An improved codon modeling approach for accurate estimation of the mutation bias

SMBE everywhere 2022

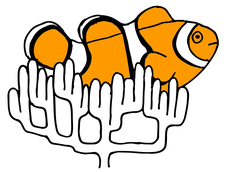
GS3: Mutational Biases and Adaptation

August 2, 2022

T. Latrille[1,2], N. Lartillot[2]

[1]Université de Lausanne (Switzerland)

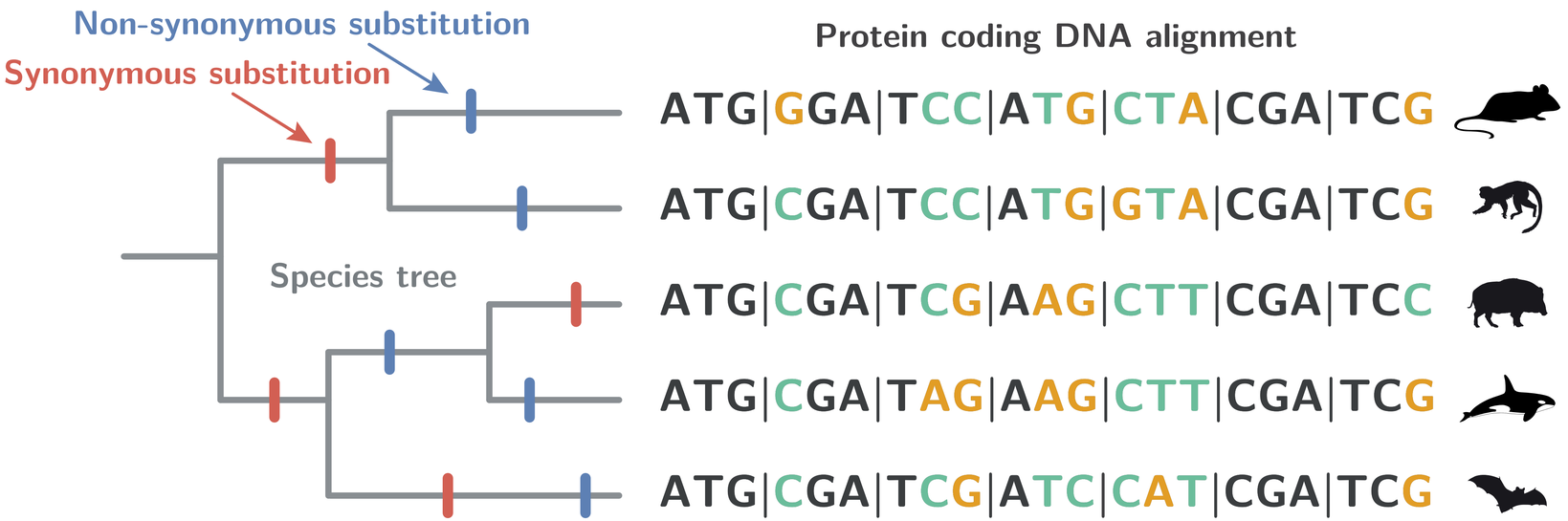
[2]Université de Lyon (France)



thibault.latrille@unil.ch

- **I. Can codon models accurately estimate selection?**
- **II. Can codon models accurately estimate mutation?**
- **III. Can codon models disentangle mutation and selection?**

Codon models take advantage of the genetic code



- **Non-synonymous substitutions are reflecting the effect of mutation, selection and drift.**
- **Synonymous substitutions are considered selectively neutral, reflecting the mutational processes.**
- **Contrasting non-synonymous and synonymous substitution rates allows estimating the strength of selection exercised on proteins.**

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

Phylogenetic codon models

- $Q_{i,j}$ is the substitution rate from codon i to j.

$$\begin{cases} Q_{i,j} = \mu_{i,j} \text{ if codons } i \text{ and } j \text{ are synonymous} \\ Q_{i,j} = \omega\, \mu_{i,j} \text{ if codon } i \text{ and } j \text{ are non-synonymous.} \end{cases}$$

ω

Scaling factor exerced on non-synonymous mutations

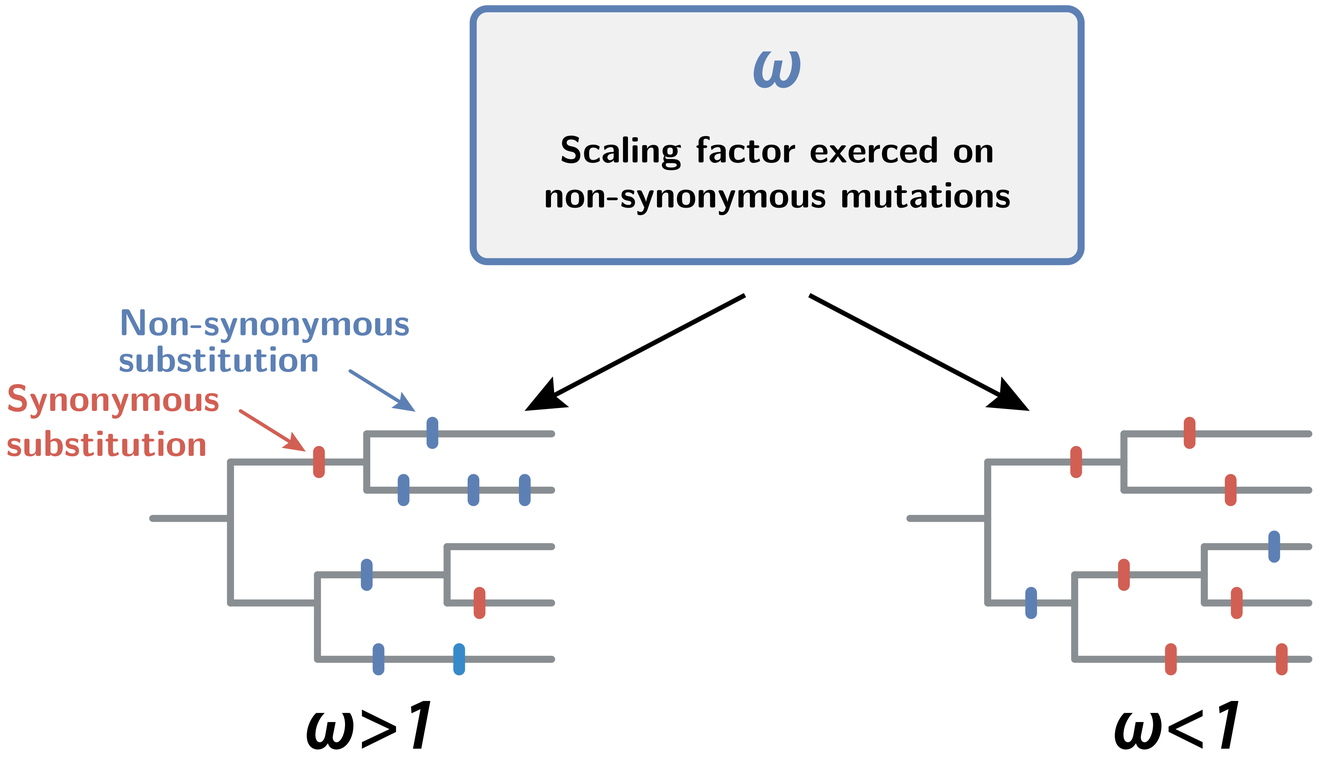
$\mu_{i,j}$

Mutation rates between nucleotides

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

Goldman & Yang (1994); Muse & Gaut (1994); Rodrigue *et al* (2008).

Phylogenetic codon models



- **Detecting fast evolving genes.**

→ Kosiol *et al* (2008).

- **Detecting rapidly changing sites.**

→ Nieslen & Yang (1998); Enard *et al* (2016).

- **Decting burst of evolution.**

→ Yang & Nielsen (1998); Zhang & Nielsen (2005).

- **Stronger selection for highly expressed proteins.**

→ Drummond (2005); Zhang & Yang (2015).

- **More constrains for buried sites inside a protein.**

→ Ramsey *et al* (2011); Echave *et al* (2016).

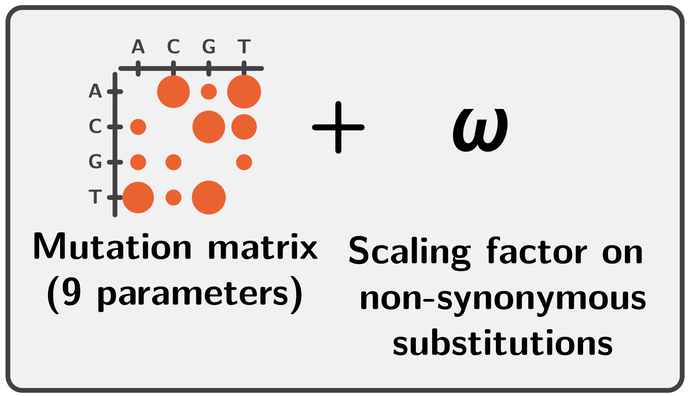
- **Weaker selection for long-lived and bigger species.**

→ Popadin *et al* (2007); Lanfear et al (2010).

Mutation and selection are modelled separately in codon models

Alignment of coding sequence

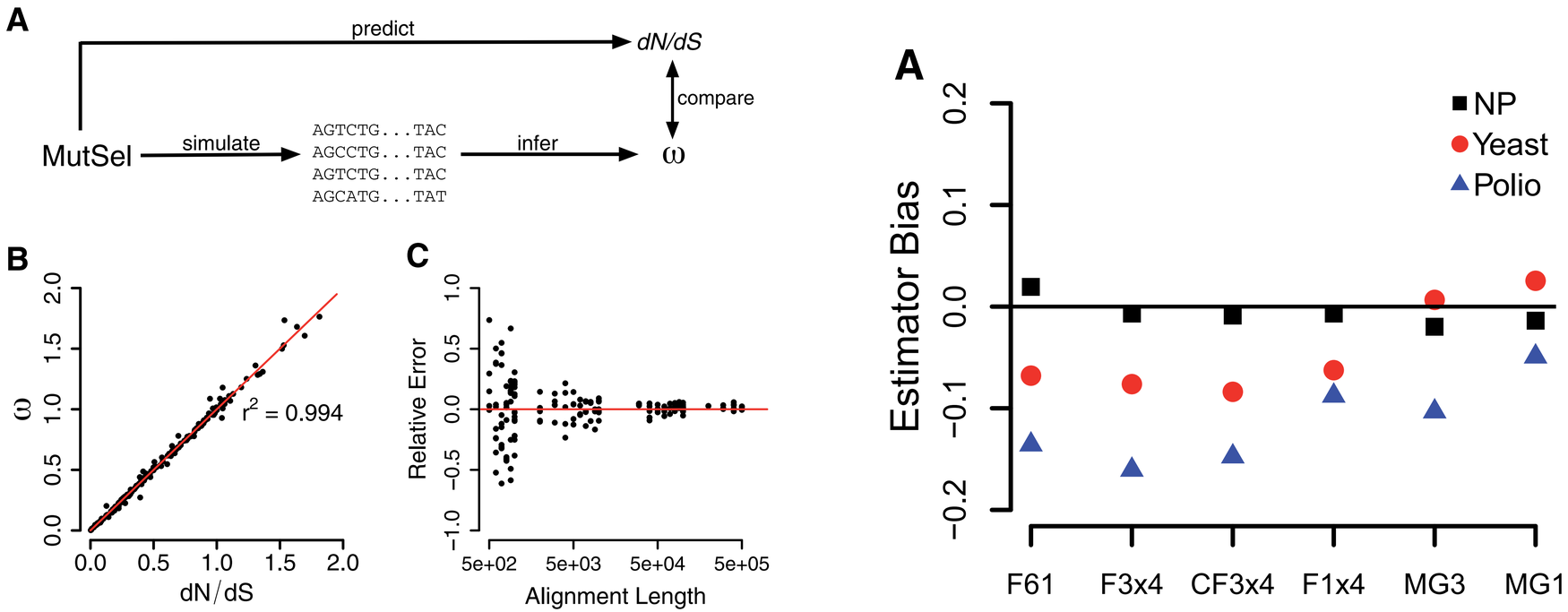
ATG|GGA|TCC|ATG|CTA|CGA|TCG
ATG|CGA|TCC|ATG|GTA|CGA|TCG
ATG|CGA|TCG|AAG|CTT|CGA|TCC →
ATG|CGA|TAG|AAG|CTT|CGA|TCG
ATG|CGA|TCG|ATC|CAT|CGA|TCG



- Codon models seek to capture mutation at the level of nucleotide and selection at the level of amino-acids.
- Can codon models disentangle mutation and selection?

Part I.
Can codon models accurately estimate selection?

Muse & Gaut codon model reliably estimate selection

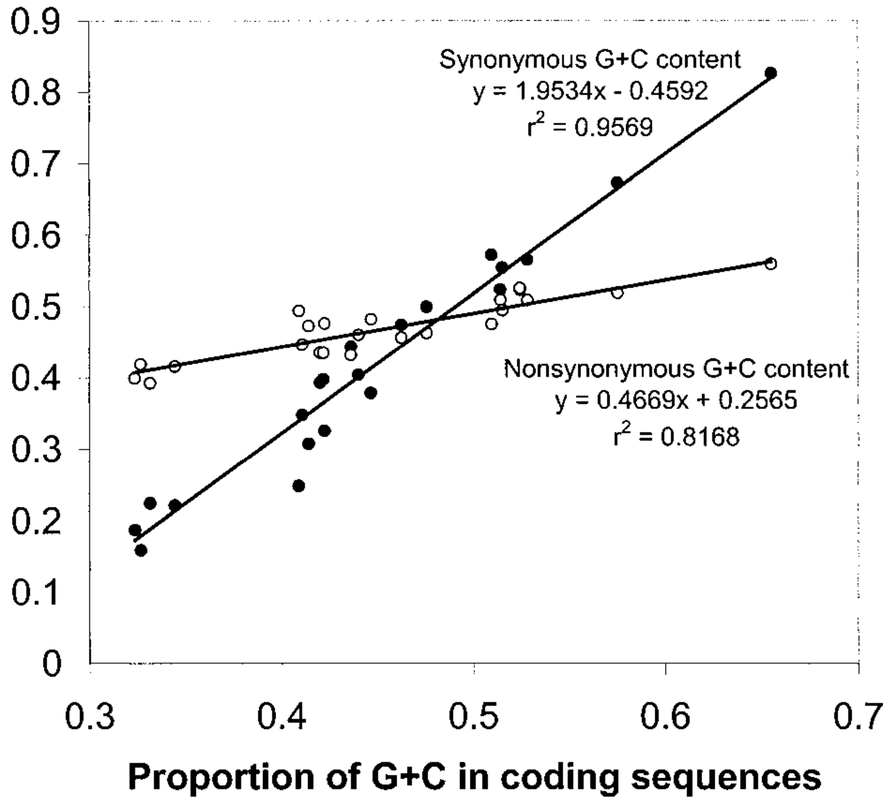


- **The Muse & Gaut (MG) model is the most accurate to infer selection.**

- **MG model predicts that the observed bias in nucleotide composition is equal to the underlying mutation bias.**

Spielman & Wilke (2015) - The relationship between dNdS and scaled selection coefficients - MBE

Observed bias in the nucleotide composition is not the underlying mutation bias

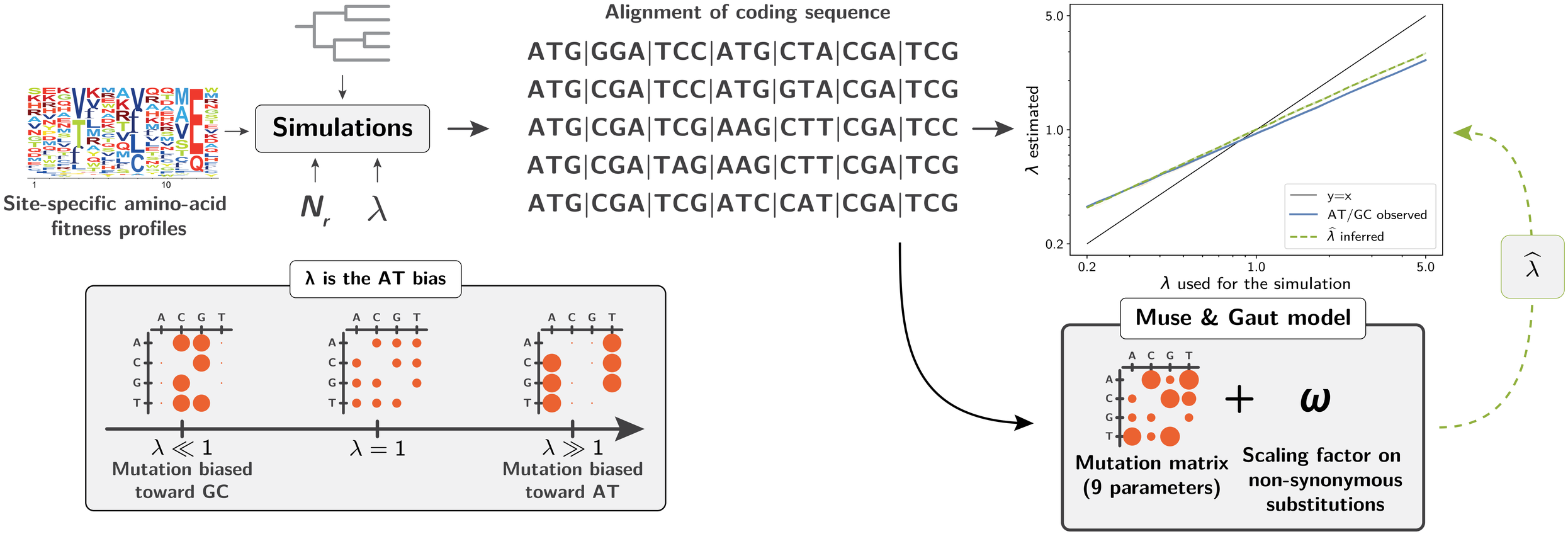


- **Then, how does the MG model performs under a mutation bias?**

Singer and Hickey (2000) - Nucleotide bias causes a genomewide bias in the amino-acid composition of proteins - MBE

Part II.
Can codon models accurately estimate mutation?

Codon models do not accurately estimate the mutation bias

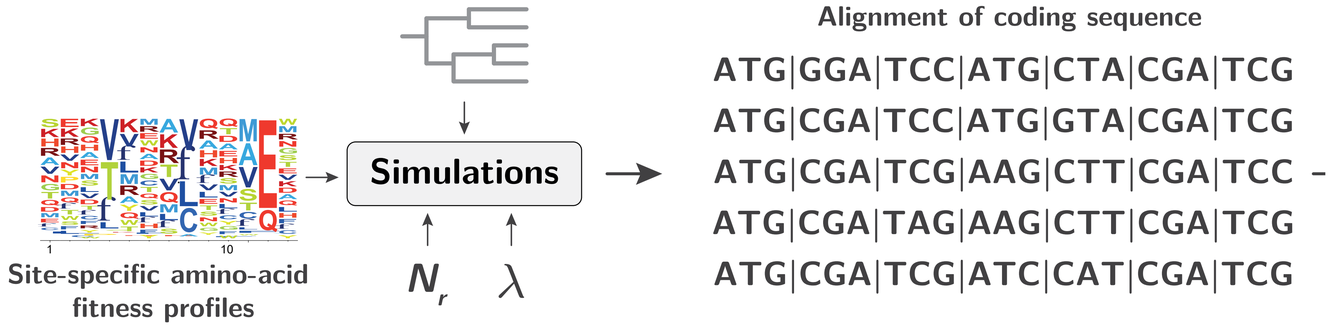


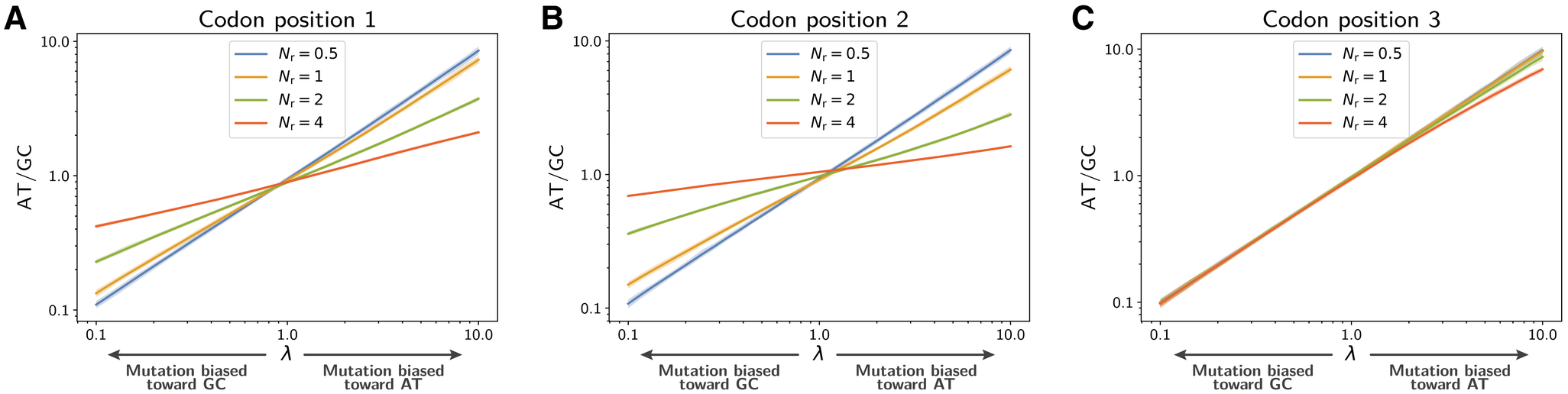
- **Codon models struggle between true mutation bias and observed bias in nucleotide composition.**

Latrille & Lartillot (2022) - An improved codon modeling approach for accurate estimation of mutation bias - MBE

Part III.
Can we construct codon models to disentangle mutation and selection?

Codon positions have different biases in observed nucleotide compositions

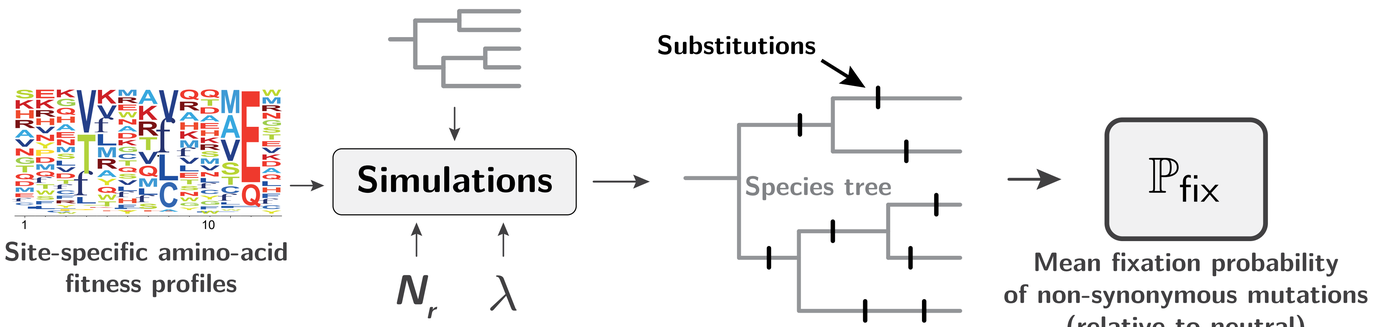




- **Are mutation selected in different direction?**

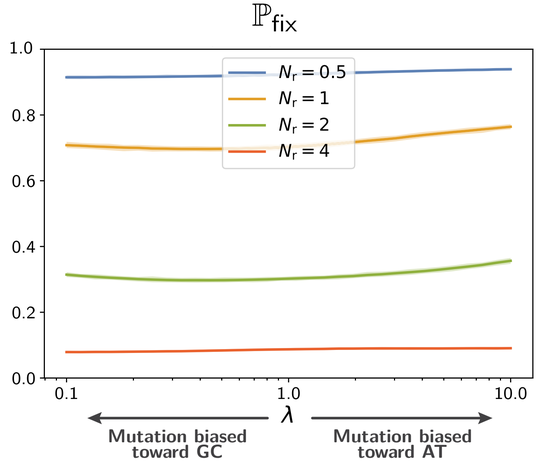
Latrille & Lartillot (2022) - An improved codon modeling approach for accurate estimation of mutation bias - MBE

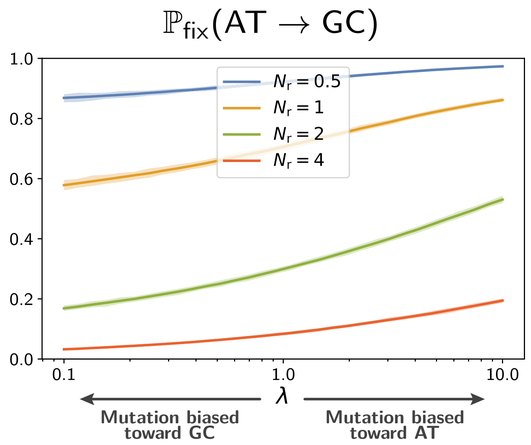
Selection is opposed to the mutational bias

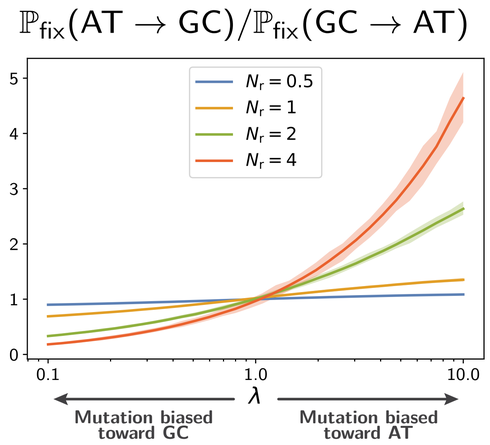


Cartoon representation.

Simulations, 5000 sites and 498 species.







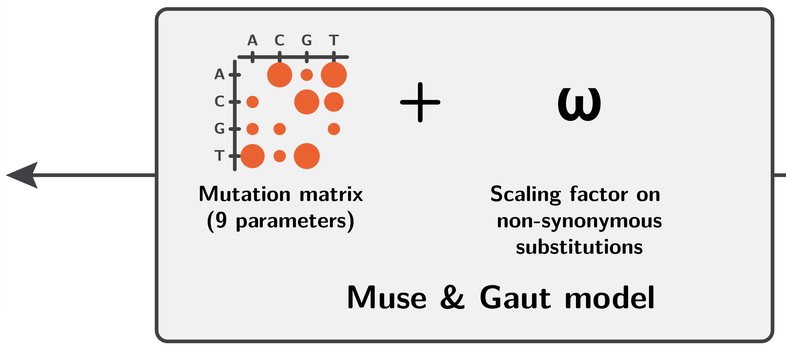
- **We need codon models with selection in different directions.**

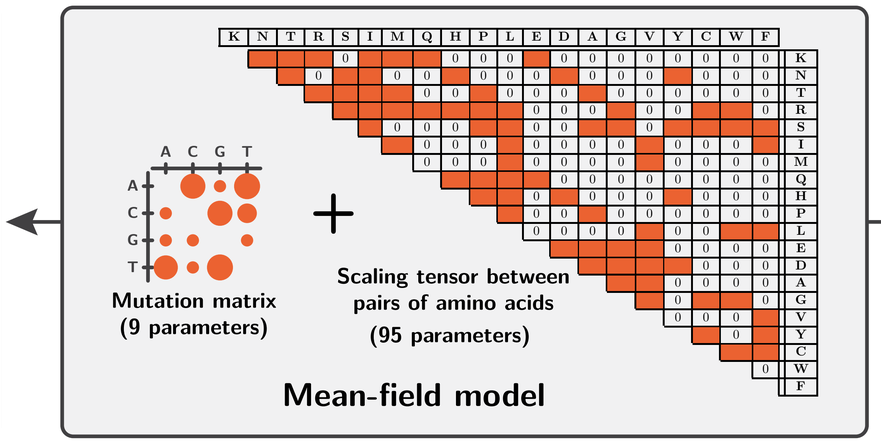
Latrille & Lartillot (2022) - An improved codon modeling approach for accurate estimation of mutation bias - MBE

Modelling selection in different directions allows to accurately infer mutation biases

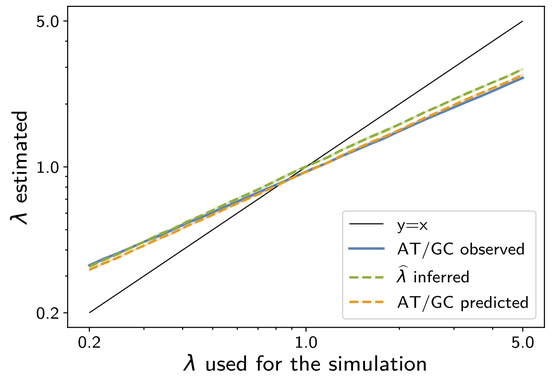
Empirical experiments

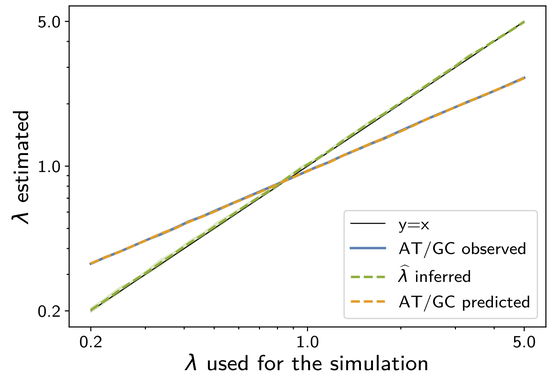
Influenza Nucleoprotein 498 sites, 180 strains	*E-coli* Lactamase 263 sites, 85 strains
$\widehat{\lambda}$=1.39	$\widehat{\lambda}$=0.85
$\widehat{\omega}$=0.085	$\widehat{\omega}$=0.29
$\widehat{\lambda}$=1.64	$\widehat{\lambda}$=0.68
$\widehat{\omega}$=0.086	$\widehat{\omega}$=0.30
$\widehat{\omega}_{AT\to GC}$=0.14	$\widehat{\omega}_{AT\to GC}$=0.31
$\widehat{\omega}_{GC\to AT}$=0.10	$\widehat{\omega}_{GC\to AT}$=0.44
$\widehat{\omega}_{AT\to GC}/\widehat{\omega}_{GC\to AT}$=1.36	$\widehat{\omega}_{AT\to GC}/\widehat{\omega}_{GC\to AT}$=0.71





Simulated experiments





Latrille & Lartillot (2022) - An improved codon modeling approach for accurate estimation of mutation bias - MBE

Conclusion

- **Muse & Gaut codon model with a single parameter of selection is accurate, although it does not reliably estimate mutational biases.**
- **Muse & Gaut codon model should be used to estimate selection.**
- **Mutational bias is balanced by a fixation bias (selection) in the opposite direction.**
- **Should not be confused with GC-biased gene conversion.**
- **Inference of mutational bias requires to model fixation bias in different directions.**
- **How can we measure the load generated by mutation bias and GC-biased gene conversion?**