A phylogenetic mutation-selection model predicts fitness effects of mutations in extant mammals

BEvAS 2023 - EPFL

April 18, 2023

T. Latrille[1,2], N. Rodrigue[3], N. Lartillot[2], J. Joseph[2], D. A. Hartasánchez[2], N. Salamin[1]

[1]Université de Lausanne (Switzerland)

[2]Université de Lyon (France)

[3]Carlton University, Ottawa (Canada)

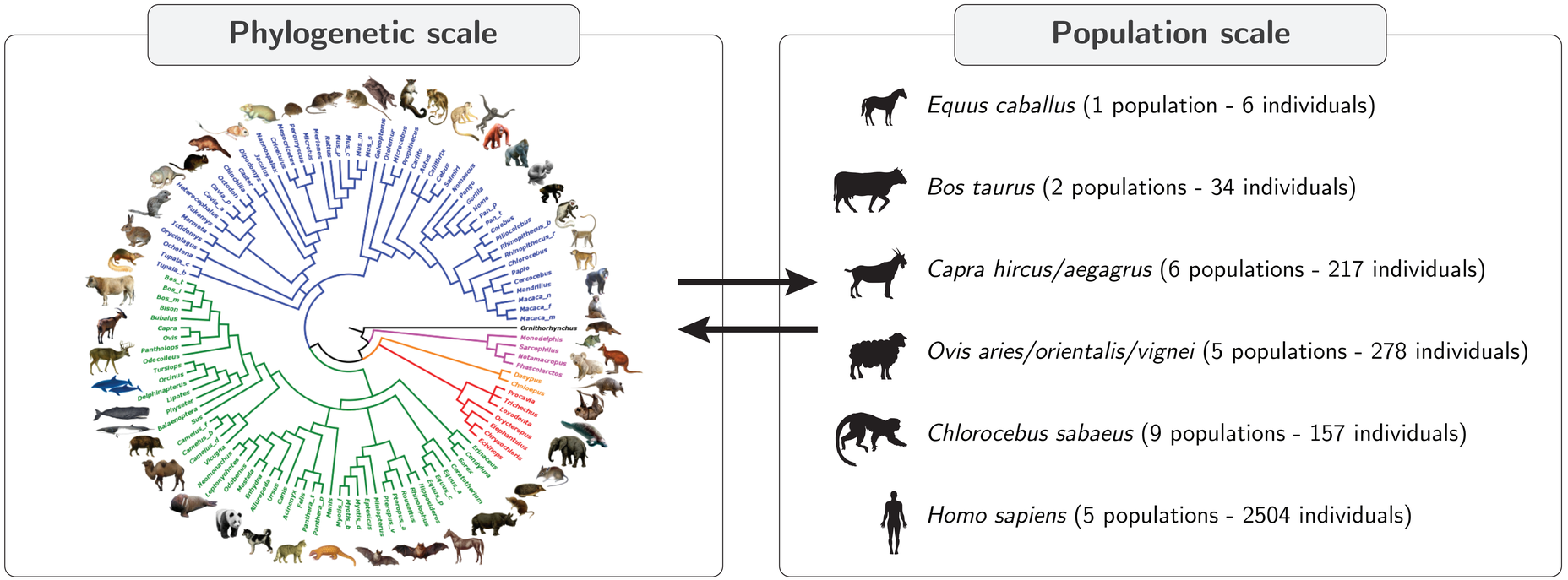
UNIL | Université de Lausanne

Département de Biologie Computationnelle

thibault.latrille@unil.ch

Evolution at different time scales

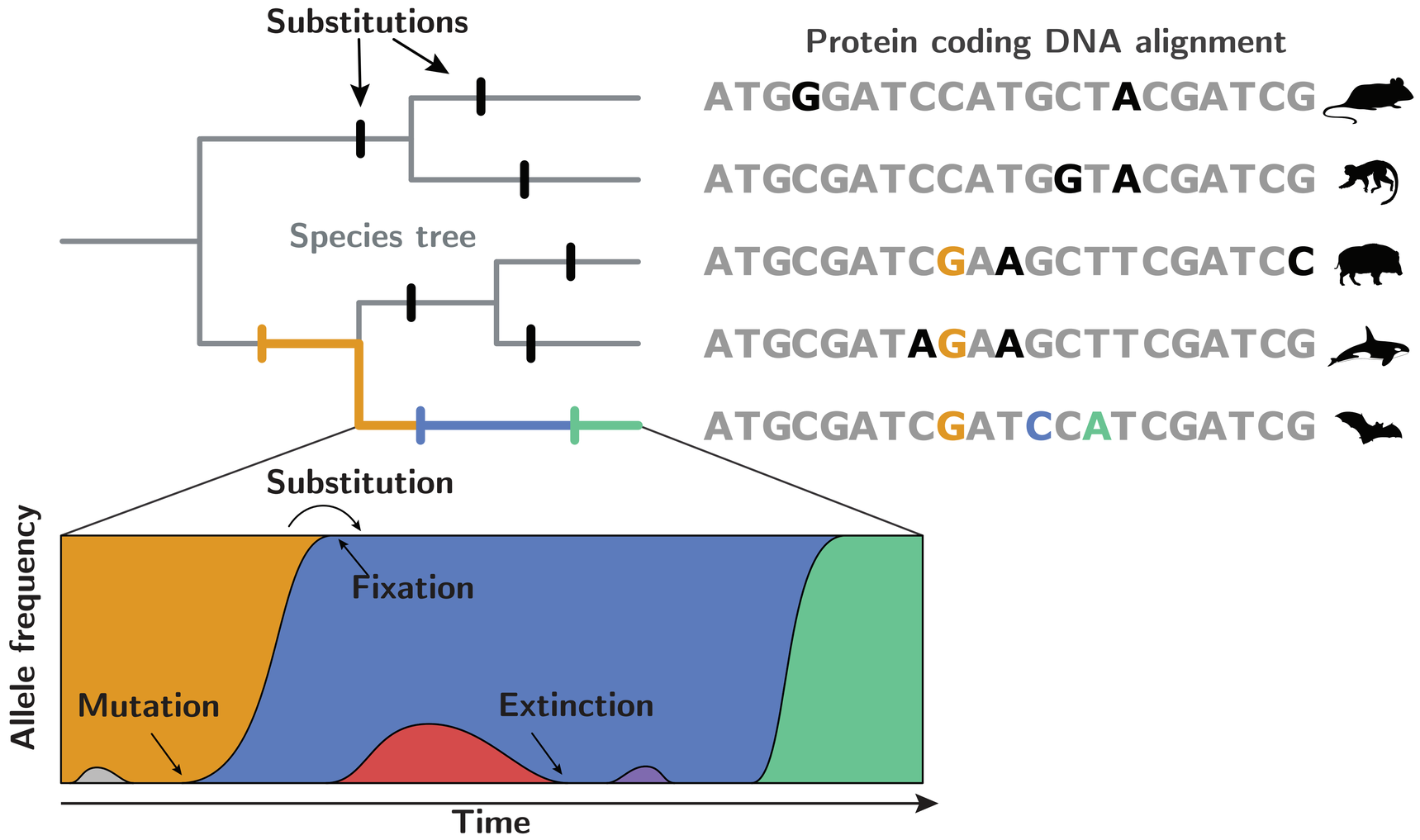
Protein coding DNA alignment for 14.509 conserved genes across mammals



- **Can phylogenetic models be derived from population-genetic principles?**
- **Can such model be used to disentangle mutation, selection and drift?**

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

Can phylogenetic models be derived from population-genetic principles?



How to derive substitution rate as a function of mutation, selection and drift?

Kimura (1969), Otha (1982)

Substitution rate for neutral mutations

- N_{e} is the number of diploid individuals.
- q is the substitution rate of new alleles.
- μ is the mutation rate of new alleles.
- $\mathbb{P}_{\mathrm{fix}}$ is the probability of fixation of new allele.

$$\begin{aligned} q &= 2N_{\mathrm{e}} \times \mu \times \mathbb{P}_{\mathrm{fix}}, \\ &= 2N_{\mathrm{e}} \times \mu \times \frac{1}{2N_{\mathrm{e}}}, \\ &= \mu. \end{aligned}$$

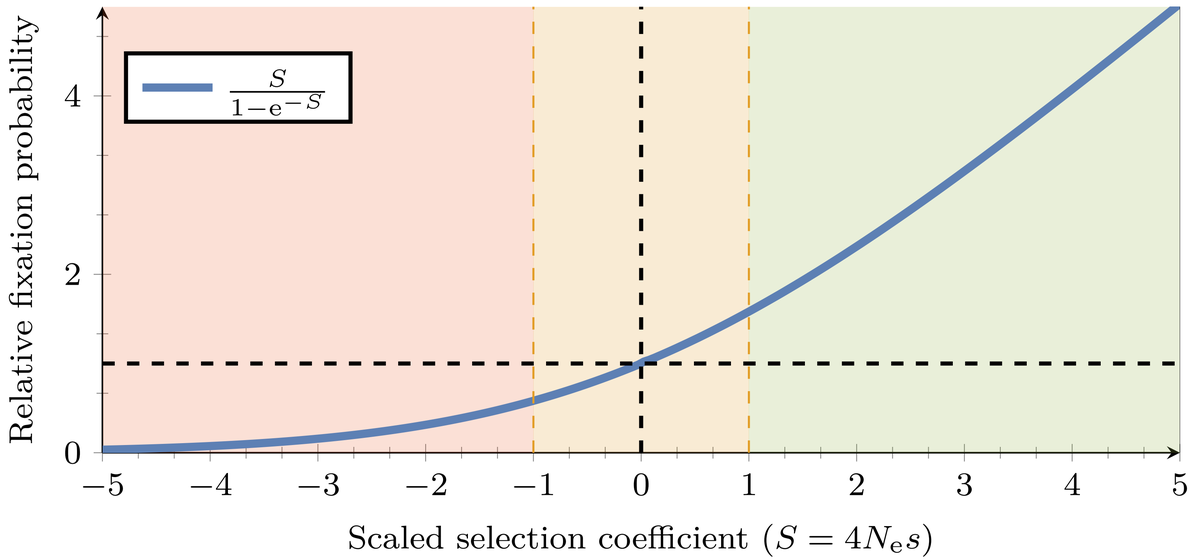
The substitution rate equals to the mutation rate

Kimura (1969), Otha (1982)

Substitution rate for mutations under selection

- N_e is the number of diploid individuals.
- q is the substitution rate of new alleles.
- μ is the mutation rate of new alleles.
- $\mathbb{P}_{\text{fix}}(s)$ is the probability of fixation of new allele, with selection coefficient s.

$$
\begin{aligned}
q &= 2N_e \times \mu \times \mathbb{P}_{\text{fix}}(s), \\
&= 2N_e \times \mu \times \frac{2s}{1 - e^{-4N_e s}}, \\
&= \mu \times \frac{S}{1 - e^{-S}} \text{ with } S = 4N_e s.
\end{aligned}
$$



The substitution rate is a function of mutation rate and scaled selection

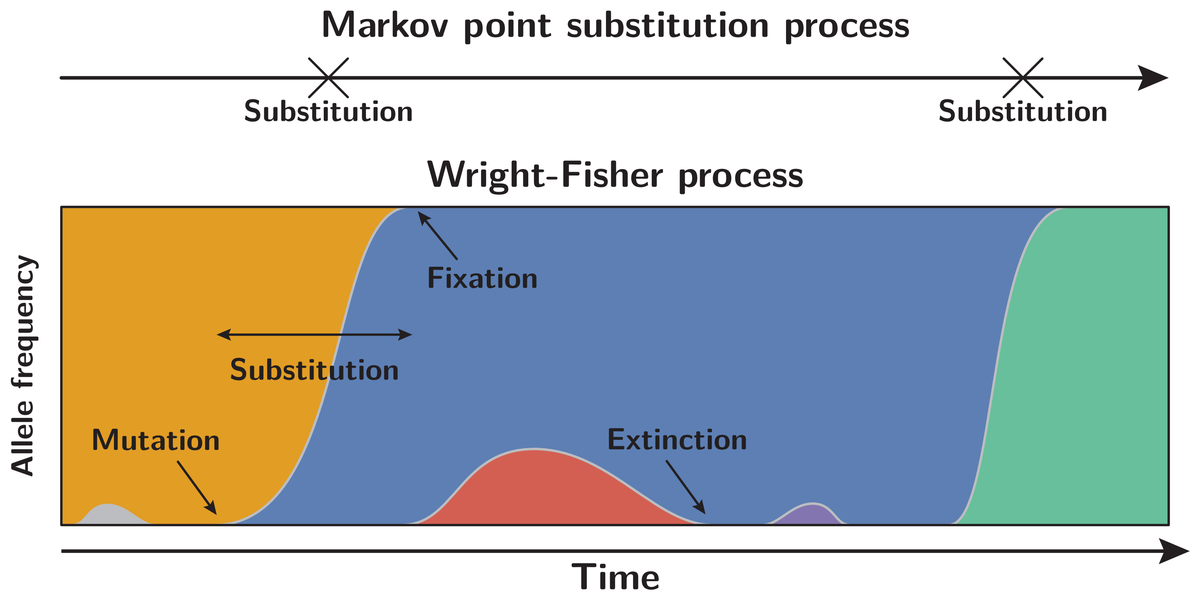
Kimura (1969), Otha (1982)

The mutation-selection process

Markov process with k states, with transition rate from a to b given as:

$$q_{a \to b} = \mu_{a \to b} \frac{F_b - F_a}{1 - \mathrm{e}^{F_a - F_b}},$$

where $\mu_{a \to b}$ is the mutational rate from a to b, and F_a the scaled fitness of a.



Halpern & Bruno (1998); McCandlish (2014)

Stationary distribution of the mutation-selection process

Markov process with k states, with transition rate from a to b given as:

$$q_{a \to b} = \mu_{a \to b} \frac{F_b - F_a}{1 - \mathrm{e}^{F_a - F_b}}, \tag{1}$$

Assuming that the mutational process is time-reversible, the detailed balance for the stationary distribution of the mutation process ($\boldsymbol{\sigma}$) satisfy:

$$\frac{\mu_{b \to a}}{\mu_{a \to b}} = \frac{\sigma_a}{\sigma_b} \tag{2}$$

Then, the Markov process for substitutions is time-reversible, the detailed balance for the stationary distribution ($\boldsymbol{\pi}$) satisfy:

$$\frac{\pi_a}{\pi_b} = \frac{q_{b \to a}}{q_{a \to b}} = \frac{\mu_{b \to a}}{\mu_{a \to b}} \frac{\mathrm{e}^{F_a}}{\mathrm{e}^{F_b}} = \frac{\sigma_a}{\sigma_b} \frac{\mathrm{e}^{F_a}}{\mathrm{e}^{F_b}}. \tag{3}$$

The stationary distribution $\boldsymbol{\pi}$ is:

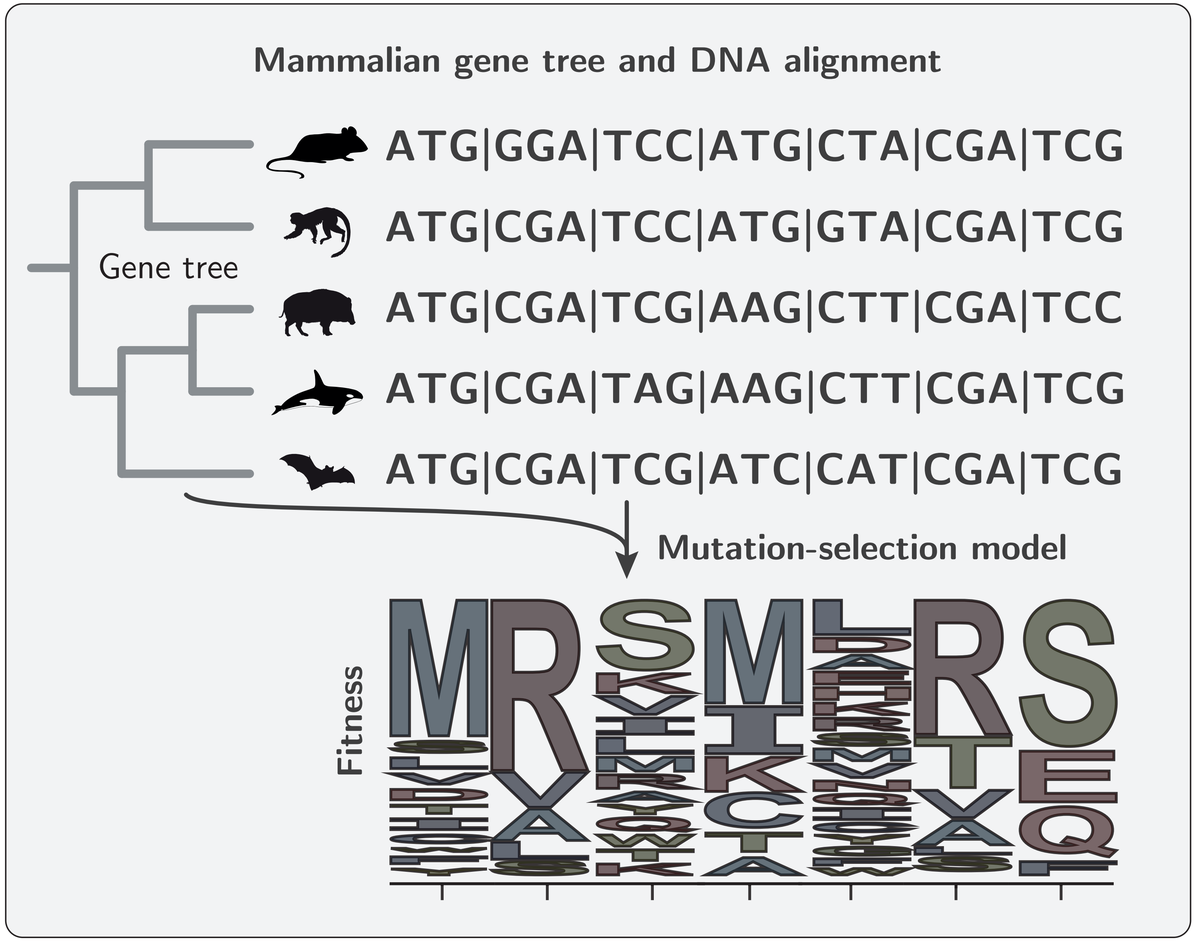
$$\pi_a = \frac{\sigma_a \mathrm{e}^{F_a}}{\sum_{b=1}^{k} \sigma_b \mathrm{e}^{F_b}}. \tag{4}$$

Halpern & Bruno (1998); McCandlish (2014)

Mutation-selection models applied to protein coding DNA sequences

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

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



- **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)

The phylogenetic mutation-selection codon model as a bridge between phylogenetics and population-genetics

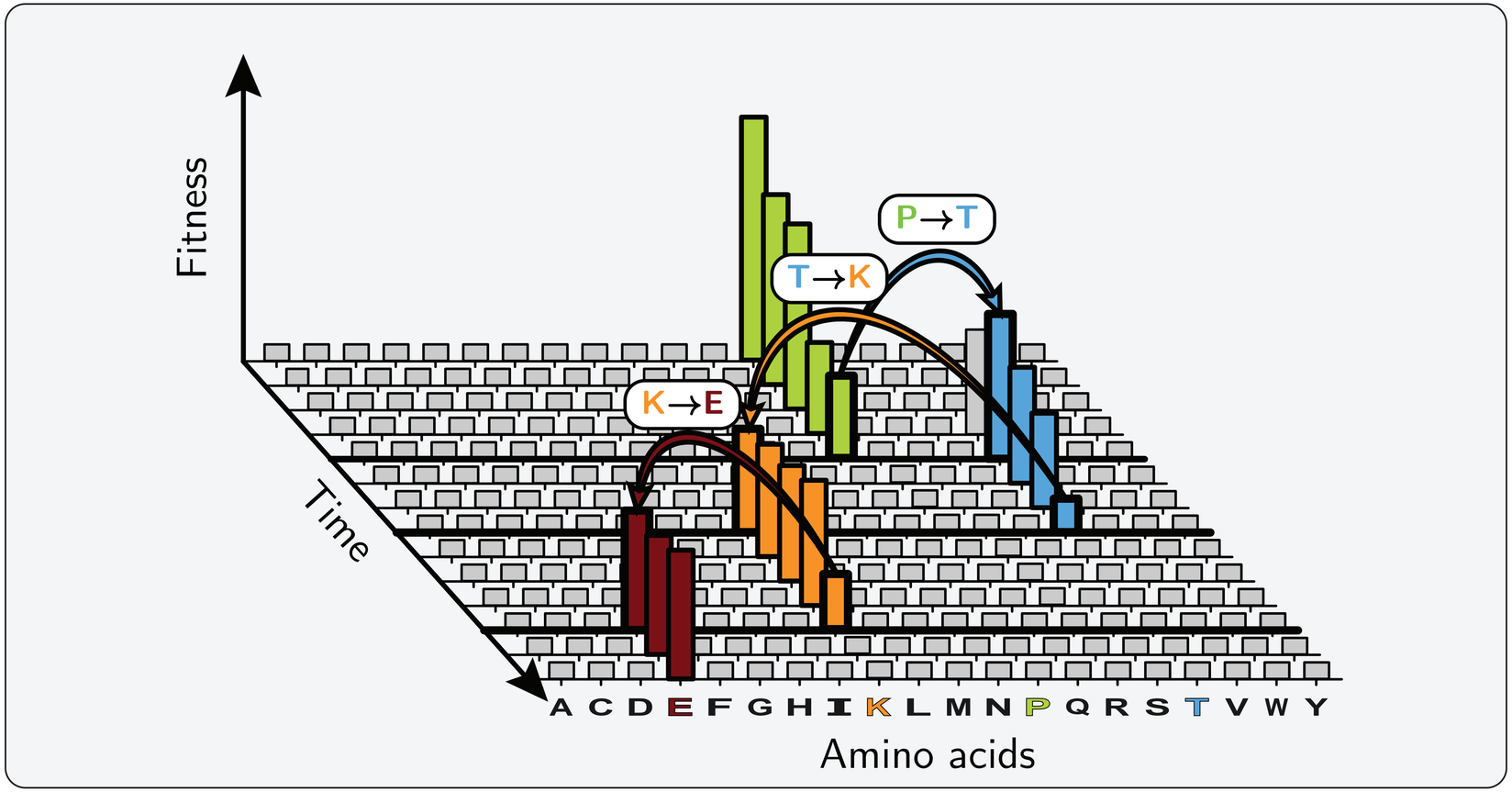
I. Mutation-selection models to detect adaptation

- Is adaptation congruent at the phylogenetic and population-genetic scale?
- *Latrille, Rodrigue & Lartillot*, PNAS (2023).

II. Mutation-selection models to predict selection coefficient of mutations

- Is a new mutation deleterious, nearly-neutral or beneficial?
- *Latrille, Joseph, Hartasánchez & Salamin*, in prep.

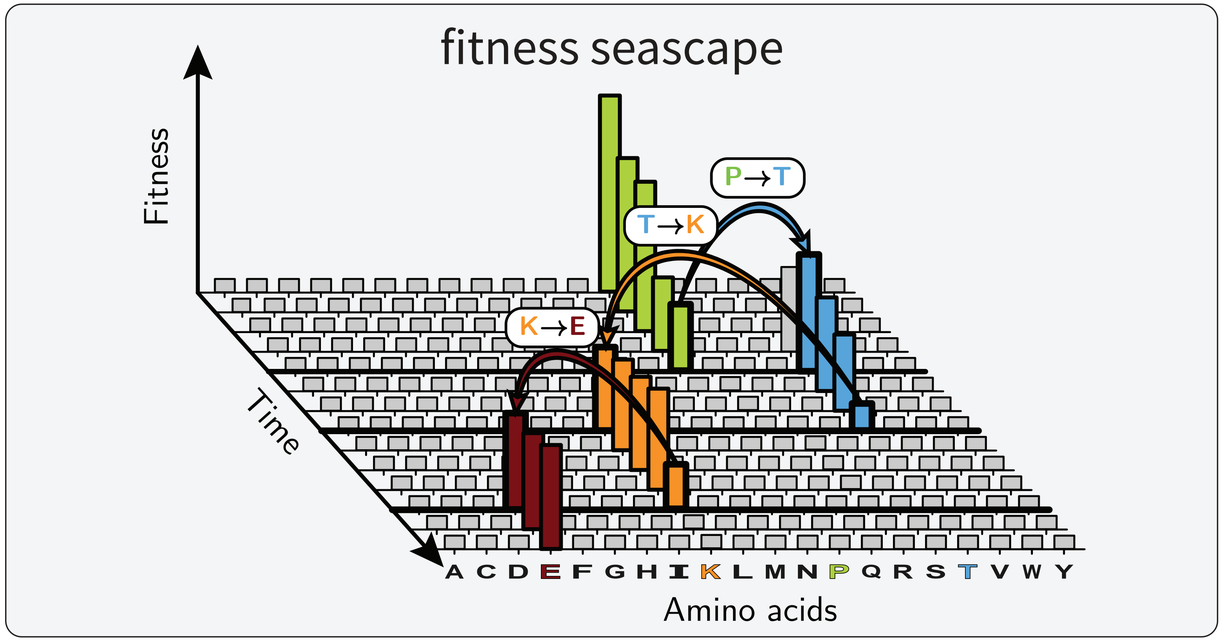
What is adaptation?

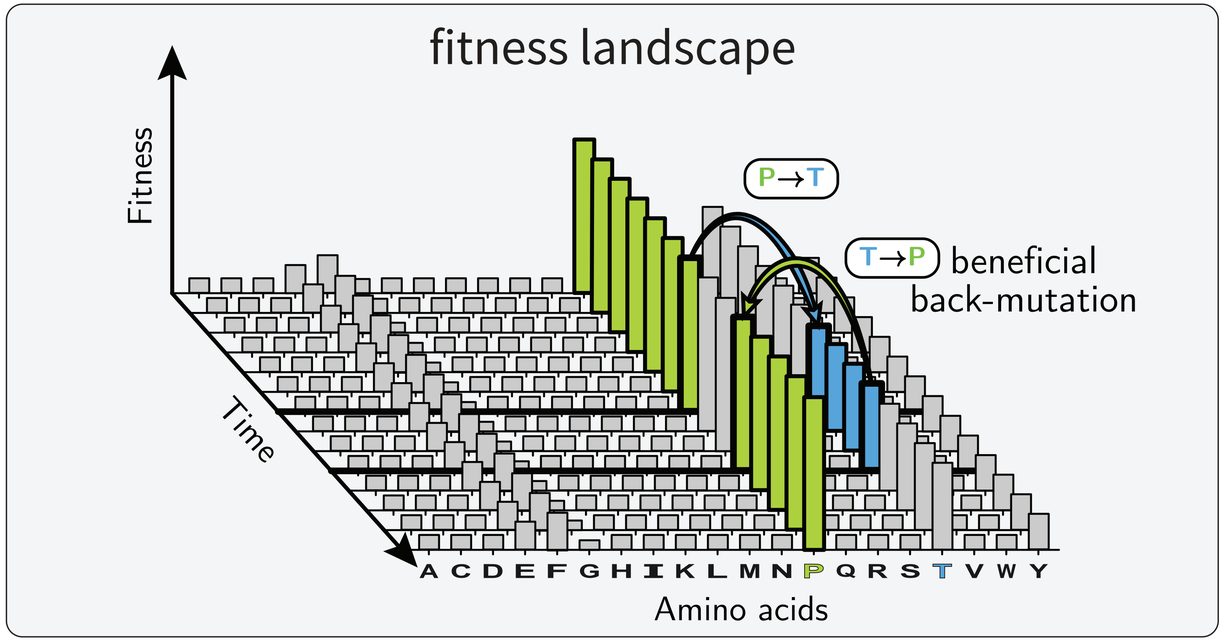


- **Environmental changes that are external to the genome.**
  - **The optimal state is a moving target.**

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

Fitness landscapes and seascapes





- **The mutation-selection model is null model of evolution without adaptation**

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

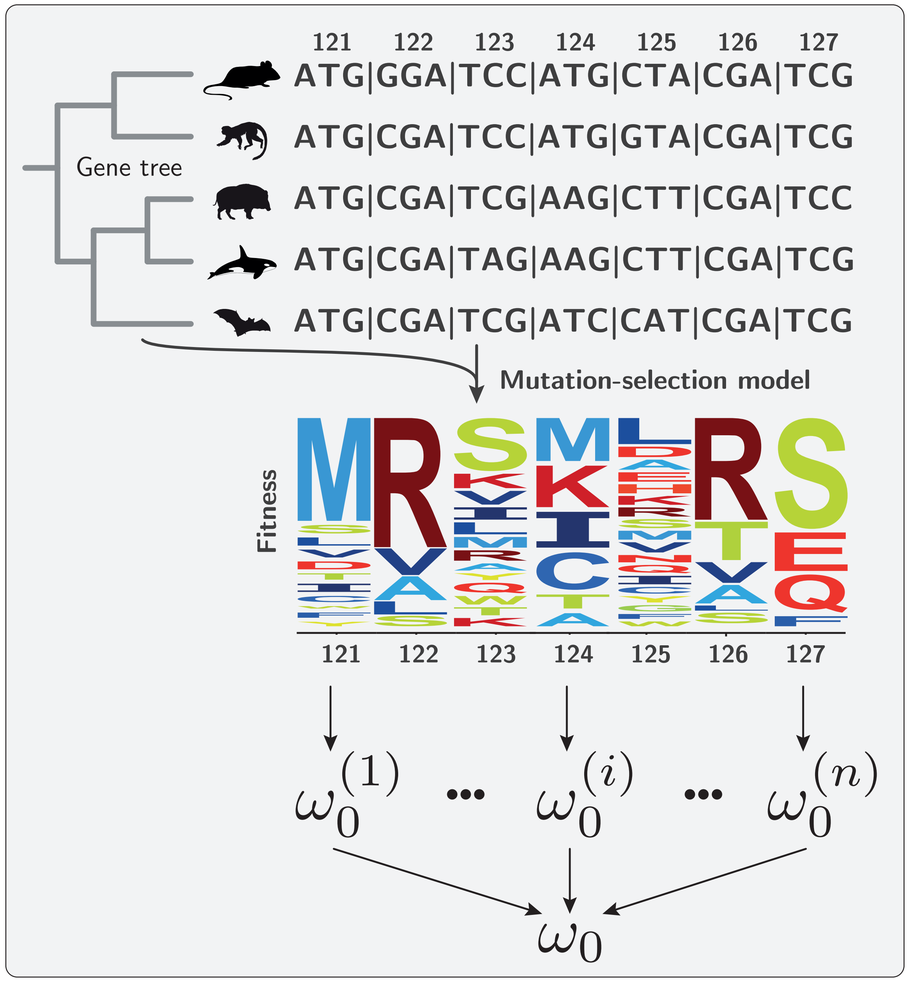
Predicting the rate of evolution of a gene

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

$$\omega_0^{(i)} = \frac{\left\langle \pi_a^{(i)} q_{a \to b}^{(i)} \right\rangle}{\left\langle \pi_a^{(i)} \mu_{a \to b} \right\rangle},$$

$$\Rightarrow \omega_0 = \frac{1}{n} \sum_{i=1}^{n} \omega_0^{(i)}.$$

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

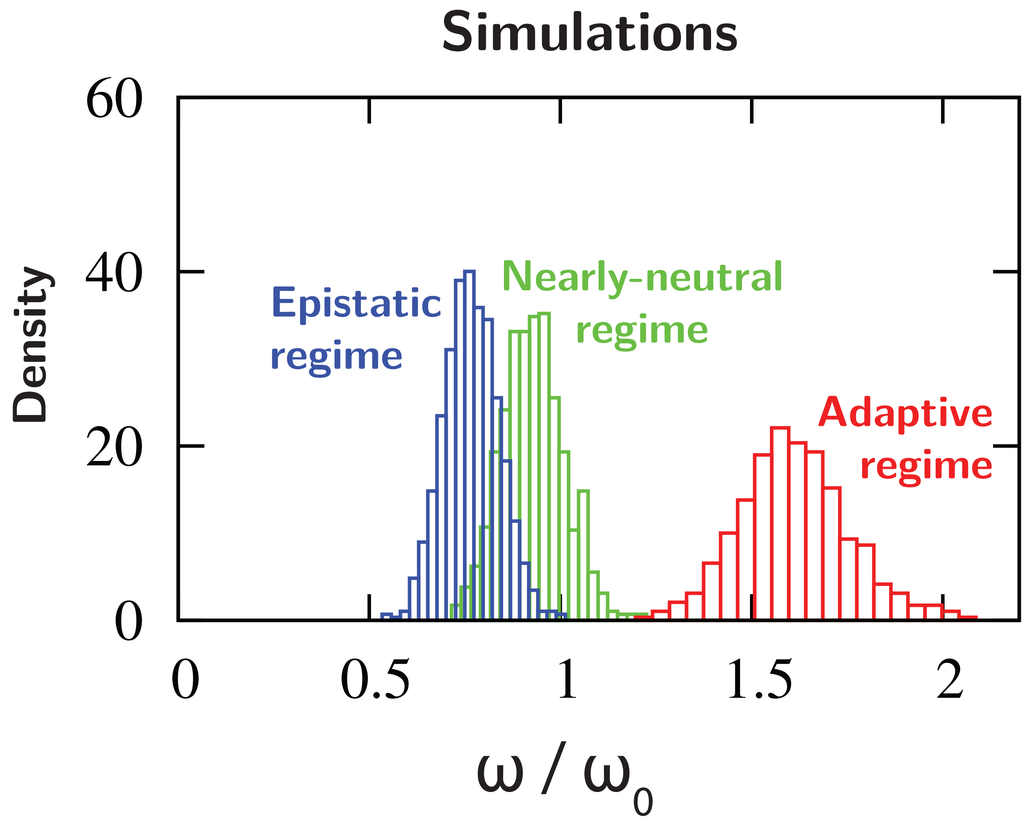


- ω_0: predicted rate of evolution under the mutation-selection model.

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

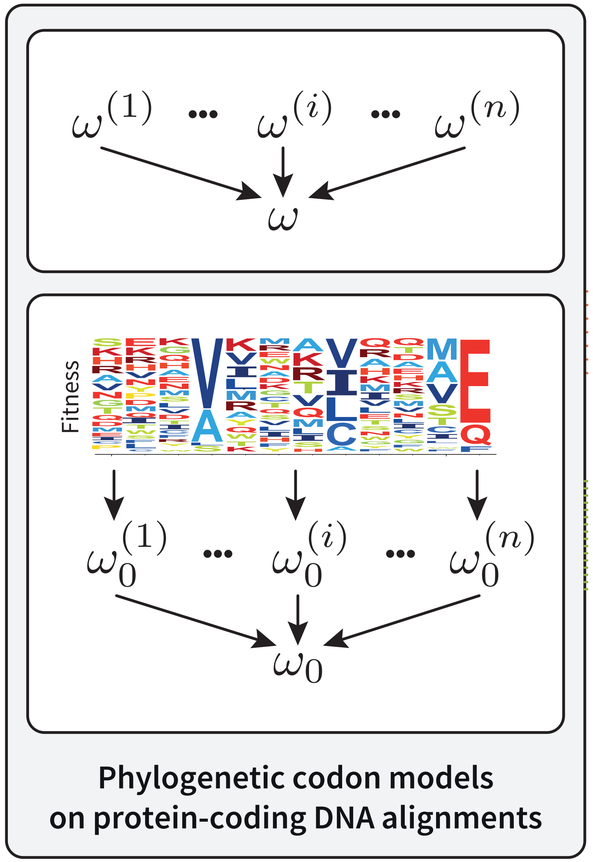
Mutation-selection model tested against simulations

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

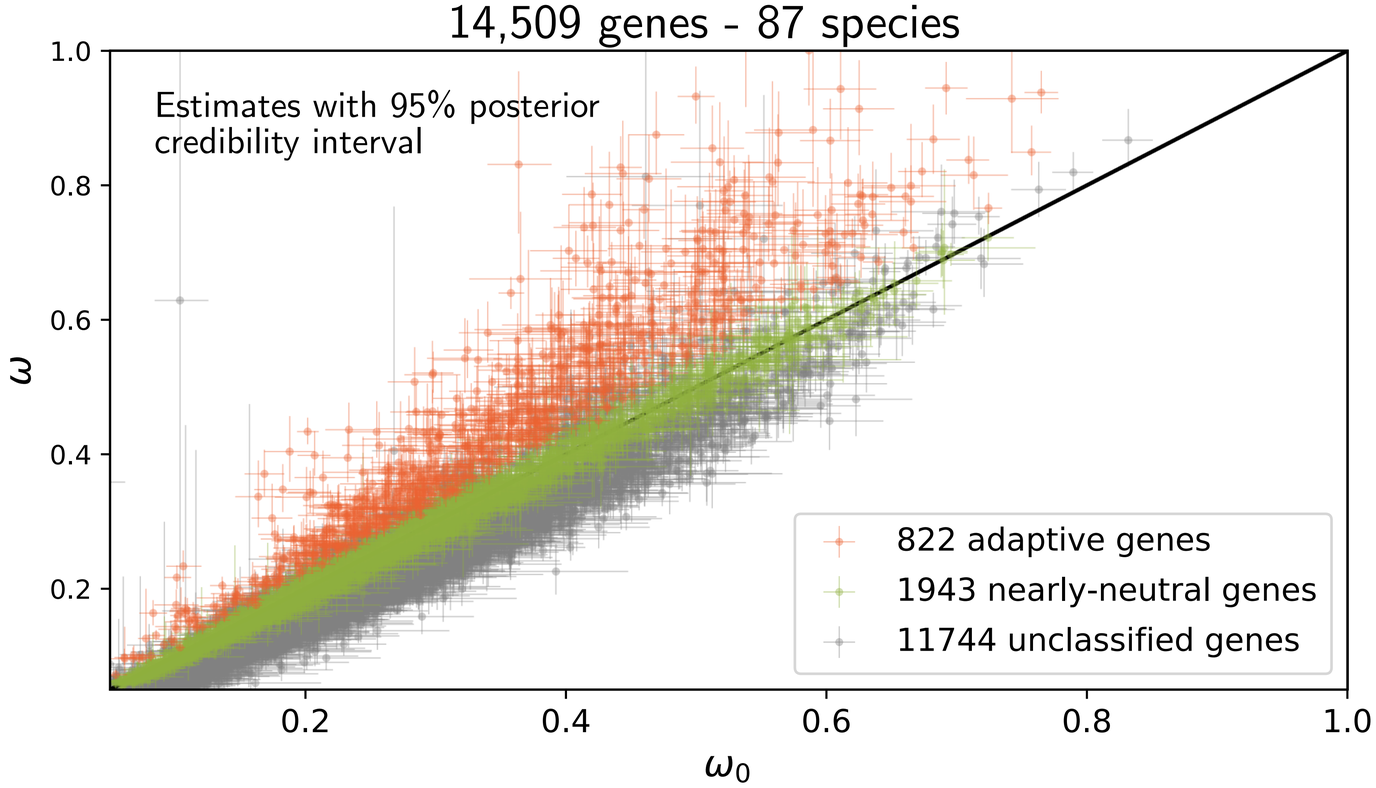


Rodrigue & Lartillot (2017)

Mutation-selection model applied to mammalian genes

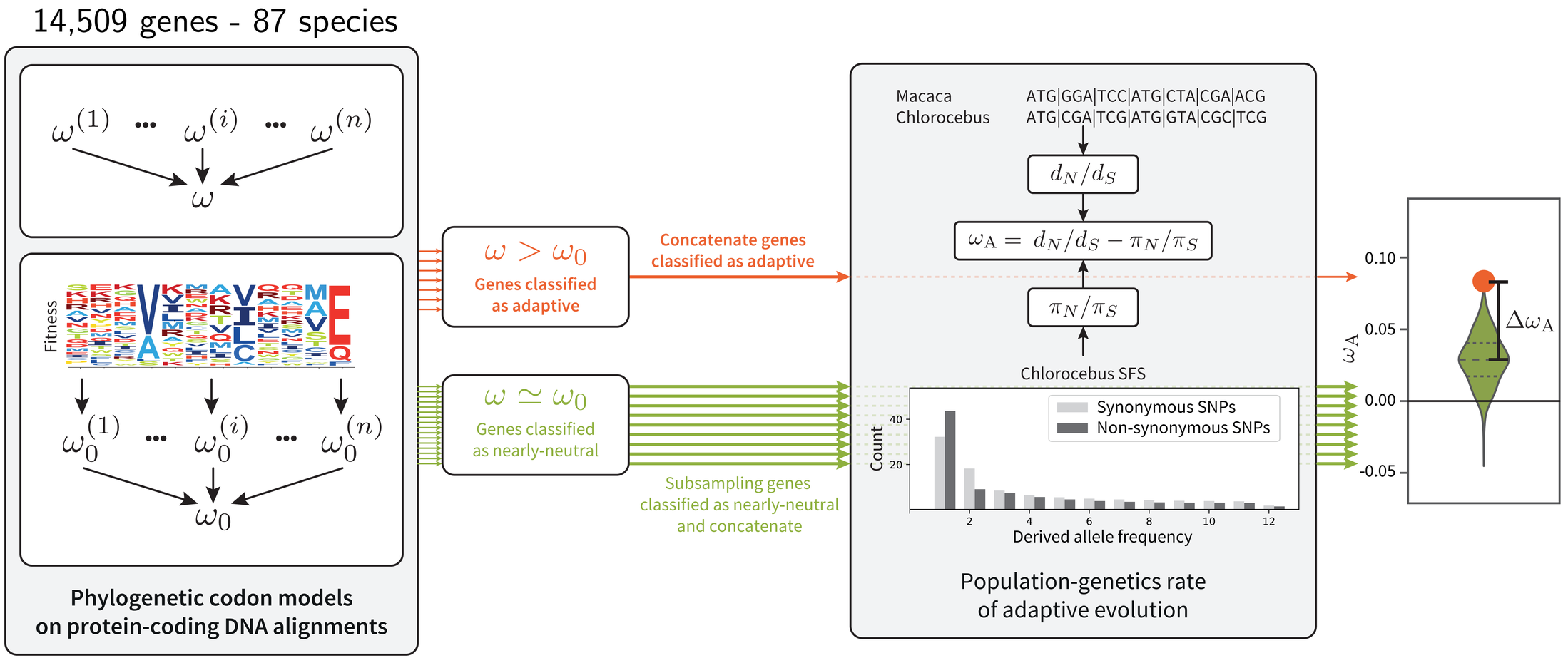


Phylogenetic codon models
on protein-coding DNA alignments



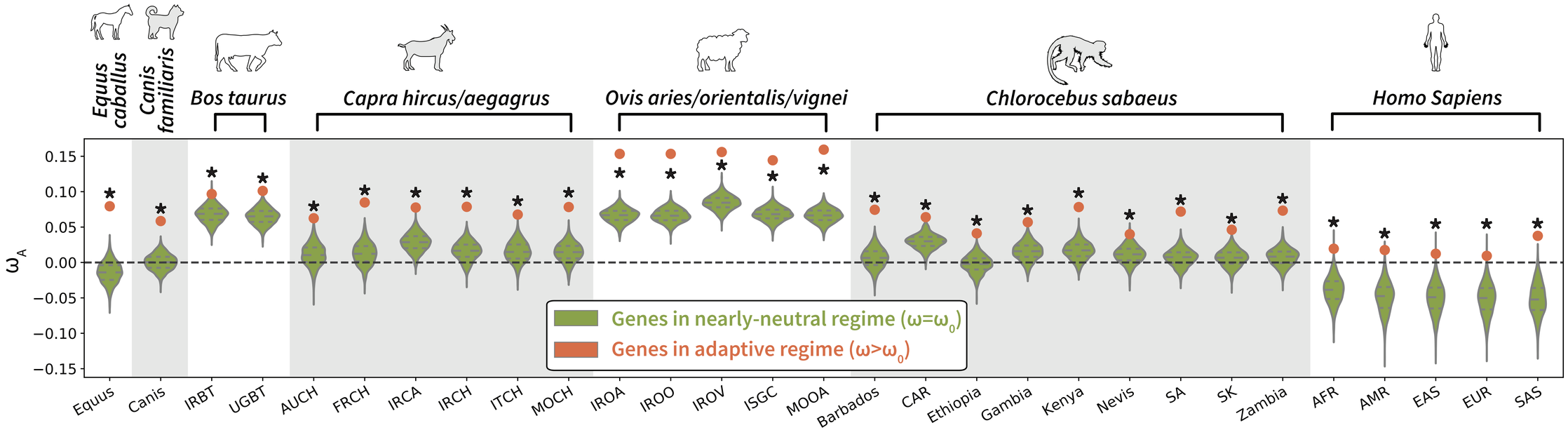
- **Genes predicted to be under adaptation at the phylogenetic scale are enriched in ontologies related to immunity, response to virus and external membrane.**

Is adaptation at different evolutionary scale comparable?



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

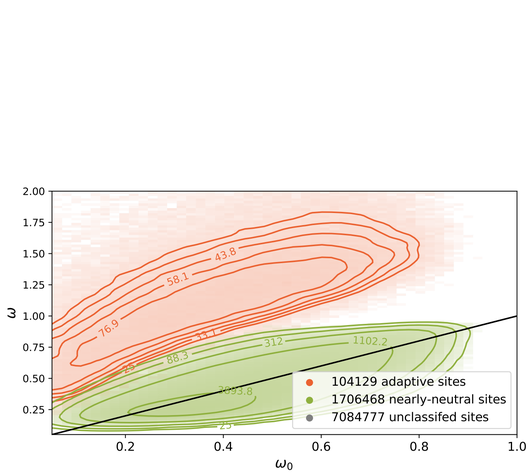
Replicability across different populations

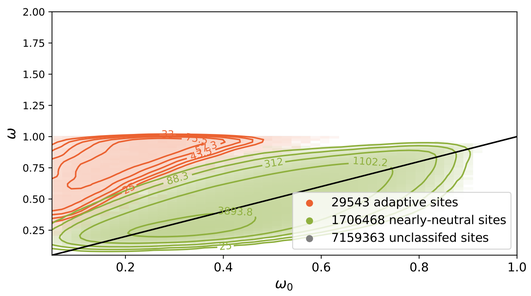


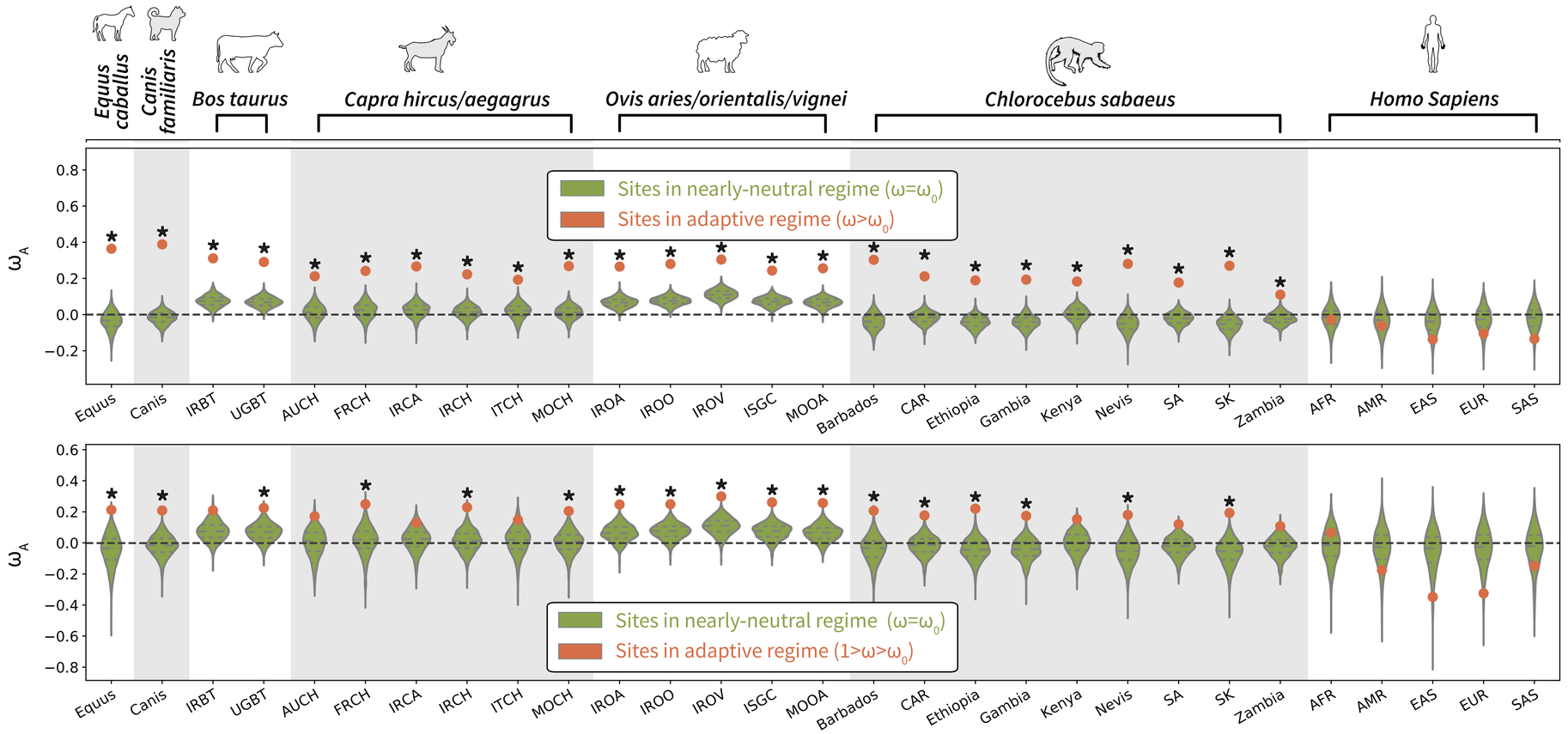
- **Genes predicted to be under adaptation at the phylogenetic scale are under adaptation at the population-genetic scale.**

- **Replicable across populations and species.**

Detection of adaptation at the site level







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

- **Replicable across populations and species.**

(partial) conclusion

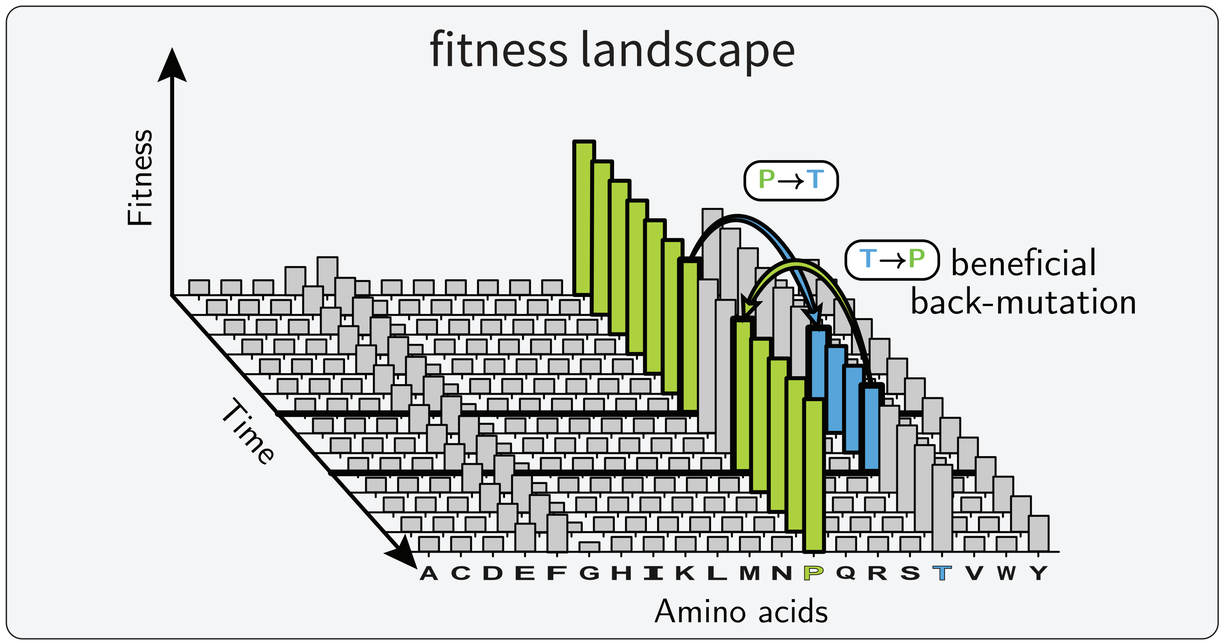
I. **Genes and sites predicted to be under adaptation at the phylogenetic scale exhibit adaptation at the population-genetic scale.**

- Mutation-selection is a null (nearly-neutral) model of evolution.
- Stastically more powerfull than detecting adaptation as $\omega > 1$ (neutral model).

II. Mutation-selection models to predict selection coefficient of mutations

- Is a new mutation deleterious, nearly-neutral, or a beneficial?
- Can we derive the proportion of beneficial mutations that are not adaptive?

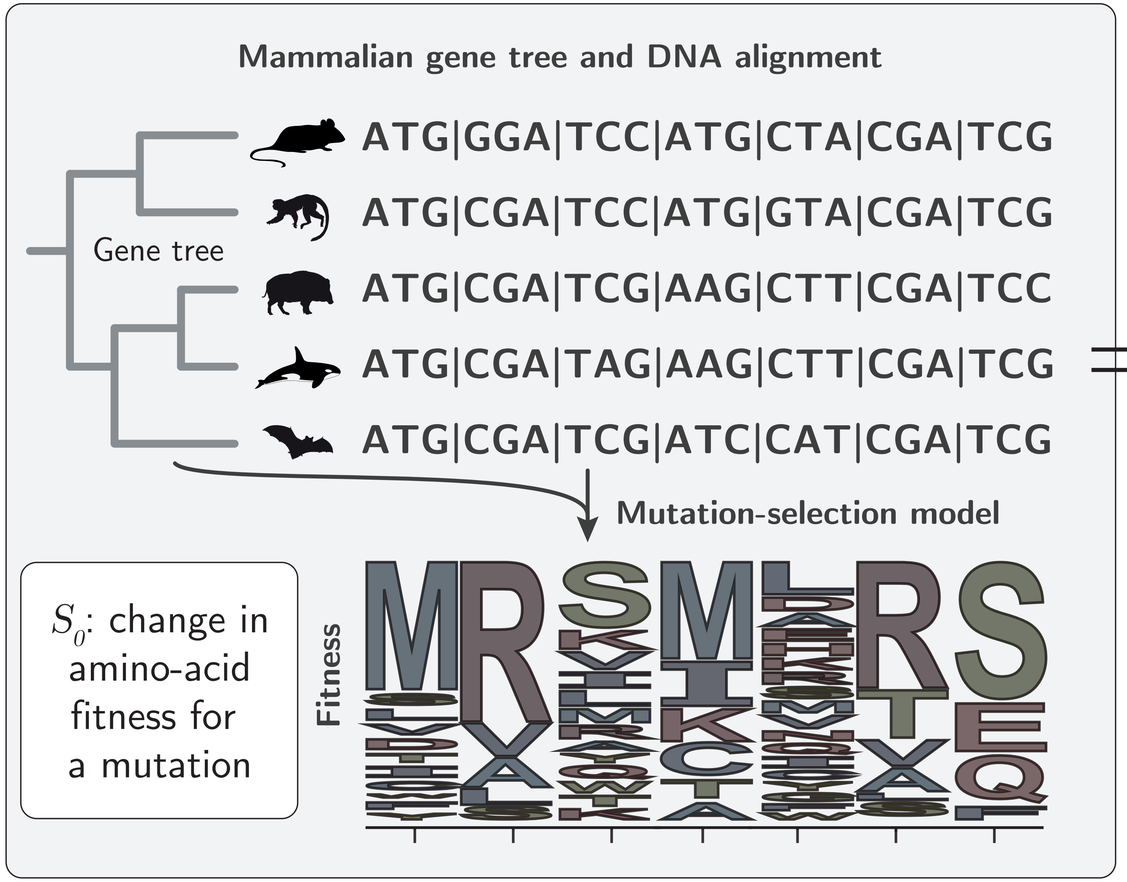
Beneficial and deleterious mutations at mutation-selection balance

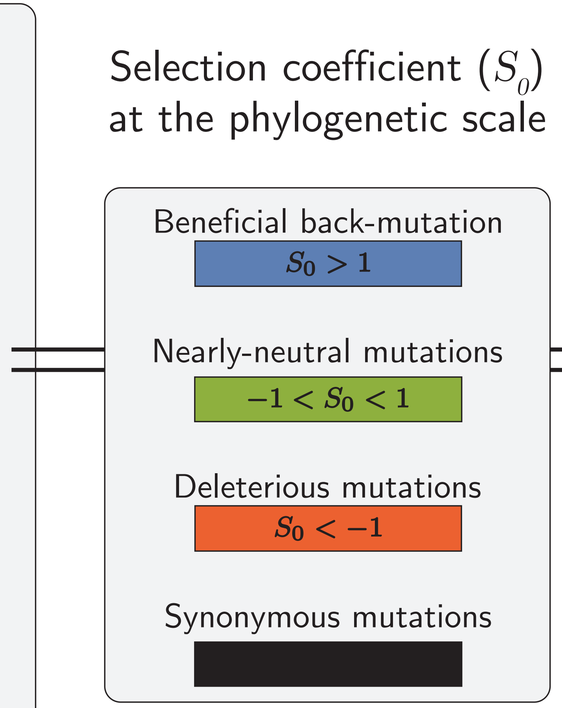


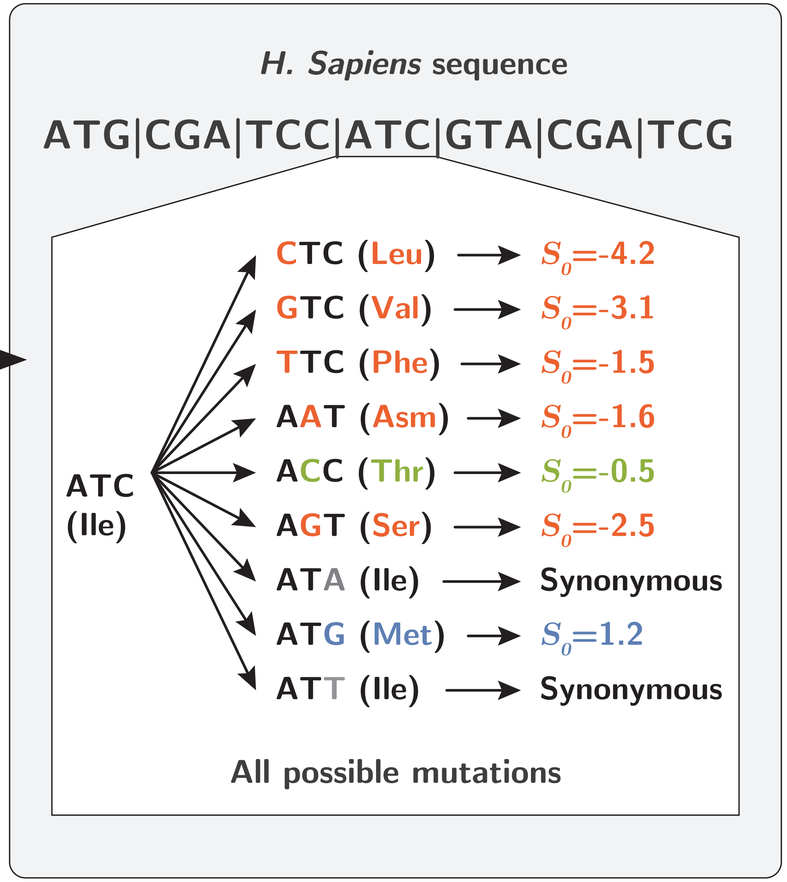
- Deleterious mutations can reach fixation due to genetic drift.
- Mutations in the opposite direction are restoring our genomes, not driving innovation.

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

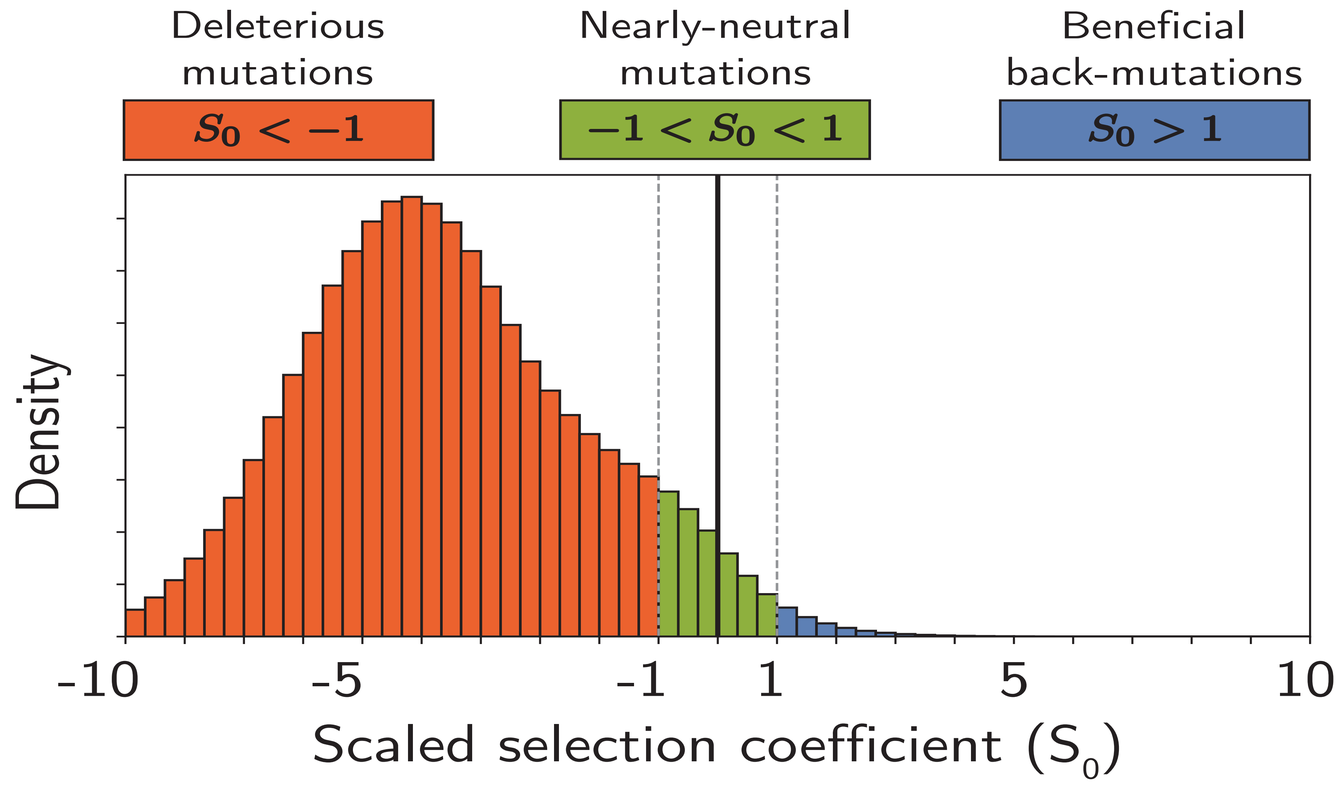
Is a new mutation deleterious, nearly-neutral, or a beneficial back-mutation?







What is the expected effect of a new mutation in *H. sapiens*?

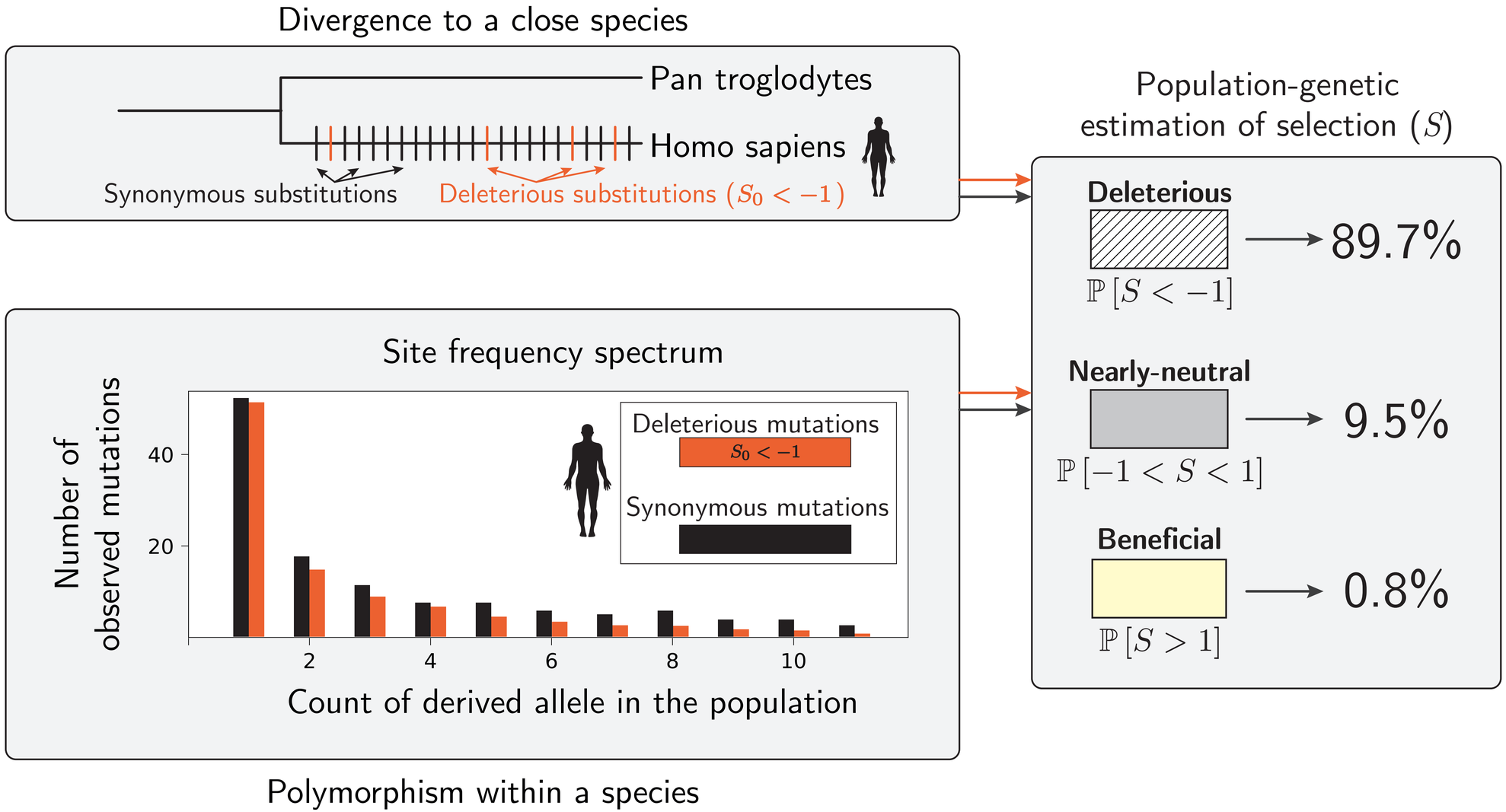


- **Are predicted deleterious mutations purified away in *H. sapiens?***
  → SNPs associated with clinical terms such as *likely pathogenic & pathogenic*

- **Are predicted beneficial back-mutations advantageous in *H. sapiens?***
  → Substitutions in the terminal lineage with $d_N/d_S=1.5$

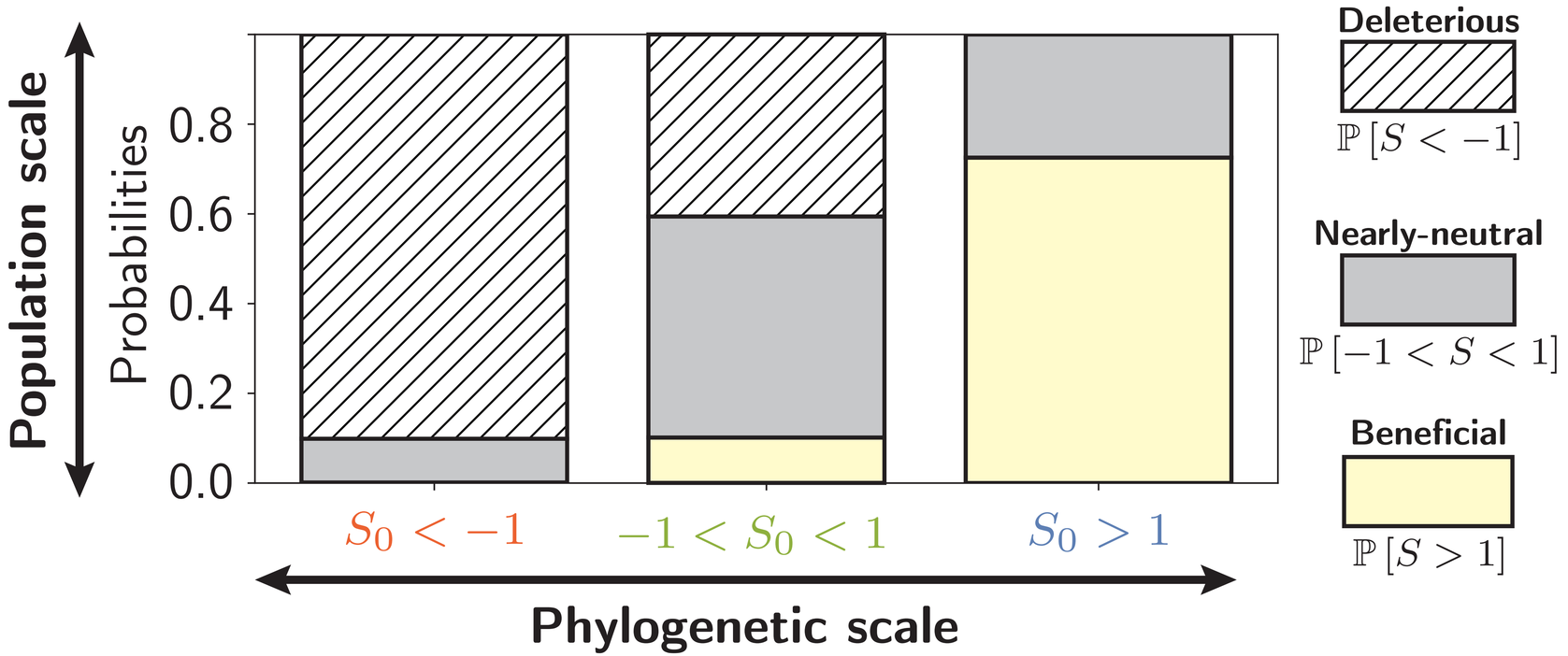
- **Is selection at the phylogenetic and population scale comparable?**

Is selection at the phylogenetic and population scale comparable?



Tataru & Bataillon (2020)

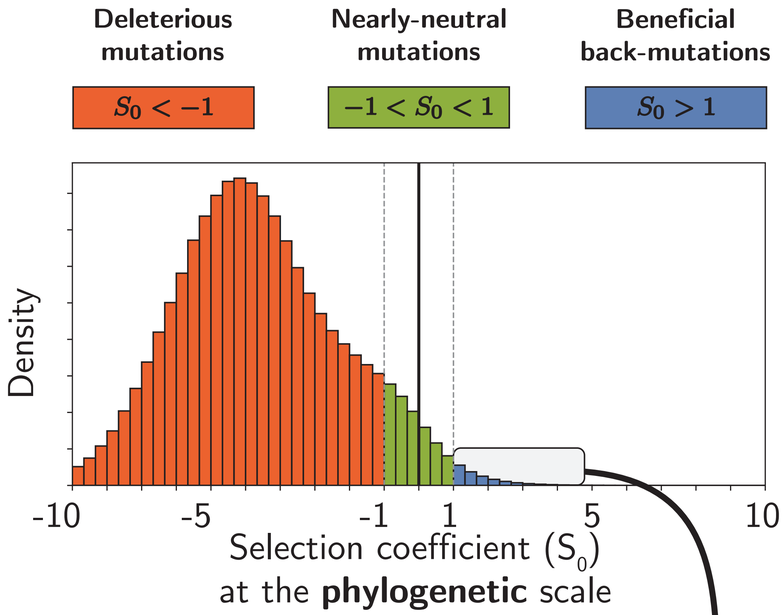
Is selection at the phylogenetic and population scale comparable?

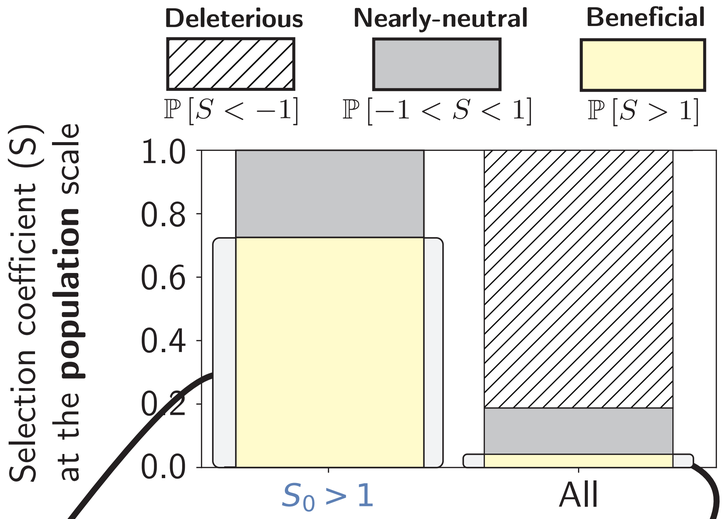


- **Predicted deleterious mutations are effectively purified away in *H. sapiens.***
  → 90% precision (false positives) and 97% recall (false negatives).

- **Predicted beneficial back-mutations are effectively beneficial in *H. sapiens.***
  → 70% precision (false positives) and 25% recall (false negatives).

What is the proportion of beneficial mutations that are not adaptive?





$$\mathbb{P}[S_0 > 1] = 0.014$$

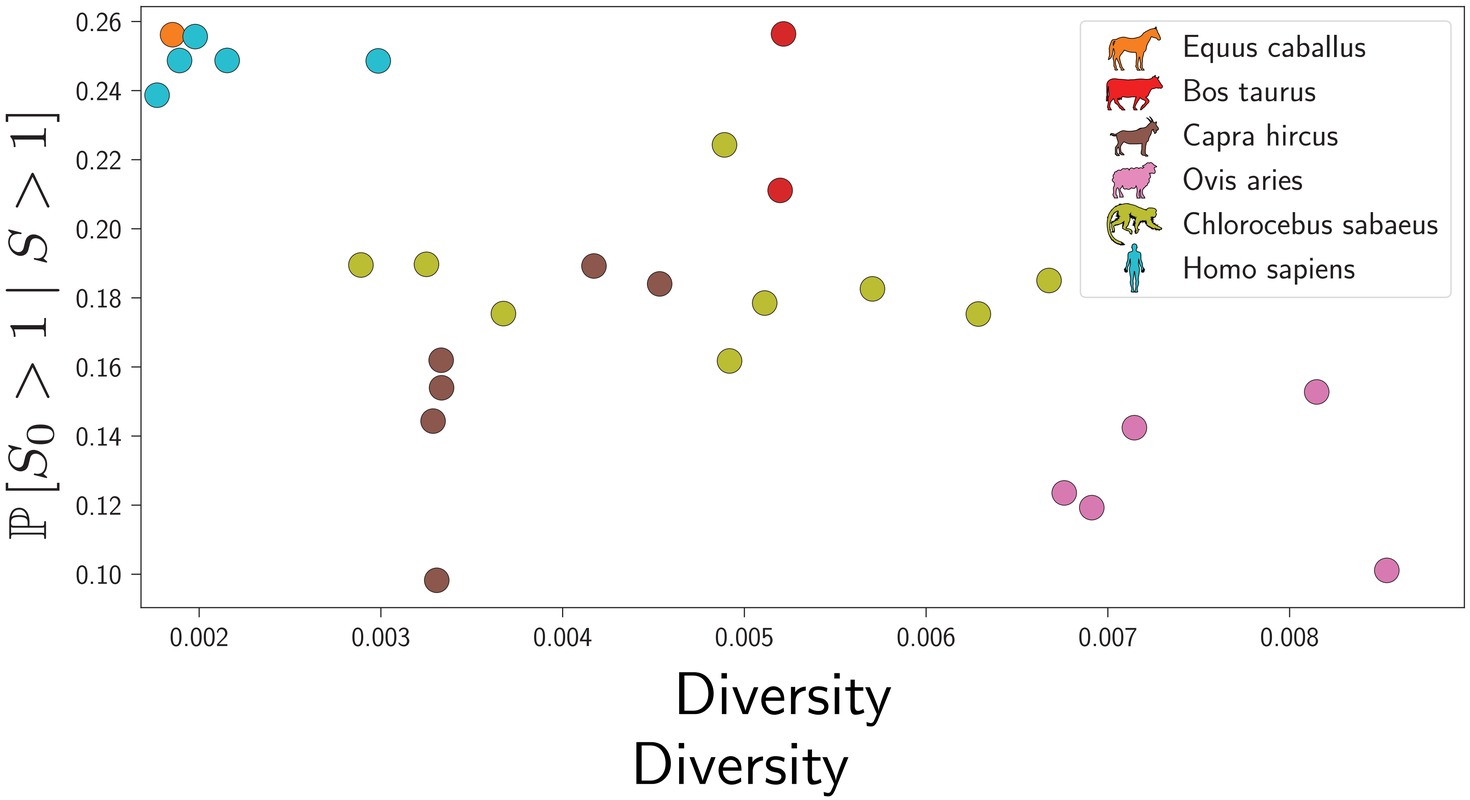
$$\mathbb{P}[S > 1 \mid S_0 > 1] = 0.72$$

$$\mathbb{P}[S > 1] = 0.042$$

$$\mathbb{P}[S_0 > 1 \mid S > 1] = \frac{\mathbb{P}[S > 1 \mid S_0 > 1]\,\mathbb{P}[S_0 > 1]}{\mathbb{P}[S > 1]} = 0.25$$

Probability for beneficial mutation to be a back-mutation

Are our results replicable?



Diversity

- **P[S_0>1 | S>1]: probability for beneficial mutation to be a back-mutation.**
- **Between 10 and 26% across different populations (mammals)**
- **Higher proportion of beneficial back-mutations for population with lower diversity.**

Take home messages

I. **Genes and sites predicted to be under adaptation at the phylogenetic scale exhibit adaptation at the population-genetic scale.**

- Mutation-selection is a null (nearly-neutral) model of evolution.
- Stastically more powerfull than detecting adaptation as $\omega > 1$ (neutral model).

II. Mammalian genes exhibit widespread beneficial mutations that are not adaptive

- Mutation-selection model can predict the selective effect of mutations.
- We can derive the proportion of beneficial mutations that are not adaptive.

Acknowledgements

Nicolas Salamin & his Group

UNIL | Université de Lausanne

Julien Joseph & LBBE

UNIVERSITÉ DE LYON

LBBE