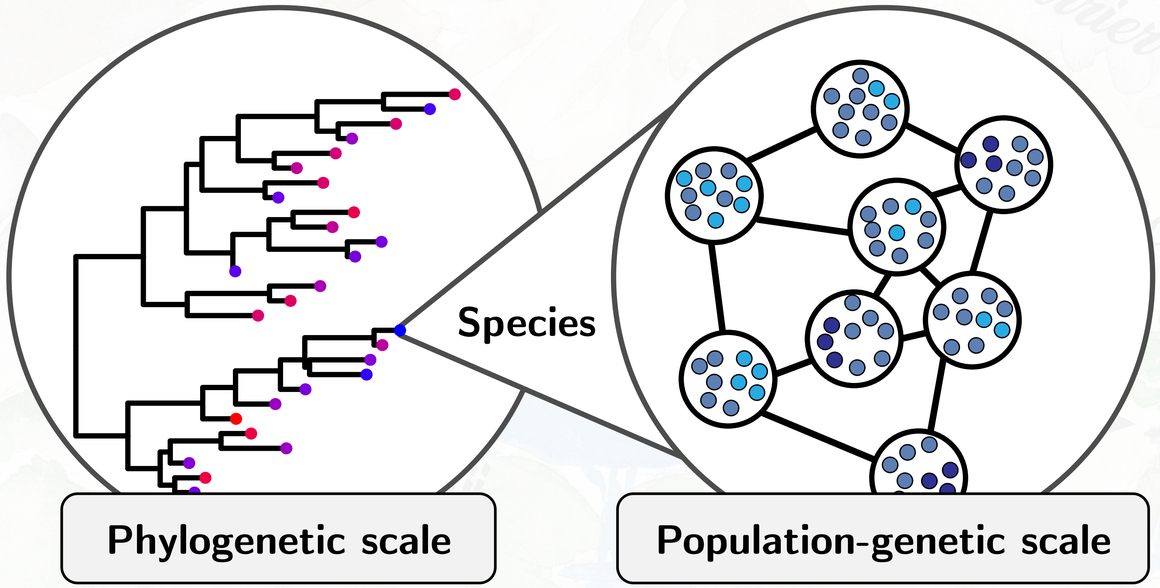
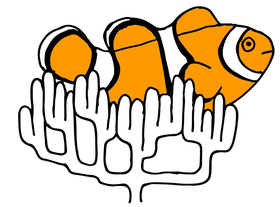
Predicting Selection on Traits and Sequences:
Contrast Across Evolutionary Scales



Thibault Latrille

Université de Lausanne

Department of Ecology & Evolution



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

Check for updates

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

Equal contribution: DAH & TL

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

T. Latrille[1], M. Bastian[2], T. Gaboriau[1], N. Salamin[1]

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

Thibault Latrille[*,1,2] Vincent Lanore,[1] and Nicolas Lartillot[1]

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

Thibault Latrille[*,1,2] and Nicolas Lartillot[1]

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

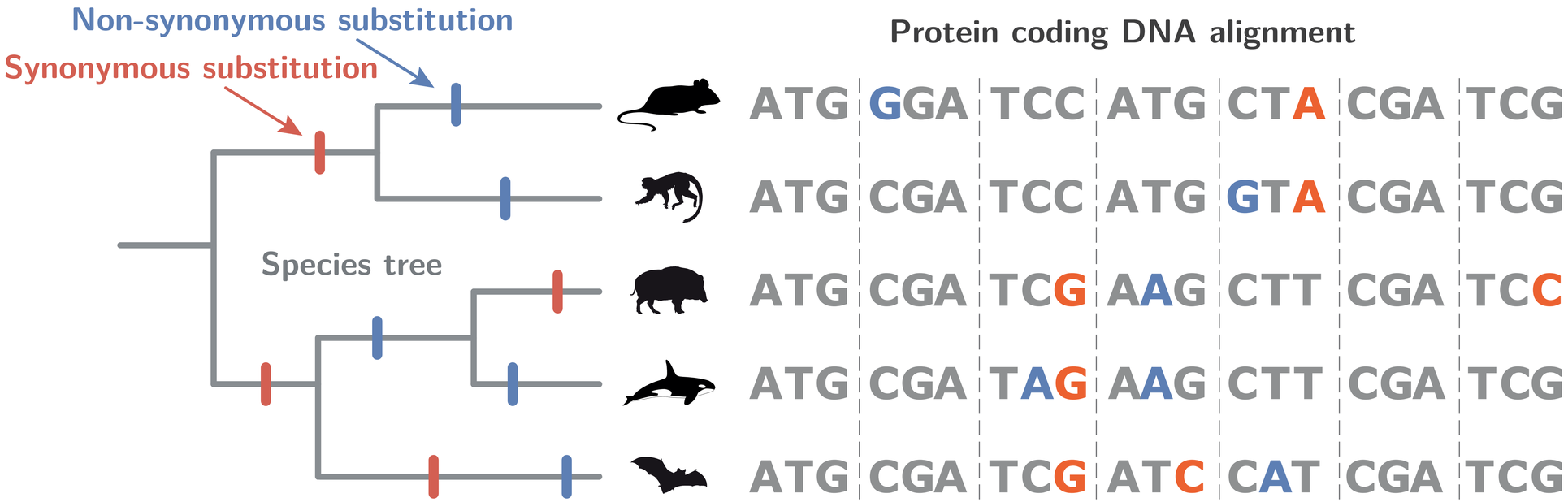
Postdoc

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).

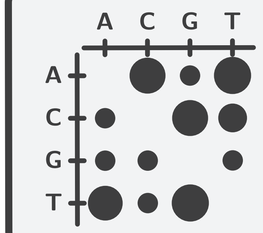
How to measure the rate of protein evolution?

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

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

$$d_N/d_S = \omega \times \mu/\mu = \omega$$

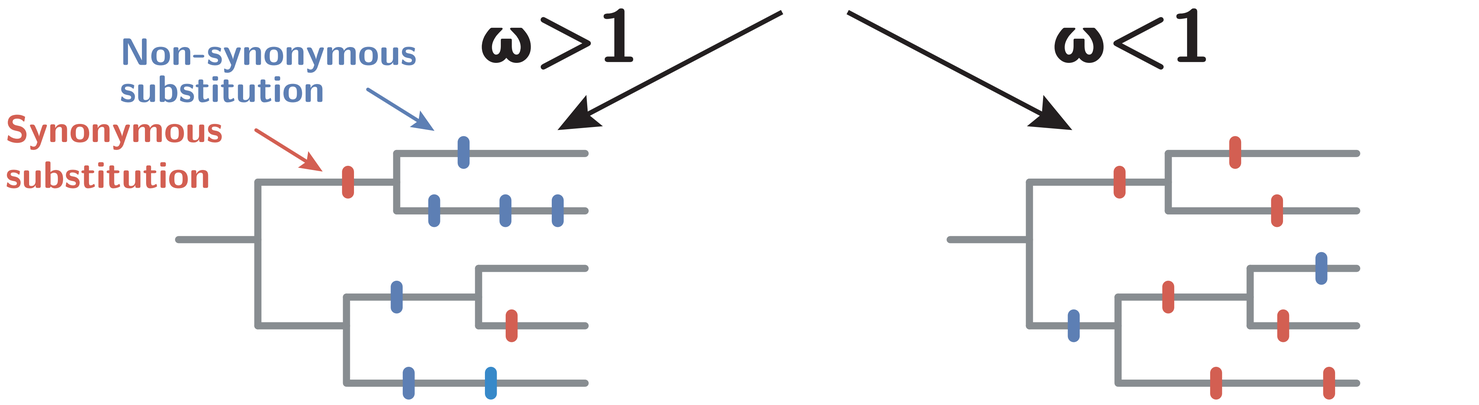
μ



Mutation rates between nucleotides

ω

Scaling factor exerced on non-synonymous mutations

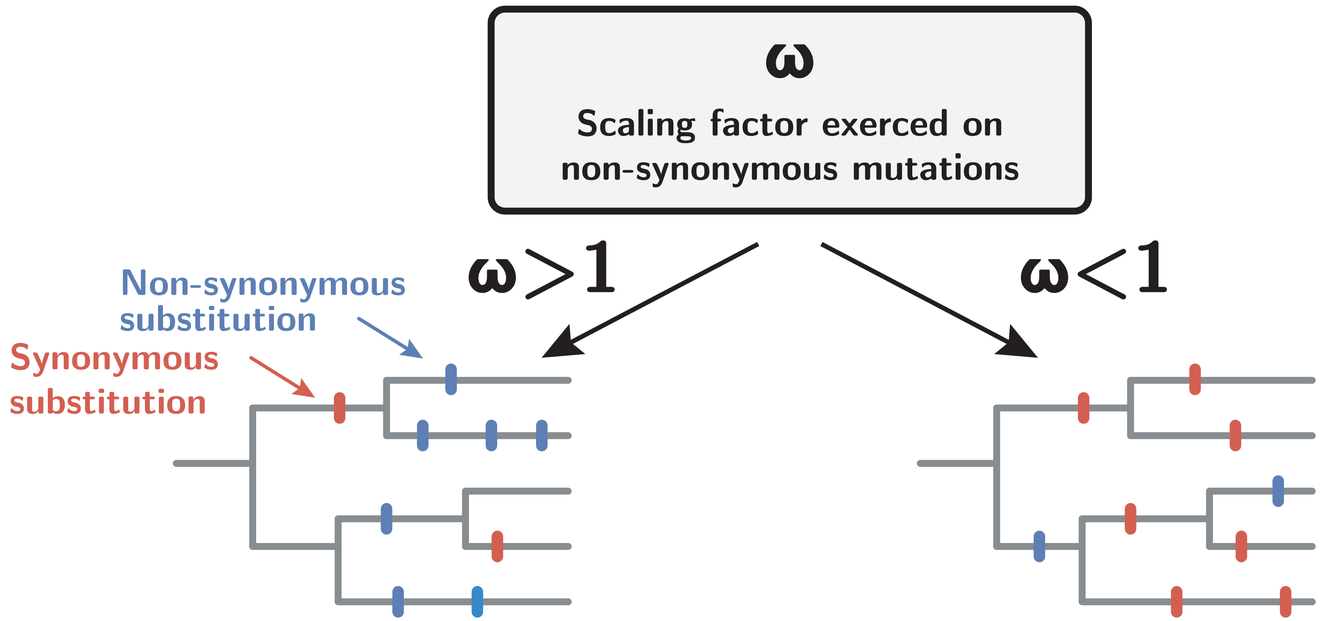


Protein is under adaptation | **Protein is constrained**

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 (ω>1), a majority are constrained (ω<1).



Protein is under adaptation

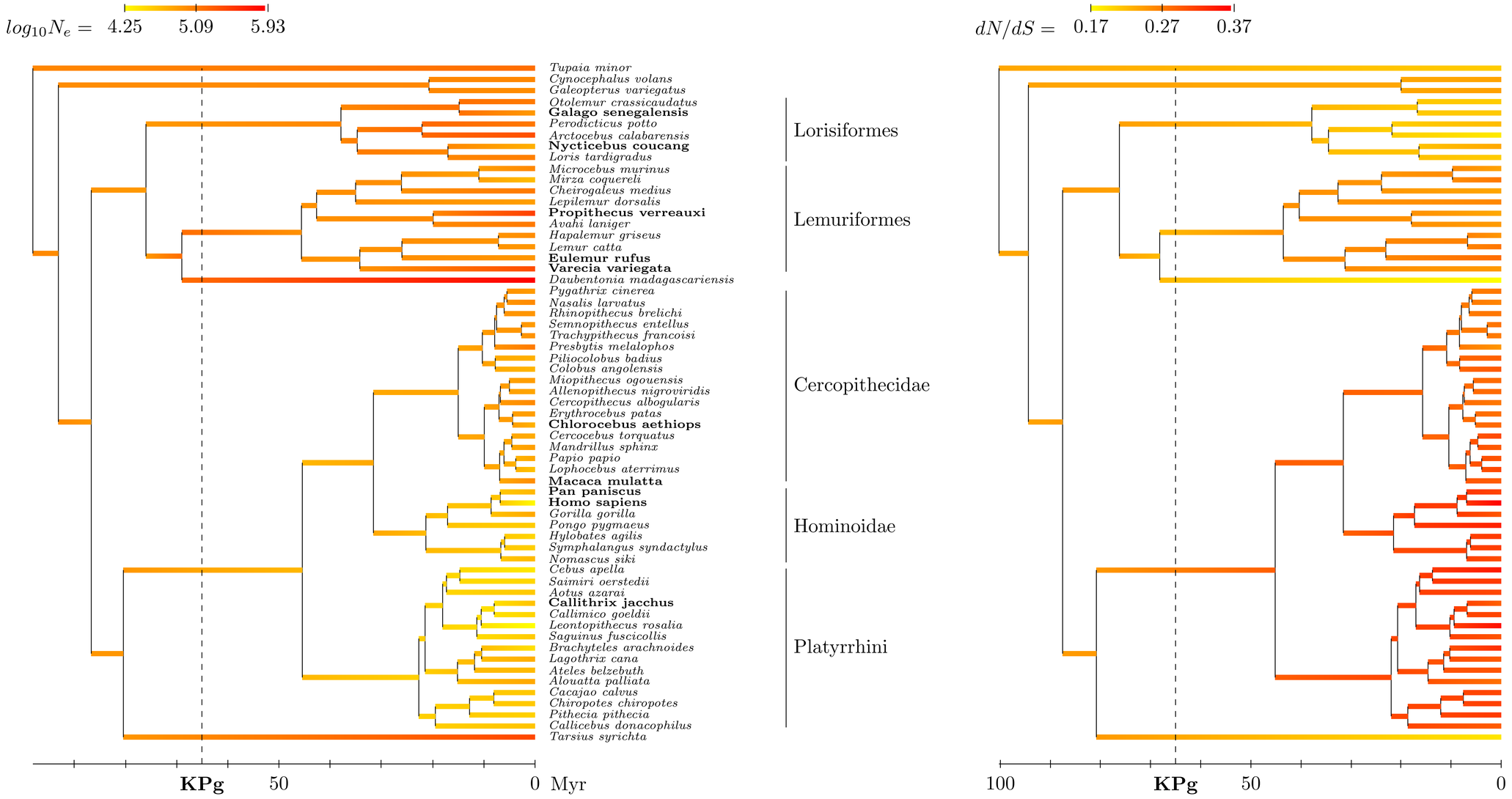
- **A very few genes have ω>1.**
  Kosiol *et al* (2008).
- **But we can detect sites with ω>1.**
  Nieslen & Yang (1998); Enard *et al* (2016).
- **Some branches can have a transient ω>1.**
  Yang & Nielsen (1998); Zhang & Nielsen (2005).

Protein is constrained

- **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?

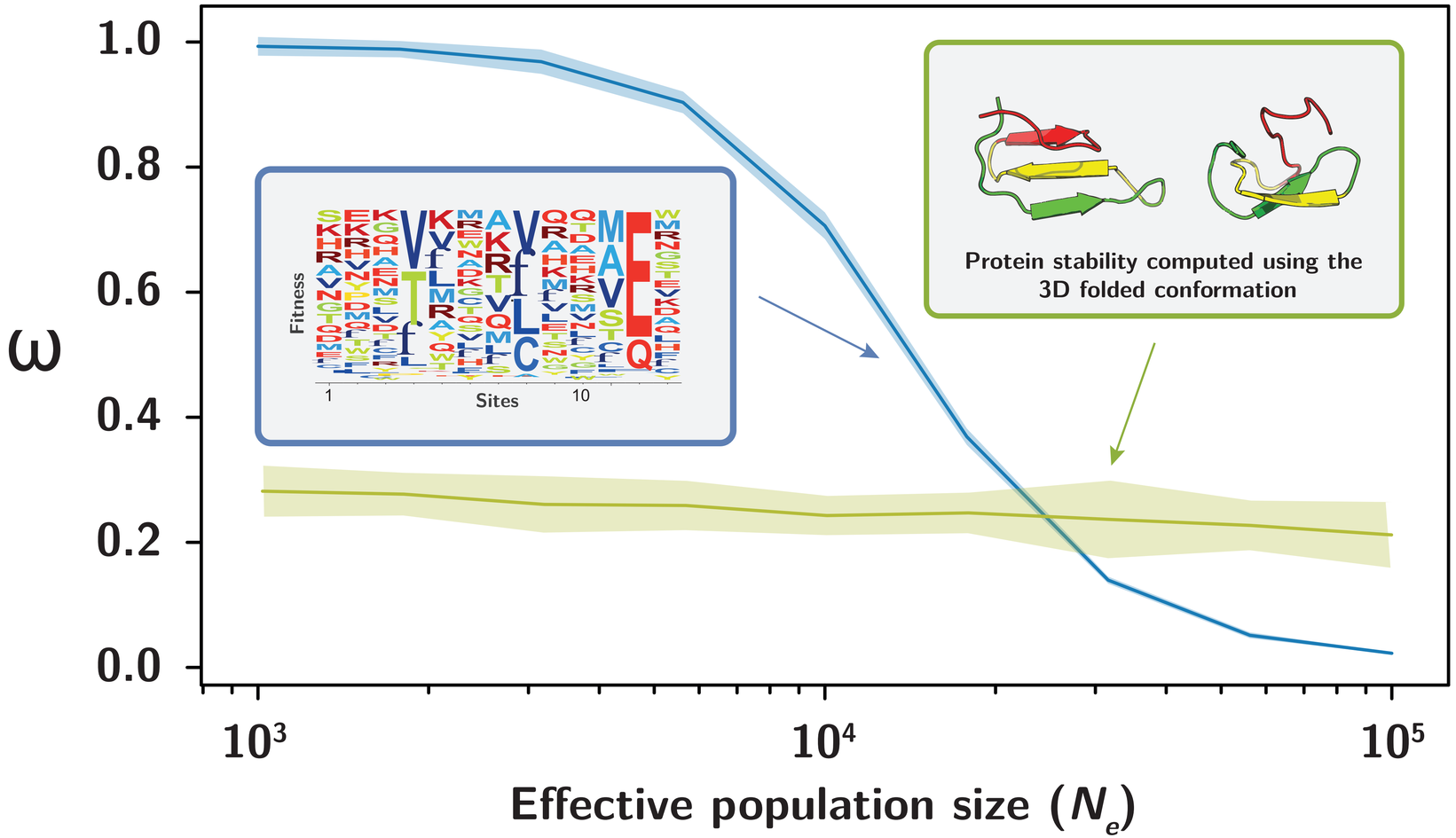
Higher N_e results in lower ω due to better efficacy of selection (r=-0.58).



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)

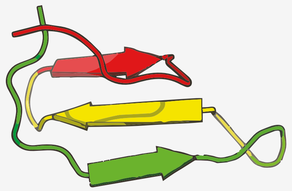
We first need to define a genotype-phenotype-fitness relationship.

Genotype

ATG|GGA| ... |TCG

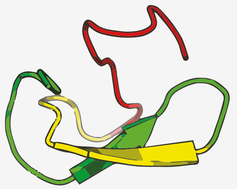
Protein coding DNA sequence

Phenotype



G_{F}

Free energy of folded state



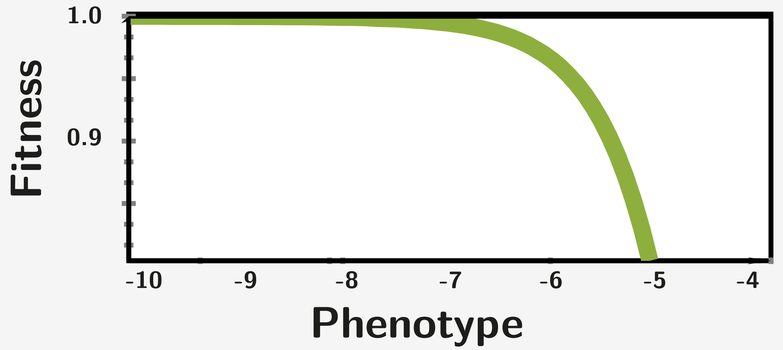
G_{U}

Free energy of unfolded state

$$\Delta G = G_{\mathrm{F}} - G_{\mathrm{U}}$$

Fitness

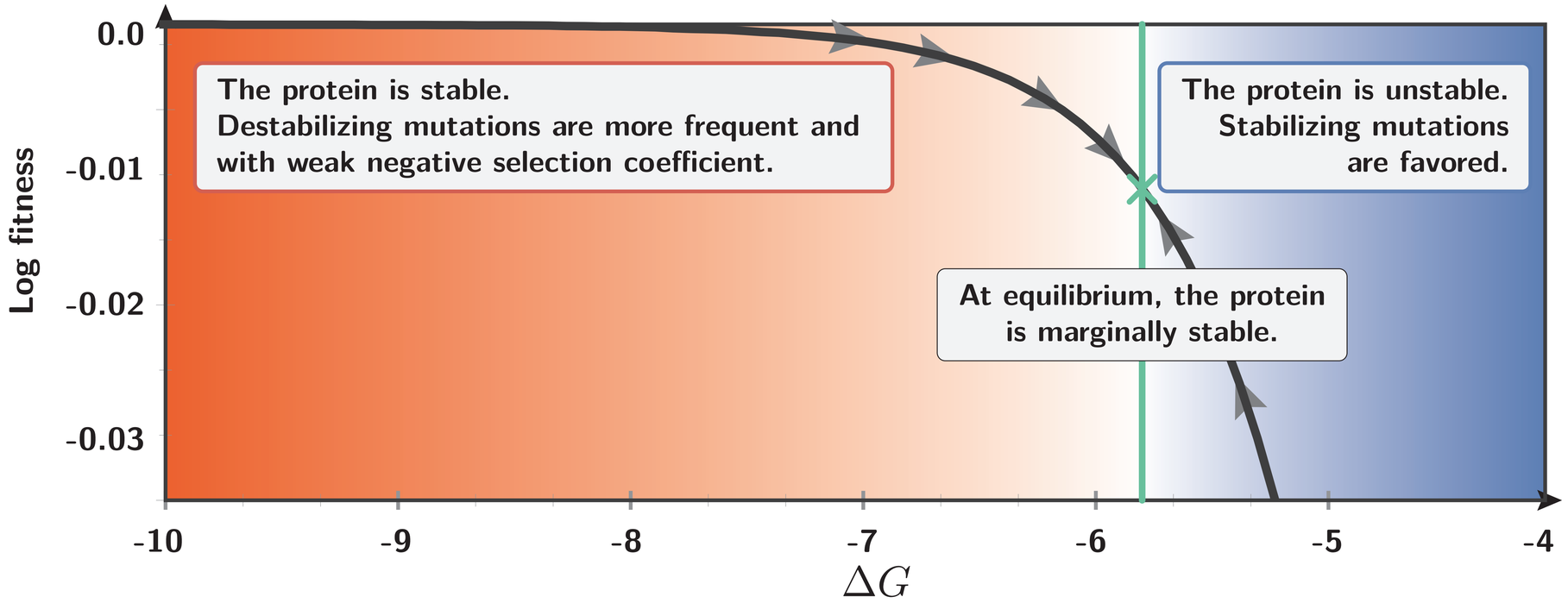
Fitness is equal to the proportion of folded proteins



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.



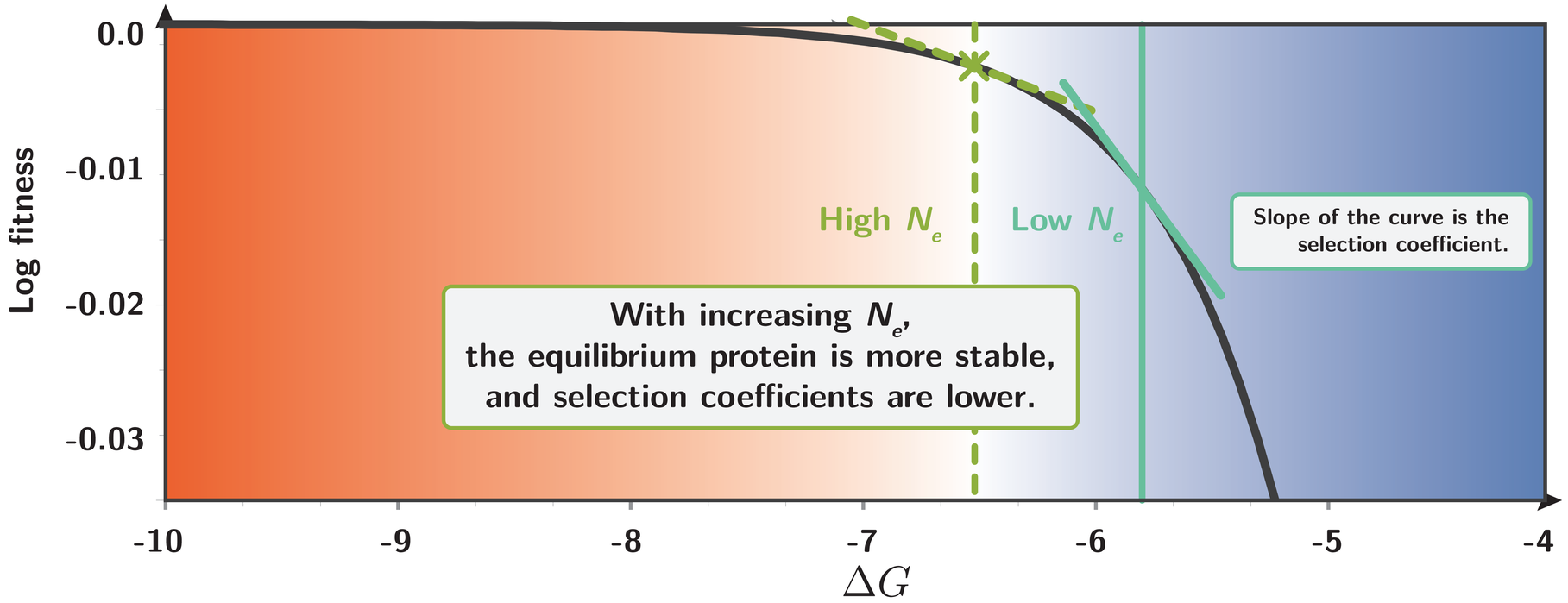
- **The optimal stability of proteins is never achieved.**

- **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)

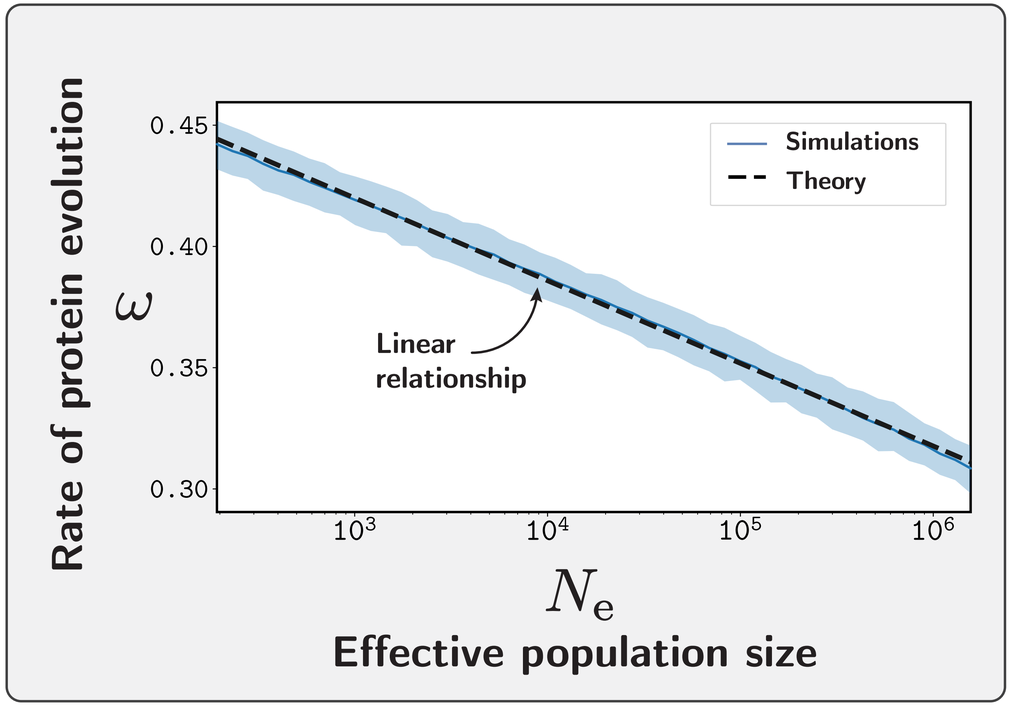
Negative linear relationship between ω and $\log(N_e)$.

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:

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



Latrille & Lartillot (2021)

What is the relationship between ω and expression level?

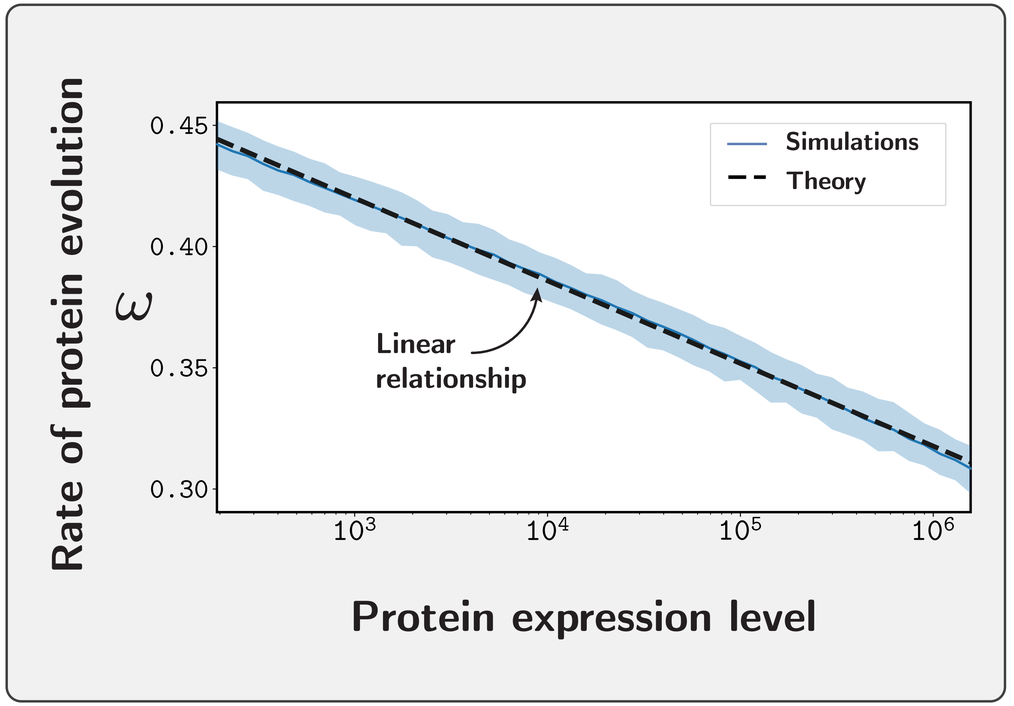
Negative linear relationship between ω and log of expression level.

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

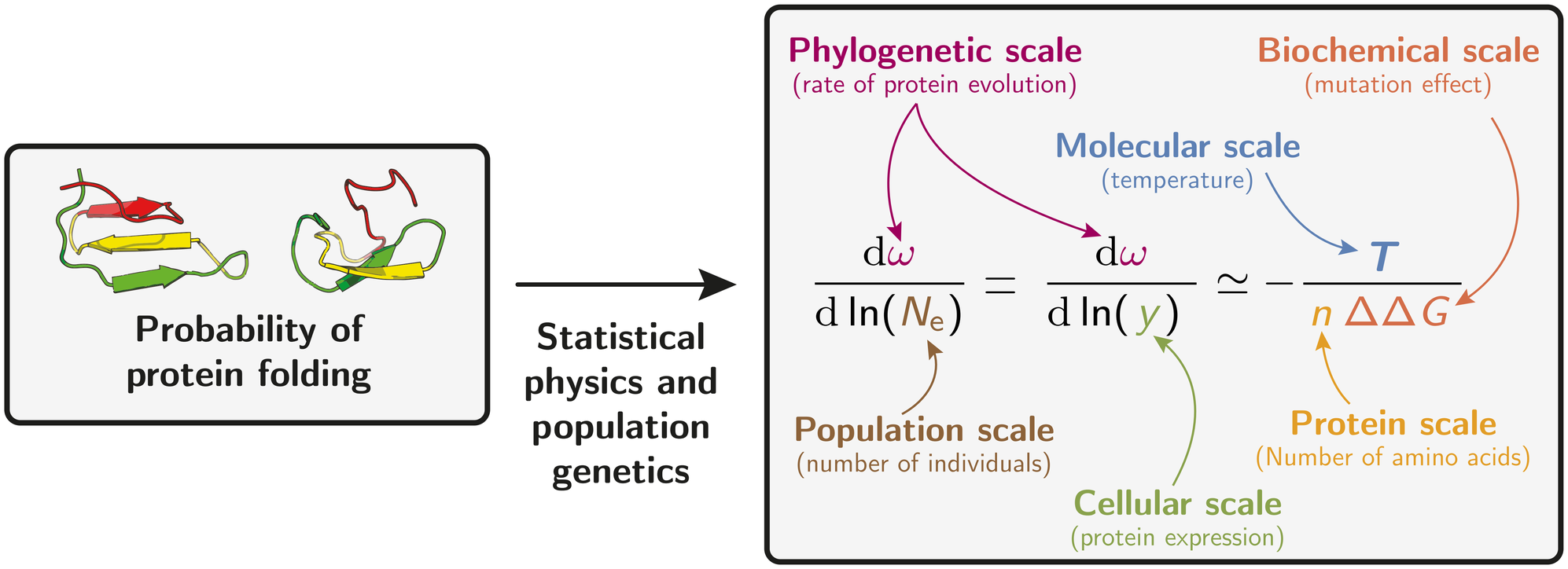
$$\frac{\mathrm{d}\omega}{\mathrm{d}\ln(y)} \simeq -\frac{\dfrac{\partial \ln f(x^*)}{\partial x^*}}{\dfrac{\partial^2 \ln f(x^*)}{\partial {x^*}^2}} \simeq -\frac{T}{n \times \Delta\Delta G}.$$



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 alway <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]