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

ISMB 2024 - Montreal

July 15, 2024

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

[1] Université de Lausanne (Switzerland)

[2] Université de Lyon (France)

UNIL | Université de Lausanne

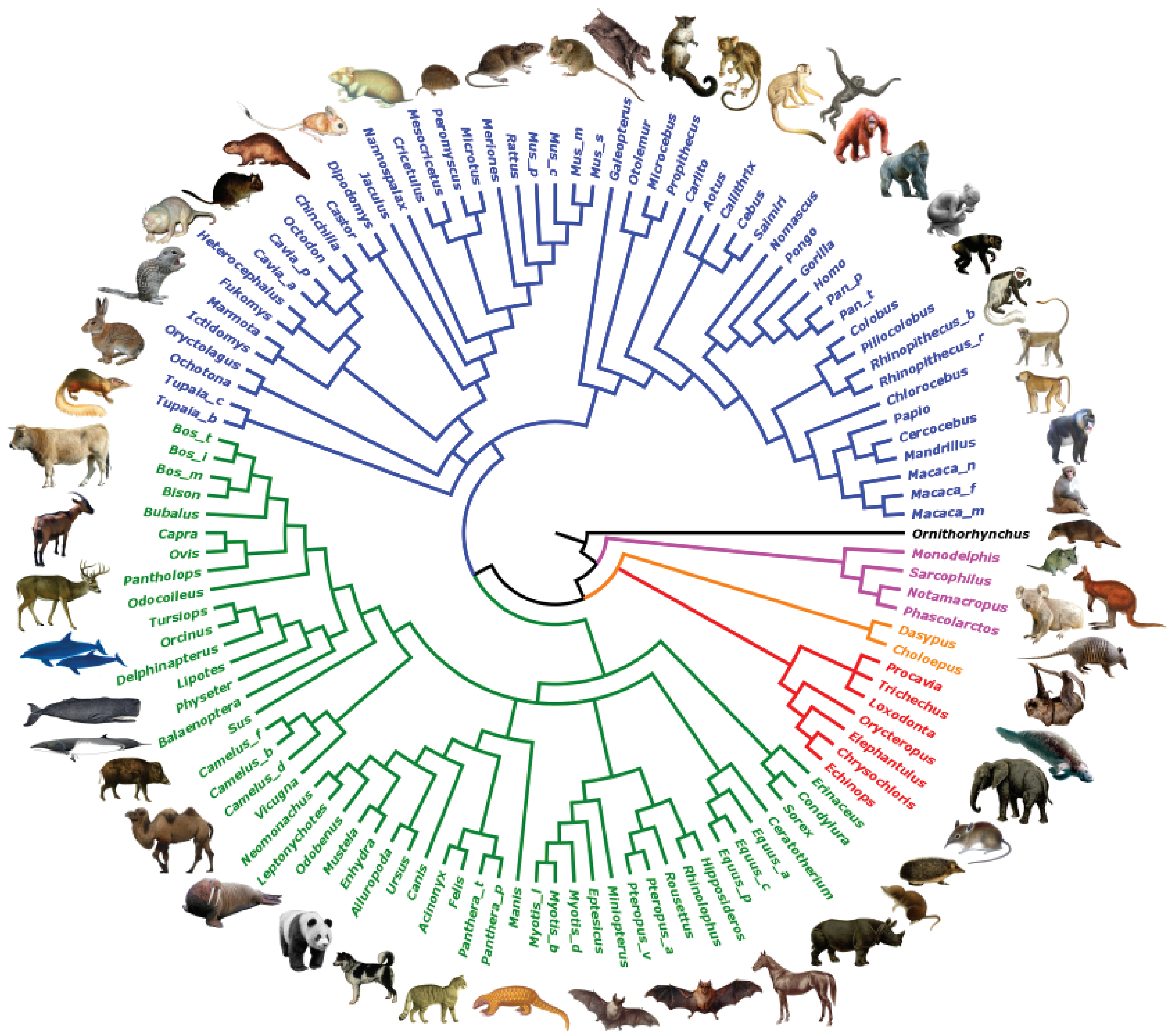
Département de
Biologie Computationnelle

thibault.latrille@unil.ch

Evolution at different time scales

Protein coding DNA alignment for 14.509 conserved genes across mammals

Phylogenetic scale



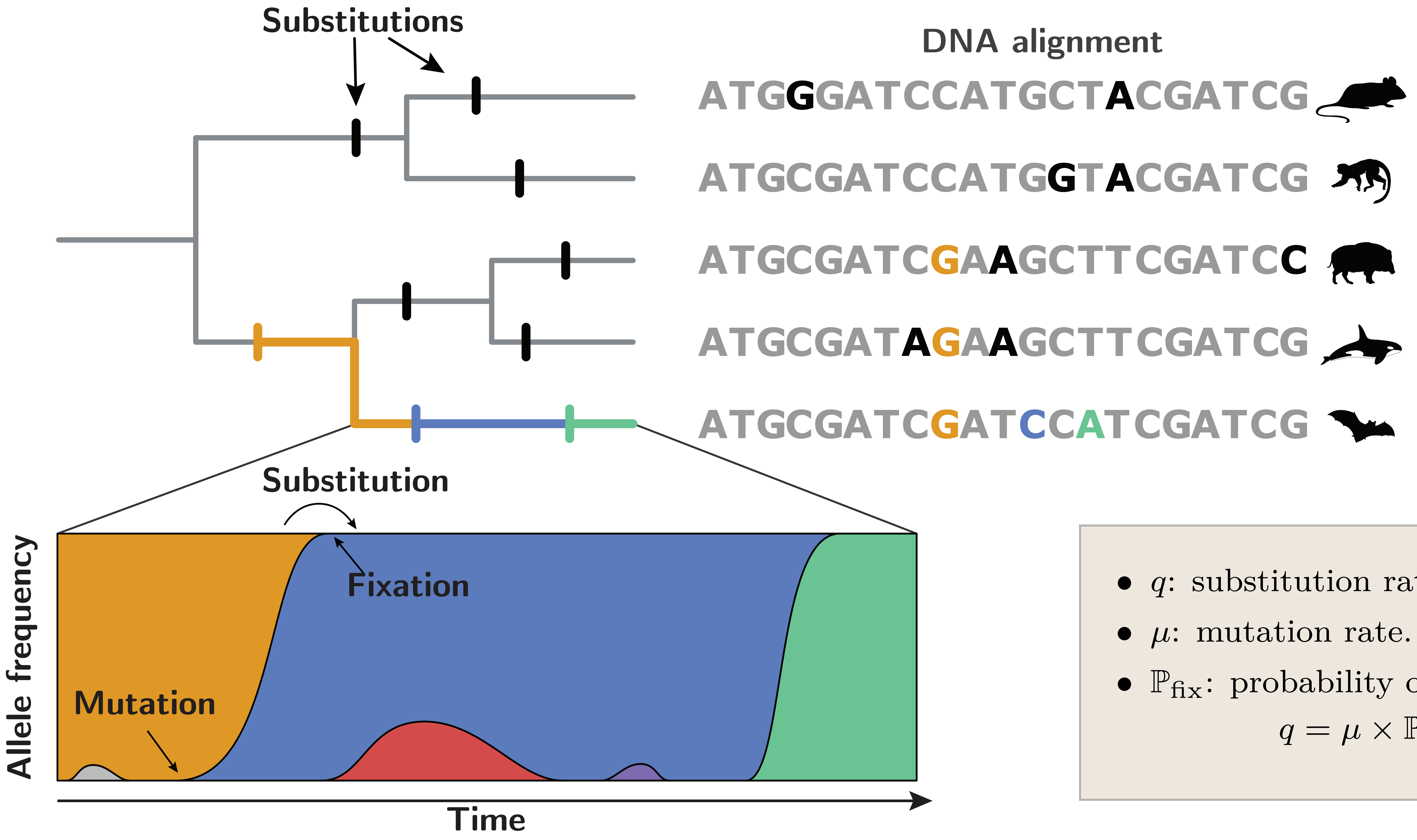
Population scale

- Equus caballus (1 population - 6 individuals)
- Bos taurus (2 populations - 34 individuals)
- Capra hircus/aegagrus (6 populations - 217 individuals)
- Ovis aries/orientalis/vignei (5 populations - 278 individuals)
- Chlorocebus sabaeus (9 populations - 157 individuals)
- Homo sapiens (5 populations - 2504 individuals)

- **Can we estimate the fitness effect of mutations from a mammalian DNA alignment?**
- **Is selection estimated across mammals predictive of selection in a current population?**

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

Substitutions are mutations that reached fixation



- q: substitution rate.
- μ: mutation rate.
- $\mathbb{P}_{\text{fix}}$: probability of fixation.

$$q = \mu \times \mathbb{P}_{\text{fix}}$$

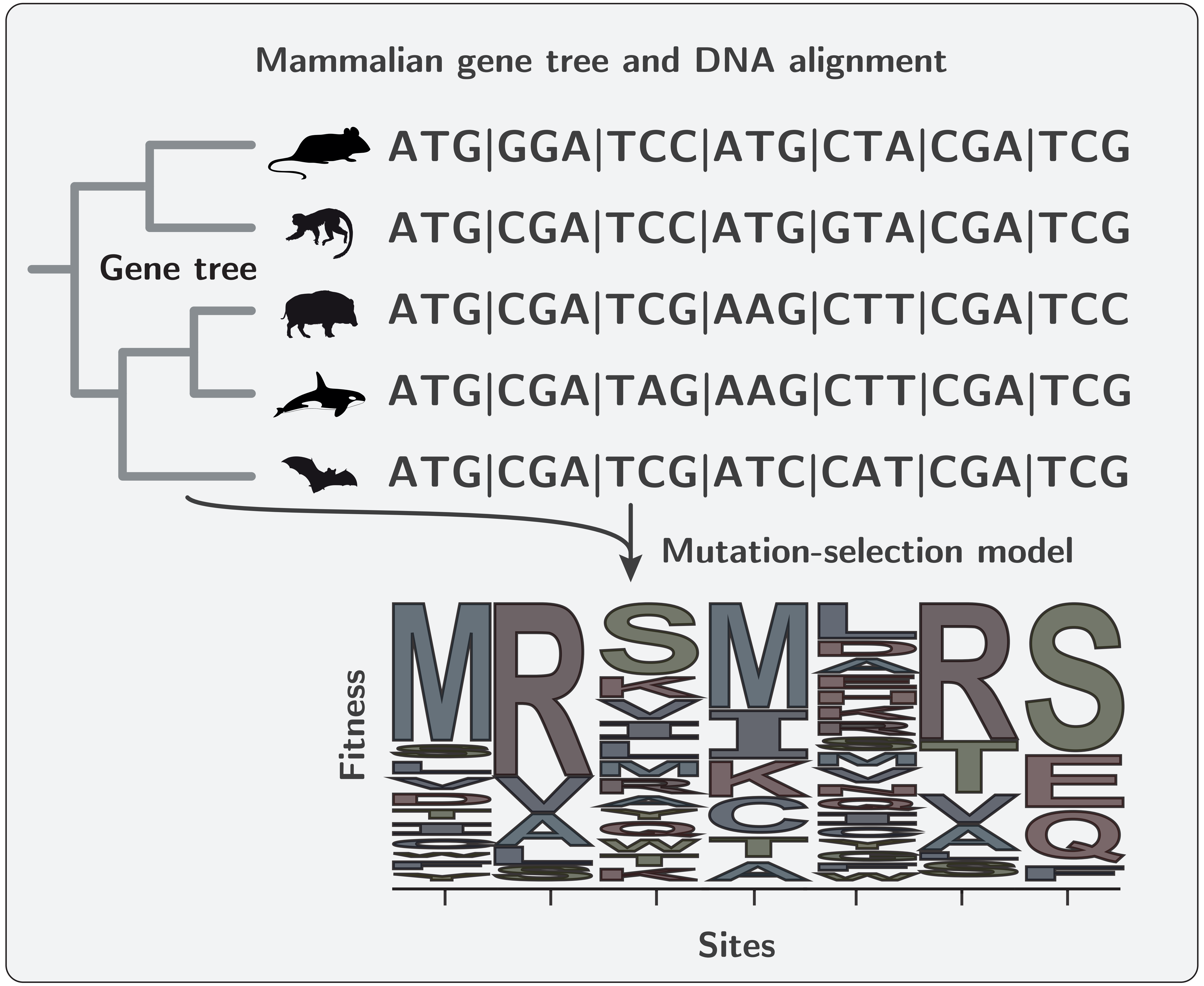
Substitution rate = Mutation rate x Probability of fixation

Kimura (1969), Otha (1982)

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).

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



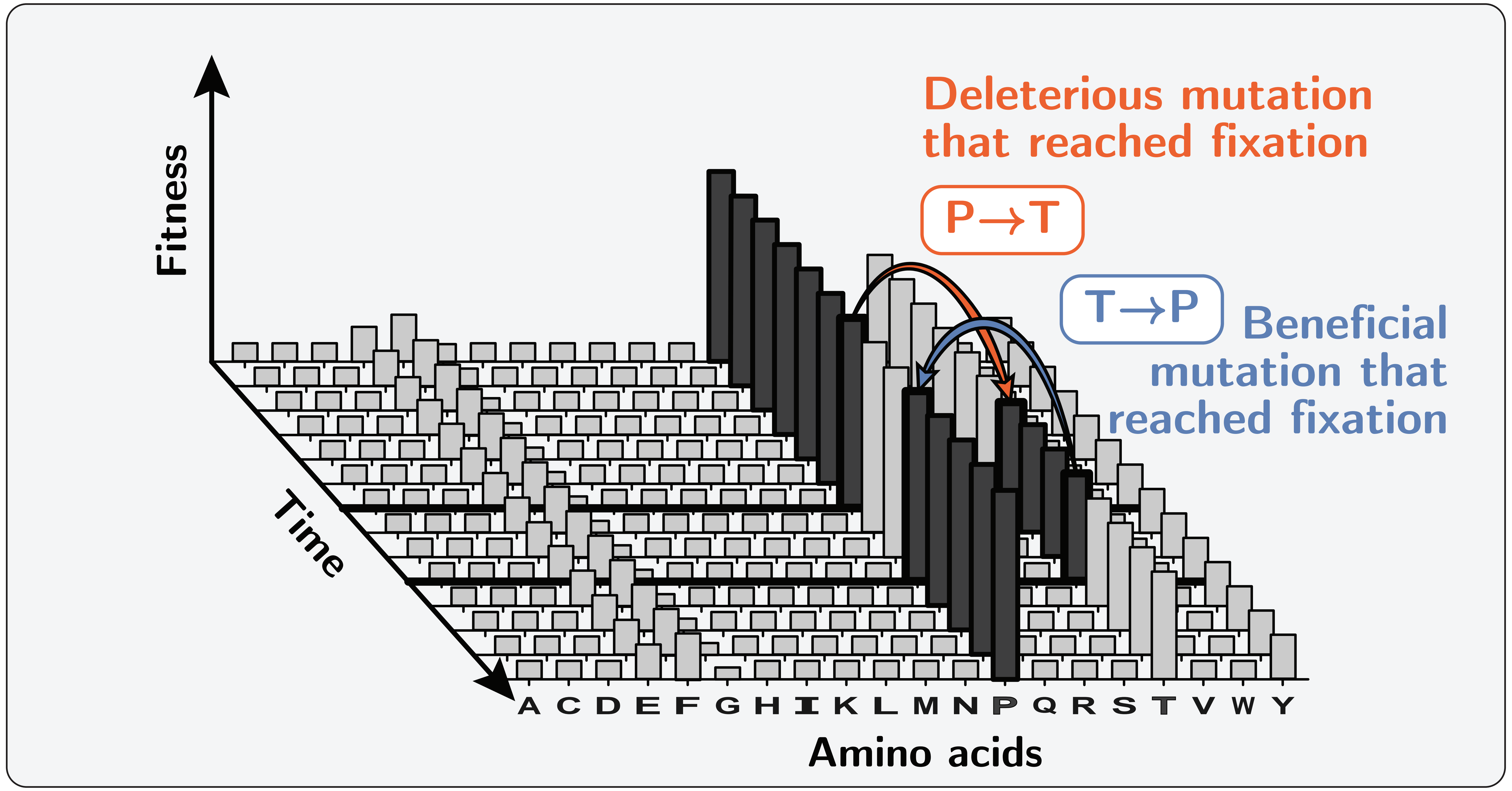
Mammalian gene tree and DNA alignment

ATG|GGA|TCC|ATG|CTA|CGA|TCG
ATG|CGA|TCC|ATG|GTA|CGA|TCG
ATG|CGA|TCG|AAG|CTT|CGA|TCC
ATG|CGA|TAG|AAG|CTT|CGA|TCG
ATG|CGA|TCG|ATC|CAT|CGA|TCG

- **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)

Is a new mutation deleterious or beneficial?

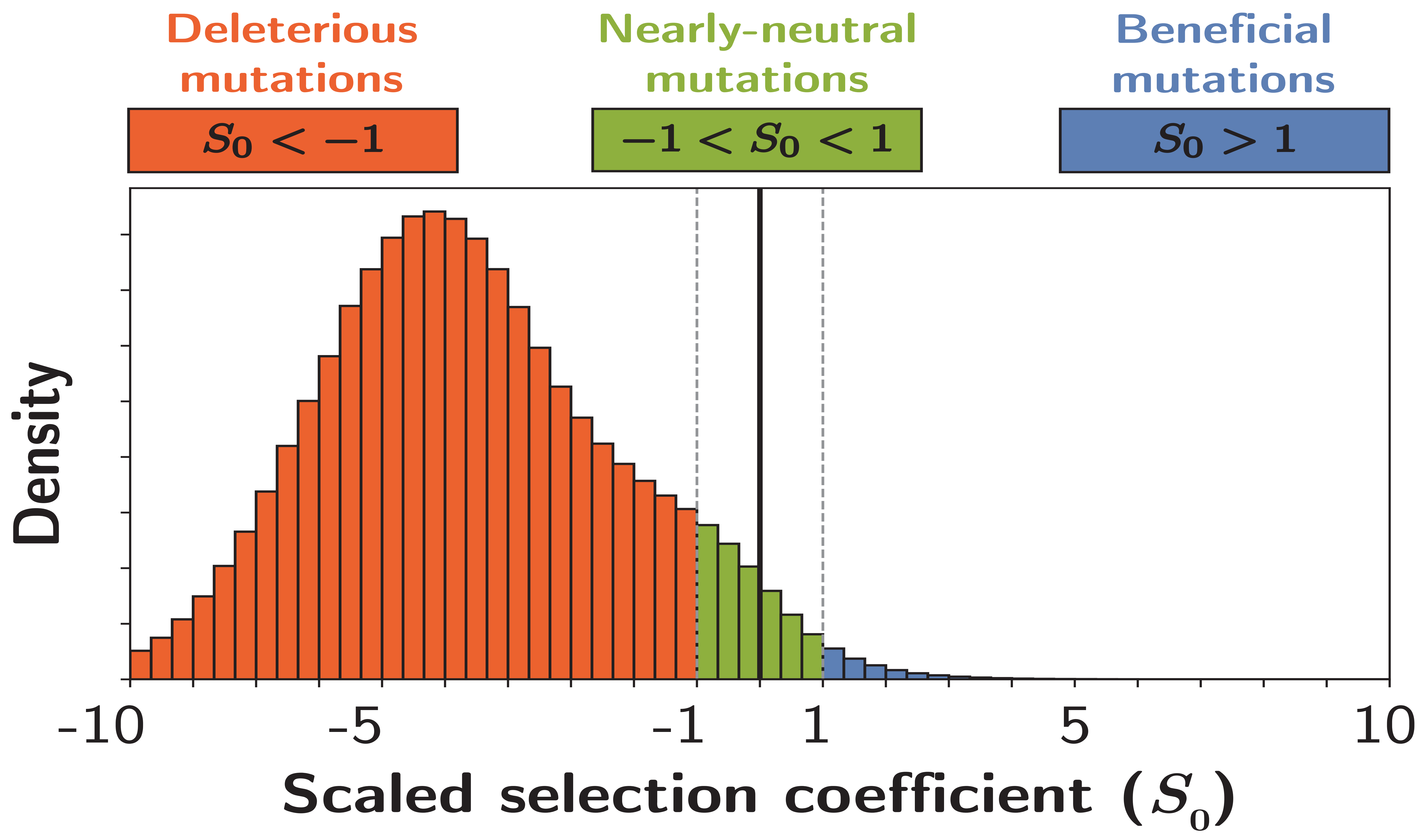


- **Deleterious mutations have negative fitness effects.**
  - » *They can however reach fixation due to genetic drift.*

- **Beneficial mutations have positive fitness effects.**
  - » They bring an advantage to their bearer.

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

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



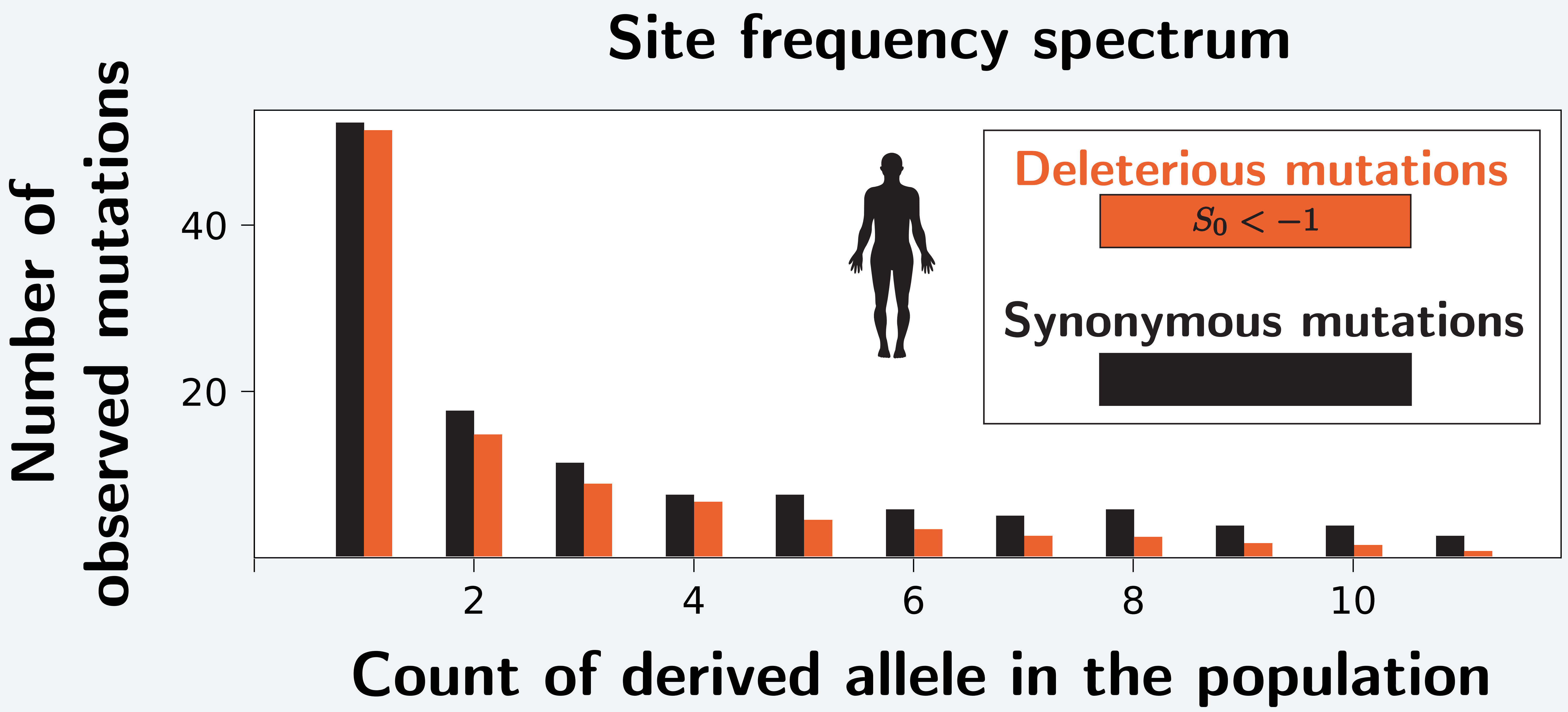
- **Are predicted deleterious mutations purified away in *H. sapiens?***
  » Mutations associated with clinical terms such as likely pathogenic & pathogenic.

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

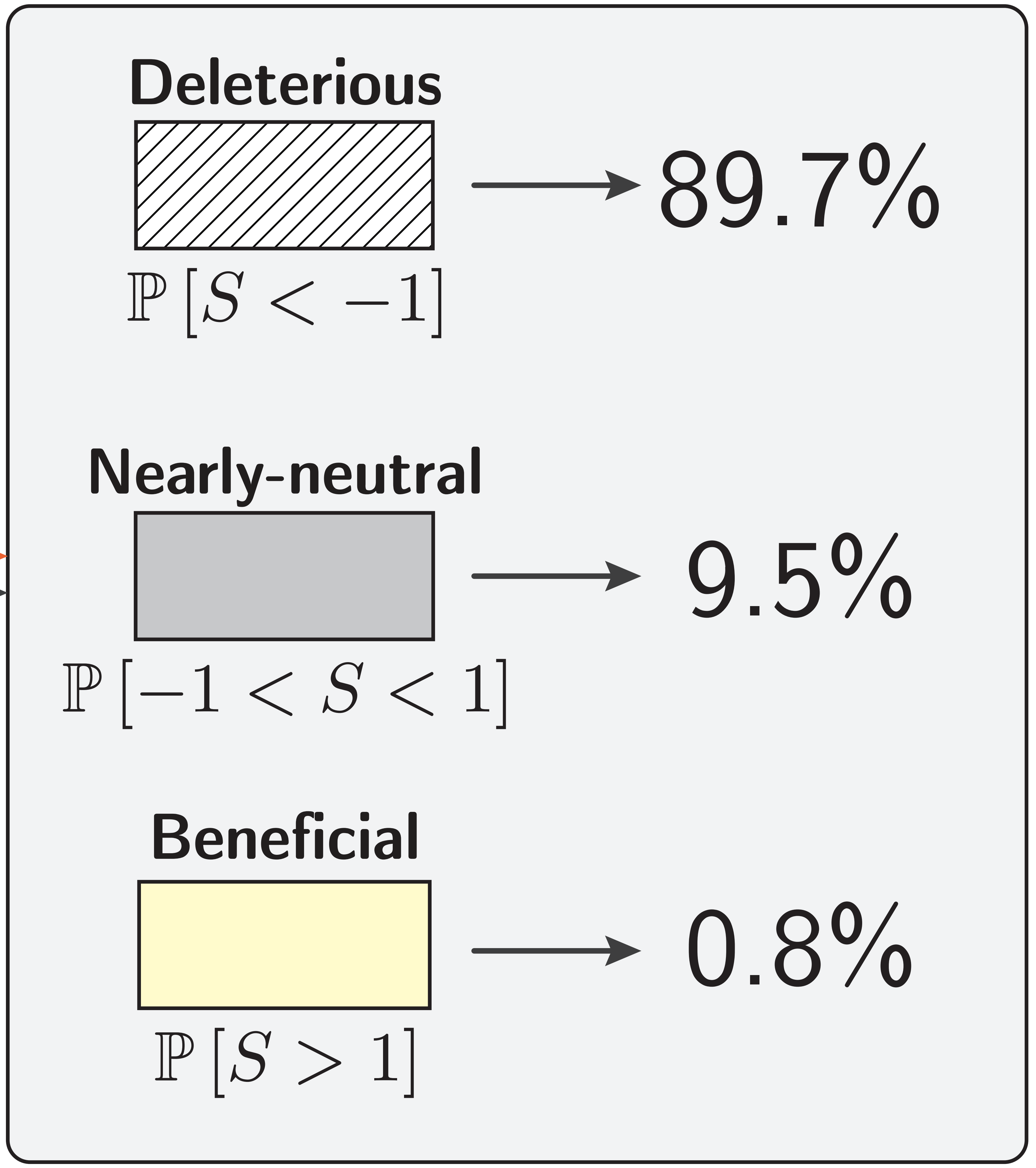
- **Is selection at the phylogenetic and population scale comparable?**

Is selection at the phylogenetic and population scale comparable?

Polymorphism within a species



Estimation of selection at the population-genetic scale(S)



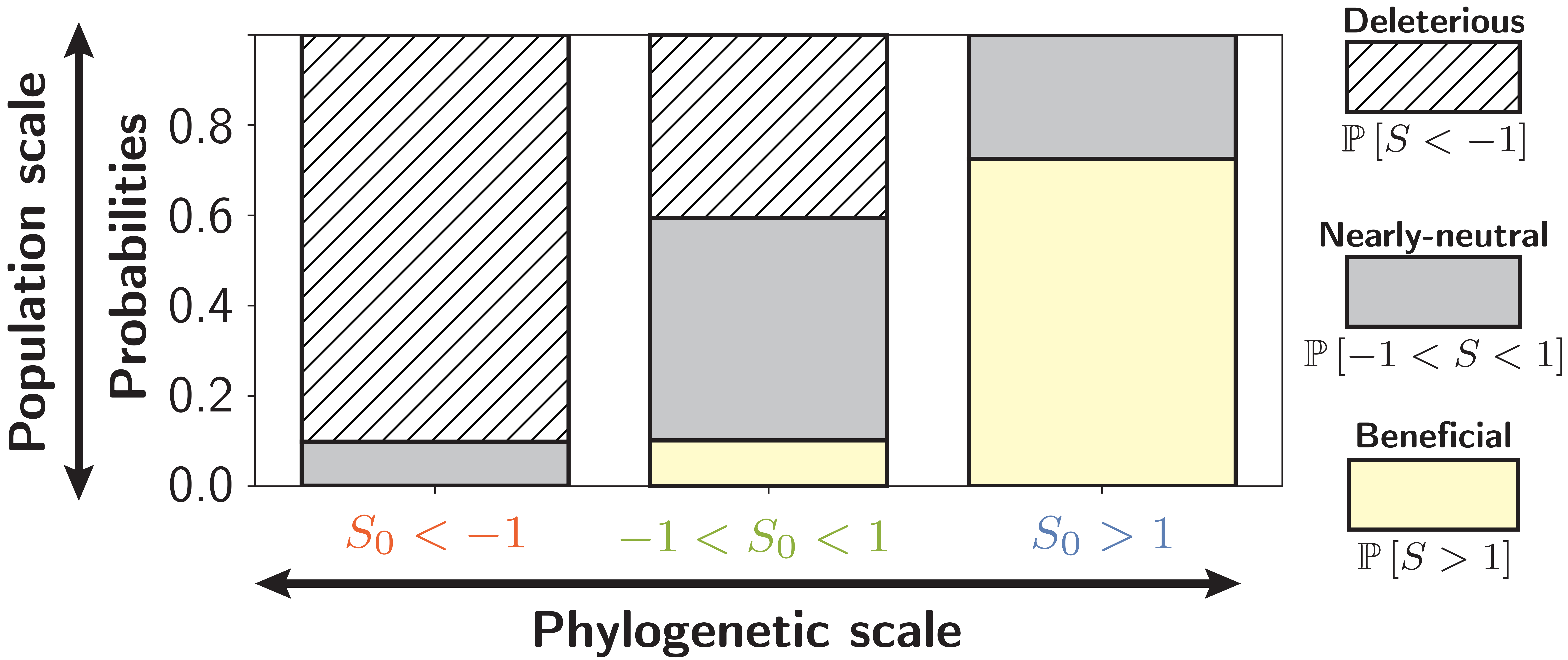
Deleterious ▨ $\mathbb{P}[S < -1]$ → 89.7%

Nearly-neutral $\mathbb{P}[-1 < S < 1]$ → 9.5%

Beneficial $\mathbb{P}[S > 1]$ → 0.8%

Tataru & Bataillon (2020)

Is selection at the phylogenetic and population scale comparable?



- **Predicted deleterious mutations are effectively purified away in *H. sapiens*.**
  » 91% precision (false positives) and 97% recall (false negatives).

- **Predicted beneficial back-mutations are effectively beneficial in *H. sapiens*.**
  » 70% precision (false positives) and 35% recall (false negatives).

Take home messages

- *Theoretical evolution* - **Can we use phylogenetics to predict the selection coefficient of a mutation?**
  - » Mutation-selection codon models formalize a bridge between phylogenetics and population genetics.
- *Evolutionary biology* - **Is the fitness landscape stable across the mammalian evolution?**
  - » The fitness effects as estimated across mammals is predictive of selection in extant populations.
- *Nearly-neutral theory* - **Are deleterious mutations reliably purged only when selection overpowers drift?**
  - » Mutations with predicted to be nearly-neutral are equivalent to neutral mutations.

Acknowledgements

Nicolas Salamin & his Group

UNIL | Université de Lausanne

Julien Joseph & LBBE

UNIVERSITÉ DE LYON

LBBE

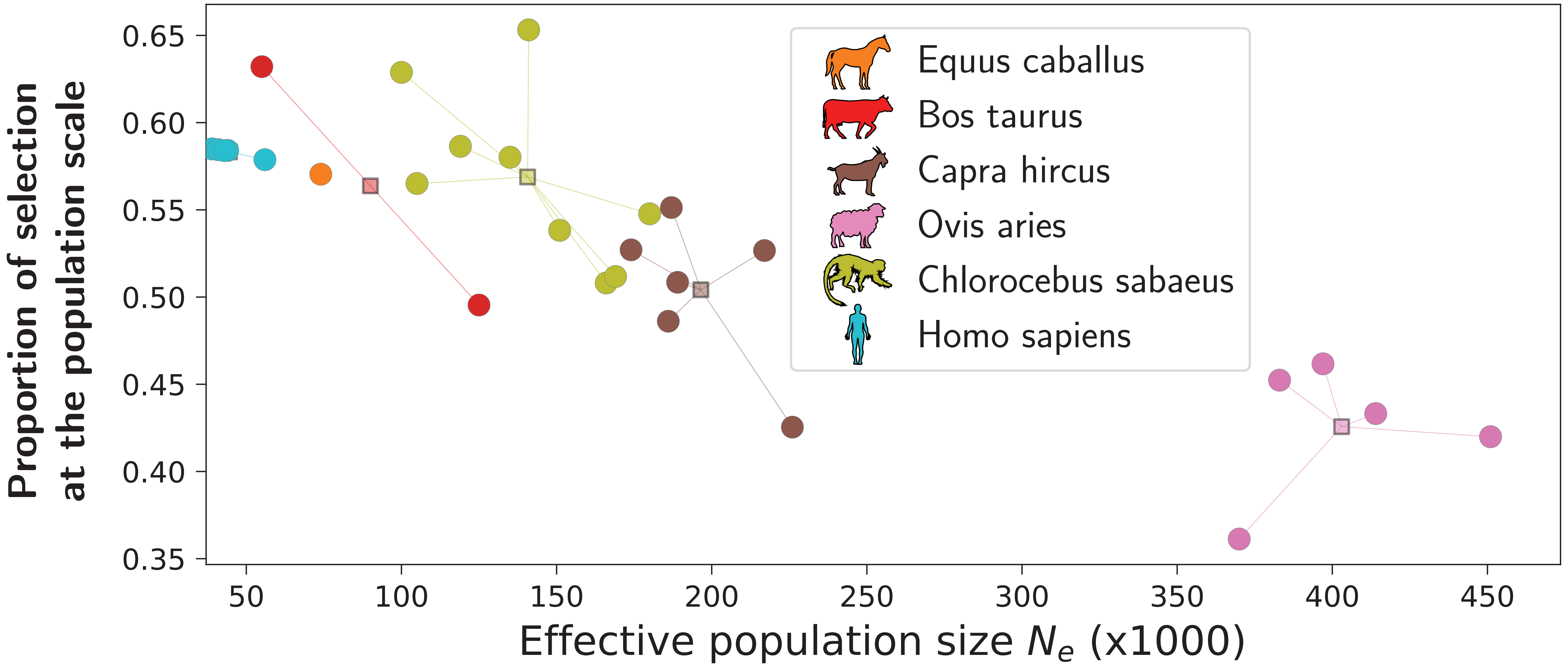
Are our results replicable across populations?

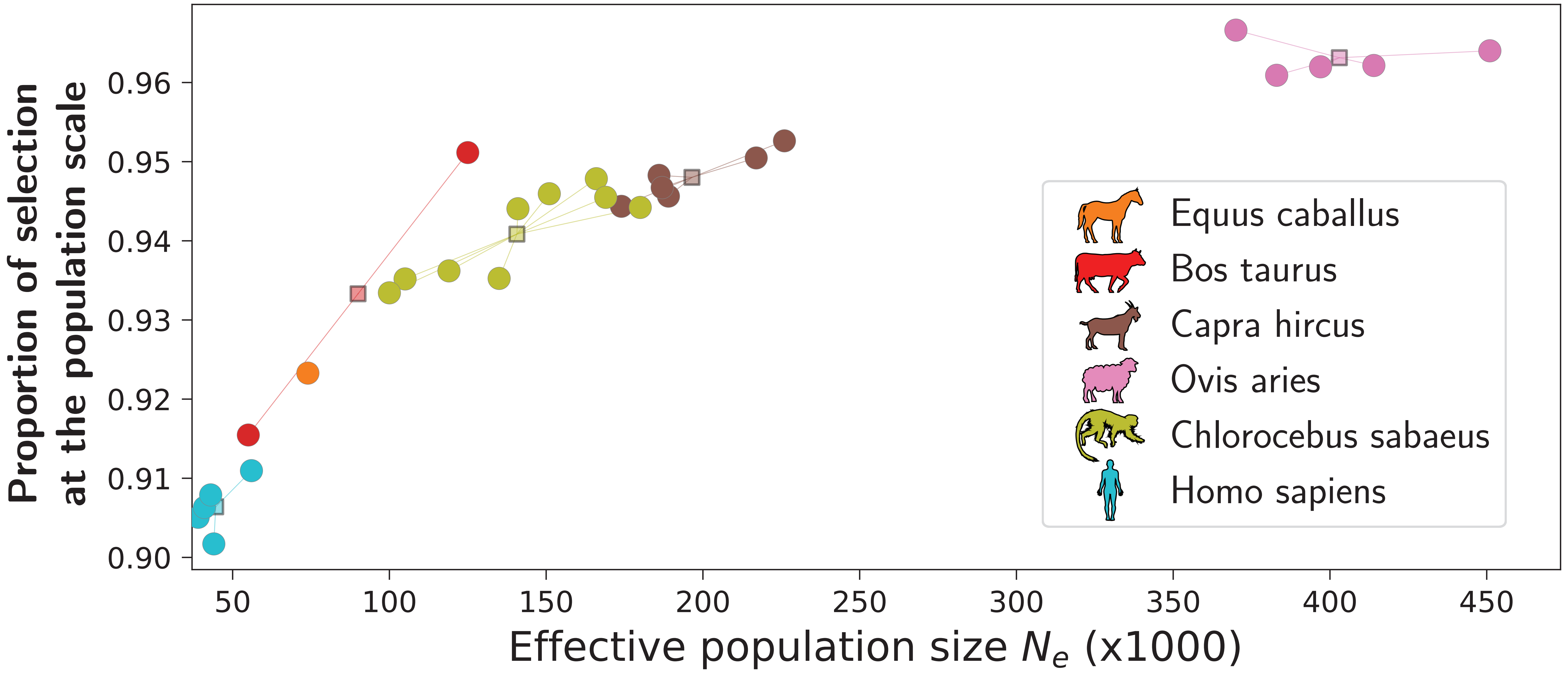
Nearly-neutral mutations

$-1 < S_0 < 1$

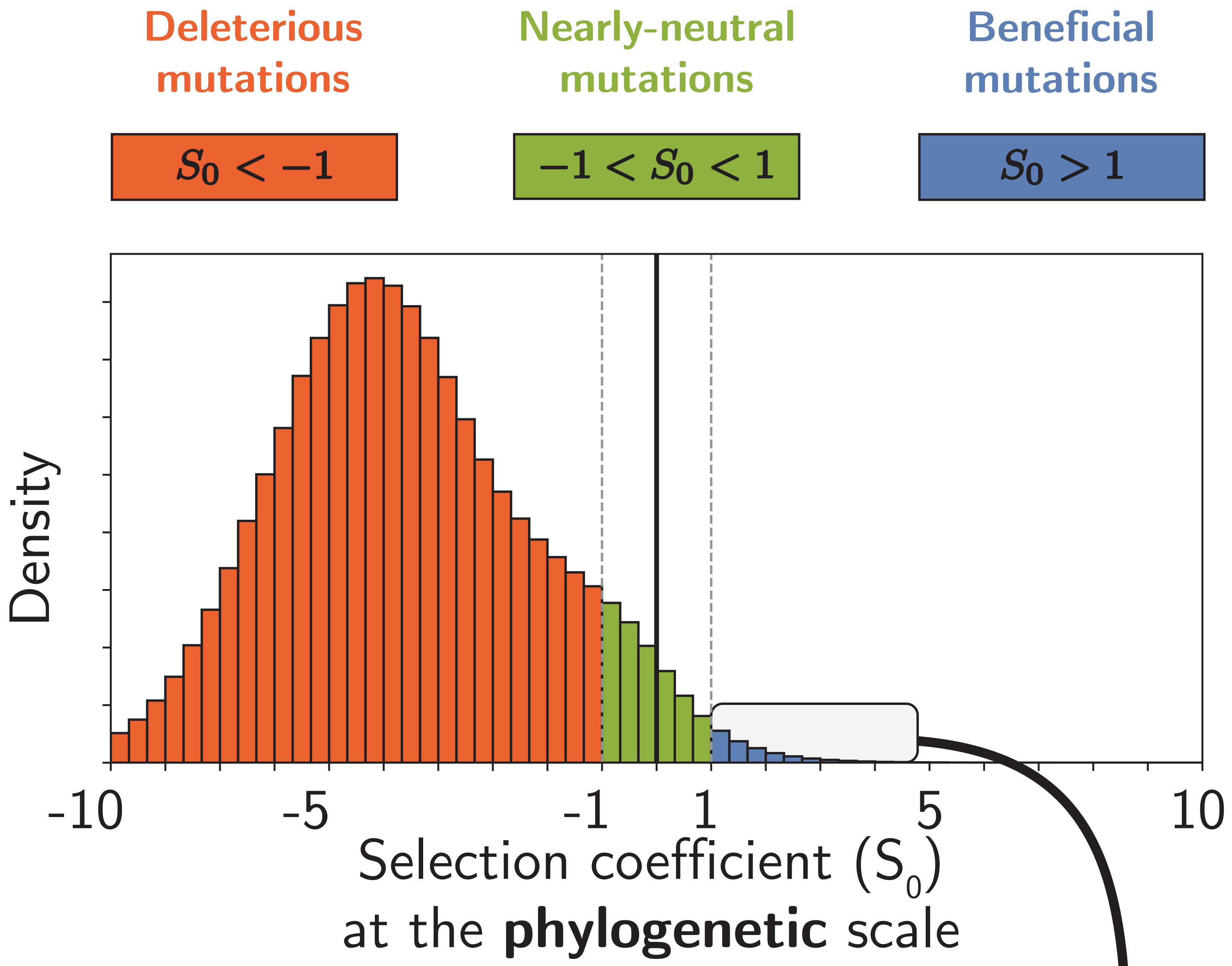
Deleterious mutations

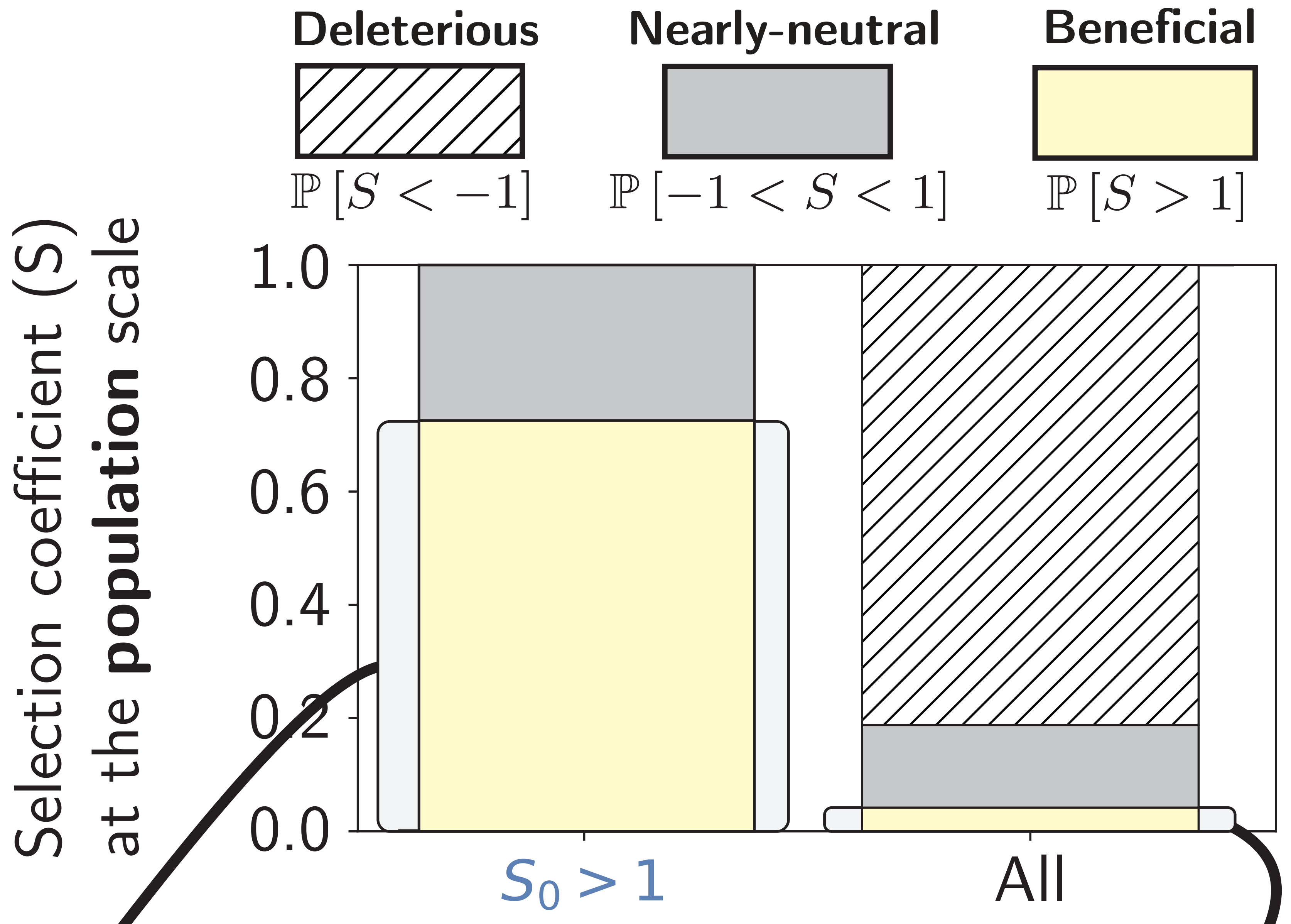
$S_0 < -1$





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





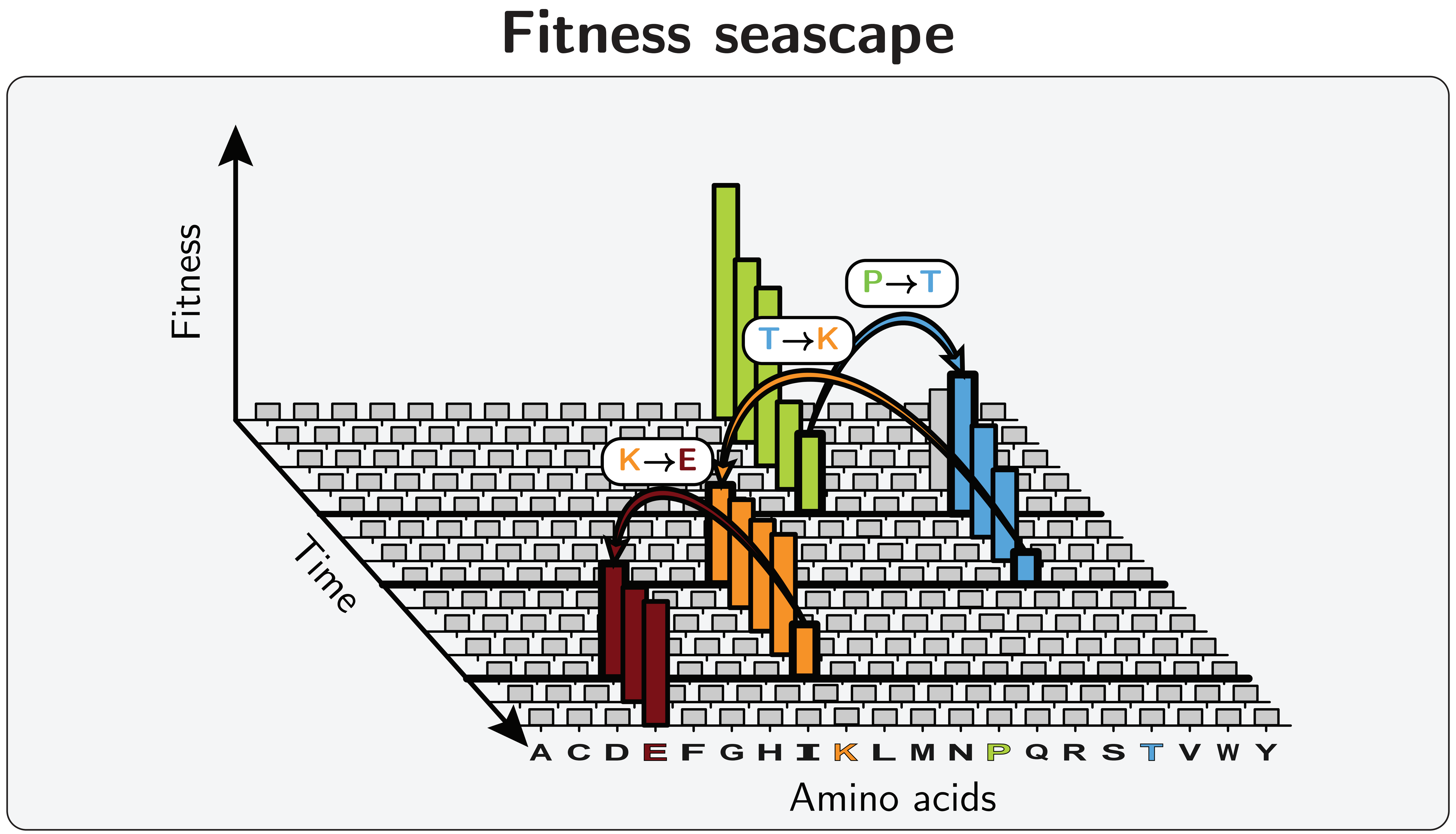
$$\mathbb{P}[S_0 > 1] = 0.010$$

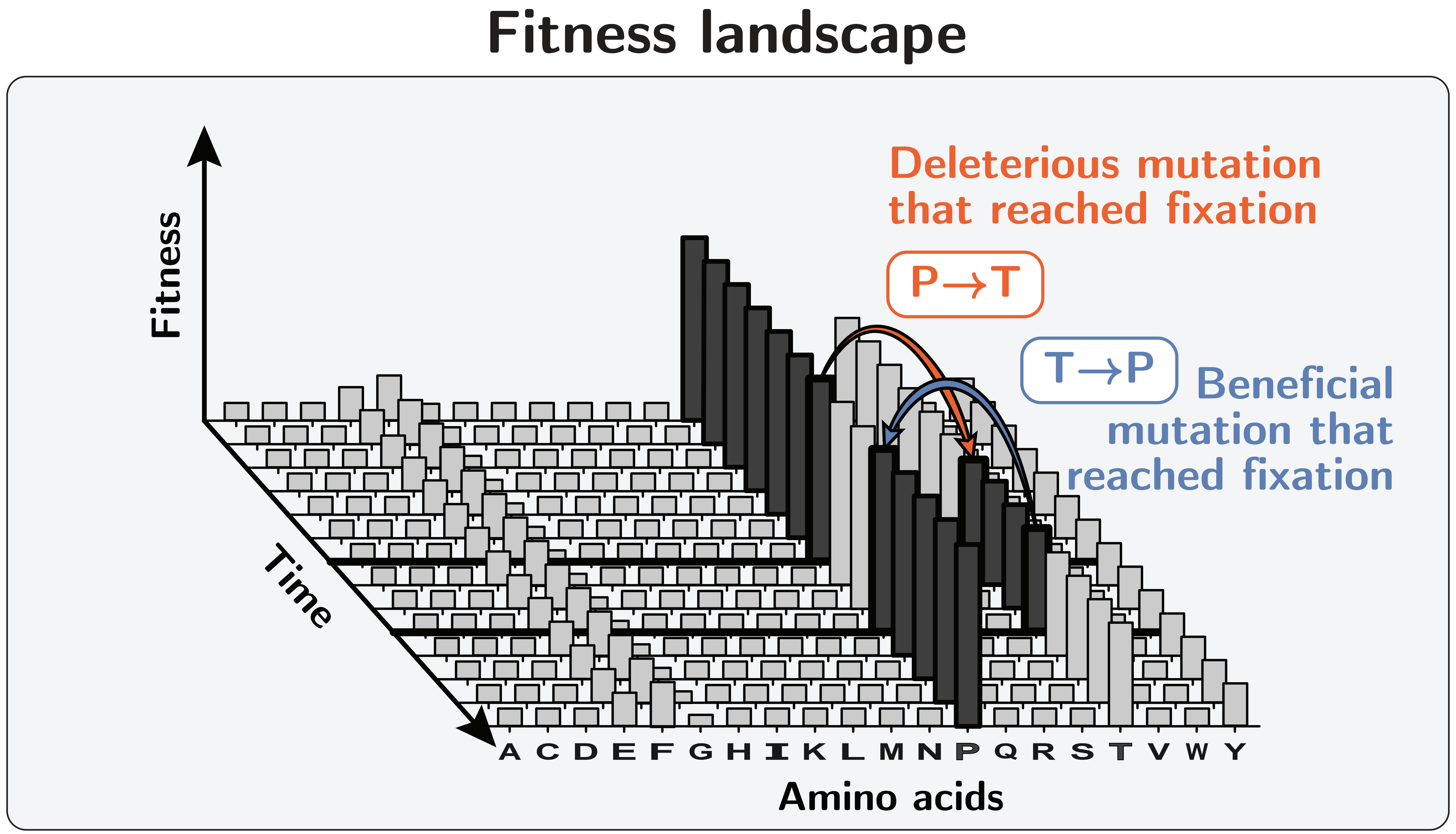
$$\mathbb{P}[S > 1 \mid S_0 > 1] = 0.72$$

$$\mathbb{P}[S > 1] = 0.020$$

$$\mathbb{P}[S_0 > 1 \mid S > 1] = \frac{\mathbb{P}[S > 1 \mid S_0 > 1]\,\mathbb{P}[S_0 > 1]}{\mathbb{P}[S > 1]} = 0.35$$

Fitness landscapes and seascapes



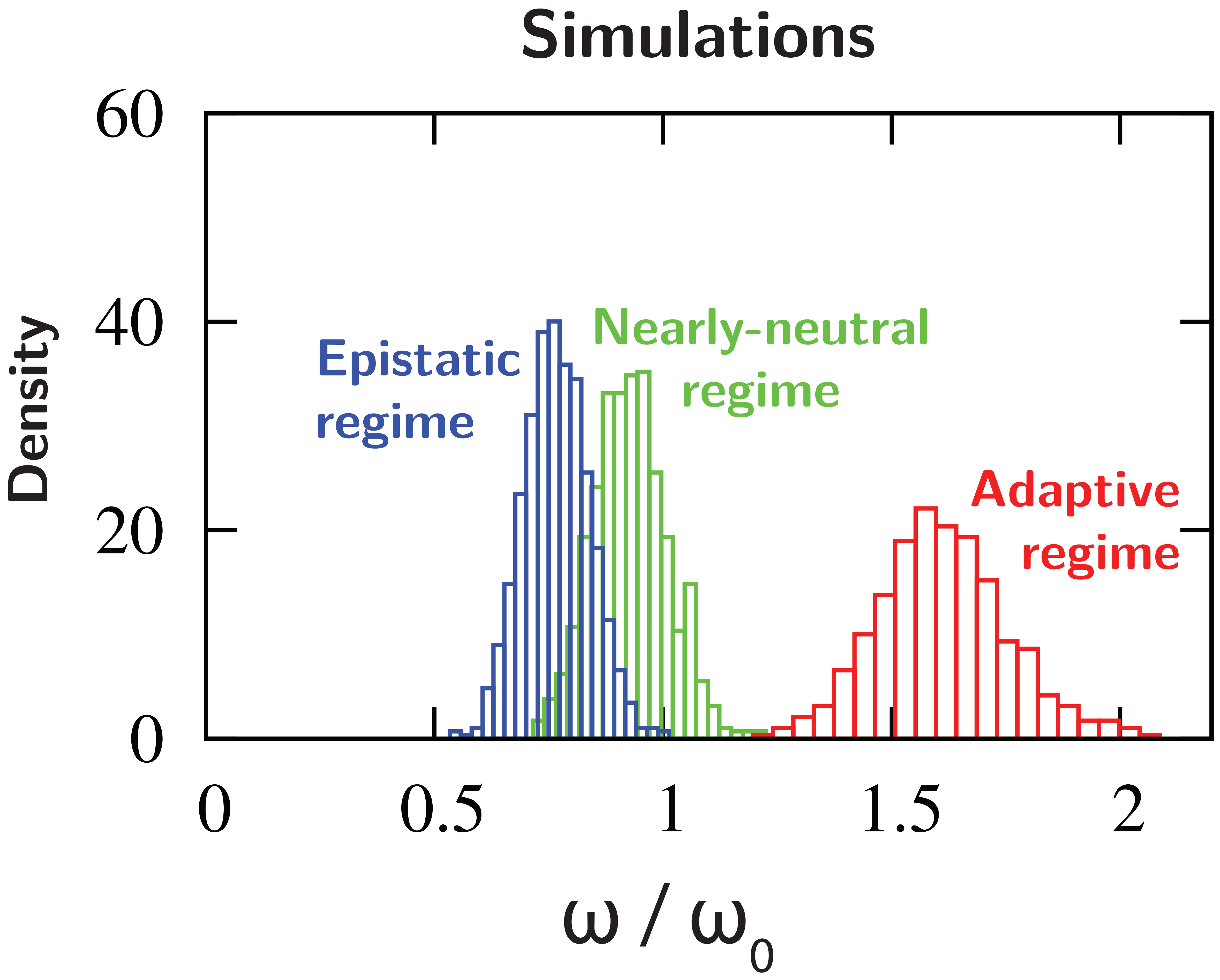


- **The mutation-selection model is null model of evolution without adaptation**

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

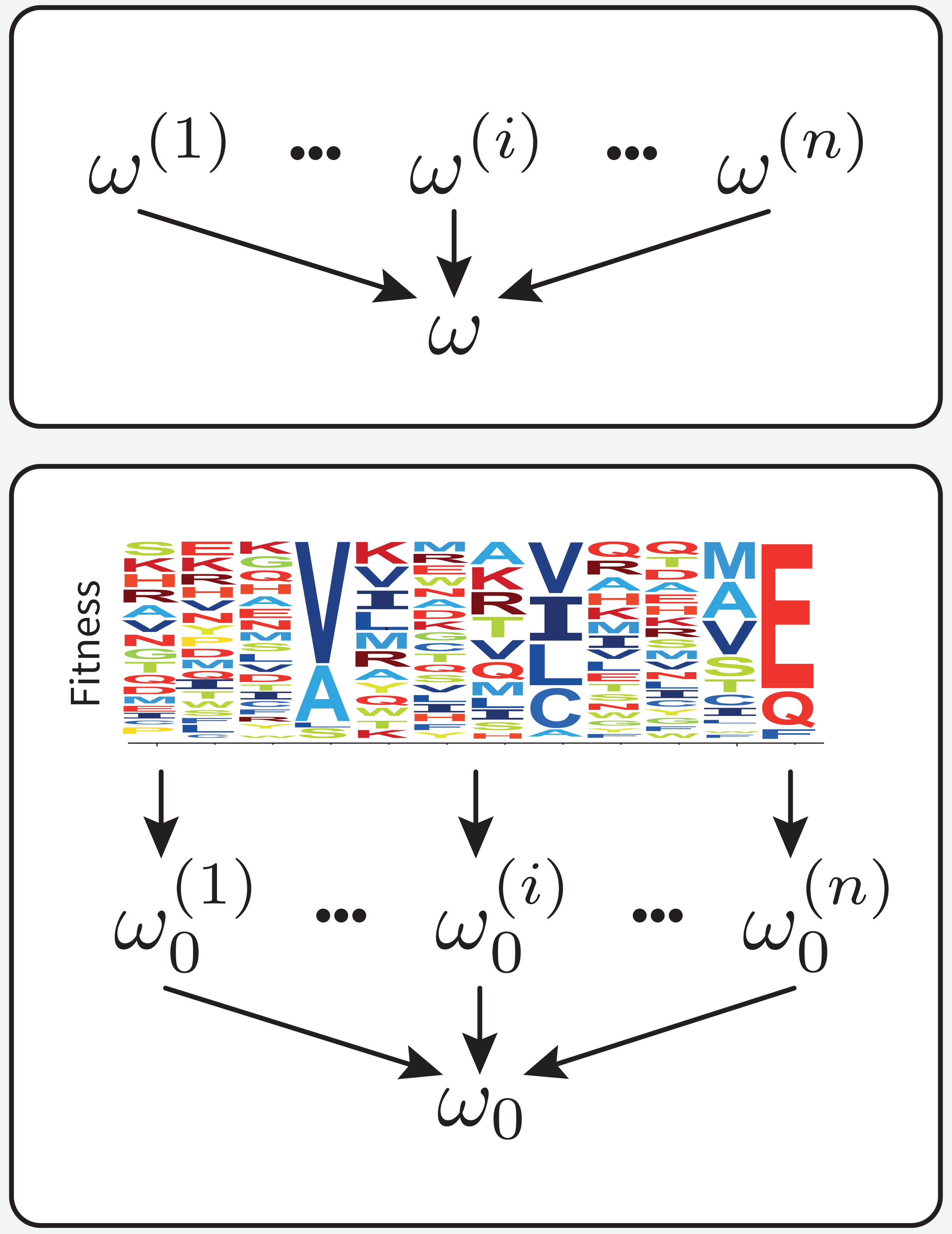
Mutation-selection model tested against simulations

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

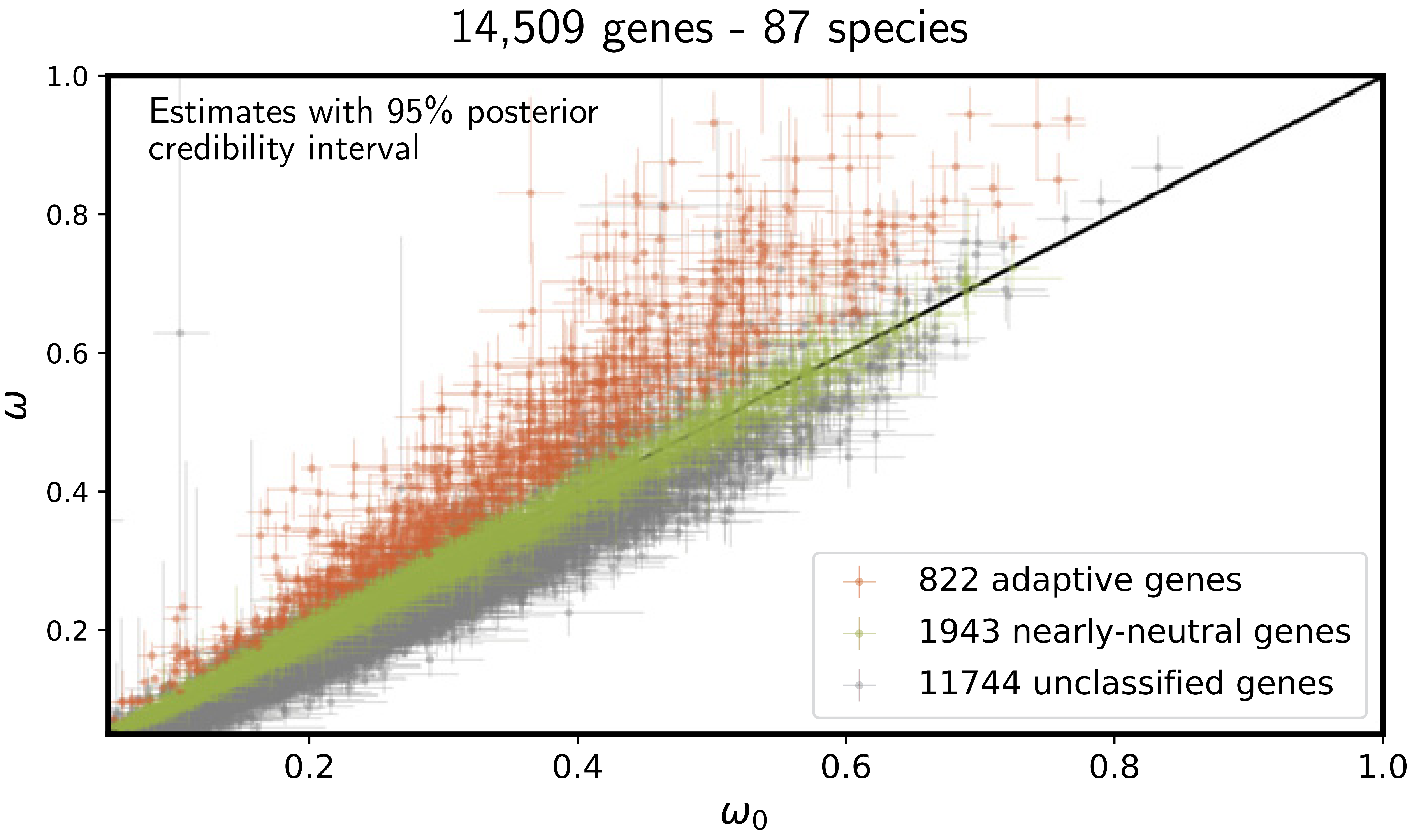


Rodrigue & Lartillot (2017)

Mutation-selection model applied to mammalian genes



Phylogenetic codon models on protein-coding DNA alignments


14,509 genes - 87 species

- **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?

14,509 genes - 87 species

Phylogenetic codon models on protein-coding DNA alignments

Population-genetics rate of adaptive evolution

- **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.**