

The effects of demography and genetics on the neutral distribution of quantitative traits

Evan M. Koch*

emkoch@uchicago.edu

Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

1 Abstract

Neutral models for quantitative trait evolution are useful for identifying phenotypes under selection in natural populations. Models of quantitative traits often assume phenotypes are normally distributed. This assumption may be violated when a trait is affected by relatively few genetic variants or when the effects of those variants arise from skewed or heavy-tailed distributions. Traits such as gene expression levels and other molecular phenotypes may have these properties. To accommodate deviations from normality, models making fewer assumptions about the underlying trait genetics and patterns of genetic variation are needed. Here, we develop a general neutral model for quantitative trait variation using a coalescent approach by extending the framework developed by SCHRAIBER and LANDIS (2015). This model allows interpretation of trait distributions in terms of familiar population genetic parameters because it is based on the coalescent. We show how the normal distribution resulting from the infinitesimal limit, where the number of loci grows large as the effect size per mutation becomes small, depends only on expected pairwise coalescent times. We then demonstrate how deviations from normality depend on demography through the distribution of coalescence times as well as through genetic parameters. In particular, population growth events exacerbate deviations while bottlenecks reduce them. This model also has practical applications, which we demonstrate by designing an approach to simulate from the null distribution of Q_{ST} , the ratio of the trait variance between subpopulations to that in the overall population. We further show that it is likely impossible to distinguish sparsity from skewed or heavy-tailed distributions of mutational effects using only trait values sampled from a population. The model analyzed here greatly expands the parameter space for which neutral trait models can be designed.

2 Introduction

Neutral models of quantitative traits provide a null distribution against which various goodness-of-fit tests can be used to test for the action of natural selection (LANDE, 1976; LEINONEN *et al.*, 2013). Neutral models can also clarify the effects of purely neutral forces such as genetic drift and

*Present address: Department of Medicine, Division of Genetics, Brigham and Womens Hospital/Harvard Medical School, Boston, MA, USA.

mutation on trait distributions (LYNCH and HILL, 1986). A common approach is to first model phenotypes as normally distributed, either among offspring within a family, among members of a population, or among species (TURELLI, 2017). Indeed, it has been suggested that the normality assumption is the defining characteristic of quantitative genetics (RICE, 2004). This might be justified if phenotypes are influenced by a large number of sufficiently independent Mendelian factors (FISHER, 1918), or normality may simply appear approximately true in practice.

Neutral models for quantitative traits have been developed in a variety of contexts. The goal of these models is to ask to whether phenotypic differentiation between groups can be reasonably explained by processes other than natural selection. On macroevolutionary time scales, models stemming from LANDE (1976) have used Brownian motion to describe the evolution of the mean value of a quantitative trait in a population. These models are used in statistical methods to test for extreme trait divergence between species (TURELLI *et al.*, 1988), test for phylogenetic signal in trait distributions (FRECKLETON *et al.*, 2002), and correct for phylogenetic dependence when calculating correlations between traits (FELSENSTEIN, 1985). On shorter time scales, neutral distributions assuming multivariate normality of trait values also underlie tests for spatially varying selection in structured populations such as the method developed by OVASKAINEN *et al.* (2011). Other neutral models for quantitative traits have not assumed normality (CHAKRABORTY and NEI, 1982; LYNCH and HILL, 1986; LANDE, 1992), and the dynamics of phenotypic evolution are examined forwards in time as a balance between mutation creating variance, migration spreading variance among subpopulations, and fixation removing it. However, these studies were limited to simple models of population structure and history. Backwards-in-time, coalescent models would allow for more general demographic scenarios.

Under the normality assumption, quantitative trait dynamics can be modeled without concern for the number of causal loci influencing the trait, the genealogies at these sites, or the distribution of mutational effects (the distribution that new mutations affecting the trait draw their effects from). However, heritable phenotypic variation is ultimately due to discrete mutations at discrete locations in the genome, and how the phenotypic variance arising from these mutations is distributed depends on the genealogies at these loci. For instance, the distribution of genealogies in the genome might be strongly influenced by recent population growth and the distribution of mutational effects could be skewed for biological reasons inherent in the details of a particular developmental pathway. When the number of mutations affecting a trait is large, the central limit theorem ensures that the distributions of genealogies and mutational effects can be ignored, but a full model of phenotypic variation would have to include them. Importantly, deviations from normality may affect the outcomes of goodness-of-fit tests that necessarily aim to identify outliers from a normal model.

A more inclusive model of neutral phenotypic variation can begin by considering the genealogies at causal loci. The principle modeling framework for genealogical variation is the coalescent process (WAKELEY, 2008), but few studies have connected the coalescent to quantitative genetics. WHITLOCK (1999) used coalescent theory to argue that measures of phenotypic (Q_{ST}) and genetic (F_{ST}) differentiation have the same expected value given general models of population subdivision. By simulating from the coalescent with recombination, GRISWOLD *et al.* (2007) investigated the effects of shared ancestry and linkage disequilibrium on the genetic covariance matrix for a set of traits (\mathbf{G} matrix). They found that linkage disequilibrium and small numbers of causal loci can cause phenotypic covariances not predicted by the mutational covariance matrix. MENDES *et al.* (2018) also used a quantitative trait model based on the coalescent to show that using a species tree based on population split times can lead to, among other problems, increased false positive rates in phylogenetic comparative methods. Although not explicitly connected to the coalescent, OVASKAINEN *et al.* (2011) developed their test for spatially varying

selection by assuming that the covariance in trait values, conditional on the G-matrix in the ancestral population, depends only on the pairwise coancestry coefficients, which have a clear interpretation in terms of the coalescent process (SLATKIN, 1991).

Two studies have asked how the shape of the distribution of mutational effect sizes, beyond just the mutational variance, impacts trait distributions. KHAITOVICH *et al.* (2005) modeled the evolution of gene expression values on phylogenetic trees assuming a single non-recombining causal locus but allowed for an arbitrary distribution of mutational effects. Using this model they were able to detect deviations from normality consistent with asymmetries in the distribution of mutational effects on gene expression in great apes. More recently, SCHRAIBER and LANDIS (2015) developed a similar general model of quantitative trait evolution at the population level based on the coalescent and allowing for any number of causal loci. They derived the characteristic function for the distribution of phenotypic values in a sample and showed how such values can deviate strongly from normality when the number of loci is small or the mutational distribution has heavy tails. SCHRAIBER and LANDIS (2015) note that the possibility for multimodal trait distributions could lead to incorrect inferences of divergent selection within populations.

SCHRAIBER and LANDIS (2015) derived their results for a panmictic, constant-size population. Natural populations rarely have stable population sizes and show considerable spatial structure, and it is unclear how these violations of the constant-size, panmictic model might influence deviations from normality. We take advantage of the ability of coalescent theory to handle nonequilibrium demographies and population structure to relax these modeling assumptions.

Extending coalescent models of quantitative traits to structured populations is important because the analysis of structured populations provides an opportunity to infer the incidence of local adaptation or stabilizing selection. Due to the build up of linkage disequilibrium between alleles affecting a trait, the neutral divergence in trait values among different subpopulations in a structured population is about as variable as the variance in allele frequencies (ROGERS and HARPENDING, 1983). The Q_{ST}/F_{ST} paradigm was developed to test whether trait divergence in structured populations could be explained by neutral forces alone (SPITZE, 1993; WHITLOCK, 2008; LEINONEN *et al.*, 2013). Q_{ST} , defined as the ratio of the trait variance between subpopulations to the total trait variance, is compared to F_{ST} , which measures the same property for genetic variation and is calculated using neutral markers to provide a null distribution. If the observed Q_{ST} is sufficiently far from the null expectation, it is concluded that natural selection has acted. OVASKAINEN *et al.* (2011) developed a modern extension of the Q_{ST}/F_{ST} paradigm for genetic values measured in breeding experiments and BERG and COOP (2014) also extended the paradigm to make use of genetic values computed from GWAS summary statistics. An advantage of the BERG and COOP (2014) approach is that by using computing genetic values from GWAS loci it makes no assumptions about normality at the population level. However, since suitably sized GWAS's have only been performed in humans, the approach has not yet been extended to other species. Understanding the neutral distribution of trait values is therefore necessary for the development of goodness-of-fit tests that are applicable to populations with complicated histories and traits with sparse genetic architectures.

We generalize the work of SCHRAIBER and LANDIS (2015) by deriving the form of the moment generating function (mgf) for arbitrary distributions of coalescent times (e.g. those arising under exponential growth or an island model of migration). The key result of SCHRAIBER and LANDIS (2015), the characteristic function of the sampling distribution of phenotypic values, is a special case of this general generating function. We then show how a normal model arises by taking the infinitesimal limit where the effect size per mutation becomes small as the number of loci potentially affecting the trait becomes large. We then calculate the third and fourth central moments of the trait distribution in panmictic populations to illustrate how departures from normality

depend both on genetic parameters and genealogical distributions. For instance, in exchangeable populations the expected third central moment is proportional to the third noncentral moment of the mutational distribution times the expected time to the first coalescent event in a sample of size three.

Finally, we discuss the consequences of these results for Q_{ST} tests and the inference of genetic parameters. We find an improved null distribution that can be derived simply by using the normal distribution that arises in the infinitesimal limit of our coalescent model. Additionally, we show that it is likely not possible to infer most features of the mutational distribution using only trait values sampled from a population. Future work will be necessary to develop tests for selection that take into account both demography and genetic parameters, but the model developed here provides the groundwork for such an undertaking.

3 Model

In the model we investigate here, there are L unlinked potentially causal loci at which mutations influence the trait value. Following KIMURA (1969), an infinite number of mutations are possible within each locus and the rate per unit of coalescent time at which mutations affecting the trait (causal mutations) arise is $\frac{\theta}{2}$. That is, $\frac{\theta}{2}$ is the mutation rate for the entire locus and not per nucleotide. An approximation for when at most one mutation per locus is likely is considered in Section 6. Mutations are randomly assigned effects from a distribution of effect sizes, and effects are additive both within and between loci. The moment generating function (mgf) of this distribution is written as ψ and the i^{th} non-central moment is m_i . Individuals are haploid and the sum of all mutations occurring in an individual's history determines the trait value of the individual. An extension to diploidy would be straightforward but is not considered here. Correlations between individuals arise when mutations fall on shared portions of genealogies at specific loci. Because the loci are unlinked we assume their genealogies are independent. This model is shown schematically in Figure 1.

The genealogy at a locus is represented by the random vector of branch lengths, \mathbf{T} . An element T_ω of \mathbf{T} is the branch length subtending only individuals in the set ω and no others in the sample. For example, $T_{\{a,b\}}$ is the length of the branch subtending only individuals a and b . If a branch subtending only a and b does not exist for a given genealogy, $T_{\{a,b\}}$ is set to zero. In this way \mathbf{T} encodes both the branch lengths and the topology of a genealogy. Ω is the set of all possible branches. If there are three sampled individuals, a , b , and c , then $\Omega = \{\{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}\}$ and $\mathbf{T} = (T_{\{a\}}, T_{\{b\}}, T_{\{c\}}, T_{\{a,b\}}, T_{\{a,c\}}, T_{\{b,c\}})$. The mgf for the distribution of branch lengths is denoted $\varphi_{\mathbf{T}}$.

Phenotypic trait values are the random quantities we are interested in and result from mutations occurring along the branches of genealogies. We will hereafter refer to the phenotypic trait simply as trait values and ignore any environmental component. The random vector of trait values in the sampled individuals is \mathbf{Y} . If we had sampled individuals a , b , and c , then $\mathbf{Y} = (Y_a, Y_b, Y_c)$. The contribution to the trait values from a single locus is the change relative to the value in the most recent common ancestor (MRCA) of the sample at that locus. Since we do not know the ancestral value, we cannot directly observe the change in trait values. Thus, for a trait controlled by multiple loci, \mathbf{Y} is the sum over contributions from these loci, each measured with respect to an arbitrary value. However, \mathbf{Y} is sufficient to determine measurable quantities such as differences in trait values between individuals as well as the sample variance. The moment generating functions for the distribution of trait values is denoted $\varphi_{\mathbf{Y}}$.

Here we refer to the genetic parameters of a trait as the combination of quantities not in-

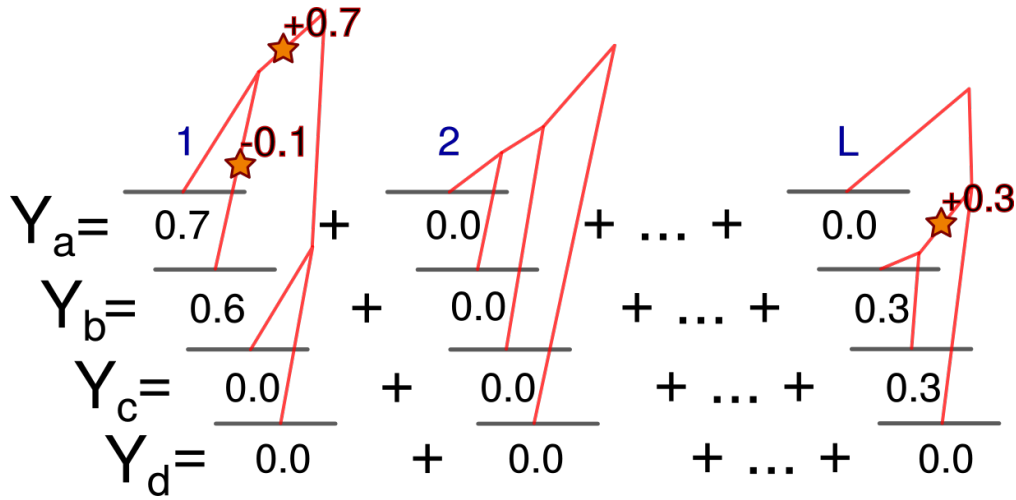


Figure 1: A schematic representation of the model for how trait distributions arise from genealogical and mutational processes. L loci potentially affect the trait in a set of individuals and have independent genealogies. Mutations occur within loci as a Poisson process and act additively to give individual trait values. Many loci with the potential to affect the trait may receive no mutations.

fluenced by the genealogical process: $(L, \frac{\theta}{2}, \psi)$. Another quantity useful for describing a trait's distribution is its sparsity. Sparsity should reflect how many mutations segregating in the population influence the trait, with a more 'sparse' trait being one affected by fewer segregating mutations. Formally, we measure sparsity as the average number of pairwise differences between two randomly chosen haplotypes at loci affecting the trait. A trait with fewer causal pairwise differences is more sparse. Sparsity thus depends both on the genetic parameters through the mutation rate, the number of potentially causal loci, and the distribution of coalescence times.

In populations of exchangeable individuals, a useful way to summarize the distribution of genealogies is through the moments of $\mathbb{T}_{k,n}$ which denotes the amount of time that k lineages remain in the genealogy of a sample of size n . The pairwise coalescent time between a lineage in individual i and in individual j is written as $\mathcal{T}_{i,j}$. When considering structured populations, $\mathcal{T}_{a,b}$ is also used to denote the coalescence time between a randomly chosen lineage from subpopulation a and a randomly chosen lineage from subpopulation b . A final set of quantities are defined for sums of branch lengths. Let τ_{a+b} be the sum of all branches ancestral to both a and b , and $\tau_{a/b}$ be the sum of all branches ancestral to a but not b . Extensions of this for more than two individuals are also used. The same notation is used when referring to sets of branch indices. So Ω_{a+b} and $\Omega_{a/b}$ would be the sets of branches added to get τ_{a+b} and $\tau_{a/b}$ respectively.

4 The moment generating function for the distribution of trait values

We first derive the mgf of the distribution of trait values following closely the approach of SCHRAIBER and LANDIS (2015) and KHAITOVICH *et al.* (2005), but generalizing to arbitrary demographies and population structure. We consider the distribution of trait values over evo-

lutionary realizations of the combined random processes of drift and mutation. The probability distribution for a trait is complex in its general form. There is a point mass at zero corresponding to the possibility that no mutations affecting the trait occur, and mutational effects could be drawn from discrete or continuous distributions. Correlations between individuals arise because of shared history in the genealogies at individual loci. An analytical expression for the probability distribution of trait values does not exist except in certain limits such as when the number of mutations affecting the trait becomes large.

However, even in the absence of a probability distribution function we can use the mgf approach to learn something about the distribution of trait values. Following the definition of the mgf for a vector-valued random variable, the mgf for a trait controlled by a single nonrecombining locus is

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \mathbb{E} [e^{\mathbf{k} \cdot \mathbf{y}}] = \int e^{\mathbf{k} \cdot \mathbf{y}} P(\mathbf{Y} = \mathbf{y}) d\mathbf{y}. \quad (1)$$

The vector \mathbf{k} contains dummy variables for each individual, and the whole operation is an integral transform of the probability distribution of trait values. Equation (1) can be rewritten by conditioning on the genealogy to give

$$\begin{aligned} \varphi_{\mathbf{Y}}(\mathbf{k}) &= \int_{\mathbf{Y}} e^{\mathbf{k} \cdot \mathbf{y}} \int_{\mathbf{T}} P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} d\mathbf{y} \\ &= \int \int e^{\mathbf{k} \cdot \mathbf{y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) d\mathbf{y} P(\mathbf{T} = \mathbf{t}) d\mathbf{t}. \end{aligned} \quad (2)$$

To proceed it is necessary to make assumptions about the mutational process. The first is that mutations occur as a Poisson process along branches and the second is that mutations at a locus are additive. Under these assumptions, the changes in the trait value along each branch are conditionally independent given the branch lengths. KHAITOVICH *et al.* (2005) and SCHRAIBER and LANDIS (2015) noted that this describes a compound Poisson process. The mgf of a compound Poisson process with rate λ over time t is $\exp(\lambda t(\psi(k) - 1))$, where ψ is the mgf of the distribution of the jump sizes caused by events in the Poisson process. In this case the jump sizes are the effects on the trait value caused by new mutations. Using this expression for the mgf of a compound Poisson process, along with the fact that the mgf of two perfectly correlated random variables with the same marginal distribution is $\varphi_{X_1}(k_1 + k_2)$, we can rewrite equation (2) as

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int \prod_{\omega \in \Omega} \exp\left(\frac{\theta}{2} t_{\omega} \left(\psi\left(\sum_{a \in \omega} k_a\right) - 1\right)\right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t}. \quad (3)$$

We can recognize equation (3) as the mgf of \mathbf{T} with the dummy variable s_{ω} for branch T_{ω} set to $\frac{\theta}{2} (\psi(\sum_{a \in \omega} k_a) - 1)$. Or,

$$\varphi_{\mathbf{T}}(\mathbf{s}) \Big|_{s_{\omega} = \frac{\theta}{2} (\psi(\sum_{a \in \omega} k_a) - 1)}. \quad (4)$$

Thus, if the mgf of the distribution of branch lengths is known, equation (3) allows us to obtain the mgf of the trait values through a simple substitution. When the trait is controlled by L independent loci, the full mgf, $\varphi_{\mathbf{Y}}(\mathbf{k})$, is obtained by raising equation (4) to the power L . This result obviates the need for separate derivations for particular models of population history and structure. LOHSE *et al.* (2011) derived the mgf of the genealogy in various population models including migration and splitting of subpopulations. Using their result for a single population we could obtain equation (1) of SCHRAIBER and LANDIS (2015) using equation (4).

5 The infinitesimal limit

This general model converges to a normal distribution when we take the infinitesimal limit. We accomplish this by first substituting a Taylor series for the genealogical and mutational distributions in equation (3) (see Appendix A). The infinitesimal limit corresponds to the situation where the effect sizes of mutations become small as the number of loci becomes large. The resulting distribution is multivariate normal where the expected trait value is $E[T_{MRC A}] \frac{\theta}{2} \mu_1$, the variance is $E[T_{MRC A}] \frac{\theta}{2} \mu_2$, and the covariance between trait values in two individuals a and b is $E[\tau_{a+b}] \frac{\theta}{2} \mu_2$. $E[T_{MRC A}]$ is the expected time to the most recent common ancestor in the sample or population. $\frac{\theta}{2} \mu_1$ is the rate per unit time per genome that mutational bias shifts the mean. $\frac{\theta}{2} \mu_2$ is the rate per unit time per genome that variance accumulates. This limit requires that the products of L and moments three and greater of the mutational distribution go to zero as the number of loci becomes large and the effect size per mutation becomes small. This can be thought of as requiring the mutational distribution to not have too heavy of tails. Details of the derivation are given in Appendix A.

Interestingly, the rate of variance accumulation is proportional to the second moment of the mutational distribution instead of the variance. We can see the intuition for this by considering a degenerate distribution where each mutation has the same effect. Here, we would still expect variation among individuals due to differences in the number of mutations each individual receives, even though the variance of the mutational distribution is zero. The variance among individual trait values thus has one component due to differences in the number of mutations and an additional component due to differences in the effects of these mutations. The first component is proportional to the square of the mean mutational effect, while the second is proportional to the mutational variance. Therefore, the sum of the two components is proportional to m_2 , the mean squared effect.

Since the trait values are normally distributed, any linear combination of sampled trait values will be as well. This includes the distributions of observable quantities like the differences in trait values from a reference individual or from a sample mean. The distribution of trait differences between individuals is multivariate normal with mean zero and covariance between any pair of trait differences given by

$$\text{Cov}[Y_i - Y_j, Y_k - Y_l] = \mu_2 \frac{\theta}{2} (E[\mathcal{T}_{i,l}] + E[\mathcal{T}_{j,k}] - E[\mathcal{T}_{i,k}] - E[\mathcal{T}_{j,l}]), \quad (5)$$

where $\mathcal{T}_{i,j} = 0$ if $i = j$. Classical theory in quantitative genetics uses a univariate normal distribution of phenotypes in a panmictic population. We can recover this by considering a population of exchangeable individuals. In this case $E[\mathcal{T}_{i,j}] = E[\mathbb{T}_{2,2}]$ for all pairs $i \neq j$. Individual trait values are then conditionally independent given the mean value in the population and are normally distributed with variance $E[\mathbb{T}_{2,2}] \frac{\theta}{2} \mu_2$.

The normal model in the infinitesimal limit provides additional theoretical justification for studies using normal models to look for differences in selection on quantitative traits between populations (OVASKAINEN *et al.*, 2011; PRAEBEL *et al.*, 2013; ROBINSON *et al.*, 2015). Additionally, equation (5) implies that a covariance matrix based on mean pairwise coalescent times rather than population split times should be used in phylogenetic models of neutral trait evolution.

6 Low-mutation-rate approximation

A useful simplification of the model considered so far is to ignore the possibility of more than one mutation per locus. This approximation is reasonable as long as the nucleotide positions affecting

the trait are loosely linked throughout the genome. The low-mutation-rate approximation greatly simplifies the mgf of the trait distribution such that it is no longer necessary to know the full form of the mgf of the genealogy:

$$\varphi_{\mathbf{Y}}(\mathbf{k}) \approx \left[1 + \sum_{\omega \in \Omega} \mathbb{E}[T_{\omega}] \frac{\theta}{2} \left(\psi \left(\sum_{a \in \omega} k_a \right) - 1 \right) + O(\theta^2) \right]^L. \quad (6)$$

Equation (6) ignores terms that are order two and above in the mutation rate. Conveniently, the mgf of the trait values depends only on the expected length of each branch, whereas equation (3) requires second order and greater moments of branch lengths (e.g. $\mathbb{E}[T_{\omega_1} T_{\omega_2}]$). We can use equation (6) to express moments of the trait distribution in terms of expected branch lengths calculated from coalescent models.

7 Moments of the trait distribution

For most population genetic models and for sample sizes greater than three, the recursive nature of the trait distribution mgf makes it computationally unfeasible to solve under general parameter values. However, it is not necessary to have an expression for the full mgf in order to derive moments of the trait distribution in terms of moments of branch lengths and mutational effects. Under the low mutation rate approximation moments can be calculated by differentiating equation (6). Even without making this approximation, moments can be calculated by taking Taylor expansions in equation (2) and only considering terms contributing to the desired moment's order. We implemented a symbolic math program to calculate trait moments using this procedure, and the details are given in Appendix B. As the normal distribution is completely defined by its first two moments, the extent to which a trait distribution deviates from normality can be measured by the extent to which its moments deviate from those of a normal distribution with the same mean and variance.

We have so far considered the distribution of a trait value Y_a over evolutionary realizations. The expectation of Y_a is $L \frac{\theta}{2} m_1 \mathbb{E}[T_{MRC A}]$, and the variance is $L \frac{\theta}{2} m_2 \mathbb{E}[T_{MRC A}] + L (m_1 \frac{\theta}{2})^2 \text{Var}[T_{MRC A}]$. Although simple to derive using computer algebra, expressions for the higher central moments of Y_a are complicated even under the low mutation rate approximation, and there is not much to be gained by showing them here.

However, in a given evolutionary realization there will be a distribution of trait values in the population. The population-level trait distribution can also be described by its moments, but since this distribution is random the moments at the population level are also random quantities. Since Y_a is relative to a value that is not directly observed, the expected population-level moments offer more insight. In particular, we are interested in how the trait distribution at the population level might deviate from normality. SCHRAIBER and LANDIS (2015) computed the expected first four central moments of a constant-size population. We derive the same expectations under an

arbitrary demographic history,

$$E[M_2] = L \frac{\theta}{2} E[\mathbb{T}_{2,2}] m_2 \quad (7a)$$

$$E[M_3] = L \frac{\theta}{2} E[\mathbb{T}_{3,3}] m_3 \quad (7b)$$

$$\begin{aligned} E[M_4] &= 3 \left(L \frac{\theta}{2} E[\mathbb{T}_{2,2}] m_2 \right)^2 \\ &+ 3L \left(\frac{\theta}{2} m_2 \right)^2 \text{Var}[\mathbb{T}_{2,2}] + \frac{1}{3} L \left(\frac{\theta}{2} m_2 \right)^2 \left(\frac{11}{9} E[\mathbb{T}_{2,4}^2] - \frac{1}{3} E[\mathbb{T}_{2,4} \mathbb{T}_{3,4}] - \frac{1}{4} E[\mathbb{T}_{3,4}^2] \right) \\ &+ L m_4 \frac{\theta}{2} (E[\mathbb{T}_{4,4}] + \frac{1}{3} E[\mathbb{T}_{3,4}] + \frac{2}{9} E[\mathbb{T}_{2,4}]). \end{aligned} \quad (7c)$$

The normal limit corresponds to equation (7b) and the second two lines of equation (7c) going to zero. To give insight into the expressions in equation (7), moment calculations done by hand under the low mutation rate approximation are presented in Appendix D.

Equation (7a) gives the expected trait variance in the population. However, the variance will vary over realizations of the evolutionary process. The variation in the population variance depends on the sparsity of the trait and the number of causal loci. The variation in the variance can be quantified using its coefficient of variation (CVV), the standard deviation of the variance divided by its expectation. For a constant-size, panmictic population

$$\text{CVV} = \sqrt{\frac{4}{3} \frac{1}{L} + \frac{1}{6} \frac{m_4/m_2^2}{L\theta E[\mathbb{T}_{2,2}]}}. \quad (8)$$

Equation (8) shows a contribution due to linked mutations occurring at a single locus ($\frac{4}{3} \frac{1}{L}$) and a contribution due to sparsity and mutation ($\frac{1}{6} \frac{m_4/m_2^2}{L\theta E[\mathbb{T}_{2,2}]}$). Even when the sparsity is low, i.e., when a large number of variants affect the trait, if the trait is only controlled by a single locus there will be considerable variation in the population variance ($\text{CVV} = \sqrt{\frac{4}{3}}$). On the other hand, the CVV of a sparse trait controlled by many loci will depend on the ratio of the fourth and squared second non-central moments of the mutational distribution (m_4/m_2^2).

8 Comparison to normal distribution

Deviations of the population distribution from normality depend on the distribution of coalescent times and the genetic parameters. One natural way to quantify a distribution's deviation from normality is through its kurtosis. The kurtosis, defined as M_4/M_2^2 , measures the tendency of a distribution to produce outliers (WESTFALL, 2014). However, since the kurtosis is a ratio of two random variables, its expectation is challenging to calculate. Rather than attempt this calculation, we compare the expected fourth central moment itself to that expected under normality. This approach qualitatively identifies the factors influencing deviations from normality.

Using the low mutation rate approximation, the ratio of the expected fourth moment to that under normality simplifies to

$$\frac{E[M_4]}{3 \left(L \frac{\theta}{2} E[\mathbb{T}_{2,2}] m_2 \right)^2} \approx 1 + \underbrace{\frac{m_4}{m_2^2}}_{\text{mutation}} \underbrace{\left(6L \frac{\theta}{2} E[\mathbb{T}_{2,2}] \right)^{-1}}_{\text{sparsity}} \underbrace{\left(\frac{2E[\mathbb{T}_{4,4}] + \frac{2}{3} E[\mathbb{T}_{3,4}] + \frac{4}{9} E[\mathbb{T}_{2,4}]}{E[\mathbb{T}_{2,2}]} \right)}_{\text{demography}} \quad (9)$$

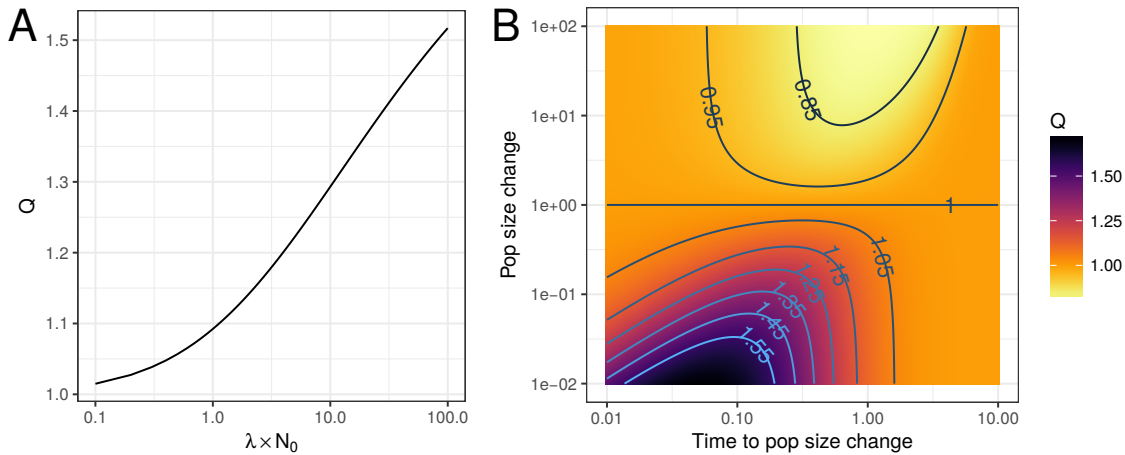


Figure 2: The effects of demography on deviations of the expected fourth central moment of the population trait distribution from normality. Q measures the effect due to demography on the expected fourth central moment (equation (9)). **(A):** Q increases as the exponential growth rate increases relative to the current population size. λ is the growth rate and N_0 is the initial effective population size. **(B):** Q values when the population undergoes an instantaneous step change at some point in the past. The time and magnitude of this change are given in units of the current effective population size. Q increases when the population grows and decreases when it declines.

The expected excess in M_4 , relative to normality, increases with sparsity and with m_4/m_2^2 . m_4/m_2^2 is equal to the kurtosis of the mutational effect distribution when mutations are unbiased and reflects its propensity to produce large effect mutations. The excess depends on demography through the factor $Q = \frac{2E[\mathbb{T}_{4,4}] + \frac{2}{3}E[\mathbb{T}_{3,4}] + \frac{4}{9}E[\mathbb{T}_{2,4}]}{E[\mathbb{T}_{2,2}]}$. The extent to which demography increases or decreases deviations from normality can be investigated by calculating Q in different models. In a constant-size and panmictic population Q is equal to one. In a population where lineages are exchangeable, $E[\mathbb{T}_{k,n}]$ can be calculated numerically using expressions from GRIFFITHS and TAVARÉ (1998) or POLANSKI and KIMMEL (2003). Values of Q in an exponentially growing population are shown in Figure 2A. Holding sparsity and the mutational distribution constant, a population which underwent exponential growth will have a greater expected deviation from normality in its trait distribution. Another example demography is a population that goes through a step change at some point in the past. Figure 2B shows that when the population size increased at some point in the past, Q is increased similarly to the exponential growth scenario. When the population experiences a bottleneck Q is decreased below one. Additionally, Q appears more sensitive to population growth than to bottlenecks.

As a concrete example, we can consider the differences in the expected fourth moment produced by different demographic histories in different human populations. In the demographic model fit by TENNESSEN *et al.* (2012), the generic European population experiences a bottleneck associated with out-of-Africa and recent growth while the generic African population experiences a more stable history also with recent growth. Differences in demographic history between the two populations has resulted in a lower heterozygosity in European populations due to the out-of-Africa bottleneck (YU *et al.*, 2002).

For a given sparsity, the African population model predicts a smaller deviation from normality

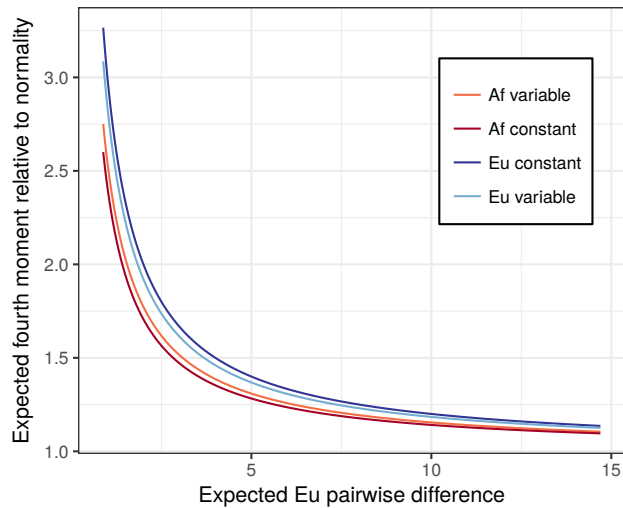


Figure 3: **A comparison between the expected fourth moment at different levels of sparsity in the African and European demographic models fit by TENNESSEN *et al.* (2012).** Trait sparsity is varied by changing the expected number pairwise differences at sites affecting the trait in the European model. The mutational kurtosis is set to six. The darker lines show the predicted relationship for populations with the same heterozygosity as the European and African models but with constant size.

than the European model (Figure 3). The expected fourth moment in constant-size populations with the same heterozygosity as the African and European models is lower for the African model and higher for the European model. This is because the African model is dominated by population growth that leads to a Q greater than one, while the European model is dominated by a bottleneck event that leads to a Q less than one (Figure 2). However, differences due to demography are small and deviations from normality are mostly driven by differences in heterozygosity at causal loci.

Even though recombination is not included in the model, we can form an idea about how linkage might impact deviations from normality. Line two of equation (7c) corresponds to the contribution from two mutations occurring at a single locus. The first quantity indicates that the expectation of the fourth moment increases with the variance of the pairwise coalescence time. The second part does not have a clear interpretation. If the sum of these two terms is positive this agrees with the intuition that linkage disequilibrium increases deviations from normality by reducing the effective number of independent loci.

Simulations of the kurtosis itself show substantial variance, with about a quarter of simulated populations having a value less than three (that of a normal distribution) even as the mean kurtosis increases to almost nine (Appendix C). This high variance in the kurtosis is likely due to a high variance in both the trait variance and fourth moment. This, along with the fact that deviations from the infinitesimal model inflate the fourth moment (equation (9)), leads to a situation where the kurtosis increases with trait sparsity but the variance is high across evolutionary realizations.

9 Trait divergence in structured populations

To demonstrate the utility of the normal distribution arising in the infinitesimal limit (equation 5), we derive a simple procedure for simulating from null distribution for the divergence between groups in structured populations. A common way to quantify the divergence in trait values between groups is Q_{ST} , defined as the variance between the group means divided by the total variance in the population (SPITZE, 1993). In a haploid model, $Q_{ST} = \frac{V_{between}}{V_{between} + V_{within}}$, where

$$V_{between} = \frac{1}{K} \sum_i (\bar{Y}_i - \bar{Y})^2$$

and

$$V_{within} = \frac{1}{\sum_k N_k} \sum_i \sum_j (Y_{i,j} - \bar{Y}_i)^2.$$

Here, $Y_{i,j}$ is the trait value of individual j in population i , K is the number of subpopulations, and N_k is the size of subpopulation k .

In the normal model, all $Y_{i,j}$ are normally distributed. Therefore, $\bar{Y}_i - \bar{Y}$ and $Y_{i,j} - \bar{Y}_i$ are also normally distributed. When population sizes are large, individual deviations from population means are nearly uncorrelated, as are $V_{between}$ and V_{within} . V_{within} is nearly constant across evolutionary realizations and is equal to $\sum_k N_k E[\mathcal{T}_{k,k}] / \sum_k N_k$ because the within population variances are approximately uncorrelated and their variances are order $1/N_k$. While we do not have an explicit form for the density function of the between-group variance, we can simulate from its distribution by drawing a vector of $\bar{Y}_i - \bar{Y}$ values from a multivariate normal distribution with mean zero and with a covariance matrix whose element between populations a and b is

$$\text{Cov}[\bar{Y}_a - \bar{Y}, \bar{Y}_b - \bar{Y}] = \mu_2 \frac{\theta}{2} (E[\mathcal{T}_{a,\cdot}] + E[\mathcal{T}_{b,\cdot}] - E[\mathcal{T}_{\cdot,\cdot}] - E[\mathcal{T}_{a,b}]). \quad (10)$$

To simulate Q_{ST} values we do not need to know μ_2 or $\frac{\theta}{2}$ because the scale of the trait variance cancels in the Q_{ST} ratio. Therefore, all that is necessary to simulate from the null distribution of Q_{ST} is to have estimates of the expected coalescent time within and between populations. No further information about the history or structure of the population is needed. Since only the relative coalescent times matter it is also not necessary to scale these estimates to units of years or generations using a mutation rate.

This procedure for simulating from the null distribution of Q_{ST} could be useful in testing whether an observed Q_{ST} is unlikely under neutrality. Current goodness-of-fit tests either compare Q_{ST} to an empirical distribution of F_{ST} values or to a χ^2 distribution (LEINONEN *et al.*, 2013). In the second case, an identical distribution to that developed by LEWONTIN and KRAKAUER (1973) is used as the null distribution for Q_{ST} . The χ^2 testing procedure was suggested by WHITLOCK and GUILLAUME (2009) and is implemented in the program *QstFstComp* (GILBERT and WHITLOCK, 2015). The Lewontin-Krakauer (LK) distribution assumes independence between subpopulations and provides a good approximation in populations without spatial structure, such as in a symmetric island model of migration (Figure 4). When subpopulations are strongly correlated, such as in a one-dimensional stepping-stone model of population structure, the LK distribution is a poor approximation. Even when the distributions appear qualitatively similar, there are substantial differences in tail probabilities (Figure 5).

Nearly identical issues to these were raised with the LK test at the time of its publishing (NEI and MARUYAMA, 1975; ROBERTSON, 1975). However, while the neutral distribution for F_{ST} depends on the precise details of population structure and history, the distribution of Q_{ST}

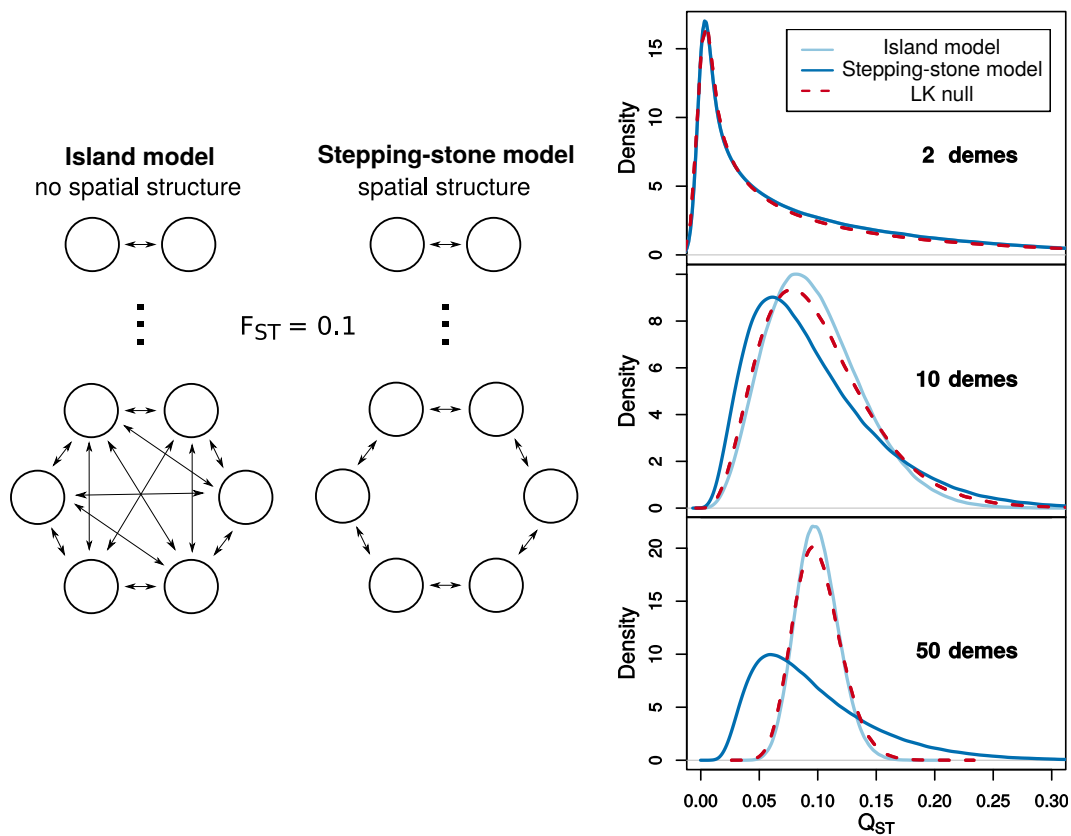


Figure 4: **A comparison of neutral sampling distributions for Q_{ST} at the population level.** The Lewontin-Krakauer (LK) distribution for Q_{ST} is compared to the null distribution in the infinitesimal limit in a case with (the stepping-stone model) and a case without (the island model) spatial structure. The case with no spatial structure assumes the migration rate is equal between all subpopulations, and the case with spatial structure arranges subpopulations in a ring with migration only between neighboring subpopulations. Migration rates and subpopulation sizes are all equal and are set such that $F_{ST} = 0.1$ even as the number of subpopulations is increased (SLATKIN, 1991). Q_{ST} values for these models are simulated by drawing vectors from a multivariate normal distribution parameterized using the expected pairwise coalescence times given by (SLATKIN, 1991). Under the LK distribution Q_{ST} is distributed as $F_{ST}/(n_d - 1)$ times a chi-square distribution with $n_d - 1$ degrees of freedom ($Q_{ST} \sim F_{ST}\chi_{n_d-1}^2/(n_d - 1)$), where n_d is the number of subpopulations. Vertical lines show the mean Q_{ST} under the different null distributions. The mean Q_{ST} values for the two normal models are nearly identical. Discordance between these lines and that for the LK distribution illustrates that $Q_{ST} \neq F_{ST}$. The neutral distributions for Q_{ST} become increasingly different as the number of subpopulations over which the genetic divergence is spread increases.

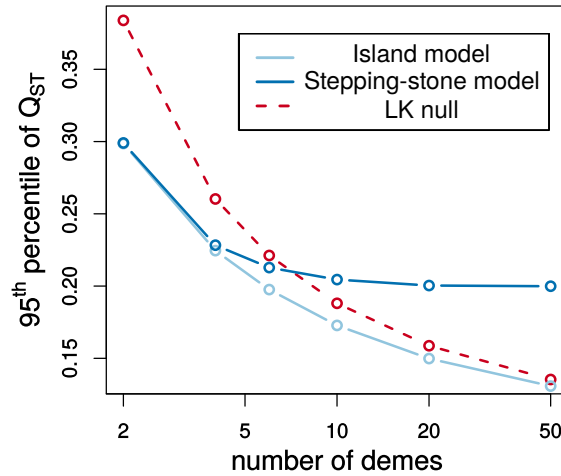


Figure 5: **Differences in the 95th percentile of different neutral sampling distributions for Q_{ST} at the population level.** The Lewontin-Krakauer (LK) distribution is compared to neutral null distributions for structured populations with and without a spatial component using the multivariate normal model that arises in the infinitesimal limit.

only depends on the set of within and between subpopulation coalescence times. The point here is not that the LK distribution is particularly bad, but rather that an improved neutral distribution can be obtained using the normal model arising in the infinitesimal limit. The neutral distribution described here is similar to the extension of the LK test developed by BONHOMME *et al.* (2010) to account for the correlation structure between subpopulations. The BONHOMME *et al.* (2010) method treats allele frequencies as multivariate normal with covariance matrix parameterized by coancestry coefficients. OVASKAINEN *et al.* (2011) use a normal model similar to that found in the infinitesimal limit here, but the covariance matrix is also based on coancestry coefficients. When phenotypic and genetic divergence is mostly driven by changes in allele frequency, the coalescent and coancestry based models should be very similar. However, the coalescent model is ultimately preferable since it is the correct null model at any scale of population divergence in the infinitesimal limit. When only allele frequency data are available, a coancestry model is the only option, but it is still better to model variable levels of shared ancestry between populations than to use a single value of F_{ST} .

Additionally, treating Q_{ST} as a random variable lets us reexamine the classic result in evolutionary quantitative genetics that $Q_{ST} = F_{ST}$ (WHITLOCK, 1999). F_{ST} , in this context, refers to a parameter of the population. In particular, $F_{ST} = \frac{\bar{t} - \bar{t}_0}{\bar{t}}$, where \bar{t} is the expected coalescent time for two loci sampled at random from the entire population and \bar{t}_0 is the expected coalescent time for two loci sampled within a subpopulation (SLATKIN, 1991). This value is constant over realizations of the evolutionary process. Q_{ST} can only refer to either state of the population or to an estimate of this state. As shown above, Q_{ST} , as a state of the population, varies across evolutionary realizations. Thus, there is no sense in which Q_{ST} can be defined as a constant parameter in the way that F_{ST} can. The expectation of $V_{between}$ is $\bar{t} - \bar{t}_0$, and the expectation of V_{within} is \bar{t}_0 . $\frac{E[V_{between}]}{E[V_{between}] + E[V_{within}]}$ is equal to F_{ST} , but due to Jensen's inequality, the expectation of this ratio ($E[Q_{ST}]$) is always less than F_{ST} .

10 Inferring genetic parameters

SCHRAIBER and LANDIS (2015) suggested it might be possible to infer the shape of the distribution of mutational effects through its moments for sparse traits whose distributions deviate from normality. Using the expressions for the expected moments of the trait value distribution in equation (7), we could try and design a method of moments estimator for the moments of the mutational distribution. If $\hat{D}_{k,n}$ is an estimator of $\frac{\theta}{2}E[\mathbb{T}_{k,n}]$, the system of equations for the first three central moments under the low mutation rate approximation is

$$\begin{aligned}\widehat{M}_2 &= \widehat{D}_{2,2}Lm_2 \\ \widehat{M}_3 &= \frac{1}{3}\widehat{D}_{3,3}Lm_3 \\ \widehat{M}_4 &= 3\widehat{M}_2^2 + (\widehat{D}_{4,4} + \frac{1}{3}\widehat{D}_{3,4} + \frac{2}{9}\widehat{D}_{2,4})Lm_4.\end{aligned}\tag{11}$$

From equation (11) we can see that the moments of the trait distribution only enter through products with the number of loci potentially affecting the trait (Lm_i). If these products were estimated as \widehat{Lm}_2 , \widehat{Lm}_3 , and \widehat{Lm}_4 , it would be possible to estimate the ratios m_2/m_4 and m_3/m_4 of the moments of the mutational distribution. These ratios are meaningless on their own because any value could be obtained by changing the scale on which the trait is measured. The quantity that is identifiable in this system of equation is the compound parameter $\frac{m_3^2}{m_2m_4}$. This quantity reflects something about the mutational bias relative to the spread of the distribution. However, it is likely not possible to distinguish sparsity from skewed or heavy-tailed distributions of mutational effects.

11 Discussion

Neutral models of quantitative trait evolution are important for establishing a baseline against which to test for selection. SCHRAIBER and LANDIS (2015) recently analyzed a neutral model of trait evolution that made few assumptions about the number of loci potentially affecting the trait and the distribution of mutational effects at these loci. However, they only derived results for constant-size, panmictic populations. We extend their results to populations with arbitrary distribution of coalescent times and therefore varying demographies and population structures. As their key result, SCHRAIBER and LANDIS (2015) derived the characteristic function of the distribution of trait values in a sample. In this paper, we instead work with the moment generating function, but the two approaches are interchangeable as long as the mgf for the distribution of mutational effects exists (which it will if the moments of the mutational distribution are all finite). Our main result, given in equation (4), shows that the characteristic function from SCHRAIBER and LANDIS (2015) is a special outcome of a general procedure whereby the moment generation function for a trait distribution can be obtained by making a simple substitution into the moment generating function for a distribution over genealogies. The moment generating functions for many demographies of interest and for sample sizes above three are sufficiently complex that solving for them is impractical (LOHSE *et al.*, 2011). However, progress can still be made by using Taylor expansions to write moments of the trait distribution in terms of moments of the genealogical and mutational distributions.

This result extends previous work using coalescent theory to investigate neutral models of quantitative traits (WHITLOCK, 1999; SCHRAIBER and LANDIS, 2015). Ours is the most general model yet analyzed. As a natural first step, we show that the infinitesimal limit suggested by

FISHER (1918) leads to a model where phenotypes are normally distributed as the number of loci becomes large and the variance of effect sizes becomes small. In the limiting distribution, the variance of the difference in trait values between two individual is proportional to the expected pairwise coalescent time between them, and the covariance between a pair of differences more generally is $\text{Cov}[Y_a - Y_b, Y_c - Y_d] \propto E[\mathcal{T}_{a,d}] + E[\mathcal{T}_{b,c}] - E[\mathcal{T}_{a,c}] - E[\mathcal{T}_{b,d}]$. The resulting covariance matrix completely specifies the neutral distribution under the infinitesimal model. This is similar to classic models in evolutionary quantitative genetics considering the neutral divergence of trait values after population splits (LANDE, 1976; LYNCH, 1989) but holds regardless of the precise details of population structure and history. SCHRAIBER and LANDIS (2015) derive essentially the same distribution using a central limit theorem argument.

MENDES *et al.* (2018) point out several problems caused by incomplete lineage sorting when species split times are used to generate the covariance matrix in phylogenetic comparative methods. Since the normal model arising in the infinitesimal limit is explicitly based on the coalescent, it is not subject to these problems. A matrix based on pairwise coalescence times rather than a species tree based on population split times takes into account the effects of all lineages at causal loci, even those that do not follow the species topology. The covariances specified by equation (5) could also be used to generate within-species contrasts in a similar manner to the method suggested by FELSENSTEIN (2002).

We compared sparse traits to the normal model by calculating how the first four expected central moments different from those expected under normality. This showed how demography and genetics separately influence the expected deviation from normality. For a fixed expected trait sparsity, population growth produces greater deviations in the fourth central moment while population bottlenecks produce lower deviations (Figures 2). However, for realistic demographic scenarios, we find that the effects attributable to demography are small (Figure 3). We only analyzed cases where the individuals in the population were exchangeable, but adding population structure would increase deviations from normality as drifting trait means between subpopulations will yield a multimodal distribution.

We next apply the above theory to two simple problems where a coalescent perspective on the neutral distribution of a quantitative trait provides useful intuition. The first of these is the question of the appropriate null distribution for Q_{ST} at the population level. We show how the null distribution under the normal model can be easily simulated from by taking advantage of the theory presented here, providing a much better approximation when populations are correlated than previous approaches (WHITLOCK and GUILLAUME, 2009) (Figure 4). In the second we show that the shape of the mutational distribution is largely confounded with the number of loci affecting the trait, with only one compound parameter identifiable. This makes it unlikely that mutational parameters could be inferred from trait values sampled from a population.

Even though we have broadened the model space for neutral traits, many features of real populations have not yet been incorporated. Linkage between loci is a particular concern as there is substantial linkage disequilibrium between causal loci (BULIK-SULLIVAN *et al.*, 2015). LOHSE *et al.* (2011) derived the form of the moment generating function for linked loci and future work will attempt to incorporate this using equation (4). In particular, it will be important for future work to show how this affects the distribution in the infinitesimal limit. Diploidy, dominance, and epistasis have also been ignored thus far. The qualitative effects described here should hold under diploidy, but having trait values within individuals summed over loci from two copies of the genome will decrease deviations from normality. Dominance will also tend to produce a normal distribution as the effects are independent between loci, but future work is needed to examine how this interacts with the distribution of genealogies to affect the trait distribution.

BARTON *et al.* (2017) recently performed a deep mathematical investigation of a more formal

“infinitesimal model” that is subtly different from the infinitesimal limit considered here. They proved conditions under which the trait values of offspring within a family are normally distributed with variance independent of the parental trait values conditional on the pedigree and segregation variance in the base population. Interestingly, they found the normality for offspring trait values still holds under some forms of pairwise epistasis that are not too extreme. This implies it may be possible to include epistasis in the infinitesimal limit considered here. It would be important to know how epistasis affects the neutral divergence of trait between populations and species.

Although GWAS of many traits have shown them to be controlled by large numbers of loci (BOYLE *et al.*, 2017), this will not necessarily be the case for every trait of interest to biologists. It has been suggested, for instance, that gene expression levels have a sparse genetic architecture (WHEELER *et al.*, 2016). Since there is much interest in testing whether natural selection has acted on gene expression levels (WHITEHEAD and CRAWFORD, 2006; GILAD *et al.*, 2006; YANG *et al.*, 2017), well-calibrated goodness-of-fit tests will need to take into account the complications that arise when trait distributions deviate from normality (KHAITOVICH *et al.*, 2005). Direct measurements of mutational distributions (GRUBER *et al.*, 2012; METZGER *et al.*, 2016) could aid in such a calibration. Finally, equation (7) suggests a means to determine whether sparsity impacts trait distributions by examining populations with different demographic histories. Populations with a greater $\mathbb{T}_{3,3}$ to $\mathbb{T}_{2,2}$ ratio are also expected to show a greater M_3 to M_2 ratio and an equivalent relationship would be expected for the fourth moment. Because gene expression studies generally measure a large number of traits, observing such a trend on average could be a sign of deviations from normality. Generally, more work is needed to develop neutrality tests that are robust to the details of genetics and population history, and to investigate whether any information about mutation can be uncovered by statistical models that share information across multiple measured traits.

Acknowledgements

We thank John Novembre, Daniel Rice, and members of the Novembre lab for helpful discussions. We would also like to thank Josh Schraiber, Daniel Rice, John Novembre, Maryn Carlson, Hussein Al-Asadi, and Joe Marcus for their comments on the manuscript. This work was supported by an NSF GRFP to the author and NIH grant GM108805 to John Novembre.

References

- BARTON, N., A. ETHERIDGE, and A. VÉBER, 2017 The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* **118**: 50–73.
- BERG, J. J., and G. COOP, 2014 A Population Genetic Signal of Polygenic Adaptation. *PLoS Genetics* **10**.
- BONHOMME, M., C. CHEVALET, B. SERVIN, S. BOITARD, J. ABDALLAH, *et al.*, 2010 Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics* **186**: 241–262.
- BOYLE, E. A., Y. I. LI, and J. K. PRITCHARD, 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**: 1177–1186.

- BULIK-SULLIVAN, B. K., P.-R. LOH, H. K. FINUCANE, S. RIPKE, J. YANG, *et al.*, 2015 LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**: 291–295.
- CHAKRABORTY, R., and M. NEI, 1982 Genetic Differentiation of Quantitative Characters Between Populations of Species. *Genetics Research* **39**: 303–314.
- FELSENSTEIN, J., 1985 Phylogenies and the comparative method. *American Naturalist* **125**: 1–15.
- FELSENSTEIN, J., 2002 Contrasts for a Within-Species Comparative Method. In M. Slatkin and M. Veuille, editors, *Modern Developments in Theoretical Population Genetics*. Oxford University Press, 118–129.
- FISHER, R. A., 1918 XV. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **52**: 399–433.
- FRECKLETON, R. P., P. H. HARVEY, and M. D. PAGEL, 2002 Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* **160**: 712–726.
- GILAD, Y., A. OSHLACK, and S. A. RIFKIN, 2006 Natural selection on gene expression. *Trends in Genetics* **22**: 456–461.
- GILBERT, K. J., and M. C. WHITLOCK, 2015 QST-FST comparisons with unbalanced half-sib designs. *Molecular Ecology Resources* **15**: 262–267.
- GRIFFITHS, R., and S. TAVARÉ, 1998 The age of a mutation in a general coalescent tree. *Communications in Statistics. Stochastic Models* **14**: 273–295.
- GRISWOLD, C. K., B. LOGSDON, and R. GOMULKIEWICZ, 2007 Neutral evolution of multiple quantitative characters: a genealogical approach. *Genetics* **176**: 455–66.
- GRUBER, J. D., K. VOGEL, G. KALAY, and P. J. WITTKOPP, 2012 Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *saccharomyces cerevisiae*: Frequency, effects, and dominance. *PLoS Genetics* **8**: e1002497.
- KELLEHER, J., A. M. ETHERIDGE, and G. MCVEAN, 2015 Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol* **12**: e1004842.
- KHAI TOVICH, P., S. PÄÄBO, and G. WEISS, 2005 Toward a Neutral Evolutionary Model of Gene Expression. *Genetics* **170**: 929–939.
- KIMURA, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* : 893–903.
- LANDE, R., 1976 Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* **30**: 314–334.
- LANDE, R., 1992 Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* **46**: 381–389.
- LEINONEN, T., R. J. MCCAIRNS, R. B. O’HARA, and J. MERILÄ, 2013 QST -FST comparisons: Evolutionary and ecological insights from genomic heterogeneity.

- LEWONTIN, R., and J. KRAKAUER, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **73**: 175–195.
- LOHSE, K., R. J. HARRISON, and N. H. BARTON, 2011 A general method for calculating likelihoods under the coalescent process. *Genetics* **189**: 977–87.
- LYNCH, M., 1989 Phylogenetic Hypotheses Under the Assumption of Neutral Quantitative- Genetic Variation. *Evolution* **43**: 1–17.
- LYNCH, M., and W. G. HILL, 1986 Phenotypic Evolution by Neutral Mutation. *Evolution* **40**: 915–935.
- MENDES, F. K., J. A. FUENTES-GONZÁLEZ, J. G. SCHRAIBER, and M. W. HAHN, 2018 A multispecies coalescent model for quantitative traits. *eLife* **7**.
- METZGER, B. P., F. DUVEAU, D. C. YUAN, S. TRYBAN, B. YANG, *et al.*, 2016 Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression. *Molecular Biology and Evolution* **33**: 1131–1146.
- MEURER, A., C. P. SMITH, M. PAPROCKI, O. ČERTÍK, S. B. KIRPICHEV, *et al.*, 2017 SymPy: symbolic computing in python. *PeerJ Computer Science* **3**: e103.
- NEI, M., and T. MARUYAMA, 1975 Letters to the editors: Lewontin-Krakauer test for neutral genes. *Genetics* **80**: 395.
- OVASKAINEN, O., M. KARHUNEN, C. ZHENG, J. M. C. ARIAS, and J. MERILÄ, 2011 A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics* **189**: 621–632.
- POLANSKI, A., and M. KIMMEL, 2003 New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. *Genetics* **165**: 427–436.
- PRAEBEL, K., R. KNUDSEN, A. SIWERTSSON, M. KARHUNEN, K. K. KAHILAINEN, *et al.*, 2013 Ecological speciation in postglacial European whitefish: rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution* **3**: 4970–4986.
- RICE, S., 2004 *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer.
- ROBERTSON, A., 1975 Letters to the editors: Remarks on the Lewontin-Krakauer test. *Genetics* **80**: 396.
- ROBINSON, M. R., G. HEMANI, C. MEDINA-GOMEZ, M. MEZZAVILLA, T. ESKO, *et al.*, 2015 Population genetic differentiation of height and body mass index across Europe. *Nature Genetics* **47**: 1357–1362.
- ROGERS, A., and H. HARPENDING, 1983 Population Structure and Quantitative Characters. *Genetics* **105**: 985–1002.
- SCHRAIBER, J. G., and M. J. LANDIS, 2015 Sensitivity of quantitative traits to mutational effects and number of loci. *Theoretical population biology* **102**: 85–93.
- SLATKIN, M., 1991 Inbreeding coefficients and coalescence times. *Genetical research* **58**: 167–175.

- SPITZE, K., 1993 Population structure in *Daphnia obtusa*: Quantitative genetic and allozymic variation. *Genetics* **135**: 367–374.
- TENNESSEN, J. A., A. W. BIGHAM, T. D. O’CONNOR, W. FU, E. E. KENNY, *et al.*, 2012 Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* (New York, N.Y.) **337**: 64–9.
- TURELLI, M., 2017 Commentary: Fisher’s infinitesimal model: A story for the ages. *Theoretical Population Biology* **118**: 46–49.
- TURELLI, M., J. H. GILLESPIE, and R. LANDE, 1988 Rate Tests for Selection on Quantitative Characters During Macroevolution and Microevolution. *Evolution* **42**: 1085–1089.
- WAKELEY, J., 2008 *Coalescent Theory: An Introduction*. W. H. Freeman.
- WESTFALL, P. H., 2014 Kurtosis as Peakedness, 1905-2014. R.I.P. *The American Statistician* **68**: 191–195.
- WHEELER, H. E., K. P. SHAH, J. BRENNER, T. GARCIA, K. AQUINO-MICHAELS, *et al.*, 2016 Survey of the Heritability and Sparse Architecture of Gene Expression Traits across Human Tissues. *PLOS Genetics* **12**: e1006423.
- WHITEHEAD, A., and D. L. CRAWFORD, 2006 Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences* **103**: 5425–5430.
- WHITLOCK, M. C., 1999 Neutral additive genetic variance in a metapopulation. *Genetical Research* **74**: 215–221.
- WHITLOCK, M. C., 2008 Evolutionary inference from QST. *Molecular Ecology* **17**: 1885–1896.
- WHITLOCK, M. C., and F. GUILLAUME, 2009 Testing for spatially divergent selection: Comparing QST to FST. *Genetics* **183**: 1055–1063.
- YANG, J.-R., C. J. MACLEAN, C. PARK, H. ZHAO, and J. ZHANG, 2017 Intra and Interspecific Variations of Gene Expression Levels in Yeast Are Largely Neutral: (Nei Lecture, SMBE 2016, Gold Coast). *Molecular Biology and Evolution* **34**: 2125–2139.
- YU, N., F. C. CHEN, S. OTA, L. B. JORDE, P. PAMILO, *et al.*, 2002 Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**: 269–274.

A A central limit theorem in the infinitesimal limit

Recall that the moment generating function for the distribution of trait values from a single locus is

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int \exp\left(\sum_{\omega \in \Omega} s_{\omega} t_{\omega}\right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \Big|_{s_{\omega} = \frac{\theta}{2} (\psi(\sum_{a \in \omega} k_a) - 1)}.$$

If we substitute in the Taylor series expansions for the moment generating function of the trait value distribution we get

$$\int \prod_{\Omega} \exp\left[t_{\omega} \frac{\theta}{2} \left(\sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a\right)^n\right)\right] P(\mathbf{T} = \mathbf{t}) d\mathbf{t}.$$

If we then write the Taylor series of each exponential function we get

$$\int \prod_{\Omega} \left[\sum_{j=0}^{\infty} \frac{t_{\omega}^j}{j!} \left(\frac{\theta}{2}\right)^j \left(\sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a\right)^n\right)^j \right] P(\mathbf{T} = \mathbf{t}) d\mathbf{t},$$

which is equivalent to

$$1 + \sum_{\Omega} E[T_{\omega}] \frac{\theta}{2} \sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a\right)^n + \sum_{\Omega \times \Omega} \frac{1}{2} E[T_{\omega_1} T_{\omega_2}] \sum_{n=1}^{\infty} \frac{\theta}{2} \frac{m_n}{n!} \left(\sum_{a \in \omega_1} k_a\right)^n \sum_{n=1}^{\infty} \frac{\theta}{2} \frac{m_n}{n!} \left(\sum_{a \in \omega_2} k_a\right)^n + \dots$$

This is raised to the power L for a trait controlled by L loci. We want the limit as the number of loci increases while the size of mutational decreases. This can be expressed by the limits $Lm_1 \rightarrow \mu_1$, $Lm_2 \rightarrow \mu_2$ and $Lm_i \rightarrow 0$ for $i > 2$ as $L \rightarrow \infty$. Knowing we will not be retaining m_3 and above we can rewrite the mgf as

$$\left(1 + \sum_{\Omega} E[T_{\omega}] \frac{\theta}{2} \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2} \left(\sum_{a \in \omega} k_a\right)^2\right)\right)^L.$$

The result of taking these limits is

$$\exp\left(\sum_{\omega \in \Omega} E[T_{\omega}] \frac{\theta}{2} \left(\mu_1 \sum_{a \in \omega} k_a + \frac{\mu_2}{2} \left(\sum_{a \in \omega} k_a\right)^2\right)\right). \quad (12)$$

This is multivariate normal distribution with mean equal to $E[T_{MRCAs}] \frac{\theta}{2} \mu_1$, variance equal to $E[T_{MRCAs}] \frac{\theta}{2} \mu_2$, and covariance between Y_a and Y_b equal to $E[\tau_{a+b}] \frac{\theta}{2} \mu_2$. This can be seen from equation (12) by noting that in the mgf for a multivariate normal distribution the coefficient in the exponential of k_a is the mean of Y_a and the coefficient of $k_a k_b$ is $2\text{Cov}[Y_a, Y_b]$ if $a \neq b$ and $\text{Var}[Y_a]$ if $a = b$.

These Y values are not directly observed because in the theory presented here they are measured as the sum of differences since the T_{MRCAs} at the causal loci affecting the trait. Rather, differences between individual trait values are what analyses would be based on. Since the Y are normally distributed differences between trait values such as $Y_a - Y_b$ will be as well.

$$\begin{aligned} \text{Var}[Y_a - Y_b] &= \text{Var}[Y_a] + \text{Var}[Y_b] - 2\text{Cov}[Y_a, Y_b] \\ &= 2(E[T_{MRCAs}] - E[\tau_{a+b}]) \mu_2. \end{aligned}$$

This also tells us that the distribution of trait differences will be the same in the infinitesimal limit with and without a low mutation rate approximation.

A much simpler heuristic derivation of the limiting normal distribution can be done by calculating the variance and covariance at a single locus. This derivation is very similar to that done by SCHRAIBER and LANDIS (2015). Using the law of total variance we can write

$$\text{Var}[Y] = \text{E}[\text{Var}[Y|T]] + \text{Var}[\text{E}[Y|T]]$$

The variance conditional on T can be calculated again using the law of total variance and conditioning on the number of mutation at the locus.

$$\begin{aligned} \text{Var}[Y|T] &= \text{E}[\text{Var}[Y|M]|T] + \text{Var}[E[Y|M]|T] \\ &= \text{E}[M(m_2 - m_1^2)|T] + \text{Var}[Mm_1|T] \\ &= \frac{\theta}{2}T(m_2 - m_1^2) + \frac{\theta}{2}Tm_1^2 \\ &= \frac{\theta}{2}Tm_2 \end{aligned}$$

$$\text{E}[Y|T] = \frac{\theta}{2}Tm_1$$

$$\text{Var}\left[\frac{\theta}{2}Tm_1\right] = \left(\frac{\theta}{2}m_1\right)^2 \text{Var}[T]$$

Therefore we have

$$\text{Var}[T] = \frac{\theta}{2}m_2\text{E}[T_{MRC A}] + \left(\frac{\theta}{2}m_1\right)^2\text{Var}[T_{MRC A}]. \quad (13)$$

The same procedure can be done for the covariance.

$$\text{Cov}[Y_a, Y_b] = \text{E}[\text{Cov}[Y_a, Y_b|T]] + \text{Cov}[E[Y_a|T], E[Y_b|T]]$$

We can break Y_a and Y_b into a shared part, Y_S and unshared parts for each, $Y_{\delta a}$ and $Y_{\delta b}$.

$$\begin{aligned} \text{Cov}[Y_S + Y_{\delta a}, Y_S + Y_{\delta b}|T] &= \text{Var}[Y_S|T] = \frac{\theta}{2}\tau_{a+b}m_2. \\ E\left[\frac{\theta}{2}\tau_{a+b}m_2\right] &= \frac{\theta}{2}m_2\text{E}[\tau_{a+b}] \end{aligned}$$

$$\text{Cov}[E[Y_a|T], E[Y_b|T]] = \text{Cov}\left[\frac{\theta}{2}Tm_1, \frac{\theta}{2}Tm_1\right] = \left(\frac{\theta}{2}m_1\right)^2 \text{Var}[T_{MRC A}]$$

Therefore we have

$$\text{Cov}[Y_a, Y_b] = \frac{\theta}{2}m_2\text{E}[\tau_{a+b}] + \left(\frac{\theta}{2}m_1\right)^2 \text{Var}[T_{MRC A}]. \quad (14)$$

The terms proportional to the variance of the $T_{MRC A}$ disappear because $Lm_1^2 \rightarrow 0$ as $L \rightarrow \infty$.

B Automatic moment derivations

Moments of the trait distribution can be calculated by differentiating equation (3). To automate the process, symbolic math programs were written using `sympy` (MEURER *et al.*, 2017). To derive moments for an arbitrary distribution of coalescent times we take Taylor series of the mgf of the mutational distribution to get

$$\left(\int \prod_{\Omega} \exp \left[t_{\omega} \frac{\theta}{2} \left(\sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a \right)^n \right) \right] P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \right)^L. \quad (15)$$

We then also Taylor expand the exponential function appearing in this integral to get

$$\left(1 + \sum_{\Omega} E[T_{\omega}] \frac{\theta}{2} \sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a \right)^n + \frac{1}{2} E \left[\left(\sum_{\Omega} T_{\omega} \frac{\theta}{2} \sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a \right)^n \right)^2 \right] + \frac{1}{6} E \left[\left(\sum_{\Omega} T_{\omega} \frac{\theta}{2} \sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a \right)^n \right)^3 \right] + \dots \right)^L. \quad (16)$$

Of course, the infinite sums appearing in this expression pose a problem. To calculate a particular moment, we only consider terms that will contribute terms of that moment's order when equation (16) is differentiated and zero is substituted for all dummy variables. For example, to calculate $E[Y_a Y_b Y_c]$ we only want terms that contain an order three product of k when the polynomial in equation (16) is expanded. For a moment of order m , we only need consider terms of size m and less in the series expansion of the exponential, and within the expansions of the mutational mgf we only need consider terms up to order $m - m_t + 1$ where m_t is the order of the exponential term being considered.

C Kurtosis simulations

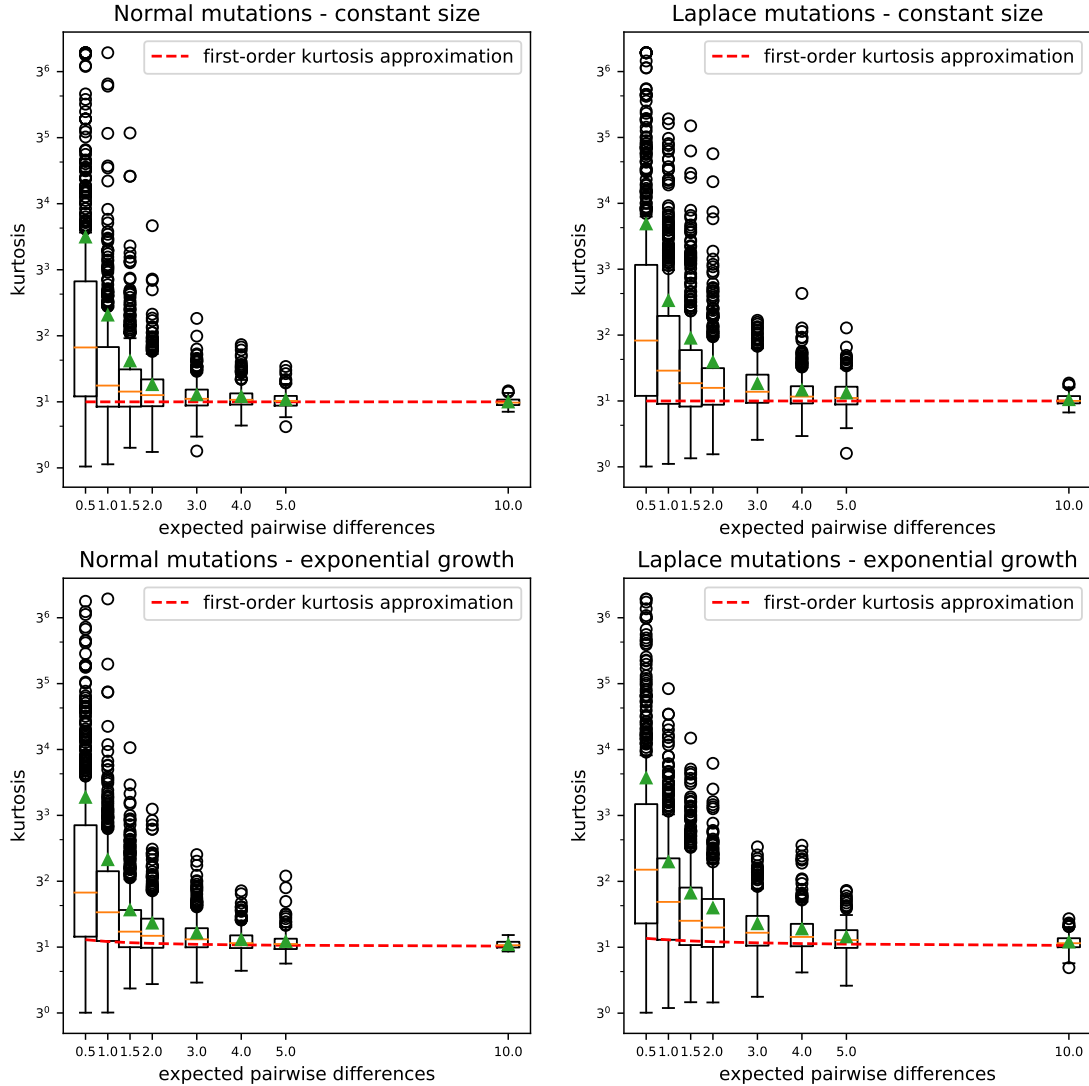


Figure S1: The distribution of the population kurtosis under different levels of sparsity, mutational kernels, and demographics. The sparsity is varied by changing the expected number of pairwise differences at sites affecting the trait. Normal and Laplace distributions of mutational effects are compared. A constant size population is compared to an exponential growth scenario with growth rate equal to the reciprocal of the final effective population size. Green triangles denote the mean kurtosis. The dashed red lines give the first order approximation to the expected kurtosis given in equation 17.

Entire populations were simulated using *msprime* (KELLEHER *et al.*, 2015) and mutations were assigned effects from a standard normal or Laplace distribution. The effective population size and mutation rate were kept constant and the expected number of pairwise difference was increased by increasing the number of loci affecting the trait.

A first order approximation to the kurtosis is

$$\kappa_Y^{(1)} = 3 + \frac{m_4/m_2^2 \frac{E[\mathbb{T}_{4,4}] - \frac{1}{6}E[\mathbb{T}_{3,4}] - \frac{1}{6}E[\mathbb{T}_{2,4}]}{E[\mathbb{T}_{2,2}]}}{L \frac{\theta}{2} E[\mathbb{T}_{2,2}] + m_4/m_2^2 \frac{\frac{1}{9}E[\mathbb{T}_{2,4}] + \frac{1}{6}E[\mathbb{T}_{3,4}]}{E[\mathbb{T}_{2,2}]}}. \quad (17)$$

Although this expression suggests that the expected κ_Y will be greater than under normality when external branches are longer ($E[\mathbb{T}_{4,4}] > \frac{1}{6}E[\mathbb{T}_{3,4}] + \frac{1}{9}E[\mathbb{T}_{2,4}]$) and less than under normality when they are longer ($E[\mathbb{T}_{4,4}] < \frac{1}{6}E[\mathbb{T}_{3,4}] + \frac{1}{9}E[\mathbb{T}_{2,4}]$), simulations show that the approximation is actually quite poor (Figure S1).

D Central moment derivations

We can use the low mutation rate approximation to the moment generating function to calculate moments of the distribution of trait values. We will start by calculating the first and second moments. We start, as we did in deriving the normal distribution, by substituting the Taylor series of the mutational mgf.

$$\varphi_Y(\mathbf{k}) \approx \left[1 + \sum_{\omega \in \Omega} E[T_\omega] \frac{\theta}{2} \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2!} \left(\sum_{a \in \omega} k_a \right)^2 + \frac{m_3}{3!} \left(\sum_{a \in \omega} k_a \right)^3 + \frac{m_4}{4!} \left(\sum_{a \in \omega} k_a \right)^4 \dots \right) \right]^L \quad (18)$$

We can expand this out using multinomial coefficients to get

$$\begin{aligned} \varphi_Y(\mathbf{k}) \approx & 1 + L \frac{\theta}{2} \sum_{\omega \in \mathcal{O}} E[T_\omega] \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2} \left(\sum_{a \in \omega} k_a \right)^2 + \dots \right) \\ & + \frac{L(L-1)}{2} \left(\frac{\theta}{2} \right)^2 \sum_{\omega \in \Omega} E[T_\omega]^2 \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2} \left(\sum_{a \in \omega} k_a \right)^2 + \dots \right)^2 \\ & + L(L-1) \left(\frac{\theta}{2} \right)^2 \sum_{\omega_1, \omega_2 \in \Omega} E[T_{\omega_1}] E[T_{\omega_2}] \left(m_1 \sum_{a \in \omega_1} k_a + \dots \right) \left(m_1 \sum_{a \in \omega_2} k_a + \dots \right) + \dots \end{aligned} \quad (19)$$

The first coefficient is $\binom{L}{L-1,1,0}$, the second is $\binom{L}{L-2,2,0}$, and the third is $\binom{L}{L-2,1,1,0}$. To calculate the moments of this distribution one takes the partial derivatives of the mgf and sets the dummy variables to zero.

$$E[Y_1^{r_1} \dots Y_n^{r_n}] = \frac{\partial^{r_1 + \dots + r_n}}{\partial k_1^{r_1} \dots \partial k_n^{r_n}} \varphi_Y(\mathbf{k}) \Big|_{\mathbf{k}=0} \quad (20)$$

Using this to calculate the first moment of the trait distribution we get

$$E[Y_a] \approx L \frac{\theta}{2} m_1 \sum_{\omega \in \Omega_a} E[T_\omega]. \quad (21)$$

The second moment is more complicated because there are k_a^2 terms in all three lines of equation 19.

$$\begin{aligned} E[Y_a^2] &\approx L \frac{\theta}{2} m_2 \sum_{\omega \in \Omega_a} E[t_\omega] \\ &\quad + \frac{L(L-1)}{2} \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega \in \Omega_a} 2E[T_\omega]^2 \\ &\quad + L(L-1) \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2 \in \Omega_{a+b}} 2E[T_{\omega_1}]E[T_{\omega_2}] \end{aligned} \quad (22)$$

Terms with $(\frac{\theta}{2})^2$ are kept because they also include a second order term of L in front of them. We can now calculate the variance using $\text{Var}[Y] = E[Y^2] - E[Y]^2$. The squared first moment can be written as

$$\begin{aligned} \left(L \frac{\theta}{2} m_1 \sum_{\omega \in \Omega_a} E[t_\omega] \right)^2 &= L^2 \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega \in \Omega_a} E[T_\omega]^2 \\ &\quad + L^2 \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2 \in \Omega_{a+b}} E[T_{\omega_1}]E[T_{\omega_2}]. \end{aligned} \quad (23)$$

Subtracting this from the second moment gives

$$\begin{aligned} \text{Var}[Y_a] &\approx L \frac{\theta}{2} m_2 \sum_{\omega \in \Omega_a} E[T_\omega] \\ &\quad - L \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega \in \Omega_a} E[T_\omega]^2 \\ &\quad - 2L \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2 \in \Omega_{a+b}} E[T_{\omega_1}]E[T_{\omega_2}] \\ &= L \frac{\theta}{2} m_2 \sum_{\omega \in \Omega_a} E[T_\omega] - L \left(\frac{\theta}{2} m_1 \sum_{\omega \in \Omega_a} E[T_\omega]\right)^2 \\ &= L \frac{\theta}{2} m_2 E[T_{MRC A}] - L \left(\frac{\theta}{2} m_1 E[T_{MRC A}]\right)^2 \\ &\approx L \frac{\theta}{2} m_2 E[T_{MRC A}]. \end{aligned} \quad (24)$$

Due to the large number of terms we only derive the fourth moment of the trait value distribution for the case when the mean mutational effect is zero. The terms of (18) appearing in the fourth moment after we apply (20) are

$$L \left(\frac{\theta}{2}\right) \frac{m_4}{24} \sum_{\omega \in \Omega_a} E[T_\omega] \left(\sum_{a \in \omega} k_a \right)^4$$

for the fourth moment along one branch,

$$\binom{L}{L-2, 2, \mathbf{0}} \left(\frac{\theta}{2}\right)^2 \left(\frac{m_2}{2}\right)^2 24 \sum_{\omega \in \Omega_a} E[T_\omega]^2 \left(\sum_{a \in \omega} k_a \right)^4$$

for the second moment of the same branch chosen twice, and

$$\binom{L}{L-2, 1, 1, \mathbf{0}} \left(\frac{\theta}{2}\right)^2 \left(\frac{m_2}{2}\right)^2 24 \sum_{\omega_1, \omega_2 \in \Omega_{a+b}} \mathbb{E}[T_{\omega_1}] \mathbb{E}[T_{\omega_2}] \left(\sum_{a \in \omega_1} k_a\right)^2 \left(\sum_{a \in \omega_2} k_a\right)^2$$

for the second moments on two different branches. Taking the fourth derivatives of these in terms of the desired branch we get

$$\begin{aligned} \mathbb{E}[Y_a^4] &= L \frac{\theta}{2} m_4 \mathbb{E}[T_{MRC A}] \\ &+ \frac{L(L-1)}{2} \left(\frac{\theta}{2}\right)^2 \left(\frac{m_2}{2}\right)^2 24 \sum_{\omega \in \Omega_a} \mathbb{E}[t_\omega]^2 \\ &+ L(L-1) \left(\frac{\theta}{2}\right)^2 \left(\frac{m_2}{2}\right)^2 24 \sum_{\omega_1, \omega_2 \in \Omega_{a+b}} \mathbb{E}[T_{\omega_1}] \mathbb{E}[T_{\omega_2}] \\ &= L \frac{\theta}{2} m_4 \mathbb{E}[T_{MRC A}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \sum_{\omega \in \Omega_a} \mathbb{E}[T_\omega]\right)^2 \\ &= L \frac{\theta}{2} m_4 \mathbb{E}[T_{MRC A}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \mathbb{E}[T_{MRC A}]\right)^2 \\ &\approx L \frac{\theta}{2} m_4 \mathbb{E}[T_{MRC A}] + 3 \left(L \frac{\theta}{2} m_2 \mathbb{E}[T_{MRC A}]\right)^2. \end{aligned} \quad (25)$$

The variance and the fourth moment derived in equation (24) and (25) can be used to derive the kurtosis of Y_a . The kurtosis of a random variable X is defined as

$$\text{Kurt}[X] = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{\mathbb{E}[(X - \mathbb{E}[X])^2]^2}.$$

This is the fourth central moment divided by the variance. It is possible to calculate the kurtosis of a single trait value over evolutionary realization. For ease of calculation, we'll examine this in the case where the mean mutation effect (and therefore trait value) is zero. If we plug (24) and (25) into the expression for the kurtosis we get

$$\begin{aligned} \text{Kurt}[Y_a] &= \frac{L \frac{\theta}{2} m_4 \mathbb{E}[T_{MRC A}]}{\left(L \frac{\theta}{2} m_2 \mathbb{E}[T_{MRC A}]\right)^2} + \frac{3L(L-1) \left(\frac{\theta}{2} m_2 \mathbb{E}[T_{MRC A}]\right)^2}{\left(L \frac{\theta}{2} m_2 \mathbb{E}[T_{MRC A}]\right)^2} \\ &= \frac{m_4}{L \frac{\theta}{2} m_2^2 \mathbb{E}[T_{MRC A}]} + \frac{3(L^2 - L)}{L^2} \\ &= \frac{\kappa}{L \frac{\theta}{2} \mathbb{E}[T_{MRC A}]} + 3 \left(1 - \frac{1}{L}\right). \end{aligned}$$

We also calculate some additional moments that have less clear interpretations but which are useful when calculating the expected fourth central moment in the population. The first of these is $\mathbb{E}[Y_a^3 Y_b]$. The terms of (18) appearing in this are

$$L \left(\frac{\theta}{2}\right) \frac{m_4}{24} 4k_a^3 k_b \sum_{\omega \in \Omega_{a+b}} \mathbb{E}[T_\omega]$$

and

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 k_a^2 \times 2k_a k_b \left(\sum_{\omega \in \Omega_{a+b}} E[T_\omega] \right) \left(\sum_{\omega \in \Omega_a} E[T_\omega] \right).$$

This ultimately gives

$$E[Y_a^3 Y_b] = L \frac{\theta}{2} m_4 E[\tau_{a+b}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRC A}] E[\tau_{a+b}]. \quad (26)$$

The next fourth moment of interest is $E[Y_a^2 Y_b Y_c]$. The terms of (18) are

$$L \frac{\theta}{2} \frac{m_4}{24} 12k_a^2 k_b k_c \sum_{\omega \in \Omega_{a+b+c}} E[T_\omega],$$

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 k_a^2 \times 2k_b k_c \left(\sum_{\omega \in \Omega_a} E[T_\omega] \right) \left(\sum_{\omega \in \Omega_{a+b}} E[T_\omega] \right),$$

and

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 2k_a k_b \times 2k_a k_c \left(\sum_{\omega \in \Omega_{a+b}} E[T_\omega] \right) \left(\sum_{\omega \in \Omega_{b+c}} E[T_\omega] \right).$$

Taking the appropriate derivatives of these gives

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c}] + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRC A}] E[\tau_{b+c}] + 2L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}] E[\tau_{a+c}]. \quad (27)$$

Individuals in the population are exchangeable as long as it is not structured. The pairwise expected shared branch lengths are in that case all equal and we can write (27) as

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c}] + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRC A}] E[\tau_{a+b}] + 2L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}]^2. \quad (28)$$

The final moment we'll look at is $E[Y_a Y_b Y_c Y_d]$ which has relevant terms

$$L \frac{\theta}{2} \frac{m_4}{24} 24k_a k_b k_c k_d \sum_{\omega \in \Omega_{a+b+c+d}} E[T_\omega],$$

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 2k_a k_b \times 2k_c k_d \left(\sum_{\omega \in \Omega_{a+b}} E[T_\omega] \right) \left(\sum_{\omega \in \Omega_{c+d}} E[T_\omega] \right),$$

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 2k_a k_c \times 2k_b k_d \left(\sum_{\omega \in \Omega_{a+c}} E[T_\omega] \right) \left(\sum_{\omega \in \Omega_{b+d}} E[T_\omega] \right),$$

and

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 2k_a k_d \times 2k_b k_c \left(\sum_{\omega \in \Omega_{a+d}} E[T_\omega] \right) \left(\sum_{\omega \in \Omega_{b+c}} E[T_\omega] \right).$$

When the appropriate fourth order partial derivatives are taken we get

$$\begin{aligned} & L \frac{\theta}{2} m_4 E[\tau_{a+b+c+d}] \\ & + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}] E[\tau_{c+d}] \\ & + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+c}] E[\tau_{b+d}] \\ & + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+d}] E[\tau_{b+c}]. \end{aligned}$$

We can again simplify this expression for populations if we assume that individual are exchangeable. This gives

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c+d}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}]^2. \quad (29)$$

The expected kurtosis in the population is a quotient of random variables and calculating a second order approximation requires calculating eight order moments of the trait distribution. Instead we will calculate the expected fourth central moment.

$$E[M_{4,Y}] = E \left[\frac{1}{N} \sum \left(Y_i - \frac{\sum Y_j}{N} \right)^4 \right]. \quad (30)$$

Examining the sum inside the expectation we see that

$$\begin{aligned} E \left[\left(Y_i - \frac{\sum Y_j}{N} \right)^4 \right] &= E[Y_i^4] - 4E \left[Y_i^3 \frac{\sum Y_j}{N} \right] + 6E \left[Y_i^2 \left(\frac{\sum Y_j}{N} \right)^2 \right] - 4E \left[Y_i \left(\frac{\sum Y_j}{N} \right)^3 \right] + \left(\frac{\sum Y_j}{N} \right)^4 \\ &= E[Y_i^4] - \frac{4}{N} \sum_j E[Y_i^3 Y_j] + \frac{6}{N^2} \sum_{j,k} E[Y_i^2 Y_j Y_k] \\ &\quad - \frac{4}{N^3} \sum_{j,k,l} E[Y_i Y_j Y_k Y_l] + \frac{1}{n^4} \sum_{j,k,l,d} E[Y_j Y_k Y_l Y_d]. \end{aligned} \quad (31)$$

In calculating these expectations we have to remember that the value depends only on the number of times each variable appears in the expectation. That is, $E[Y_1^2 Y_2 Y_3]$ is equivalent to $E[Y_1 Y_2^3 Y_3]$ as long as all individuals in the population are exchangeable. The resulting expansion of (31) can be simplified by only considering terms of order one. Other terms can be ignored since we are assuming there are a large number of individuals in the population. This yields

$$\begin{aligned} E \left[\left(Y_i - \frac{\sum Y_j}{N} \right)^4 \right] &= E[Y_i^4] - \frac{4(n-1)}{n} E[Y_i^3 Y_j] + \frac{6(n-1)(n-2)}{n^2} E[Y_i^2 Y_j Y_k] \\ &\quad - \frac{4(n-1)(n-2)(n-3)}{n^3} E[Y_i Y_j Y_k Y_l] \\ &\quad + \frac{(n-1)(n-2)(n-3)(n-4)}{n^4} E[Y_j Y_k Y_l Y_d] + O(n^{-1}) \\ &\approx E[Y_i^4] - 4E[Y_i^3 Y_j] + 6E[Y_i^2 Y_j Y_k] - 3E[Y_i Y_j Y_k Y_l]. \end{aligned} \quad (32)$$

The first term, $E[Y_i^4]$ was derived in equation (25) as

$$L\frac{\theta}{2}m_4E[T_{MRC A}] + 3L(L-1)\left(\frac{\theta}{2}m_2E[T_{MRC A}]\right)^2.$$

The second term, $E[Y_i^3Y_j]$ was derived in equation (26) as

$$L\frac{\theta}{2}m_4E[\tau_{a+b}] + 3L(L-1)\left(\frac{\theta}{2}m_2\right)^2 E[T_{MRC A}]E[\tau_{a+b}].$$

The third term, $E[Y_i^2Y_jY_k]$ was derived in equation (27) as

$$L\frac{\theta}{2}m_4E[\tau_{a+b+c}] + L(L-1)\left(\frac{\theta}{2}m_2\right)^2 E[T_{MRC A}]E[\tau_{a+b}] + 2L(L-1)\left(\frac{\theta}{2}m_2\right)^2 E[\tau_{a+b}]^2.$$

The fourth term, $E[Y_iY_jY_kY_l]$ was derived in equation (29) as

$$L\frac{\theta}{2}m_4E[\tau_{a+b+c+d}] + 3L(L-1)\left(\frac{\theta}{2}m_2\right)^2 E[\tau_{a+b}]^2.$$

Plugging these into (32) we get

$$\begin{aligned} E[M_4] &\approx L\frac{\theta}{2}m_4 (E[T_{MRC A}] - 4E[\tau_{a+b}] + 6E[\tau_{a+b+c}] - 3E[\tau_{a+b+c+d}]) \\ &\quad + 3\left(L\frac{\theta}{2}m_2\right)^2 (E[T_{MRC A}] - E[\tau_{a+b}])^2. \end{aligned} \quad (33)$$

With some simple manipulation this can be rewritten in terms of $\mathbb{T}_{k,n}$ to give

$$\begin{aligned} E[M_4] &= 3\left(L\frac{\theta}{2}E[\mathbb{T}_{2,2}]m_2\right)^2 \\ &\quad + Lm_4\frac{\theta}{2}(E[\mathbb{T}_{4,4}] + \frac{1}{3}E[\mathbb{T}_{3,4}] + \frac{2}{9}E[\mathbb{T}_{2,4}]). \end{aligned} \quad (34)$$