# Sensitivity of quantitative traits to mutational effects and number of loci

Joshua G. Schraiber [a],[*], Michael J. Landis [b]

[a] Department of Genome Sciences, University of Washington, Seattle, WA 98195, United States
[b] Department of Integrative Biology, University of California, Berkeley, CA 94720, United States

## ABSTRACT

When models of quantitative genetic variation are built from population genetic first principles, several assumptions are often made. One of the most important assumptions is that traits are controlled by many genes of small effect. This leads to a prediction of a Gaussian trait distribution in the population, via the Central Limit Theorem. Since these biological assumptions are often unknown or untrue, we characterized how finite numbers of loci or large mutational effects can impact the sampling distribution of a quantitative trait. To do so, we developed a neutral coalescent-based framework, allowing us to gain a detailed understanding of how number of loci and the underlying mutational model impacts the distribution of a quantitative trait. Through both analytical theory and simulation we found the normality assumption was highly sensitive to the details of the mutational process, with the greatest discrepancies arising when the number of loci was small or the mutational kernel was heavy-tailed. In particular, skewed mutational effects will produce skewed trait distributions and fat-tailed mutational kernels result in multimodal sampling distributions, even for traits controlled by a large number of loci. Since selection models and robust neutral models may produce qualitatively similar sampling distributions, we advise extra caution should be taken when interpreting model-based results for poorly understood systems of quantitative traits.

## 1. Introduction

Questions about the distribution of traits that vary continuously in populations were critical in motivating early evolutionary biologists. The earliest studies of quantitative trait variation relied on phenomenological models, because the underlying nature of heritable variation was not yet well understood (Galton, 1883, 1889; Pearson, 1894, 1895). Despite the rediscovery of the work of Mendel (1866), researchers studying continuous variation in natural populations were initially skeptical that Mendel's laws could explain what they observed (Weldon, 1902; Pearson, 1904). These views were reconciled when Fisher (1918) showed that the observations of correlation and variation between phenotypes in natural populations could be explained by a model in which many genes made small contributions to the phenotype of an individual.

The insights of Fisher (1918) made it possible to build models of quantitative trait evolution from population genetic first principles. Early work focused primarily on the interplay between mutation and natural selection in the maintenance of quantitative genetic variation in natural populations, while typically ignoring the effects of genetic drift (Fisher, 1930; Haldane, 1954; Latter, 1960; Kimura, 1965).

However, genetic drift plays an important role in shaping variation in natural populations. While earlier work assumed that a finite number of alleles control quantitative genetic variation (e.g. Latter (1970)), Lande (1976) used the continuum-of-alleles model proposed by Kimura (1965) to model the impact of genetic drift on differentiation within and between populations. A key assumption of Lande's models is that the additive genetic variance in a trait is constant over time. In fact, in finite populations the genetic variance itself is random; at equilibrium, there are still stochastic fluctuations around the deterministic value assumed by Lande, even if none of the underlying genetic architecture changes (Bürger and Lande, 1994).

Several later papers explored more detailed models to understand how genetic variance changes through time due to the joint effects of mutation and drift (e.g. Chakraborty and Nei (1982)). Lynch and Hill (1986) undertook an extremely thorough analysis

* Corresponding author.
*E-mail address:* schraib@uw.edu (J.G. Schraiber).

of the evolution of neutral quantitative traits. They analyzed the moments (e.g. mean and variance) of trait distributions that arise due to mutation and genetic drift and provided several quantities that can be used to interpret variation within and between species and analyze mutation accumulation experiments.

Much of this earlier work has made several simplifying assumptions about the distribution of mutational effects and the genetic architecture of the traits in question. For instance, Lynch and Hill (1986), despite analyzing quite general models of dominance and epistasis, ignored the impact of heavy tailed or skewed mutational effects. While, in many cases, such properties of the mutational effect distribution are not expected to have an impact if a large number of genes determine the phenotype in question, it is unknown what impact they may have when only a small number of genes determine the genetic architecture of the trait. Moreover, when mutational effects display "power-law" or "fat-tailed" behavior, the impact of the details of the mutational effects may persist even in the so-called infinitesimal limit of a large number of loci with small effects. Finally, mutation accumulation experiments have produced skewed and/or leptokurtic samples of quantitative traits (Mackay et al., 1992), which is a direct motivation to relax assumptions on the mutational effects distribution.

Such deviations that stem from the violations of common modeling assumptions have the potential to influence our understanding of variation in natural populations. For instance, leptokurtic trait distributions may be a signal of some kind of diversifying selection (Kopp and Hermisson, 2006) but are also possible under neutrality when the number of loci governing a trait is small. Similarly, multimodal trait distributions may reflect some kind of underlying selective process (Doebeli et al., 2007) but may also be due to rare mutations of large effect.

We have two main goals in this work. Primarily, we want to assess the impact of violations of common assumptions on properties of the sampling distribution of a quantitative trait (e.g. variance, kurtosis, modality). Secondly, we believe that the formalism that we present here can be useful in a variety of situations in quantitative trait evolution, particularly in the development of robust null models for detecting selection at microevolutionary time scales. To this end, we introduce a novel framework for computing sampling distributions of quantitative traits. Our framework builds upon the coalescent approach of Whitlock (1999), but allows us to recover the full sampling distribution, instead of merely its moments.

First, we outline the biological model and explain how we can compute quantities of interest using a formalism based on characteristic functions. We then use this approach to compute the sample central moments. While much previous work focuses on only the first two central moments (mean and variance), we are able to compute arbitrarily high central moments, which are related to properties such as skewness and kurtosis. By doing so, we are able to determine the regime in which the details of the mutational effect distribution are visible in a sample from a natural population. Additionally, we explore the convergence to the infinitesimal limit and find that when "fat-tailed" effects are present, traditional theory based on the assumption of normality can lead to misleading predictions about phenotypic variation.

## 2. Model

The mechanistic model we construct has two components: a coalescent process, and a genetic mutational process that acts upon the controlling quantitative trait loci by sampling effect sizes from a mutational kernel. Together, these processes generate the values of quantitative traits sampled from the study population while explicitly modeling their shared genetic ancestry. Although we opt for simple model components during this exposition, the model generally supports more realistic and complex extensions, such as population structure and epistasis.

We assume that we sample $n$ haploid individuals from a randomly mating population of size $N$. Initially, we consider a trait governed by a single locus and we will later extend the theory to traits governed by multiple loci. Let $\mu$ be the mutation rate per generation at the locus, and $\theta = 2N\mu$ be the coalescent-scaled mutation rate. We model mutation as a process by which a new mutant adds an independent and identically distributed random effect to the ancestral state. Note that when the distribution of random effects is continuous, this corresponds to the Kimura (1965) continuum-of-alleles model. However, it is also possible for the effect distribution to be discrete, similar to the discrete model of Chakraborty and Nei (1982). While this model does not capture the impact of a biallelic locus with exactly two effects, the following theory could easily be modified to analyze that case.

Fig. 1 shows one realization of both the coalescent and mutational processes for a sample of size 5. Given the phenotype at the root of the tree and the locations and effects of each mutation on the tree, the phenotypes at the tips are determined by adding mutant effects from the root to tip. To specify the root, we can assume without loss of generality that the ancestral phenotype for the entire population has a value 0 (this is similar to the common assumption in quantitative genetics literature that the ancestral state at each locus can be assigned a value of 0).

This mutational process can be described as a compound Poisson process (see also Khaitovich et al. (2005b); Chaix et al. (2008); Landis et al. (2013) for compound Poisson processes in a phylogenetic context). To ensure that this paper is self contained, we briefly review relevant facts about compound Poisson processes in Appendix A.1.

In the following, we ignore the impact of non-genetic variation and focus on the breeding value of individuals, i.e. the average phenotype of an individual harboring a given set of mutations.

## 3. Results

### 3.1. Computing the characteristic function of a sample

In many analyses, the object of interest is the joint probability of the data. If we let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be the vector representing the values of the quantitative trait observed in a sample of $n$ individuals, we denote the joint probability of the data as $p(x_1, x_2, \ldots, x_n)$. Note that, in general, $X_i$ and $X_j$ are correlated due to shared ancestry, and that $p$ must be computed by integrating over all mutational histories consistent with the data. Hence, computing $p$ directly is extremely difficult.

Instead, we compute the characteristic function of $\mathbf{X}$. For a one-dimensional random variable, $X$, the characteristic function is defined as $\mathbb{E}(e^{ikX})$ where $i$ is the imaginary unit, and $k$ is a dummy variable. Generalizing this definition to an $n$-dimensional random variable, we are interested in computing

$$\lambda_n(\mathbf{k}) = \mathbb{E}(e^{i\mathbf{k}^T\mathbf{X}})$$
$$= \mathbb{E}(e^{i(k_1X_1+k_2X_2+\cdots+k_nX_n)})$$

where $\mathbf{k} = (k_1, k_2, \ldots, k_n)$ is a vector of dummy variables. Like a probability density function, the characteristic function of $\mathbf{X}$ contains all the information about the distribution of $\mathbf{X}$. Moreover, computing moments of $\mathbf{X}$ is reduced to calculating derivatives of the characteristic function, which will prove useful in the following.

We calculate the characteristic function of $\mathbf{X}$ in two parts. First, we compute a recursive formula for $\phi_n$, the characteristic function given that ancestral phenotype of the *sample* is equal to 0. Then, we compute $\rho_n$, the characteristic function of the ancestral
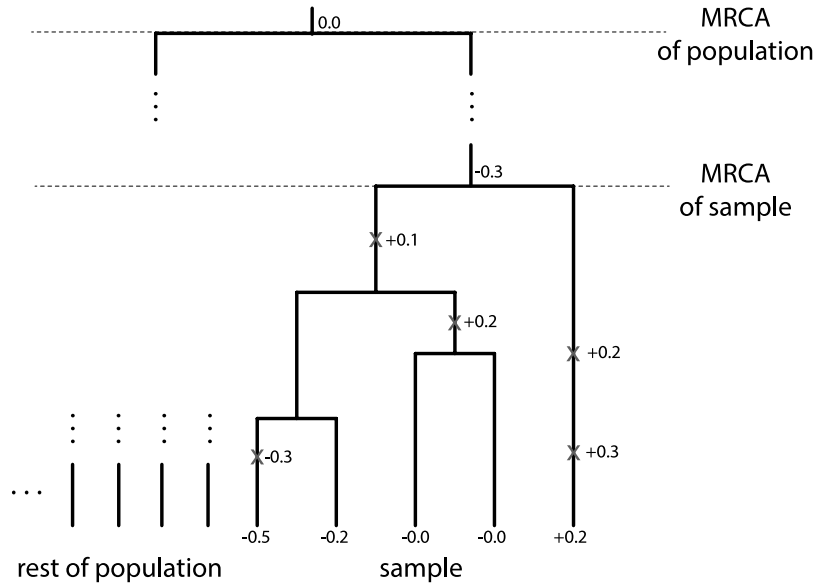
**Fig. 1.** Example realization of coalescent process for a sample of size 5. Mutations (marked as light gray X's), are placed upon the genealogy representing each individual in the population. Effects of each mutation are drawn from a probability distribution and are added along each branch length. The model is specified such that the most recent common ancestor (MRCA) of the population has phenotype 0.0, while the MRCA of the sample may have a phenotype different from zero, due to mutations that accumulate between the MRCA of the sample and the MRCA of the population.

phenotype of the sample, assuming that the ancestral phenotype of the *population* is equal to 0. As we show in Appendix A.2, we can then multiply these characteristic functions to obtain the characteristic function of **X**.

We use a backward–forward argument to compute the recursive formula, first conditioning on the state when the first pair of lineages coalesce (backward in time) and then integrating (forward in time) to obtain the characteristic function for a sample of size $n$, $\phi_n$. This results in

$$\phi_n(\mathbf{k}) = \frac{2}{n(n-1) - \theta\left(\sum\limits_{u=1}^{n} \psi(k_u) - n\right)} \sum_{u<v} \phi_{n-1}(\mathbf{k}^{(u,v)}) \qquad (1)$$

where $\mathbf{k}^{(u,v)}$ is the vector of length $n-1$ made by removing $k_u$ and $k_v$ and adding $k_u + k_v$ to the vector of dummy variables and $\psi(\cdot)$ is the characteristic function of the mutational effect distribution.

This equation has a straight-forward interpretation. The characteristic function for a sample of size $n$, $\phi_n$, is simply the characteristic function for a sample of size $n-1$, $\phi_{n-1}$, averaged over all possible pairs that could coalesce first, multiplied by the characteristic function for the amount of trait change that occurs more recently than the first coalescent. The multiplication comes from the fact that the characteristic function of a sum of independent random variables is the product of the characteristic functions of those random variables. We prove this result in Appendix A.3.

In Appendix A.4, we also show that the characteristic function for the phenotype at the root of the sample is

$$\rho_n(k) = n!(n-1) \sum_{u=1}^{\infty} \prod_{v=2}^{u} \frac{v(v-1)}{v(v-1) - \theta(\psi(k)-1)} \frac{u!}{(n+u)!}. \qquad (2)$$

Intuitively, this equation arises by conditioning on whether $u$ lineages are left in the *population* when the sample reaches its common ancestor and then averaging over the (random) time between when the individuals in the sample coalesce and when everyone in the population coalesces.

Hence, the characteristic function for a sample of size $n$ is

$$\lambda_n(\mathbf{k}) = \rho_n(k_1 + k_2 + \cdots + k_n)\phi_n(\mathbf{k}).$$

### 3.2. Sampling traits controlled by a small number of loci

It is common practice in both theoretical and applied quantitative genetics to summarize information about the phenotypic distribution within a population by computing central moments. However, care must be taken when interpreting theoretical predictions about central moments estimated from a sample. This is because the phenotypes in the sample are not independent, but instead correlated due to their shared genealogical history. Hence, in any *particular* population, an estimate of a central moment may deviate from its expected value, even as the number of individuals sampled grows to infinity (Aldous, 1985).

With this caveat in mind, we computed the first four expected central moments for a sample of phenotypes taken from this model (see Appendix A.5 for details). They are

$$\mathbb{E}(h_2) = \frac{1}{2}\theta L m_2$$

$$\mathbb{E}(h_3) = \frac{1}{6}\theta L m_3 \qquad (3)$$

$$\mathbb{E}(h_4) = \frac{3}{4}\theta^2 L^2 m_2^2 + \frac{1}{4}\theta L(2\theta m_2^2 + m_4),$$

where $h_k$ is the unique minimum variance unbiased estimator of the $k$th central moment (Halmos et al., 1946), $m_k$ is the $k$th moment of the mutational effect distribution (which can be calculated by differentiating the characteristic function of the effect distribution, $\psi$) and $L$ is the number of loci that influence the trait.

These equations reveal that it may be possible to construct method-of-moments estimators for the moments of the mutation effect distribution and/or the number of loci that govern a trait.

### 3.3. "Infinitesimal" limits for large numbers of loci

Many traits are assumed to be governed by a large number of loci, each individually of small effect. This is known as an infinitesimal model (Falconer and Mackay, 1996). Typically, the sampling distribution in the infinitesimal limit is assumed to be Gaussian, by appealing to the central limit theorem. Here, we find that under certain circumstances traits may not be normally distributed, even in the limit.
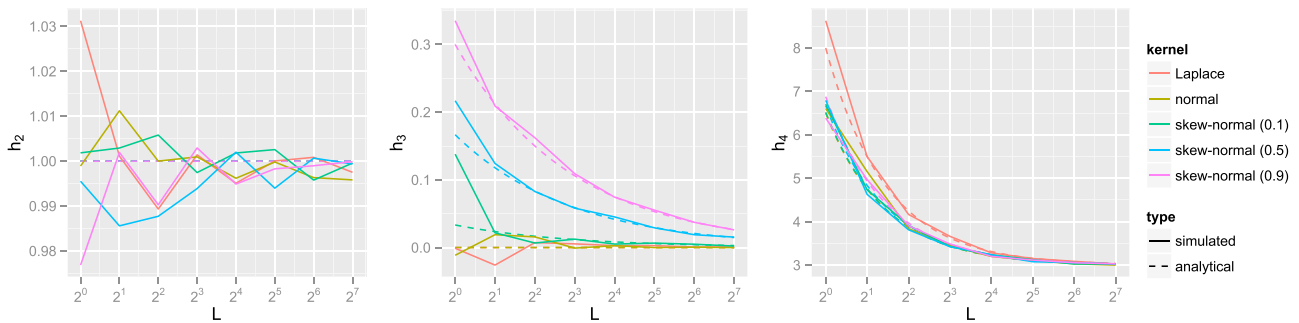
**Fig. 2.** Central moments. From left to right, the panels correspond to the central moments, $h_2$, $h_3$, and $h_4$, respectively, for the sampling distributions evolving under various mutational kernels. Data were simulated for 1024 sampled individuals and 2000 replicates for eight values of $L$, the number of loci. Colors distinguish the mutational kernel and relevant kernel parameters (if any). Solid lines correspond to moment values computed from the simulated data. Dashed lines correspond to the expected moment values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To obtain a non-trivial limit, we must assume that as the number of loci controlling the trait increases, the effect of each individual locus decreases in an appropriate way. Then, computing the characteristic function for a trait governed by a large number of independent loci is simple due to the fact the characteristic function of the sum of independent random variables is the product of their characteristic functions. Thus, assuming that each locus has the same effect distribution (this assumption can be relaxed relatively easily) the characteristic function of the limiting distribution of trait values is given by

$$\Lambda_n(\mathbf{k}) = \lim_{L \to \infty} \lambda_n(\mathbf{k})^L$$
$$= \lim_{L \to \infty} \rho_n(k_1 + k_2 + \cdots + k_n)^L \phi_n(\mathbf{k})^L$$
$$= R_n(k_1 + k_2 + \cdots + k_n)\Phi_n(\mathbf{k})$$

where $R_n$ is the limiting distribution of the phenotype at the root and $\Phi_n$ is the limiting distribution of the evolution in the sample.

In Appendices A.6 and A.7, we show that mutation effect distributions with power law behavior instead converge to a limiting *stable* distribution. A random variable $X$ is said to have a power law distribution if $P(X > x) \sim \kappa x^{-\alpha}$ for large $x$, some $\kappa > 0$ and some $\alpha \in [0, 2)$. In this limit, individuals with shared genealogy may still have highly correlated phenotypes, due to rare mutations of large effect.

On the other hand, all mutation effect distributions without power law behavior converge to a Gaussian limit, due to the central limit theorem. To obtain the non-trivial limit, we assume that the variance of mutational effects per locus, $\tau^2$, goes to zero as the number of loci, $L$, goes to infinity in such a way that the total variance summed over all loci converges to a constant, i.e. $L\tau^2 \to \sigma^2$. In Appendix A.8, we show that samples taken from a population in this limit can be represented as a sample from a normal distribution with a random mean. In particular,

$$X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(M, \frac{1}{2}\theta\sigma^2\right)$$
$$M \sim \mathcal{N}\left(0, \frac{1}{2}\theta\sigma^2\right),$$

where $\mathcal{N}(m, s^2)$ represents a normal distribution with mean $m$ and variance $s^2$.

### 3.4. Simulation

To gain a more intuitive picture of how trait distributions change due to the underlying mutational kernel and the number of QTL, we conducted simulations. First, we used ms (Hudson, 2002) to generate coalescent genealogies and then generated and mapped mutational effects using custom scripts in R (R Core Team,

2013). Specifically, for each segregating site generated in ms, we drew a mutational effect from the appropriate distribution and added that effect to every individual who had a derived allele at that segregating site. We held the sample variance constant across our simulations; thus, we decreased the variance of the mutational effects as the number of loci increased. All code is available at http://github.com/Schraiber/quant_trait_coalescent or http://dx.doi.org/10.6084/m9.figshare.1337954.

We first assessed the signature left by various mutational kernels on the sampling distribution by computing the central moments across simulation replicates. We used a variety of mutational kernels in an attempt to capture different kinds of biologically relevant behavior: the (symmetric) normal distribution, the skew-normal distribution (with skewness of 0.1, 0.5, and 0.9), and the Laplace distribution. We omit power law mutational kernels from this portion of the analysis because they do not have finite moments.

We simulated data while varying the number of loci from $L = 2$ to $L = 256$, while holding the sample size constant at $n = 1024$. For each mutational kernel and each $L$, we simulated 2000 replicates. Afterwards, we computed the mean values of the minimum variance unbiased estimators of the second, third and fourth central moments (i.e. $h_2$, $h_3$, and $h_4$) across simulation replicates. We then compared these to their expected values, as computed in (3). See Fig. 2.

By design, $h_2$ remains constant regardless of the mutational kernel of $L$. The normal and Laplace distributions are symmetric, and produce $h_3$ of 0, regardless of the number of QTL. This result is consistent with our analytical analysis, which shows that the trait distribution should be symmetric if the underlying mutational kernel is symmetric. However, skew-normal mutational effects result in non-zero skewness even for traits controlled by over 100 loci if the underlying mutational kernel is sufficiently strongly skewed. Thus, the rate at which the sampling distribution's third central moment, $h_3$, converges to zero is in inverse proportion to the magnitude of the mutational kernel's skewness. All mutational effects result in non-zero $h_4$ values when $L$ is small, due to the randomness of the mutational process. Nonetheless, the Laplace distribution, the sole leptokurtic kernel in this comparison, is the slowest of all to converge to the normally distributed limit, suggesting the importance of the kurtosis of the mutational kernel for determining the kurtosis of the trait distribution.

We predicted, based on the multivariate stable limit derived for power law mutational effects, that power law mutational kernels may result in multimodal sampling distributions, even for traits controlled by a large number of QTL. We set out to determine the frequency of multimodality using the dip test (Hartigan and Hartigan, 1985). Briefly, the dip test computes a statistic measuring departure from unimodality and compares it to a conservative null distribution. Because the dip test is conservative, we expected to
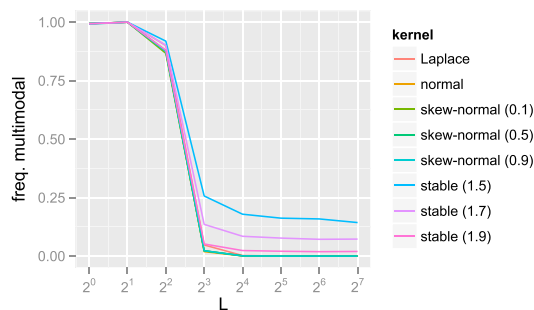
**Fig. 3.** Frequency to reject unimodal sampling distribution. Solid lines report the frequency the null hypothesis of the dip-test, that the sampling distribution was unimodal, was rejected for $p < 0.05$ when evolving under various mutational kernels. Data were simulated for 1024 sampled individuals and 2000 replicates for eight values of $L$, the number of loci. Colors distinguish the mutational kernel and relevant kernel parameters (if any). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reject unimodality less than 5% of the time for large $L$ and kernels without power law behavior. Conversely, for kernels with power law behavior, we expect a larger fraction of tests to reject the null hypothesis, even for large numbers of loci.

We complemented the earlier simulated data with data simulated under three additional, $\alpha$-stable, mutational kernels. We chose $\alpha \in \{1.5, 1.7, 1.9\}$ to assess a wide range of heavy tailed behavior, while ensuring that the distributions retained a finite mean.

Fig. 3 shows that all mutational kernels result in multimodal sampling distributions when there are a small number of QTL. However, trait distributions with $\alpha$-stable mutational kernels remained multimodal even for large $L$ and the frequency of rejecting unimodality is proportional to the $\alpha$ value of the mutational kernel. As $\alpha$ decreases, the large effect mutations responsible for multimodality become more prominent, and cause a larger proportion of simulations to reject unimodality. In contrast, mutational kernels without power law behavior become unimodal as $L$ increases. Nonetheless, we again see that heavier tailed distributions take longer to converge to the normal limit; in particular, Laplace distributed mutations converge to unimodality more slowly than normally distributed mutations.

## 4. Discussion

The natural world is replete with quantitative trait variation, and understanding the forces governing the evolution of quantitative traits is a central goal of evolutionary biology. The model of Fisher (1918), which explained how quantitative variation can be generated by Mendelian inheritance, provides an underpinning for understanding the generation and maintenance of variation in continuous characters. A primary assumption of much of this work is that traits are controlled by a large number of loci and that new mutations have a very small, symmetric effect on the trait value.

In this work, we introduced a coalescent framework for modeling neutral evolution in quantitative traits. This stands in contrast to past work, which has typically taken a forward-in-time approach based on classical population genetics (but see Whitlock (1999) who also utilized a coalescent model). Our backward-in-time, sample-focused approach enabled us to derived an expression for the joint distribution of the data with arbitrary mutational effects and numbers of loci. We found that traits governed by a large number of loci with small effects are well-modeled by a Gaussian distribution, as expected. However, we saw that with small numbers of loci, significant departures from normality can be observed. Moreover, for fat-tailed (or power-law) mutational kernels, there are significant departures from normality (including multimodality), even when the number of loci becomes large.

We assessed departure from normality in traits governed by a small number of loci by exploring the central moments of three different mutational kernels (normal, skew-normal and Laplace distributions) both analytically and by simulation. We showed that although all three mutational kernels converge to a Gaussian distribution, traits controlled by a small number of loci retain the signature of their underlying mutational kernel in their 3rd and 4th central moments. Hence, it may be possible to reconstruct aspects of the mutational effect distribution by observing phenotypes in natural populations. This may be particularly interesting for analyzing variation in gene expression, because mutational effects in *cis* may be strongly skewed (Khaitovich et al., 2005a; Chaix et al., 2008; Gruber et al., 2012). Our theory suggests that the distribution of gene expression in a population might therefore be skewed.

We were also interested in the circumstances under which multi-modal phenotypic distributions can arise. When a trait has a simple genetic architecture, it is easy to see that there must be discrete phenotypic clusters, corresponding to groups of individuals sharing the same mutations. As the number of loci increases, there are more mutational targets (and thus more mutation events), which smooths the distribution, causing the sampling distribution to converge to the appropriate limiting distribution. For mutational effects with finite variance, this ultimately results in a limiting Gaussian distribution, consistent with the central limit theorem. However, when the mutational kernel is fat-tailed, the marginal effects of each locus do not vanish as the number of loci grows. Thus, some clade-specific mutations will always be of large effect despite the number of loci assumed by the model, resulting in a multi-modal sampling distribution.

These results show that even under the assumption of neutrality, significant departures from normality are possible and can be detected in empirical data. It is possible that these deviations from normality may be conflated with signatures of selection acting on quantitative variation. Several recent studies have claimed that evidence of non-Gaussianity may be evidence for non-neutral evolution at macroevolutionary time scales. For instance, Khaitovich et al. (2005a) and Chaix et al. (2008) found that the distribution of gene expression differences between great apes is strongly positively skewed. Similarly, Uyeda et al. (2011) argued that there is a one million year wait between bursts of evolution in the fossil record and numerous studies have explored non-Gaussian trait divergence in a phylogenetic context (Landis et al., 2013; Eastman et al., 2013). While it is unlikely that the population genetic model we developed can be directly applied to macroevolutionary data of this sort (Estes and Arnold, 2007), it is important to recognize that such effects can be due to purely neutral processes.

On shorter time scales, there is significant interest in detecting non-neutral quantitative trait evolution among closely related species or populations. One powerful method compares a measure of quantitative trait divergence, $Q_{st}$, to the fixation index, $F_{st}$ (McKay and Latta, 2002; Ovaskainen et al., 2011). However, this requires estimates of breeding values from common-garden experiments, and may be difficult to achieve. In other cases (e.g. Lemos et al. (2005)) more phenomenological approaches are taken, by comparing within and between species phenotypic diversity. The null distributions of these approaches typically rely on assumptions of the infinitesimal model, which we have shown may be violated due to mutations of large effect and/or loci with relatively simple genetic bases. To address these issues and leverage the abundance of modern quantitative trait data, Berg and Coop (2014) developed a method that explicitly uses breeding values estimated from quantitative trait mapping studies. When such effect size estimates are unavailable, it may be possible to use our formalism to develop robust null models to detect selection.

This framework, which provides a generative model based on an explicit characterization of the underlying mutational

kernel, maybe useful for inferring parameters of mutational effect distributions from phenotypes sampled in natural populations. In particular, a large number of studies are now quantifying thousands of phenotypes across a large number of individuals by assaying molecular phenotypes such as gene expression (e.g Lappalainen et al. (2013)) or chromatin accessibility (e.g Degner et al. (2012)). Because these traits are thought to evolve subject to relatively few QTL of relatively large effect, we believe that it will be possible to use our model to make inferences about the mutational effects that shape phenotypic variation at the molecular level.

Our coalescent approach can be extended in several ways. Notably, we consider only haploid populations. In principle, an extension to diploid individuals is straight-forward using the result of Möhle (1998) that diploid, diecious populations of size $N$ are readily modeled by pairing random chromosomes from a haploid population of size $2N$. To incorporate diploidy, we would also need to incorporate a model of dominance, of which several exist in the literature (e.g. the model of independent dominance of Lynch and Hill (1986)).

From the point of view of the coalescent process, it is straight-forward to apply our model to populations that have undergone complex demographic histories. This is because the dynamics of a coalescent under population size fluctuations and population structure are well known. Moreover, we explored only unlinked, neutral loci and it may be possible to obtain some analytical results for linked loci and/or weak natural selection by using the ancestral recombination graph and ancestral selection graph, respectively. While analytical results are difficult within these frameworks, we believe that they can be used to perform simple simulations of quantitative traits evolving in complex scenarios, thus enabling Approximate Bayesian Computation.

## Acknowledgments

## Appendix. Mathematical derivations

### A.1. Compound Poisson processes

To obtain the probability of the data under this model, we must be able to compute the probability of the change in phenotype along a branch of the tree. Unfortunately, except for very simple mutational models, this probability is impossible to compute analytically. Instead, we compute the characteristic function of the change along a branch.

Using standard results for compound Poisson processes (Kingman, 1992), we see that the characteristic function of the change along a branch of length $t$ (in coalescent units) is

$$\varphi_t(k) = e^{\frac{\theta}{2}t(\psi(k)-1)} \tag{4}$$

where $\psi$ is the characteristic function of the mutational effect distribution.

### A.2. The phenotype at the root of the sample genealogy and the subsequent evolution within the sample are subindependent

Note that
$$
\begin{aligned}
\lambda_n(\mathbf{k}) &= \mathbb{E}(e^{i\mathbf{k}^T\mathbf{X}}) \\
&= \mathbb{E}(e^{i(k_1 X_1 + k_2 X_2 + \cdots + k_n X_n)}) \\
&= \mathbb{E}(e^{i(k_1(\mathcal{R}+\mathcal{E}_1) + k_2(\mathcal{R}+\mathcal{E}_2) + \cdots + k_n(\mathcal{R}+\mathcal{E}_n))})
\end{aligned}
$$

where $\mathcal{R}$ is the phenotype at the root of the sample genealogy and $\mathcal{E}_u$ is the subsequent evolution leading to lineage $u$ in the sample. So,

$$
\begin{aligned}
&\mathbb{E}(e^{i(k_1(\mathcal{R}+\mathcal{E}_1) + k_2(\mathcal{R}+\mathcal{E}_2) + \cdots + k_n(\mathcal{R}+\mathcal{E}_n))}) \\
&= \mathbb{E}_{\mathcal{R}}(\mathbb{E}(e^{i(k_1(\mathcal{R}+\mathcal{E}_1) + k_2(\mathcal{R}+\mathcal{E}_2) + \cdots + k_n(\mathcal{R}+\mathcal{E}_n))}|\mathcal{R})) \\
&= \mathbb{E}_{\mathcal{R}}(e^{i(k_1+k_2+\cdots+k_n)\mathcal{R}}\mathbb{E}(e^{i(k_1\mathcal{E}_1 + k_2\mathcal{E}_2 + \cdots + k_n\mathcal{E}_n)}|\mathcal{R})) \\
&= \mathbb{E}(e^{i(k_1+k_2+\cdots+k_n)\mathcal{R}})\mathbb{E}(e^{i(k_1\mathcal{E}_1 + k_2\mathcal{E}_2 + \cdots + k_n\mathcal{E}_n)})
\end{aligned}
$$

where the last line follows by independent and stationary increments of the compound Poisson process. Thus, $\mathcal{R}$ and $(\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n)$ subindependent, and hence their joint characteristic function is the product of their characteristic functions.

### A.3. Proof of recursive formula for the characteristic function

In this section we use $\mathbf{X}$ to indicate the vector of trait value conditional on the common ancestor of the sample having trait value 0. First, we condition on the state at the first coalescence (going back in time). The state consists of three components: (1) which pair of individuals coalesce, $(u, v)$, (2) the time of the coalescent event, $T_c$, and (3) the trait value in each lineage at that time, $\mathbf{X}'$ (note that, given $(u, v)$, we have that $X'_u = X'_v$, since those two lineages have coalesced and hence had the same trait value at the time of coalescence). Then,

$$
\begin{aligned}
\mathbb{E}(e^{i\mathbf{k}^T\mathbf{X}}) &= \mathbb{E}_{(u,v),\mathbf{X}',T_c}\left(\mathbb{E}\left(e^{i\mathbf{k}^T\mathbf{X}}|(u,v),\mathbf{X}',T_c\right)\right) \\
&= \frac{2}{n(n-1)}\sum_{u<v}\mathbb{E}_{\mathbf{X}',T_c}\left(\mathbb{E}\left(e^{i\mathbf{k}^T\mathbf{X}}|(u,v),\mathbf{X}',T_c\right)\right) \\
&= \frac{2}{n(n-1)}\sum_{u<v}\mathbb{E}_{\mathbf{X}',T_c}\left(\mathbb{E}\left(e^{i\mathbf{k}^T(\mathbf{X}'+\mathbf{Y}(T_c))}|(u,v),\mathbf{X}',T_c\right)\right) \\
&= \frac{2}{n(n-1)}\sum_{u<v}\mathbb{E}_{\mathbf{X}',T_c}\left(e^{i\mathbf{k}^T\mathbf{X}'}\mathbb{E}\left(e^{i\mathbf{k}^T\mathbf{Y}(T_c)}|T_c\right)\right) \quad (5)
\end{aligned}
$$

where $\mathbf{Y}(t) = (Y_1(t), Y_2(t), \ldots, Y_n(t))$ is the vector accounting for the evolution on each lineage that occurs during time $t$. The second line follows by the fact that each pair is equally likely to coalesce (with probability $\binom{n}{2}^{-1}$) and the third line by independent increments of a compound Poisson process.

Now, we compute the internal expectation going forward in time. Noticing that $\mathbb{E}(e^{i\mathbf{k}^T\mathbf{Y}(T_c)}|T_c)$ is simply the characteristic function of a compound Poisson process run for length $T_c$, we see from (4) that

$$\mathbb{E}\left(e^{i\mathbf{k}^T\mathbf{Y}(T_c)}|T_c\right) = \exp\left\{\frac{\theta}{2}T_c\left(\sum_{i=1}^{n}\psi(k_i) - n\right)\right\}.$$

Because $T_c$ and $\mathbf{X}'$ are independent, we can integrate over $T_c$ analytically in the outer expectation. The distribution of the time to the first coalescent event in a sample of size $n$ is Exponential with rate $\binom{n}{2}$, hence,

$$
\begin{aligned}
&\mathbb{E}_{T_c}\left(\exp\left\{\frac{\theta}{2}T_c\left(\sum_{i=1}^{n}\psi(k_i) - n\right)\right\}\right) \\
&= \frac{n(n-1)}{n(n-1) - \theta\left(\sum_{i=1}^{n}\psi(k_i) - n\right)}.
\end{aligned}
$$

Plugging this result into (5) results in

$$\mathbb{E}\left(e^{i\mathbf{k}^T \mathbf{X}}\right) = \frac{2}{n(n-1) - \theta\left(\sum_{i=1}^{n} \psi(k_i) - n\right)} \sum_{u<v} \mathbb{E}_{\mathbf{X}'}\left(e^{i\mathbf{k}^T \mathbf{X}'}\right),$$

but since $\mathbf{X}'$ is simply the result of the same process where two of the entries are identical, we obtain the recursive formula (1) with $\phi_n = \mathbb{E}\left(e^{i\mathbf{k}^T \mathbf{X}}\right)$ when $\mathbf{X}$ is a vector of size $n$.

To initialize the recursion, we must compute the characteristic function for a sample of size 2. This is

$$\phi_2(\mathbf{k}) = \frac{2}{2 - \theta(\psi(k_1) + \psi(k_2) - 2)}. \tag{6}$$

### A.4. The phenotype at the root of the sample genealogy

First, we define $\Delta$ to be the time between when the sample genealogy finds a common ancestor and when the population genealogy finds a common ancestor. Then, we note that, conditional on $\Delta$, the characteristic function of the phenotype at the root of the sample genealogy is

$$e^{\frac{\theta}{2}(\psi(k)-1)\Delta},$$

by using Eq. (4). Thus, the after integrating over $\Delta$, the desired quantity is *moment generating function* of $\Delta$, defined by

$$M(z) = \mathbb{E}(e^{z\Delta})$$

evaluated at $z = \frac{\theta}{2}(\psi(k) - 1)$.

We compute $M(z)$ by conditioning on how many lineages are left in the population genealogy when the sample reaches its most recent common ancestor. To do this, we make use of a result of Saunders et al. (1984),

$$\mathbb{P}(u \text{ lineages left in population|sample coalesced})$$
$$= n!(n-1)\frac{u!}{(n+u)!}.$$

Given that $u$ lineages are left in the population when the sample reaches its most recent common ancestor, the remaining time until the whole population reaches its common ancestor is simply the time it takes for a coalescent started with $u$ to reach its most recent common ancestor, $C_u$. Thus by conditioning on the number of lineages left in the population and using the result from Saunders et al. (1984),

$$M(z) = \mathbb{E}(e^{z\Delta})$$
$$= \mathbb{E}_u(\mathbb{E}(e^{z\Delta}|u \text{ lineages left when sample coalesces}))$$
$$= \sum_{u=1}^{\infty} \mathbb{E}(e^{zC_u})n!(n-1)\frac{u!}{(n+u)!}$$
$$= \sum_{u=1}^{\infty} \prod_{v=2}^{u} \frac{v(v-1)}{v(v-1)-2z}n!(n-1)\frac{u!}{(n+u)!}$$

where the final line follows by recognizing that $C_u$ is the sum of $u-1$ independent exponential random variables with means $\binom{u}{2}, \binom{u-1}{2}, \ldots, \binom{2}{2}$. Substituting $\frac{\theta}{2}(\psi(k)-1)$ for $z$ yields the desired result.

### A.5. Computing sample central moments

While it is difficult to compute the expectation of any sample central moments for a particular sample, it is possible to average over replicate populations to compute expectations over replicate samples. We first begin by defining the *h*-statistics, which are

the unique minimum variance unbiased estimators of the central moments (Halmos et al., 1946). In particular, letting $X_i$ be the phenotype of individual $i$ in a sample of size $n$ (n.b. that these labels are arbitrary because individuals are exchangeable), and putting $S_p = \sum_{i=1}^{n} X_i^p$,

$$h_2 = \frac{nS_2 - S_1^2}{(n-1)n}$$

$$h_3 = \frac{2S_1^3 - 3nS_1S_2 + n^2S_3}{(n-2)(n-1)n}$$

$$h_4 = \frac{6nS_1^2S_2 + 3(3-2n)S_2^2 - 4(n^2 - 2n + 3)S_1S_3 + (n^3 - 2n^2 + 3n)S_4 - 3S_1^4}{(n-3)(n-2)(n-1)n}.$$

We now compute expectations over these quantities by computing expectations over the products of the $S_p$. Because the phenotypes of the samples are correlated, this results in formulas that are different from the case of independent and identically distributed random variables. For instance,

$$\mathbb{E}(S_1^2) = \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^2\right)$$
$$= \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_iX_j\right)$$
$$= \mathbb{E}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_iX_j\right)$$
$$= n\mathbb{E}(X_1^2) + n(n-1)\mathbb{E}(X_1X_2).$$

Computing similar formulas as required and substituting into the definitions of the *h*-statistics results in

$$\mathbb{E}(h_2) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1X_2)$$
$$\mathbb{E}(h_3) = \mathbb{E}(X_1^3) + 2\mathbb{E}(X_1X_2X_3) - 3\mathbb{E}(X_1^2X_2)$$
$$\mathbb{E}(h_4) = \mathbb{E}(X_1^4) + 6\mathbb{E}(X_1^2X_2X_3) - 4\mathbb{E}(X_1^3X_2) - 3\mathbb{E}(X_1X_2X_3X_4),$$

where the expectations on the right hand sides are over the *correlated* phenotypes in the sample. It is possible to compute these expectations by taking derivatives of the characteristic function (1). We demonstrate this for the case of $h_2$; all other results follow similarly.

First, note that because all of these moments are central moments, the phenotype at the root of the sample genealogy will always cancel out and we only are concerned with the characteristic function relating to the evolution within the sample (1). Next, using the fact that the characteristic function of a sum of random variables is the product of their characteristic functions, let $\phi_n^{(L)}(\mathbf{k}) = \phi_n(\mathbf{k})^L$, i.e. the characteristic function of a trait governed by $L$ identical loci. A basic property of the characteristic function is that

$$\mathbb{E}(X_1^{p_1}X_2^{p_2}\cdots X_n^{p_n})$$
$$= (-i)^{\sum_u p_u} \frac{\partial^{\sum_u p_u}}{\partial k_1^{p_1}\partial k_2^{p_2}\cdots \partial k_n^{p_n}}\phi_n^{(L)}(\mathbf{k})\Bigg|_{k_1=k_2=\cdots=k_n=0}.$$

So,

$$\mathbb{E}(X_1^2) = (-i)^2 \frac{\partial^2}{\partial k_1^2}\phi_n^{(L)}(\mathbf{k})\Bigg|_{k_1=k_2=\cdots=k_n=0}$$
$$= C_n + \frac{n-1}{n}\theta L m_2$$

and

$$\mathbb{E}(X_1X_2) = (-i)^2 \frac{\partial^2}{\partial k_1 \partial k_2}\phi_n^{(L)}(\mathbf{k})\Bigg|_{k_1=k_2=\cdots=k_n=0}$$
$$= C_n + \frac{n-2}{2n}\theta L m_2,$$

where $C_n$ is a term that depends on $n$ and is identical in the two computations, and $m_2 = -\psi''(0)$ is the second moment of the mutational effect kernel. Finally,

$$
\begin{aligned}
\mathbb{E}(h_2) &= \mathbb{E}(X_1^2) - \mathbb{E}(X_1 X_2) \\
&= C_n + \frac{n-1}{n}\theta L m_2 - \left(C_n + \frac{n-2}{2n}\theta L m_2\right) \\
&= \frac{1}{2}\theta L m_2.
\end{aligned}
$$

### A.6. Derivation of multivariate stable limit for sample distribution

Recall that a random variable $X$ is said to have a fat-tailed (or power-law) distribution if

$$
P(X > x) \sim \kappa x^{-\alpha} \tag{7}
$$

for large $x$ and some $\kappa > 0$. As is typical in the literature, we reserve the term "fat-tailed" for distributions with $\alpha \in (0, 2)$.

To obtain an appropriate scaling limit, we assume that there is a parameter $t$, related to the parameter $\kappa$ in (7) by

$$
t = \kappa \frac{\pi}{\sin(\alpha\pi/2)\Gamma(\alpha)\alpha}, \tag{8}
$$

such that $Lt \to s$ as $n \to \infty$. Note that $s^{1/\alpha}$ is proportional to the scale parameter of the resulting limit distribution.

We provide a heuristic derivation, rather than a rigorous proof. First, we argue by induction that the (per locus) characteristic function for a sample of size $n$ is

$$
\tilde{\phi}_n(\mathbf{k}) \sim 1 - \frac{\theta s}{L}\left(\sum_{\mathbf{j}\in\mathcal{P}^*(\mathbf{k})} c_{n,|\mathbf{j}|}\left|\sum_{z\in\mathbf{j}} z\right|^{\alpha}\right)
$$

for large $L$, where $\mathcal{P}^*(\mathbf{k})$ is the power set of the elements in $\mathbf{k}$, *except* the set $\{k_1, k_2, \ldots, k_n\}$, and $c_{n,|\mathbf{j}|}$ is a combinatorial constant that depends only on the sample size $n$ and $|\mathbf{j}|$, the size of the set $\mathbf{j}$.

Note that for $n = 2$, this can be seen by observing that for large $L$, the characteristic function of a fat-tailed distribution is asymptotically

$$
\begin{aligned}
\psi(k) &\sim 1 - \frac{s}{L}|k| \\
&\equiv \tilde{\psi}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\phi_2(\mathbf{k}) &\sim 1 - \frac{\theta}{2}\frac{s}{L}(|k_1|^{\alpha} + |k_2|^{\alpha}) \\
&\equiv \tilde{\phi}_2.
\end{aligned}
$$

Now, assume that the formula holds for $\tilde{\phi}_{n-1}$. Using the recursion (1), we have

$$
\begin{aligned}
\phi_n(\mathbf{k}) &\sim \frac{1}{\binom{n}{2} - \frac{\theta}{2}\left(\sum_{u=1}^{n}\tilde{\psi}(k_u) - n\right)}\sum_{u<v}\tilde{\psi}_{n-1}(\mathbf{k}^{(u,v)}) \\
&= \frac{\binom{n}{2} - \frac{\theta s}{L}\left(\sum_{\mathbf{j}\in\mathcal{P}^*(\mathbf{k})} \tilde{c}_{n,|\mathbf{j}|}\left|\sum_{z\in\mathbf{j}} z\right|^{\alpha}\right)}{\binom{n}{2} + \frac{\theta s}{2L}\left(\sum_{u=1}^{n}|k_u|^{\alpha}\right)}.
\end{aligned}
$$

The second line follows from plugging $\tilde{\psi}$ and $\tilde{\phi}$, and $\tilde{c}_{n,|\mathbf{j}|}$ arises by summing over the appropriate terms coming from all

characteristic functions in the sum. Again looking for an asymptotic for large $L$, we see that

$$
\begin{aligned}
&\frac{\binom{n}{2} - \frac{\theta s}{L}\left(\sum_{\mathbf{j}\in\mathcal{P}^*(\mathbf{k})} \tilde{c}_{n,|\mathbf{j}|}\left|\sum_{z\in\mathbf{j}} z\right|^{\alpha}\right)}{\binom{n}{2} + \frac{\theta s}{2L}\left(\sum_{u=1}^{n}|k_u|^{\alpha}\right)} \\
&\sim 1 - \frac{\theta s}{\binom{n}{2}L}\left(\sum_{\mathbf{j}\in\mathcal{P}^*(\mathbf{k})} \tilde{c}_{n,|\mathbf{j}|}\left|\sum_{z\in\mathbf{j}} z\right|^{\alpha} + \frac{1}{2}\sum_{u=1}^{n}|k_u|^{\alpha}\right) \\
&= 1 - \frac{\theta s}{L}\left(\sum_{\mathbf{j}\in\mathcal{P}^*(\mathbf{k})} c_{n,|\mathbf{j}|}\left|\sum_{z\in\mathbf{j}} z\right|^{\alpha}\right) \\
&= \tilde{\phi}_n(\mathbf{k}).
\end{aligned}
$$

Finally, we note that by raising $\tilde{\phi}_n$ to the $L$th power, and taking the limit as $L \to \infty$, we obtain the log characteristic function

$$
\log\Phi_n(\mathbf{k}) = -\theta s\left(\sum_{\mathbf{j}\in\mathcal{P}^*(\mathbf{k})} c_{n,|\mathbf{j}|}\left|\sum_{z\in\mathbf{j}} z\right|^{\alpha}\right), \tag{9}
$$

where all terms are defined as before.

The characteristic function in (9) can be recognized to be that of a multivariate $\alpha$-stable distribution (Press, 1972). These multivariate distributions are fat-tailed generalizations of the familiar multivariate normal distribution, and this limit corresponds to a generalized multivariate central limit theorem for sums of random vectors with fat-tailed distributions.

### A.7. Limiting distribution of the phenotype at the root of the sample genealogy

Again, we proceed heuristically rather than rigorously. First, note that for large $L$,

$$
\frac{v(v-1)}{v(v-1) - \theta(\psi(k) - 1)} \sim 1 - \frac{\theta s}{Lv(v-1)}|k|^{\alpha}
$$

so that

$$
\prod_{v=2}^{u}\left(1 - \frac{\theta s}{Lv(v-1)}|k|^{\alpha}\right) \sim 1 - \frac{u-1}{u}\frac{s}{L}\theta|k|^{\alpha}
$$

for large $L$. Thus,

$$
\begin{aligned}
\rho_n(k) &\sim n!(n-1)\sum_{u=1}^{\infty}\left(1 - \frac{u-1}{u}\frac{s}{L}\theta|k|^{\alpha}\right)\frac{u!}{(n+u)!} \\
&= 1 - \frac{s\theta}{Ln}|k|^{\alpha}.
\end{aligned}
$$

So, by definition of the exponential function, we have that

$$
\begin{aligned}
R_n(k) &= \lim_{L\to\infty}\rho_n(k)^L \\
&= e^{-\frac{s\theta}{n}|k|^{\alpha}}, \tag{10}
\end{aligned}
$$

which is the characteristic function of a univariate $\alpha$-stable distribution, arising from the fact that the phenotype at the root of the sample genealogy is itself a limit of a sum of random variables. Note that as $n \to \infty$ (i.e. the sample becomes the whole population), $R(k) \to 1$, because the root of the sample genealogy is the same as the root of the population genealogy and the root value has been specified to be equal to 0.

## A.8. Multivariate Gaussian limits

For the case where the mutation distribution is not fat-tailed, we can use the multivariate central limit theorem to more efficiently derive the limiting distribution. The appropriate scaling in this case is to assume that if $\tau^2$ is the variance of the mutation effect kernel, then $L\tau^2 \to \sigma^2$ as $L \to \infty$.

To apply the multivariate central limit theorem, we must derive the pairwise covariances between samples. While the required covariances could be computed by taking derivatives of the characteristic function, it is more instructive to compute these moments directly. For simplicity, we assume that the mutation effect distribution has mean 0 and variance $\tau^2$.

Assume that the population genealogy at a single locus, $\mathcal{G}$, is fixed. Noting that the variance per unit time accrued by the mutational process is $\theta/2\tau^2$ and using the rules for calculating covariance structure on a phylogeny, it is easy to see that for samples $i$ and $j$ we have

$$\text{Cov}(X_i, X_j | \mathcal{G}) = \begin{cases} \dfrac{\theta}{2}\tau^2 T & \text{if } i = j \\ \dfrac{\theta}{2}\tau^2 (T - T_{ij}) & \text{if } i \neq j \end{cases}$$

where $T$ is the height of $\mathcal{G}$ and $T_{ij}$ is the height of the most recent common ancestor of samples $i$ and $j$. We can then use the law of total covariance,

$$\text{Cov}(X_i, X_j) = \mathbb{E}(\text{Cov}(X_i, X_j | \mathcal{G})) + \text{Cov}(\mathbb{E}(X_i | \mathcal{G}), \mathbb{E}(X_j | \mathcal{G}))$$

to see that

$$\text{Cov}(X_i, X_j) = \begin{cases} \theta\tau^2 & \text{if } i = j \\ \dfrac{1}{2}\theta\tau^2 & \text{if } i \neq j. \end{cases}$$

This arises because $\mathbb{E}(T) = 2$ and $\mathbb{E}(T_{ij}) = 1$.

Hence, as the number of loci increases to infinity in such a way that $L\tau^2 \to \sigma^2$, the sampling distribution converges to a multivariate normal distribution with mean 0 and variance covariance matrix $\Sigma$ having elements

$$\Sigma_{ij} = \begin{cases} \theta\sigma^2 & \text{if } i = j \\ \dfrac{1}{2}\theta\sigma^2 & \text{if } i \neq j. \end{cases}$$

Because the pairwise covariances are equal, the random vector **X** is an exchangeable Gaussian random vector. Hence, using well-known facts about the representation of exchangeable Gaussian random vectors, one arrives at the representation in the main text.

## References

Aldous, D.J., 1985. Exchangeability and Related Topics. Springer.

Berg, J.J., Coop, G., 2014. A population genetic signal of polygenic adaptation. PLoS Genet. 10, e1004412.

Bürger, R., Lande, R., 1994. On the distribution of the mean and variance of a quantitative trait under mutation-selection-drift balance. Genetics 138, 901–912.

Chaix, R., Somel, M., Kreil, D.P., Khaitovich, P., Lunter, G., 2008. Evolution of primate gene expression: drift and corrective sweeps? Genetics 180, 1379–1389.

Chakraborty, R., Nei, M., 1982. Genetic differentiation of quantitative characters between populations or species: I. Mutation and random genetic drift. Genet. Res. 39, 303–314.

Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al., 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 482, 390–394.

Doebeli, M., Blok, H.J., Leimar, O., Dieckmann, U., 2007. Multimodal pattern formation in phenotype distributions of sexual populations. Proc. R. Soc. B Biol. Sci. 274, 347–357.

Eastman, J.M., Wegmann, D., Leuenberger, C., Harmon, L.J., 2013. Simpsonian 'evolution by jumps' in an adaptive radiation of anolis lizards. arXiv preprint arXiv:1305.4216.

Estes, S., Arnold, S.J., 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am. Natur. 169, 227–244.

Falconer, D., Mackay, T., 1996. Introduction to Quantitative Genetics, fourth ed. American Genetic Association.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52, 399–433.

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

Galton, F., 1883. Inquiries into Human Faculty and its Development. Macmillan.

Galton, F., 1889. Natural Inheritance. Macmillan.

Gruber, J.D., Vogel, K., Kalay, G., Wittkopp, P.J., 2012. Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. PLoS Genet. 8, e1002497.

Haldane, J., 1954. The statics of evolution, Evolution as a Process, pp. 109–121.

Halmos, P.R., et al., 1946. The theory of unbiased estimation. Ann. Math. Stat. 17, 34–43.

Hartigan, J.A., Hartigan, P., 1985. The dip test of unimodality. Ann. Statist. 70–84.

Hudson, R.R., 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics 18, 337–338.

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., Pääbo, S., 2005a. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309, 1850–1854.

Khaitovich, P., Pääbo, S., Weiss, G., 2005b. Toward a neutral evolutionary model of gene expression. Genetics 170, 929–939.

Kimura, M., 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. USA 54, 731.

Kingman, J.F.C., 1992. Poisson Processes, third ed. Oxford University Press.

Kopp, M., Hermisson, J., 2006. The evolution of genetic architecture under frequency-dependent disruptive selection. Evolution 60, 1537–1550.

Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 314–334.

Landis, M.J., Schraiber, J.G., Liang, M., 2013. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Syst. Biol. 62, 193–204.

Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A, Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al., 2013. Transcriptome and genome sequencing uncovers functional variation in humans. Nature.

Latter, B., 1960. Natural selection for an intermediate optimum. Aust. J. Biol. Sci. 13, 30–35.

Latter, B., 1970. Selection in finite populations with multiple alleles. ii. Centripetal selection, mutation, and isoallelic variation. Genetics 66, 165.

Lemos, B., Meiklejohn, C.D., Cáceres, M., Hartl, D.L., 2005. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. Evolution 59, 126–137.

Lynch, M., Hill, W.G., 1986. Phenotypic evolution by neutral mutation. Evolution 915–935.

Mackay, T., Lyman, R.F., Jackson, M.S., 1992. Effects of P element insertions on quantitative traits in *Drosophila melanogaster*. Genetics 130, 315–332.

McKay, J.K., Latta, R.G., 2002. Adaptive population divergence: markers, qtl and traits. Trends Ecol. Evol. 17, 285–291.

Mendel, G., 1866. Versuche über pflanzenhybriden. Verhandlungen des natur-forschenden Vereines in Brunn 4 (3), 44.

Möhle, M., 1998. Coalescent results for two-sex population models. Adv. Appl. Probab. 513–520.

Ovaskainen, O., Karhunen, M., Zheng, C., Arias, J.M.C., Merilä, J., 2011. A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. Genetics 189, 621–632.

Pearson, K., 1894. Contributions to the mathematical theory of evolution. Philos. Trans. R. Soc. Lond. Ser. A 71–110.

Pearson, K., 1895. Contributions to the mathematical theory of evolution. III. regression, heredity, and panmixia. Proc. R. Soc. Lond 59, 69–71.

Pearson, K., 1904. Mathematical contributions to the theory of evolution. XII. On a generalised theory of alternative inheritance, with special reference to Mendel's laws. Philos. Trans. R. Soc. Lond. Ser. A 53–86.

Press, S.J., 1972. Multivariate stable distributions. J. Multivariate Anal. 2, 444–462.

R Core Team 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria.

Saunders, I.W., Tavaré, S., Watterson, G., 1984. On the genealogy of nested subsamples from a haploid population. Adv. Appl. Probab. 471–491.

Uyeda, J.C., Hansen, T.F., Arnold, S.J., Pienaar, J., 2011. The million-year wait for macroevolutionary bursts. Proc. Natl. Acad. Sci. 108, 15908–15913.

Weldon, W.F.R., 1902. Mendel's laws of alternative inheritance in peas. Biometrika 228–254.

Whitlock, M.C., 1999. Neutral additive genetic variance in a metapopulation. Genet. Res. 74, 215–221.