Detecting diversifying selection for a trait from within and between-species genotypes and phenotypes

T. Latrille[1], **M. Bastian**[2], **T. Gaboriau**[1], **N. Salamin**[1]

[1]Department of Computational Biology, Université de Lausanne, Lausanne, Switzerland

[2]Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Université Lyon 1, Villeurbanne, France

thibault.latrille@ens-lyon.org

October 2, 2023

Abstract

To quantify selection acting on a trait, methods have been developed using either within or between-species variation. However, methods using within-species variation do not integrate the changes at the macro-evolutionary scale. Conversely, current methods using between-species variation usually discard within-species variation, thus not accounting for processes at the micro-evolutionary scale. The main goal of this study is to define a neutrality index for a quantitative trait, by combining within- and between-species variation. This neutrality index integrates nucleotide polymorphism and divergence for normalizing trait variation. As such, it does not require estimation of population size nor of time of speciation for normalization. Our index can be used to seek deviation from the null model of neutral evolution, and test for diversifying selection. Applied to brain mass and body mass at the mammalian scale, we show that brain mass is under diversifying selection. Finally, we show that our test is not sensitive to the assumption that population sizes, mutation rates and generation time are constant across the phylogeny, and automatically adjust for it.

***K*eywords** Quantitative genetics · Trait evolution · Selection · Phylogenetics · Population genetics

Introduction

Determining whether a trait is under a particular regime of selection has been a long-standing goal in evolutionary biology. Fundamentally, distinguishing neutral evolution from selection requires determining which selective regime is supported by the observed variation of traits or sequences. The variation of phenotypes (traits) and genotypes (sequences) can be observed at different scales, across different development stages at the individual level, across different individuals and populations at the species level, and finally across different species at the phylogenetic level. All these systems require different assumptions and methodologies, and

the endeavor to determine the selective regime for a given trait has thus incorporated theories, methods, and developments across various fields of evolutionary biology such as quantitative genetics, population genetics, phylogenetics and comparative genomics (Lynch & Walsh, 1998; Walsh & Lynch, 2018).

Leveraging individual variations within the same species, Genome-Wide Association Studies (GWAS) in humans have shown that traits are mostly polygenic (many loci associated with a given trait) and under stabilizing selection, while the loci affecting those traits are mostly pleiotropic (many traits associated with a given locus) with additive effects (Sella & Barton, 2019; Simons et al., 2018). Across several populations, by contrasting both trait and genetic differentiation, Q_{ST}–F_{ST} methods have been used to determine the selective regime and to quantify the strength of selection acting on a trait (Leinonen et al., 2008; Merilä & Crnokrak, 2001). A trait differentiation (Q_{ST}) higher than genetic differentiation (F_{ST}) is interpreted as a signature of diversifying selection due to adaptation in different optimum trait value in the different populations (Lamy et al., 2012). Contrarily, Q_{ST} lower than F_{ST} is interpreted as a signature of stabilizing selection. However, Q_{ST}–F_{ST} methods have been found to require many populations (O'Hara & Merilä, 2005), and that various factors can generate a spurious signal of selection (Edelaar et al., 2011; Pujol et al., 2008). Moreover, the test for diversifying selection is limited to recent local adaptation since the test is based on the variation observed within a single species. To disentangle selection from neutral evolution, trait variation can also be observed at a larger time scale. For example, change in mean trait value accumulates linearly with time of divergence from a sister species, and also proportionally to the trait variance (Lande, 1980a; Turelli, 1984). Empirically, this effect can be observed for genes with larger within-species variation in gene expression level, which exhibits a faster accumulation of divergence in mean expression level (Khaitovich et al., 2004). Altogether, both the trait variance and the evolution in mean value can be used to test for trait selection in a pair of species (Walsh & Lynch, 2018).

To disentangle neutral evolution and selection, trait evolution can also be observed at a larger time scale. For example, change in mean trait value accumulates linearly with time of divergence from a sister species, and also proportionally to the trait variance (Lande, 1980a; Turelli, 1984). Gene expression exhibits a similar accumulation as divergence in expression accumulates faster for genes with large within-species variation (Khaitovich et al., 2004). Altogether, both the trait variance and the evolution in mean value can be used to test for trait selection in a pair of species (Walsh & Lynch, 2018).

Alternatively, by accounting for the underlying relationships between several species, the selective regime for a quantitative trait can also be tested at the phylogenetic scale (Felsenstein, 1985). Under neutral evolution, the change in mean trait value along a given branch of the tree is normally distributed, with a variance proportional to divergence time (Hansen & Martins, 1996). As a result, the mean trait value can be modeled as a Brownian process branching at every node of the tree (Hansen & Martins, 1996; Harmon, 2018). Reconstructing the trait variation along the whole phylogeny as a Brownian process can thus constitute a null model of neutral trait evolution. Deviations from the assumptions of the Brownian process are however well known. When trait variation is constraint because of optimum mean trait values across or between species, the pattern of evolution can be modeled by the Ornstein-Uhlenbeck processes, which is often interpreted as a signature of stabilizing selection (Catalán et al., 2019). Alternatively, a trend in the Brownian process

(the tendency of a trait to evolve in a certain direction without fixed optimum) is interpreted as a signature of directional selection at the phylogenetic scale (Silvestro et al., 2019). However, studies have shown that such comparative approaches are subject to different biases (Harmon, 2018). First, a trait under stabilizing selection for which the optimal trait value is also evolving as a Brownian process will not deviate from a Brownian process, and thus be wrongly classified as neutral (Hansen & Martins, 1996). In other words, the better fit of a Brownian process does not necessarily constitute proof of the neutral model. Second, even for a trait evolving under a neutral regime, the Ornstein-Uhlenbeck process might sometimes be statistically preferred over a Brownian process due to sampling artifacts (Cooper et al., 2016; Price et al., 2022; Silvestro et al., 2015). Those limitations, altogether with the use of mean trait estimates leaving out the variance in traits between individuals, easily generate misclassification of selection from methods at the phylogenetic scale.

At the frontier between micro and macroevolution, comparative methods at the phylogenetic scale have acknowledged the importance of modeling within-species variation together with changes in mean trait value to either describe measurement errors (Hansen & Bartoszek, 2012; Lynch, 1991), incorporate values for individuals (Felsenstein, 2008) or to scale the rate of change in mean trait value (Gaboriau et al., 2020; Gaboriau et al., 2023; Kostikova et al., 2016). Within-species variation has also been used to infer diversifying selection by estimating the ratio of between to within variation of many traits and test for deviation from the average ratio across traits (Rohlfs et al., 2014; Rohlfs & Nielsen, 2015). Here, our goal was again to use both variances between and within species to determine the selective regime of a quantitative trait. We build a novel framework that integrates trait variation at the phylogenetic and population scales together with estimates of molecular divergence at both scales. It allowed us to define an expected ratio of normalized variance between and within species while setting the threshold of this ratio for neutral, stabilizing, and diversifying selection. The ratio that we propose can be considered as a neutrality index for a any quantitative trait articulating trait and nucleotide variation within and between species. Importantly, our neutrality index also leverages nucleotide divergence and polymorphism to normalize trait variation at both scales, such that it does not require estimating population size (within-species) or speciation time (between species). From the field of population genetics, our study can be seen as the macro-evolutionary generalization of Q_{ST}–F_{ST} methods to account for phylogenetic relationships between species. From the field of phylogenetics, our study can be seen as an alternative to the EVE model (Rohlfs et al., 2014; Rohlfs & Nielsen, 2015) for a single trait, where we set a threshold for neutral evolution by leveraging species nucleotide polymorphism and divergence.

Materials and Methods

Neutrality index for a quantitative trait

Prior to developing our neutrality index, we review theoretical expectations for variations of quantitative traits and genomic sequences under neutral evolution for both within- and between-species variation.

Within-species trait variations

For a given trait, the genetic architecture is mainly defined by the number of loci encoding the trait (L) and the random additive effect of a mutation on the trait (a). New mutations are generating trait variance and the average effect of a mutation on the trait is $\sigma^2_{\mathrm{M}} = L \cdot \mathbb{E}[a^2]$. At the individual level, the mutational variance (V_{M}) is the rate at which new mutations contribute to the trait variance per generation. As shown in Lande (1979, 1980b), V_{M} is a function of the mutation rate per loci per generation (μ) and σ^2_{M}:

$$V_{\mathrm{M}} = 2\mu \cdot \sigma^2_{\mathrm{M}}. \tag{1}$$

While in an infinitesimal model mutations supply new genetic variants, random genetic drift depletes standing variation (Barton et al., 2017; Sella & Barton, 2019; Turelli, 2017). For a neutral trait at equilibrium between mutation and drift (Lynch et al., 1998), the additive genetic variance in a species (V_{A}) is a function of the mutational variance (V_{M}) and the effective number of individuals in the population (N_{e}):

$$V_{\mathrm{A}} = 2N_{\mathrm{e}} \cdot V_{\mathrm{M}}, \tag{2}$$

$$= 4N_{\mathrm{e}} \cdot \mu \cdot \sigma^2_{\mathrm{M}} \text{ from eq. 1.} \tag{3}$$

For any neutral genomic region of interest, the nucleotide diversity, π, is measured as the number of mutations segregating in the population divided by the length of the region. Any segregating mutations will eventually reach fixation or extinction due to random genetic drift and π is also at a balance between mutations and drift. As shown in Tajima (1989), π is a function of the mutation rate per loci per generation (μ) and the effective population size (N_{e}):

$$\pi = 4N_{\mathrm{e}} \cdot \mu. \tag{4}$$

We define σ^2_{W} as the ratio of additive genetic variance of the trait (V_{A}) over π of any neutral genomic region of interest. This ratio allows removing the effect of N_{e} and μ, which are parameters not related to the genetic architecture of the trait, giving σ^2_{W} as a proxy of σ^2_{M}:

$$\sigma^2_{\mathrm{W}} \overset{\text{def}}{=} \frac{V_{\mathrm{A}}}{\pi}, \tag{5}$$

$$= \frac{4N_{\mathrm{e}} \cdot \mu \cdot \sigma^2_{\mathrm{M}}}{4N_{\mathrm{e}} \cdot \mu} \text{ from eq. 1 and 4,} \tag{6}$$

$$= \sigma^2_{\mathrm{M}}. \tag{7}$$

The additive genetic variance is also equal to the observed phenotypic variance (V_P) multiplied by narrow-sense heritability (h^2; (Hill et al., 2008)), which leads to σ^2_{W} being a function of V_P and h^2:

$$\sigma^2_{\mathrm{W}} = \frac{h^2 \cdot V_P}{\pi}. \tag{8}$$

Between-species trait variations

For a given species, we denote by $\bar{P}_t$ the mean value of the trait across the individuals in the species at generation t. If the trait is neutral and encoded by many loci as assumed by the infinitesimal model, $\bar{P}_t$

evolves as a Brownian process (Felsenstein, 1985; Hansen & Martins, 1996). The variance of $\bar{P}_t$ after t generations, $\mathrm{Var}\left[\bar{P}_t\right]$ is given (Hansen & Martins, 1996) by:

$$\mathrm{Var}\left[\bar{P}_t\right] = \frac{V_\mathrm{A}}{N_\mathrm{e}} \cdot t \tag{9}$$

$$= 4t \cdot \mu \cdot \sigma^2_\mathrm{M}, \text{ from eq. 3,} \tag{10}$$

Moreover, for any genomic region under neutral evolution, some mutations will eventually reach fixation due to random genetic drift, resulting in a substitution of a nucleotide at the species level. The probability of fixation ($\mathbb{P}_\mathrm{fix}$) of a neutral mutation is $1/2N_\mathrm{e}$ (Kimura, 1962). We can derive the substitution rate per generation q as the number of mutations per generation ($2N_\mathrm{e} \cdot \mu$) multiplied by the probability of fixation for each newly arisen mutations $\mathbb{P}_\mathrm{fix}$ (McCandlish & Stoltzfus, 2014), giving:

$$q = 2N_\mathrm{e} \cdot \mu \cdot \mathbb{P}_\mathrm{fix}, \tag{11}$$

$$= 2N_\mathrm{e} \cdot \mu \cdot \frac{1}{2N_\mathrm{e}}, \tag{12}$$

$$= \mu. \tag{13}$$

That is, if mutations are neutral, the rate of substitution within a genomic region equals the rate at which new mutations arise per generation for the same genomic region (Kimura, 1968).

After t generations and assuming that no multiple substitutions occurred at the same site, the nucleotide divergence d, which is the fraction of the genomic region that generated a substitution, will be t multiplied by the nucleotide substitution rate per generation (q):

$$d = t \cdot q \tag{14}$$

$$= t \cdot \mu \text{ from eq. 13.} \tag{15}$$

We define σ^2_B as the variance in the mean trait value ($\mathrm{Var}\left[\bar{P}_t\right]$) normalized by the nucleotide divergence of any neutral genomic region (d). This ratio allows removing the effect of t and μ, which are parameters not related to the genetic architecture of the trade, giving σ^2_B as another proxy of σ^2_M:

$$\sigma^2_\mathrm{B} \overset{\mathrm{def}}{=} \frac{\mathrm{Var}\left[\bar{P}_t\right]}{4d}, \tag{16}$$

$$= \frac{4t \cdot \mu \cdot \sigma^2_\mathrm{M}}{4t \cdot \mu} \text{ from eq. 10 and 15,} \tag{17}$$

$$= \sigma^2_\mathrm{M}. \tag{18}$$

Neutrality index

The variability between either individuals or species can be obtained for both quantitative traits and genomic sequences. At the population level, the variability of the trait between individuals can be combined with the nucleotide diversity of any neutrally evolving genomic region to obtain σ^2_W, which equals σ^2_M if the trait is neutrally evolving (see above). At the phylogenetic level, the variability of the mean trait value between species can be combined with the nucleotide divergence of any neutrally evolving genomic region to obtain σ^2_B. Similarly, $\sigma^2_\mathrm{B} = \sigma^2_\mathrm{M}$ if the trait is neutrally evolving and the genetic architecture of the trait has not

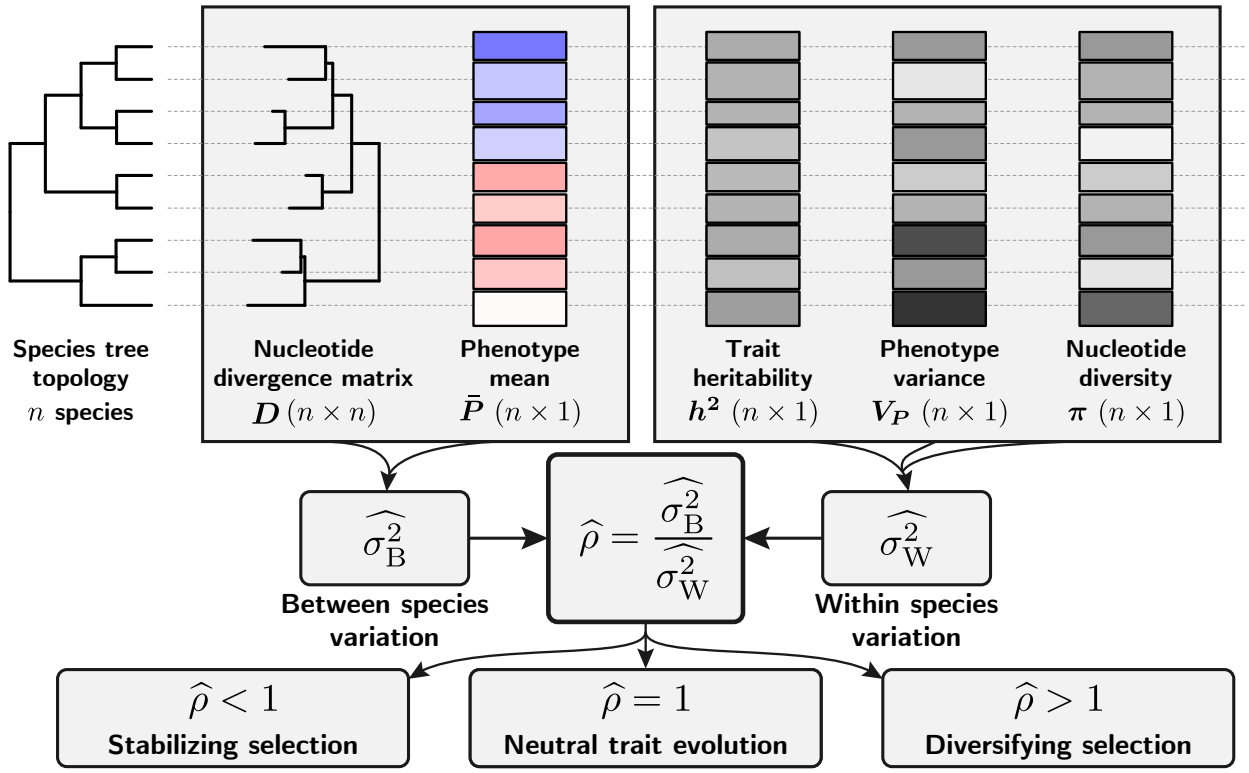


Figure 1: Between species, the change along the phylogeny of the mean phenotypic trait allows the estimation of between-species trait variation, $\widehat{\sigma^2_{\mathrm{B}}}$, which is normalized by nucleotide divergence. Within species, the genetic variance allows the estimation of within-species trait variation, $\widehat{\sigma^2_{\mathrm{W}}}$, which is normalized by nucleotide diversity. $\widehat{\rho}$ is the ratio of $\widehat{\sigma^2_{\mathrm{B}}}$ over $\widehat{\sigma^2_{\mathrm{W}}}$. Under neutral evolution, $\widehat{\rho}$ is expected to be equal to one. Under diversifying selection, the trait is heterogeneous between species, but homogeneous within species, leading to $\widehat{\rho}$ greater than one. Under stabilizing selection, the trait is homogeneous between species, leading to $\widehat{\rho}$ smaller than one. Importantly, the sequence from which nucleotide diversity and divergence are estimated should be neutrally evolving, but they are not necessarily linked to the quantitative trait under study, they allow for discarding the confounding effect on mutation rate diversity, population size and divergence time.

changed along the phylogenetic tree. We thus have, for a neutrally evolving trait:

$$\sigma^2_{\mathrm{W}} = \sigma^2_{\mathrm{B}} \text{ from eq. 7 and 18,} \tag{19}$$

$$\Rightarrow \rho \overset{\text{def}}{=} \frac{\sigma^2_{\mathrm{B}}}{\sigma^2_{\mathrm{W}}} = 1. \tag{20}$$

We define a neutrality index $\rho = \sigma^2_{\mathrm{B}}/\sigma^2_{\mathrm{W}}$ that will equal 1 for a trait evolving neutrally. Both σ^2_{B} and σ^2_{W} can be estimated using quantitative trait and genomic sequences within and between species, while neither the mutation rate (μ), nor the effective population size (N_{e}) or time of divergence (t) need to be estimated. Moreover, the sequence from which π and d are estimated should be neutrally evolving, but they are not necessarily linked to the quantitative trait under study.

Estimate

Based on the comparative framework that can account for phylogenetic inertia (Felsenstein, 1985; O'Meara et al., 2006), we provide a maximum likelihood estimate for ρ as well as a Bayesian estimate to derive posterior probabilities that the null model of neutrality (i.e. $\rho = 1$) is rejected.

Maximum likelihood estimate

At the phylogenetic scale, for n taxa in the tree, $\boldsymbol{D}$ $(n \times n)$ is the distance matrix computed from the branch lengths (d as nucleotide divergence in units of substitutions per site) and the topology of the phylogenetic tree. The diagonal $\boldsymbol{D}_{i,i}$ represents the total distances from the root of the tree to each taxon (i). The off-diagonal elements ($\boldsymbol{D}_{i,j} = \boldsymbol{D}_{j,i}$) are the distances between the root and the most recent common ancestor of taxa i and j. The state P_0 at the root of the tree for the trait can be estimated from the $n \times 1$ vector of mean trait values $\bar{\boldsymbol{P}}$ at the tips of the tree using maximum likelihood (O'Meara et al., 2006):

$$P_0 = \left(\mathbf{1}^\intercal \times \boldsymbol{D}^{-1} \times \mathbf{1}\right)^{-1} \cdot \left(\mathbf{1}^\intercal \times \boldsymbol{D}^{-1} \times \bar{\boldsymbol{P}}\right), \tag{21}$$

where $\mathbf{1}$ is an $n \times 1$ column vector of ones.

Finally, between-species variation $\widehat{\sigma^2_{\mathrm{B}}}$ is estimated as (O'Meara et al., 2006):

$$\widehat{\sigma^2_{\mathrm{B}}} = \frac{1}{4} \frac{\left(\bar{\boldsymbol{P}} - P_0 \cdot \mathbf{1}\right)^\intercal \times \boldsymbol{D}^{-1} \times \left(\bar{\boldsymbol{P}} - P_0 \cdot \mathbf{1}\right)}{n - 1}. \tag{22}$$

For a given species i with inter-individual data available, additive genetic variance of a trait ($V_{\mathrm{A},i}$) is the product of heritability (h^2_i) and phenotypic variance ($V_{P,i}$). The ratio of $V_{\mathrm{A},i}$ over nucleotide diversity of neutrally evolving sequences (π_i) is a sample estimate of σ^2_{W}. Averaged across all species, we obtain the estimate $\widehat{\sigma^2_{\mathrm{W}}}$ as:

$$\widehat{\sigma^2_{\mathrm{W}}} = \frac{1}{n} \sum_{i=1}^{n} \frac{V_{\mathrm{A},i}}{\pi_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{V_{P,i} \cdot h^2_i}{\pi_i}. \tag{23}$$

As depicted in fig. 1, the neutrality index is estimated as:

$$\widehat{\rho} = \frac{\widehat{\sigma^2_{\mathrm{B}}}}{\widehat{\sigma^2_{\mathrm{W}}}}. \tag{24}$$

Bayesian estimate

The Bayesian framework allows obtaining the posterior distribution of the neutrality index ($\widehat{\rho}$) for a given trait. Even though $\widehat{\rho}$ is estimated independently for each trait of interest in the maximum likelihood framework (previous section), here we generalize to K traits co-varying along the phylogenetic tree using the *BayesCode* software (Latrille et al., 2021). Trait variation along the phylogenetic tree is modeled as a K-dimensional Brownian process $\boldsymbol{\mathcal{B}}$ $(1 \times K)$ starting at the root and branching along the tree topology (Huelsenbeck & Rannala, 2003; Lartillot & Poujol, 2011; Lartillot & Delsuc, 2012; Latrille et al., 2021). The rate of change of the Brownian process is determined by the positive semi-definite and symmetric covariance matrix between traits $\boldsymbol{\Sigma}$ $(K \times K)$. The off-diagonal elements of $\boldsymbol{\Sigma}$ are the covariance between traits, and the diagonal elements are the variance of each trait, thus corresponding to $\widehat{\sigma^2_{\mathrm{B}}}$ (see section S2.1). With an

inverse Wishart distribution as the prior on the covariance matrix, the posterior on $\mathbf{\Sigma}$, conditional on $\mathcal{B}$ is also an invert Wishart distribution (see section S2.2). We used Metropolis-Hastings algorithm to sample $\boldsymbol{\mathcal{B}}$, while the posterior distribution of $\mathbf{\Sigma}$ is sampled using Gibbs sampling. For each trait and each species, the prior on heritability (h^2) for each species is set as a uniform distribution with user-defined boundaries. Heritability and phenotypic variance for each trait are combined with nucleotide diversity to compute $\widehat{\sigma^2_{\mathrm{W}}}$ for each species before being averaged across species (as in eq. 23). From $\widehat{\sigma^2_{\mathrm{W}}}$ and $\mathbf{\Sigma}$, the posterior distribution of $\widehat{\rho}$ (as in eq. 24) is obtained for each trait. The posterior distribution of $\widehat{\rho}$ thus allows testing for deviation from neutrality (Fig. 1), for example, by computing $\mathbb{P}[\widehat{\rho} > 1]$ to test for evidence of diversifying selection and $\mathbb{P}[\widehat{\rho} < 1]$ to test for evidence of stabilizing selection.

Applicability to empirical data

Our method assumes that the narrow-sense heritability (h^2) of a trait is known such as to estimate additive genetic variance (V_{A}) from phenotypic variance (V_P) as $V_{\mathrm{A}} = h^2 \cdot V_P$. Fortunately, if heritability is not known, the test for diversifying selection can still be performed, although it is underpowered. Indeed, if the additive genetic variance is substituted by phenotypic variance, it is equivalent to assuming complete heritability ($h^2 = 1$). Because $h^2 \leq 1$ by definition, we overestimate the within-species variation and thus underestimate $\widehat{\rho}$. It is, however, possible to test for diversifying selection because testing for $\widehat{\rho} > 1$ while using phenotypic variance instead of additive genetic variance means that knowing the additive genetic variance would have only increased the evidence for diversifying selection. Similarly, using the broad-sense heritability (H^2) instead of narrow-sense heritability (h^2) results in an underestimation of $\widehat{\rho}$ since $h^2 \leq H^2$. In contrast, the test for stabilizing selection is invalid if $\widehat{\rho}$ is underestimated. Several assumptions made by our test might not hold on empirical data and their consequences on the neutrality index and the test that can be performed are shown in Table 2.

Simulation

We tested the performance of our neutrality index (ρ) to detect selection on a quantitative trait using simulations. We performed simulations under different selective regimes (neutral, stabilizing, diversifying), different demographic histories (constant or fluctuating population size) and different evolution of the mutation rate (constant or fluctuating). Simulations were individual-based and followed a Wright-Fisher model with mutation, selection and drift for a diploid population including speciation along a predefined ultrametric phylogenetic tree (Fig. 2A&B). Each individual phenotypic value was the sum of genotypic value and an environmental effect. The environmental effect was normally distributed with variance V_{E}. We assumed that the genotypic value was encoded by $L = 5,000$ loci, with each locus contributing an additive effect that was normally distributed with standard deviation $a = 1$ (Fig. 2A and section S1.1 for theoretical formulation). We assumed a trait with a narrow-sense heritability of $h^2 = 0.2$ and computed the theoretical V_{E} accordingly (see section S1.1). Assuming a diploid panmictic population of size $N_{\mathrm{e}} = 50$ at the root of the tree, and with non-overlapping generations, we simulated explicitly each generation along an ultrametric phylogenetic tree. For each offspring, the number of mutations was drawn from a Poisson distribution with mean $2 \cdot \mu \cdot L$, with the mutation rate per generation μ. From the empirical mammalian dataset (see next section), we computed

an average nucleotide divergence from the root to leaves of 0.18 and average genetic diversity of 0.00276. We scaled parameters in our simulations to fit plausible values for mammals. We thus used a mutation rate of $\mu = 0.00276/4N_e = 1.38 \times 10^{-5}$ per generation per locus and a total of $t = 0.18/1.38 \times 10^{-5} = 13,500$ generations from root to leaves, and the number of generations along each branch was proportional to the branch length.

The changes in log-μ and log-N_e along the lineages were both modeled by a geometric Brownian process ($\mathcal{B}(0, \sigma_\mu = 0.0086)$ and $\mathcal{B}(0, \sigma_{N_e} = 0.0086)$, which led to a standard deviation of $0.0086 \cdot \sqrt{13,500} = 1.0$ in log-space from root to leaves. An Ornstein-Uhlenbeck process was overlaid to the instant value of log-N_e provided by the geometric Brownian process to account for short-term changes between generations ($\mathrm{OU}(0, \sigma_{N_e} = 0.1, \theta_{N_e} = 0.9)$). The geometric Brownian motion accounted for long-term fluctuations (low rate of changes σ_{N_e} but unbounded), while the Ornstein-Uhlenbeck introduced short-term fluctuations (high rate of changes σ_{N_e} but bounded and mean-reverting). The simulation started from an initial sequence at equilibrium at the root of the tree and, at each node, the process was split until it finally reached the leaves of the tree. From a speciation process perspective, this was equivalent to an allopatric speciation over one generation.

A random genetic drift was introduced by resampling individuals at each generation, with each parent having a probability of being sampled that was proportional to its fitness (W). Selection was modeled as a one-dimensional Fisher's geometric landscape, with the fitness of an individual being a monotonously decreasing function of the distance between the individual and the optimal phenotype (Blanquart & Bataillon, 2016; Tenaillon, 2014). More specifically, the fitness of an individual was given by $W = e^{(P-\lambda)^2/\alpha}$, where P was the trait value of the individual, $\lambda = 0.0$ was the optimal trait value, and $\alpha = 0.02$ was the strength of selection. Mutations were considered as a displacement of the phenotype in the multidimensional space. Beneficial mutations moved the phenotype closer to the optimum, while deleterious mutations moved it further away. Stabilizing selection was implemented by fixing the optimum phenotype to a single value ($\lambda = 0.0$). Diversifying selection was implemented by allowing the optimum phenotype to move along the phylogenetic tree as a geometric Brownian process (Hansen, 1997) ($\lambda \sim \mathcal{B}(0, \sigma_\lambda = 1.0)$). Neutral evolution was implemented by fixing the fitness landscape ($W = 1$), which meant that each individual had the same probability of being sampled at each generation.

Nucleotide diversity (π) was measured as the heterozygosity of neutral markers that were simulated along the phylogenetic tree but not linked to the trait simulated. Nucleotide divergence (d) was measured as the number of substitutions per site of neutral markers along the branches of the phylogenetic tree. The additive genetic variance was measured as phenotypic variance multiplied by heritability. Heritability was estimated from the slopes of the regression of offspring's phenotypic trait values on parental phenotypic trait values (Lynch & Walsh, 1998) averaged over the last 10 simulated generations. Heritability was thus not a given parameter of the simulations, but rather measured as it would be in empirical data.

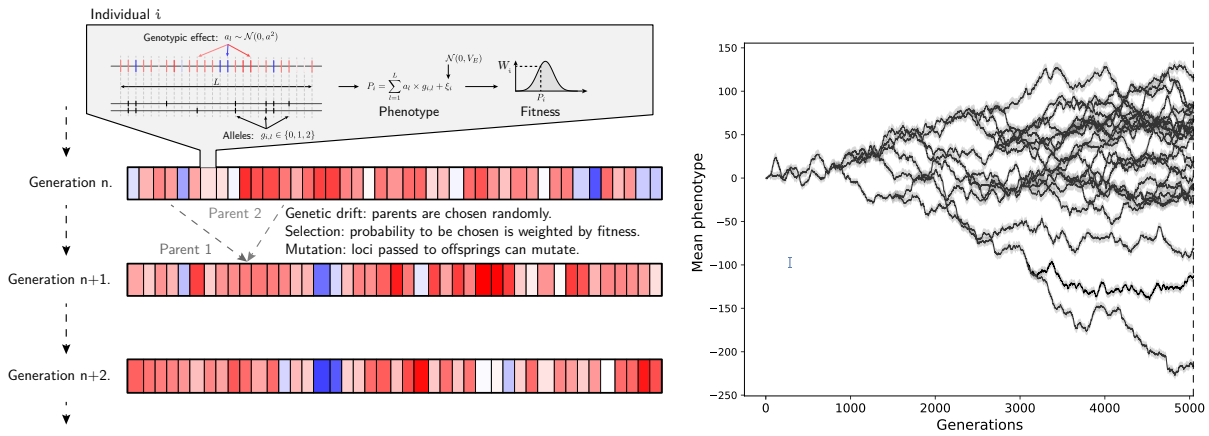


Figure 2: Wright-Fisher simulations with mutation, selection and drift. Left panel: For a given individual, the trait phenotypic value is the sum of genotypic value and a environmental effect (standard deviation V_{E}). The trait's genotypic value is encoded by L loci, with each locus contributing additively to the genotypic value. Parents are selected for reproduction to the next generation according to their phenotypic value, with a probability proportional to their fitness. Mutations are drawn from a Poisson distribution, with each locus having a probability μ to mutate. Drift is modeled by the resampling of parents. Right panel: examples of a trait evolving along a phylogenetic tree, with the mean phenotype (black line) and the variance of the trait genotypic value (gray area).

Empirical dataset

We analyzed a dataset of body and brain masses from mammals. The log-transformed values of body and brain masses were taken from Tsuboi et al. (2018). We removed individuals not marked as adults and split the data into males and females due to sexual dimorphism in body and brain masses. We discarded species with only one representative sample. The mammalian nucleotide diversity was obtained from the Zoonomia project (Genereux et al., 2020), with nucleotide divergence obtained on a set of neutral markers in Foley et al. (2023), and with nucleotide diversity measured as heterozygosity in Wilder et al. (2023).

We also analyzed a dataset of primate species, with the nucleotide variation obtained from Kuderna et al. (2023) and the quantitative trait variation also from Tsuboi et al. (2018), using the same filtering as for the mammalian dataset. However, the primate nucleotide divergence was not obtained on a set of neutral markers as for the mammalian dataset, but across the whole genome.

Results

Neutrality index

For a neutral trait, the genetic architecture, meaning the number of loci encoding the trait and the average effect of a mutation on the trait, is formally related to both within and between-species variation of the trait. We defined the neutrality index as $\rho = \sigma^2_{\mathrm{B}}/\sigma^2_{\mathrm{W}}$, which equals 1 for a neutral trait (see Materials and Methods), suggesting that traits for which this relationship was not verified were putatively under

selection. Under stabilizing selection, the variation between species is depleted because the mean trait value is maintained similar between different species, which leads to $\rho < 1$. In contrast, under diversifying selection, the variation between species is inflated because species will have potentially different trait values (Hansen, 1997), which leads to $\rho > 1$. Our neutrality index for a quantitative trait leveraged the data for any number of species, and took advantage of the signal over the whole phylogenetic tree, while at the same time taking into account phylogenetic inertia and addressing the non-independence between species (Fig. 1). This statistic was obtained as a maximum likelihood estimate ($\widehat{\rho}$), from eq. 23 and 22. We also devised a Bayesian estimate to obtain the posterior distribution of the neutrality index, and test for diversifying selection as $\mathbb{P}[\widehat{\rho} > 1]$, and stabilizing selection as $\mathbb{P}[\widehat{\rho} < 1]$.

Our neutrality index made a series of assumptions that we described in details in the Material and Methods section. Table 2 summarized these assumptions and outlined possible consequences for the neutrality test that we proposed.

Results against simulations

The inference framework was first tested on independently simulated datasets matching an empirically relevant mammalian empirical regime (see Materials and Methods). Under constant population size (N_e) and constant mutation rate (μ) across the phylogenetic tree (fig. 3, top row), we found no false negative for simulations of stabilizing ($\mathbb{P}[\widehat{\rho} < 1] > 0.975$; blue in fig. 3) or diversifying ($\mathbb{P}[\widehat{\rho} > 1] > 0.975$; red in fig. 3) selection. For simulations under neutral evolution, 77% of those were correctly identified ($0.025 \leq \mathbb{P}[\widehat{\rho} > 1] \leq 0.975$; yellow in fig. 3), while 21% and 2% were wrongly detected as stabilizing or diversifying selection, respectively. Once we introduced fluctuating N_e and μ (Fig. 3, bottom row), our ability to identify simulations under either diversifying or stabilizing selection remained the same with all cases detected correctly. For simulations under neutral evolution, 51% of the simulations were correctly detected ($0.025 \leq \mathbb{P}[\widehat{\rho} > 1] \leq 0.975$), while 49% were detected as stabilizing selection ($\mathbb{P}[\widehat{\rho} < 1] > 0.975$) and none as diversifying selection.

Results on empirical data

For mammalian body and brain mass, we obtained male (♂) and female (♀) trait variations. Combined with nucleotide diversity and divergence, we estimated $\widehat{\rho}$ and posterior probabilities of diversifying selection under different assumptions for trait heritability as shown in the Table 1. Assuming complete heritability, brain mass was found to be under diversifying selection with posterior probabilities of 0.0 for both males and females. If we assumed that heritability (h^2) of body mass was uniformly distributed between 20% and 40% (Hu et al., 2022), posterior probabilities of diversifying selection became 0.635 for males and 0.324 for females. Mammalian brain mass was found to be under diversifying selection with posterior probabilities of 0.877 for males and 0.972 for females when complete heritability was assumed. Assuming a uniform distribution between 20% and 40% for heritability led to posterior probabilities of diversifying selection of 1.0 for both males and females.

We also analyzed a similar dataset for body mass focusing this time only at Primates (Table 1). For primates body mass, we found posterior probabilities of diversifying selection of 1.0 for males and 0.914

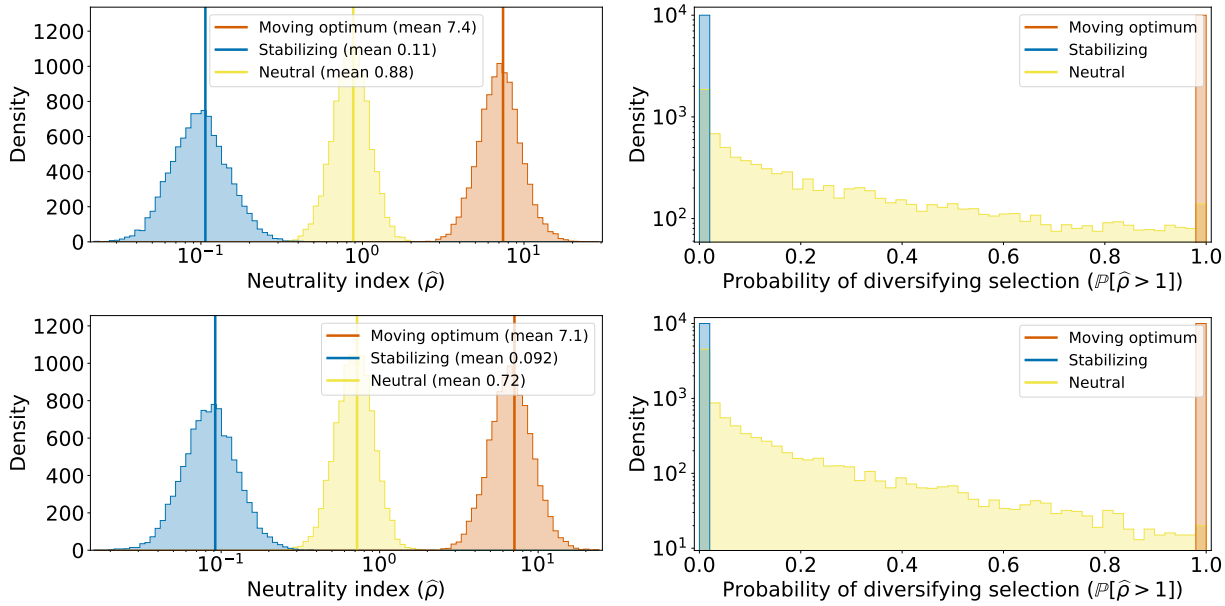


Figure 3: $10,000$ simulations of trait evolution along a phylogenetic tree under different selection regimes. Traits simulated under stabilizing selection (blue), under a neutral evolution (yellow), and under a moving optimum (red). Histogram of ratio of between-species trait variation ($\widehat{\sigma^2_{\mathrm{B}}}$) over within-species trait variation $\widehat{\sigma^2_{\mathrm{W}}}$ with $\widehat{\rho} = \widehat{\sigma^2_{\mathrm{B}}}/\widehat{\sigma^2_{\mathrm{W}}}$ estimated from each simulated data (left) and probabilities of $\widehat{\rho}$ being greater than 1 (right). Effective population size (N_{e}) and mutation rate (μ) were either constant (top row), or fluctuating as a Brownian process along the phylogenetic tree (bottom row).

for females when assuming a uniform distribution for the heritability of body mass between 20% and 40%. Assuming complete heritability of body mass did not change the posterior probability for males, but increased the one for female to 1.0. Evidence for diversifying selection on body mass was therefore more pronounced in Primates than in mammals. However, the genetic markers used to normalize trait variance with nucleotide divergence were not necessarily neutral, which could create spurious false positives by artificially inflating $\widehat{\rho}$ (Table 2 and methods).

Discussion

In this study, we proposed a neutrality index for a quantitative trait that can be used within a statistical framework to test for selection. Our neutrality index for a trait, ρ, is calculated as the ratio of the normalized within- to between-species variation and it allowed the identification of the evolutionary regime of a quantitative trait. At the phylogenetic scale, trait variation between species was normalized by sequence divergence obtained from a neutral set of markers. Similarly, trait variation within species was normalized by sequence polymorphism obtained also from a neutral set of markers. Our estimate of $\widehat{\rho}$ could be tested for deviation from the value of 1.0 expected under the null hypothesis of neutrality. Technically, the neutrality index can be estimated either as a maximum likelihood point estimate, or as a mean posterior estimate from a Bayesian implementation (see section S3). The latter also enabled the estimation of the posterior credible interval to

Dataset	Trait	h^2	Sex	n	$\widehat{\rho}$	95% CI for $\widehat{\rho}$	$\mathbb{P}[\widehat{\rho} > 1]$
Mammals	Body mass	1.0	♂	36	0.340	0.217-0.523	0.000
Mammals	Body mass	1.0	♀	26	0.277	0.160-0.490	0.000
Mammals	Body mass	$\mathcal{U}(0.2, 0.4)$	♂	36	1.124	0.721-1.754	0.635
Mammals	Body mass	$\mathcal{U}(0.2, 0.4)$	♀	26	0.936	0.523-1.715	0.324
Mammals	Brain mass	1.0	♂	36	1.351	0.851-2.173	0.877
Mammals	Brain mass	1.0	♀	26	1.727	0.991-2.938	0.972
Mammals	Brain mass	$\mathcal{U}(0.2, 0.4)$	♂	36	4.527	2.831-7.091	1.000
Mammals	Brain mass	$\mathcal{U}(0.2, 0.4)$	♀	26	6.001	3.288-10.941	1.000
Primates	Body mass	1.0	♂	71	0.558	0.401-0.784	0.000
Primates	Body mass	1.0	♀	65	0.389	0.278-0.547	0.000
Primates	Body mass	$\mathcal{U}(0.2, 0.4)$	♂	71	1.875	1.288-2.695	1.000
Primates	Body mass	$\mathcal{U}(0.2, 0.4)$	♀	65	1.296	0.899-1.821	0.914
Primates	Brain mass	1.0	♂	71	1.929	1.395-2.616	1.000
Primates	Brain mass	1.0	♀	65	1.950	1.399-2.790	1.000
Primates	Brain mass	$\mathcal{U}(0.2, 0.4)$	♂	71	6.479	4.658-8.944	1.000
Primates	Brain mass	$\mathcal{U}(0.2, 0.4)$	♀	65	6.522	4.664-9.294	1.000

Table 1: Test of diversifying selection on a mammal and a primate dataset, by splitting males (♂) and females (♀). Traits considered were body mass or brain mass (log-transformed). Heritability (h^2) was either assumed complete ($h^2 = 1.0$) or uniformly distributed between 20% and 40% ($h^2 \sim \mathcal{U}(0.2, 0.4)$). n was the number of species in the dataset. $\widehat{\rho}$ was the posterior estimate of our neutrality index, with the 95% credible interval (CI) for $\widehat{\rho}$ also computed. $\mathbb{P}[\widehat{\rho} > 1]$ was the estimated posterior probability of diversifying selection.

test for departure from a neutrally evolving trait (e.g. $\mathbb{P}[\widehat{\rho} > 1]$). We tested our statistical procedure against simulated data and showed that our test was able to correctly detect simulations under diversifying selection (test of $\widehat{\rho} > 1$) or under stabilizing selection (test of $\widehat{\rho} < 1$). However, our test detected a spurious signal of stabilizing selection ($\widehat{\rho} < 1$) when we simulated the evolution of a neutral trait. We thus argue that our method should be used to detect diversifying selection, but that it had low accuracy to detect stabilizing selection due to false positives.

Our results showed that our method significantly improved over currently available methods to detect selection acting on a trait at the phylogenetic scale. Current methods relying on evolution of the mean trait value between species also tend to statistically prefer a model of stabilizing selection over a Brownian process when the trait is neutral (Cooper et al., 2016; Price et al., 2022; Silvestro et al., 2015). Our approach could in theory be applied to detect stabilizing selection at the phylogenetic scale, but we showed that it did not have the statistical power to identify those cases. In contrast, we showed that our method was able to

Broken assumption	Consequences	$\widehat{\sigma^2_{\mathrm{W}}}$	$\widehat{\sigma^2_{\mathrm{B}}}$	Test $\rho > 1$	Test $\rho < 1$
Trait encoded by few loci	Between-species trait variation is underestimated	–	Underestimated	Conservative	Invalid
Sexual dimorphism	Within-species trait variation is overestimated	Overestimated	–	Conservative	Invalid
Inbreeding	Nucleotide diversity (π) is underestimated	Overestimated	–	Conservative	Invalid
Markers for polymorphism are negatively selected	Nucleotide diversity (π) is underestimated	Overestimated	–	Conservative	Invalid
Markers for polymorphism are positively selected	Nucleotide diversity (π) is underestimated	Overestimated	–	Conservative	Invalid
Markers for divergence are positively selected	Nucleotide divergence (d) is overestimated	–	Underestimated	Conservative	Invalid
Markers for polymorphism under balanced selection	Nucleotide diversity (π) is overestimated	Underestimated	–	Invalid	Conservative
Markers for divergence are negatively selected	Nucleotide divergence (d) is underestimated	–	Overestimated	Invalid	Conservative
Multiple nucleotide substitutions at the same locus	Nucleotide divergence (d) is underestimated	–	Overestimated	Invalid	Conservative

Table 2: Assumptions breaks and their consequences on the estimation of within-species variation ($\widehat{\sigma^2_{\mathrm{W}}}$), between-species variation ($\widehat{\sigma^2_{\mathrm{B}}}$), and on the neutrality index $\rho = \widehat{\sigma^2_{\mathrm{B}}}/\widehat{\sigma^2_{\mathrm{W}}}$. The last two columns indicate whether the test for diversifying selection ($\rho > 1$) and for stabilizing selection $\rho < 1$ are conservative or invalid due to violated assumptions.

identify correctly cases of diversifying selection, which is a clear an improvement over current methods that model only mean trait value. Indeed, under diversifying selection, mean trait value will not deviate from a Brownian process, and thus cannot be distinguished from neutral evolution (Hansen & Martins, 1996; Harmon, 2018). For example, testing the selective regime in the expression level of the majority of genes led to the selection of a Brownian process as the prefered model and the interpretation that the expression was evolving neutrally (Catalán et al., 2019). Our diversity index has the advantage to discriminate the alternative model of diversifying selection from the neutral case by comparing within- and between-species variation correctly normalized to remove confounding factors. Our approach is not the first one to normalize between-species variation to detect selection, but this was done by using within-species variations (Rohlfs et al., 2014; Rohlfs & Nielsen, 2015) and not estimates of neutral molecular divergence as done in our study. These studies have further compared their statistic across a pool of traits, which allowed them to identify outlier traits putatively under diversifying selection but without testing for selection on a single trait at a time (Gillard et al., 2021; Rohlfs & Nielsen, 2015). Instead, our procedure can be applied to a single trait, estimating the neutrality index and giving a statistical test for departures from the null model of neutral evolution for a single test. Our diversity index opens new avenues to revisit these studies and better test for the selective regime affecting the quantitative traits, assuming we have access to genomic datasets to estimate nucleotide divergence and polymorphism.

The main novelty of our study was to use the nucleotide divergence and polymorphism to normalize trait variation between and within species. In the context of within species variation, $\mathrm{Q_{ST}}$–$\mathrm{F_{ST}}$ tests have been developed to compare trait and sequence across several populations to test for selection (Leinonen et al., 2013; Martin et al., 2008). Our neutrality index also used the genetic sequences from which nucleotide divergence and polymorphism are estimated. Although the sequences should be neutrally evolving, they do not have to be necessarily linked to the quantitative trait under study. Nucleotide variation allows normalizing for diversity driven by confounding factors such as population sizes (N_{e}), mutation rates (μ) and generation time (Hansen & Martins, 1996; Harmon, 2018). Thus our test avoids the estimation of the parameters, which are complex to correctly infer, and it also bypasses the estimation of divergence time, which was necessary in previous

approaches (Walsh & Lynch, 2018). But importantly, by normalizing with sequence variation, we also showed using simulated data that our test was not sensitive to the assumption that N_e, μ and generation time were constant across the phylogenetic tree, an unmet assumption empirically (Bergeron et al., 2023; Wilder et al., 2023). Indeed, under the neutral case of evolution, changes in N_e, μ and generation time impacted similarly trait and sequence variation. The normalization by nucleotide divergence and polymorphism automatically absorbed long-term and short-term changes in N_e, μ and generation time, which canceled out in the ratio of trait variation $\widehat{\rho}$.

Even though our test was developed for a quantitative trait, analogies with other tests of selection developed for molecular sequences also provided insight into its behavior. First, we acknowledge that our test took inspiration from the McDonald and Kreitman (1991) test devised for protein-coding DNA sequences, where synonymous mutations were used to determine the neutral expectation, and the inflation of divergence was compared to polymorphism within species. Second, because ρ was compared to 1, our test ultimately bear analogy to the codon-based test of selection, where the ratio of non-synonymous to synonymous substitutions (ω) is compared to 1 (Goldman & Yang, 1994; Muse & Gaut, 1994). As $\omega < 1$ is interpreted as purifying selection acting on the protein, $\rho < 1$ is interpreted as stabilizing selection acting on the trait. Similarly, the interpretation of adaptation for $\omega > 1$ is analogous to diversifying selection for $\rho > 1$. With this analogy in mind, we could leverage the vast literature discussing and interpreting the results of these tests and their pitfalls (Anisimova & Kosiol, 2009; Jensen et al., 2019; Nielsen, 2005). First, not rejecting the neutral null model of $\rho = 1$ did not necessarily imply that the trait was effectively neutral, since diversifying and stabilizing selection could compensate each other resulting in $\rho = 1$, analogously to $\omega = 1$ under a mix of adaptation and purifying selection (Nielsen, 2005). Second, empirical evidence for $\rho < 1$ did not rule out diversifying selection, but rather that this diversifying selection was not strong enough to overcome the stabilizing selection, similarly to strong purifying selection resulting $\omega < 1$ even though those genes and sites are under adaptation (Latrille et al., 2023). By explicitly modeling stabilizing selection as a moving optimum, it would theoretically be possible to tease apart the effect of diversifying and stabilizing selection in the context of quantitative traits to obtain a statistically more powerful test.

In the context of detecting diversifying selection on a trait, we argue that the main drawback of our method is that the additive genetic variance of the trait is required instead of the phenotypic variance. If phenotypic variance was used instead of additive genetic variance to estimate $\widehat{\rho}$, meaning that we assumed complete heritability, the neutrality index $\widehat{\rho}$ was ultimately underestimated. Similarly, using broad-sense heritability instead of narrow-sense heritability would result in underestimated $\widehat{\rho}$. In such context, the test of stabilizing selection ($\widehat{\rho} < 1$]) would be statistically invalid. However, the test of diversifying selection ($\widehat{\rho} > 1$) was underpowered although not invalided, meaning that absence of evidence would not be evidence of absence. As an example, even though we assumed complete heritability for brain mass, we uncovered diversifying selection in mammals since $\widehat{\rho} > 1$.

The development of our neutrality index was also based on several assumptions that could be relaxed in future studies. First, we cannot predict the behavior of our test in the context of population structures, gene flow and introgression. These factors should be thoroughly investigated using simulations. Second, loci

were assumed to contribute additively to the phenotype. Although the effects of dominance and epistasis is typically weak compared to the additive effects on the quantitative traits, their influence should be assessed (Crow, 2010; Hill et al., 2008). Third, the genetic architecture of the trait was assumed to be constant across the phylogenetic tree, whereas it might actually be variable among individuals and species (Huber et al., 2015; Tung et al., 2015). Such an assumption can theoretically be relaxed and changes in genetic architecture along the phylogenetic tree could jointly be estimated (Arnold et al., 2008; Gaboriau et al., 2020; Hohenlohe & Arnold, 2008; Kostikova et al., 2016). Finally, our Bayesian estimation could integrate uncertainty from the estimation of genetic variation, using sequences as input instead of estimated values of nucleotide diversity and divergence.

From an empirical point of view, our method required integrating genomic and trait variation, which could reduce the possible datasets to be used. However, such datasets will become more and more accessible and we showed the applicability of our method by applying it to the illustrative example of mammals brain and body mass. Because our test was also based on several assumptions that might not hold on empirical data, we also provided a table containing the main assumptions and their consequences on the neutrality index and the test that can be performed (Table 2). For example, at the primate scale, the evidence for $\widehat{\rho} > 1$ does not necessarily imply that the brain mass was evolving under diversifying selection since the markers used for nucleotide divergences were not neutral, which can lead to a spurious $\widehat{\rho} > 1$. In conclusion, our study provided a statistical framework to test for diversifying selection acting on a quantitative trait while integrating the trove of genomic data available both within and between species, and we believe that our new approach is a promising tool to investigate the evolution of quantitative traits.

Acknowledgements

We gratefully acknowledge the help of Nicolas Lartillot, Philippe Veber, Isabela Jeronimo do Ó, Anna Marcionetti, Julien Clavel, and Daniele Silvestro for their insightful discussions and Julien Joseph for his advice and reviews concerning this manuscript.

Competing interests:

The authors declare no conflicts of interest.

Data and materials availability:

The data that support the findings of this study are openly available in GitHub at github.com/ThibaultLatrille/MicMac. Snakemake pipeline, analysis scripts and documentation are available in the repository to replicate the study.

References

Anisimova, M., & Kosiol, C. (2009). Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Molecular Biology and Evolution*, *26*(2), 255–271. https://doi.org/10.1093/molbev/msn232

Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C., & Jones, A. G. (2008). Understanding the evolution and stability of the G-Matrix. *Evolution*, *62*(10), 2451–2461. https://doi.org/10.1111/j.1558-5646.2008.00472.x

Barton, N. H., Etheridge, A. M., & Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, *118*, 50–73. https://doi.org/10.1016/j.tpb.2017.06.001

Bergeron, L. A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M. F., Quintard, B., Hoffman, J. I., Li, Z., St. Leger, J., Shao, C., Stiller, J., Gilbert, M. T. P., Schierup, M. H., & Zhang, G. (2023). Evolution of the germline mutation rate across vertebrates. *Nature*, 1–7. https://doi.org/10.1038/s41586-023-05752-y

Blanquart, F., & Bataillon, T. (2016). Epistasis and the structure of fitness landscapes: Are experimental fitness landscapes compatible with fisher's geometric model? *Genetics*, *203*(2), 847–862. https://doi.org/10.1534/genetics.115.182691

Catalán, A., Briscoe, A. D., & Höhna, S. (2019). Drift and directional selection are the evolutionary forces driving gene expression divergence in eye and brain tissue of Heliconius butterflies. *Genetics*, *213*(2), 581–594. https://doi.org/10.1534/genetics.119.302493

Cooper, N., Thomas, G. H., Venditti, C., Meade, A., & Freckleton, R. P. (2016). A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, *118*(1), 64–77. https://doi.org/10.1111/bij.12701

Crow, J. F. (2010). On epistasis: Why it is unimportant in polygenic directional selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1544), 1241–1244. https://doi.org/10.1098/rstb.2009.0275

Edelaar, P., Burraco, P., & Gomez-Mestre, I. (2011). Comparisons between QST and FST—how wrong have we been? *Molecular Ecology*, *20*(23), 4830–4839. https://doi.org/10.1111/j.1365-294X.2011.05333.x

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, *125*(1), 1–15. https://doi.org/10.1086/284325

Felsenstein, J. (2008). Comparative methods with sampling error and within-species variation: Contrasts revisited and revised. *The American Naturalist*, *171*(6), 713–725. https://doi.org/10.1086/587525

Foley, N. M., Mason, V. C., Harris, A. J., Bredemeyer, K. R., Damas, J., Lewin, H. A., Eizirik, E., Gatesy, J., Karlsson, E. K., Lindblad-Toh, K., Zoonomia Consortium, Springer, M. S., & Murphy, W. J. (2023). A genomic timescale for placental mammal evolution. *Science*, *380*(6643), eabl8189. https://doi.org/10.1126/science.abl8189

Gaboriau, T., Mendes, F. K., Joly, S., Silvestro, D., & Salamin, N. (2020). A multi-platform package for the analysis of intra- and interspecific trait evolution. *Methods in Ecology and Evolution*, *11*(11), 1439–1447. https://doi.org/10.1111/2041-210X.13458

Gaboriau, T., Tobias, J. A., Silvestro, D., & Salamin, N. (2023, March 1). *Exploring the Macroevolutionary Signature of Asymmetric Inheritance at Speciation*. https://doi.org/10.1101/2023.02.28.530448

Genereux, D. P., Serres, A., Armstrong, J., Johnson, J., Marinescu, V. D., Murén, E., Juan, D., Bejerano, G., Casewell, N. R., Chemnick, L. G., Damas, J., Di Palma, F., Diekhans, M., Fiddes, I. T., Garber, M., Gladyshev, V. N., Goodman, L., Haerty, W., Houck, M. L., … Zoonomia Consortium. (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature*, *587*(7833), 240–245. https://doi.org/10.1038/s41586-020-2876-6

Gillard, G. B., Grønvold, L., Røsæg, L. L., Holen, M. M., Monsen, Ø., Koop, B. F., Rondeau, E. B., Gundappa, M. K., Mendoza, J., Macqueen, D. J., Rohlfs, R. V., Sandve, S. R., & Hvidsten, T. R. (2021). Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biology*, *22*(1), 103. https://doi.org/10.1186/s13059-021-02323-0

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *11*(5), 725–736.

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*, *51*(5), 1341–1351. https://doi.org/10.1111/j.1558-5646.1997.tb01457.x

Hansen, T. F., & Bartoszek, K. (2012). Interpreting the evolutionary regression: The interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology*, *61*(3), 413–425. https://doi.org/10.1093/sysbio/syr122

Hansen, T. F., & Martins, E. P. (1996). Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. *Evolution*, *50*(4), 1404–1417. https://doi.org/10.1111/j.1558-5646.1996.tb03914.x

Harmon, L. (2018). Phylogenetic comparative methods: Learning from trees.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLOS Genetics*, *4*(2), e1000008. https://doi.org/10.1371/journal.pgen.1000008

Hohenlohe, P. A., & Arnold, S. J. (2008). MIPoD: A hypothesis-testing framework for microevolutionary inference from patterns of divergence. *The American Naturalist*, *171*(3), 366–385. https://doi.org/10.1086/527498

Hu, Z.-L., Park, C. A., & Reecy, J. M. (2022). Bringing the Animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. *Nucleic Acids Research*, *50*(D1), D956–D961. https://doi.org/10.1093/nar/gkab1116

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., Shah, A., Gilles, B., Wirth, T., McMillan, W. O., & Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. *Heredity*, *114*(5), 515–524. https://doi.org/10.1038/hdy.2015.22

Huelsenbeck, J. P., & Rannala, B. (2003). Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution*, *57*(6), 1237–1247. https://doi.org/10.1111/j.0014-3820.2003.tb00332.x

Jensen, J. D., Payseur, B. A., Stephan, W., Aquadro, C. F., Lynch, M., Charlesworth, D., & Charlesworth, B. (2019). The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution*, *73*(1), 111–114. https://doi.org/10.1111/evo.13650

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W., & Pääbo, S. (2004). A neutral model of transcriptome evolution. *PLOS Biology*, *2*(5), e132. https://doi.org/10.1371/journal.pbio.0020132

Kimura, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics*, *47*(6), 713–719.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*(5129), 624–626.

Kostikova, A., Silvestro, D., Pearman, P. B., & Salamin, N. (2016). Bridging inter- and intraspecific trait evolution with a hierarchical bayesian approach. *Systematic Biology*, *65*(3), 417–431. https://doi.org/10.1093/sysbio/syw010

Kuderna, L. F. K., Gao, H., Janiak, M. C., Kuhlwilm, M., Orkin, J. D., Bataillon, T., Manu, S., Valenzuela, A., Bergman, J., Rousselle, M., Silva, F. E., Agueda, L., Blanc, J., Gut, M., de Vries, D., Goodhead, I., Harris, R. A., Raveendran, M., Jensen, A., … Marques Bonet, T. (2023). A global catalog of whole-genome diversity from 233 primate species. *Science*, *380*(6648), 906–913. https://doi.org/10.1126/science.abn7829

Lamy, J.-B., Plomion, C., Kremer, A., & Delzon, S. (2012). QST $<$ FST As a signature of canalization. *Molecular Ecology*, *21*(23), 5646–5655. https://doi.org/10.1111/mec.12017

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. *Evolution*, *33*(1), 402–416. https://doi.org/10.2307/2407630

Lande, R. (1980a). Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist*, *116*(4), 463–479.

Lande, R. (1980b). Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution*, *34*(2), 292–305. https://doi.org/10.2307/2407393

Lartillot, N., & Poujol, R. (2011). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Molecular Biology and Evolution*, *28*(1), 729–744. https://doi.org/10.1093/molbev/msq244

Lartillot, N., & Delsuc, F. (2012). Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution*, *66*(6), 1773–1787. https://doi.org/10.1111/j.1558-5646.2011.01558.x

Latrille, T., Rodrigue, N., & Lartillot, N. (2023). Genes and sites under adaptation at the phylogenetic scale also exhibit adaptation at the population-genetic scale. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(11), e2214977120. https://doi.org/10.1073/pnas.2214977120

Latrille, T., Lanore, V., & Lartillot, N. (2021). Inferring long-term effective population size with mutation–selection models. *Molecular Biology and Evolution*, *38*(10), 4573–4587. https://doi.org/10.1093/molbev/msab160

Leinonen, T., O'hara, R. B., Cano, J. M., & Merilä, J. (2008). Comparative studies of quantitative trait and neutral marker divergence: A meta-analysis. *Journal of Evolutionary Biology*, *21*(1), 1–17. https://doi.org/10.1111/j.1420-9101.2007.01445.x

Leinonen, T., McCairns, R. J. S., O'Hara, R. B., & Merilä, J. (2013). QST–FST comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, *14*(3), 179–190. https://doi.org/10.1038/nrg3395

Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution*, *45*(5), 1065–1080. https://doi.org/10.1111/j.1558-5646.1991.tb04375.x

Lynch, M., Latta, L., Hicks, J., & Giorgianni, M. (1998). Mutation, selection, and the maintenance of life-history variation in a natural population. *Evolution*, *52*(3), 727–733. https://doi.org/10.1111/j.1558-5646.1998.tb03697.x

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1). Sinauer Sunderland, MA.

Martin, G., Chapuis, E., & Goudet, J. (2008). Multivariate Qst–Fst Comparisons: A Neutrality Test for the Evolution of the G Matrix in Structured Populations. *Genetics*, *180*(4), 2135–2149. https://doi.org/10.1534/genetics.107.080820

McCandlish, D. M., & Stoltzfus, A. (2014). Modeling evolution using the probability of fixation: History and implications. *Quarterly Review of Biology*, *89*(3), 225–252. https://doi.org/10.1086/677571

McDonald, J. H., & Kreitman, M. (1991). Adaptative protein evolution at Adh locus in Drosophila. *Nature*, *351*(6328), 652–654. https://doi.org/10.1038/351652a0

Merilä, J., & Crnokrak, P. (2001). Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, *14*(6), 892–903. https://doi.org/10.1046/j.1420-9101.2001.00348.x

Muse, S. V., & Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, *1*(5), 715–724.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*(1), 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

O'Hara, R. B., & Merilä, J. (2005). Bias and Precision in QST Estimates: Problems and Some Solutions. *Genetics*, *171*(3), 1331–1339. https://doi.org/10.1534/genetics.105.044545

O'Meara, B. C., Ané, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for Different Rates of Continuous Trait Evolution Using Likelihood. *Evolution*, *60*(5), 922–933. https://doi.org/10.1111/j.0014-3820.2006.tb01171.x

Price, P. D., Palmer Droguett, D. H., Taylor, J. A., Kim, D. W., Place, E. S., Rogers, T. F., Mank, J. E., Cooney, C. R., & Wright, A. E. (2022). Detecting signatures of selection on gene expression. *Nature Ecology & Evolution*, 1–11. https://doi.org/10.1038/s41559-022-01761-8

Pujol, B., Wilson, A. J., Ross, R. I. C., & Pannell, J. R. (2008). Are QST–FST comparisons for natural populations meaningful? *Molecular Ecology*, *17*(22), 4782–4785. https://doi.org/10.1111/j.1365-294X.2008.03958.x

Rohlfs, R. V., Harrigan, P., & Nielsen, R. (2014). Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. *Molecular Biology and Evolution*, *31*(1), 201–211. https://doi.org/10.1093/molbev/mst190

Rohlfs, R. V., & Nielsen, R. (2015). Phylogenetic ANOVA: The expression variance and evolution model for quantitative trait evolution. *Systematic Biology*, *64*(5), 695–708. https://doi.org/10.1093/sysbio/syv042

Sella, G., & Barton, N. H. (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics*, *20*(1), 461–493. https://doi.org/10.1146/annurev-genom-083115-022316

Silvestro, D., Kostikova, A., Litsios, G., Pearman, P. B., & Salamin, N. (2015). Measurement errors should always be incorporated in phylogenetic comparative analysis. *Methods in Ecology and Evolution*, *6*(3), 340–346. https://doi.org/10.1111/2041-210X.12337

Silvestro, D., Tejedor, M. F., Serrano-Serrano, M. L., Loiseau, O., Rossier, V., Rolland, J., Zizka, A., Höhna, S., Antonelli, A., & Salamin, N. (2019). Early Arrival and Climatically-Linked Geographic Expansion of New World Monkeys from Tiny African Ancestors. *Systematic Biology*, *68*(1), 78–92. https://doi.org/10.1093/sysbio/syy046

Simons, Y. B., Bullaughey, K., Hudson, R. R., & Sella, G. (2018). A population genetic interpretation of GWAS findings for human quantitative traits. *PLOS Biology*, *16*(3), e2002985. https://doi.org/10.1371/journal.pbio.2002985

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tenaillon, O. (2014). The utility of fisher's geometric model in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 179–201. https://doi.org/10.1146/annurev-ecolsys-120213-091846

Tsuboi, M., van der Bijl, W., Kopperud, B. T., Erritzøe, J., Voje, K. L., Kotrschal, A., Yopak, K. E., Collin, S. P., Iwaniuk, A. N., & Kolm, N. (2018). Breakdown of brain–body allometry and the encephalization of birds and mammals. *Nature Ecology & Evolution*, *2*(9), 1492–1500. https://doi.org/10.1038/s41559-018-0632-1

Tung, J., Zhou, X., Alberts, S. C., Stephens, M., & Gilad, Y. (2015). The genetic architecture of gene expression levels in wild baboons (E. T. Dermitzakis, Ed.). *eLife*, *4*, e04729. https://doi.org/10.7554/eLife.04729

Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology*, *25*(2), 138–193. https://doi.org/10.1016/0040-5809(84)90017-0

Turelli, M. (2017). Commentary: Fisher's infinitesimal model: A story for the ages. *Theoretical Population Biology*, *118*, 46–49. https://doi.org/10.1016/j.tpb.2017.09.003

Walsh, B., & Lynch, M. (2018, June 21). *Evolution and Selection of Quantitative Traits*. Oxford University Press.

Wilder, A. P., Supple, M. A., Subramanian, A., Mudide, A., Swofford, R., Serres-Armero, A., Steiner, C., Koepfli, K.-P., Genereux, D. P., Karlsson, E. K., Lindblad-Toh, K., Marques-Bonet, T., Munoz Fuentes, V., Foley, K., Meyer, W. K., Zoonomia Consortium, Ryder, O. A., & Shapiro, B. (2023). The contribution of historical processes to contemporary extinction risk in placental mammals. *Science*, *380*(6643), eabn5856. https://doi.org/10.1126/science.abn5856

Detecting diversifying selection for a trait from within and between-species genotypes and phenotypes

T. Latrille[1], **M. Bastian**[2], **T. Gaboriau**[1], **N. Salamin**[1]

[1]Department of Computational Biology, Université de Lausanne, Lausanne, Switzerland

[2]Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Université Lyon 1, Villeurbanne, France

thibault.latrille@ens-lyon.org

October 2, 2023

Supplementary materials

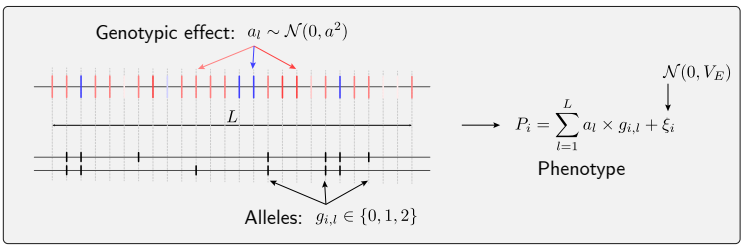
Contents

1 Genetic architecture of the trait

1.1 Genotype-phenotype map

- L is the number of loci encoding the trait.
- $a_l \sim \mathcal{N}(0, a^2)$ is the effect of a mutation on the trait at locus $l \in \{1, \ldots, L\}$.
- N_e is the effective number of individuals.
- $g_{i,l} \in \{0, 1, 2\}$ is the genotypic value at locus l for individual $i \in \{1, \ldots, N_\mathrm{e}\}$.
- $G_i = \sum_{l=1}^{L} a_l \times g_{i,l}$ is the genotypic value for individual i.
- $\xi_i \sim \mathcal{N}(0, V_\mathrm{E})$ is the effect of environment on the trait for individual i.
- $P_i = G_i + \xi_i$ is the phenotype for individual i.

Figure S1: summary of trait's genetic architecture.



within-species, the mean ($\bar{G}$) and variance (V_A) of the genotype are:

$$\bar{G} = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} G_i \quad \text{and} \quad V_\mathrm{A} = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} \left(G_i - \bar{G}\right)^2 \tag{25}$$

The theoretical additive genetic variance (V_A) is a function of the number of loci (L) and the effect of a mutation (a) as:

$$V_\mathrm{A} = 4N_\mathrm{e} \cdot \mu \cdot L \cdot a^2 \tag{26}$$

The mean ($\bar{P}$) and variance (V_P) of the phenotype are:

$$\bar{P} = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} P_i \quad \text{and} \quad V_P = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} \left(P_i - \bar{P}\right)^2 \tag{27}$$

Heritability (h^2) is defined as:

$$h^2 = \frac{V_\mathrm{A}}{V_P} = \frac{V_\mathrm{A}}{V_\mathrm{A} + V_\mathrm{E}} \tag{28}$$

Altogether, effective population size (N_e), the number of loci (L) and the effect of a mutation (a), we can compute the variance of the environment (V_E) that is required to reach a given heritability (h^2) as:

$$V_\mathrm{E} = V_\mathrm{A} \cdot \left(\frac{1}{h^2} - 1\right) = 4N_\mathrm{e} \cdot \mu \cdot L \cdot a^2 \cdot \left(\frac{1}{h^2} - 1\right) \tag{29}$$

2 Bayesian estimate

2.1 Multivariate Brownian process

Here we generalize to K traits evolving along the phylogeny and are correlated between them. Their variation along the phylogeny is modeled as a K-dimensional Brownian process $\boldsymbol{\mathcal{B}}$ $(1 \times K)$ starting at the root and branching along the tree topology. The rate of change of the Brownian process is determined by the positive semi-definite and symmetric covariance matrix between traits $\boldsymbol{\Sigma}$ $(K \times K)$. Along branch j with length d_j, the Brownian process start at the ancestral node $\mathcal{A}(j)$ with value $\boldsymbol{\mathcal{B}}(\mathcal{A}(j))$, and ends at node $\mathcal{R}(j)$ with value $\boldsymbol{\mathcal{B}}(\mathcal{R}(j))$. The independent contrast $\boldsymbol{C}_j$ defined as change in trait along the branch normalized by $\sqrt{d_j}$ is a multivariate Gaussian:

$$\boldsymbol{C}_j = \frac{\boldsymbol{\mathcal{B}}(\mathcal{R}(j)) - \boldsymbol{\mathcal{B}}(\mathcal{A}(j))}{\sqrt{d_j}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}). \tag{30}$$

2.2 Sampling the covariance matrix

From the independent contrast at each branch of the tree ($\boldsymbol{C}_j$), we can define the $K \times K$ scatter matrix, $\boldsymbol{A}$, as:

$$\boldsymbol{A} = \sum_{j=1}^{2n-2} \boldsymbol{C}_j \times [\boldsymbol{C}_j]^\intercal, \tag{31}$$

where $2n-2$ is the number of branches in the tree and n the number of taxa.

The prior on the covariance matrix is an inverse Wishart distribution, with $K+1$ degrees of freedom:

$$\boldsymbol{\Sigma} \sim \text{Wishart}^{-1}(\boldsymbol{I}, K+1). \tag{32}$$

By Bayes theorem, the posterior on $\boldsymbol{\Sigma}$, conditional on a particular realization of $\boldsymbol{\mathcal{B}}$ (and thus of $\boldsymbol{C}$) is an invert Wishart distribution, of parameter $\boldsymbol{I} + \boldsymbol{A}$ and with $2n+1$ degrees of freedom.

$$\boldsymbol{\Sigma} \sim \text{Wishart}^{-1}(\boldsymbol{I} + \boldsymbol{A}, 2n+1) \tag{33}$$

This invert Wishart distribution can be obtained by sampling $2n+1$ independent and identically distributed multivariate normal random variables $\boldsymbol{Z}_k$ defined by

$$\boldsymbol{Z}_k \sim \mathcal{N}\left(\mathbf{0}, [\boldsymbol{I} + \boldsymbol{A}]^{-1}\right). \tag{34}$$

And from these multivariate samples, $\boldsymbol{\Sigma}$ is Gibbs sampled as:

$$\boldsymbol{\Sigma} = \left(\sum_{k=1}^{2n+1} \boldsymbol{Z}_k \times [\boldsymbol{Z}_k]^\intercal\right)^{-1} \tag{35}$$

3 Bayesian and Maximum-likelihood implementation

Implementation is included within the *BayesCode* software, available at https://github.com/ThibaultLatrille/bayescode.

3.1 Data formatting

Running the analysis on your dataset and compute posterior probabilities requires three files:

1. A phylogenetic tree in newick format, with branch lengths in number of substitutions per site (neutral markers).
2. A file containing the mean trait values for each species.
3. A file containing the variation within-species for each trait and the genetic variation within-species (neutral markers).

3.1.1 Phylogenetic tree

The phylogenetic tree must be in newick format, with branch lengths in substitutions per site (neutral markers).

3.1.2 Mean trait for each species

The file containing mean trait values for each species must be in a tab-delimited file with the following format:

TaxonName	Body_mass	Brain_mass
Panthera_tigris	12.26	5.676
Pithecia_pithecia	7.256	3.436
Colobus_angolensis	9.176	4.284
Saimiri_boliviensis	6.845	3.279
⋮	⋮	⋮

The columns are:

- *TaxonName*: the name of the taxon matching the name in the alignment and the tree.
- As many columns as traits, without spaces or special characters in the trait.
- The values can be `NaN` to indicate that the trait is not available for that taxon.

3.1.3 Trait variation for each species

The file containing trait variation for each species must be in a tab-delimited file with the following format:

TaxonName	Nucleotide_diversity	Body_mass_variance	Body_mass_heritability	Brain_mass_variance	Brain_mass_heritability
Pithecia_pithecia	0.0016	0.22871	0.2	0.00737	0.2
Colobus_angolensis	0.0017	0.00393	0.2	0.00416	0.2
Saimiri_boliviensis	0.0013	0.00022	0.2	0.00045	0.2
Pygathrix_nemaeus	0.0016	0.00347	0.2	0.00097	0.2
⋮	⋮	⋮	⋮	⋮	⋮

- *TaxonName*: the name of the taxon matching the name in the alignment and the tree.
- *Nucleotide_diversity*: the nucleotide diversity within-species (neutral markers), cannot be `NaN`.
- As many columns as traits, without spaces or special characters in the trait.
- *TraitName_variance*: the phenotypic variance of the trait within-species, can be `NaN` to indicate that the trait variance is not available for that taxon.
- *TraitName_heritability* (optional): the heritability of the trait within-species, between 0 and 1, cannot be `NaN`.
- The columns with the suffix `_variance` and `_heritability` are repeated for each trait.
- *TraitName_heritability_lower* (optional): the lower bound of the heritability of the trait within-species, between 0 and 1, cannot be `NaN`.
- *TraitName_heritability_upper* (optional): the upper bound of the heritability of the trait within-species, between 0 and 1, cannot be `NaN`.
- If the columns with the suffix `_heritability_lower` and `_heritability_upper` are present, the heritability is randomly drawn from a uniform distribution between the lower and upper bounds.
- If the columns with the suffix `_heritability` is present, it is taken as is.
- If the additive genetic variance (instead of phenotypic variance) is available for a trait, the heritability can be omitted and will automatically be set to 1.0.

3.2 Bayesian estimation

The executable `nodetraits` from *BayesCode* is used to run the Bayesian estimation of the model, and the executable `readnodetraits` is used to read the results.

Assuming that the file `data/body_size/mammals.male.tsv` contains the mean trait values for each species, the file `data/body_size/mammals.male.var_trait.tsv` contains the variation within-species for each trait and the genetic variation within-species (neutral markers), and the file `data/body_size/mammals.male.tree` contains the phylogenetic tree, the following commands are used to run the model and read the results.

3.2.1 Running the model

`nodetraits` is run with the following command:

```
nodetraits  --until 2000
            --tree data/body_size/mammals.male.tree
            --traitsfile data/body_size/mammals.male.tsv
            run_mammals_male
```

3.2.2 Reading the results

Once the model has run, the chain `run_mammals_male` is used to compute the posterior distribution of the ratio of between-species variation over within-species variation with `readnodetraits`:

```
readnodetraits --burnin 1000
               --var_within data/body_size/mammals.male.var_trait.tsv
               --output results_mammals_male.tsv
               run_mammals_male
```

The file `data_empirical/chain_name.ratio.tsv` then contains the posterior mean of the ratio of between-species variation over within-species variation, the 95% and 99% credible interval, and the posterior probability that the ratio is greater than 1.

3.3 Maximum likelihood estimation

To obtain the ratio (without the posterior credible interval and probability) using maximum likelihood computation, the following python script can be used:

```
python3 utils/neutrality_index.py --tree data/body_size/mammals.male.tree
                                  --traitsfile data/body_size/mammals.male.tsv
                                  --var_within data/body_size/mammals.male.var_trait.tsv
                                  --output results_ML_mammals_male.tsv
```