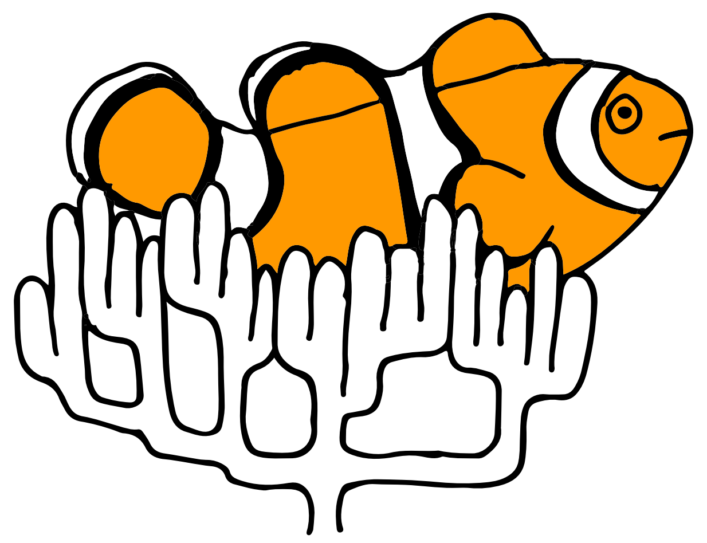
A test of diversifying selection for a trait from within and between species genotypes and phenotypes

Thibault Latrille[1], *Mélodie Bastian*[2], *Théo Gaboriau*[1], *Nicolas Salamin*[1]

[1]Université de Lausanne; [2]Université de Lyon

July 28, 2024

Evolution Joint Congress, Montréal

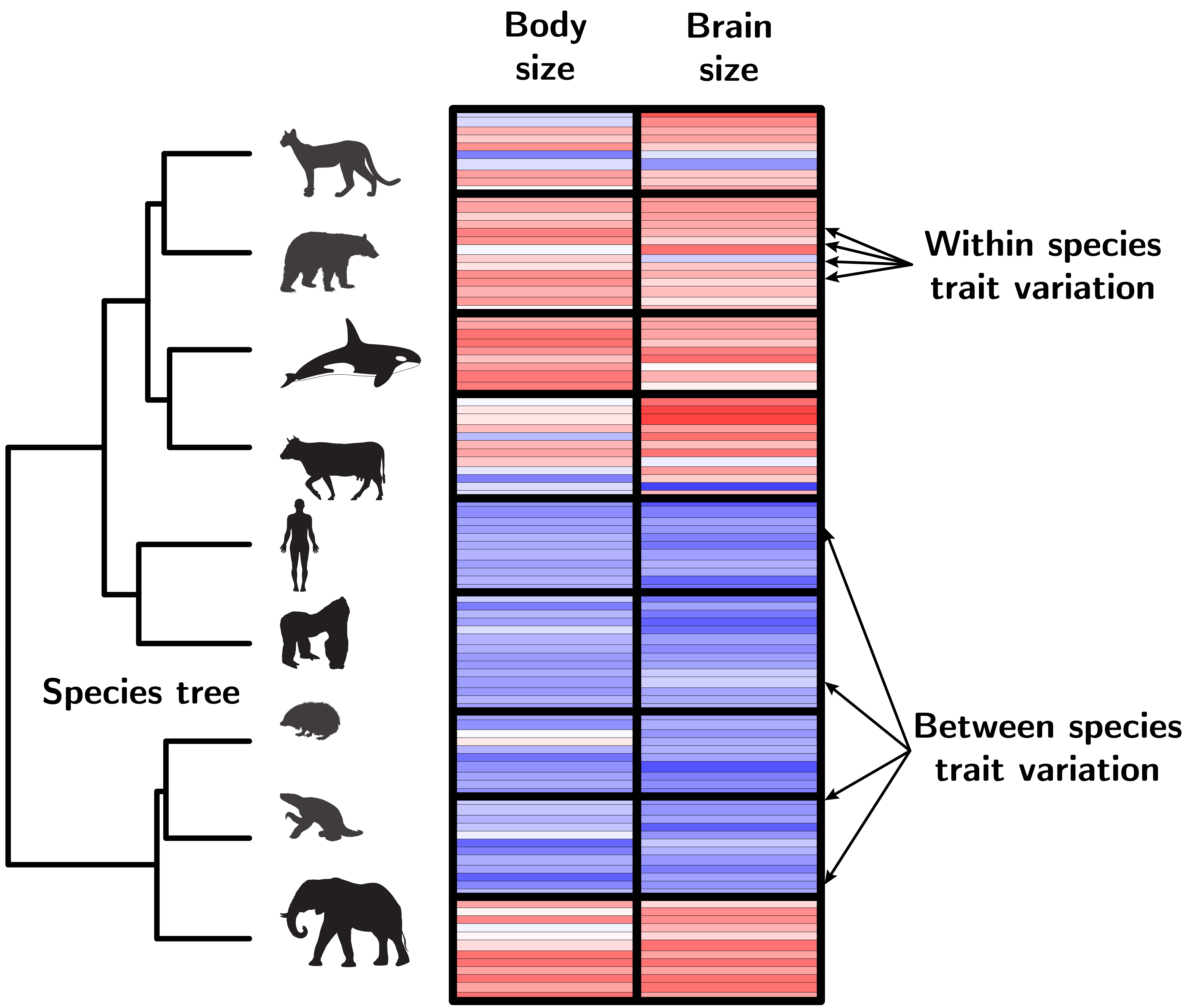


@phylogenetrips

thibault.latrille@unil.ch

- **Is a trait neutrally evolving or under selection?**
- **If not neutral, is the trait changing too fast or too slow?**
- **At which scale, between or within species variations?**
- **How to compute variations, and how to normalize it?**

Trait evolution between and within species

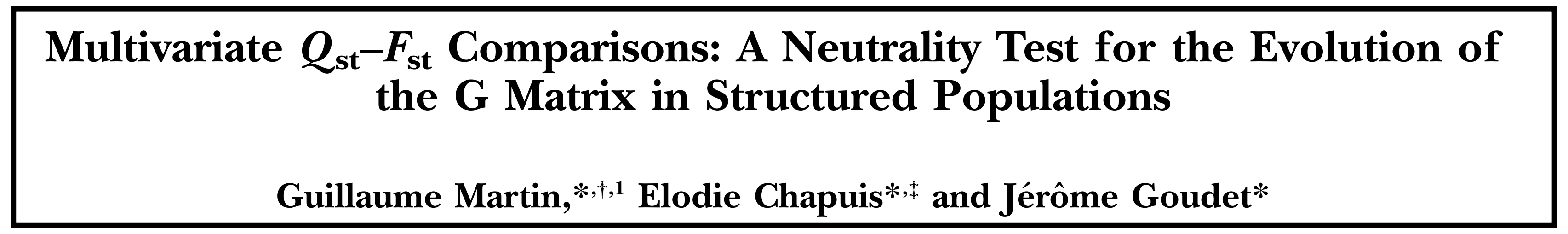


Neutral and adaptive variation in gene expression

Andrew Whitehead[*,†] and Douglas L. Crawford[‡]

Q_{ST}–F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity

Tuomas Leinonen[1], R. J. Scott McCairns[1], Robert B. O'Hara[2] and Juha Merilä[1]



Multivariate Q_{st}–F_{st} Comparisons: A Neutrality Test for the Evolution of the G Matrix in Structured Populations

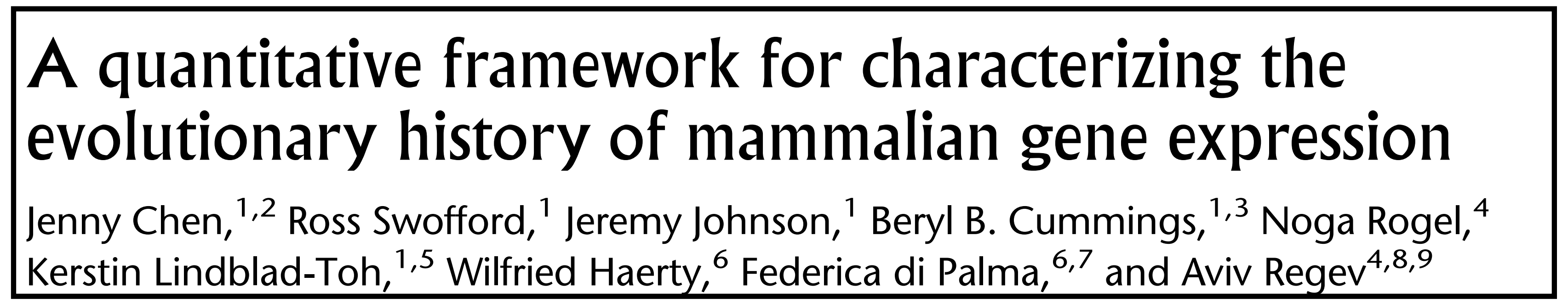
Guillaume Martin,[*,†,1] Elodie Chapuis[*,‡] and Jérôme Goudet[*]

Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution

Marguerite A. Butler[*] and Aaron A. King[†]

Drift and Directional Selection Are the Evolutionary Forces Driving Gene Expression Divergence in Eye and Brain Tissue of *Heliconius* Butterflies

Ana Catalán,[*,†,1] Adriana D. Briscoe,[‡] and Sebastian Höhna[†,§,,1]**

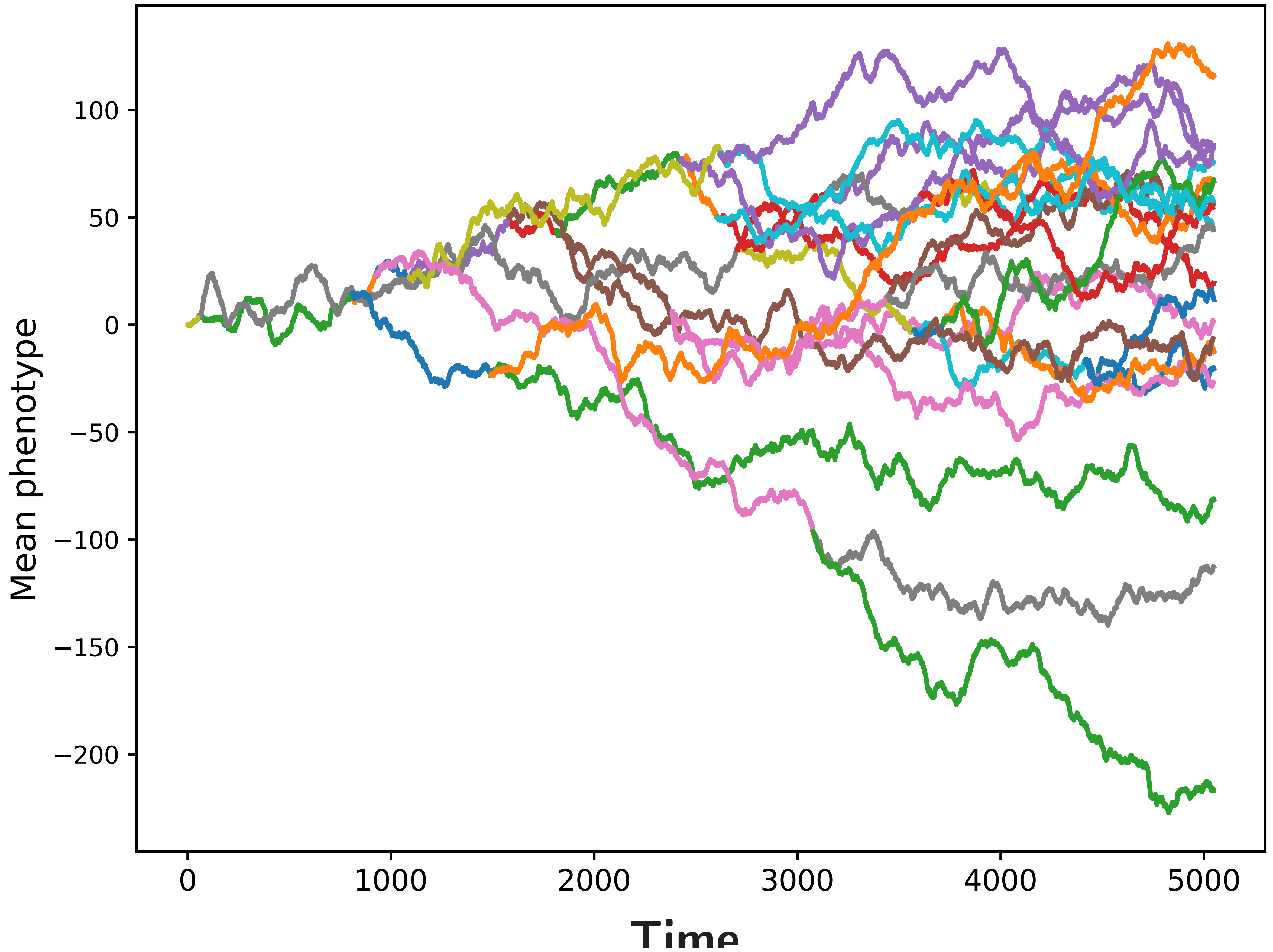


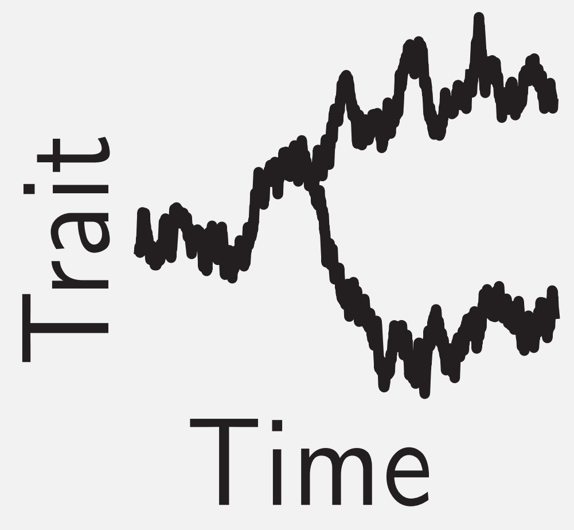
A quantitative framework for characterizing the evolutionary history of mammalian gene expression

Jenny Chen,[1,2] Ross Swofford,[1] Jeremy Johnson,[1] Beryl B. Cummings,[1,3] Noga Rogel,[4] Kerstin Lindblad-Toh,[1,5] Wilfried Haerty,[6] Federica di Palma,[6,7] and Aviv Regev[4,8,9]

How fast should evolve a neutral trait?

Modelling (mean) trait evolution between species

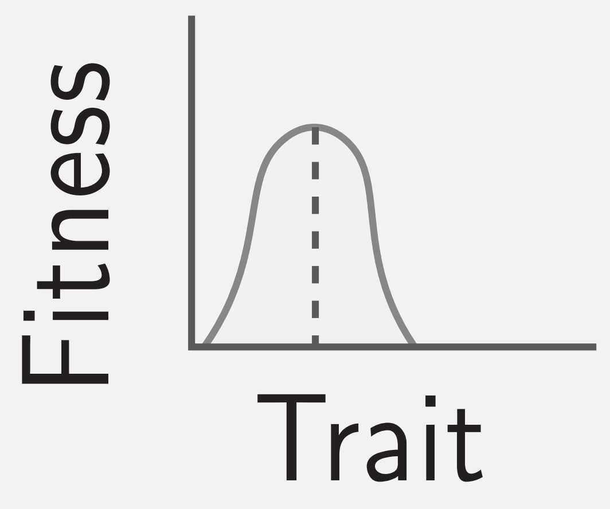




Neutral trait

Random walk of the trait

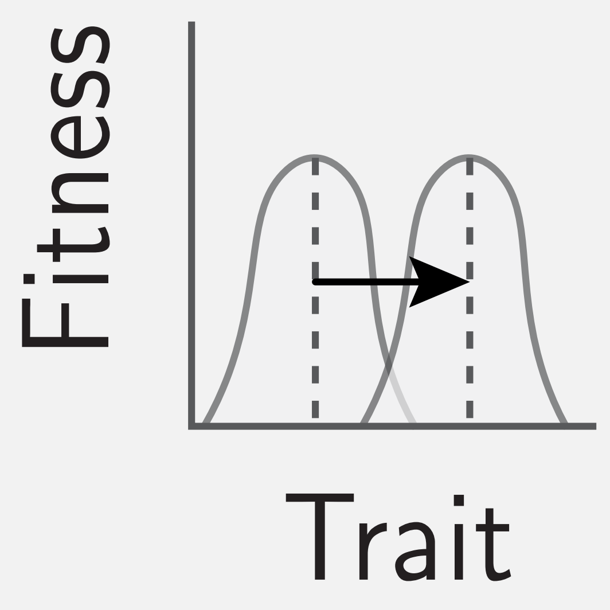
Brownian process



Stabilizing selection

An optimal value for the trait.

Ornstein-Uhlenbeck (OU) process.

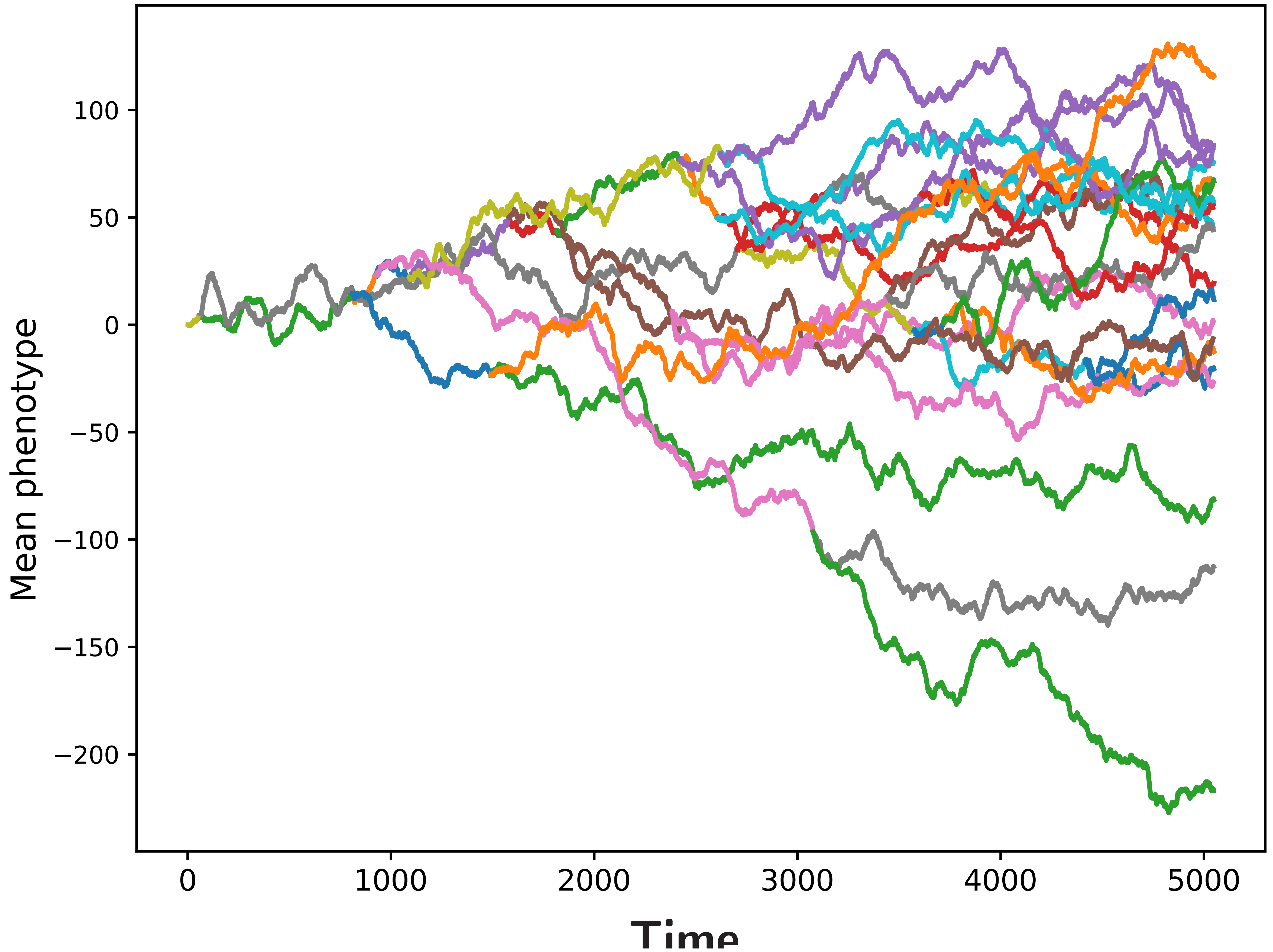


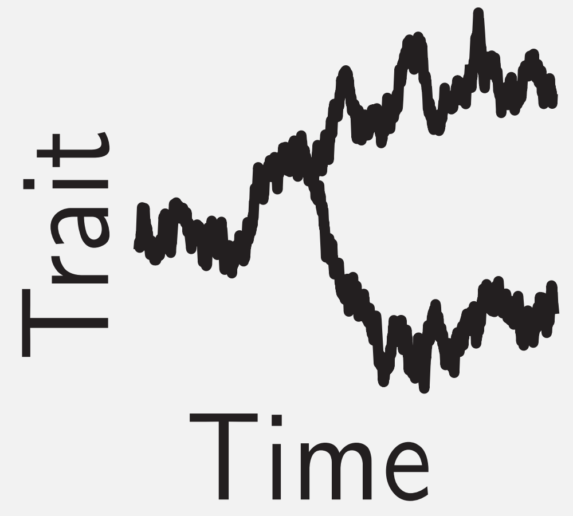
Diversifying selection

Different optimal value for the trait, changing along the phylogeny.

Shifted OU process.

Pitfalls of modelling (mean) trait evolution between species

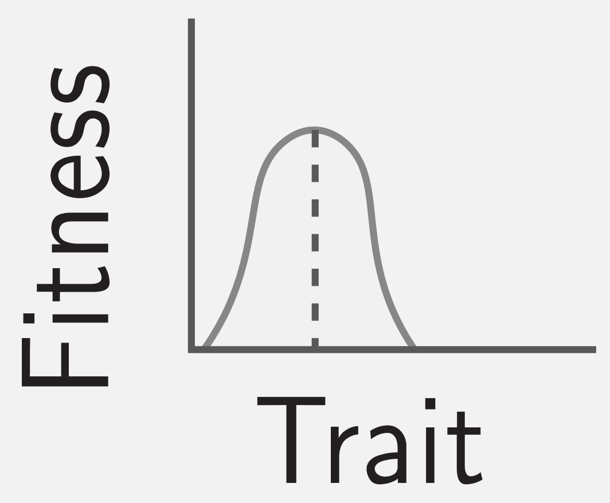




Neutral trait

Random walk of the trait

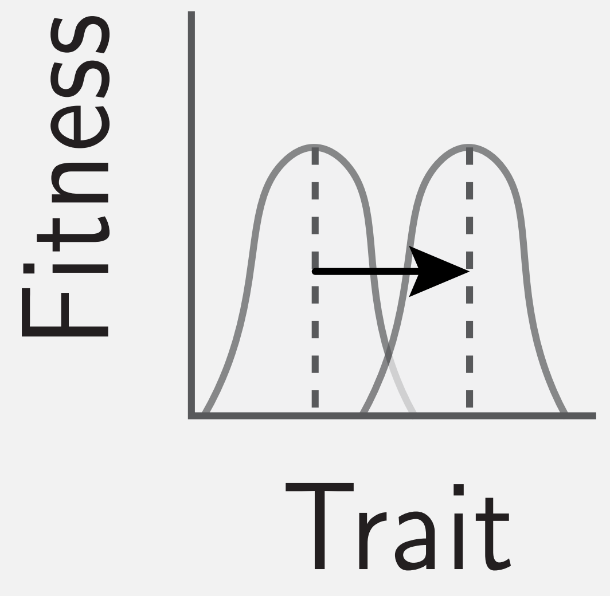
Brownian process



Stabilizing selection

An optimal value for the trait.

Ornstein-Uhlenbeck (OU) process.



Diversifying selection

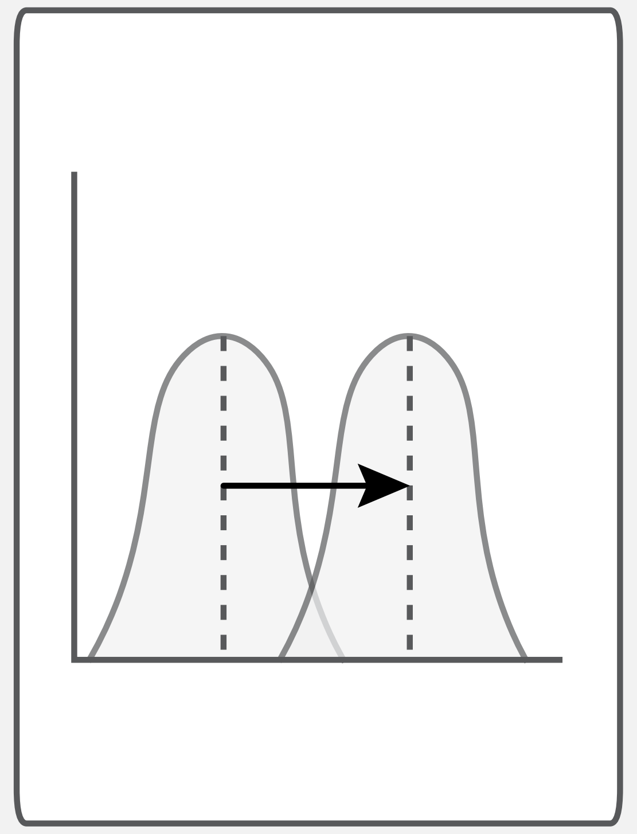
Different optimal value for the trait, changing along the phylogeny.

Shifted OU process.

- **OU process can be favored a neutral trait** [1,2]
  → Selection but it's not.
- **Brownian process can be favored for trait under diversifying selection**[3]
  → Neutral evolution but it's not.

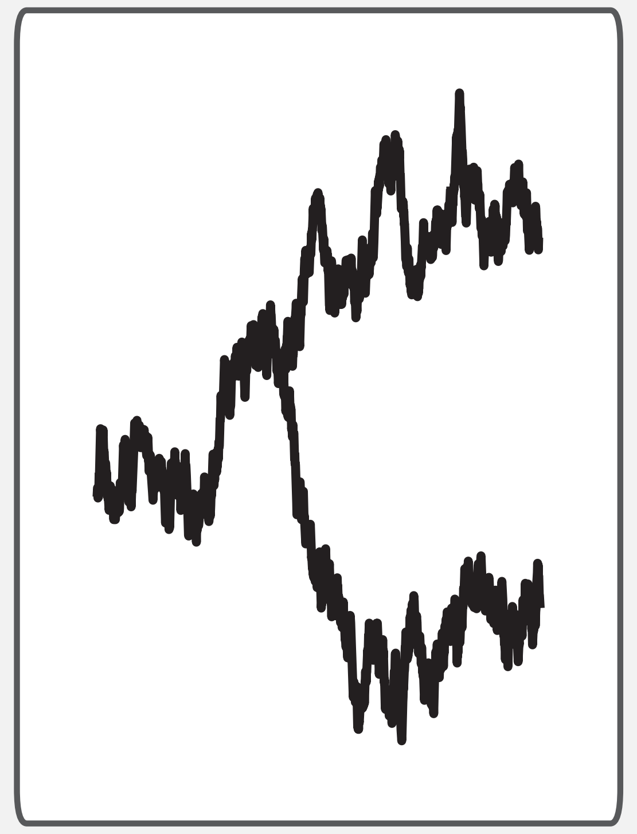
[1]Silvestro *et al* (2015); [2]Copper *et al* (2016); [3]Hansen & Martins (1996)

What are you familiar with?



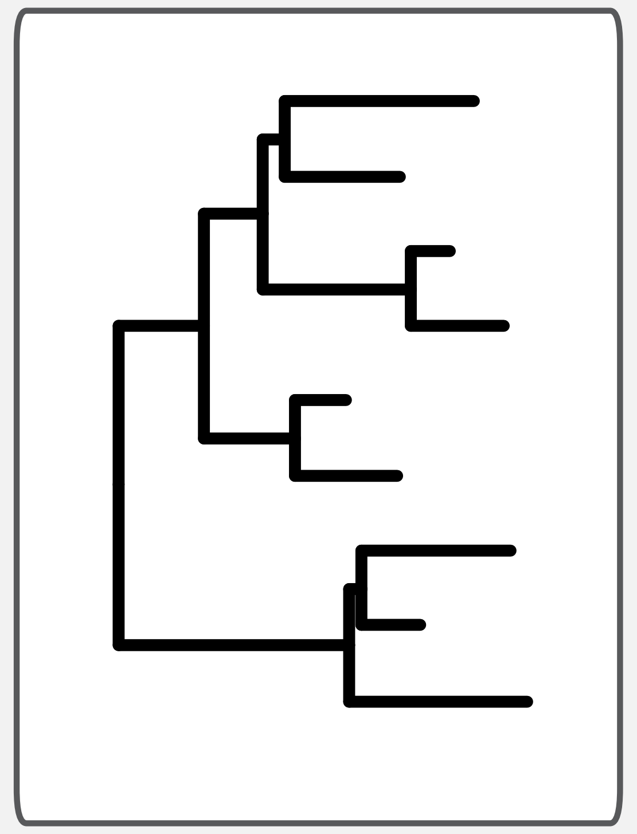
Quantitative-genetics across populations.

→ How to adapt Q_{ST}-F_{ST} methods across species?



Phylogenetic comparative method.

→ What is the expected rate of evolution for a neutral trait?

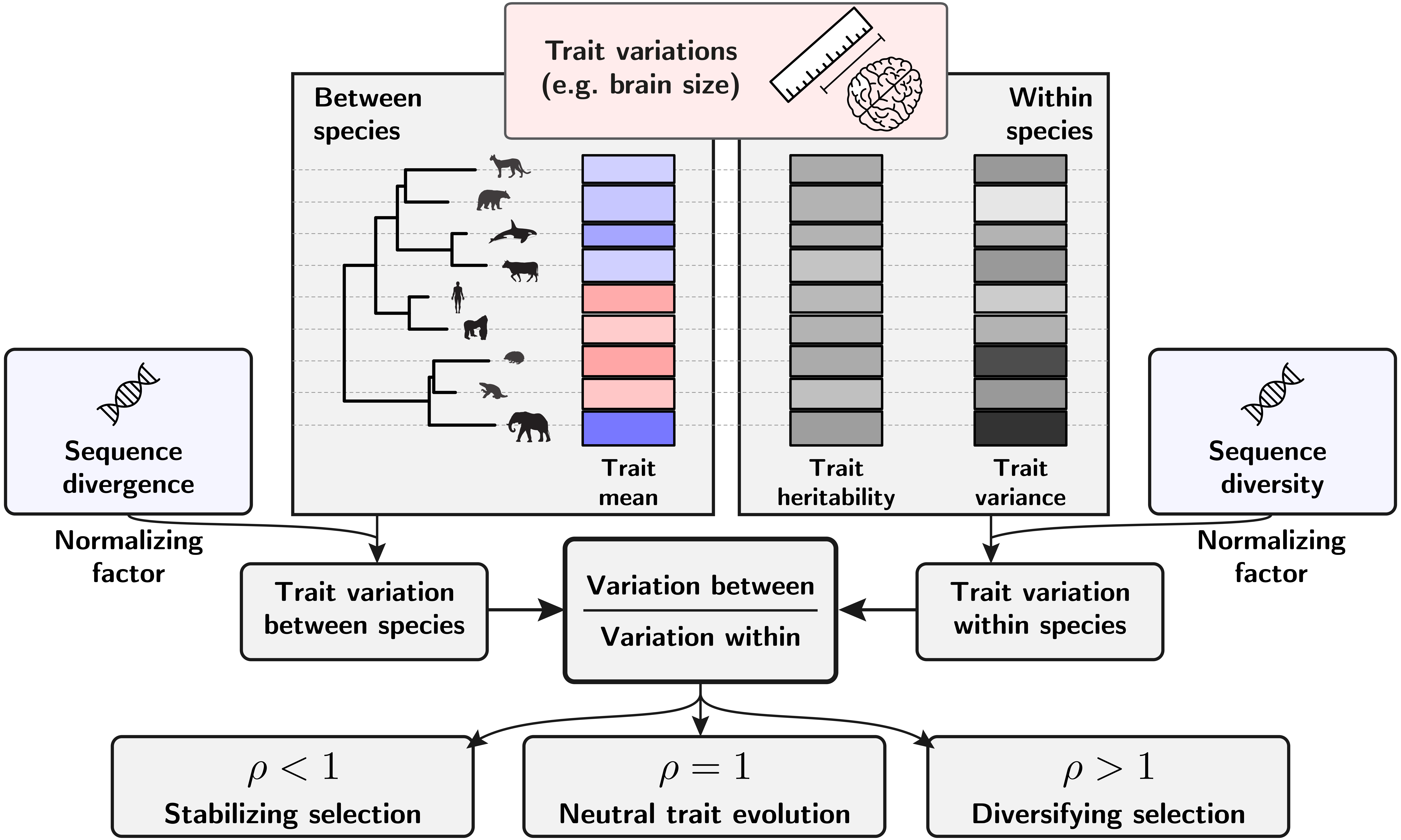


Phylogenetic DNA evolution.

→ How to a derive a d_N/d_S ratio but for a trait instead of protein coding DNA sequences?

Contrast polymorphism & divergence.

→ How to adapt McDonald & Kreitman test ($d_N/d_S > p_N/p_S$) for trait changes along a phylogeny?



Trait variations (e.g. brain size)

Between species

Within species

Trait mean

Trait heritability

Trait variance

Sequence divergence

Normalizing factor

Trait variation between species

$$\frac{\text{Variation between}}{\text{Variation within}}$$

Trait variation within species

Sequence diversity

Normalizing factor

$\rho < 1$
Stabilizing selection

$\rho = 1$
Neutral trait evolution

$\rho > 1$
Diversifying selection

- **Ratio of between over within species trait variations**
- **Normalize using nucleotide variations (i.e. divergence and polymorphism)**

When are between and within species variations equal?

Within species variation:

- V_P: phenotypic variance.
- h^2: heritability for the trait.
- π: mutations per site between two haplotypes.

Between species variation:

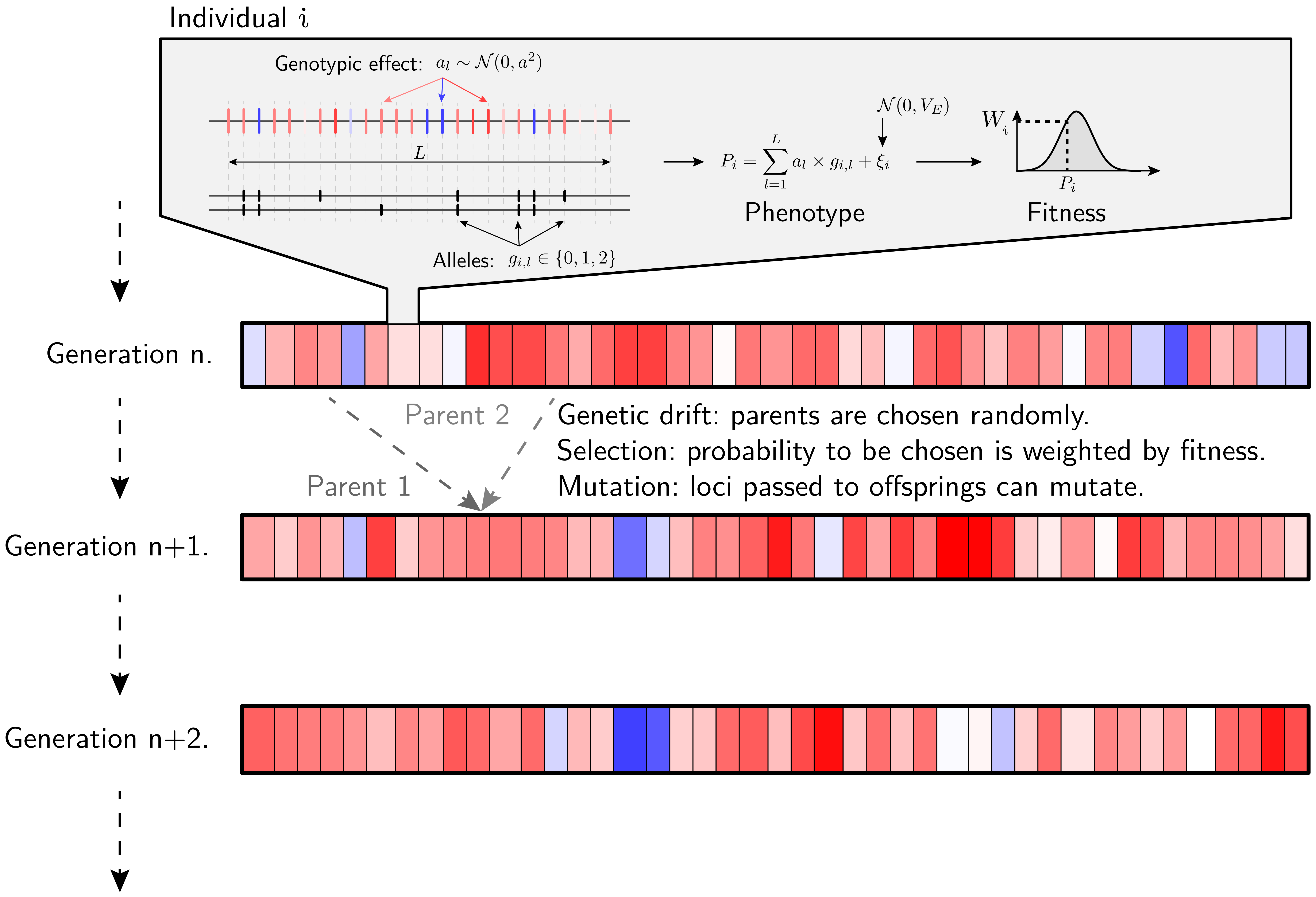
- $\text{cov}\left(\bar{P}_i, \bar{P}_j\right)$: covariance in mean trait value between a pair of species.
- d: shared nucleotide divergence between a pair of species.

$$\begin{cases} \sigma^2_{\text{W}} \overset{\text{def}}{=} \dfrac{V_P \cdot h^2}{\pi} = \dfrac{\cancel{4N_{\text{e}}} \cdot \cancel{\mu} \cdot \sigma^2_{\text{M}}}{\cancel{4N_{\text{e}}} \cdot \cancel{\mu}} = \sigma^2_{\text{M}}, \\ \sigma^2_{\text{B}} \overset{\text{def}}{=} \dfrac{\text{cov}\left(\bar{P}_i, \bar{P}_j\right)}{4d} = \dfrac{\cancel{4t} \cdot \cancel{\mu} \cdot \sigma^2_{\text{M}}}{\cancel{4t} \cdot \cancel{q}} = \sigma^2_{\text{M}}. \end{cases} \implies \boldsymbol{\sigma^2_{\text{B}} / \sigma^2_{\text{W}} = 1}.$$

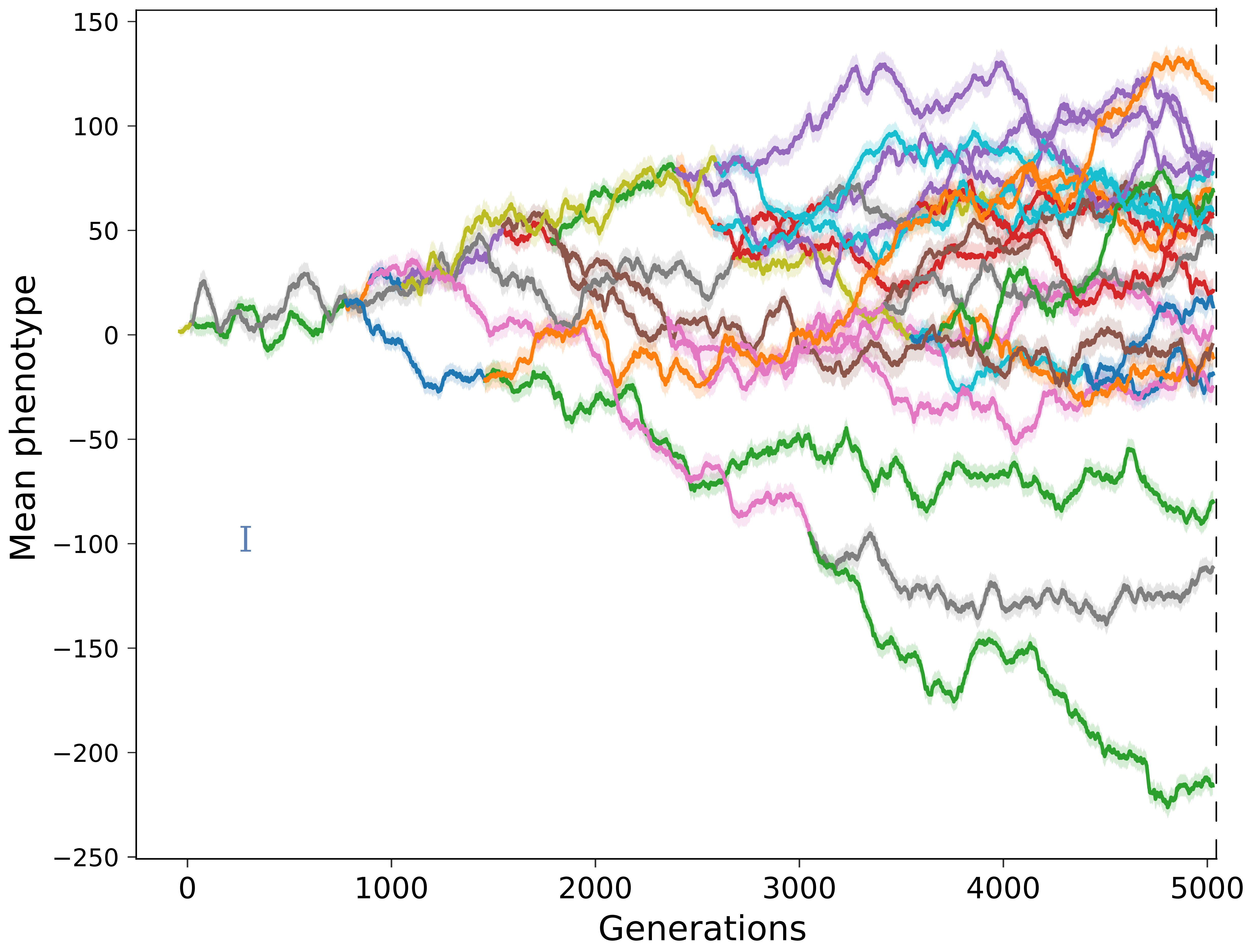
- μ: mutation rate per generation.
- q: substitution rate per generation.
- N_{e}: effective population size.
- σ^2_{M}: effect on the trait per mutation.

Lynch (1998); Hansen & Martins (1996); Kimura (1968); Tajima (1989)

Can we test our estimate against simulated data?



Genetic drift: parents are chosen randomly.
Selection: probability to be chosen is weighted by fitness.
Mutation: loci passed to offsprings can mutate.



- **Simulator across the phylogeny under different scenarios.**

Can we test our estimate against simulated data?

🟦 **Stabilizing selection**

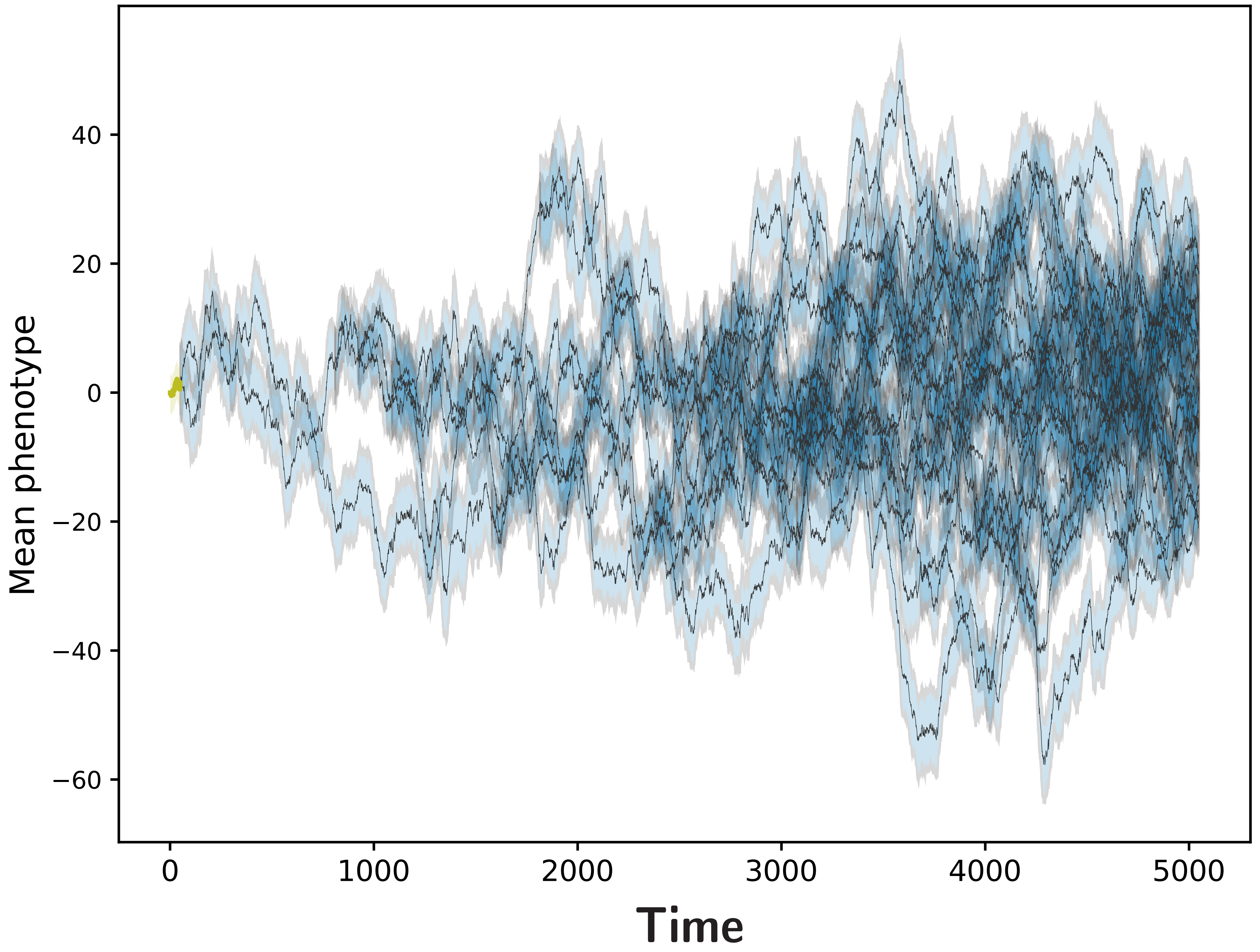
An optimal value for the trait.

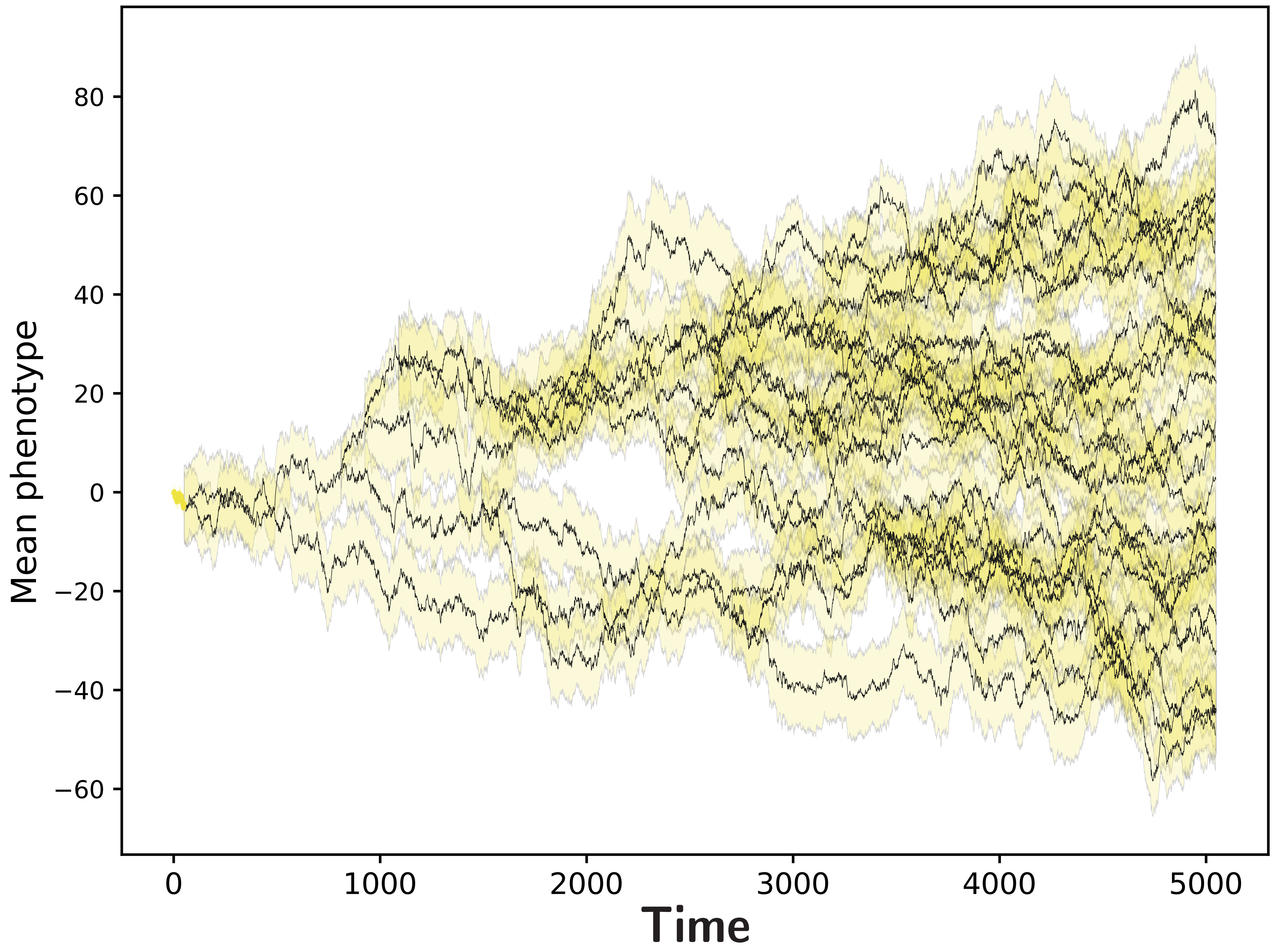
🟨 **Neutral trait**

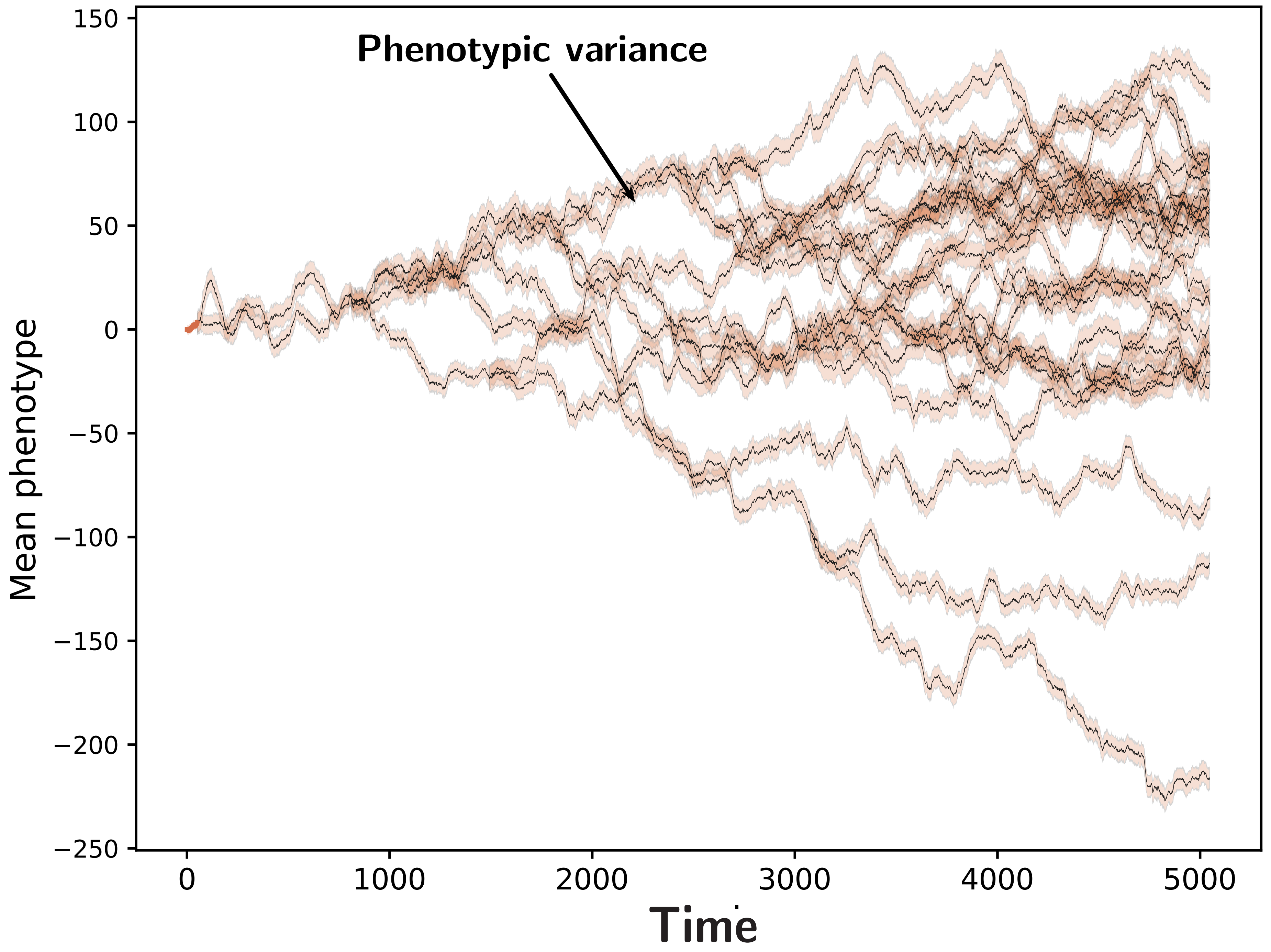
No fitness function.

🟧 **Diversifying selection**

An optimal value for the trait, changing randomly.







- **Simulations across the phylogeny under different scenarios.**

Test of neutrality against simulations

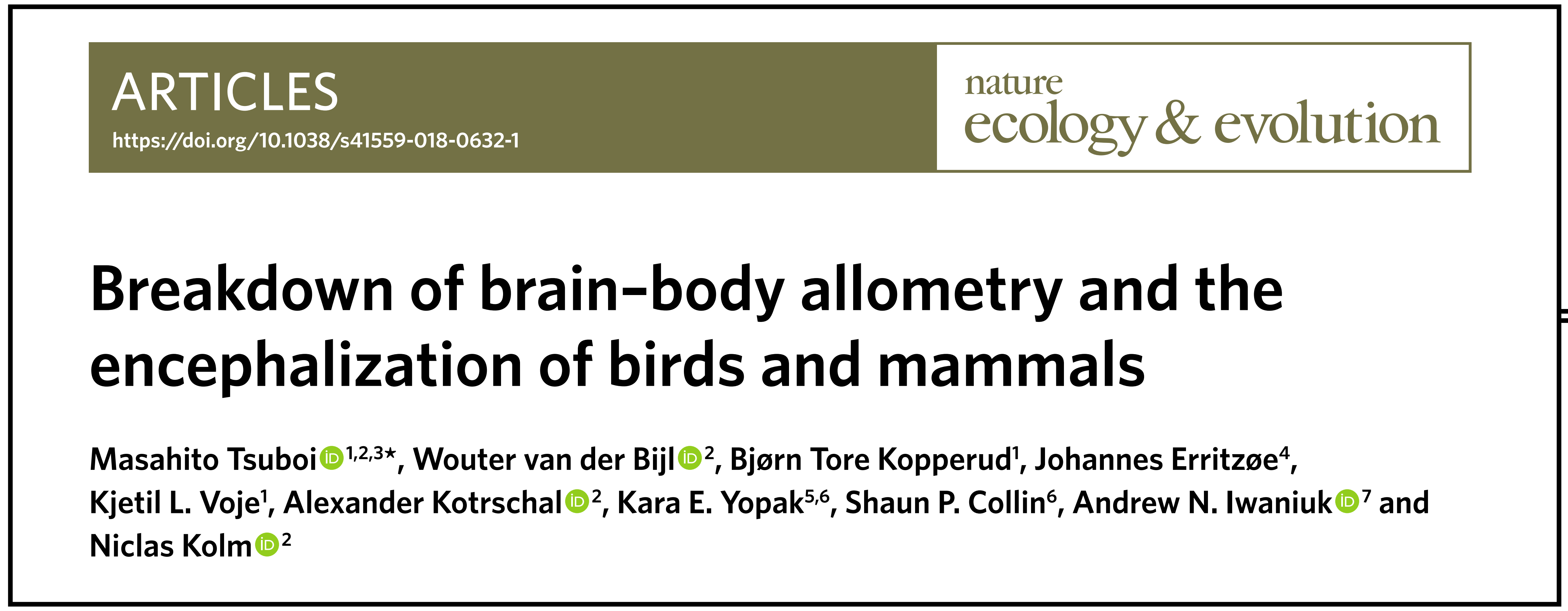
Stabilizing selection
An optimal value for the trait.

Neutral trait
No fitness function.

Diversifying selection
An optimal value for the trait, changing randomly.

- **Can test for diversifying selection acting on a trait.**
- **False positive for detection of stabilizing selection.**

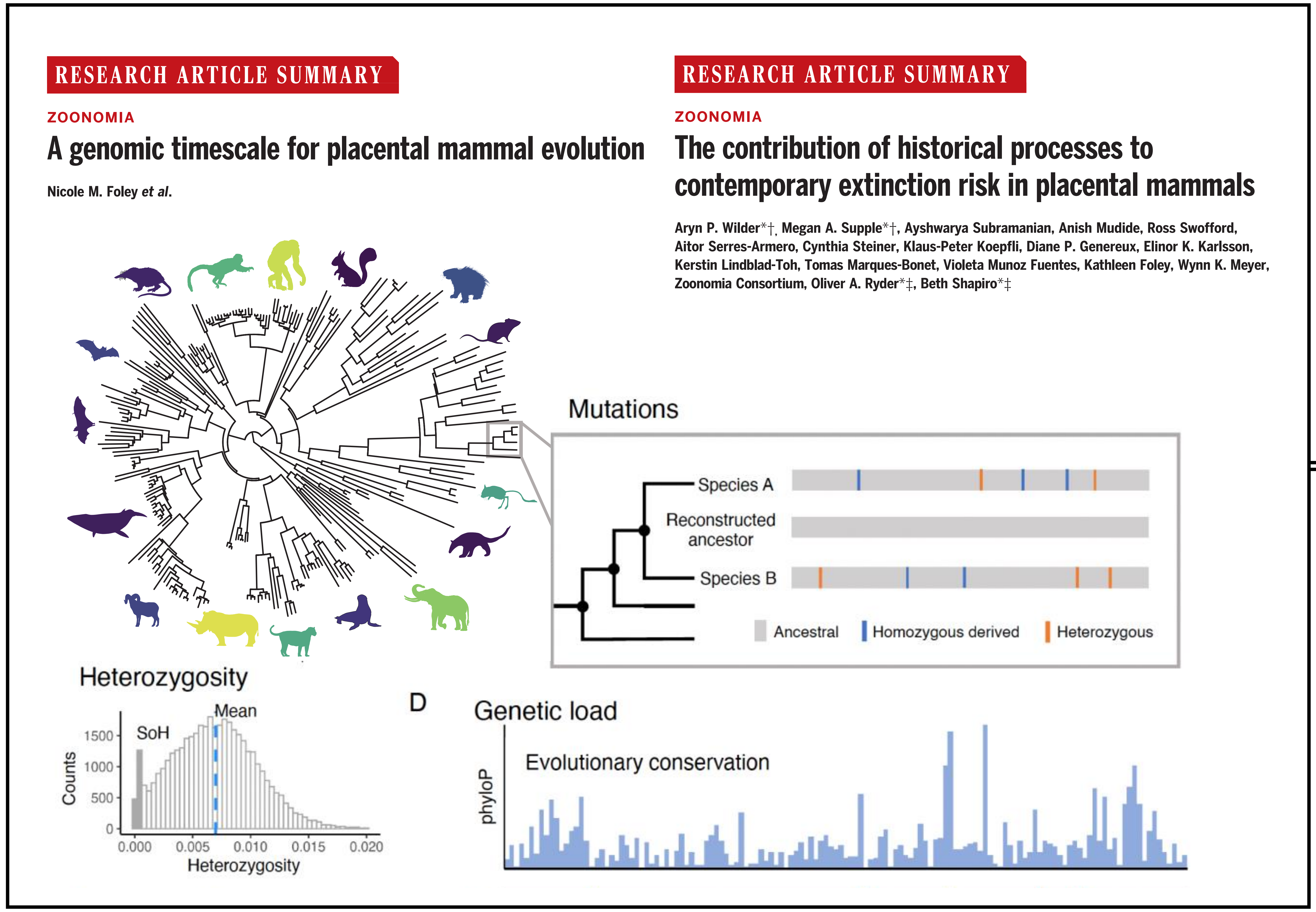
Application to brain and body size in mammals



ARTICLES
https://doi.org/10.1038/s41559-018-0632-1
nature ecology & evolution

Breakdown of brain–body allometry and the encephalization of birds and mammals

Masahito Tsuboi[1,2,3]*, Wouter van der Bijl[2], Bjørn Tore Kopperud[1], Johannes Erritzøe[4], Kjetil L. Voje[1], Alexander Kotrschal[2], Kara E. Yopak[5,6], Shaun P. Collin[6], Andrew N. Iwaniuk[7] and Niclas Kolm[2]



RESEARCH ARTICLE SUMMARY

ZOONOMIA

A genomic timescale for placental mammal evolution

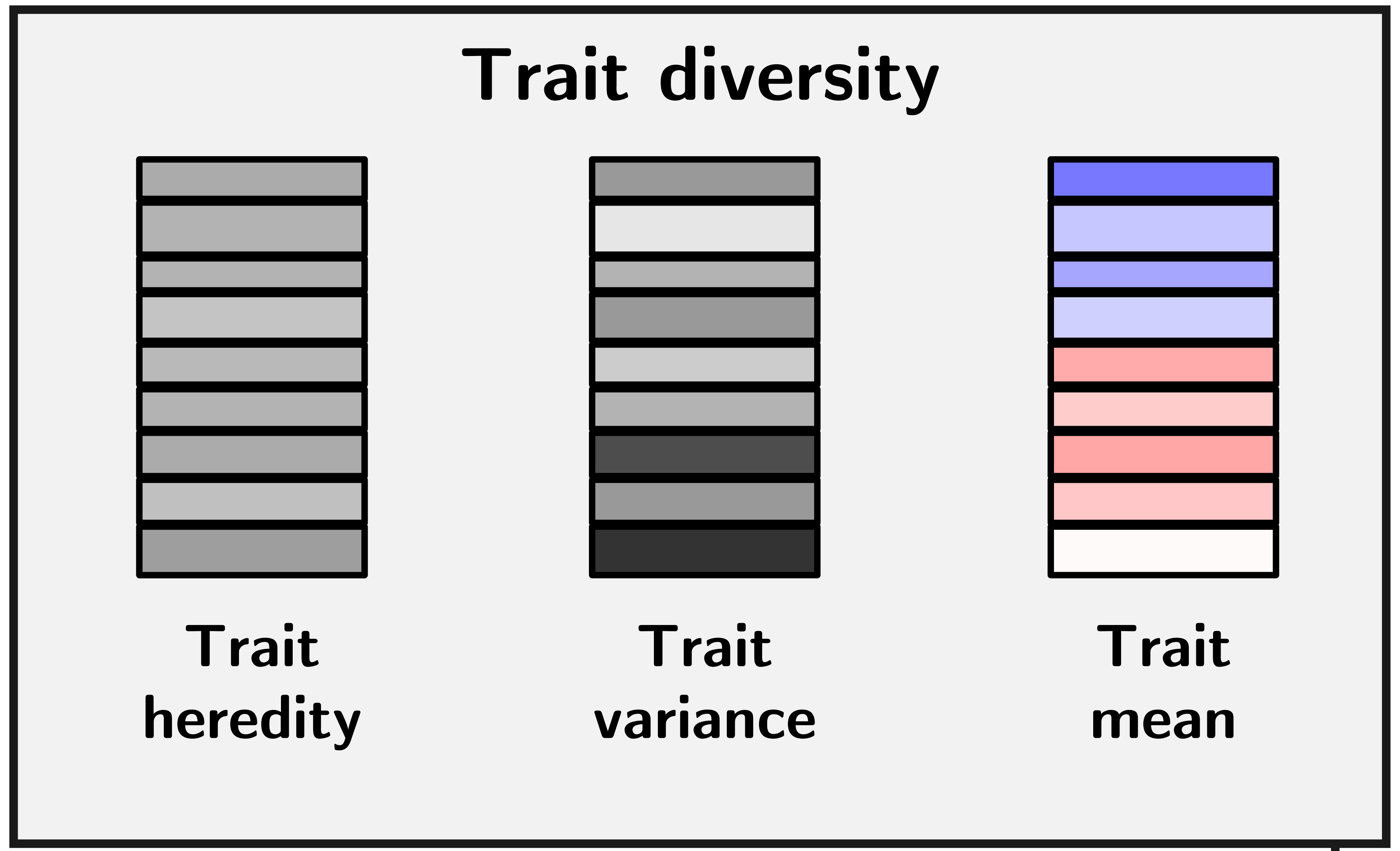
Nicole M. Foley *et al.*

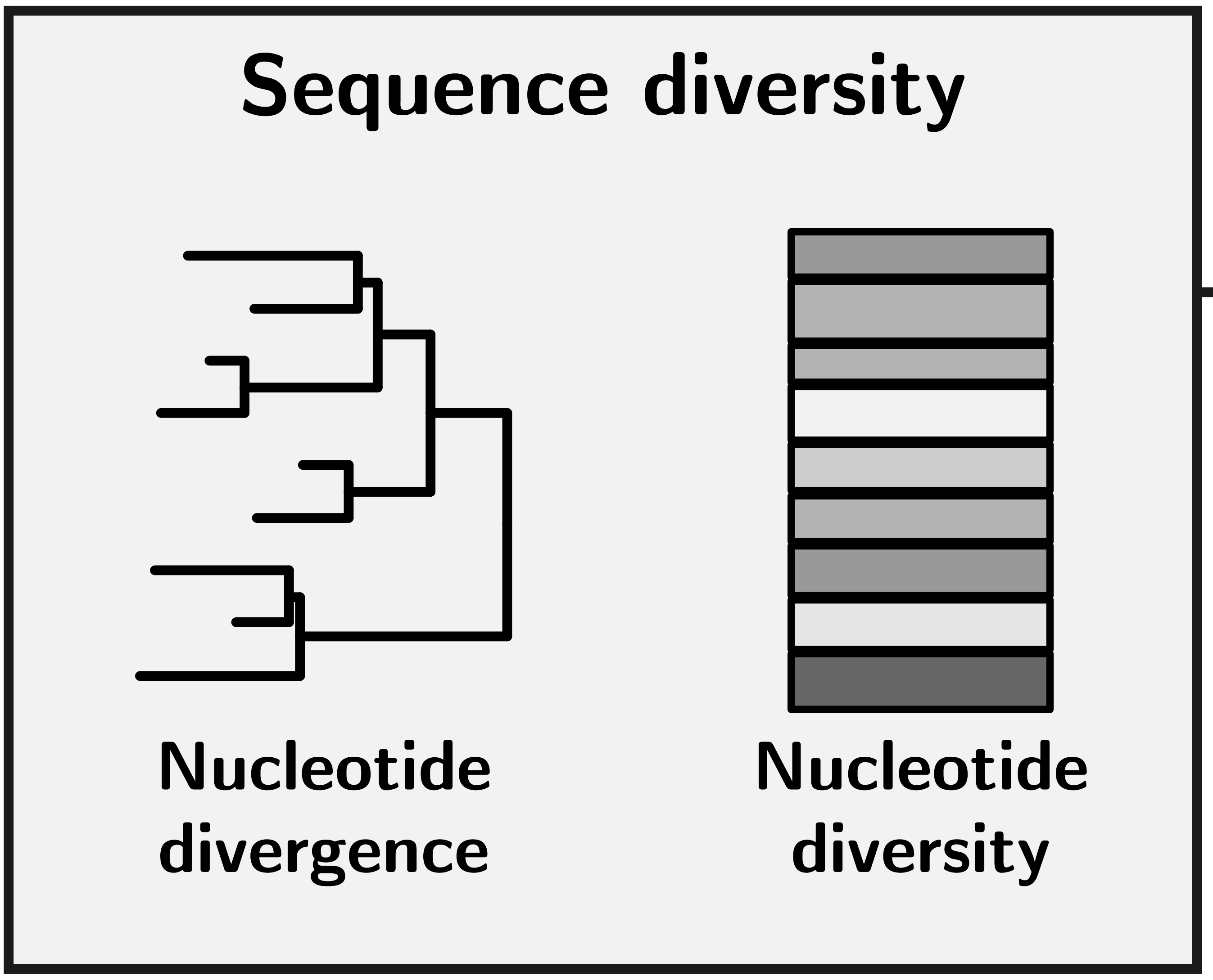
RESEARCH ARTICLE SUMMARY

ZOONOMIA

The contribution of historical processes to contemporary extinction risk in placental mammals

Aryn P. Wilder*†, Megan A. Supple*†, Ayshwarya Subramanian, Anish Mudide, Ross Swofford, Aitor Serres-Armero, Cynthia Steiner, Klaus-Peter Koepfli, Diane P. Genereux, Elinor K. Karlsson, Kerstin Lindblad-Toh, Tomas Marques-Bonet, Violeta Munoz Fuentes, Kathleen Foley, Wynn K. Meyer, Zoonomia Consortium, Oliver A. Ryder*‡, Beth Shapiro*‡





ρ	Body size	Brain size
Males	1.24	4.53**
Females	0.94	6.00**

* *significant p-value with risk α=0.05

Take home messages

- Use the ratio of between over within species trait variations.
- Normalized using nucleotide variations (i.e. divergence and polymorphism).
- Not good to detect of stabilizing selection (false positives).
- Good to detect diversifying selection.

JOURNAL OF
Evolutionary Biology

eseb
European Society for Evolutionary Biology

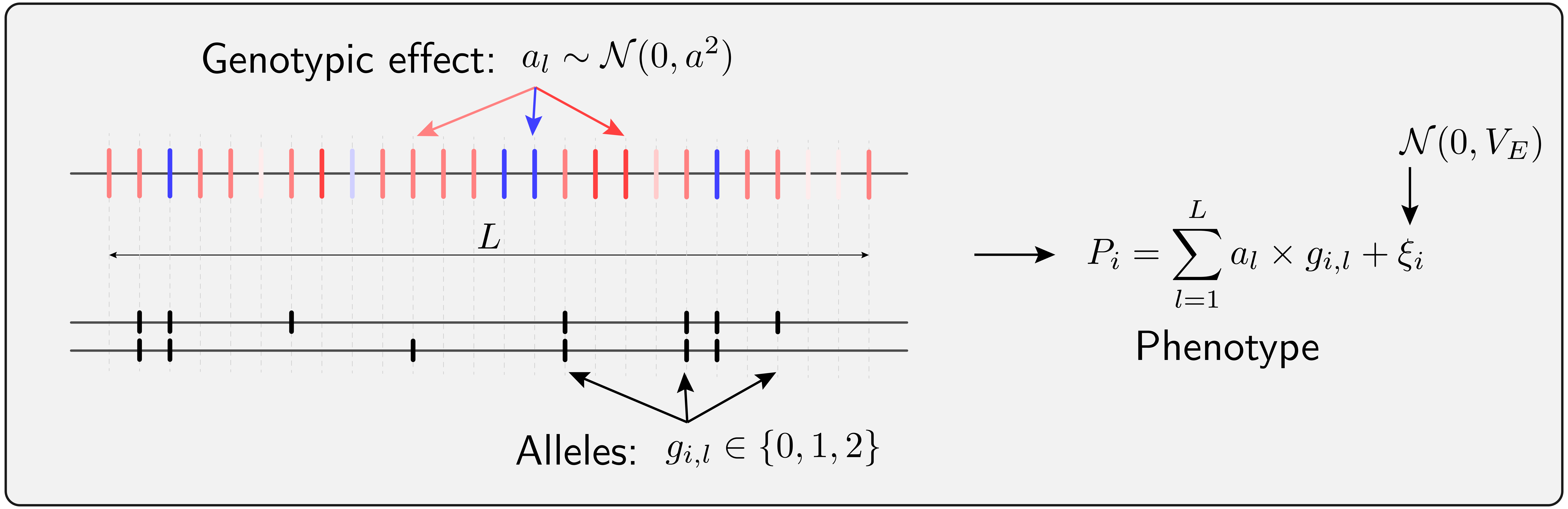
Mélodie Bastian, Théo Gaboriau, Nicolas Salamin

UNIL | Université de Lausanne

Département de
Biologie Computationnelle

@phylogenetrips
thibault.latrille@unil.ch

Genetic architecture of the trait and individual phenotype



- 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 $l \in \{1, \dots, L\}$.
- $g_{i,l} \in \{0, 1, 2\}$ is the genotype at locus l for individual $i \in \{1, \dots, N_{\text{e}}\}$.
- $G_i = \sum_{l=1}^{L} a_l \times g_{i,l}$ is the influence of genotype on the trait for individual i.

Lande (1978); Lynch (1998)

Normalizing trait variation using polymorphism

For any neutral genomic region of interest, the genetic diversity (π), the fraction of the region that are different between two randomly sampled haplotypes, is also a balance between mutations and drift[1]:

$$\pi = 4N_{\mathrm{e}} \cdot \mu. \tag{3}$$

At the population level, we have:

$$\sigma^2_{\mathrm{W}} \overset{\mathrm{def}}{=} \frac{V_{\mathrm{A}}}{\pi} = \frac{4N_{\mathrm{e}} \cdot \mu \cdot L \cdot a^2}{4N_{\mathrm{e}} \cdot \mu} = L \cdot a^2. \tag{4}$$

V_{A} is also the phenotypic variance (V_P) multiplied by heritability (h^2), giving:

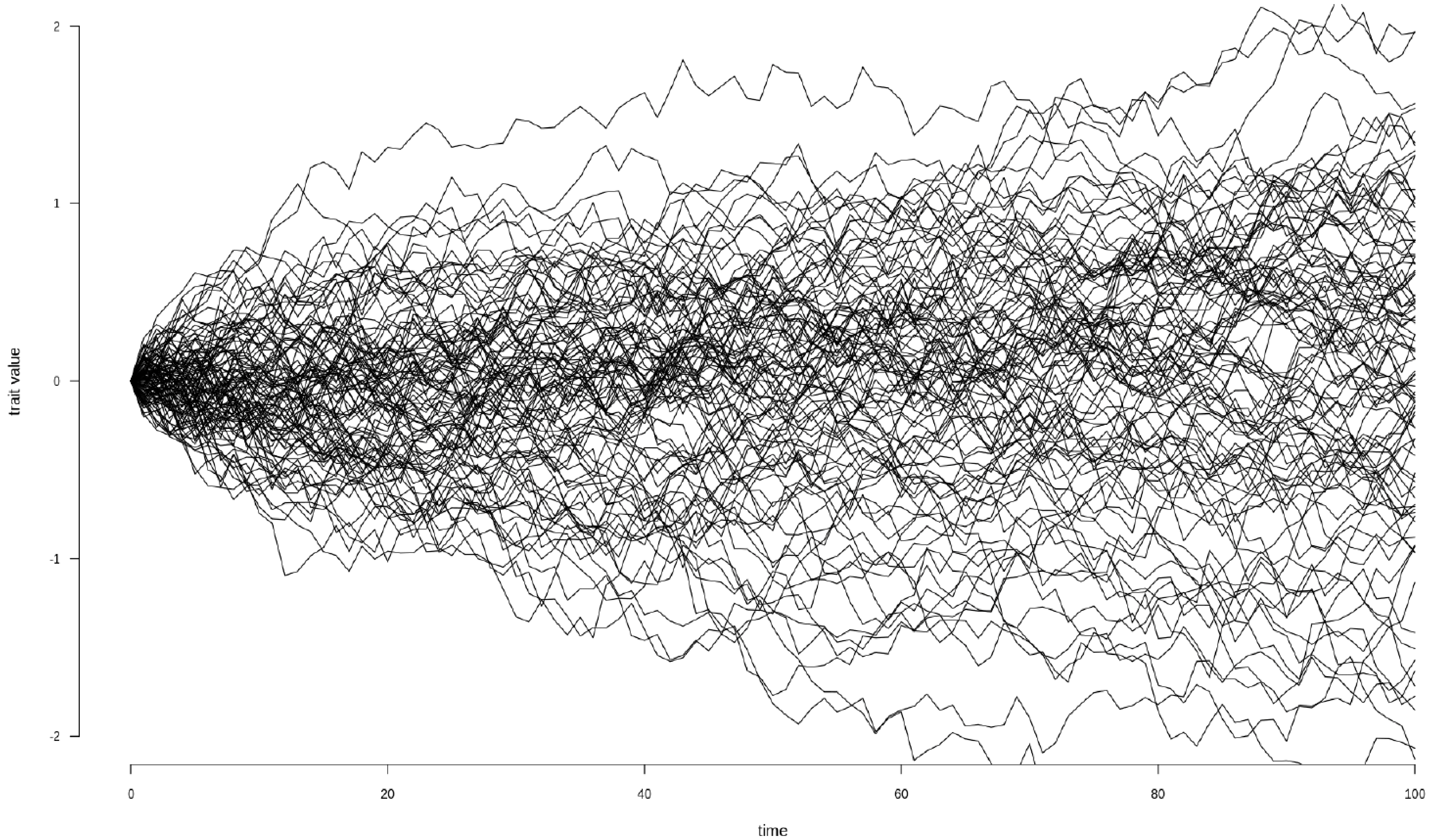
$$\sigma^2_{\mathrm{W}} = \frac{V_P \cdot h^2}{\pi} = L \cdot a^2. \tag{5}$$

[1] Tajima (1989)

Evolution of mean trait value as a function of trait architecture

$\bar{P}$ is the trait mean value. If the trait is neutral and encoded by many loci, $\bar{P}$ evolves as Brownian process. $\text{cov}\left(\bar{P}_i, \bar{P}_j\right)$ is the covariance in mean trait value between a pair of species[1]:

$$\begin{aligned} \text{cov}\left(\bar{P}_i, \bar{P}_j\right) &= \frac{V_{\text{A}}}{N_{\text{e}}} \cdot t = \frac{4N_{\text{e}} \cdot \mu \cdot L \cdot a^2}{N_{\text{e}}} \cdot t, && (6) \\ &= 4t \cdot \mu \cdot L \cdot a^2. && (7) \end{aligned}$$



---

[1] Hansen & Martins (1996)

Normalizing mean trait value evolution using divergence

For a neutral genomic region of interest, some mutations will eventually reach fixation in the population due to genetic drift ($\mathbb{P}_{\text{fix}} = 1/2N_{\text{e}}$)[1], resulting in a substitution. The substitution rate per generation (q) is[2,3]:

$$q = 2N_{\text{e}} \cdot \mu \cdot \mathbb{P}_{\text{fix}} = 2N_{\text{e}} \cdot \mu \cdot \frac{1}{2N_{\text{e}}} = \mu. \tag{8}$$

For t generations, the genetic distance (d) measured as the number of substitution per site is:

$$d = t \cdot q = t \cdot \mu. \tag{9}$$

We thus have for a pair of species:

$$\sigma^2_{\text{B}} \overset{\text{def}}{=} \frac{\text{cov}\left(\bar{P}_i, \bar{P}_j\right)}{4d} = \frac{4t \cdot \mu \cdot L \cdot a^2}{4t \cdot \mu} = L \cdot a^2. \tag{10}$$

[1] Kimura (1962); [2] Kimura (1968); [3] McCandlish & Stoltzfus (2014)[

Estimation at the phylogenetic scale

Input:

- n is the number of taxa.
- $\boldsymbol{D}$ $(n \times n)$ is the symmetric distance matrix computed from the branch lengths and the topology of the phylogenetic tree.
- $\bar{\boldsymbol{P}}$ $(n \times 1)$ is the vector of mean trait values at the tips of the tree.
- $\mathbf{1}$ $(n \times 1)$ is a vector of ones.

Output:

- ϕ is the estimate mean trait value at the root.
- σ_{B}^2 is the estimated trait variance measured at the phylogenetic scale.

At the phylogenetic scale, the maximum likelihood estimates of ϕ and σ_{B}^2 are[1]:

$$\begin{cases} \phi = \left(\mathbf{1}^{\intercal} \times \boldsymbol{D}^{-1} \times \mathbf{1}\right)^{-1} \cdot \left(\mathbf{1}^{\intercal} \times \boldsymbol{D}^{-1} \times \bar{\boldsymbol{P}}\right), \\ \sigma_{\mathrm{B}}^2 = \dfrac{1}{4} \dfrac{\left(\bar{\boldsymbol{P}} - \phi \cdot \mathbf{1}\right)^{\intercal} \times \boldsymbol{D}^{-1} \times \left(\bar{\boldsymbol{P}} - \phi \cdot \mathbf{1}\right)}{n-1}. \end{cases} \tag{11}$$

[1] O'Meara *et al.* (2006)