QUANTITATIVE TRAIT EVOLUTION WITH ARBITRARY MUTATIONAL MODELS

JOSHUA G. SCHRAIBER AND MICHAEL J. LANDIS

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 experiment freely with the number of loci and the underlying mutational model. 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, fat-tailed mutational kernels result in multimodal sampling distributions for any 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 the 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 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 primary 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

60 Whitlock (1999), but allows us to recover the full sampling distribution, instead of merely
61 its moments.
62   First, we outline the biological model and explain how we can compute quantities of
63 interest using a formalism based on characteristic functions. We then use this approach to
64 compute the sample central moments. While much previous work focuses on only the first
65 two central moments (mean and variance), we are able to compute arbitrarily high central
66 moments, which are related to properties such as skewness and kurtosis. By doing so, we
67 are able to determine the regime in which the details of the mutational effect distribution
68 are visible in a sample from a natural population. Finally, we explore the convergence to
69 the infinitesimal limit and find that when "fat-tailed" effects are present, traditional theory
70 based on the assumption of normality can lead to misleading predictions about phenotypic
71 variation.

2. Model

73   The mechanistic model we construct has three major components: a coalescent process,
74 a genetic mutational process that acts upon the controlling quantitative trait loci, and
75 a mutational kernel that samples quantitative trait effect sizes. Together these processes
76 generate the quantitative traits sampled from the study population while explicitly mod-
77 eling their shared genetic ancestry. Although we opt for simple model components during
78 this exposition, the model generally supports more realistic and complex extensions, such
79 as population structure and epistasis.
80   We assume that we sample n haploid individuals from a randomly mating population
81 of size N. Initially, consider a trait governed by a single locus and we will later extend
82 the theory to traits governed by multiple loci. Let μ be the mutation rate per generation
83 at the locus, and $\theta = 2N\mu$ be the coalescent-scaled mutation rate. We model mutation
84 as a process by which a new mutant adds an independent and identically distributed
85 random effect to the ancestral state. Note that when the distribution of random effects is
86 continuous, this corresponds to the Kimura (1965) continuum of alleles model. However,
87 it is also possible for the effect distribution to be discrete, similar to the discrete model of
88 Chakraborty and Nei (1982). While this model does not capture the impact of a biallelic
89 locus with exactly two effects, the following theory could easily be modified to analyze that
90 case.

91   [Figure 1 goes here]

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

99 This mutational process can be described as a compound Poisson process (see also
100 Khaitovich et al. (2005b); Chaix et al. (2008); Landis et al. (2013) for compound Pois-
101 son processes in a phylogenetic context). To ensure that this paper is self contained, we
102 briefly review relevant facts about compound Poisson processes in the Appendix.
103 In the following, we ignore the impact of non-genetic variation and focus on the breeding
104 value of individuals, i.e. the average phenotype of an individual harboring a given set of
105 mutations.

106   3. RESULTS

107 3.1. **Computing the characteristic function of a sample.** In many analyses, the
108 object of interest is the joint probability of the data. If we let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be
109 the vector representing the quantitative traits observed in a sample of n individuals, we
110 denote the joint probability of the data as $p(x_1, x_2, \ldots, x_n)$. Note that, in general, X_i and
111 X_j are correlated due to shared ancestry, and that p must be computed by integrating over
112 all mutational histories consistent with the data. Hence, computing p directly is extremely
113 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,
k is a dummy variable. Generalizing this definition to an n-dimensional random variable,
we are interested in computing

$$\lambda_n(\mathbf{k}) = \mathbb{E}(e^{i\mathbf{k}^T\mathbf{X}})$$
$$= \mathbb{E}(e^{i(k_1 X_1 + k_2 X_2 + \ldots + k_n X_n)})$$

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

118 We calculate this formula in two parts. First, we compute a recursive formula for ϕ_n,
119 the characteristic function given that ancestral phenotype of the *sample* is equal to 0.
120 Then, we compute ρ_n, the characteristic function of the ancestral phenotype of the sample,
121 assuming that the characteristic function of the *population* is equal to 0. As we show in the
122 Appendix, we can then multiply these characteristic functions to obtain the characteristic
123 function of $\mathbf{X}$.

124 We use a backward-forward argument to compute the recursive formula, first condi-
125 tioning on the state when the first pair of lineages coalesce (backward in time) and then
126 integrating (forward in time) to obtain the characteristic function for a sample of size n,
127 ϕ_n. This results in

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

128 where $\mathbf{k}^{(u,v)}$ is the vector of length $n-1$ made by removing k_u and k_v and adding $k_u + k_v$
129 to the vector of dummy variables.

130   This equation has a straight-forward interpretation. The characteristic function for a
131 sample of size n, ϕ_n, is simply the characteristic function for a sample of size $n-1$, ϕ_{n-1},
132 averaged over all possible pairs that could coalesce first, multiplied by the characteristic
133 function for the amount of trait change that occurs more recently than the first coalescent.
134 The multiplication comes from the fact that the characteristic function of a sum of inde-
135 pendent random variables is the product of the characteristic functions of those random
136 variables. We prove this result in the Appendix (Section 5.3).

137   In the Appendix (Section 5.4), we also show that the characteristic function for the
138 phenotype at the root of the sample is

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

139 Intuitively, this equation arises by conditioning on whether u lineages are left in the *popu-*
140 *lation* when the sample reaches its common ancestor and then averaging over the (random)
141 time between when the individuals in the sample coalesce and when everyone in the pop-
142 ulation coalesces.

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

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

144 3.2. **Sampling traits controlled by a small number of loci.** It is common practice
145 in both theoretical and applied quantitative genetics to summarize information about the
146 phenotypic distribution within a population by computing central moments. However, care
147 must be taken when interpreting theoretical predictions about central moments estimated
148 from a sample. This is because the phenotypes in the sample are not independent, but
149 instead correlated due to their shared genealogical history. Hence, in any *particular* pop-
150 ulation, an estimate of a central moment may deviate from its expected value, even as the
151 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 for details). They are

$$\mathbb{E}(h_2) = \frac{1}{2}\theta L m_2$$
$$\mathbb{E}(h_3) = \frac{1}{6}\theta L m_3$$
$$\mathbb{E}(h_4) = \frac{3}{4}\theta^2 L^2 m_2^2 + \frac{1}{4}\theta L(2\theta m_2^2 + m_4),$$

152 where h_k is the unique minimum variance unbiased estimator of the kth central moment,
153 m_k is the kth moment of the mutational effect distribution and L is the number of loci
154 that influence the trait.

155   These equations reveal that it may be possible to construct method-of-moments estima-
156 tors for the moments of the mutation effect distribution and/or the number of loci that
157 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.

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. 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 limit distribution is given by

$$
\begin{aligned}
\Lambda_n(\mathbf{k}) &= \lim_{L \to \infty} \lambda_n(\mathbf{k})^L \\
&= \lim_{L \to \infty} \rho_n(k_1 + k_2 + \ldots + k_n)^L \phi_n(\mathbf{k})^L \\
&= R_n(k_1 + k_2 + \ldots + k_n)\Phi_n(\mathbf{k}).
\end{aligned}
$$

In the Appendix, 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. In the Appendix, 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,

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

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

3.4. **Simulation.** We simulated data to verify our analytical results and obtain some insight into the nature of the stable limiting distribution that arises for power law mutational effects. We first wanted to confirm that the trait distributions converge to univariate Gaussian limiting distributions as $n \to \infty$ and $L \to \infty$ when mutational kernels are not fat-tailed. To explore how the moments of the sampling distribution change with respect to n, L, and the mutational kernel, we asked for which values of L do the moments of the various mutational kernels leave a signature in the sampled quantitative traits. Finally, we conjectured that fat-tailed mutation kernels result in trait distributions that remain multimodal as $L \to \infty$, which we verified by simulation rather than by mathematical proof.

179   For these simulation studies, we selected four mutational kernels: (1) the symmetric
180 normal distribution for it's simplicity, (2) the Laplace distribution because it is heavier-
181 tailed (or more leptokurtic) than the normal distribution yet has finite variance, (3) the
182 skew-normal distribution for it's skewness parameter and tractability, and (4) the symmet-
183 ric α-stable distribution because of its power-law behavior. To ensure that simulations of
184 different non-fat-tailed distributions were comparable, we set the variance per locus to be
185 $\tau^2 = \sigma^2/L$ when we simulated L loci, meaning the trait distribution would have constant
186 variance $\theta\sigma^2/2$. Note the symmetric normal distribution is a special case of both the skew-
187 normal distribution when the skewness parameter is zero and the α-stable distribution
188 when the "fat-tailedness" parameter is $\alpha = 2$.
189   For all simulations, we generated coalescent genealogies and mutations using the program
190 `ms` (Hudson, 2002). We then generated and mapped mutational effects using custom scripts
191 in `R` (R Core Team, 2013).
192   Code is available at `http://github.com/Schraiber/quant_trait_coalescent`.

193   [Figure 2 goes here.]

194   [Figure 3 goes here.]

195 3.4.1. *Univariate Gaussian limit.* For mutational kernels of small effect size and variance
196 τ^2 per locus, the sampling distribution converges to a normal distribution with variance
197 $\theta\sigma^2/2$ where $L\tau^2 \to \sigma^2$ as $L \to \infty$. We simulated 100 replicates of trait data for $L \in$
198 $\{1, 2, 4, \ldots, 256\}$ and $n \in \{2, 4, 8 \ldots, 512\}$ with mutation parameters $\theta = 2$ and $\tau = 1$
199 for the normal, the skew-normal (skewness=0.9), and the Laplace distributions. We then
200 assessed convergence to the normal limit using the Kolmogorov-Smirnov (KS) test statistic,
201 D, which equals zero when two distributions are identical. Figure 2 reports the frequency
202 we reject the null hypothesis—that the limiting and sampled distributions are identical—
203 for each batch of 100 replicates per value of n and L for p-values less than 0.05. For
204 $n \leq 4$, the KS test lacked power to reject the null hypothesis whatsoever. For $n \geq 8$, the
205 three mutational kernels converge to the limiting normal distribution in a similar fashion,
206 with the sampling and limiting distributions bearing strong resemblance when $L > 16$.
207 Distinctly, the Laplace distribution converges to normality at a slower rate than other
208 mutational kernels, likely resulting from it being leptokurtic (Figure 3).

209   [Figure 4 goes here.]

210 3.4.2. *Central moments.* We assessed the signature left by various mutational kernels on
211 the sampling distribution by computing the central moments across simulation replicates.
212 While the variance (h_2) remains constant for all values of L regardless of the mutational
213 kernel (by experimental design), the third central moment (h_3) and the fourth central
214 moment (h_4) depend on the mutational kernel for small values of L. As $L \to \infty$, the
215 sample moments converge to those of a normal distribution. Here, we characterize the
216 deviation from the normally-distributed moments under a variety of mutational kernels:
217 the (symmetric) normal distribution; the skew-normal distribution for skewnesses 0.1, 0.5
218 and 0.9; and the Laplace distribution. We omitted the α-stable distribution from this

219 portion of the study since its moments h_2, h_3, and h_4 only exist when $\alpha = 2$, i.e. when it
220 is Gaussian.
221   We simulated data while varying the number of loci, $L \in \{1, 2, 4, \ldots, 256\}$, holding
222 the sample size constant, $n = 1024$, for 2000 replicates for each of the five mutation
223 kernels. Afterwards, we computed the mean h_2, h_3, and h_4 statistics across replicates of
224 each mutation kernel and value of L for comparison with their expected h-statistic values
225 (Figure 4). As expected, h_2 remains constant regardless of the mutation kernel or L. The
226 normal and Laplace distributions are symmetric and produce sample h_3 values near zero,
227 indicating no skewness. The skew-normal mutation kernel result in non-zero skewness
228 even for traits controlled by over 100 loci so long as the kernel is sufficiently strongly
229 skewed. The speed the sampling distribution's third central moment, h_3, converges to
230 zero in inverse proportion to the magnitude of its mutational kernel's skewness value. All
231 distributions produce non-zero h_4 values when L is small, due to the randomness of the
232 mutation process. The h_4 value of the Laplace distribution, the sole leptokurtic mutational
233 kernel in this comparison, is the slowest of all kernels to converge to the normal limit.

234   [Figure 5 goes here.]

235 3.4.3. *Multimodality.* As $n \to \infty$ and $L \to \infty$, we proved that sampling distributions
236 generated by finite-variance mutational kernels converge to the unimodal normal distri-
237 bution and conjectured that power-law mutational kernels, such as the α-stable, converge
238 to multimodal stable distributions. Here, we demonstrate by simulation our proven and
239 conjectured modality results hold as $L \to \infty$.
240   To do so, we test for unimodality using the dip statistic, D (Hartigan and Hartigan,
241 1985). Briefly, $D(F_0, F_1)$ gives the minimized maximum difference between an empirical
242 distribution, F_1, and some unimodal (null) distribution, F_0, where F_0 is typically taken
243 to be the uniform distribution. D approaches zero when F_1 is unimodal and equals $\frac{1}{4}$
244 when the distribution is perfectly bimodal (i.e. two point masses). We used the R package
245 `diptest` (Maechler and Ringach, 2012) to compute p-values for each simulated dataset,
246 recording the frequency of replicates whose p-value is less than 0.05 for each mutational
247 kernel and value of L. If the limiting distribution is unimodal, we expect this frequency to
248 be less than 0.05 as L increases. Conversely, we expect multimodal limiting distributions
249 to converge in frequency to some value greater than 0.05 as L increases.
250   We complemented the simulated data from Section 3.4.2 with three additional α-stable
251 mutational kernels for $\alpha \in \{1.5, 1.7, 1.9\}$, keeping the coalescent-mutation process variance
252 equal across all datasets.
253   Figure 5 shows the trait distributions under α-stable mutation kernels remained mul-
254 timodal as $L \to \infty$ and stratify according to their respective α values: as α decreases
255 the large-effect mutations responsible for multimodality grow more prominent. Mutation
256 kernels of small effect size become unimodal as $L \to \infty$. Notably, Laplace-distributed mu-
257 tations converge to unimodality more slowly the normally-distributed mutations, echoing
258 the results reported in Sections 3.4.1 and 3.4.2. Also note that when the number of loci

259 is small ($L \leq 4$) the sampling distribution is multimodal regardless of the mutation ker-
260 nel. This corroborates our earlier KS tests (Section 3.4.1), which found simulated data for
261 $L \leq 4$ bore little to no resemblance to a unimodal normal distribution.

262   4. DISCUSSION

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

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

281   We assessed departure from normality in traits governed by a small number of loci by
282 exploring the central moments of three different mutational kernels (normal, skew-normal
283 and Laplace distributions) both analytically and by simulation. We showed that although
284 all three mutational kernels converge to a Gaussian distribution, traits controlled by a
285 small number of loci retain the signature of their underlying mutational kernel in their
286 3rd and 4th central moments. Hence, it may be possible to reconstruct aspects of the
287 mutational effect distribution by observing phenotypes in natural populations. This may
288 be particularly interesting for analyzing variation in gene expression, because mutational
289 effects in *cis* may be strongly skewed (Khaitovich et al., 2005a; Chaix et al., 2008; Gruber
290 et al., 2012). Our theory suggests that the distribution of gene expression in a population
291 might therefore be skewed.

292   We were also interested in the circumstances under which multi-modal phenotypic dis-
293 tributions can arise. When a trait has a simple genetic architecture, it's easy to see that
294 there must be discrete phenotypic clusters, corresponding to groups of individuals sharing
295 the same mutations. As the number of loci increases, there are more mutational targets
296 (and thus more mutation events), which smooths the distribution, causing the sampling
297 distribution to converge to the appropriate limiting distribution. For mutational effects
298 with finite variance, this ultimately results in a limiting Gaussian distribution, consistent
299 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 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); 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.

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, dioecious 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 straightforward 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.

5. Appendix

5.1. **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

$$(3) \qquad \varphi_t(k) = e^{\frac{\theta}{2}t(\psi(k)-1)}$$

where ψ is the characteristic function of the mutational effect distribution.

5.2. **The phenotype at the root of the sample genealogy and the subsequent evolution within the sample are subindepenent.** Note that

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

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,

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

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.

5.3. **Proof of recursive formula for the characteristic function.** 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 ilneages have coalesced and hence had the same

trait value at the time of coalescence). Then,

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

(4)

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 (3) 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,

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

Plugging this result into (4) 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).

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

(5)
$$
\phi_2(\mathbf{k}) = \frac{2}{2 - \theta(\psi(k_1) + \psi(k_2) - 2)}.
$$

5.4. **The phenotype at the root of the sample genealogy.** First, we note that, conditional on the time between when the sample genealogy finds a common ancestral and

372 the population genealogy finds a common ancestor, Δ, the characteristic function of the
373 phenotype at the root of the sample genealogy is

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

374 by using equation (3). Thus, the after integrating over Δ, the desired quantity is *moment*
375 *generating function* of Δ, defined by

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

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

377   We compute $M(z)$ by conditioning on how many lineages are left in the population
378 genealogy when the sample reaches its most recent common ancestor. To do this, we make
379 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 an-
cestor is simply the time it takes for a coalescent started with u to reach its most recent
common ancestor, C_u. Thus,

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

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

5.5. **Computing sample central moments.** While it is difficult to compute the expec-
tation of any sample central moments for a particular sample, it is possible to average over
replicate populations to compute expectations. This results in

$$
\begin{aligned}
\mathbb{E}(h_2) &= \mathbb{E}(X_1^2) - \mathbb{E}(X_1 X_2) \\
\mathbb{E}(h_3) &= \mathbb{E}(X_1^3) + 2\mathbb{E}(X_1 X_2 X_3) - 3\mathbb{E}(X_1^2 X_2) \\
\mathbb{E}(h_4) &= \mathbb{E}(X_1^4) + 6\mathbb{E}(X_1^2 X_2 X_3) - 4\mathbb{E}(X_1^3 X_2) - 3\mathbb{E}(X_1 X_2 X_3 X_4),
\end{aligned}
$$

383 where the expectations on the right hand sides are over the *correlated* phenotypes in the
384 sample. It is possible to compute these expectations by taking derivatives of the char-
385 acteristic function (1). After simplifying, one then arrives at the formulas in the main
386 text.

387 5.6. **Derivation of multivariate stable limit for sample distribution.** Recall that a
388 random variable X is said to have a fat-tailed (or power-law) distribution if

$$(6) \qquad\qquad P(X > x) \sim \kappa x^{-\alpha}$$

389 for large x and some $\kappa > 0$. As is typical in the literature, we reserve the term "fat-tailed"
390 for distributions with $\alpha \in (0, 2)$.
391   To obtain an appropriate scaling limit, we assume that there is a parameter t, related
392 to the parameter κ in (6) by

$$(7) \qquad\qquad t = \kappa \frac{\pi}{\sin(\alpha\pi/2)\Gamma(\alpha)\alpha},$$

393 such that $Lt \to s$ as $n \to \infty$. The parameter s is related to the scale parameter of the
394 resulting limit distribution.
395   We provide a heuristic derivation, rather than a rigorous proof. First, we argue by
396 induction that the (per locus) characteristic function for a sample of size n is

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

397 for large L, where $\mathcal{P}^*(\mathbf{k})$ is the power set of the elements in $\mathbf{k}$, *except* the set $\{k_1, k_2, \ldots, k_n\}$,
398 and $c_{n,|\mathbf{j}|}$ is a combinatorial constant that depends only on the sample size n and $|\mathbf{j}|$, the
399 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

$$\psi(k) \sim 1 - \frac{s}{L}|k|$$
$$\equiv \tilde{\psi}$$

Thus,

$$\phi_2(\mathbf{k}) \sim 1 - \frac{\theta}{2}\frac{s}{L}\left(|k_1|^{\alpha} + |k_2|^{\alpha}\right)$$
$$\equiv \tilde{\phi}_2$$

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

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

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

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

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

$$
= \tilde{\phi}_n(\mathbf{k}).
$$

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

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

402 where all terms are defined as before.
403   The characteristic function in (8) can be recognized to be that of a multivariate α-stable
404 distribution (Press, 1972). These multivariate distributions are fat-tailed generalizations
405 of the familiar multivariate normal distribution, and this limit corresponds to a generalized
406 multivariate central limit theorem for sums of random vectors with fat-tailed distributions.

407 5.7. **Limiting distribution of the phenotype at the root of the sample genealogy.**
408 Again, we proceed heuristically rather than rigorously. First, note that for large L,

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

409 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,

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

$$
= 1 - \frac{s\theta}{Ln}|k|^{\alpha}.
$$

So by definition of the exponential function, we have that

$$
R_n(k) = \lim_{L\to\infty} \rho_n(k)^L
$$

$$
(9) \qquad = e^{-\frac{s\theta}{n}|k|^{\alpha}}.
$$

410 which the characteristic function of a univariate α-stable distribution, arising from the fact
411 that the phenotype at the root of the sample genealogy is itself a limit of a sum of random

412 variables. Note that as $n \to \infty$ (i.e. the sample becomes the whole population), $R(k) \to 1$,
413 because the root of the sample genealogy is the same as the root of the population genealogy
414 and the root value has been specified to be equal to 0.

415 5.8. **Multivariate Gaussian limits.** For the case where the mutation distribution is not
416 fat-tailed, we can use the multivariate central limit theorem to more efficiently derive the
417 limiting distribution. The appropriate scaling in this case is to assume that if τ^2 is the
418 variance of the mutation effect kernel, then $L\tau^2 \to \sigma^2$ as $L \to \infty$.

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

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

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

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

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

430 to see that

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

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

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

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

435 Because the pairwise covariances are equal, the random vector $\mathbf{X}$ is an exchangeable Gauss-
436 ian random vector. Hence, using well-known facts about the representation of exchangeable
437 Gaussian random vectors, one arrives at the representation in the main text.

438   6. ACKNOWLEDGMENTS

References

443

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

445 Berg, J. J. and G. Coop. 2014. A population genetic signal of polygenic adaptation. PLoS
446   Genetics 10:e1004412.

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

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

451 Chakraborty, R. and M. Nei. 1982. Genetic differentiation of quantitative characters be-
452   tween populations or species: I. Mutation and random genetic drift. Genetical Research
453   39:303–314.

454 Doebeli, M., H. J. Blok, O. Leimar, and U. Dieckmann. 2007. Multimodal pattern formation
455   in phenotype distributions of sexual populations. Proceedings of the Royal Society B:
456   Biological Sciences 274:347–357.

457 Eastman, J. M., D. Wegmann, C. Leuenberger, and L. J. Harmon. 2013. Simpsonian 'evolu-
458   tion by jumps' in an adaptive radiation of anolis lizards. arXiv preprint arXiv:1305.4216
459   .

460 Estes, S. and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing
461   selection explain evolutionary divergence on all timescales. The American Naturalist
462   169:227–244.

463 Falconer, D. and T. Mackay. 1996. Introduction to Quantitative Genetics. 4 ed. American
464   Genetic Association.

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

466 Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian
467   inheritance. Transactions of the Royal Society of Edinburgh 52:399–433.

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

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

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

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

474 Hartigan, J. A. and P. Hartigan. 1985. The dip test of unimodality. The Annals of Statistics
475   Pages 70–84.

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

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

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

483 Kimura, M. 1965. A stochastic model concerning the maintenance of genetic variability in
484   quantitative characters. Proceedings of the National Academy of Sciences of the United

485 States of America 54:731.

486 Kingman, J. F. C. 1992. Poisson Processes. 3 ed. Oxford University Press.

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

489 Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evo-
490 lution Pages 314–334.

491 Landis, M. J., J. G. Schraiber, and M. Liang. 2013. Phylogenetic analysis using Lévy
492 processes: finding jumps in the evolution of continuous traits. Systematic biology 62:193–
493 204.

494 Latter, B. 1960. Natural selection for an intermediate optimum. Australian Journal of
495 Biological Sciences 13:30–35.

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

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

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

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

505 Maechler, M. and D. Ringach. 2012. diptest: Hartigans dip test statistic for unimodality.

506 McKay, J. K. and R. G. Latta. 2002. Adaptive population divergence: markers, qtl and
507 traits. Trends in Ecology & Evolution 17:285–291.

508 Mendel, G. 1866. Versuche über pflanzenhybriden. Verhandlungen des naturforschenden
509 Vereines in Brunn 4: 3 44.

510 Möhle, M. 1998. Coalescent results for two-sex population models. Advances in Applied
511 Probability Pages 513–520.

512 Ovaskainen, O., M. Karhunen, C. Zheng, J. M. C. Arias, and J. Merilä. 2011. A new method
513 to uncover signatures of divergent and stabilizing selection in quantitative traits. Genetics
514 189:621–632.

515 Pearson, K. 1894. Contributions to the mathematical theory of evolution. Philosophical
516 Transactions of the Royal Society of London A. Pages 71–110.

517 Pearson, K. 1895. Contributions to the mathematical theory of evolution. III. Regression,
518 heredity, and panmixia. Proceedings of the Royal Society of London 59:69–71.

519 Pearson, K. 1904. Mathematical contributions to the theory of evolution. XII. On a gener-
520 alised theory of alternative inheritance, with special reference to Mendel's laws. Philo-
521 sophical Transactions of the Royal Society of London A. Pages 53–86.

522 Press, S. J. 1972. Multivariate stable distributions. Journal of Multivariate Analysis 2:444–
523 462.

524 R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foun-
525 dation for Statistical Computing Vienna, Austria.

526 Saunders, I. W., S. Tavaré, and G. Watterson. 1984. On the genealogy of nested subsamples
527 from a haploid population. Advances in Applied probability Pages 471–491.

528 Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for
529   macroevolutionary bursts. Proceedings of the National Academy of Sciences 108:15908–
530   15913.
531 Weldon, W. F. R. 1902. Mendel's laws of alternative inheritance in peas. Biometrika
532   Pages 228–254.
533 Whitlock, M. C. 1999. Neutral additive genetic variance in a metapopulation. Genetical
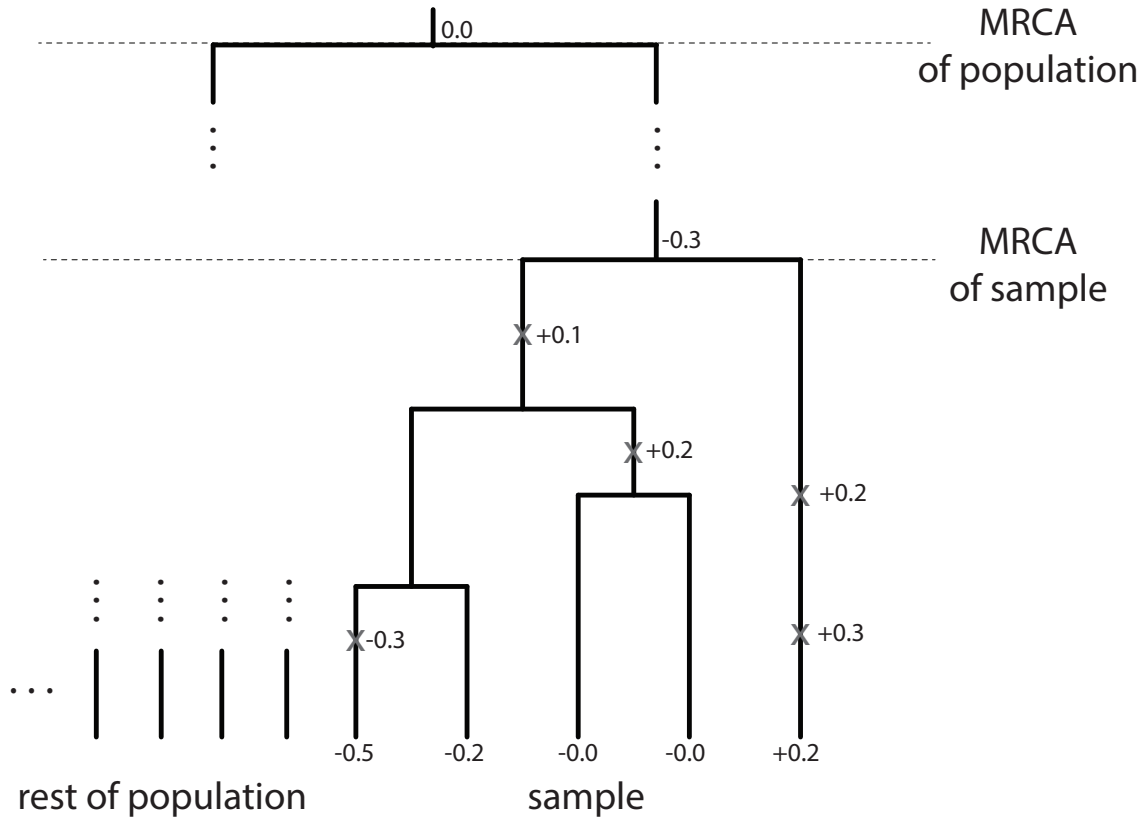534   research 74:215–221.

7. Figures

FIGURE 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 population may have a phenotype different from zero, due to mutations that accumulate between the MRCA of the sample and the MRCA of the population.
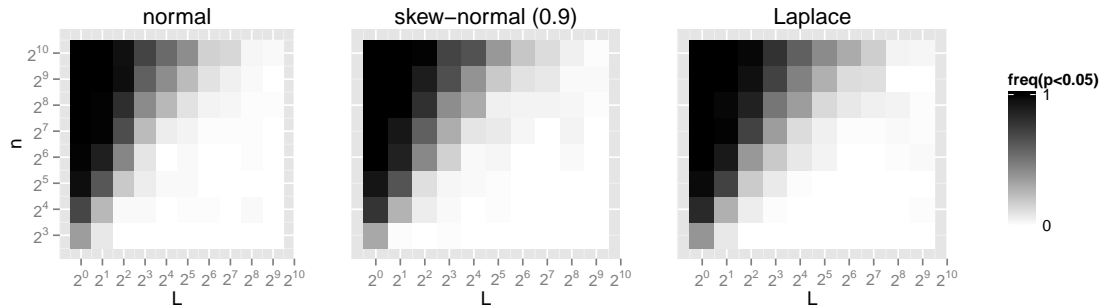
FIGURE 2. Frequency to reject normal sampling distribution. Heatmap cells correspond to number of sampled individuals, n, and number of loci, L. Panels are labeled with their respective mutational kernels. For 100 simulated replicates per cell, heatmap values correspond to the frequency the Kolmogorov-Smirnov test rejects the null hypothesis ($p < 0.05$) that the sampling distribution and the limiting normal distribution are equal. White cells indicate the sampling distribution looks normal. Black cells indicate the sampling distribution does not look normal.
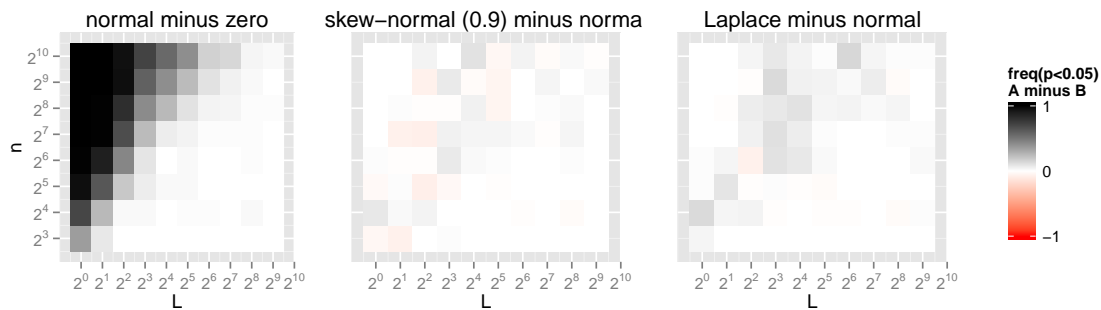

FIGURE 3. Comparison of frequency to reject normal sampling distribution. Heatmap cells correspond to number of sampled individuals, n, and number of loci, L. Panels are labeled with their respective mutational kernels. For 100 simulated replicates per cell, heatmap values correspond to the frequency the Kolmogorov-Smirnov test rejects the null hypothesis ($p < 0.05$) that the sampling distribution, A, and the limiting normal distribution are equal *minus* the frequencies computed for a second distribution, B. White cells indicate both distributions report equal frequencies. Black cells indicate A looks normal more often than B. Red cells indicate A looks normal less often than B.
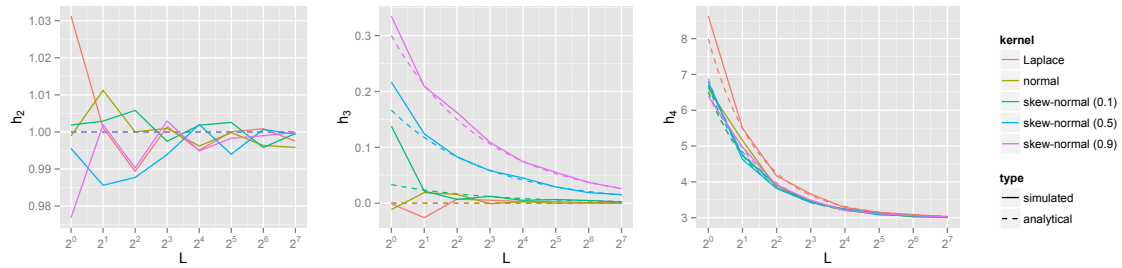
FIGURE 4. Central moments. From left to right, the panels correspond to the central moments, h_2, h_3, and h_4, respectively, for the sampling distributions evolving under various mutational kernels. Data were simulated for 1024 sampled individuals and 2000 replicates for eight values of L, the number 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.
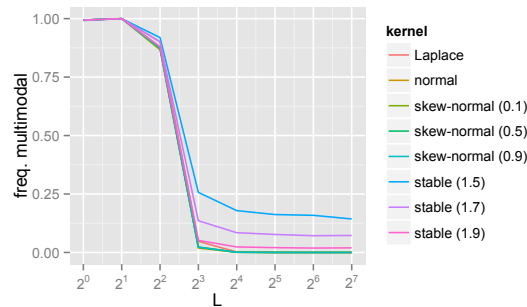

FIGURE 5. Frequency to reject unimodal sampling distribution. Solid lines report the frequency the null hypothesis of the dip-test, that the sampling distribution was unimodal, was rejected for $p < 0.05$ when evolving under various mutational kernels. Data were simulated for 1024 sampled individuals and 2000 replicates for eight values of L, the number of loci. Colors distinguish the mutational kernel and relevant kernel parameters (if any).