strength of stabilizing selection from $N_e\sigma(t)\gamma^2 = 100$ for $t < 0$ to 10 for $t \geq 0$ to avoid artifacts due to excessively strong directional selection for $t \geq 0$. $N_e = 10,000$, 10,000 replicates. Note the differences in scaling of the y-axes.

quantities, can be written as

$$\frac{dt}{dc_Z} = \frac{d-1}{(d-c_Z)(Z_{opt}^{new}/\gamma - L + d - c_Z)}$$

where time t is measured in units of $[\Theta_{bg}\sigma\gamma^2]^{-1}$. The solution is

$$t(c_Z) = C + \frac{(d-1)}{L - Z_{opt}^{new}/\gamma} \log\left[\frac{d - c_Z}{Z_{opt}^{new}/\gamma - L + d - c_Z}\right]. \quad (A1)$$
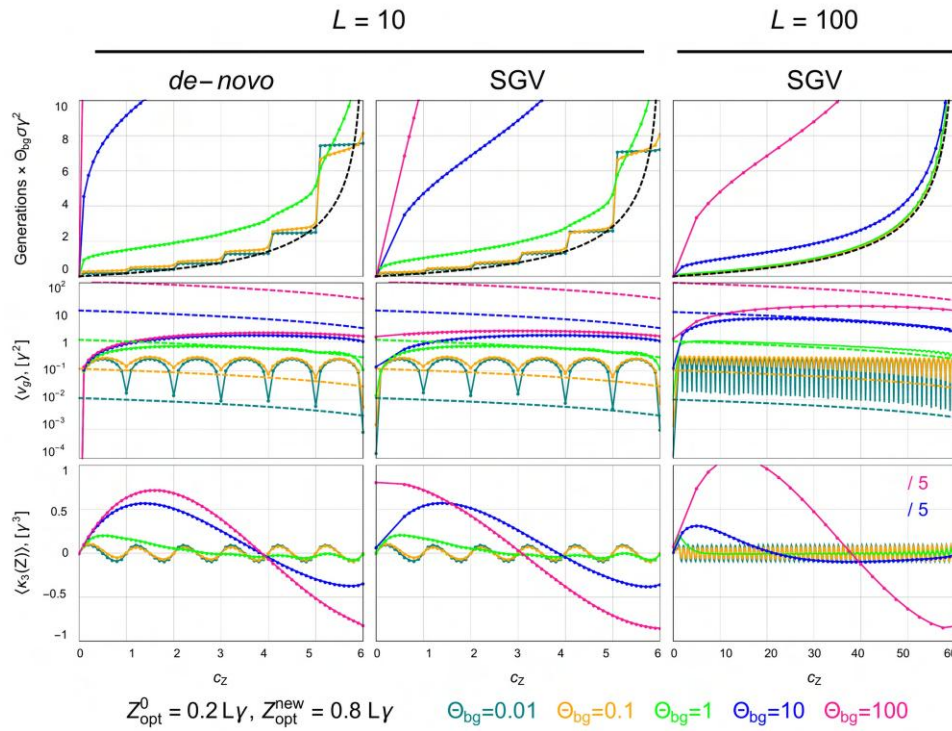
Fig. A2. Genetic variance and skew for QTs with 10 and 100 loci. QTs with 10 loci (two left columns: without/with SGV) and 100 loci (right: with SGV) adapt from $Z_{opt}^0 = 0.2L\gamma$ to $Z_{opt}^{new} = 0.8L\gamma$ (6 mutational steps for $L = 10$ and 60 steps for $L = 100$). The top row shows the average number of generations to reach passage points of phenotypic adaptation (real time in terms of pheno-time). The dashed black line indicates an approximate theoretical expectation (for all Θ_{bg}, see Equation A1). The middle and bottom row show the dynamics of the genetic variance (with expectations as dashed lines, see Equation A2) and the skew in pheno-time. $N_e = 10{,}000$, $N_e\sigma\gamma^2 = 100$ (switch to $N_e\sigma\gamma^2 = 10$ for $t > 0$ for 100 loci, cf., Fig. A1), 10,000 replicates (1,000 and 5,000 for $\Theta_{bg} = 0.01$, respectively, 0.1 in first row). The insets "/5" indicate that the corresponding lines are 1/5th of the actual data.
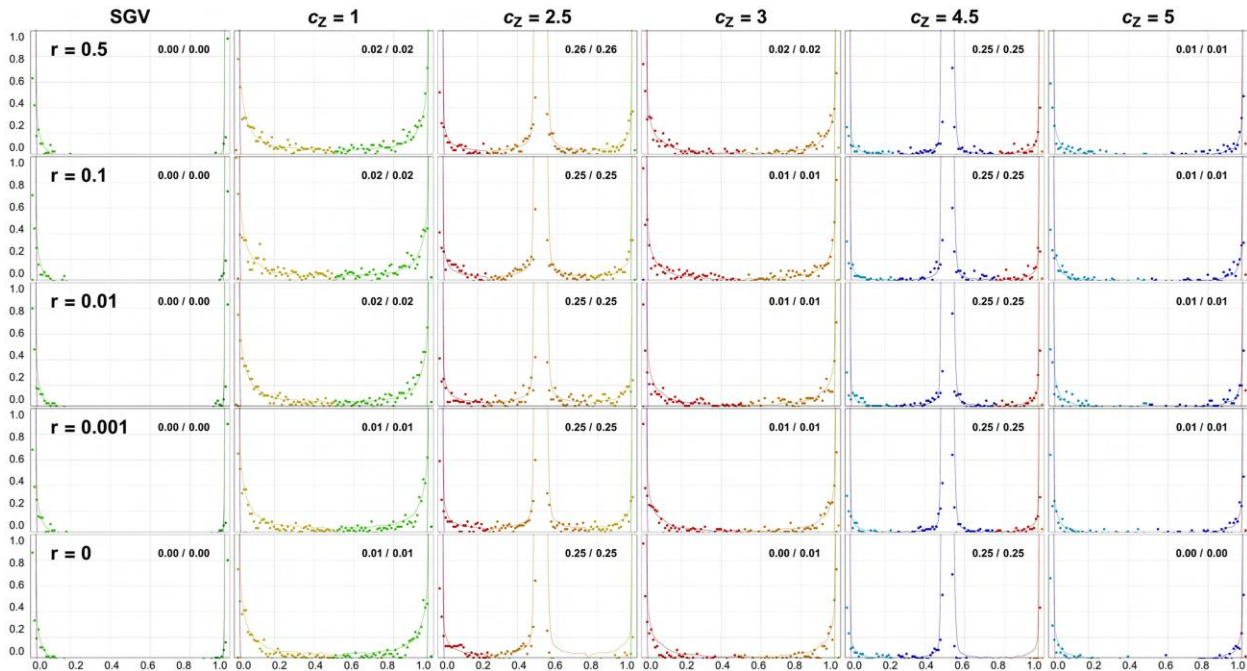

Fig. A3. Effect of linkage on adaptive architecture of a QT (haploids, $\Theta_{bg} = 0.01$). With very low background mutation rates, only complete linkage causes apparent changes in the adaptive architectures. Other parameters are as in Fig. 6.
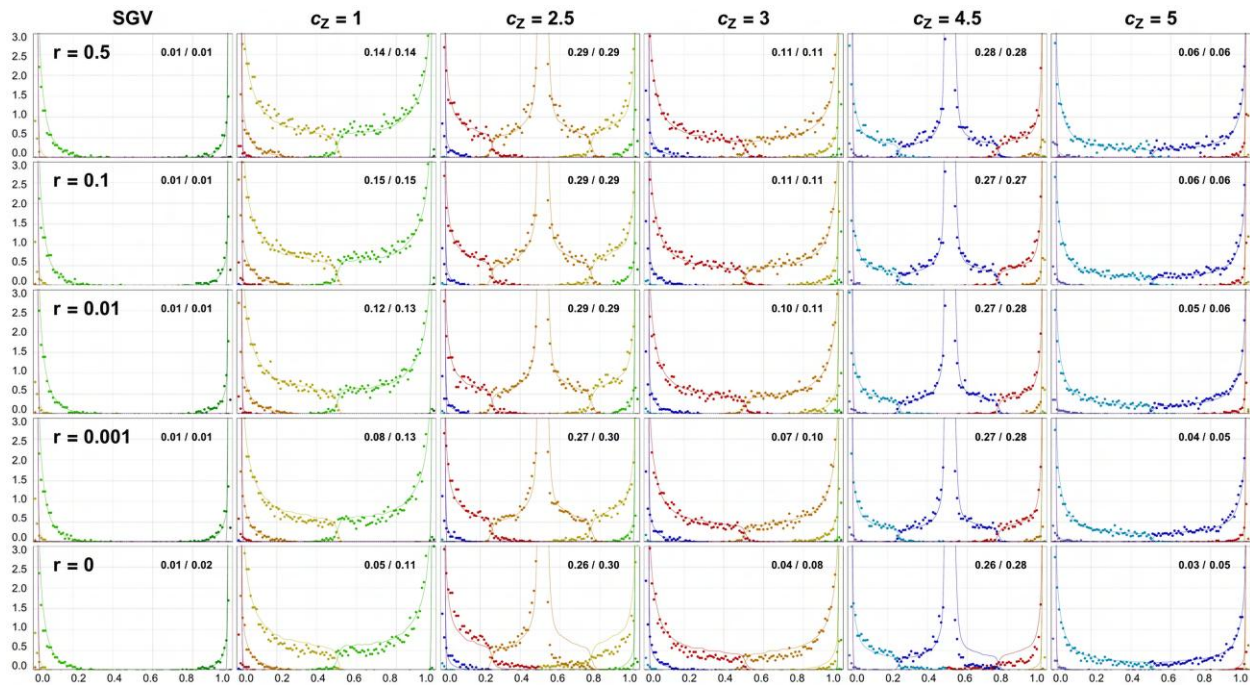
Fig. A4. Effect of linkage on adaptive architecture of a QT (haploids, $\Theta_{bg} = 0.1$). With low background mutation rates, the adaptive architectures are still highly robust, even under tight linkage. Other parameters are as in Fig. 6.
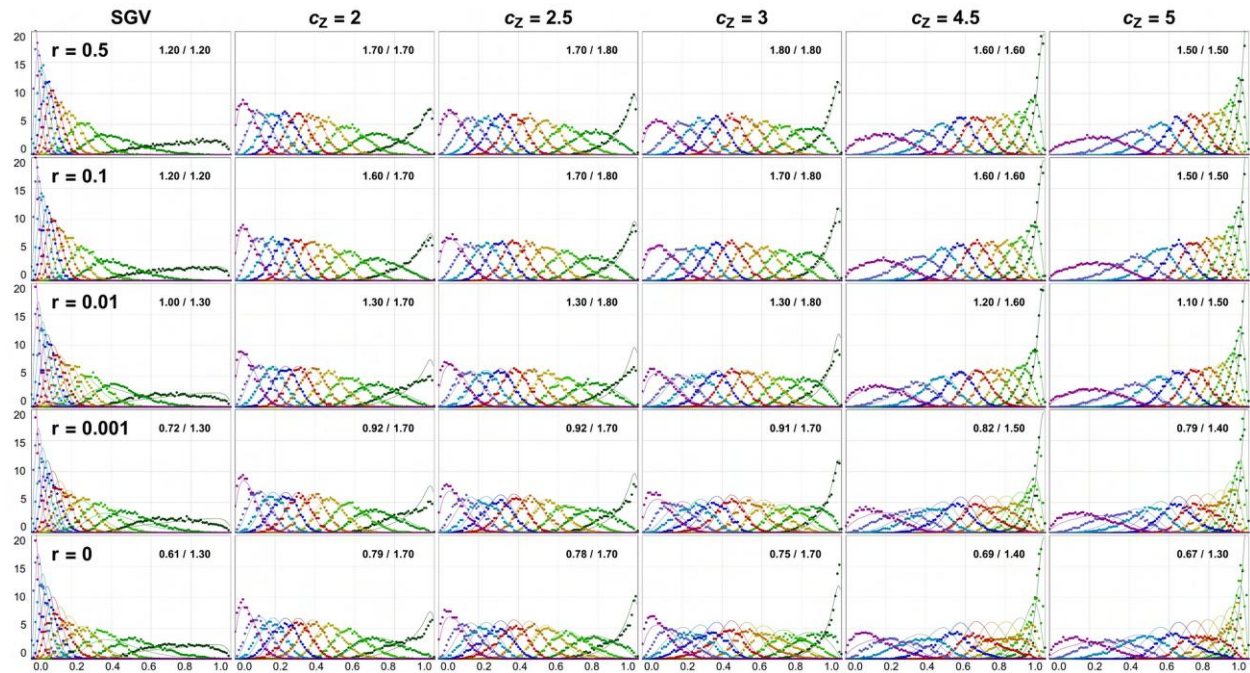

Fig. A5. Effect of linkage on adaptive architecture of a QT (haploids, $\Theta_{bg} = 10$). With high background mutation rates, we can detect mild alterations under tight and complete linkage. Other parameters are as in Fig. 6.

This curve is displayed as a black dashed line in the first row of Fig. A2 (with C chosen such that it passes through $(0, 0)$). Similarly, the analytical expectations (dashed curves) for the equilibrium variance in the second row of Fig. A2 are based on Götsch and Bürger (2023, Eq. 4.24). In terms of our model parameters and with a variable mutation rate, this reads,

$$\langle v_g \rangle (c_Z) = \Theta_{bg} \frac{d - c_Z}{d - 1}. \tag{A2}$$
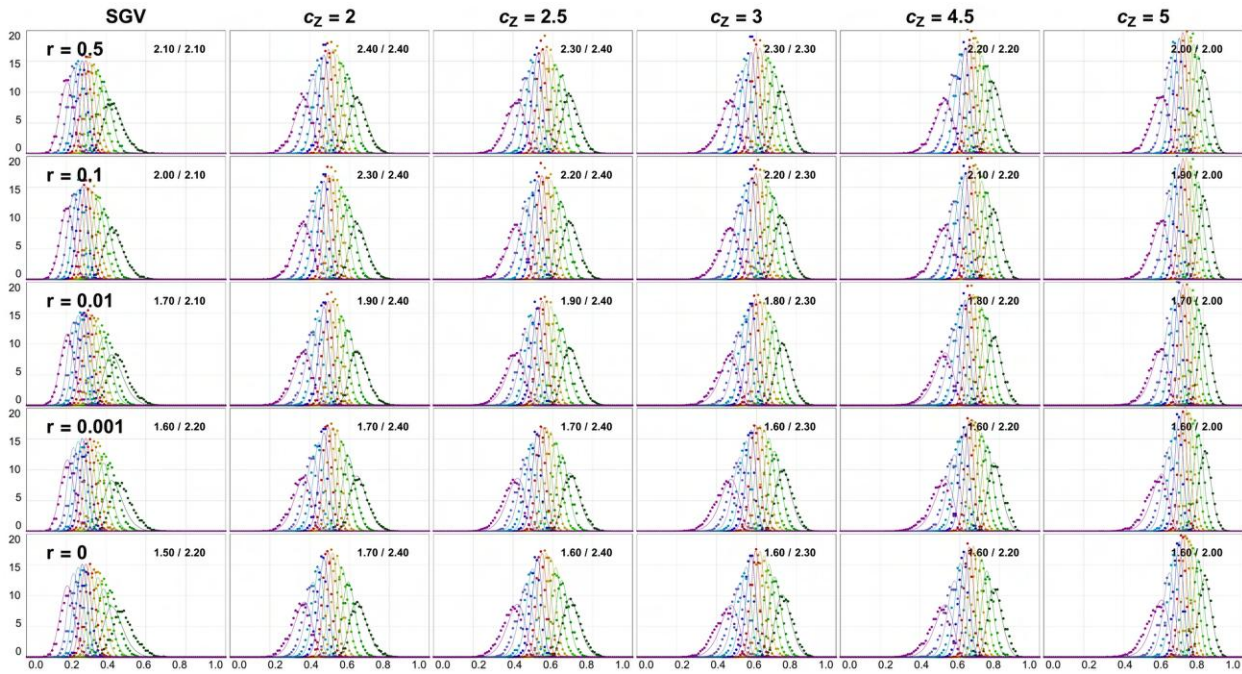
We observe the following:

Fig. A6. Effect of linkage on adaptive architecture of a QT (haploids, $\Theta_{bg} = 100$). With very high background mutation rates, linkage has no apparent effect on the adaptive architectures. Other parameters are as in Fig. 6.
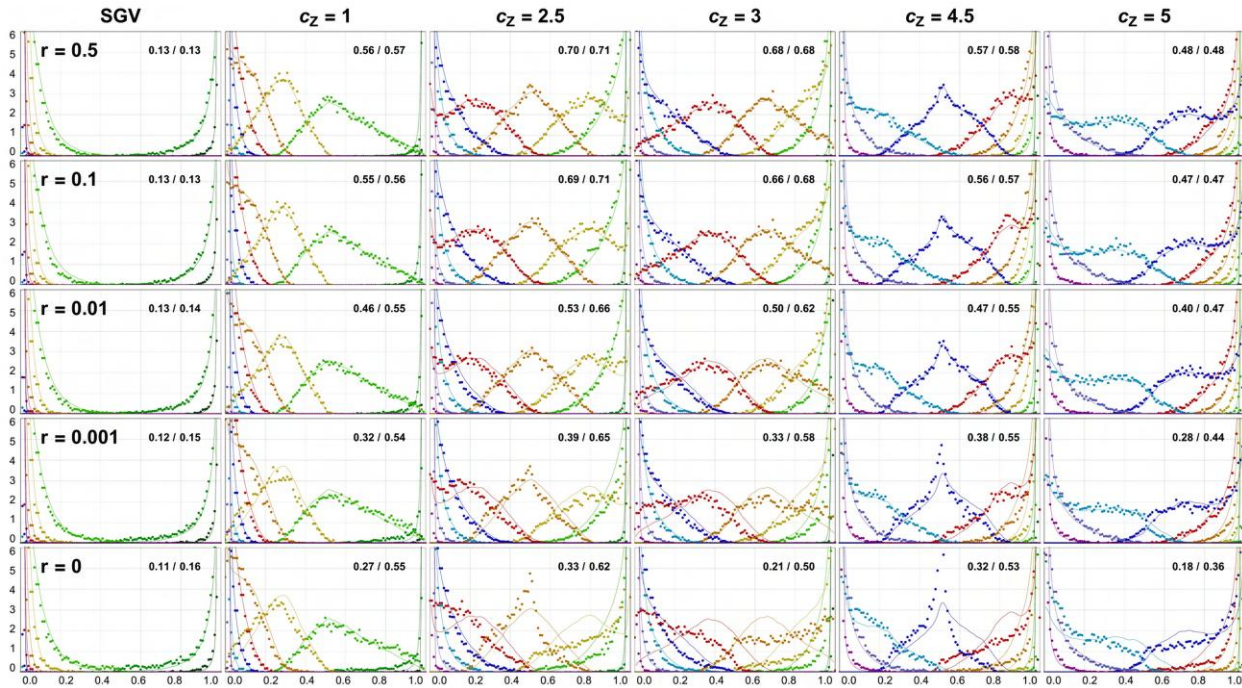

Fig. A7. Effect of linkage on adaptive architecture of a QT (diploids, $\Theta_{bg} = 1$). The shape of the marginal distributions is virtually identical for diploids without dominance ($N_e^{dip} = 500$) and haploids ($N_e = 1,000$, Fig. 6). Importantly, c_Z corresponds to steps of size $2 \cdot \gamma/\sqrt{2} = \sqrt{2}\gamma$. All other parameters and the curves (LE) are the same as in Fig. 6.

- The analytical prediction (Equation A1) (which assumes an infinite-locus model) matches the simulation data in the top row whenever the background mutation rate Θ_{bg} is small enough that recurrent mutation at the same locus is unlikely. For small L and/or high Θ_{bg}, new mutations frequently occur at loci where mutant alleles are already established, which results in slower dynamics (steeper increase).

- For the variance (second row), we see that the estimate for the average variance in real time (Equation A2) is consistently lower than the simulation data for small $\Theta_{bg} < 1$. The reason (also observed by Götsch and Bürger 2023) is that sampling at fixed phenotype values does not represent time averages. Indeed, for very small Θ_{bg}, there is no segregating variation for the trait most of the time, which is only possible for sampling at integer values of c_Z (where the approximation matches more closely). For higher Θ_{bg}, the pheno-time values no longer oscillate and approach the time average—until saturation effects due to recurrent mutation at the same loci lead to lower variances in a finite-locus model compared to an infinite-locus model for large Θ_{bg}. See also Fig. 5 for $L = 10{,}000$ in the main text, where the real time averages from the infinite-locus model ($v_g \approx \Theta_{bg}$) are approximately met once the oscillations become small for $\Theta_{bg} \geq 1$.
- The initial variance in SGV is smaller than the equilibrium long-term variance in all cases shown, even for $\Theta_{bg} = 100$. Until the dynamic equilibrium is reached, the rate of adaptation is slower (which leads to a shift to larger y-values in the top row, see $\Theta_{bg} = 10$ for $L = 100$ and $\Theta_{bg} = 1$ for $L = 10$). Since initial variation is absent in the *de novo* scenario, the rate of adaptation is even slower (compare first and second column, top row). However, differences in variance and skew due to the starting conditions vanish quickly ($\lesssim 1$ mutational step for $\Theta_{bg} \leq 10$).
- Positive initial skew (bottom row) results from the asymmetric starting conditions (more mutations to increase the trait than to decrease it). Negative genetic skew indicates a dwindling supply of fresh loci, which is either transient (yielding oscillations) for low effective $\Theta_{bg,eff}$, or persistent when the finite trait basis has been exhausted.

Adaptation with linkage
Haploids with linkage
We extend Fig. 6 to other settings of Θ_{bg}: 0.01 (Fig. A3), 0.1 (Fig. A4), 10 (Fig. A5) and 100 (Fig. A6). The effect of linkage is strongest for intermediate values of Θ_{bg}:

- For low mutation rates, tightly linked loci are rarely polymorphic simultaneously. For $\Theta_{bg} = 0.01$ (Fig. A3), the adaptive architecture remains unaltered even for $r = 0$, while for $\Theta_{bg} = 0.1$ (Fig. A4) slight changes are visible.
- While the effect of linkage for $\Theta_{bg} = 10$ (Fig. A5) is similar compared to $\Theta_{bg} = 1$ (Fig. 6), all differences vanish for $\Theta_{bg} = 100$ (Fig. A6). With very high mutation rates, the randomizing effect of recurrent, bi-directional mutation eliminates LD.

For large $\Theta_{bg} \geq 10$, the first phenotypic checkpoint is placed at $c_Z = 2$ since the trait mean in the SGV exceeds $c_Z = 1$ in some replicates (with asymmetric starting condition, mutation biases the trait mean upwards).

Diploids with linkage
For simplicity, we have derived and presented all our results so far for a haploid population. Here we show how the formalism extends to diploids without dominance. To match the haploid dynamics, we consider a diploid model with the same number of gametes, $N_e^{dip} = N_e/2$ and maintain the same mutation rates μ per locus, such that $\Theta = 2N_e\mu = 4N_e^{dip}\mu$. We assume the same number of loci L underlying the trait and the same proportion of loci with a_i-majority vs. A_i-majority allele at the initial optimum Z_{opt}^0 (constant d). We thus have $\Theta_{bg} = (d-1)2N_e\mu = (d-1)4N_e^{dip}\mu$. To maintain the same level of SGV (with expected frequency $2\mu/(\sigma\gamma^2)$), we also hold the selection strength $\sigma\gamma^2/2$ prior to the environmental change constant (compare the first columns, "SGV," between Figs. 6 and A7).

Lastly, we also want to make the genic, $v_g^{LE} = \sum_i p_i(1-p_i)[\gamma^2]$, and genetic variance, $v_g = (\sum_Z Z^2 q_Z - (\sum_Z Z q_Z)^2)$ (where q_Z is the frequency of phenotype Z in the population), comparable between haploids and diploids (see insets in Figs. 6 and A7). To that end, we need to set $\gamma^{dip} = \gamma/\sqrt{2}$ because the trait is a sum over the contributions of $2L = 20$ alleles in the diploid case, instead of $L = 10$ for haploids. Accordingly, the phenotype range is $0 \leq Z \leq 2L\gamma/\sqrt{2} = \sqrt{2}L\gamma$ and the optimum shifts by 6 mutational steps of size $\sqrt{2}\gamma$ from $Z_{opt}^0 = 2 \cdot \sqrt{2}\gamma$ to $Z_{opt}^{new} = 8 \cdot \sqrt{2}\gamma$.

Editor: K. Jain