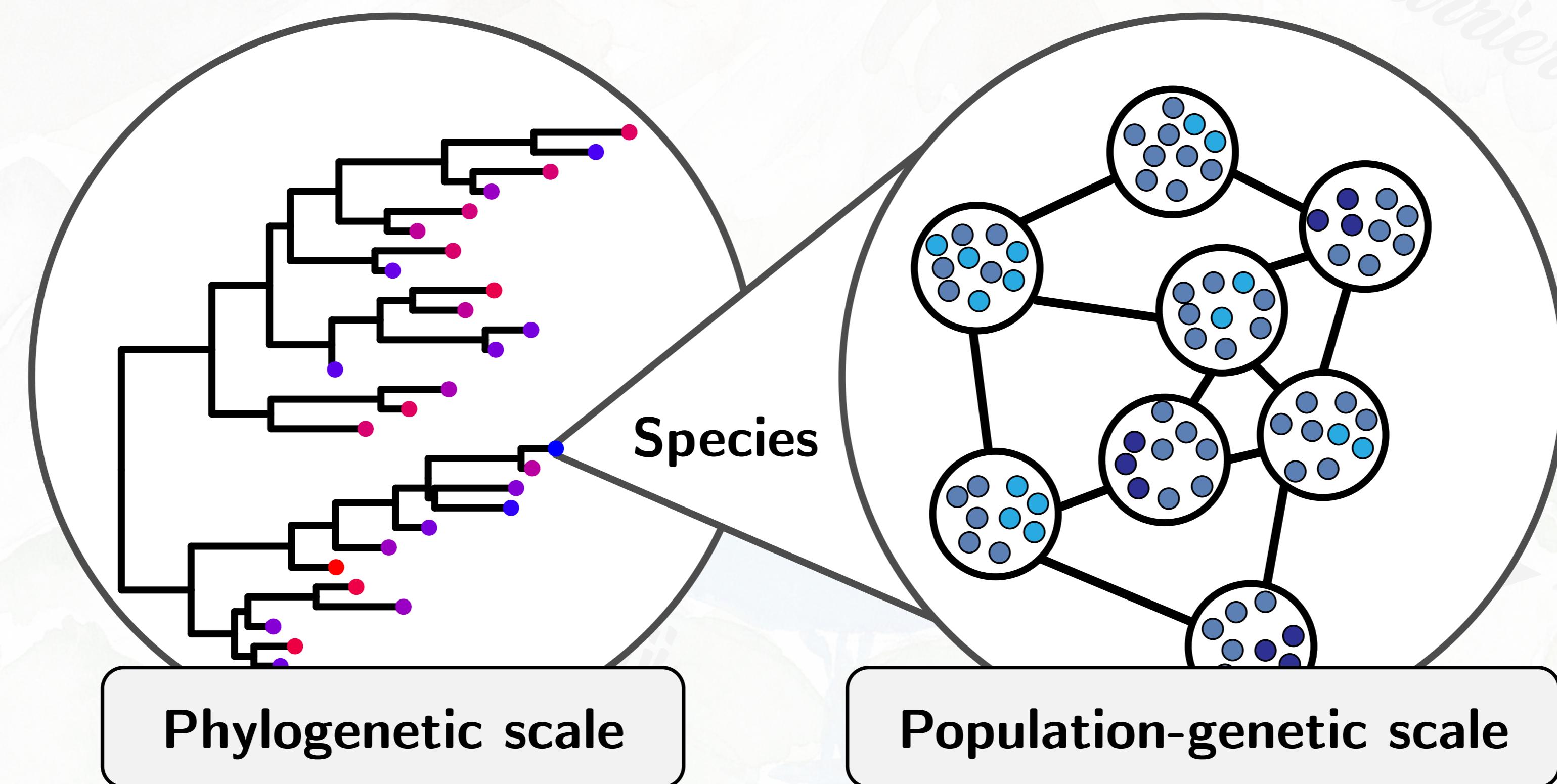
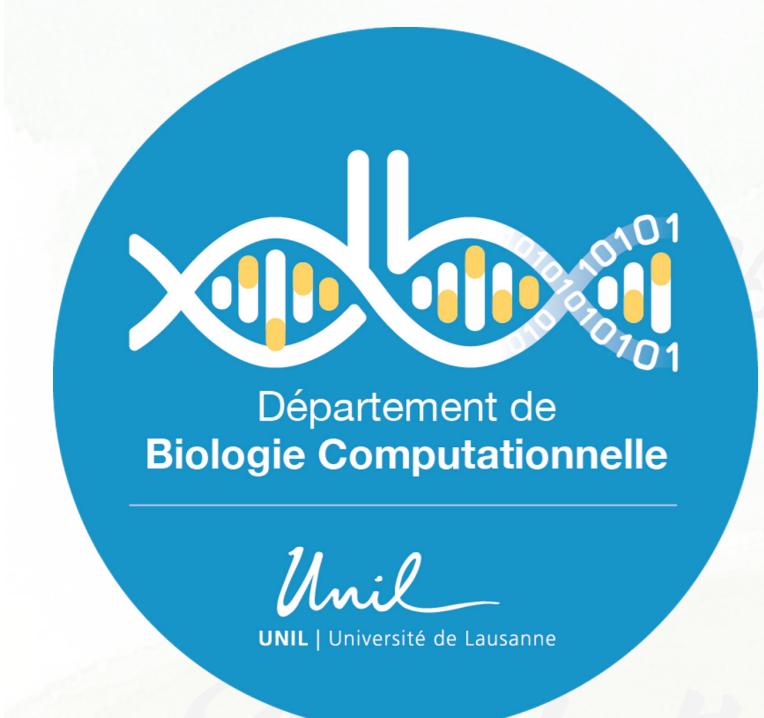


Predicting Selection on Traits and Sequences: Contrast Across Evolutionary Scales



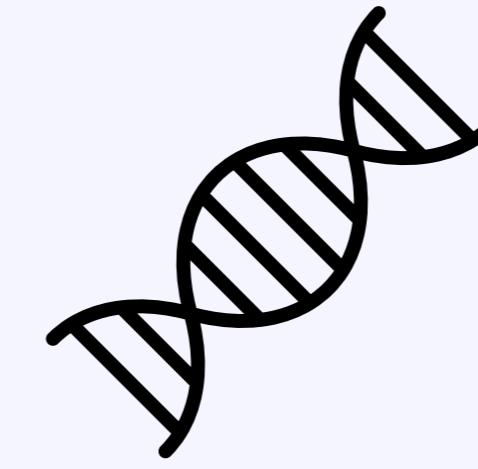
Thibault Latrille
Université de Lausanne
Department of Ecology & Evolution



What is the outline of the presentation?

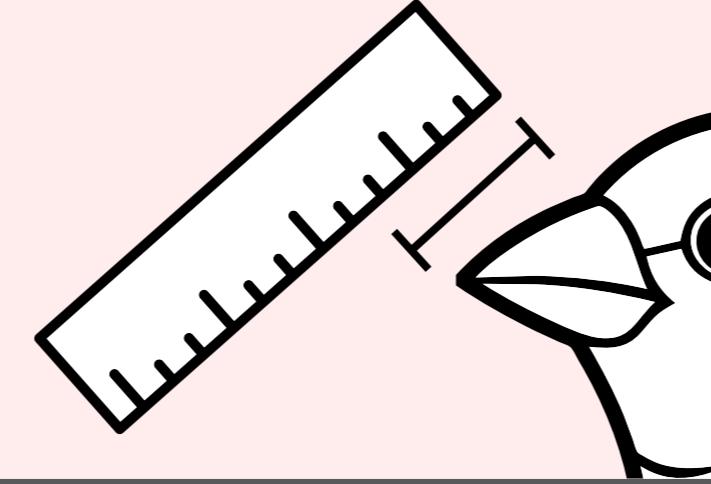
3 questions, 3 parts.

Sequences



- 1) Can we predict the rate of protein evolution?
- 2) Is adaptation on proteins predictable across evolutionary scales?

Traits



- 3) Can we integrate different scales to detect selection on a trait?

What tools to bridge evolutionary scales?

A combination of theoretical models and empirical studies.

Empirical studies

Genes and sites under adaptation at the phylogenetic scale also exhibit adaptation at the population-genetic scale

Thibault Latrille^{a,b,c,1}, Nicolas Rodrigue^d, and Nicolas Lartillot^a

Estimating the proportion of beneficial mutations that are not adaptive in mammals

T. Latrille, J. Joseph, D. A. Hartasánchez, N. Salamin

This article is a preprint

Bridging Time Scales in Evolutionary Biology



Diego A. Hartasánchez, Thibault Latrille, Marina Brasó-Vives, and Arcadi Navarro

Equal contribution: DAH & TL

Inferring Long-Term Effective Population Size with Mutation-Selection Models

Thibault Latrille^{1,*}, Vincent Lanore,¹ and Nicolas Lartillot¹

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

T. Latrille¹, M. Bastian², T. Gaboriau¹, N. Salamin¹

An Improved Codon Modeling Approach for Accurate Estimation of the Mutation Bias

Thibault Latrille^{1,*}, Nicolas Lartillot¹

Quantifying the impact of changes in effective population size and expression level on the rate of coding sequence evolution

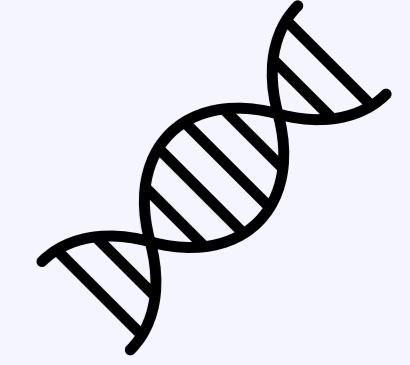
T. Latrille^{a,b,*}, N. Lartillot^a

Theoretical models

PhD

Postdoc

Sequences

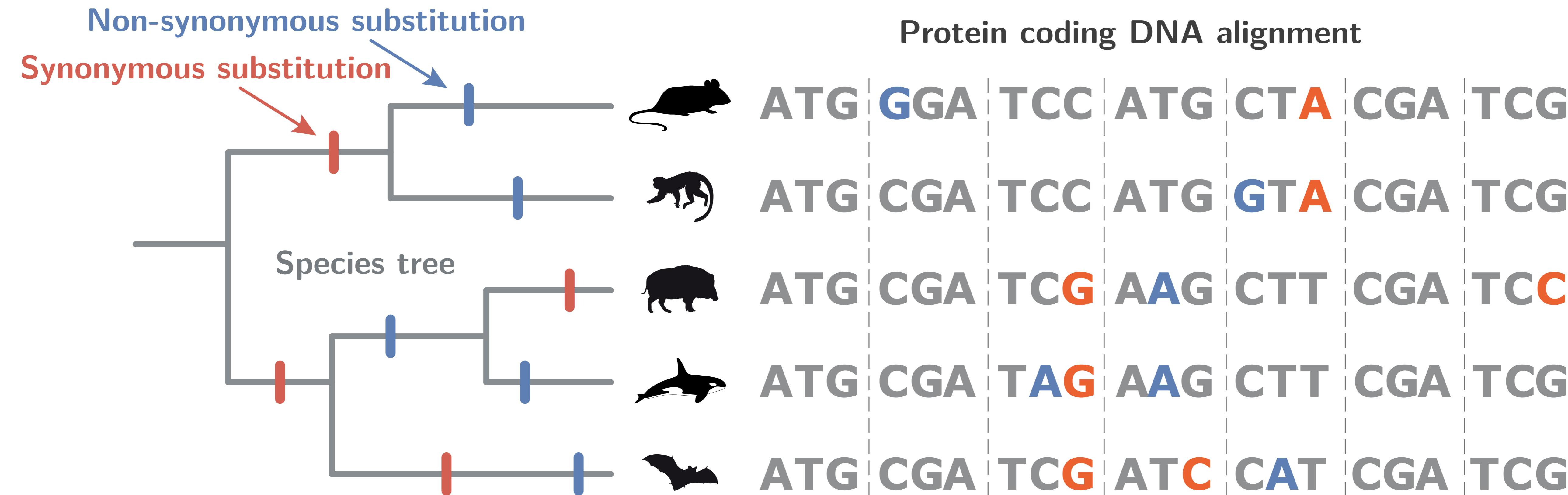


Part I

Can we predict the rate of protein evolution?

How to quantify changes in protein evolution?

With both synonymous and non-synonymous substitutions.



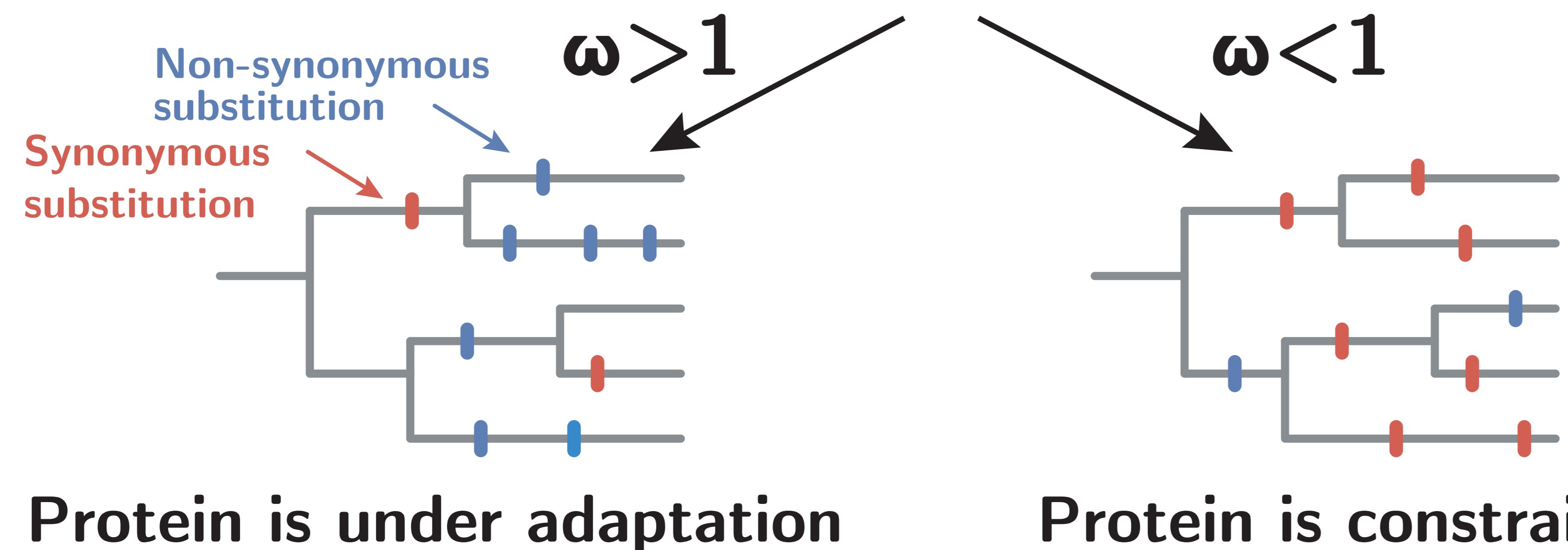
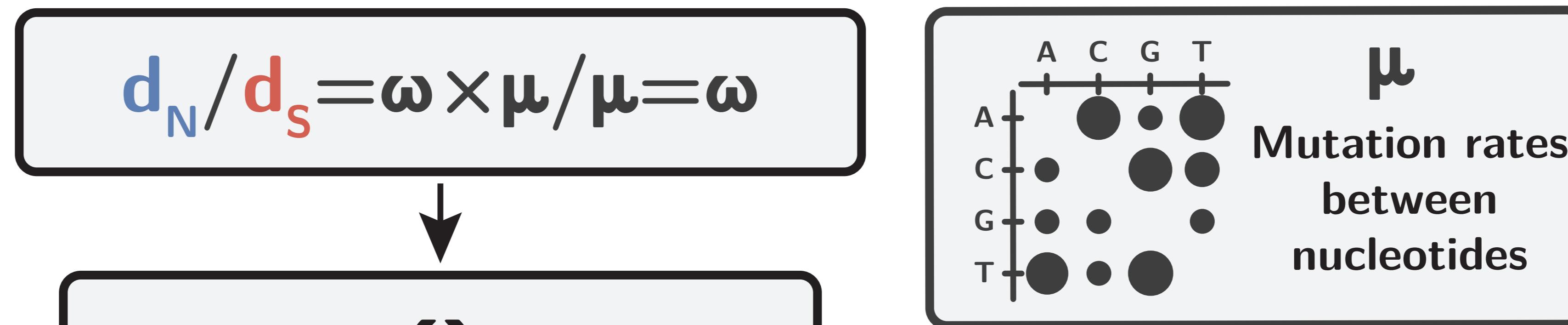
- **Non-synonymous** substitutions are reflecting the effect of mutation, selection and drift.
- **Synonymous** substitutions are considered selectively neutral, reflecting the mutational processes.

King & Jukes (1969); Kimura (1983); Goldman & Yang (1994); Muse & Gaut (1994).

How to measure the rate of protein evolution?

$d_N/d_S = \omega$ as the rate of protein evolution.

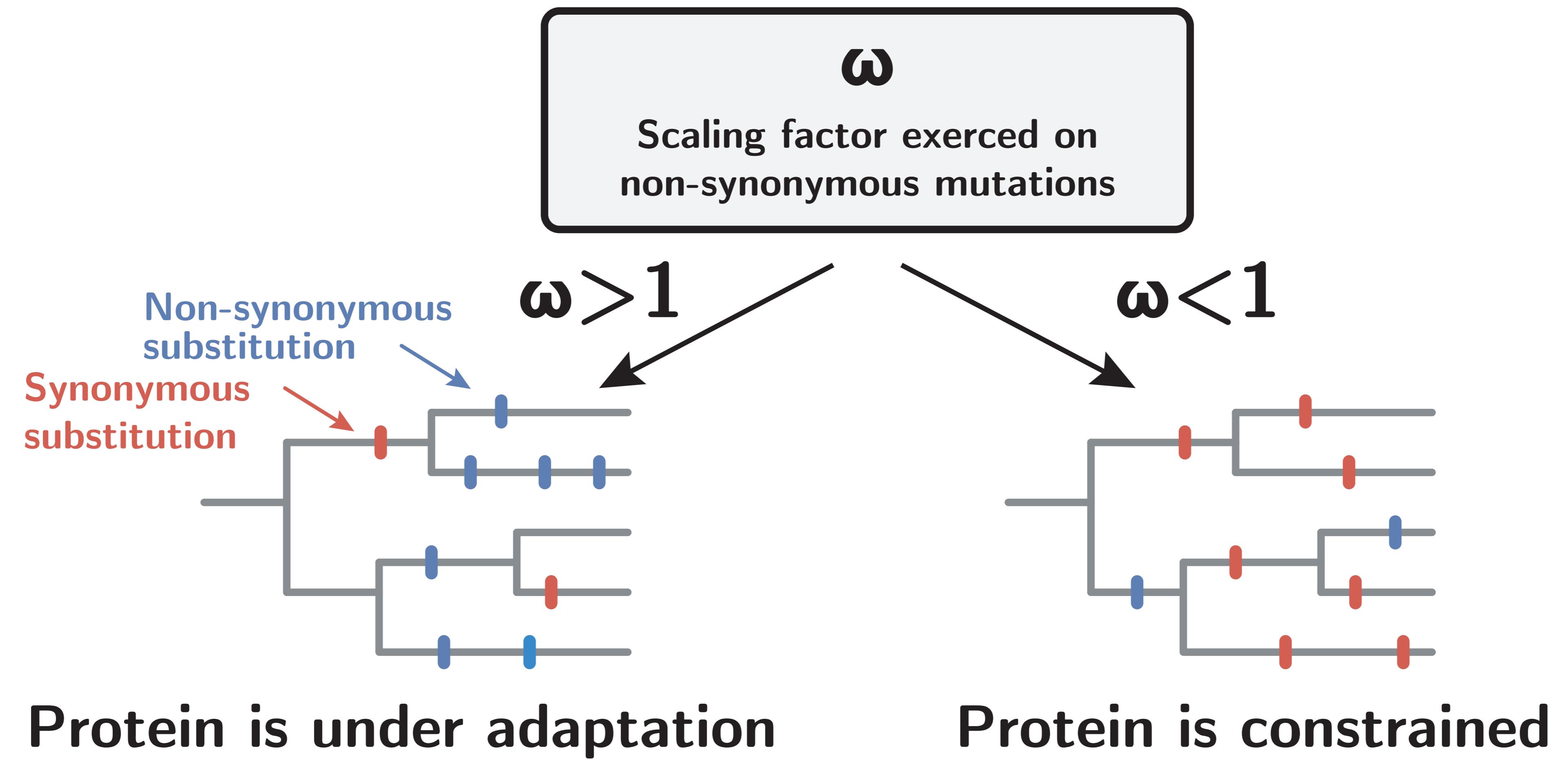
$$\begin{cases} d_S = \mu \text{ for synonymous substitutions.} \\ d_N = \omega \times \mu \text{ for non-synonymous substitutions.} \end{cases}$$



ω can be interpreted as the average fixation probability of non-synonymous mutations, relative to neutral mutations.

What are the predictors of ω ?

Few genes/sites under adaptation ($\omega > 1$), a majority are constrained ($\omega < 1$).

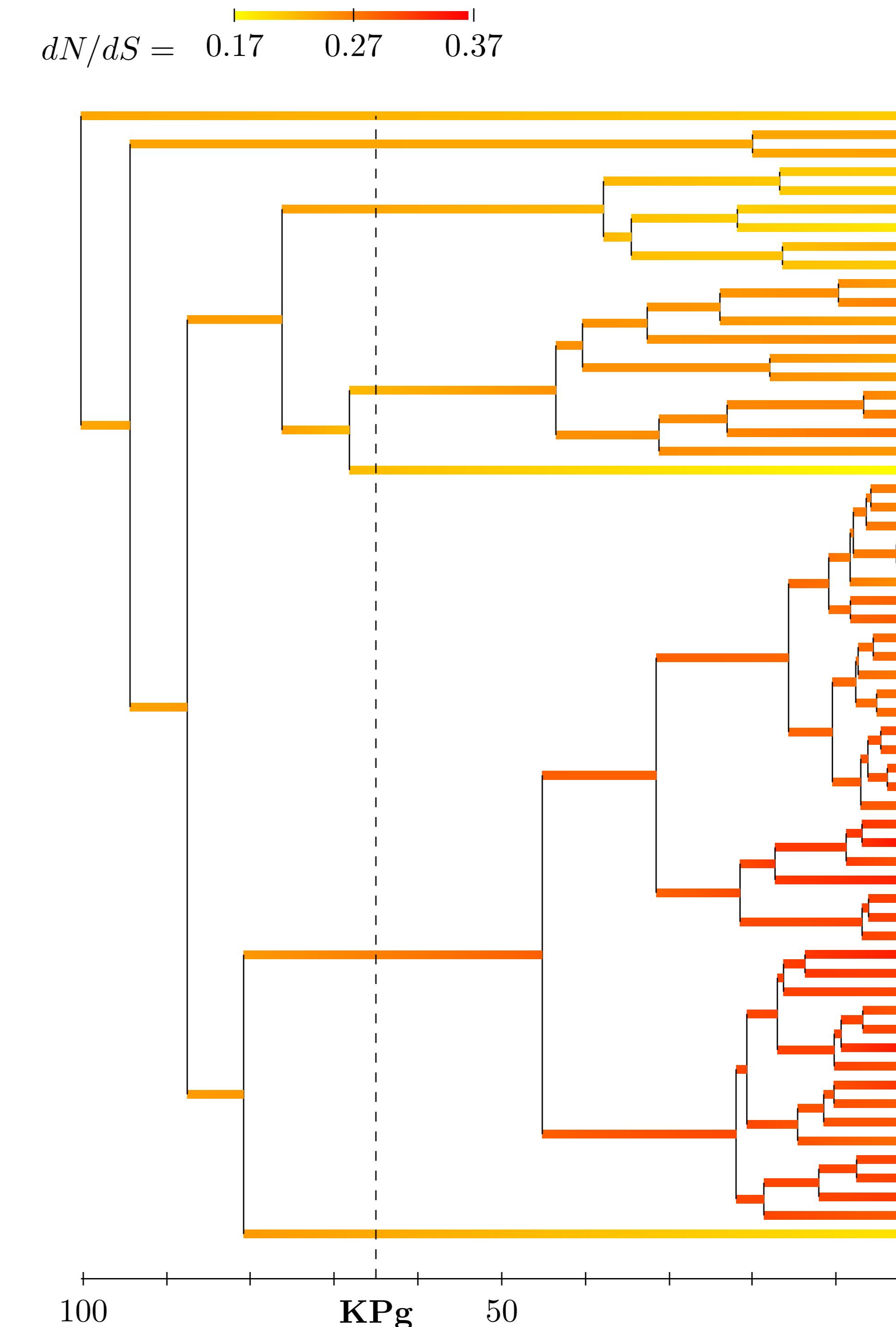
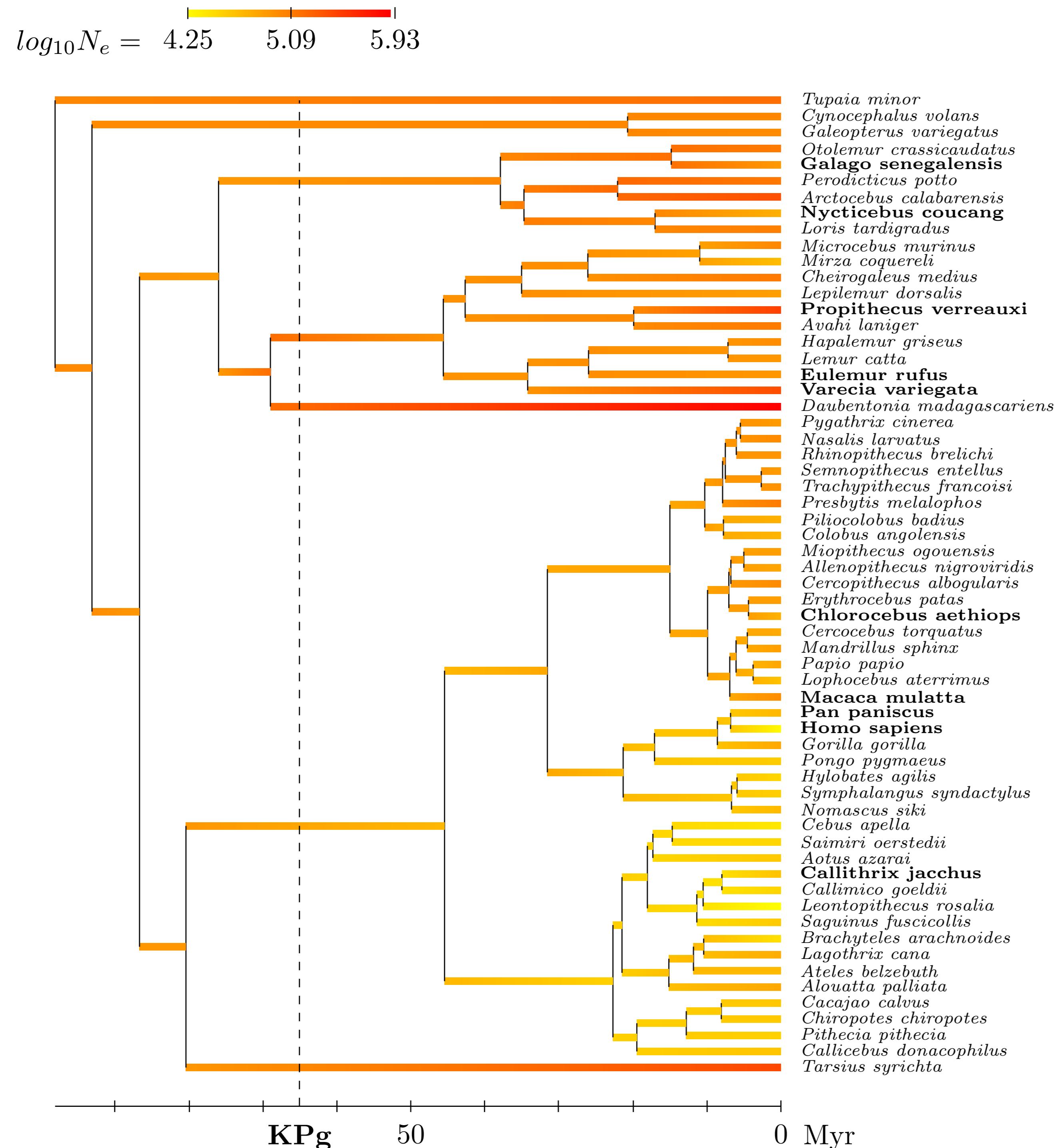


- A very few genes have $\omega > 1$.
Kosiol *et al* (2008).
- But we can detect sites with $\omega > 1$.
Nieslen & Yang (1998); Enard *et al* (2016).
- Some branches can have a transient $\omega > 1$.
Yang & Nielsen (1998); Zhang & Nielsen (2005).

- Lower ω for highly expressed proteins.
Drummond (2005); Zhang & Yang (2015).
- Lower ω for buried sites inside a protein.
Ramsey *et al* (2011); Echave *et al* (2016).
- Lower ω for short-lived and smaller species.
Popadin *et al* (2007); Lanfear *et al* (2010).

Is effective population size (N_e) predicting ω ?

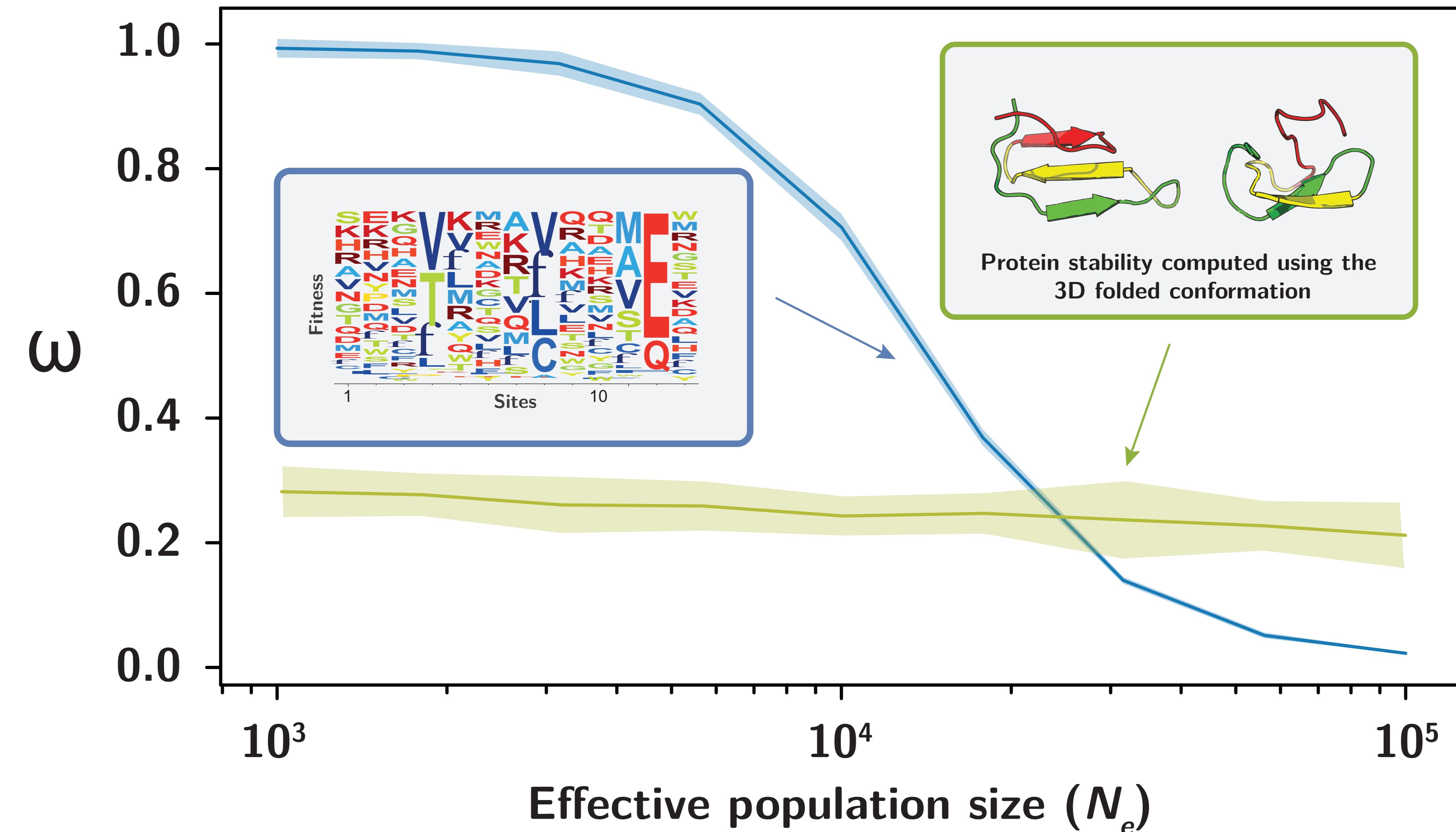
Higher N_e results in lower ω due to better efficacy of selection ($r=-0.58$).



Brevet & Lartillot (2021)

Can we theoretically use ω to predict N_e ?

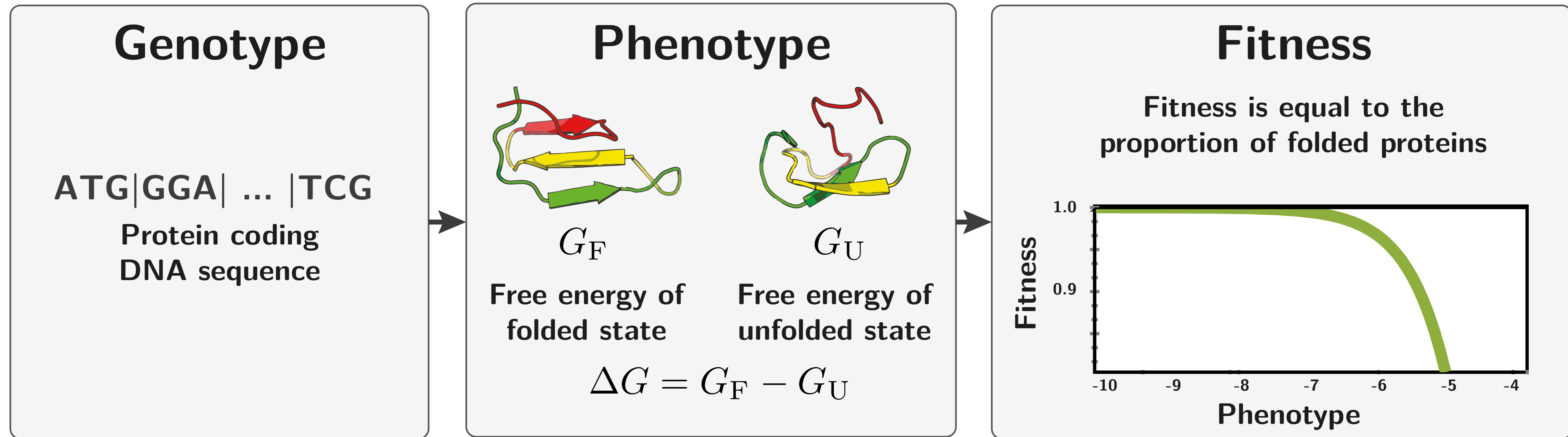
Not directly because the relationship depends on the model of protein evolution.



Latrille et al. (2021)

What is the expected relationship between ω and N_e ? (1/4)

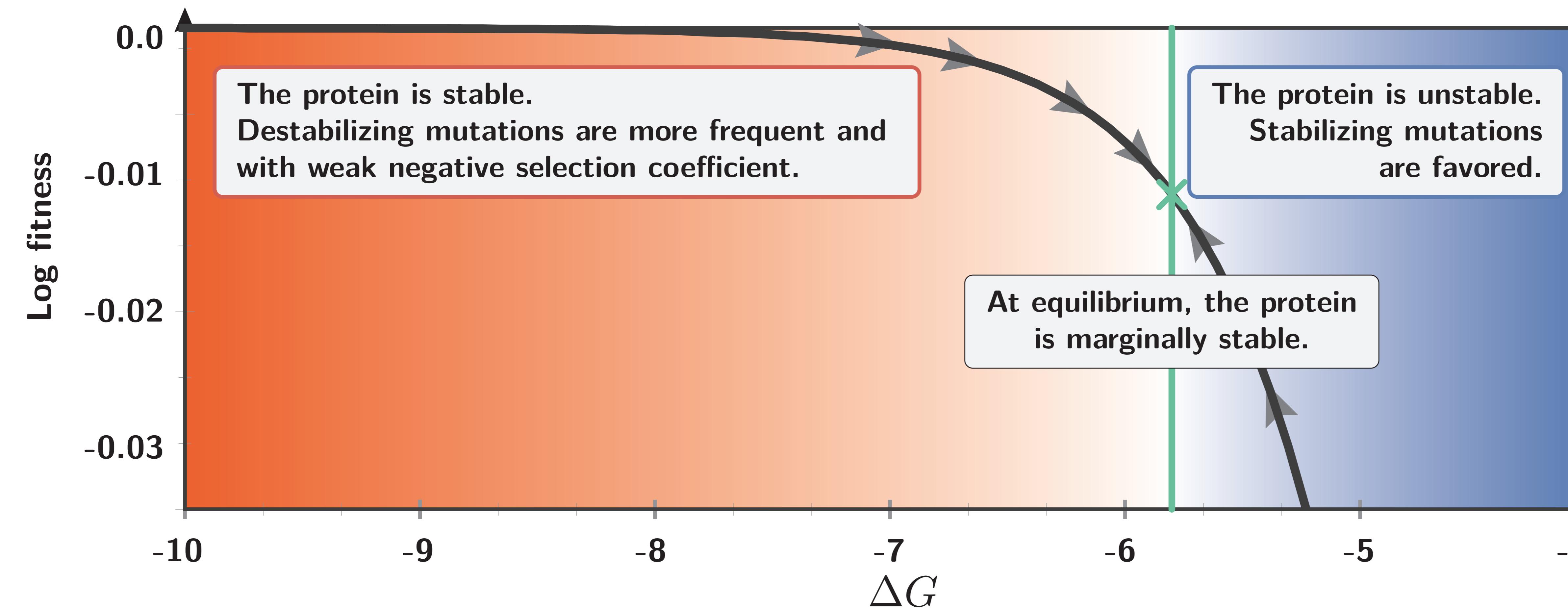
We first need to define a genotype-phenotype-fitness relationship.



Miyazawa and Jernigan (1985), Williams et al (2006), Goldstein (2011), Pollock et al (2012)

What is the expected relationship between ω and N_e ? (2/4)

Then we need to find the equilibrium and ω at this equilibrium.

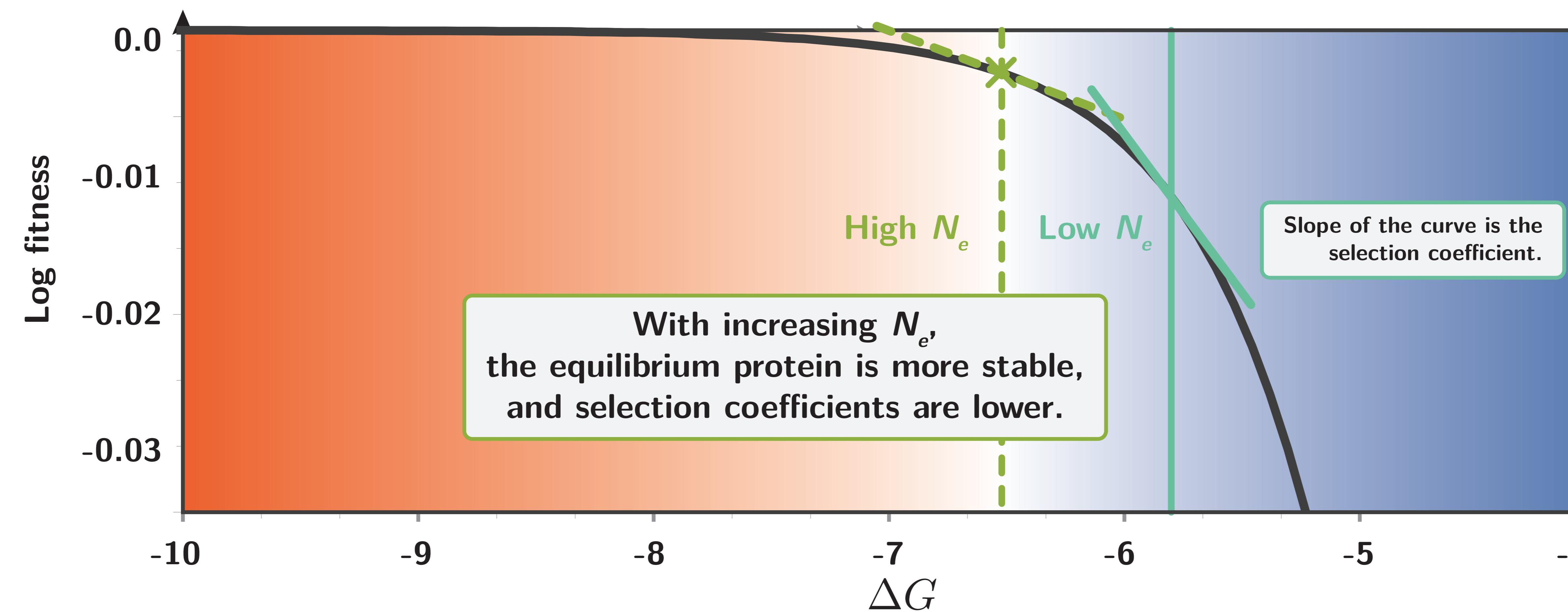


- The optimal stability of proteins is never achieved.
- Marginal stability is the default expectation of the mutation-selection balance even under directional selection for stability.

Taverna & Goldstein (2002)

What is the expected relationship between ω and N_e ? (3/4)

Then we derive how changes in N_e shift the equilibrium.



- Selection coefficient is dependent on the position in the fitness landscape.
- We can then derive the relationship between N_e and ω as a function of the microscopic molecular parameters of the model.

Cherry (1998); Goldstein (2013).

What is the expected relationship between ω and N_e ? (4/4)

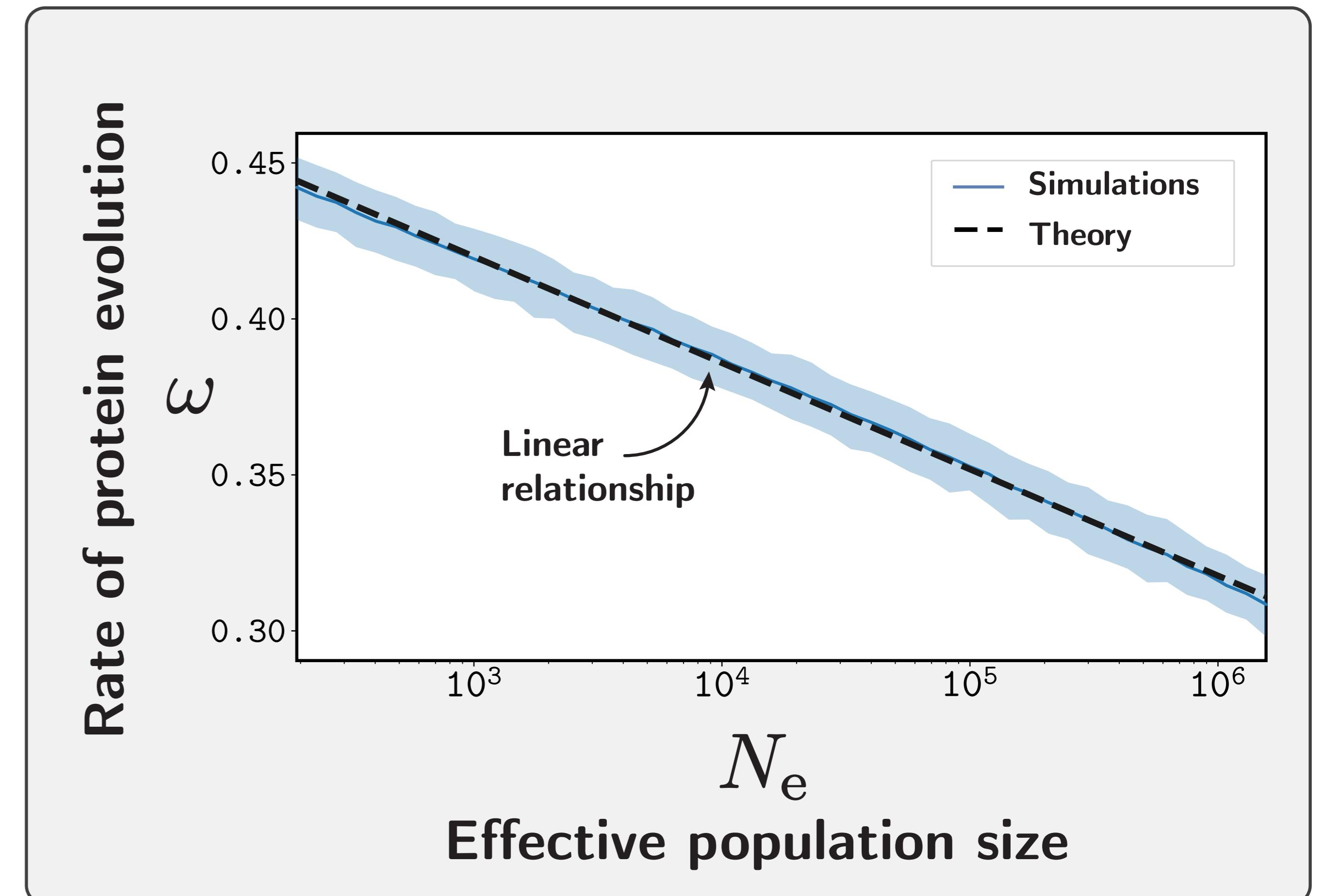
Negative linear relationship between ω and $\log(N_e)$.

Given:

- T the temperature.
- n the number of sites in the protein.
- $\Delta\Delta G > 0$ the destabilizing effect of a mutation.
- x the proportion of destabilizing sites (phenotype).
- $f(x)$ the phenotype-fitness map.
- x^* the equilibrium of x .

The response in ω after a change in N_e is:

$$\frac{d\omega}{d \ln(N_e)} \simeq - \frac{\frac{\partial \ln f(x^*)}{\partial x^*}}{\frac{\partial^2 \ln f(x^*)}{\partial x^{*2}}} \simeq - \frac{T}{n \times \Delta\Delta G}.$$



Latrille & Lartillot (2021)

What is the relationship between ω and expression level?

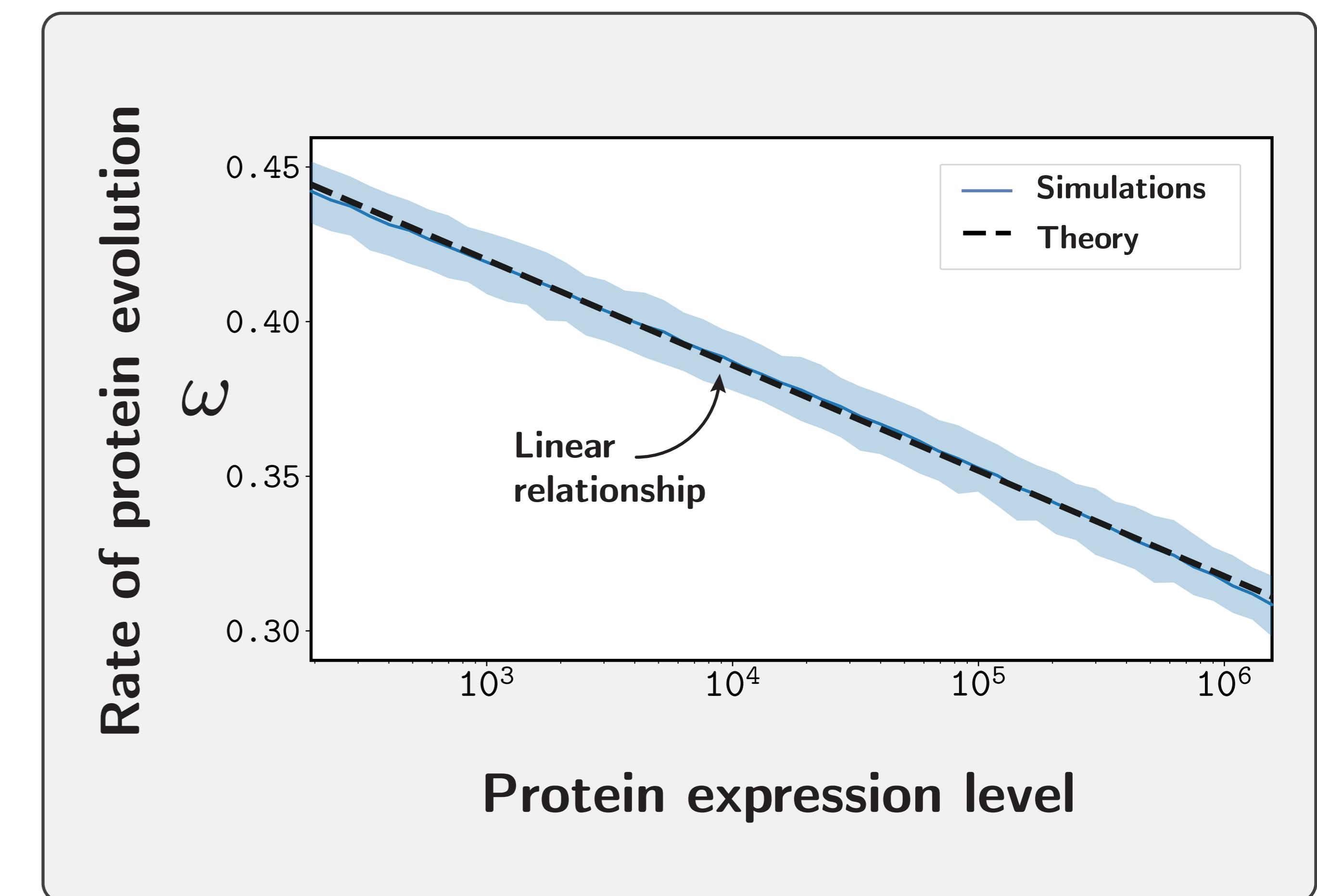
Negative linear relationship between ω and log of expression level.

If misfolded proteins are toxic, the decrease in fitness is proportional to protein expression level.

- T the temperature.
- n the number of sites in the protein.
- $\Delta\Delta G > 0$ the destabilizing effect of a mutation.
- x the proportion of destabilizing sites (phenotype).
- $f(x)$ the phenotype-fitness map.
- x^* the equilibrium of x .

The response in ω after a change in protein expression

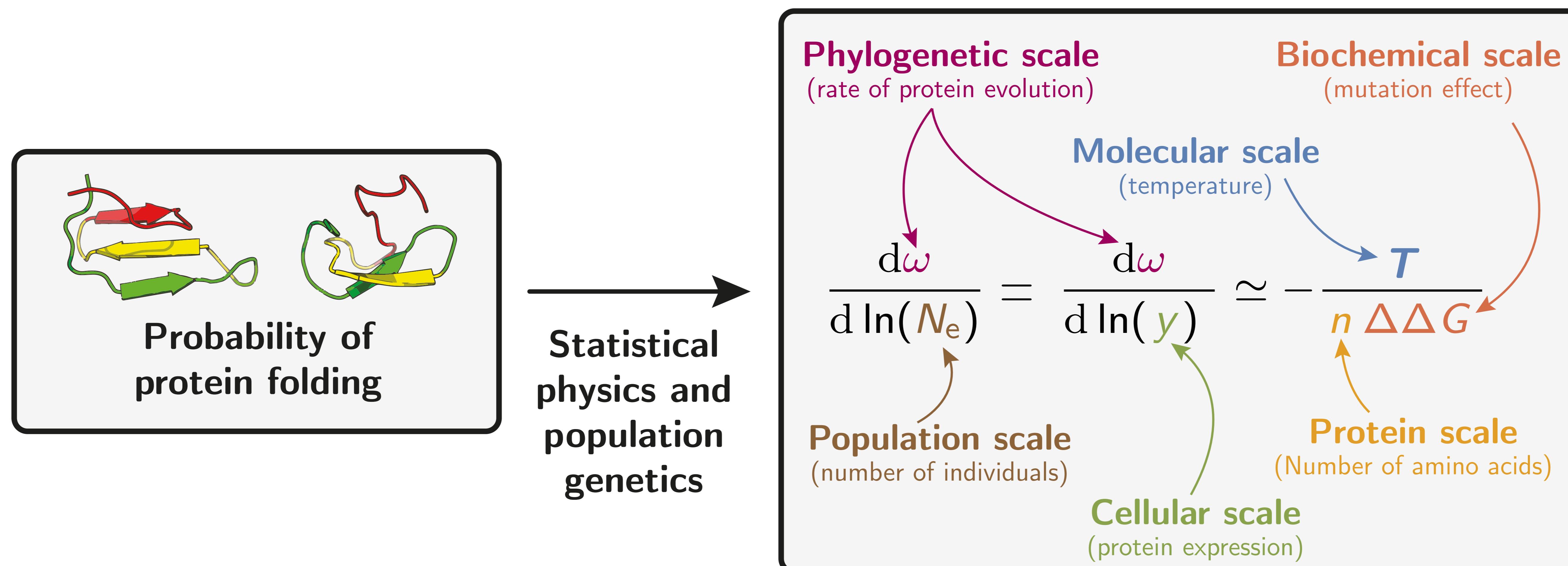
$$\frac{d\omega}{d \ln(y)} \simeq - \frac{\frac{\partial \ln f(x^*)}{\partial x^*}}{\frac{\partial^2 \ln f(x^*)}{\partial x^{*2}}} \simeq - \frac{T}{n \times \Delta\Delta G}.$$



Latrille & Lartillot (2021)

Can theoretical models of protein folding predict rate of evolution?

Models form a bridge across different scales and can be tested.



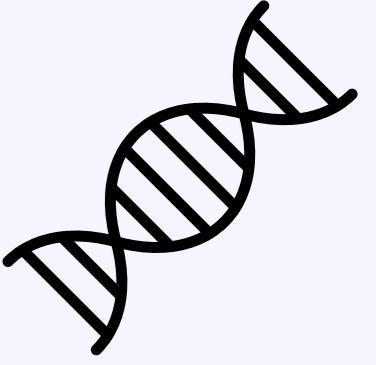
Chapter I

Can we predict the rate of protein evolution (ω)?

- With a theoretical model for selection on protein folding, ω is linearly decreasing with N_e and expression level (on log scale).
- This model forms a bridge across different scales and can be tested.
- In our model, there is no adaptation possible, ω is always < 1 .
- How to detect adaptation when proteins are generally constrained?

Quantifying the impact of changes in effective population size and expression level on the rate of coding sequence evolution

T. Latrille ^{a,b,*}, N. Lartillot ^a



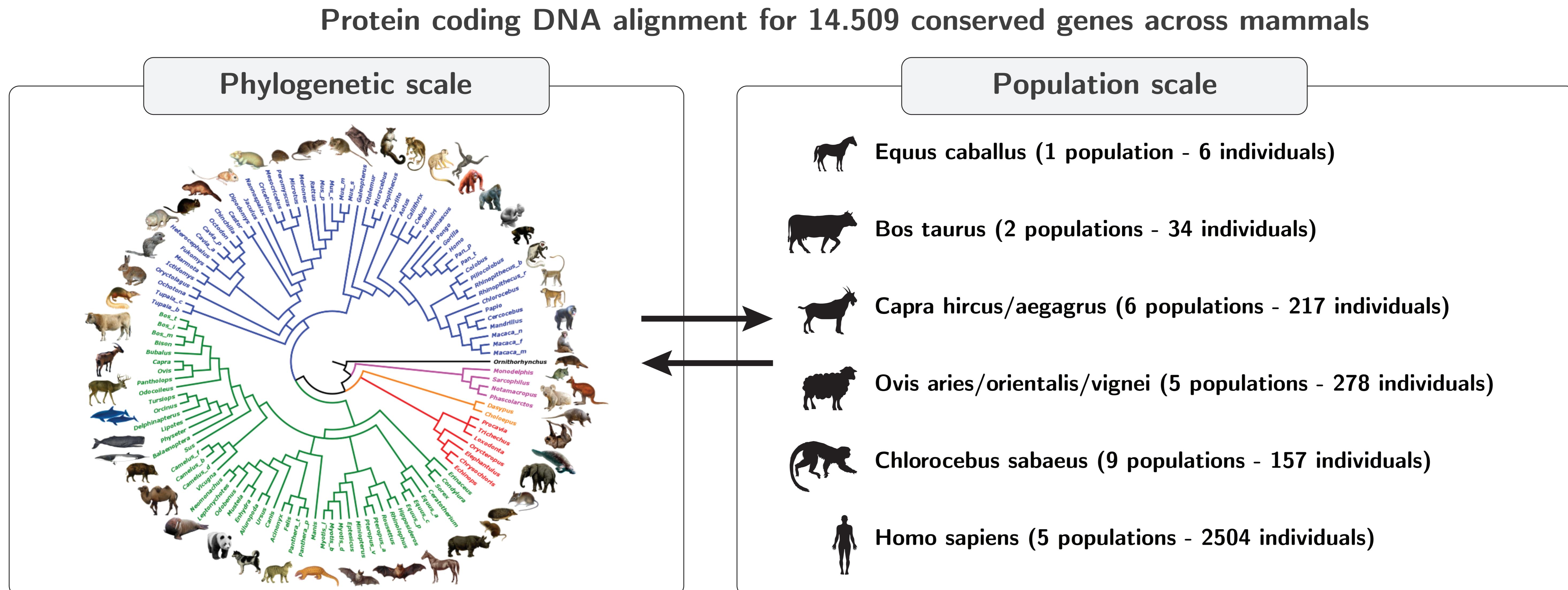
Part II

How to detect adaptation?

Is adaptation predictable across evolutionary scales?

Is adaptation predictable across evolutionary scales?

We first need to test for adaptation at different scales

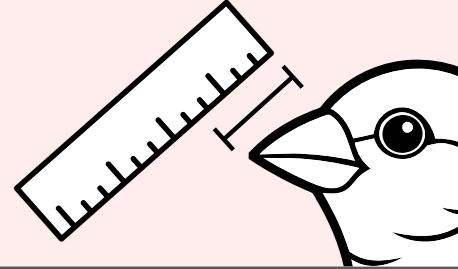


Scornavacca et al (2019); Howe et al (2021)

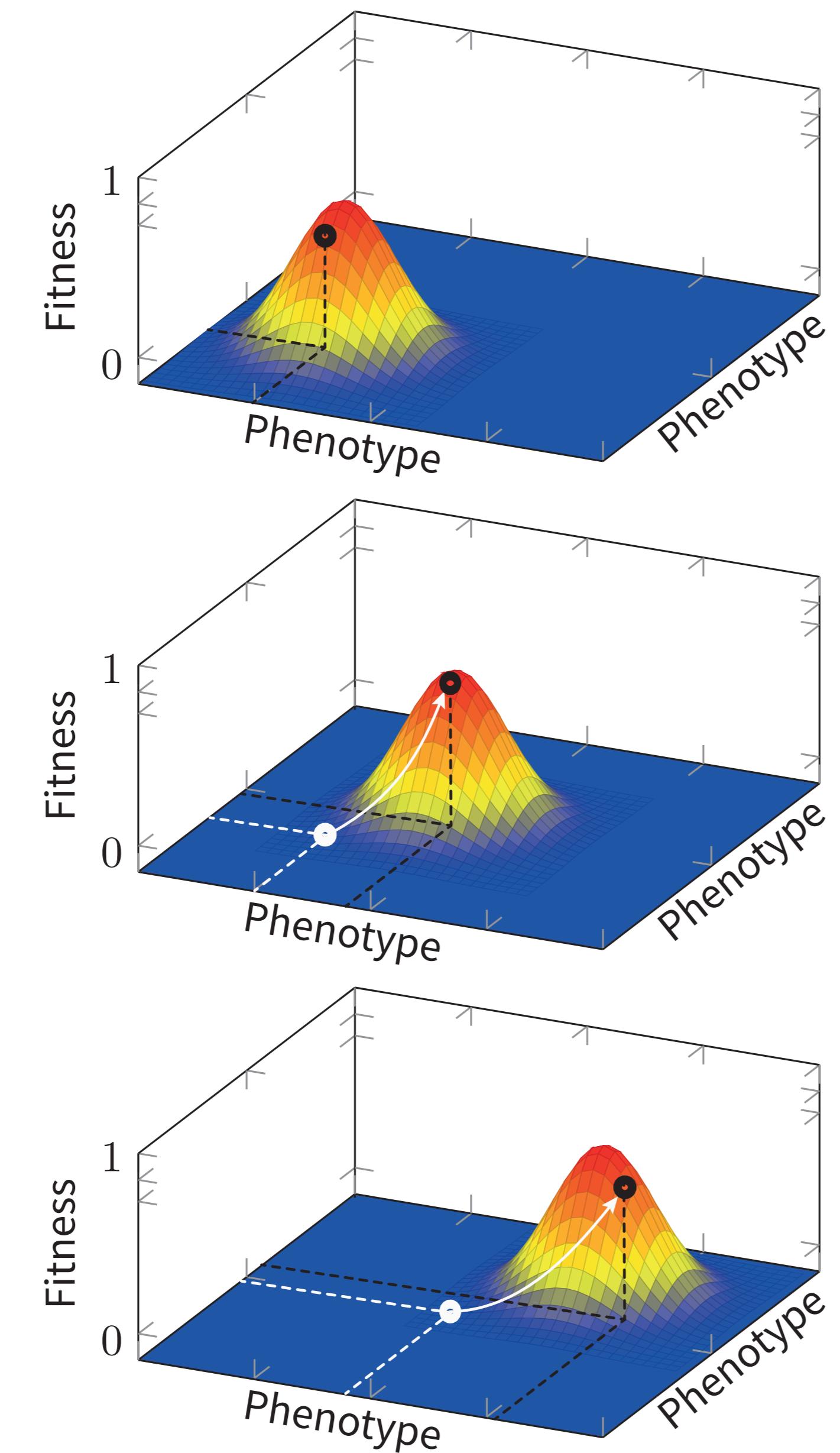
What is adaptation?

Adaptation occurs on a changing fitness landscape

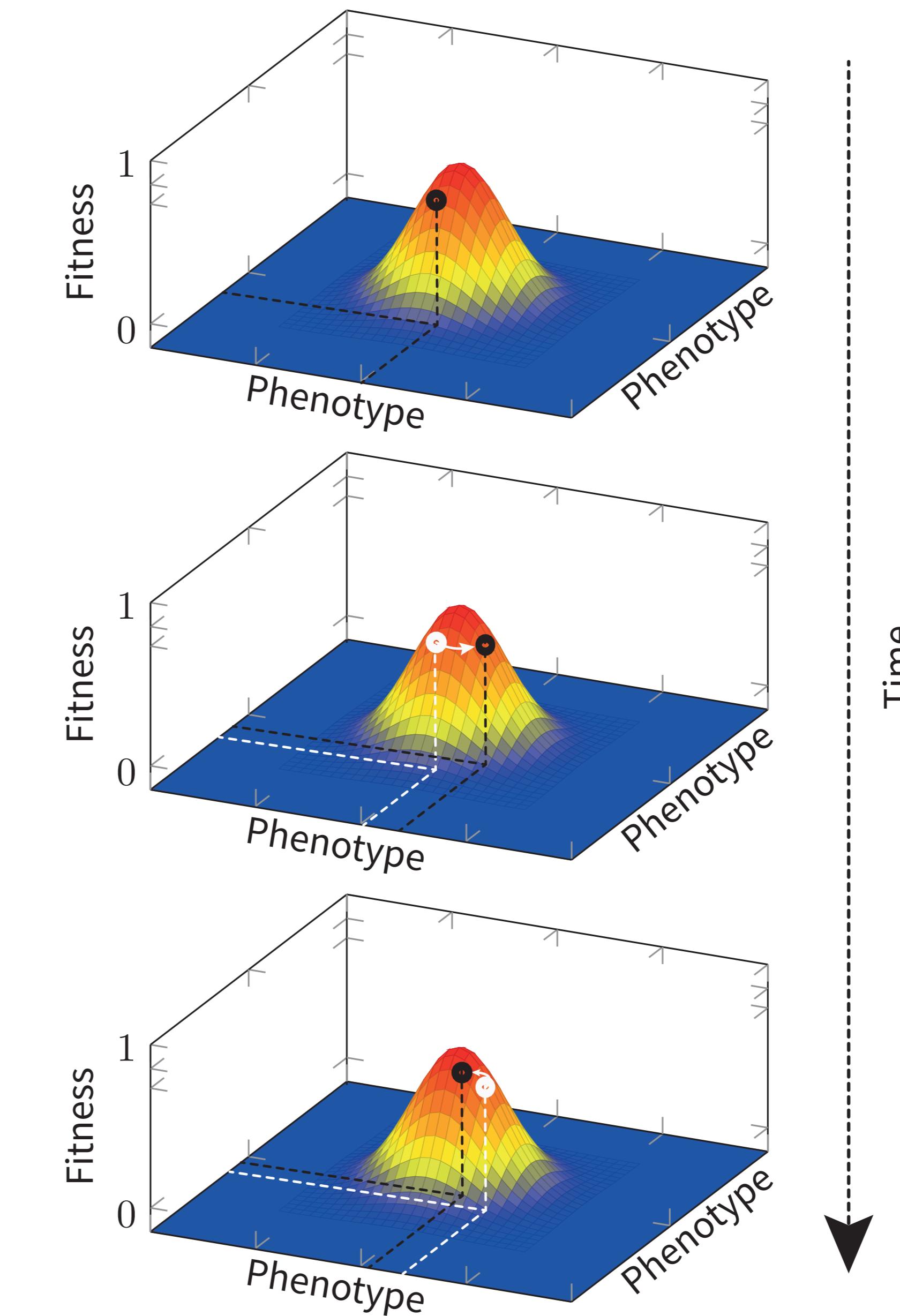
Traits



Changing fitness landscape



Stable fitness landscape



- The optimal state is a moving target.
- Environmental changes that are external.

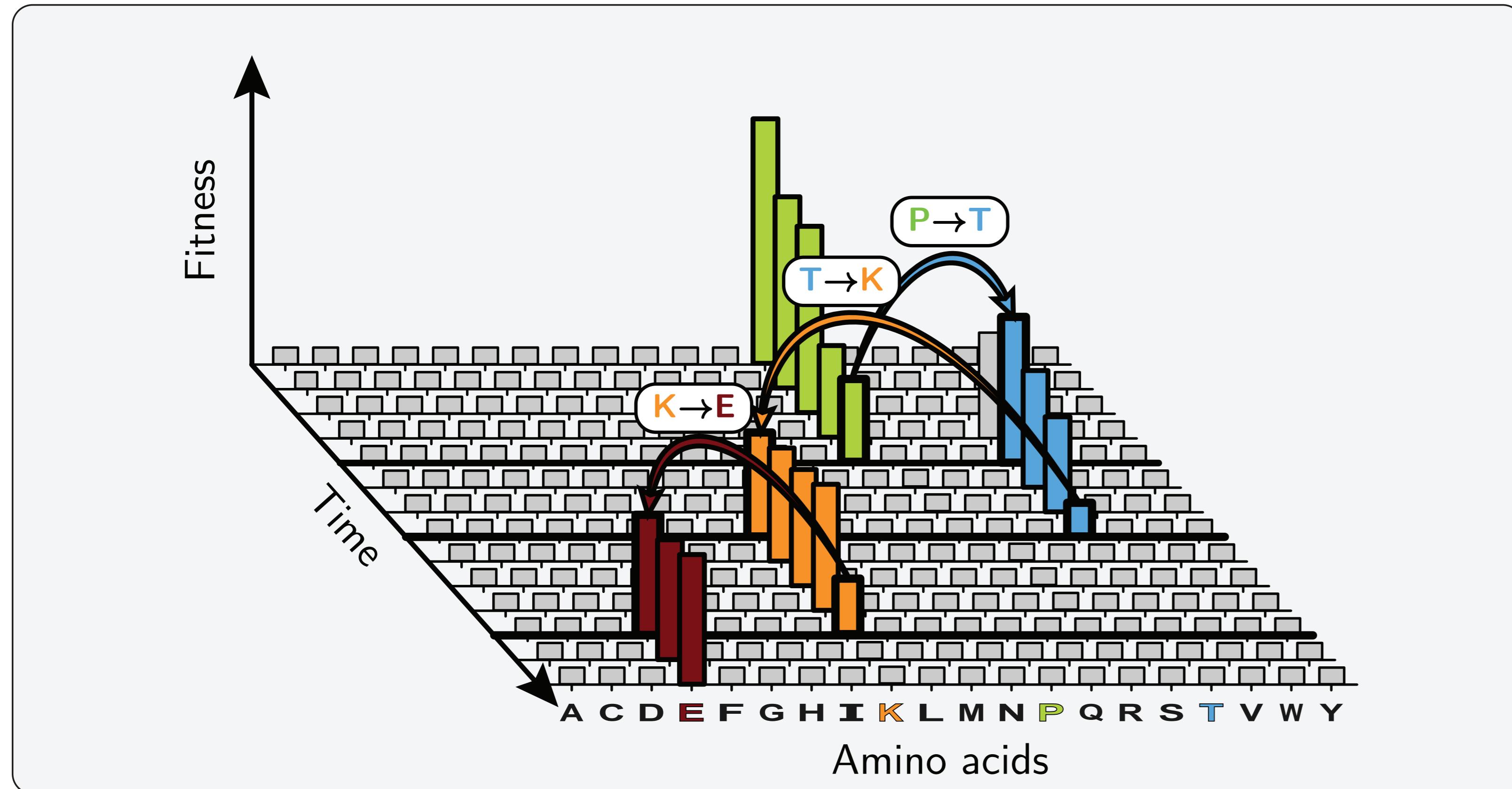
Sella & Hirsh (2005); Mustonen & Lässig (2009)

What is adaptation? (protein coding DNA case)

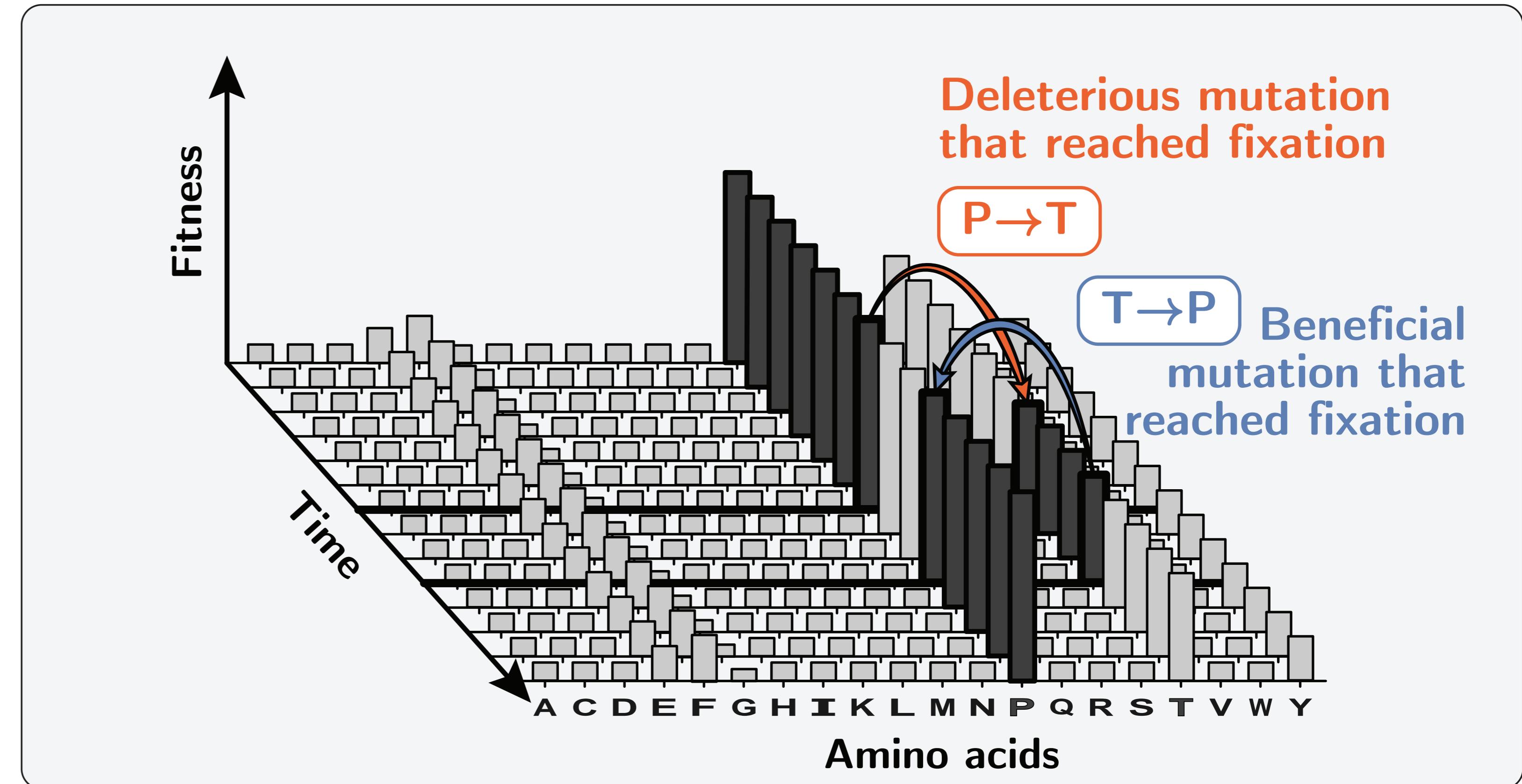
Adaptation occurs on a changing fitness landscape



Changing fitness landscape



Stable fitness landscape



- The optimal state is a moving target.
- Environmental changes that are external.

- A stable fitness landscape is a null model of evolution without adaptation.

Sella & Hirsh (2005); Mustonen & Lässig (2009)

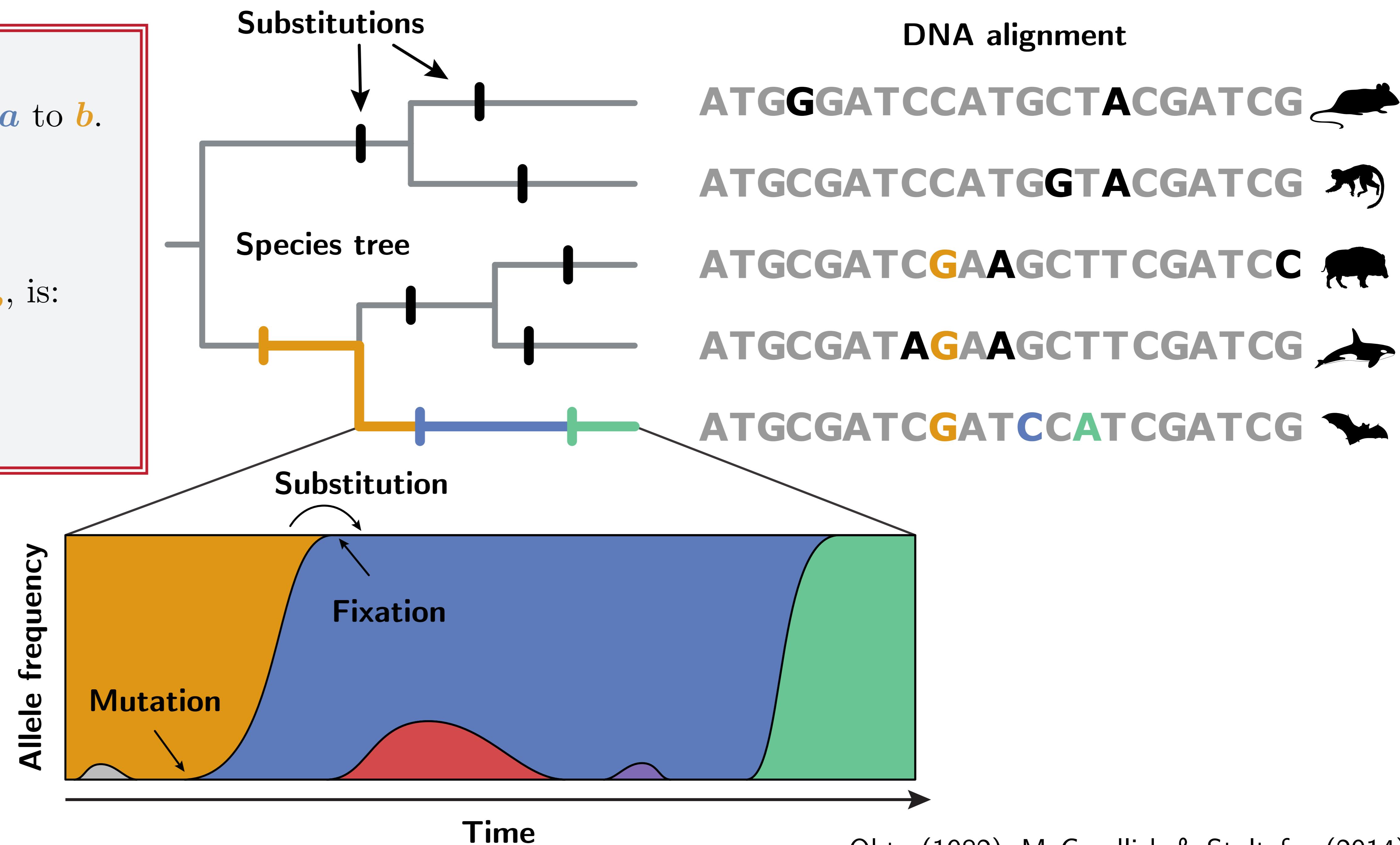
Can we estimate fitnesses from the pattern of substitutions?

Whether a mutation reached fixation depends on its fitness effect.

- $\mu_{a \rightarrow b}$ the mutational rate from state a to b .
- F_a the scaled fitness of state a .
- F_b the scaled fitness of state b .

The substitution rate from a to b , $q_{a \rightarrow b}$, is:

$$q_{a \rightarrow b} = \mu_{a \rightarrow b} \frac{F_b - F_a}{1 - e^{F_a - F_b}}.$$



Ohta (1982); McCandlish & Stoltzfus (2014)

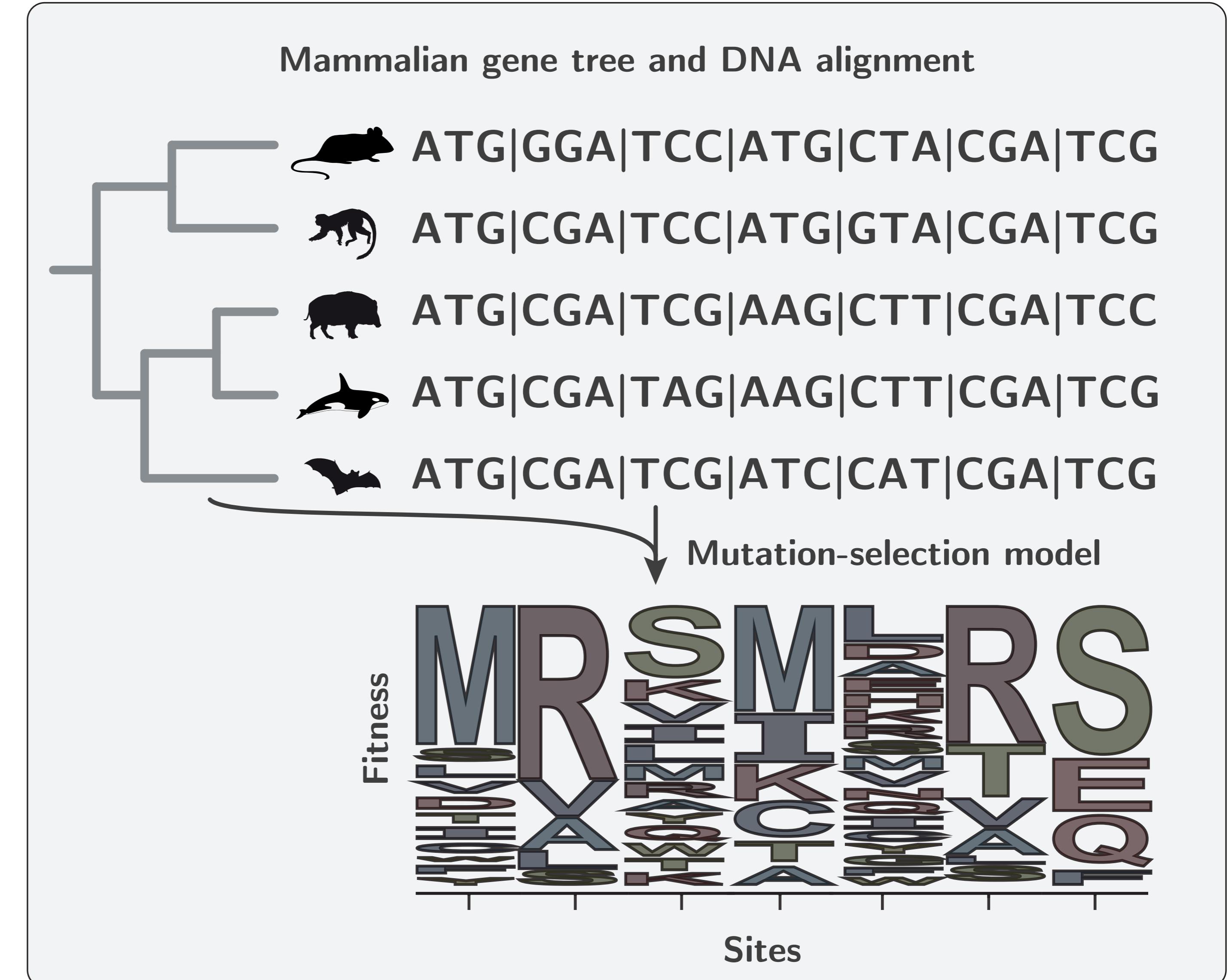
How to estimate the contribution of amino acids to fitness?

By fitting a mutation-selection model at the phylogenetic scale.

Mutation-selection model

- $\mu_{a \rightarrow b}$: mutation rate from codon a to b .
- $q_{a \rightarrow b}$: substitution rate from codon a to b .
- F_a : scaled fitness of the amino-acid encoded by codon a (F_b for codon b).

$$\begin{cases} q_{a \rightarrow b} = \mu_{a \rightarrow b} & \text{if synonymous,} \\ q_{a \rightarrow b} = \mu_{a \rightarrow b} \times \frac{F_b - F_a}{1 - e^{F_a - F_b}} & \text{if non-synonymous.} \end{cases}$$



- **Input: alignment of protein-coding DNA sequences and phylogenetic tree.**
- **Output: amino-acid fitness profiles estimated by mutation-selection models.**

Halpern & Bruno (1998); Tamuri & Goldstein (2012); Rodrigue & Lartillot (2017); Rodrigue et al (2021)

Can we predict the rate of protein evolution while assuming no adaptation?

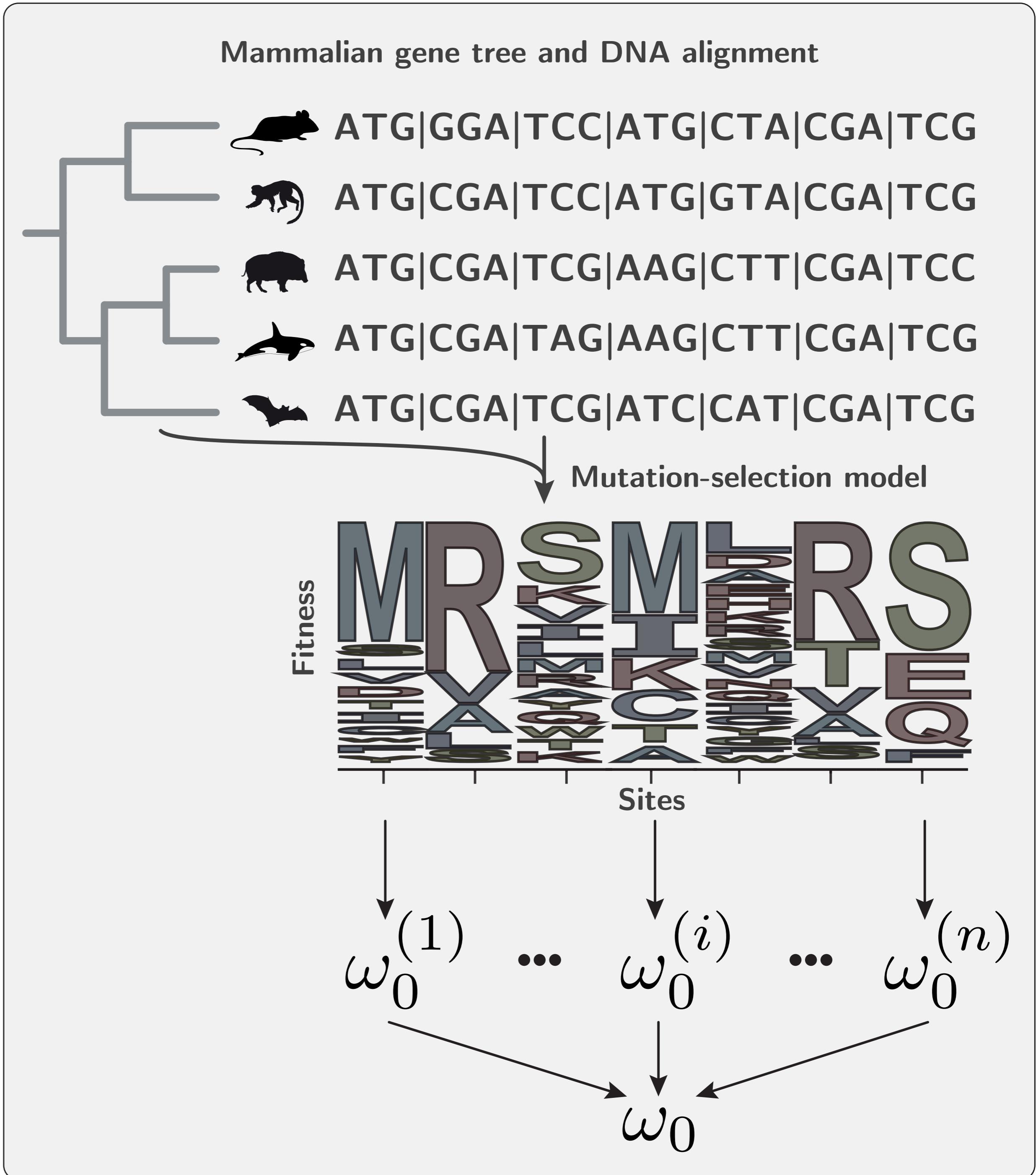
We first fit a stable first landscape then predict the rate of protein evolution.

Mutation-selection model

- $\mu_{a \rightarrow b}$: mutation rate from codon a to b .
- $q_{a \rightarrow b}^{(i)}$: substitution rate from codon a to b at site i .
- $\pi_a^{(i)}$: equilibrium frequency of codon a at site i .

$$\omega_0^{(i)} = \frac{\langle \pi_a^{(i)} q_{a \rightarrow b}^{(i)} \rangle}{\langle \pi_a^{(i)} \mu_{a \rightarrow b} \rangle},$$
$$\Rightarrow \omega_0 = \frac{1}{n} \sum_{i=1}^n \omega_0^{(i)}.$$

- $\langle \cdot \rangle$ is the average over all pairs of non-synonymous codons.
- n : number of codon sites in the DNA alignment.

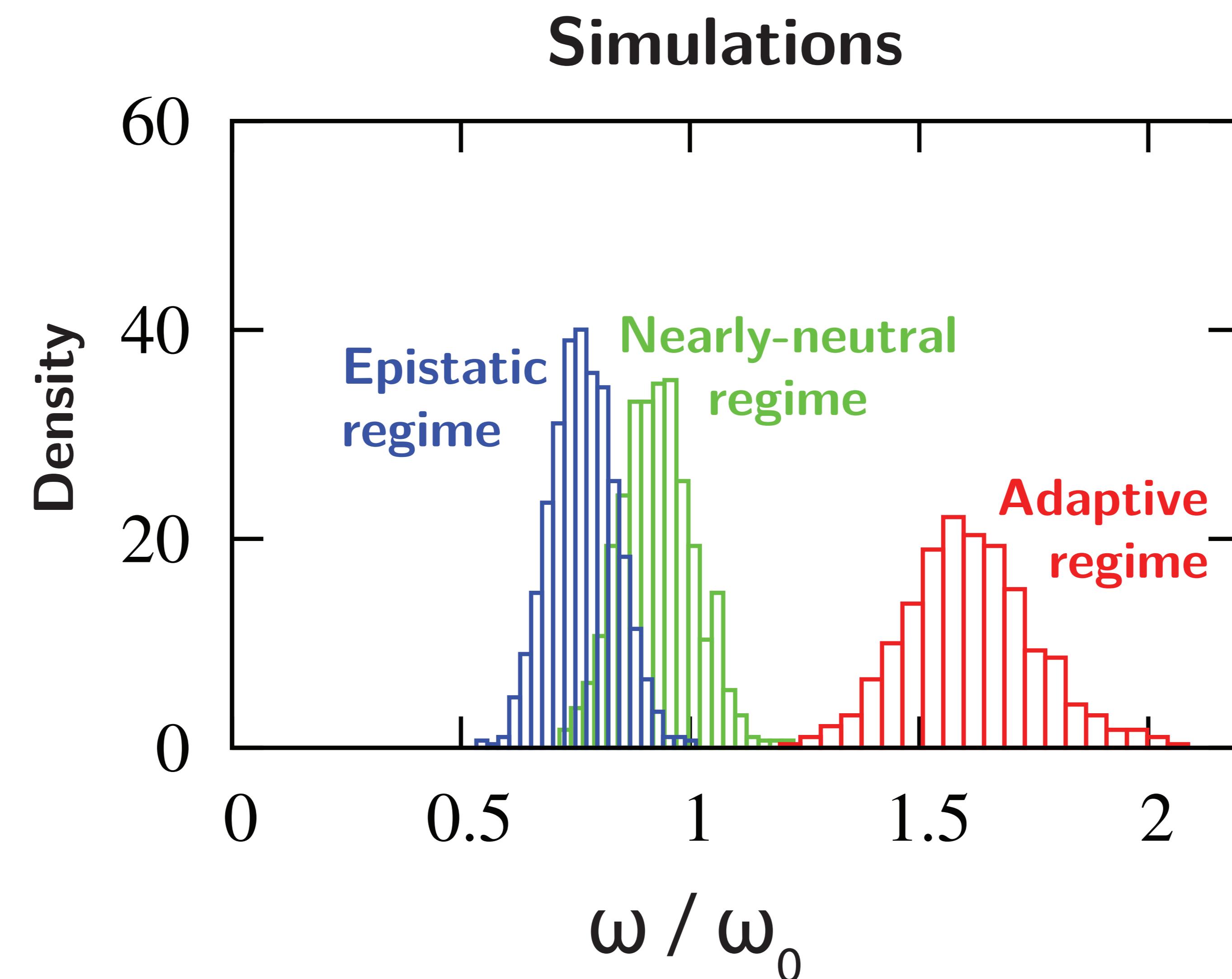


Spielman & Wilke (2015); Dos Reis (2015); Rodrigue & Lartillot (2017)

How to use the mutation-selection model to detect adaptation?

Contrasting ω and ω_0 to detect a changing fitness landscape.

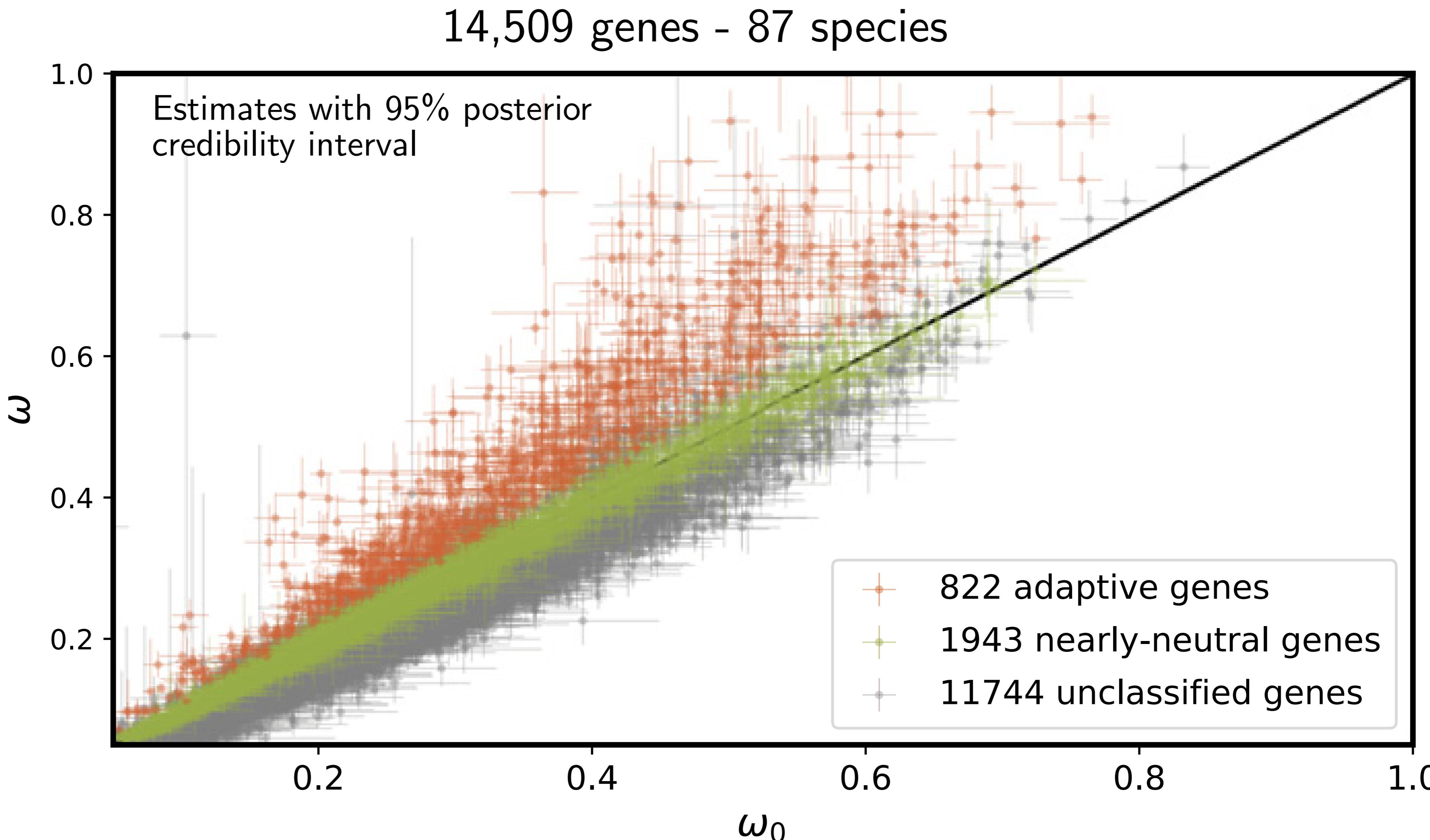
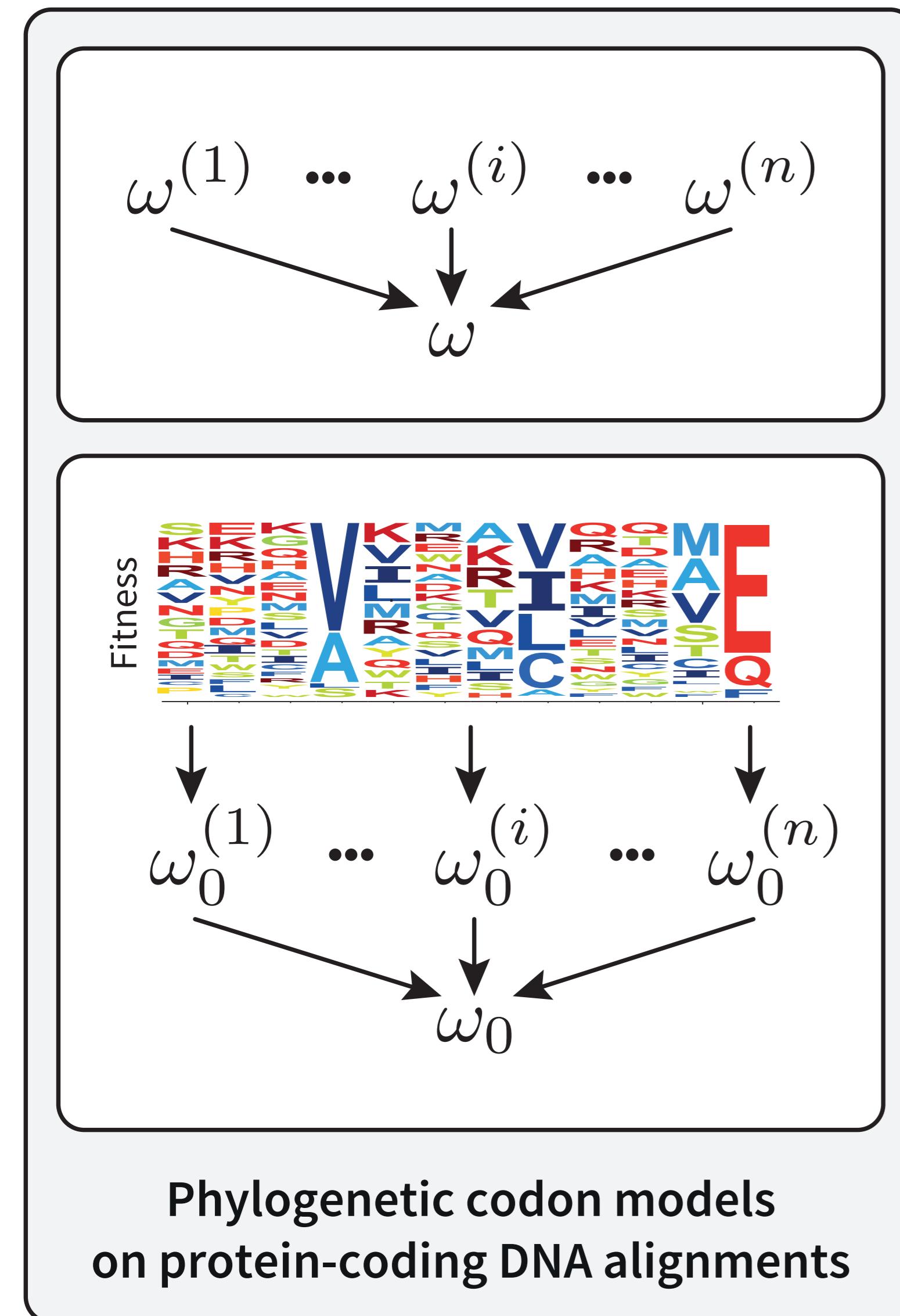
- ω : estimated rate of evolution under classical codon model.
- ω_0 : predicted rate of evolution under the mutation-selection model.



Rodrigue & Lartillot (2017)

Can the mutation-selection model detect adaptation?

We can detect genes with $\omega > \omega_0$ across mammals while they still have $\omega < 1$.

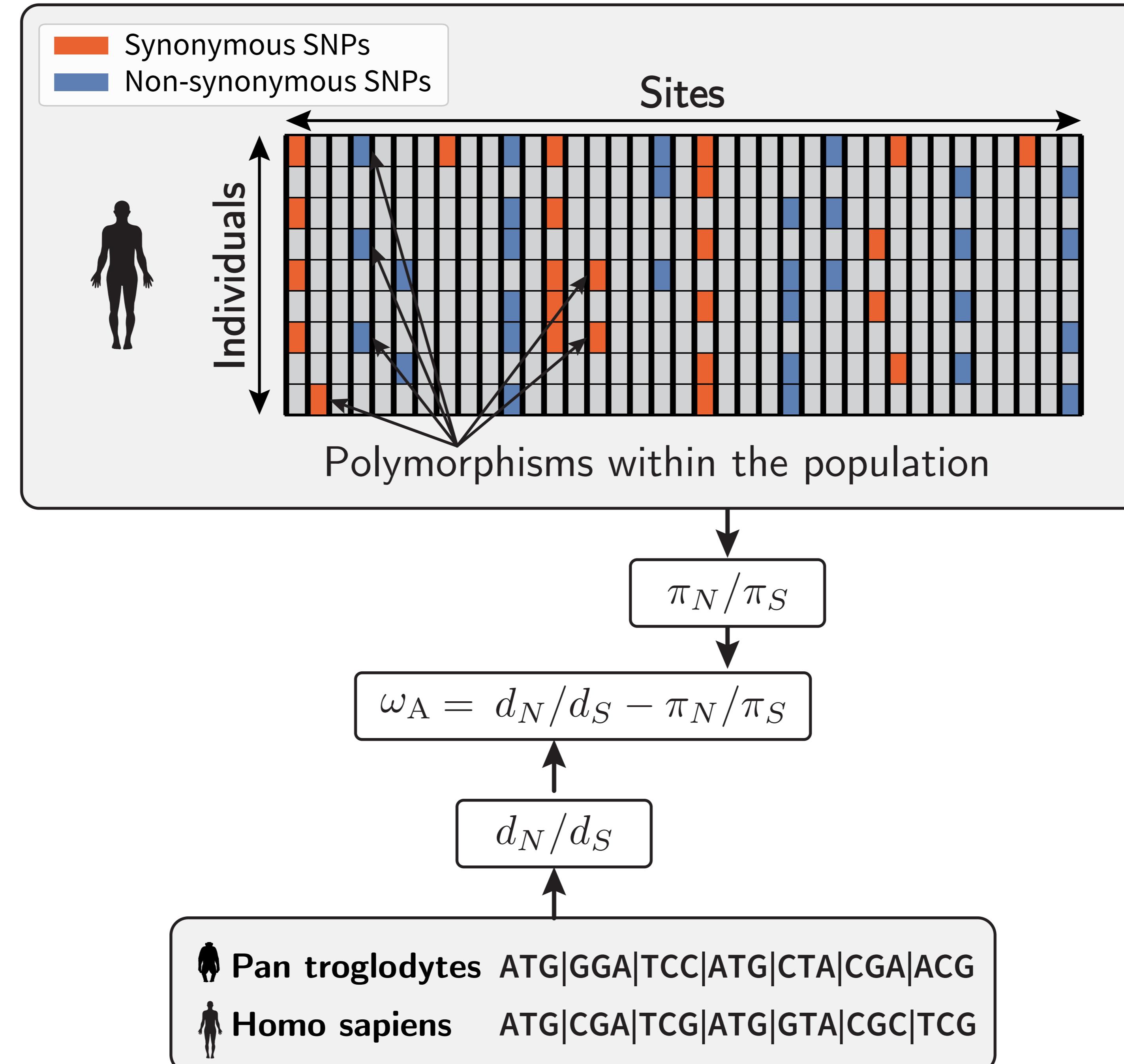


- Genes predicted to be under adaptation at the phylogenetic scale are enriched in ontologies related to immunity, response to virus and external membrane.

Latrille et al. (2023)

How to test for adaptation in a lineage?

Contrasting substitutions to polymorphism with the McDonald & Kreitman test.

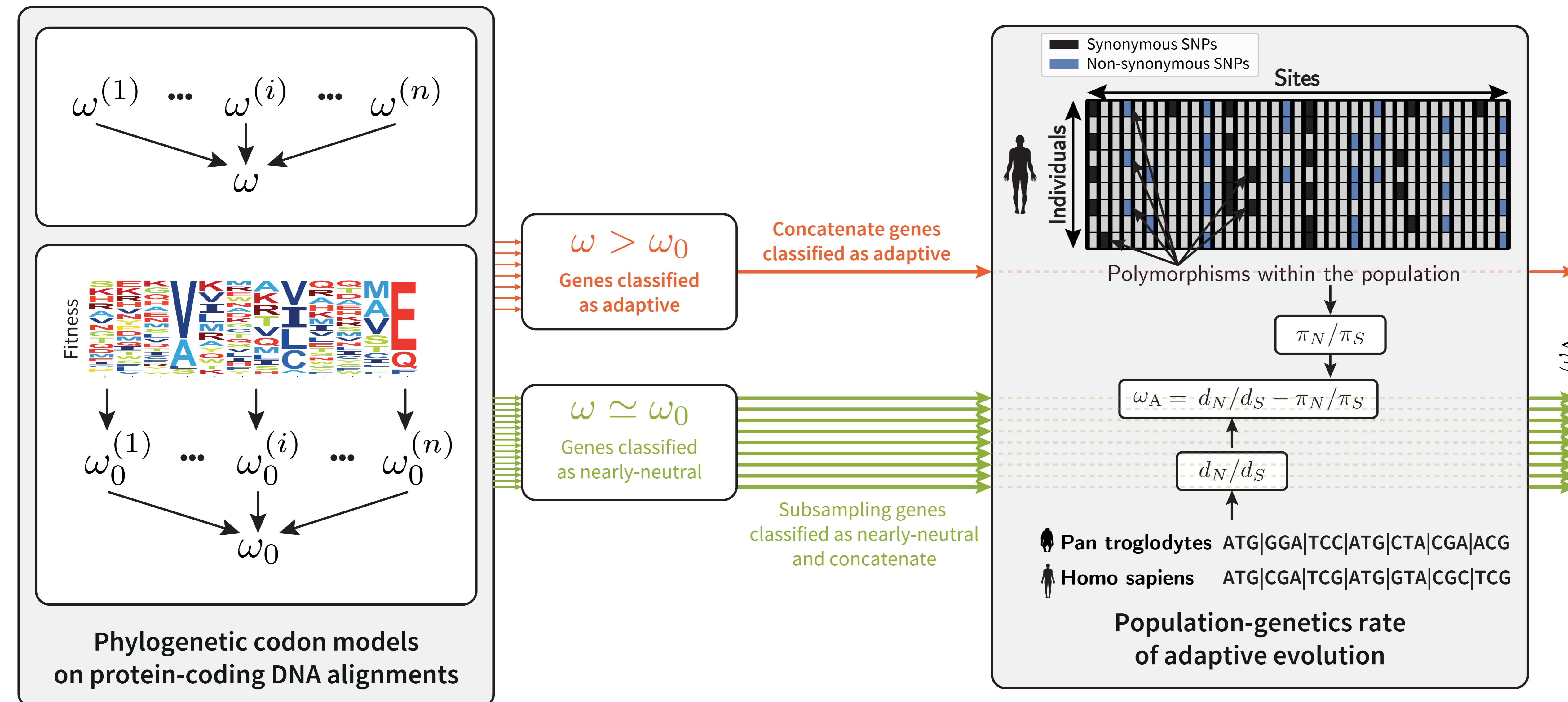


McDonald & Kreitman (1991), Messer & Petrov (2013), Tataru et al. (2017)

Is adaptation at different evolutionary scale comparable?

Adaptation at the phylogenetic scale predicts adaptation in a terminal lineage.

14,509 genes - 87 species

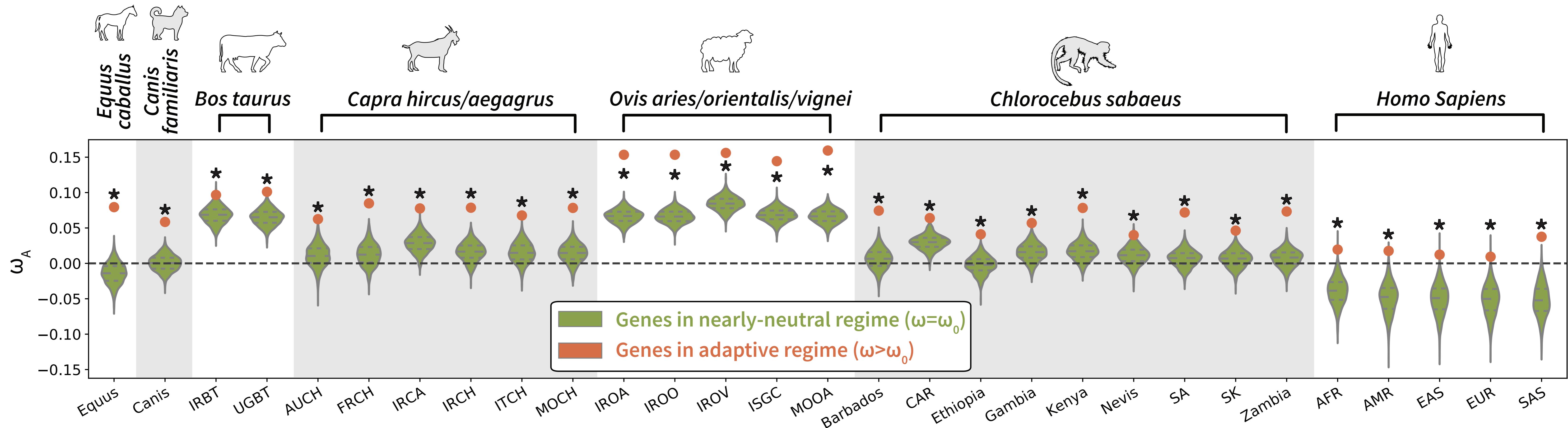


- Genes predicted to be under adaptation at the phylogenetic scale are under adaptation at the population-genetic scale.

Latrille et al. (2023)

Are the results replicable?

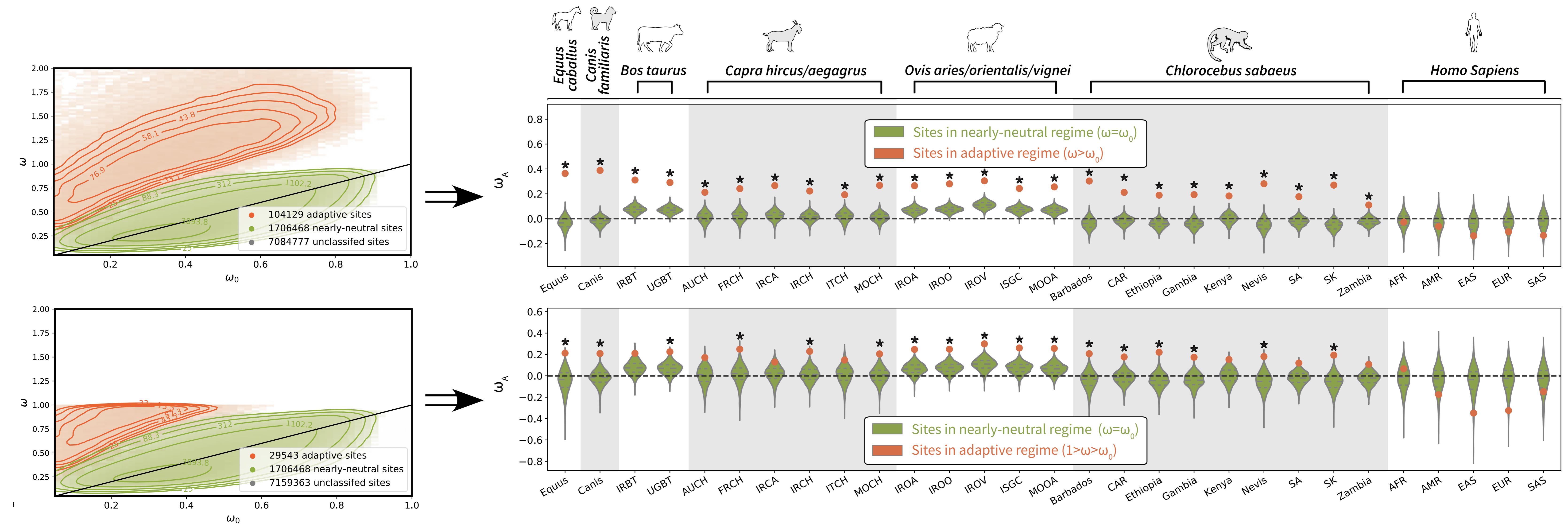
Adaptation at the phylogenetic scale predicts adaptation in terminal lineages.



Latrille et al. (2023)

Are the results generalisable to sites instead of sites?

Sites under adaptation across mammals are under adaptation in terminal lineages.



Latrille et al. (2023)

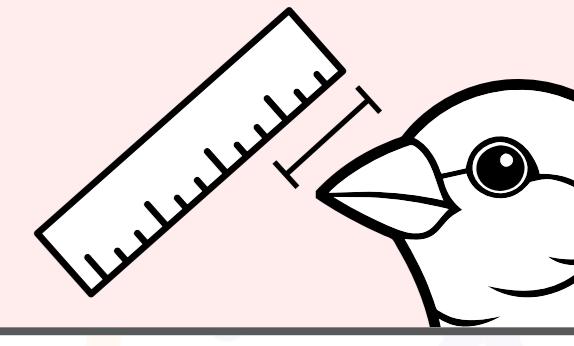
Part II

How to detect adaptation?

Is adaptation predictable across evolutionary scales?

- A stable fitness landscape is a null model of evolution.
- Adaptation as deviation from this null model.
- Adaptation at the phylogenetic scale predicts adaptation in terminal lineages and populations.



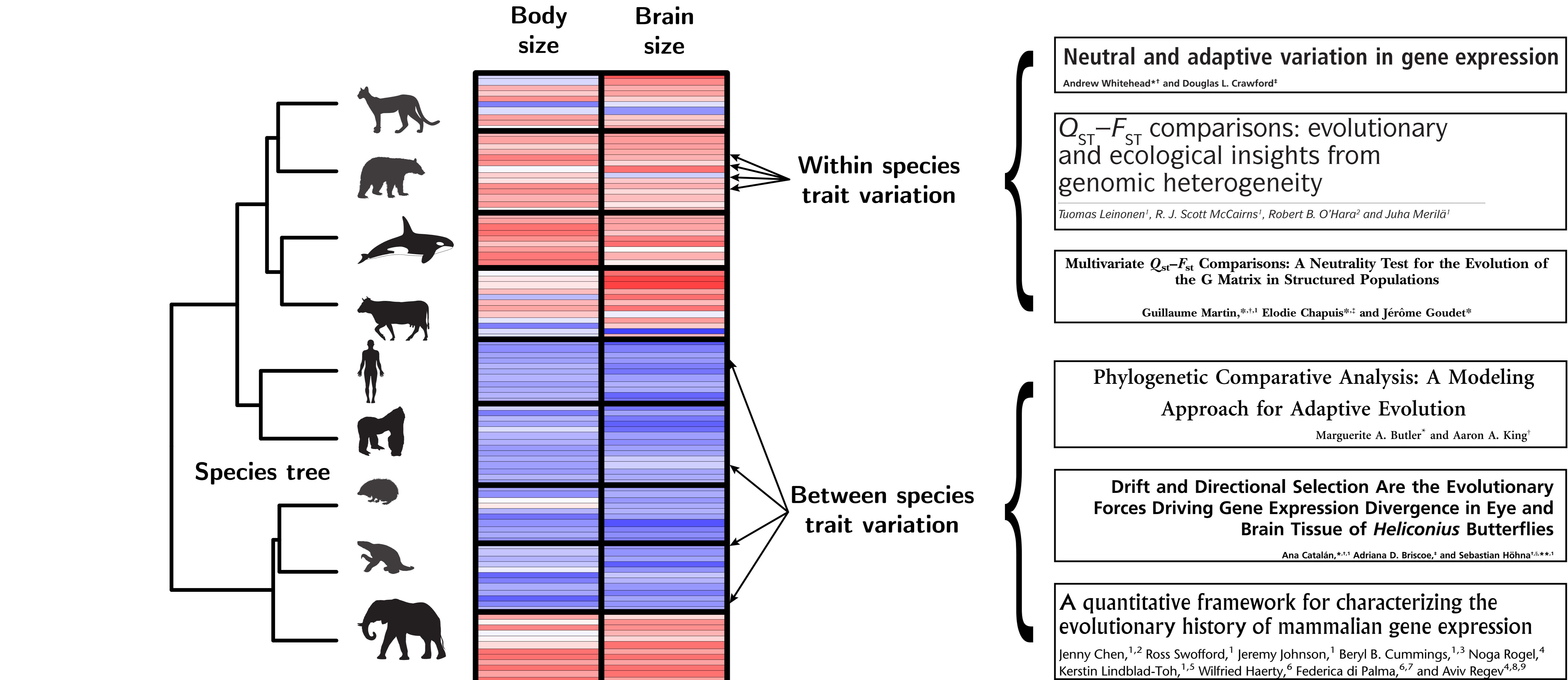


Part III

Can we integrate evolutionary scales
to detect selection on a trait?

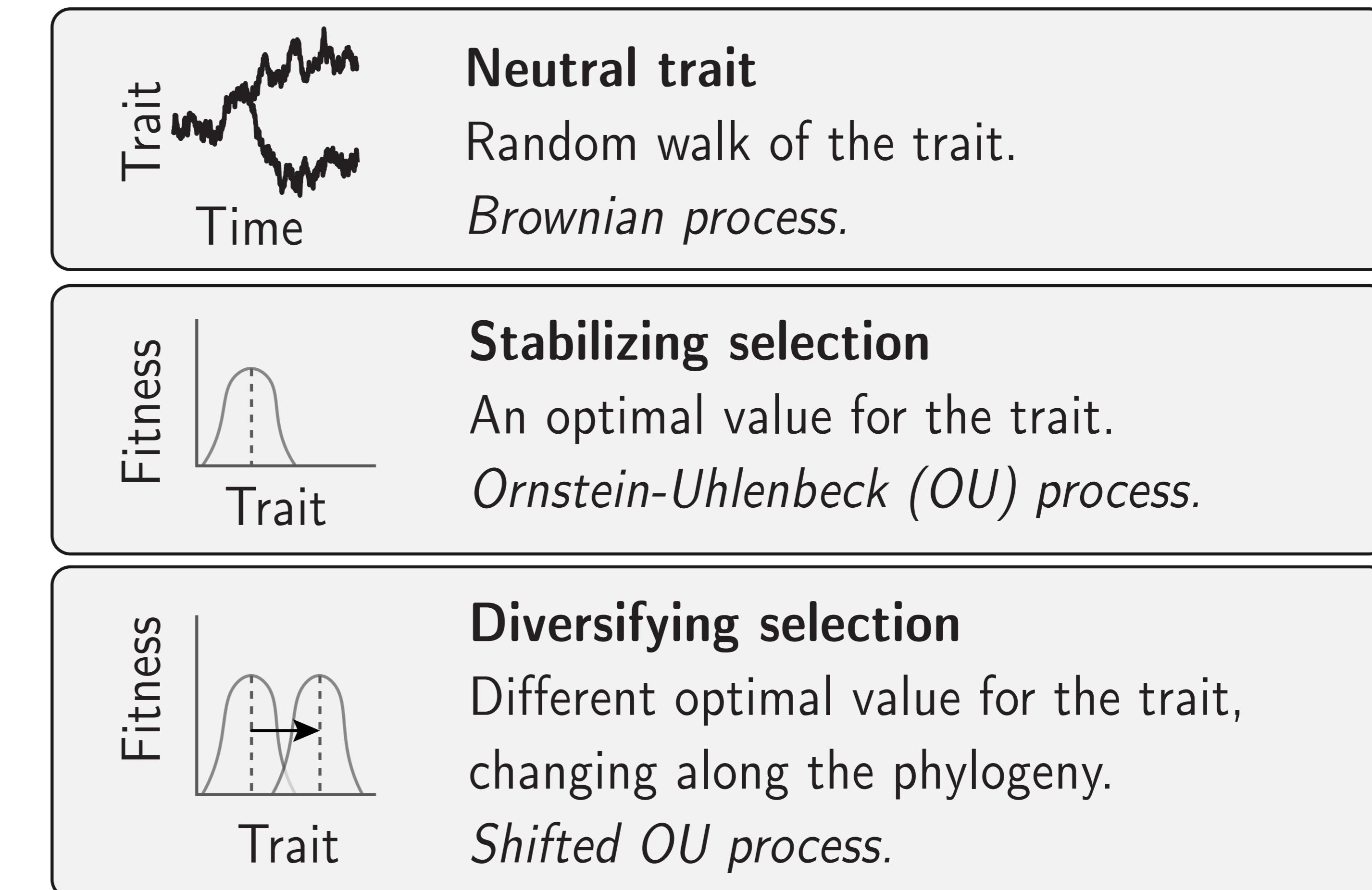
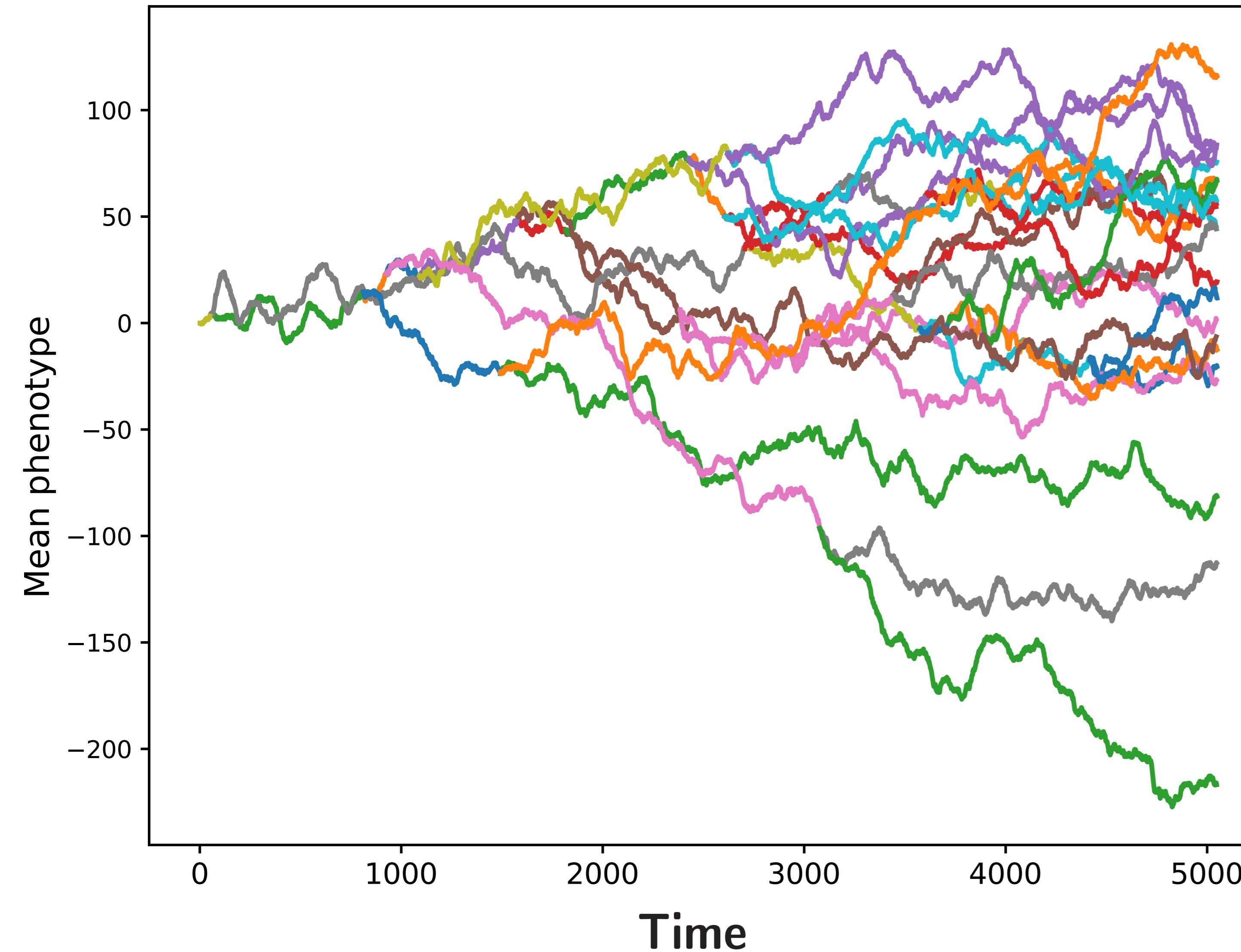
Can we use trait variation between and within species to detect selection?

Usually one or the other, but not together.



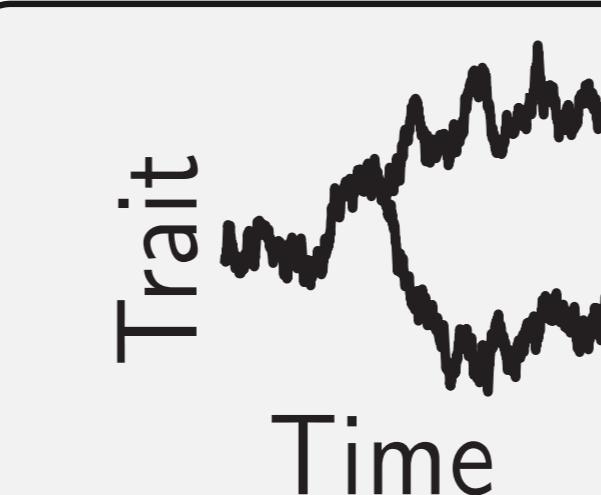
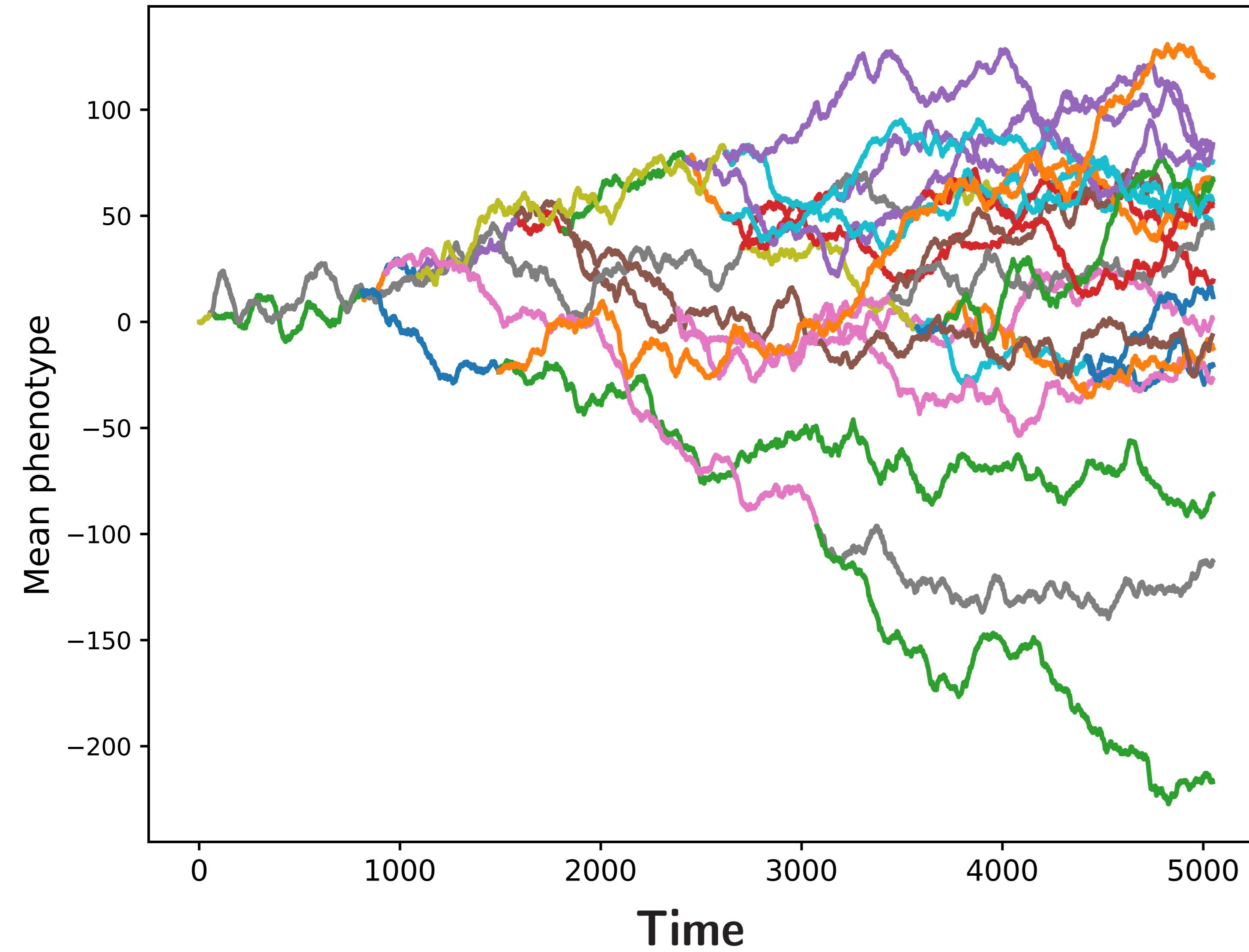
How can we use trait variation between species to detect selection?

Modeling mean trait changes along the phylogeny.



Are these models well behaved?

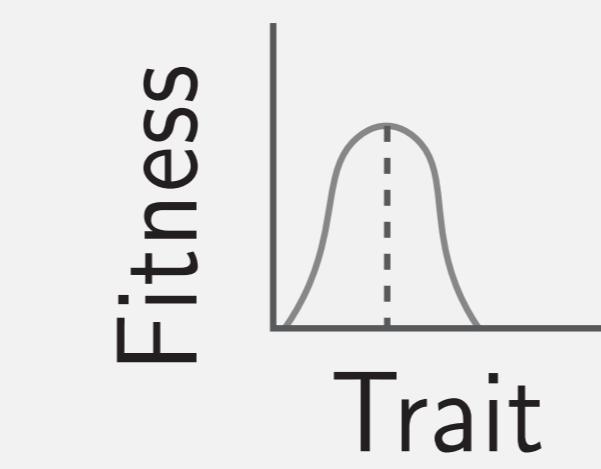
Current models produce both false negatives and false positives.



Neutral trait

Random walk of the trait.

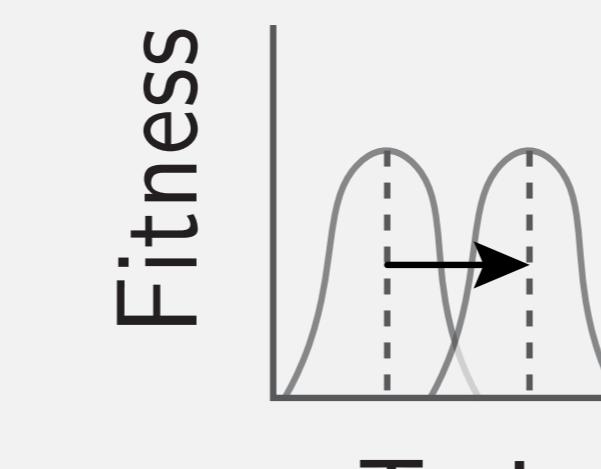
Brownian process.



Stabilizing selection

An optimal value for the trait.

Ornstein-Uhlenbeck (OU) process.



Diversifying selection

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

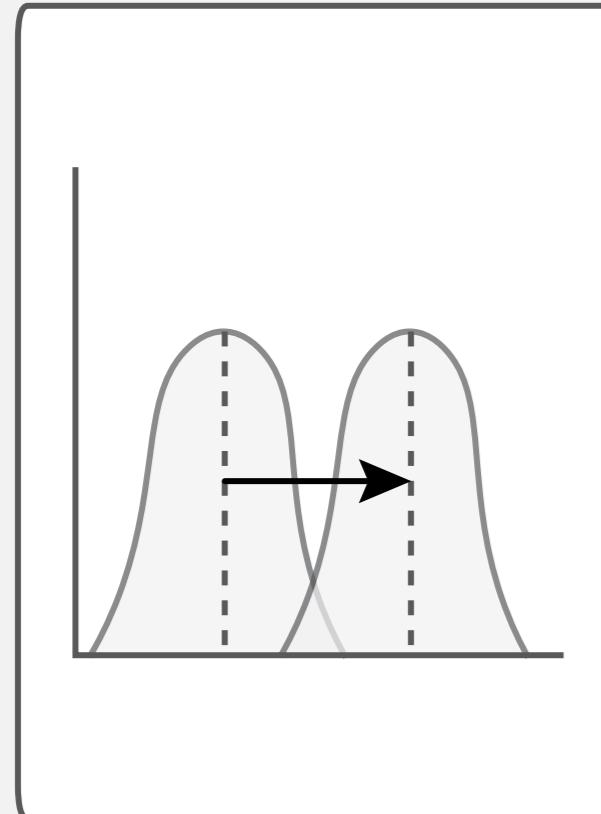
Shifted OU process.

- **OU process can be favored for 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)

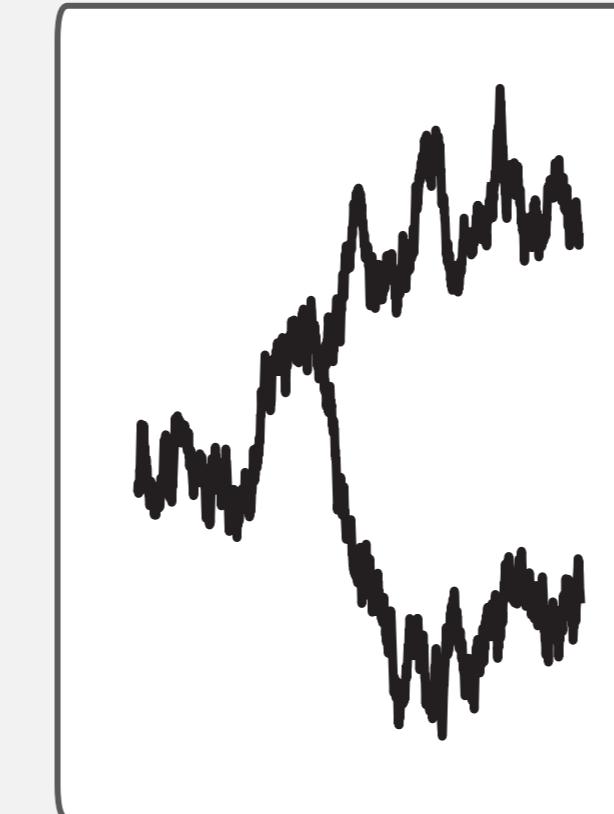
Can we take inspiration from other methods?

Several methods that can be adapted to our problem.



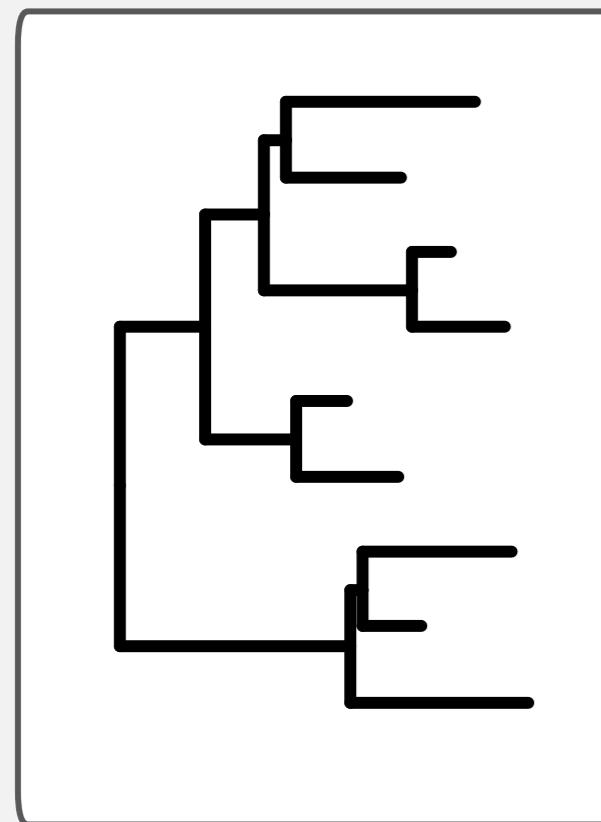
**Quantitative-genetics
across populations.**

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



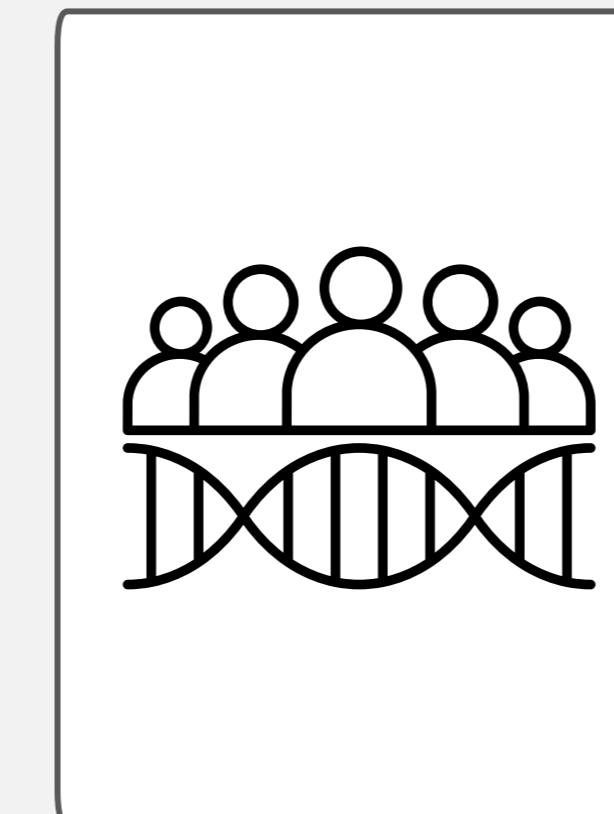
**Phylogenetic comparative
method.**

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



**Phylogenetic DNA
evolution.**

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

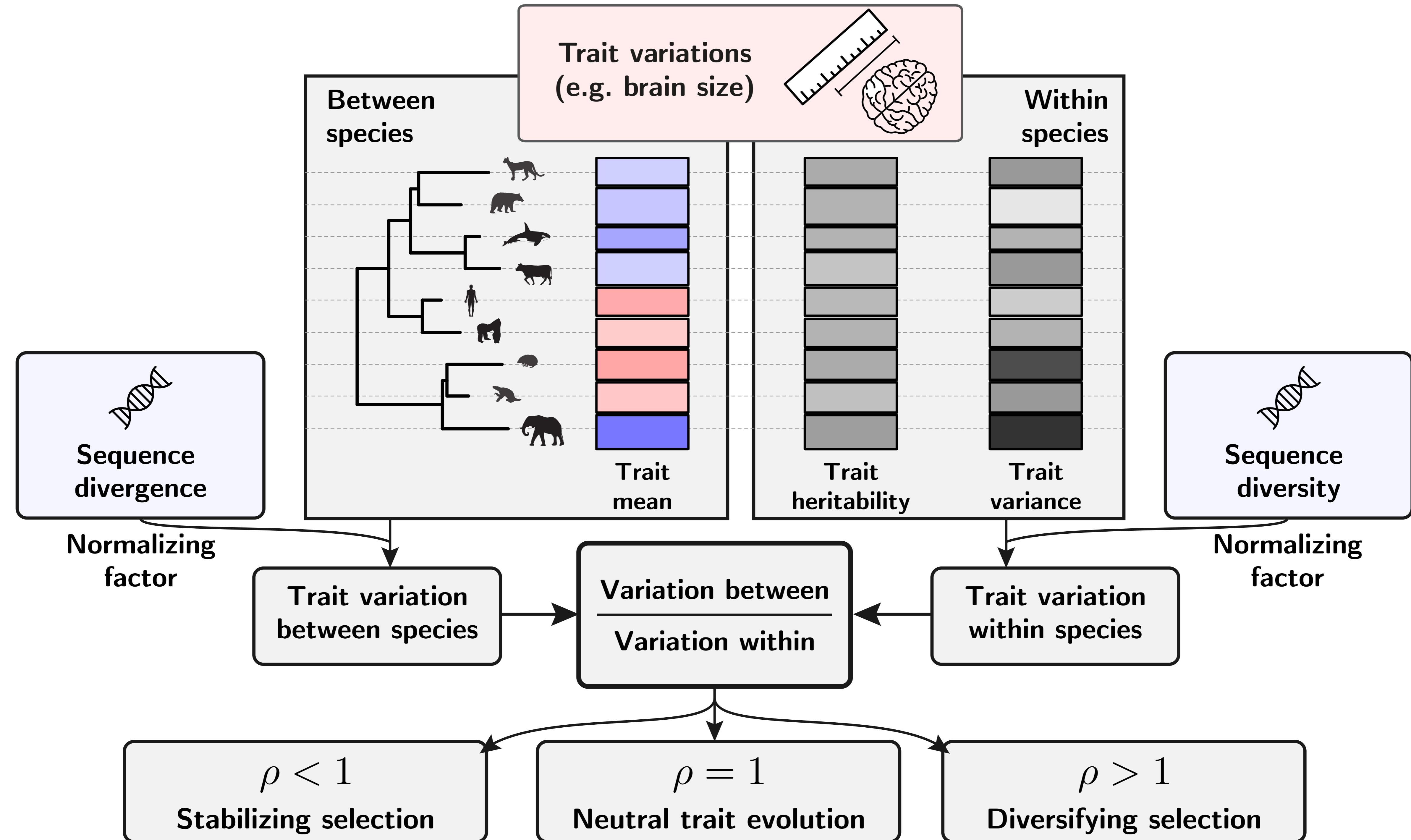


**Contrast polymorphism
& divergence.**

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

Can we integrate between and within species variations?

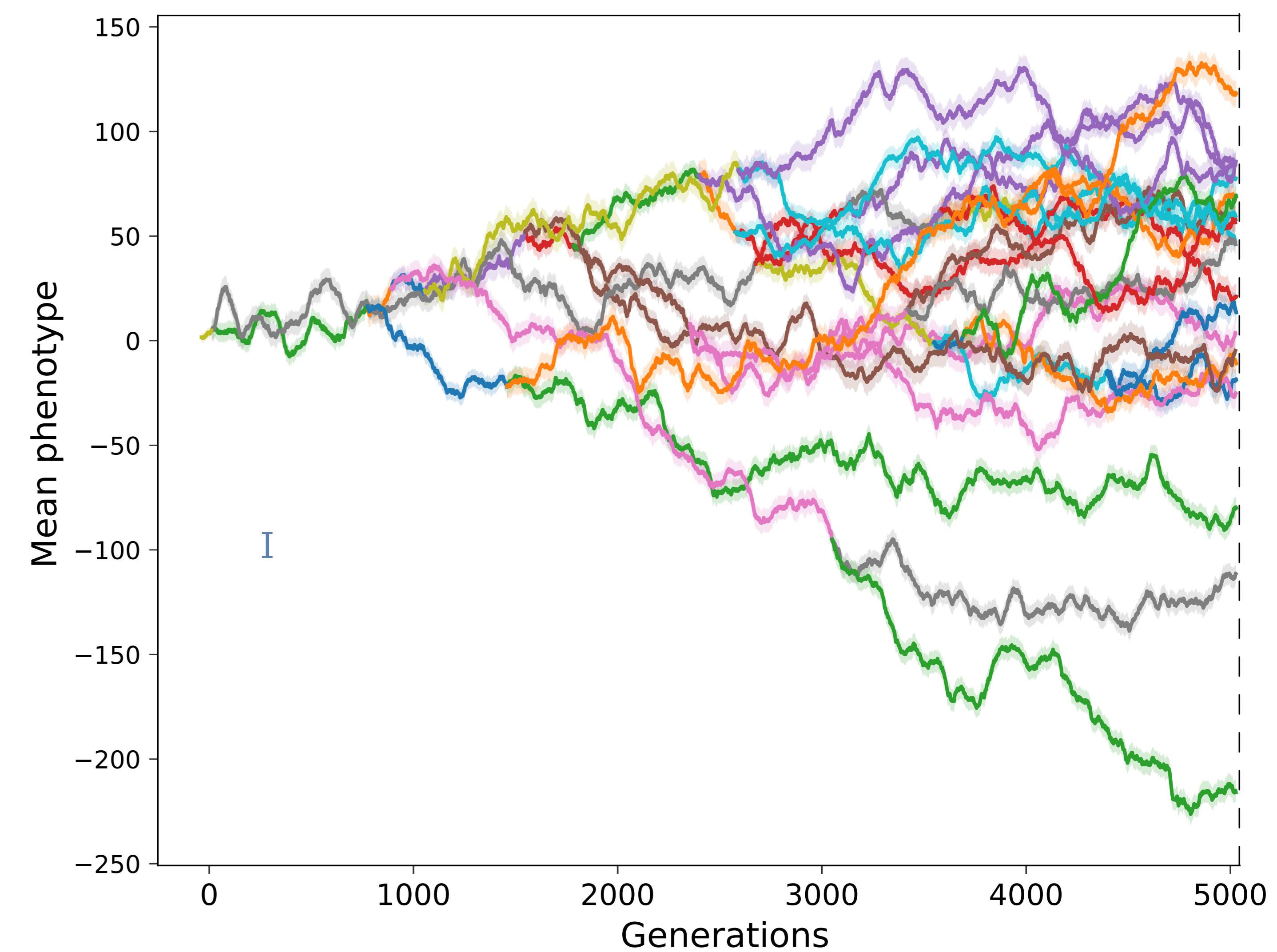
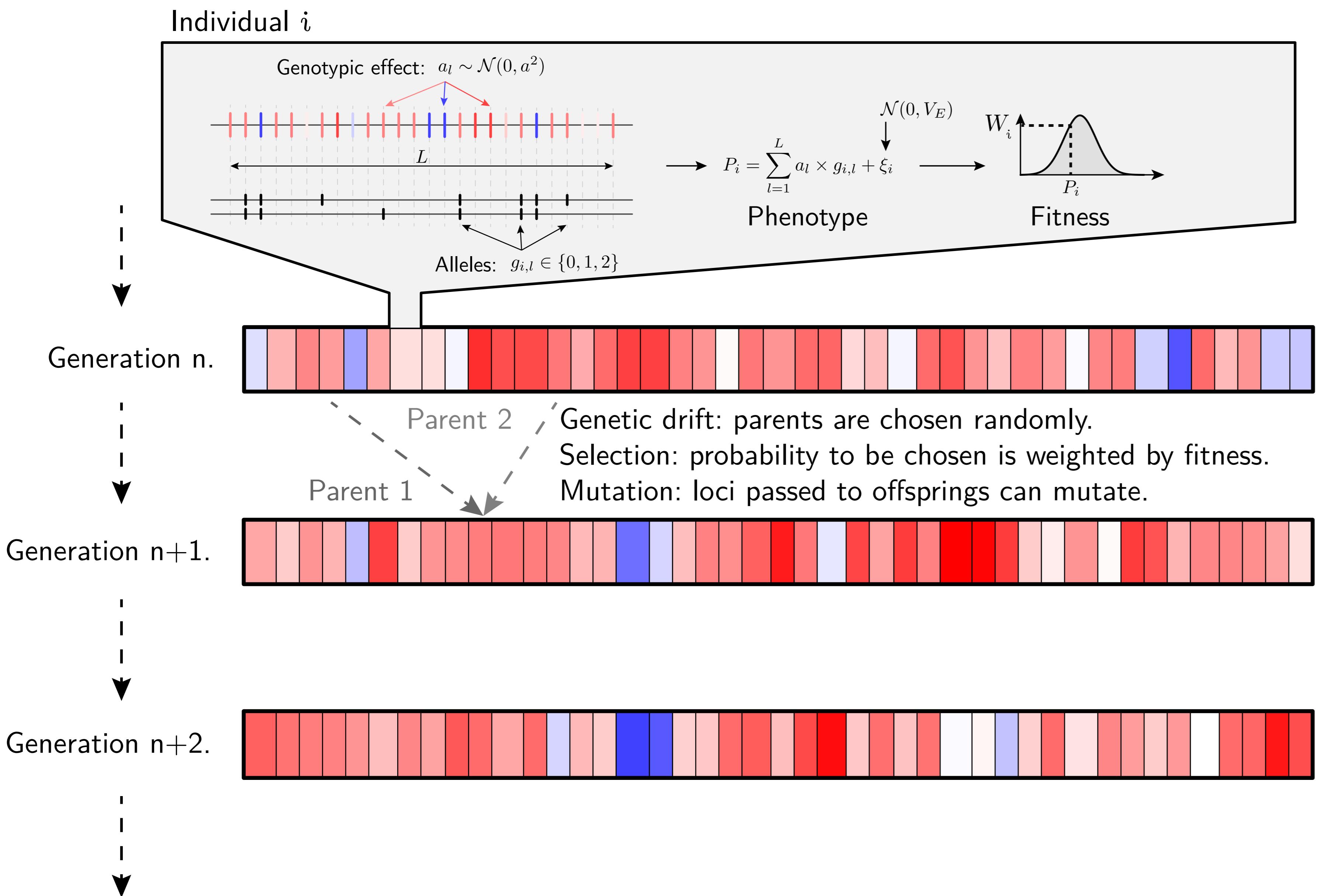
Ratio of between- over within-species trait variations (normalized by DNA changes)



Latrille et al. (2024)

Can we test our estimate on simulated data? (1/3)

Wright-fisher simulator across the phylogeny.

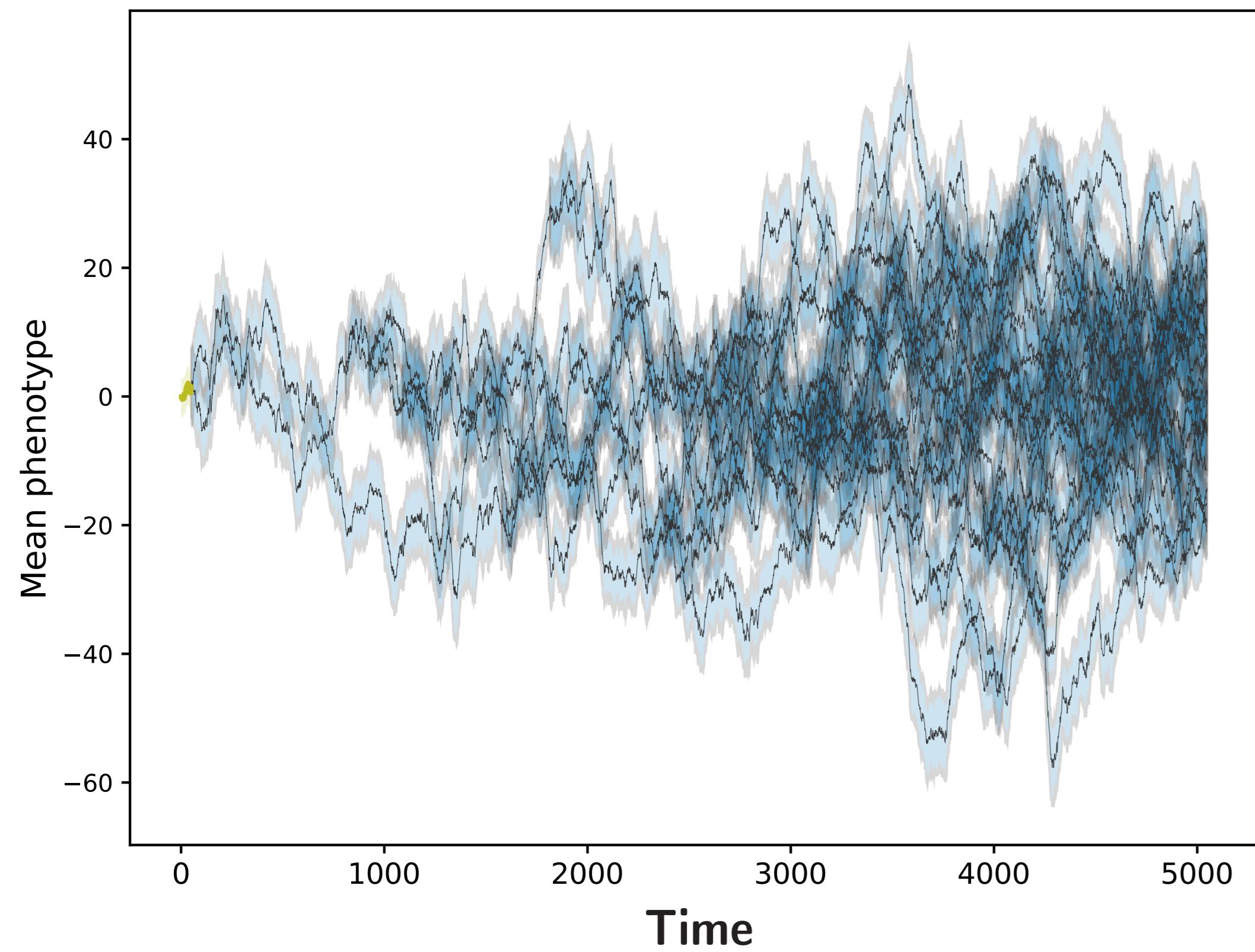


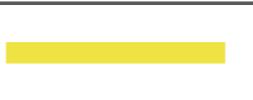
Can we test our estimate on simulated data? (2/3)

Simulator across the phylogeny under different scenarios.

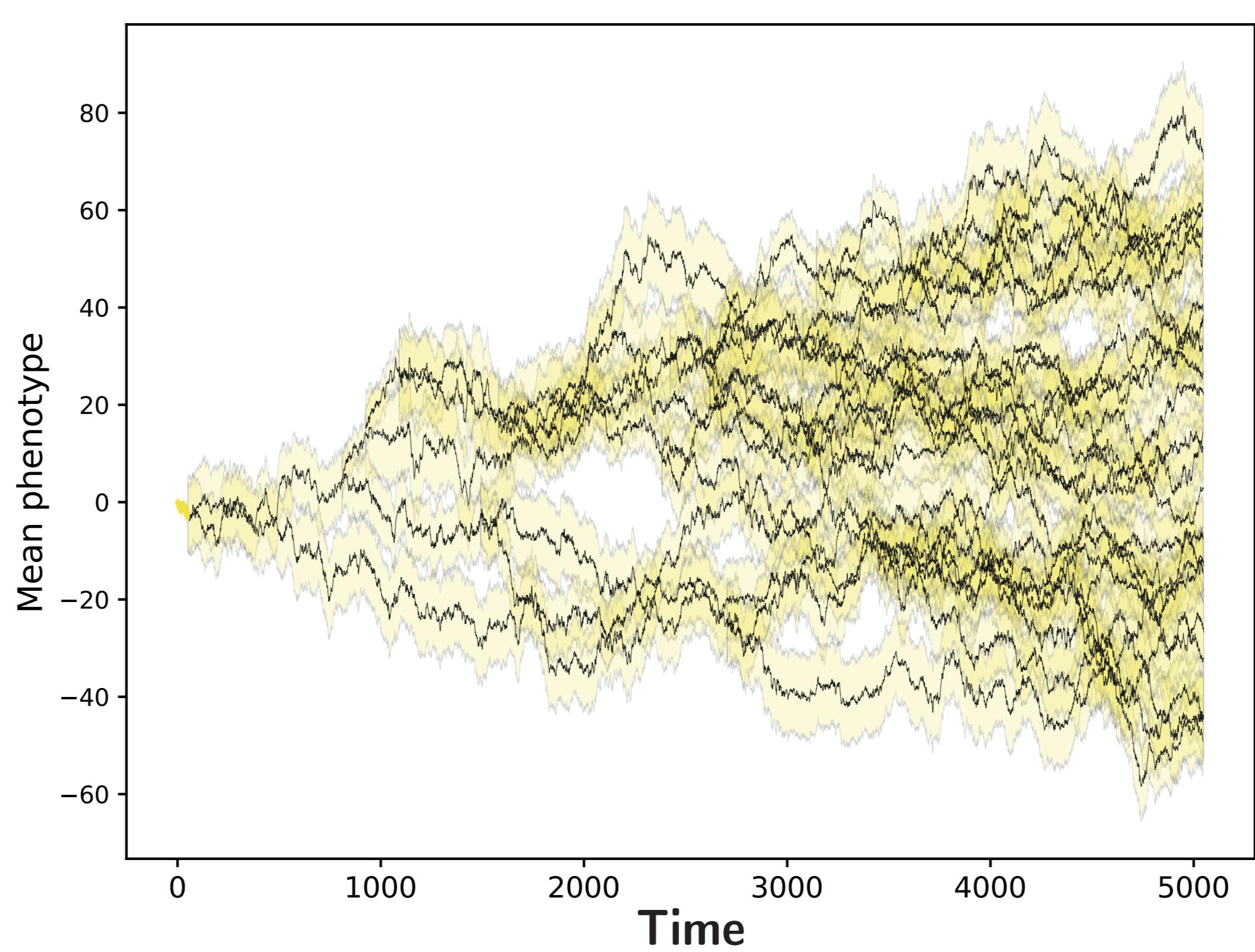
 **Stabilizing selection**

An optimal value for the trait.



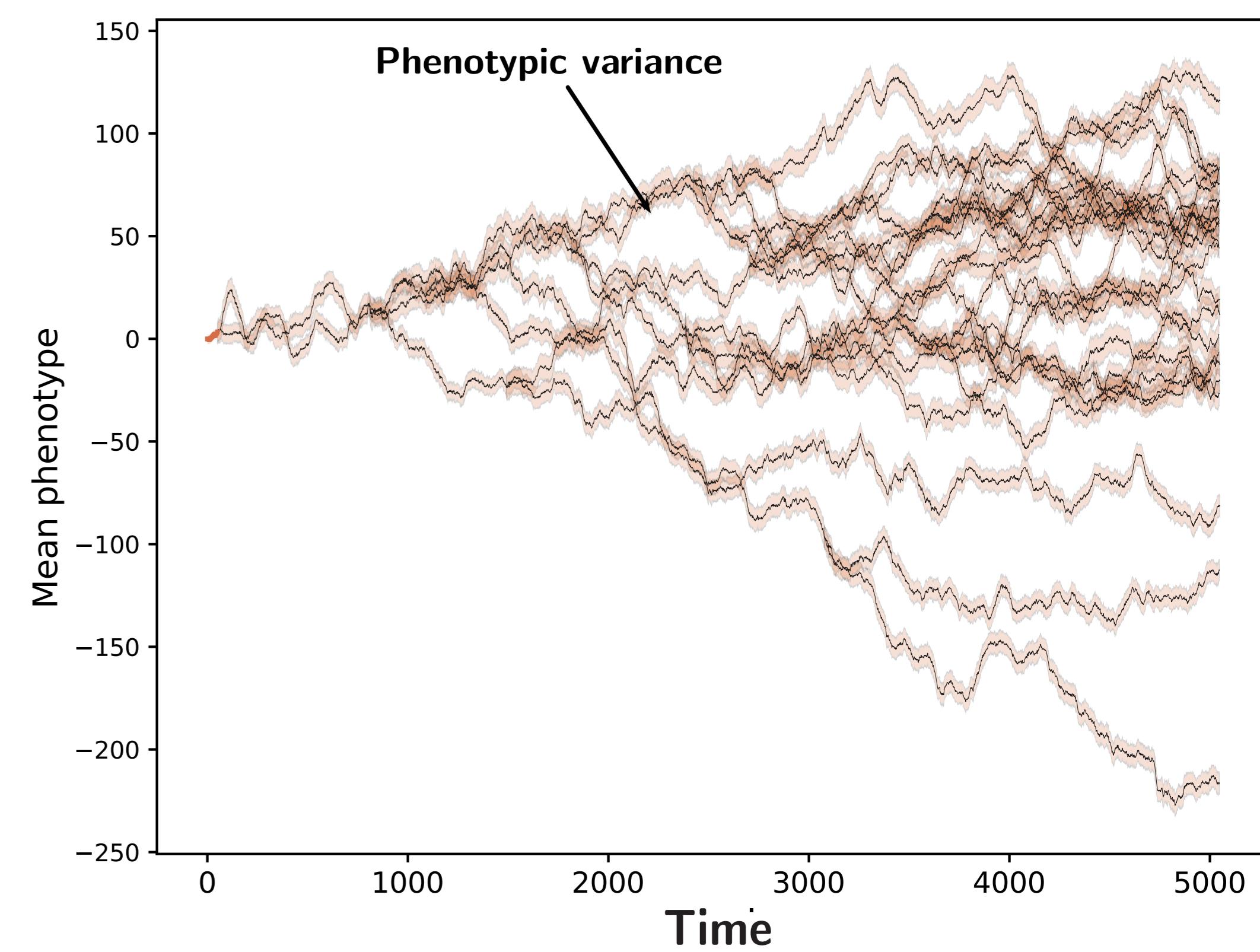
 **Neutral trait**

No fitness function.



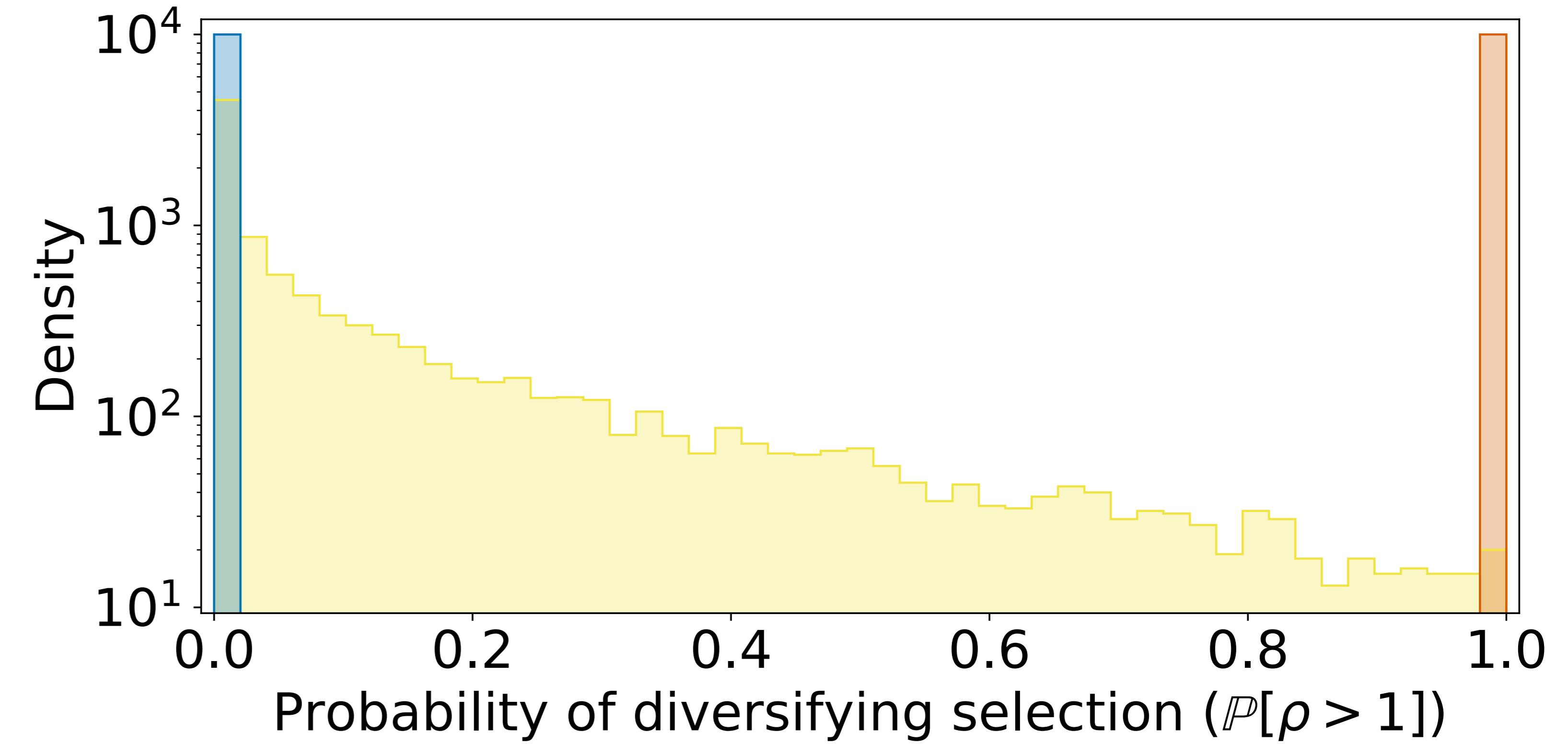
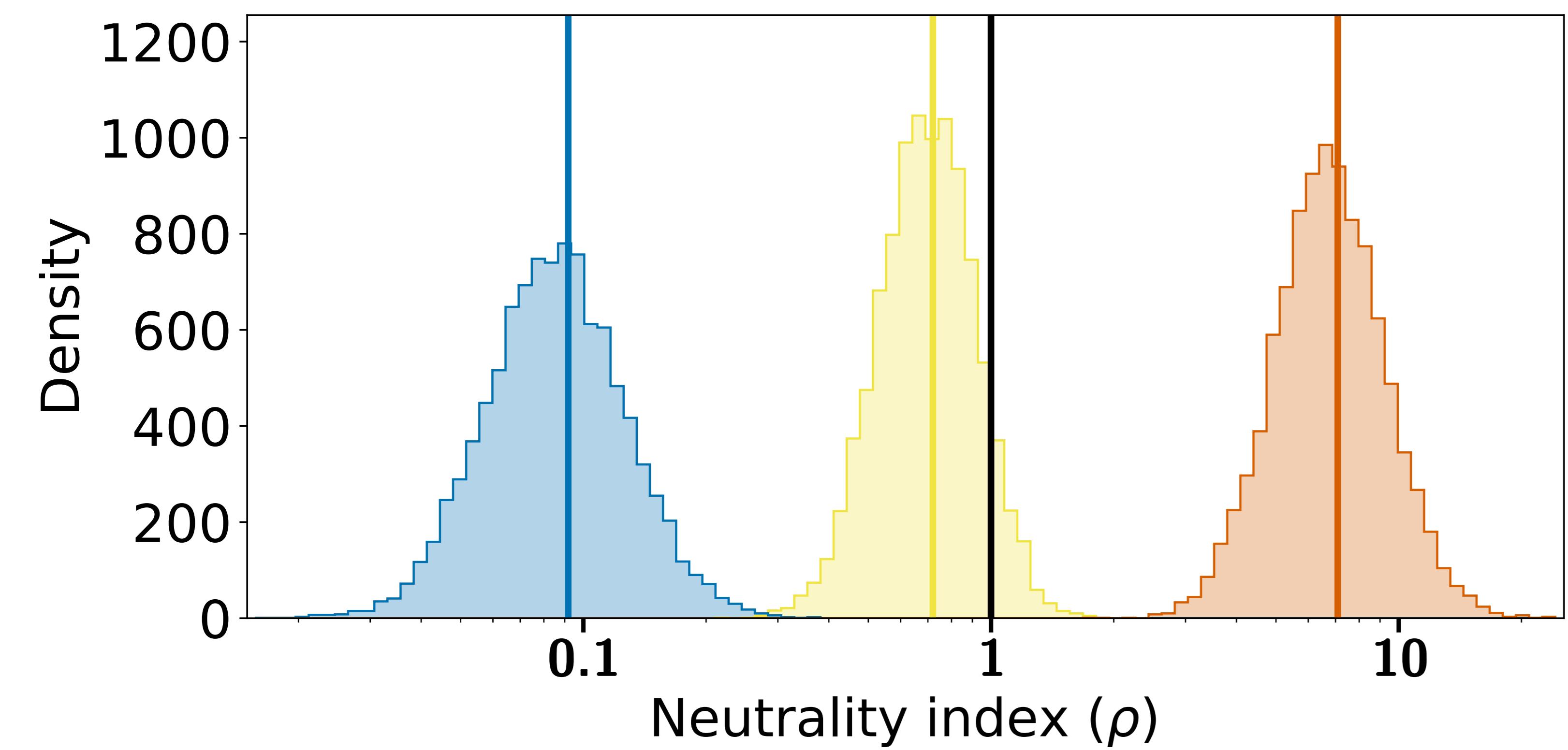
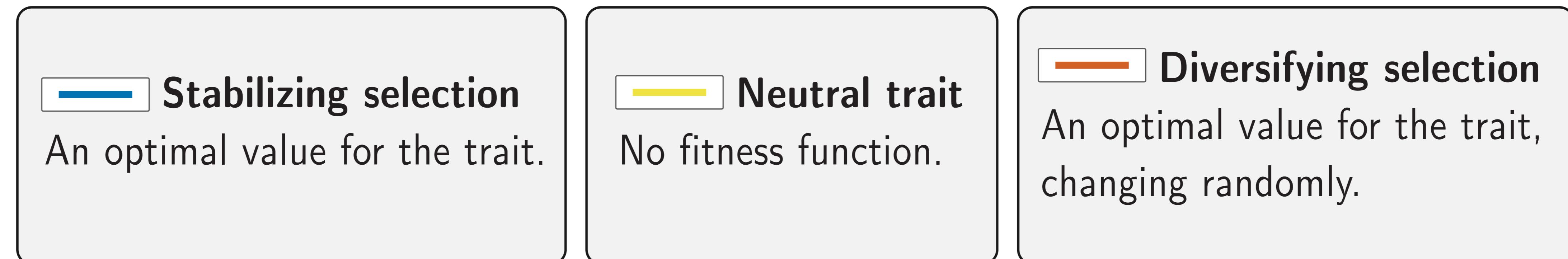
 **Diversifying selection**

An optimal value for the trait, changing randomly.



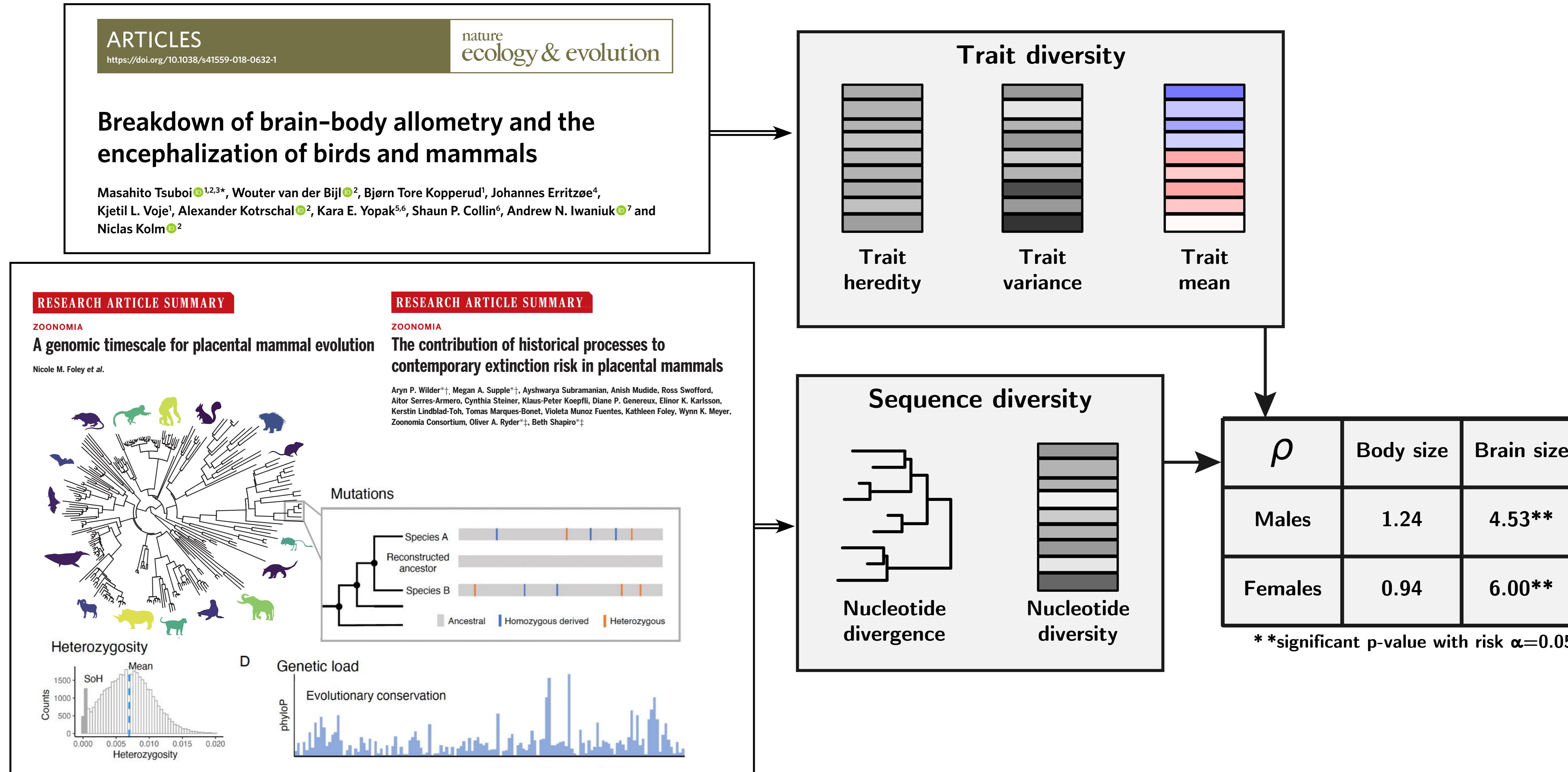
Can we test our estimate on simulated data? (3/3)

Test for diversifying selection, false positive for stabilizing selection.



Can we test our estimate on empirical data?

Brain size is evolving under diversifying selection, not body size.



Latrille et al. (2024)

Part III

Can we integrate evolutionary scales to detect selection on a trait?

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

JOURNAL OF
Evolutionary Biology

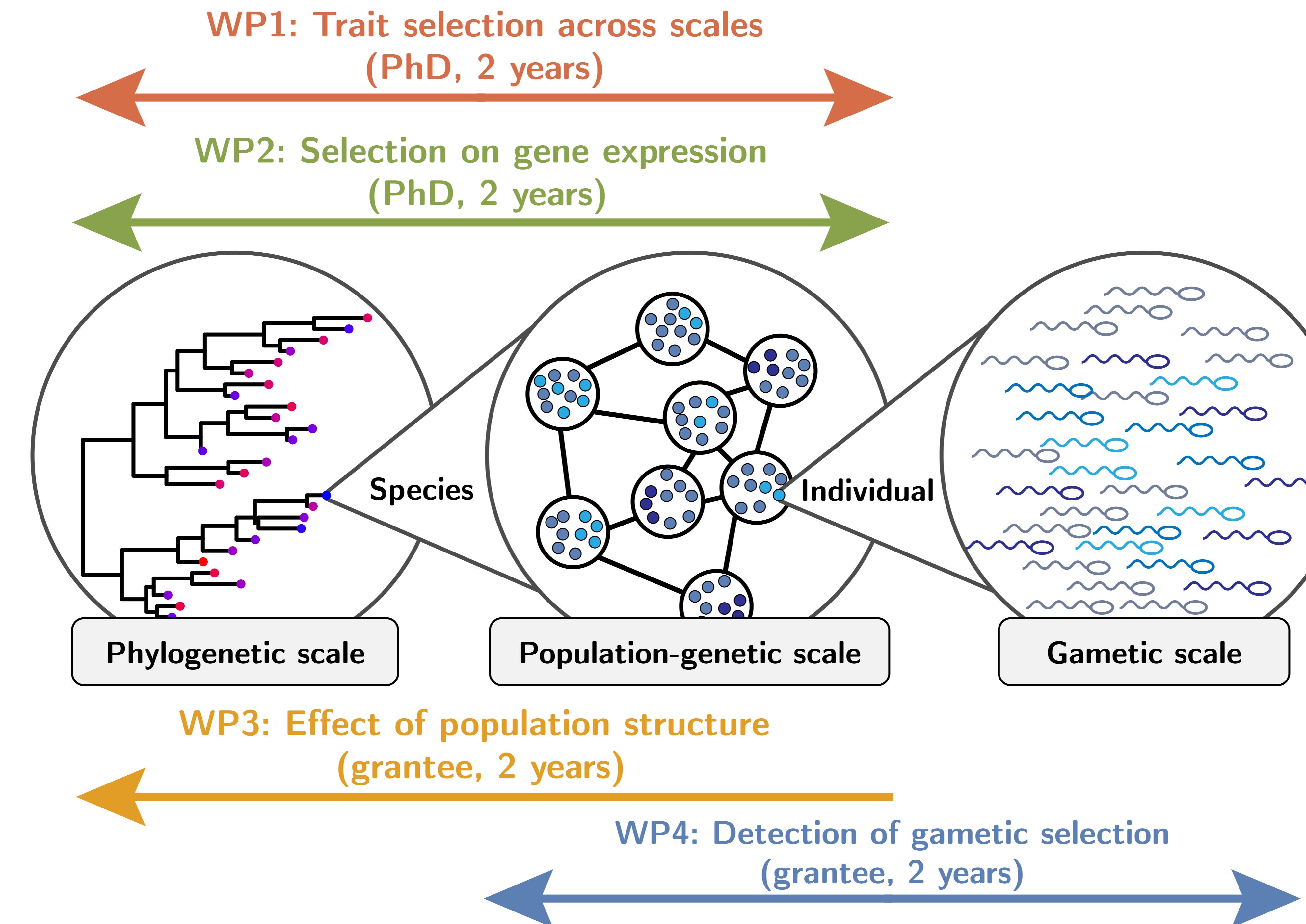


Detecting diversifying selection for a trait from within
and between-species genotypes and phenotypes

T. Latrille¹ M. Bastian², T. Gaboriau¹, N. Salamin¹

What is my Ambizione research proposal?

Predicting selection on traits and sequences



Thank you for your attention

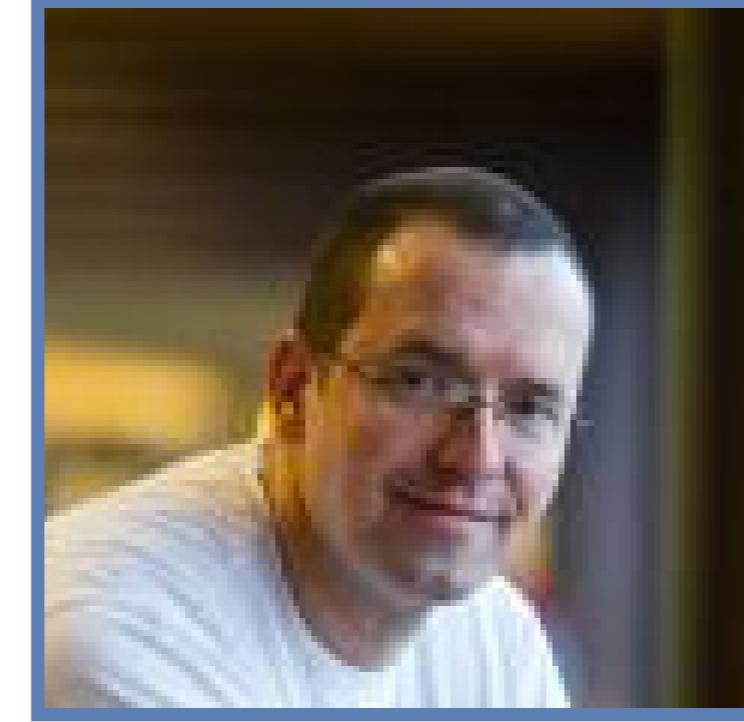
They did the work



Nicolas
Lartillot



Mélodie
Bastian



Nicolas
Rodrigue



Nicolas Salamin Group

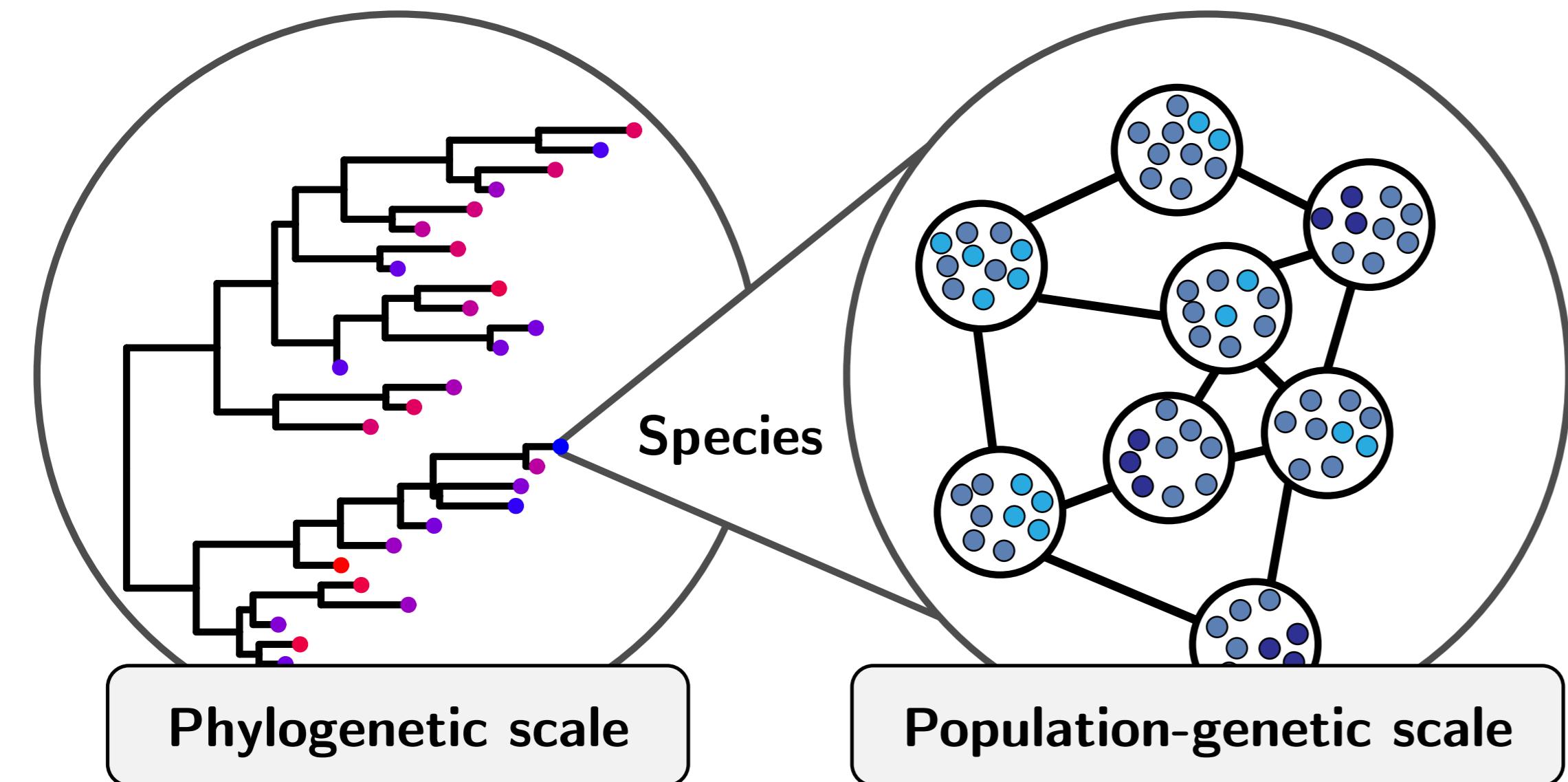


Adapted from PhD thesis (2020)

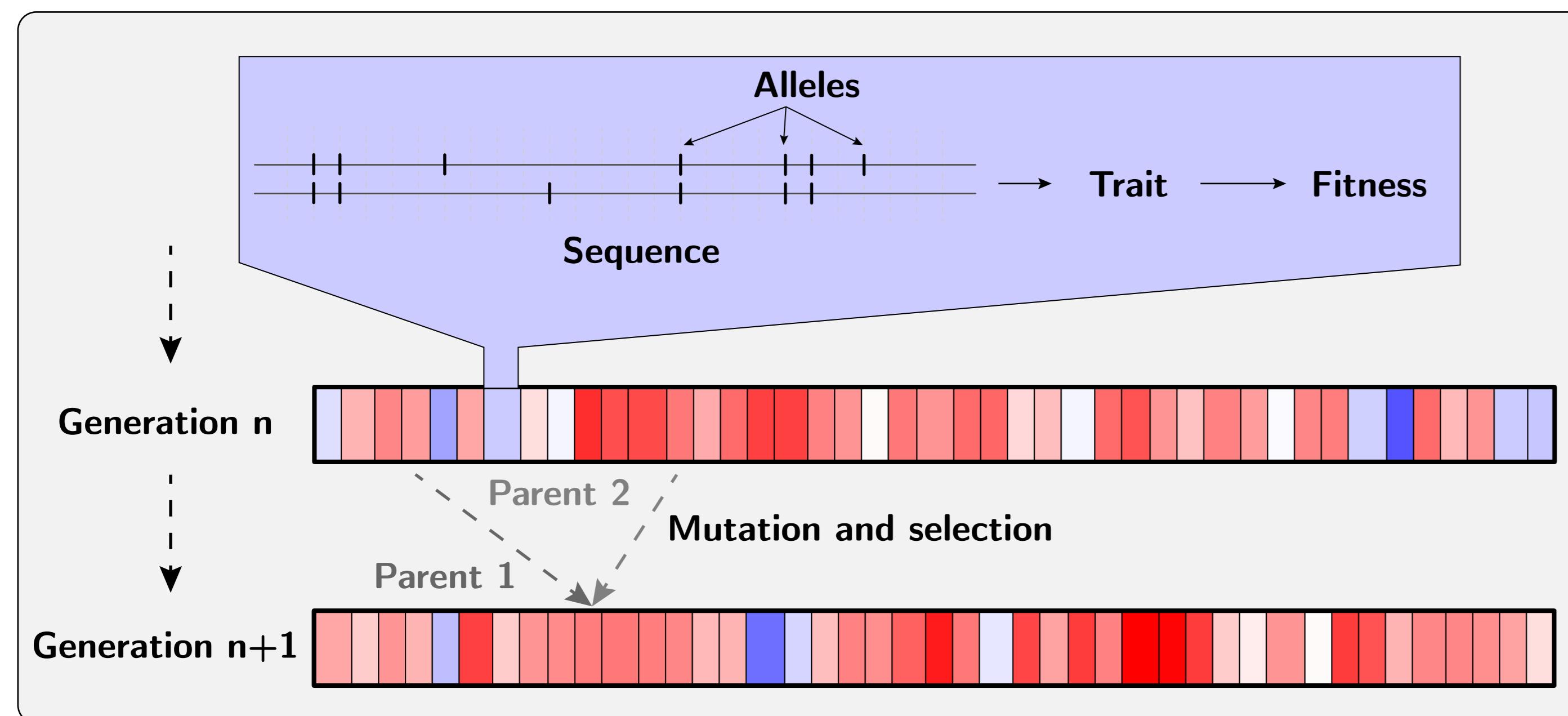
WP1: Trait selection across scales



Are signatures of diversifying selection congruent across scales?



Simulator



Rationale:

- Diversifying selection can be detected across species but also across populations within the same species.
- Whether selection at a given scale can be predictive of selection at another scale has not been tested.

Tasks & challenges:

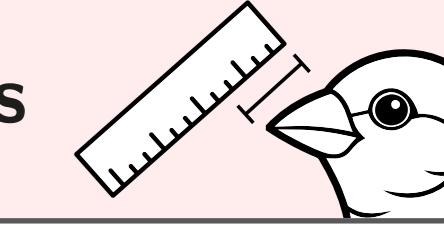
- Need a simulator which jointly works across populations as well as across species.
- Compare how different methods can detect diversifying selection.

Outcomes:

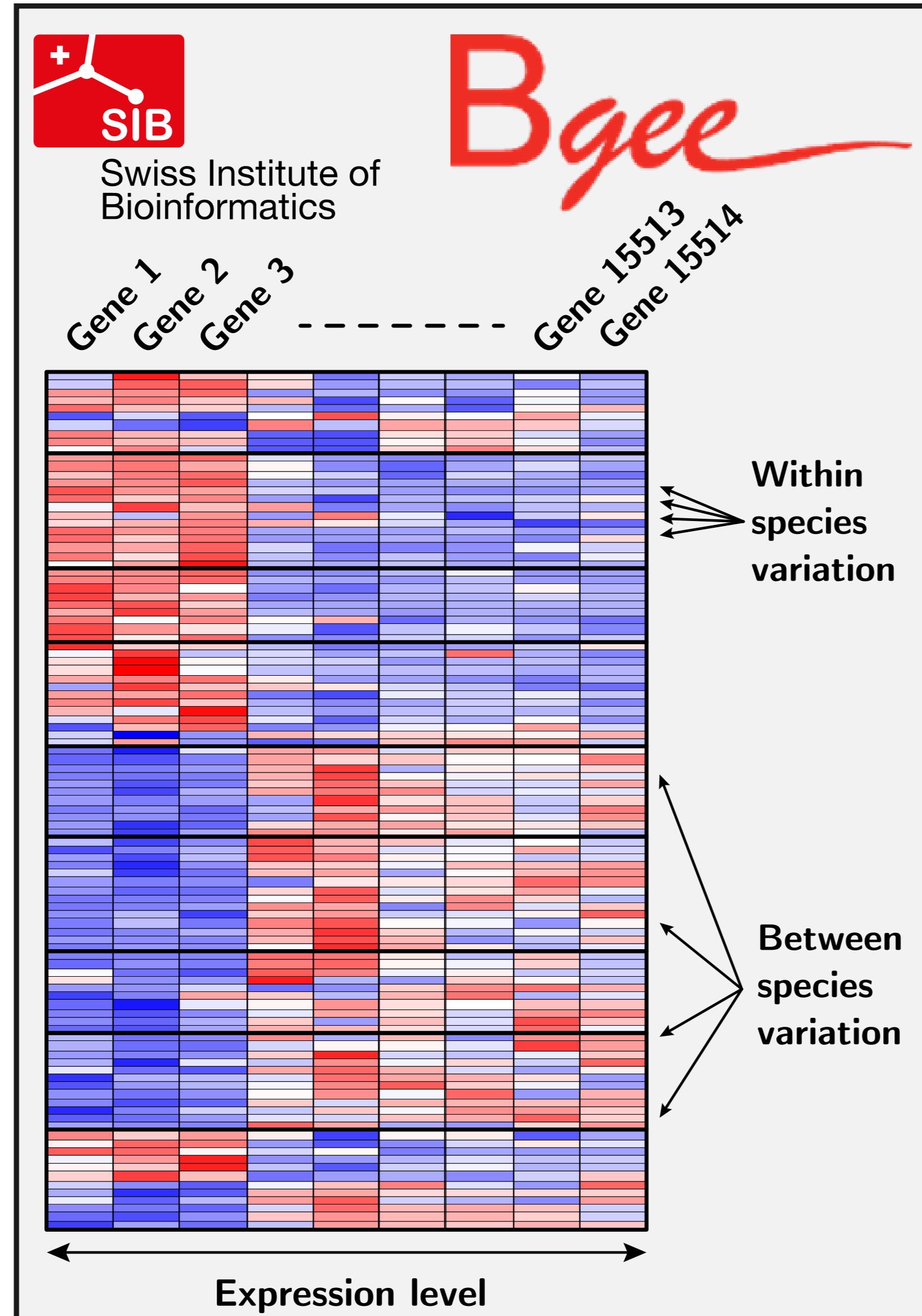
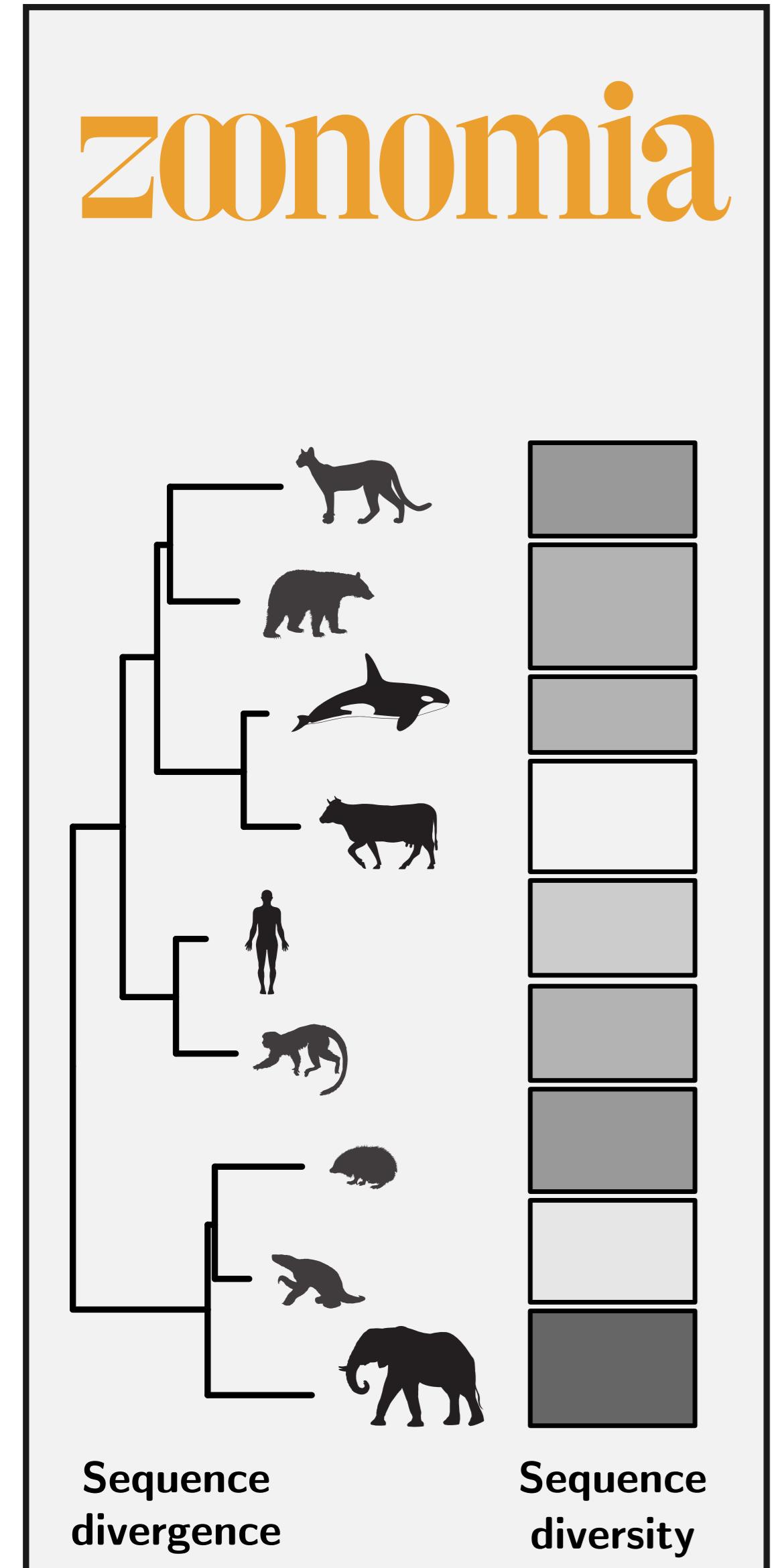
- Quantifying the accuracy and power of methods to detect diversifying selection at different scales.

WP2: Selection on gene expression

Traits



Is gene expression evolving under purifying or diversifying selection?



JOURNAL OF
Evolutionary Biology

..eseb
European Society for Evolutionary Biology

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

T. Latrille¹ , M. Bastian², T. Gaboriau¹, N. Salamin¹

Rationale:

- We can detect diversifying selection using between and within species variation in gene expression level.
- Preliminary analyses concluded there is enough signal in curated mammalian databases.

Tasks & challenges:

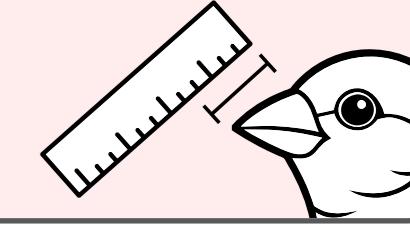
- Develop reliable pipelines and process all genes while applying different methods.

Outcomes:

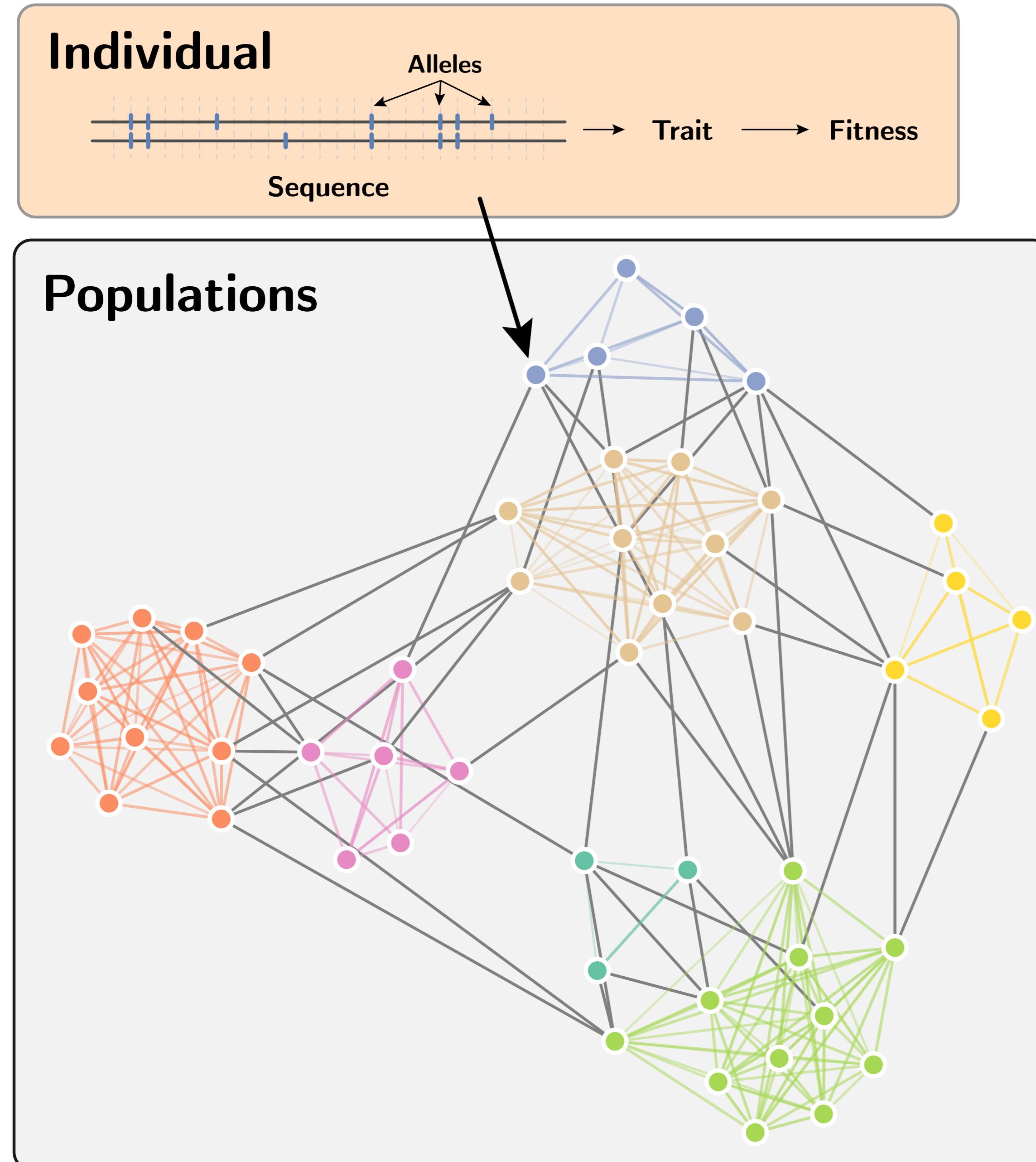
- Quantifying the extent of diversifying selection on gene expression.
- Relating this to biological function and sequence evolution.

WP3: Effect of population structure

Traits



Does population structure influence tests of diversifying selection?



Rationale:

- Methods to detect selection at the phylogenetic scale do not account for population structure.
- Structure can be modeled as a graph.

Tasks & challenges:

- Simulate trait evolution on a graph-structured population and quantify trait evolution across species.

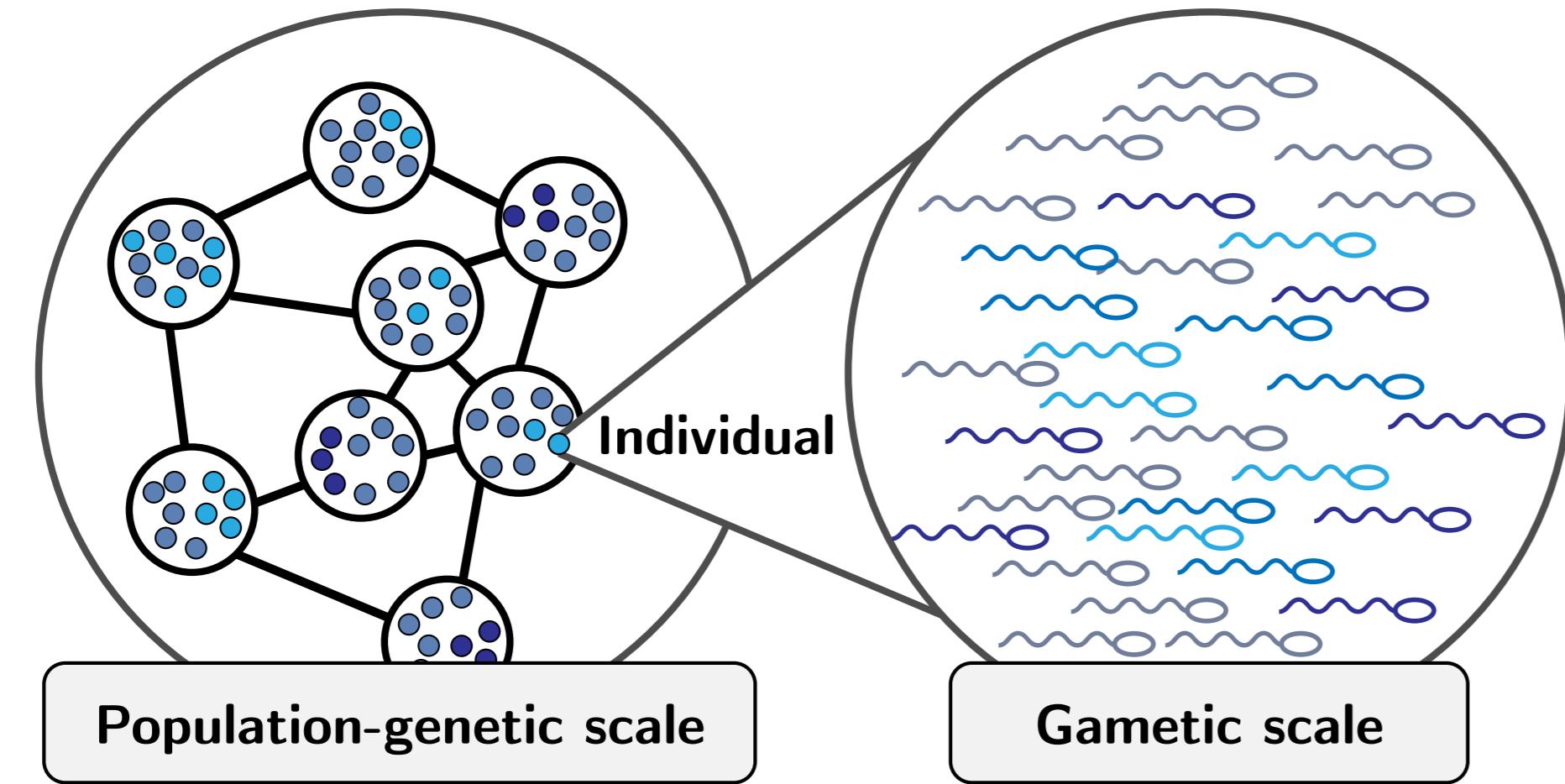
Outcomes:

- Validating/invalidating the methods to detect selection at the phylogenetic scale.

WP4: Detection of gametic selection

Sequences 

Is there selection for gametes in addition to selection for individuals?



Rational:

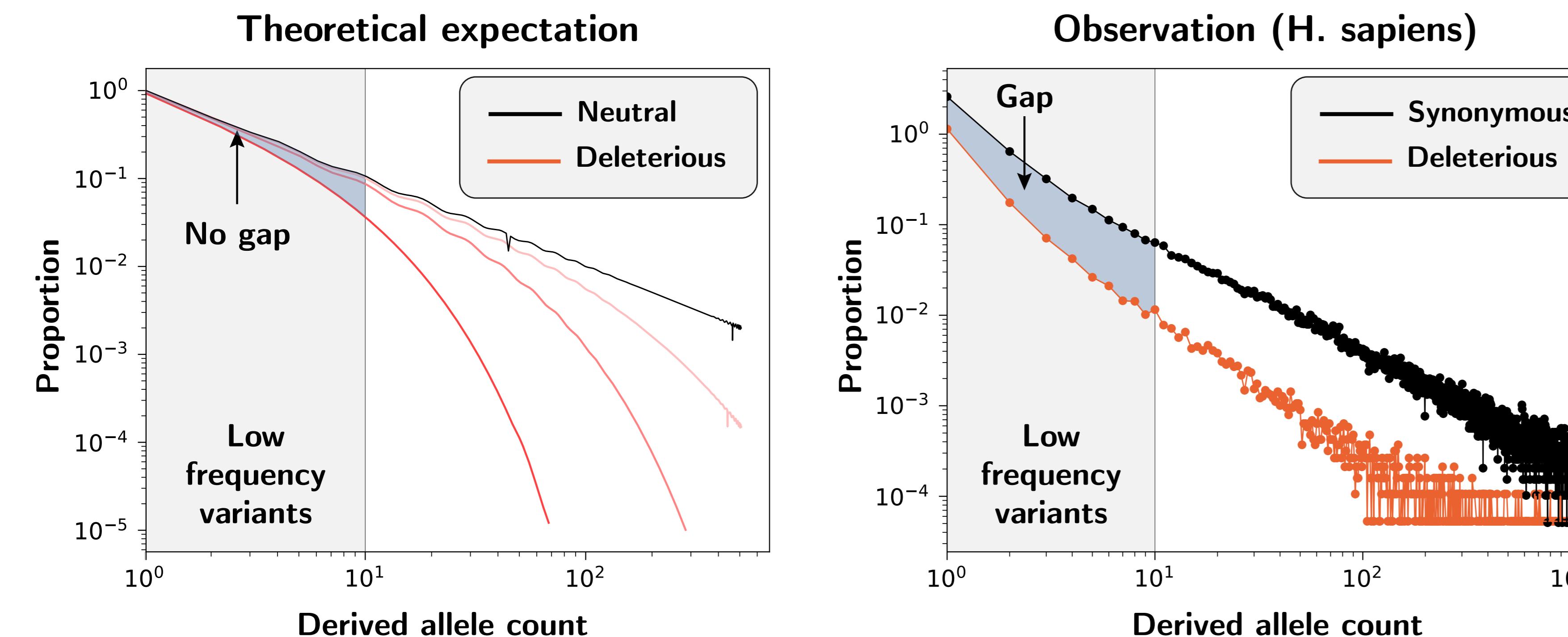
- Selection during the gametic stage is largely ignored.
- Preliminary analyses show gap in empirical data, not accounted for by population-genetic formalism.

Tasks & challenges:

- Analyze sequencing data from the UK BioBank and GnomAD database (141,456 individuals).
- Theoretical framework still to be developed.

Outcomes:

- Quantifying gametic selection in humans.
- Locating the genes associated with gametic selection, application to human infertility.



Why applying for an Ambizione?

Makes sense based on my career plan

- Journey from being supervised (PhD) to being independent (Post-docs).
- Mentoring PhD student as a next step in this journey.
- The DEE as a base camp:
 - DEE is interdisciplinary, as the project is.
 - Collaborations inside and outside.
 - Counsel from several PI on PhD supervision.



Adapted from Latrille (PhD thesis, 2020)