Simulation plateau

## **The Adam** peak

PhD defense tay **Predicting Selection on Traits and Sequences:** Contrast<sup>®</sup>Across Evolutionary Scales



### Employada dadd swamp **Thibault Latrille Université de Lausanne Department of Ecolology & Evolution**

Ambizione proposal de contrast Across Evolutionary Scales (p. 1988).<br>Per-coding and the Contrast Across Evolutionary Scales (p. 1988).





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

Thibault Latrille<sup>a,b,c,1</sup> ID, Nicolas Rodrigue<sup>d</sup>, and Nicolas Lartillot<sup>a</sup> ID

### **alignments across species or alternatively by leveraging polymorphism data within Population. According to a population. The adaptive rate rate rate rate relies on property codon models for the ratio of the rat over synonymous synonymous substitution rates. Evidence of an accelerated nonsynonymous substitution rates. substitution rate is considered a signature of personal periodic systems** Bridging Time Scales in Evolutionary

selection of the models and models and models and models and models are more detailed at  $\mathbf{R}$ **the interpretation, purifying interpretation, purifying, and positive selection.** In this study, we **conducted a large-scale exome-wide and Arcadi Navarro** and the placental material material material mutation– **selection models, assessing their performance at detecting proteins and sites under** Diego A. Hartasánchez, Thibault Latrille, Marina Brasó-Vives,

**of the background of purifying selection, these models are potentially limited in**

**genetic test of adaptation can be reconciled and are congruent, paving the way for** Thibault Latrille  $\mathbf{D}^{*,1,2}$  Vincent Lanore,<sup>1</sup> and Nicolas Lartillot<sup>1</sup>

**their sensitivity. Recent developments have led to more sophisticated mutation–**

both positive and purifying selection acting on protein-coding sequences. One method, izione proposal alignment comprision in Ne between lineages. In Ne between lineages. In Ne between lineages fro

**between phylogenetic and population genetics analyses, we integrated divergence and**

and nonsynonymous changes, two different types of methods have emerged to  $\alpha$  methods have emerged to  $\alpha$ 



Ambizione proposal and a strength composition of the strength of the strength of the strength of the strength o Department of Computational Biology, Université de Lausanne, Lausanne, Received 27 January 2021 models allow for different nucleotide rates at the three positions, an approach conflating the effects of mutation and Conceptually, the problem comes from the fact that phylogenetic codon models do not correctly capture the fixation population size (Ne) across the phylogeny, clearly an unrealistic assumption. This assumption can be alleviated by a<br>a

Estimating the proportion of beneficial mutations that are not adaptive in

Received 8 July 2022; Revised 25 Oct. 2022

D. A. Hartasánchez (✉) • T. Latrille

journal homepage: www.elsevier.com/locate/tpb

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

## **What tools to bridge evolutionary scales?** A combination of theoretical models and empirical studies.



 $M_{\rm H}$  sequences are shaped by selection, where the strength of selection relative to d

Mutation–selection phylogenetic codon models are grounded on population genetics first principles and represent a

polymorphism data across the externa Dominism C **show that we can site that the site of the pulletion is a signal of the pulletion of are also under adaptation at the population-genetic scale. Altogether, our exome-wide Selection Models and the population models and the population of the populati** Inferring Long-Term Effective Population Size with Mutationsity in which two main types of opposing processes are involved. Diversity is generated by means of mutation, migration, migration, migration, migration, and recombination; diversity is di<br>The combination, and recombination; diversity is diversity in the combination; diversity is diversity in the c **This article is a preprint**

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

## An Improved Codon Modeling Approach for Accurate



# **Can we predict the rate of protein evolution? Part I protein folding**



## **Part I**

Ambizione proposal and a contrast Across Evolutionary Scales by the p. 4





**• 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).

Ambizione proposal and a contrast Across Evolutionary Scales Contrast Across Evolutionary Scales

## **How to quantify changes in protein evolution?** With both synonymous and non-synonymous substitutions.

**Mutation rates between nucleotides**



- 
- 

Ambizione proposal and a contrast Across Evolutionary Scales

**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 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 under adaptation 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).
- Popadin *et al* (2007); Lanfear et al (2010).

**• Lower ω for short-lived and smaller species.**

### **Is effective population size (***N e* **) predicting ω?** Higher *N e* results in lower  $\omega$  due to better efficacy of selection (r=-0.58).



Ambizione proposal and a constructed phylogenetic contrast Across Evolutionary Scales and phylogenetic ph



Ambizione proposal estimate in the median estimate of Ne (posterior median estimated phylogenetic median estimat

### **Can we theoretically use ω to predict** *N e* **?** Not directly because the relationship depends on the model of protein evolution.



Ambizione proposal and a contrast Across Evolutionary Scales

### Latrille *et al.* (2021)



### **What is the expected relationship between ω and** *N e* We first need to define a genotype-phenotype-fitness relationship.

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

Ambizione proposal and a contrast Across Evolutionary Scales



p. 11

Taverna & Goldstein (2002)

**• 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.**



### **What is the expected relationship between ω and** *N e* **? (2/4)** Then we need to find the equilibrium and ω at this equilibrium.

Cherry (1998); Goldstein (2013).

**• Selection coefficient is dependent on the position in the fitness landscape.**

**• We can then derive the relationship between** *N e* **microscopic molecular parameters of the model.**

### **and ω as a function of the**





### **What is the expected relationship between ω and** *N e* Then we derive how changes in *N e* shift the equilibrium.

Latrille & Lartillot (2021)



### **What is the expected relationship between ω and** *N e* Negative linear relationship between ω and log(*N e* ).

### Given:

- $\blacksquare$   $T$  the temperature.
- $\bullet$  n the number of sites in the protein.
- $\blacktriangle$  ∆ $\triangle$  → 0 the destabilizing effect of a mutation.
- $x$  the proportion of destabilizing sites (phenotype).
- $\bullet$   $f(x)$  the phenotype-fitness map.
- $\parallel \bullet x^*$  the equilibrium of  $x$ .

The response in  $\omega$  after a change in  $N_e$  is:



$$
\frac{\mathrm{d}\omega}{\mathrm{d}\ln(N_{\mathrm{e}})} \simeq -\frac{\partial \ln f(x^*)}{\partial^2 \ln f(x^*)} \simeq -\frac{1}{n}
$$

Latrille & Lartillot (2021)

## **What is the relationship between ω and expression level?** Negative linear relationship between ω and log of expression level.



- $\parallel \bullet n$  the number of sites in the protein.
- $\blacktriangle$  ∆ $\triangle$  → 0 the destabilizing effect of a mutation.
- $x$  the proportion of destabilizing sites (phenotype).
- $\bullet$   $f(x)$  the phenotype-fitness map.
- $\parallel \bullet x^*$  the equilibrium of  $x$ .

The response in  $\omega$  after a change in protein expression  $\|\cdot\|$   $\mathbf{g}$   $\in$ 



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

$$
\frac{\mathrm{d}\omega}{\mathrm{d}\ln(y)} \simeq -\frac{\partial \ln f(x^*)}{\partial^2 \ln f(x^*)} \simeq
$$

$$
\frac{\partial \ln f(x^*)}{\partial x^{*2}}
$$

## **Can theoretical models of protein folding predict rate of evolution?**  Models form a bridge across different scales and can be tested.





## **Probability** C e and the context of the c **• This model forms a bridge across different scales and can be tested.**

- 
- 

# Chapter I<br> **Chapter I**

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

## **proteinal in the constructs**, where is no adaptation possible, wis alway<br>The adaptation when proteins are generally co **•** In our model, there is no adaptation possible, **ω** is alway <1.

Ambizione proposal **Latrille Thibault Contrast Across Evolutionary Scales** p. 16 <sup>a</sup> *Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR 5558, F-69622 Villeurbanne, France*

### **• With a theoretical model for selection on protein folding, ω is linearly decreasing with** *N e* **and expression level (on log scale).**

### **• How to detect adaptation when proteins are generally constrained?** Theoretical Population Biology journal homepage: www.elsevier.com/locate/tpb

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

T. Latrille <sup>a,b,\*</sup>, N. Lartillot <sup>a</sup>