



# Protein Paleontology

**Thomas Dupic**

*July 1st, 2025 – CIML – SAB visit*



# Background



2015

2018

2020

2025



# Background



2015

2018

2020

2025



PhD, Extended Conformal  
Field Theories

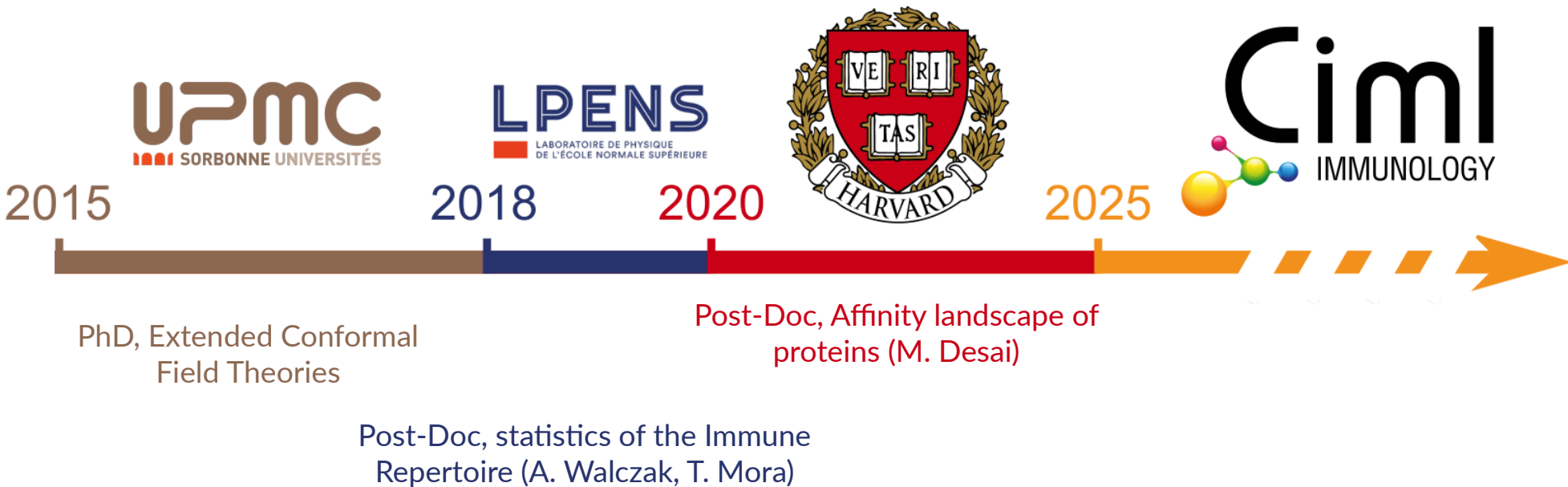
# Background



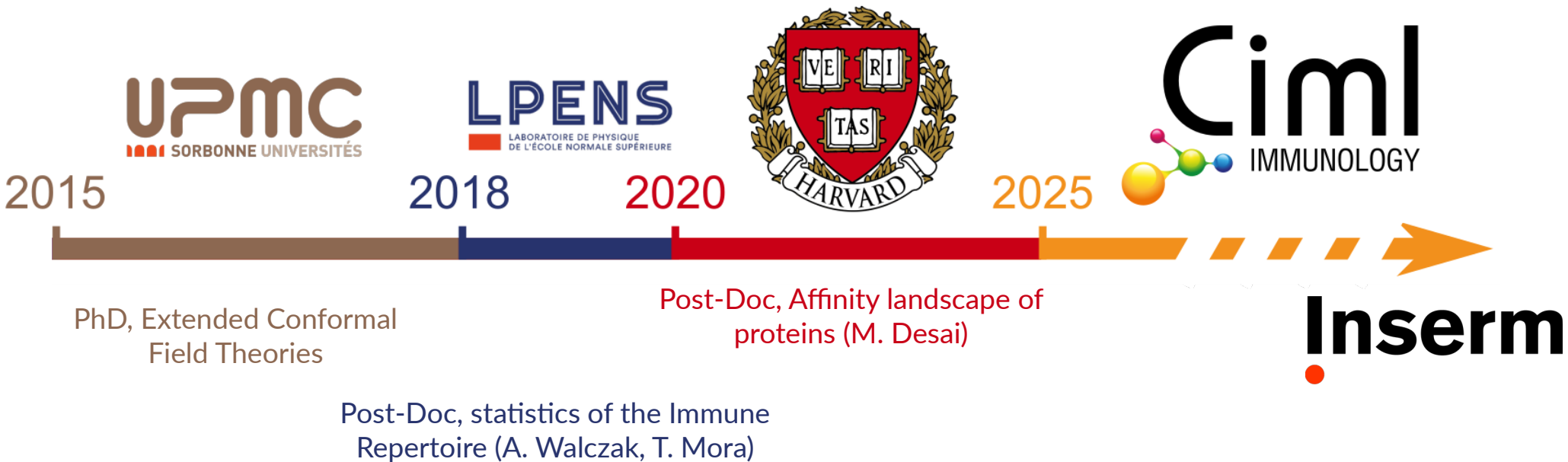
PhD, Extended Conformal  
Field Theories

Post-Doc, statistics of the Immune  
Repertoire (A. Walczak, T. Mora)

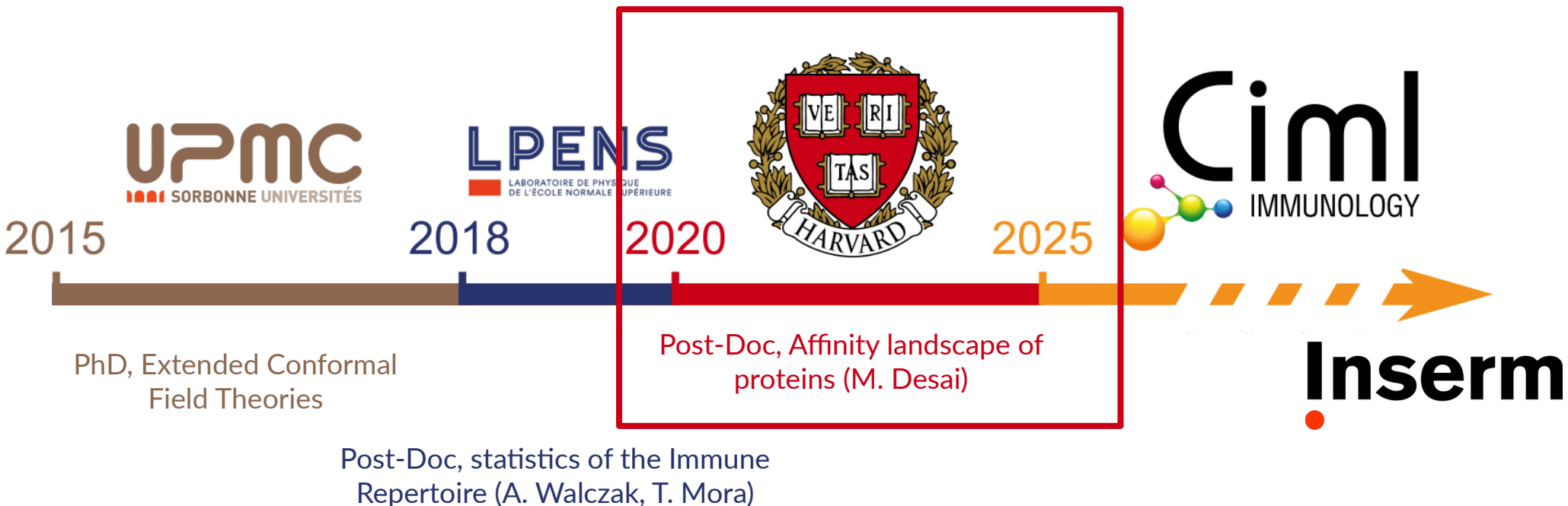
# Background



# Background



# Background



## Protein evolution

Viral proteins and antibodies evolve rapidly, under strong selection.

### *Le Monde*

Les scientifiques intrigués par de nouvelles mutations du SARS-CoV-2

Publié le 16 décembre 2020

## Protein evolution

Viral proteins and antibodies evolve rapidly, under strong selection.

*Le Monde*

Les scientifiques intrigués par de nouvelles mutations du SARS-CoV-2

Publié le 16 décembre 2020

**SARS-CoV-2: 0.03 mut. / replications, but 25 mut. / year**

## Protein evolution

Viral proteins and antibodies evolve rapidly, under strong selection.

*Le Monde*

Les scientifiques intrigués par de nouvelles mutations du SARS-CoV-2

Publié le 16 décembre 2020

**SARS-CoV-2:** 0.03 mut. / replications, **but 25 mut. / year**

**Antibodies:** 1 mut. / replications, 40 mut. total

# Protein evolution

Viral proteins and antibodies evolve rapidly, under strong selection.

*Le Monde*

Les scientifiques intrigués par de nouvelles mutations du SARS-CoV-2

Publié le 16 décembre 2020

**SARS-CoV-2:** 0.03 mut. / replications, **but 25 mut. / year**

**Antibodies:** 1 mut. / replications, 40 mut. total

Mutations that fix in the population are largely deterministic.

# Protein evolution

Viral proteins and antibodies evolve rapidly, under strong selection.

*Le Monde*

Les scientifiques intrigués par de nouvelles mutations du SARS-CoV-2

Publié le 16 décembre 2020

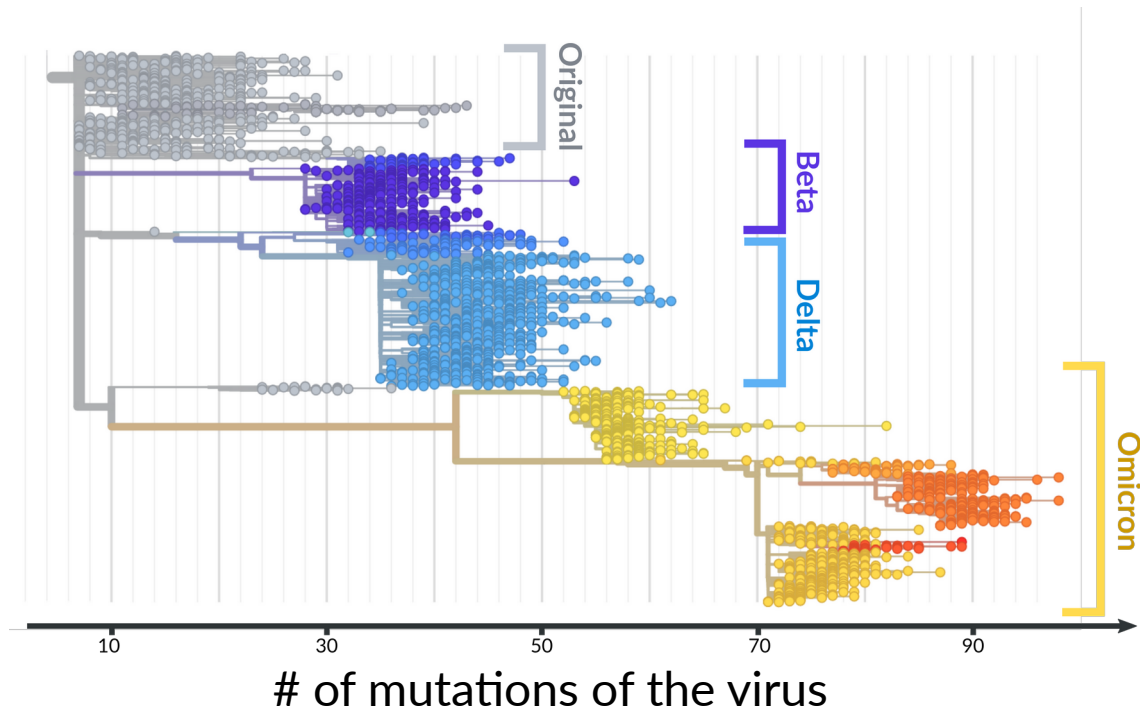
**SARS-CoV-2:** 0.03 mut. / replications, **but 25 mut. / year**

**Antibodies:** 1 mut. / replications, 40 mut. total

Can we predict / understand the evolution of these proteins ?

Mutations that fix in the population are largely deterministic.

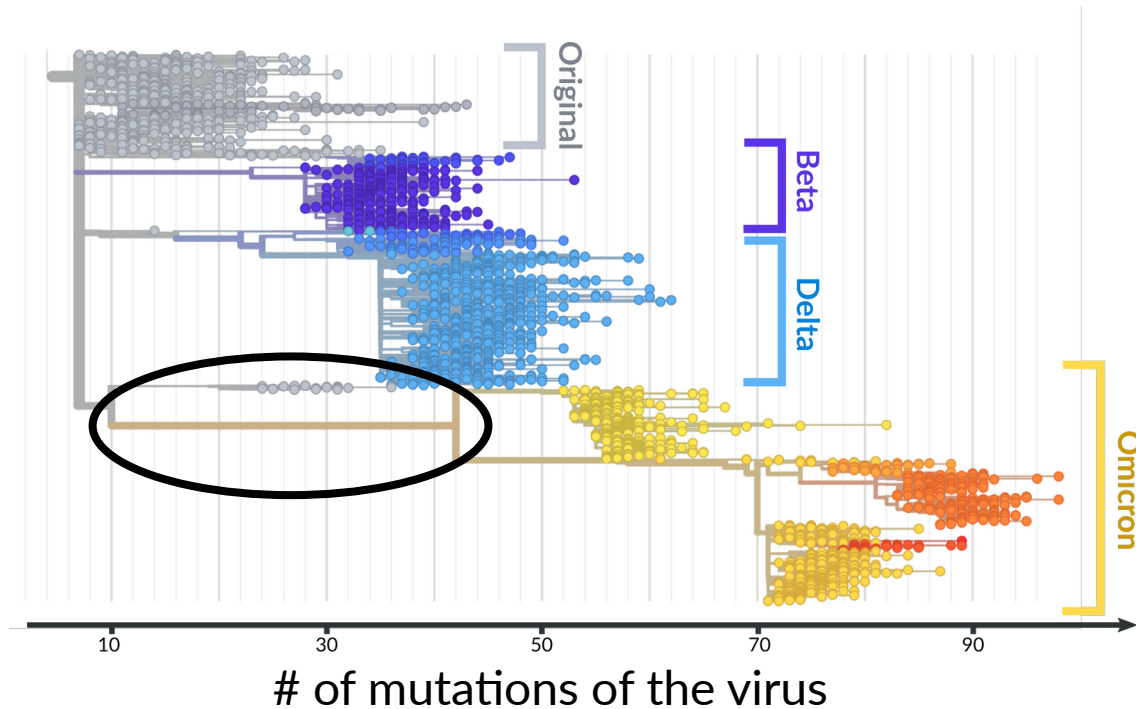
# Protein evolution



# of mutations of the virus  
SARS-CoV-2 phylogeny

Phylogenetic reconstruction can help us reconstruct the evolution of a protein. But not always...

# Protein evolution



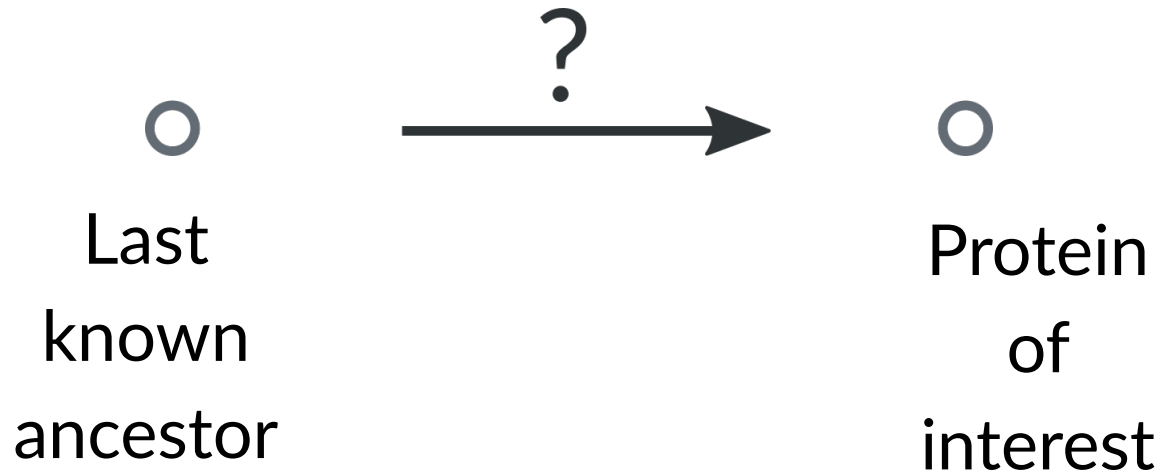
# of mutations of the virus  
SARS-CoV-2 phylogeny

Phylogenetic reconstruction can help us reconstruct the evolution of a protein. But not always...

**Evolutionary jump of the variant**

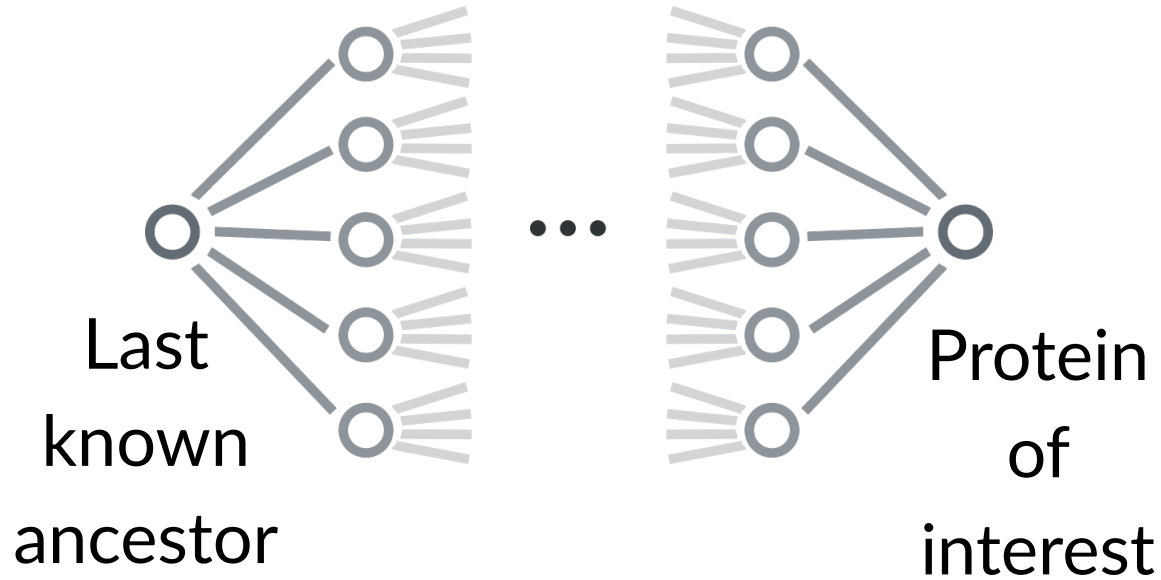
# Protein Paleontology

# Protein Paleontology



Understand protein  
evolution without  
phylogenetic tree

# Protein Paleontology



Understand protein evolution without phylogenetic tree

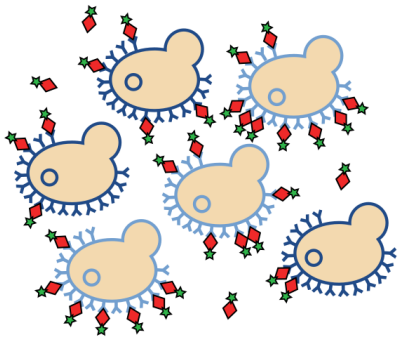
Combinatorial strategy, reconstructing all possible intermediates to measure their affinity.

## Tite-Seq

**Tite-Seq:** Measure binding energies for up to 100'000 protéines against one ligand.

# Tite-Seq

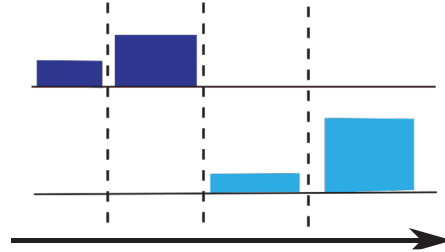
**Tite-Seq:** Measure binding energies for up to 100'000 protéines against one ligand.



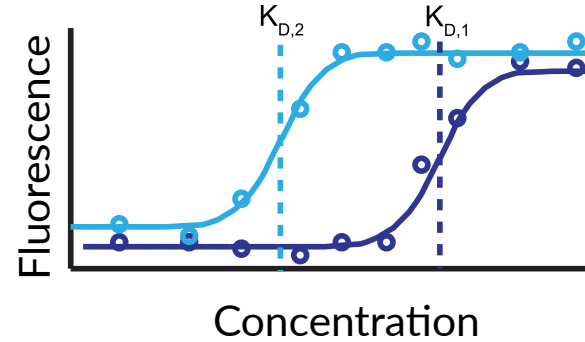
Expression on the yeast with fluorescent antigen



Cell sorting, sequencing and statistics

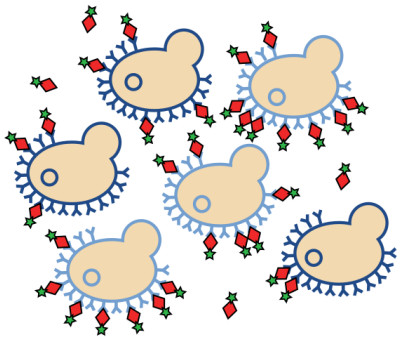


Fluorescence



# Tite-Seq

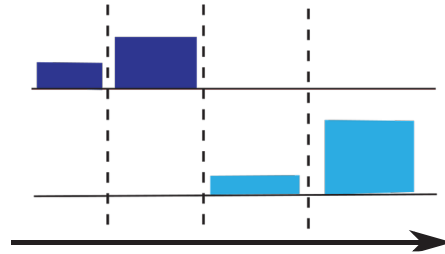
**Tite-Seq:** Measure binding energies for up to 100'000 protéines against one ligand.



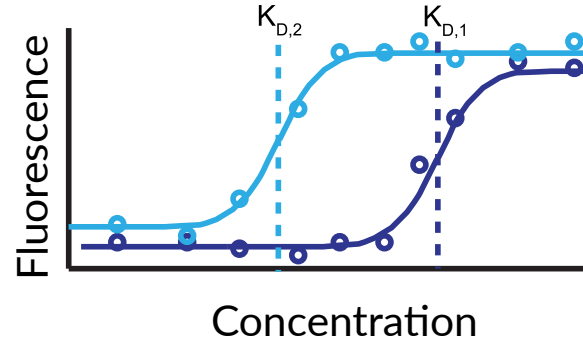
Expression on the yeast with fluorescent antigen



Cell sorting, sequencing and statistics



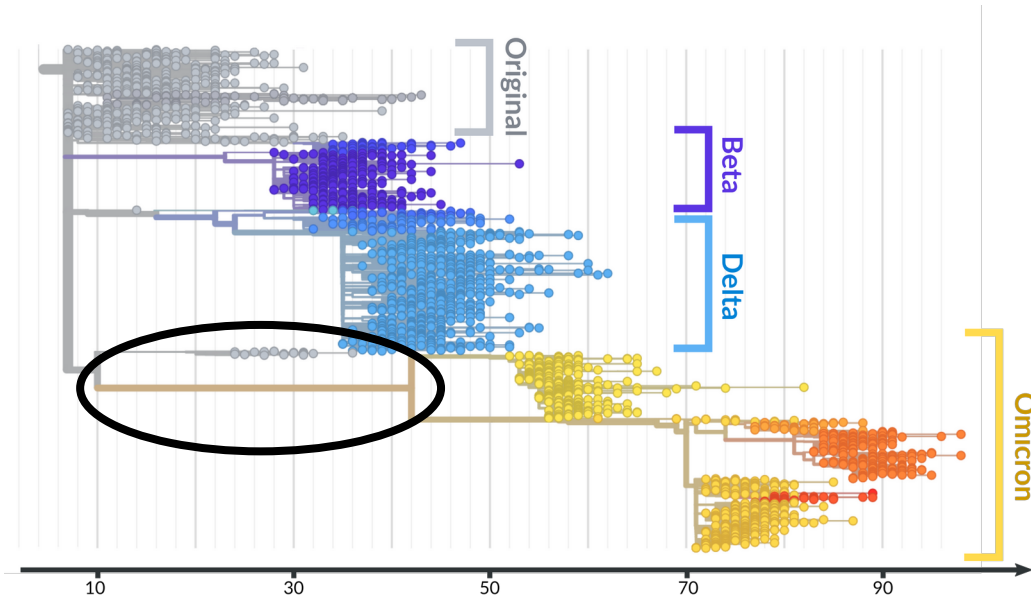
Fluorescence



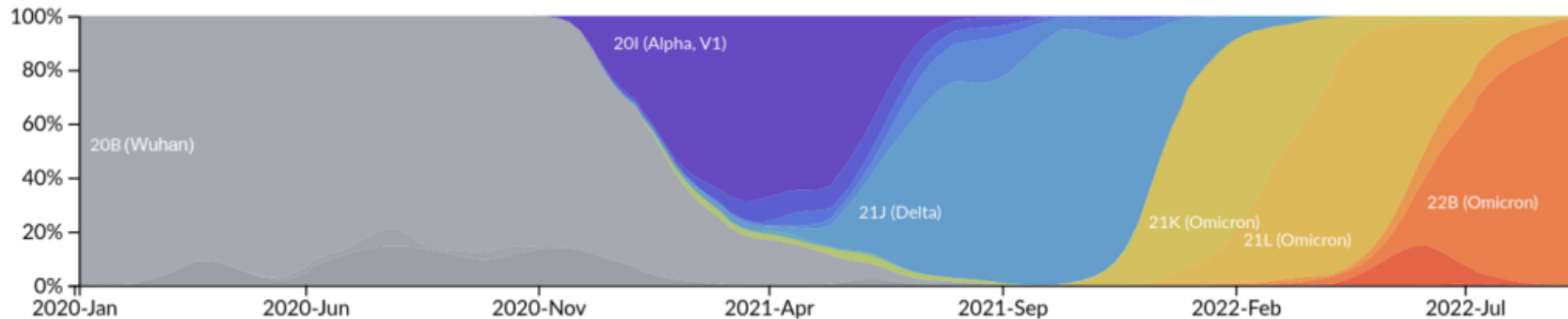
**Caveat:**

- Yeast-made  $\neq$  human-made
- In the wild, interactions don't happen "at equilibrium"

# Apparition of the SARS-CoV-2 Omicron variant



Why did Omicron take so long to appear?



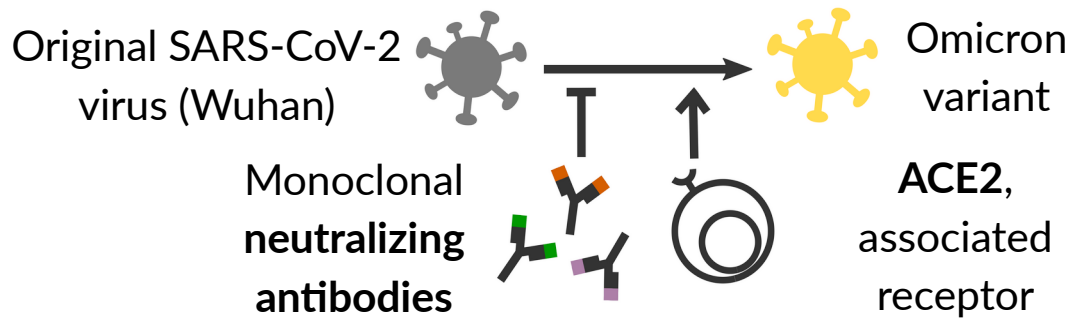
# Apparition of the SARS-CoV-2 Omicron variant

15 mutations on the binding region of the spike.



# Apparition of the SARS-CoV-2 Omicron variant

15 mutations on the binding region of the spike.

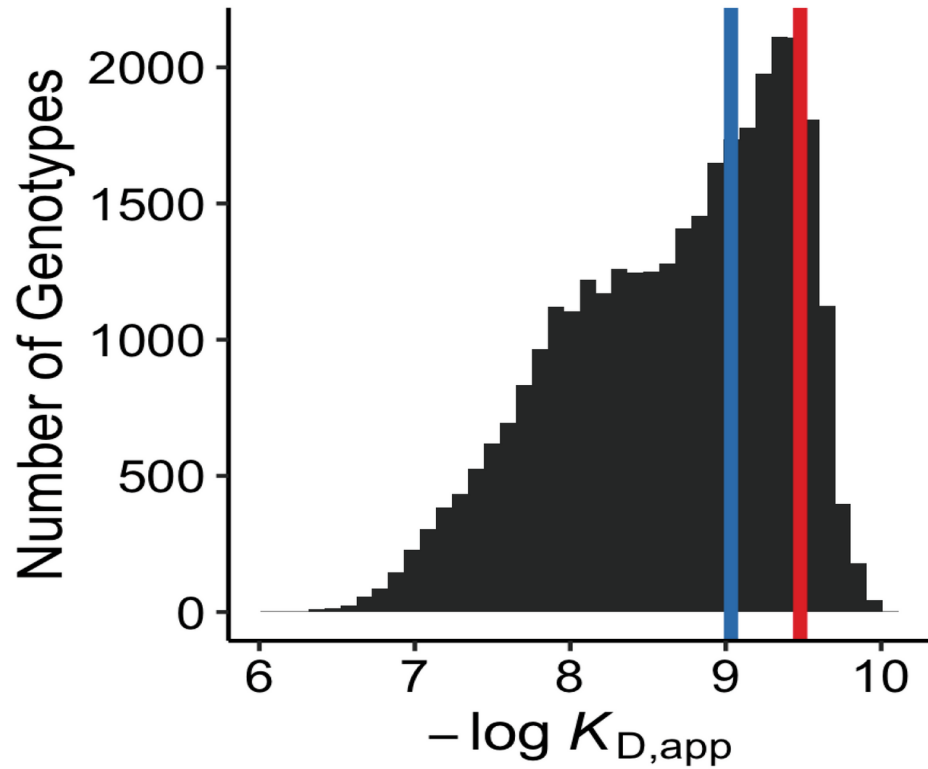


We measured the affinity of 32 768 variants of the spike protein<sup>1,2</sup>.



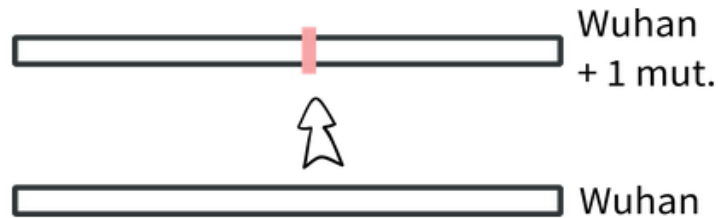
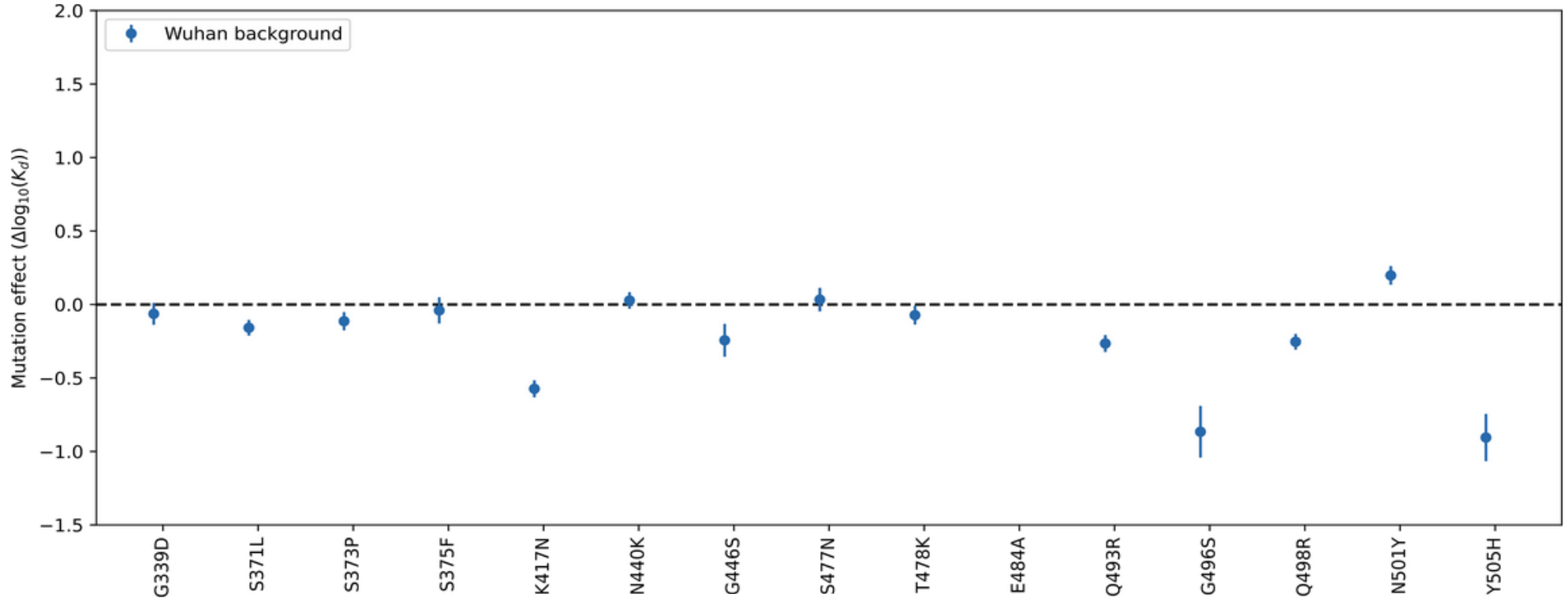
1. Moulana\*, Dupic\* et al., Nat. Comm. (2022)  
 2. Moulana\*, Dupic\* et al., eLife (2023)

# Binding affinity against ACE2

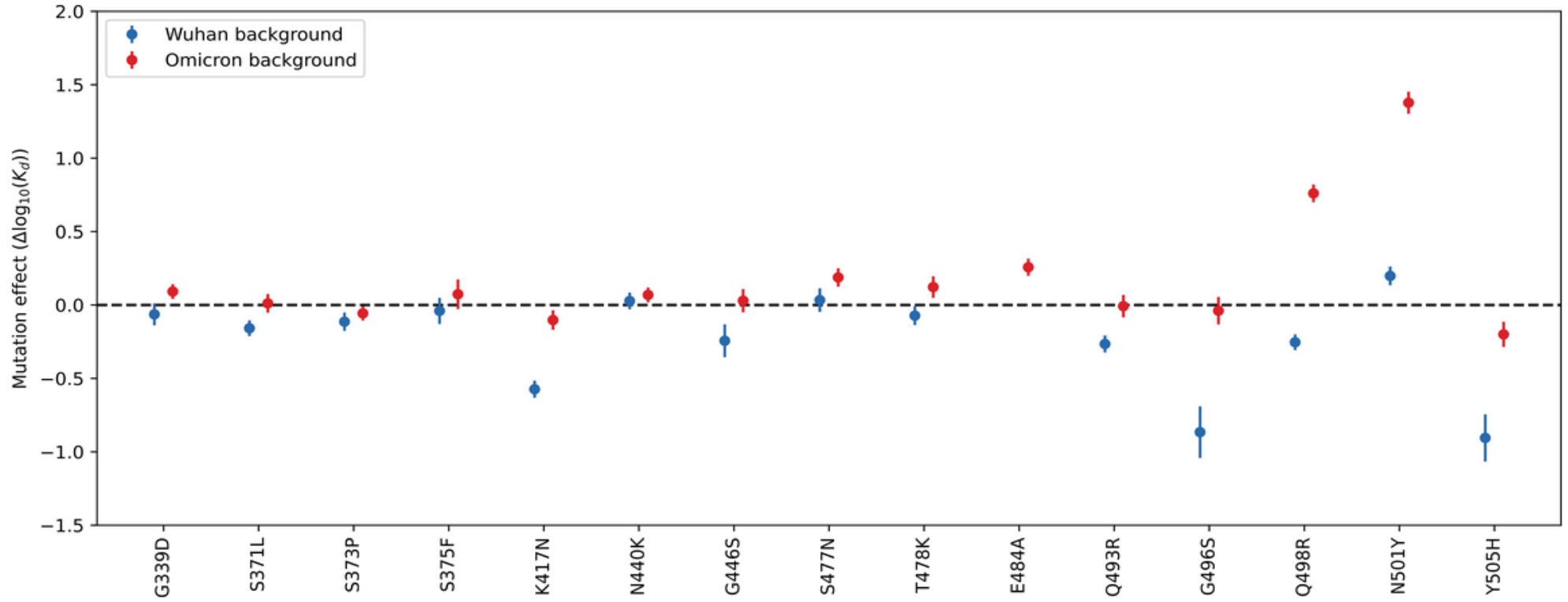


Omicron and Wuhan have similar binding affinity against ACE2.

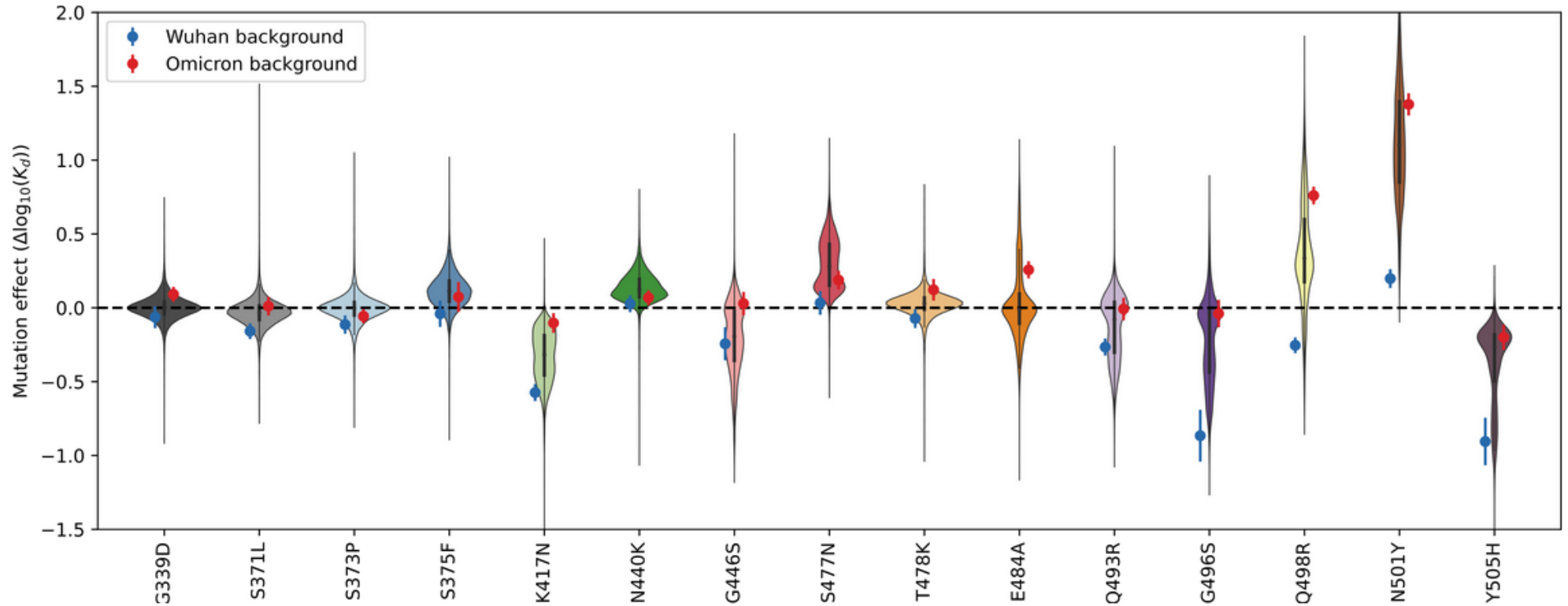
# Binding affinity against ACE2



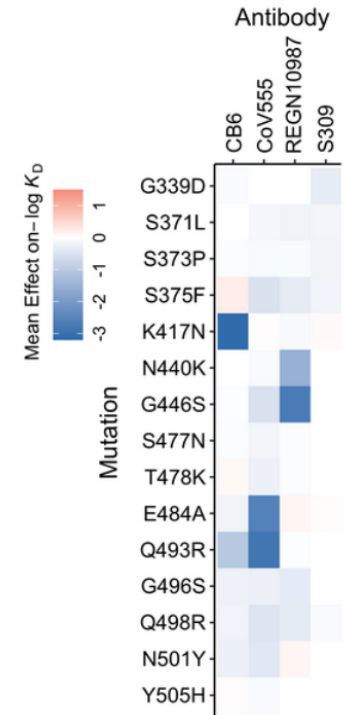
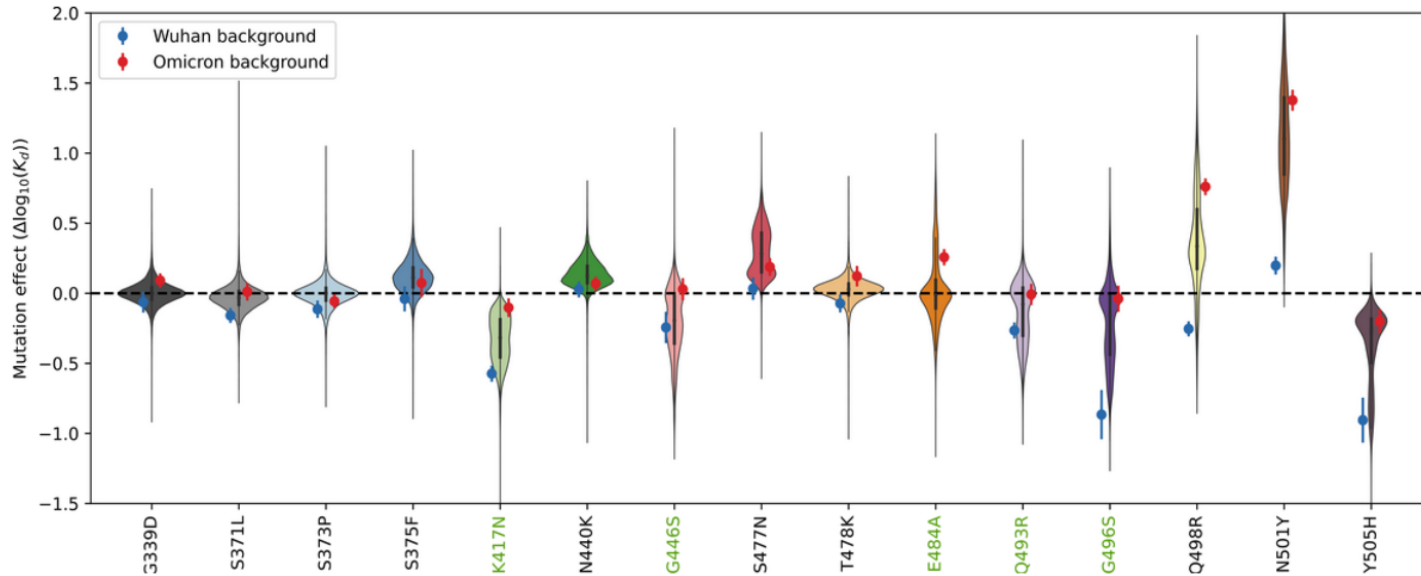
# Binding affinity against ACE2



# Binding affinity against ACE2



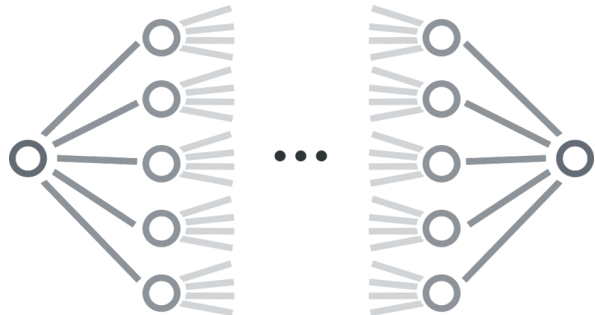
# Binding affinity against ACE2 and antibodies



# Past reconstruction

Markov Model. Jump probability (Kimura):

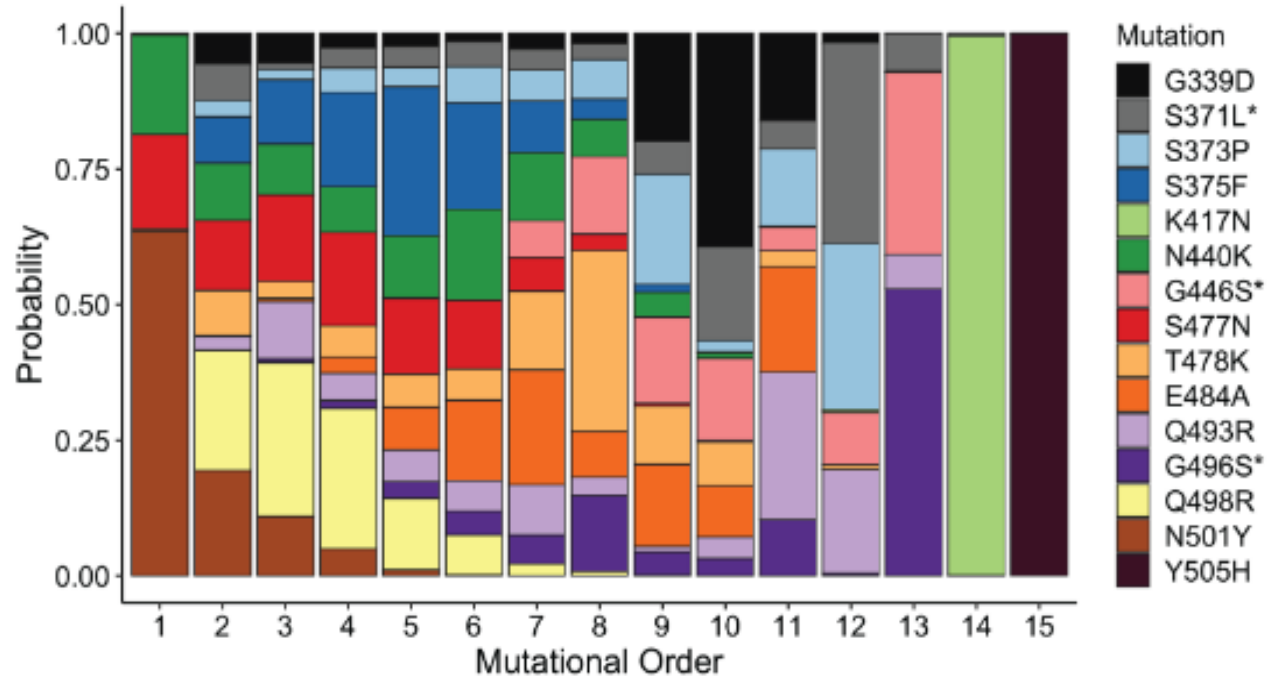
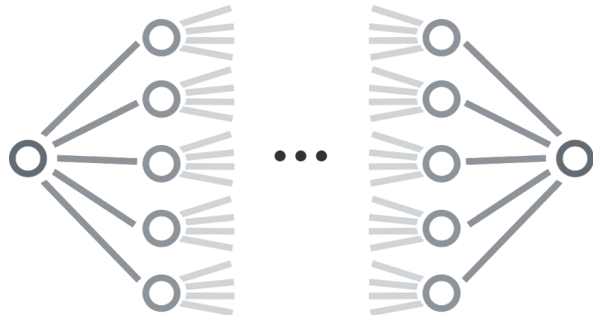
$$P_{i,j} \propto \frac{1 - e^{f_i - f_j}}{1 - e^{N(f_i - f_j)}}$$



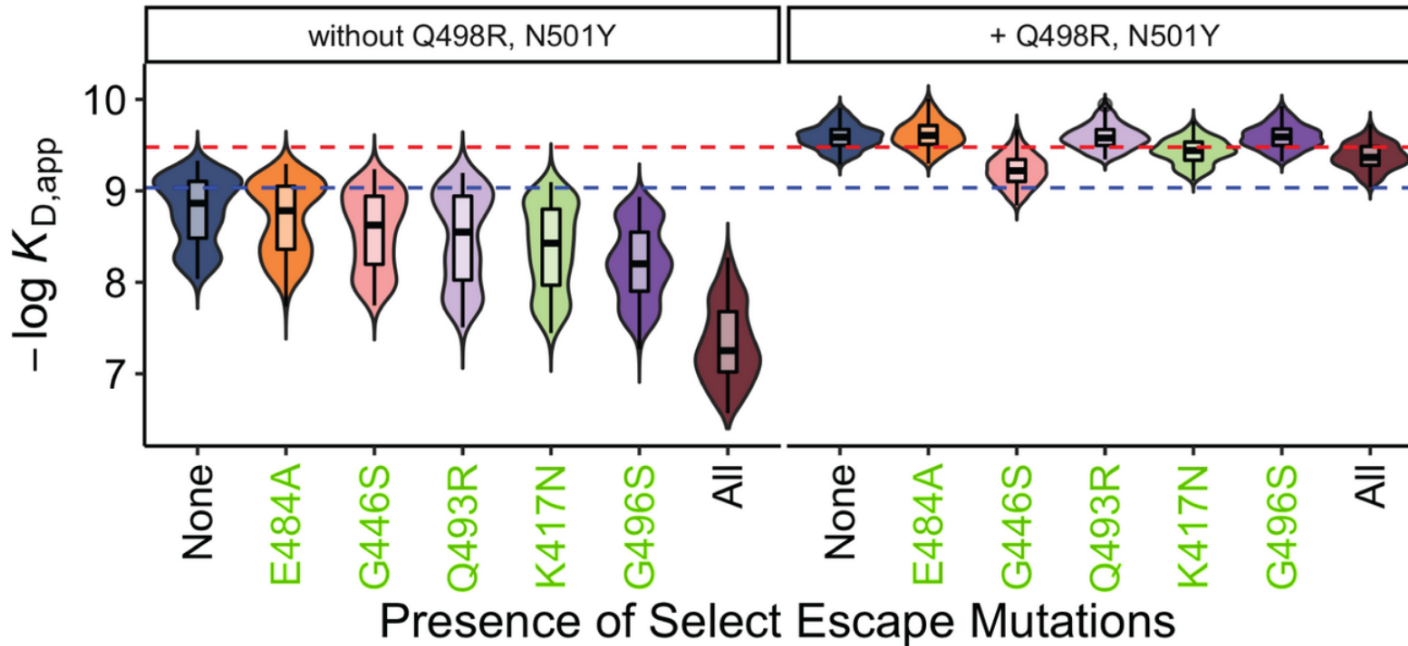
# Past reconstruction

Markov Model. Jump probability (Kimura):

$$P_{i,j} \propto \frac{1 - e^{f_i - f_j}}{1 - e^{N(f_i - f_j)}}$$



# Past reconstruction



The same two mutations improve drastically the binding affinity of all immune-escape mutations !

# Evolutionary landscape of Omicron



Ridge evolution against  
ACE2

# Evolutionary landscape of Omicron

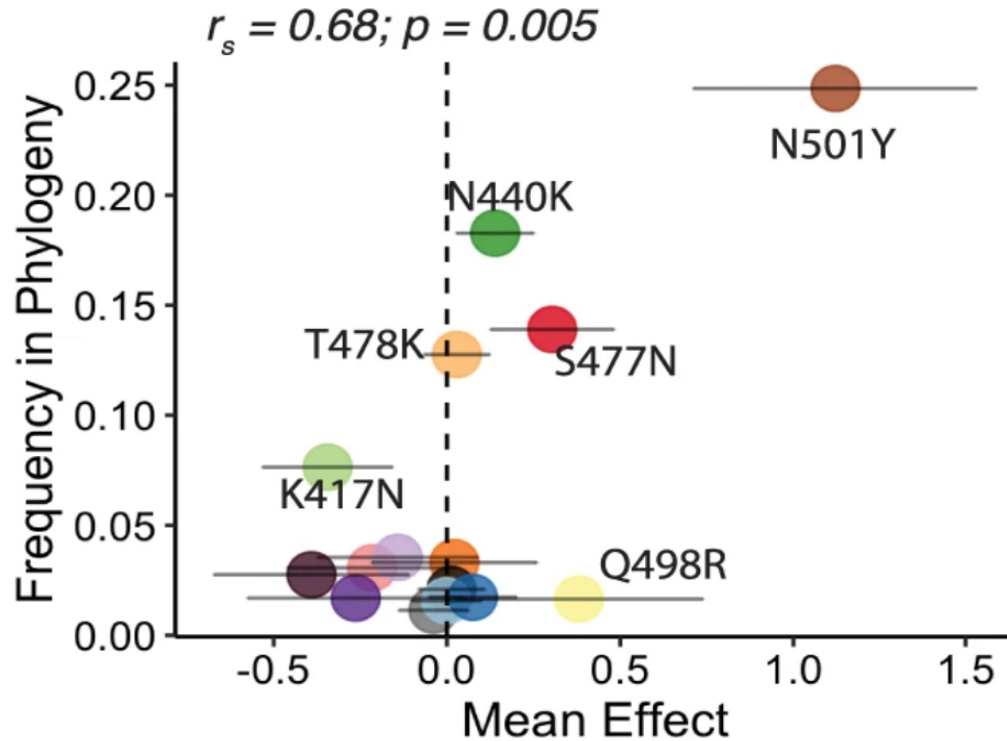


Ridge evolution against  
ACE2

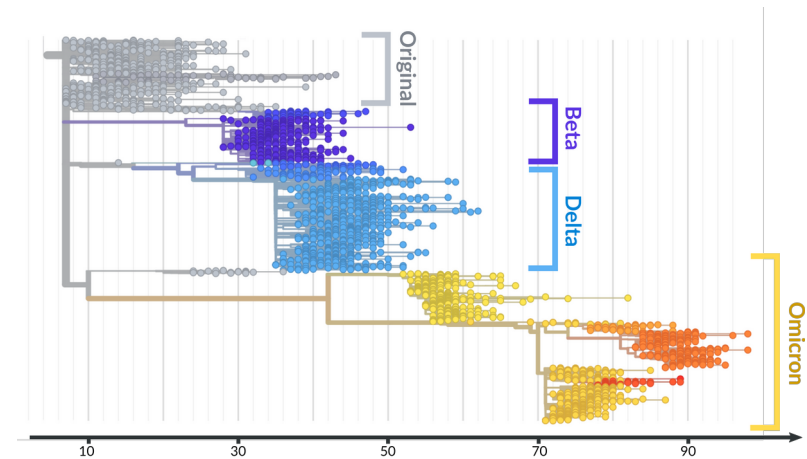


Cliff evolution against  
antibodies

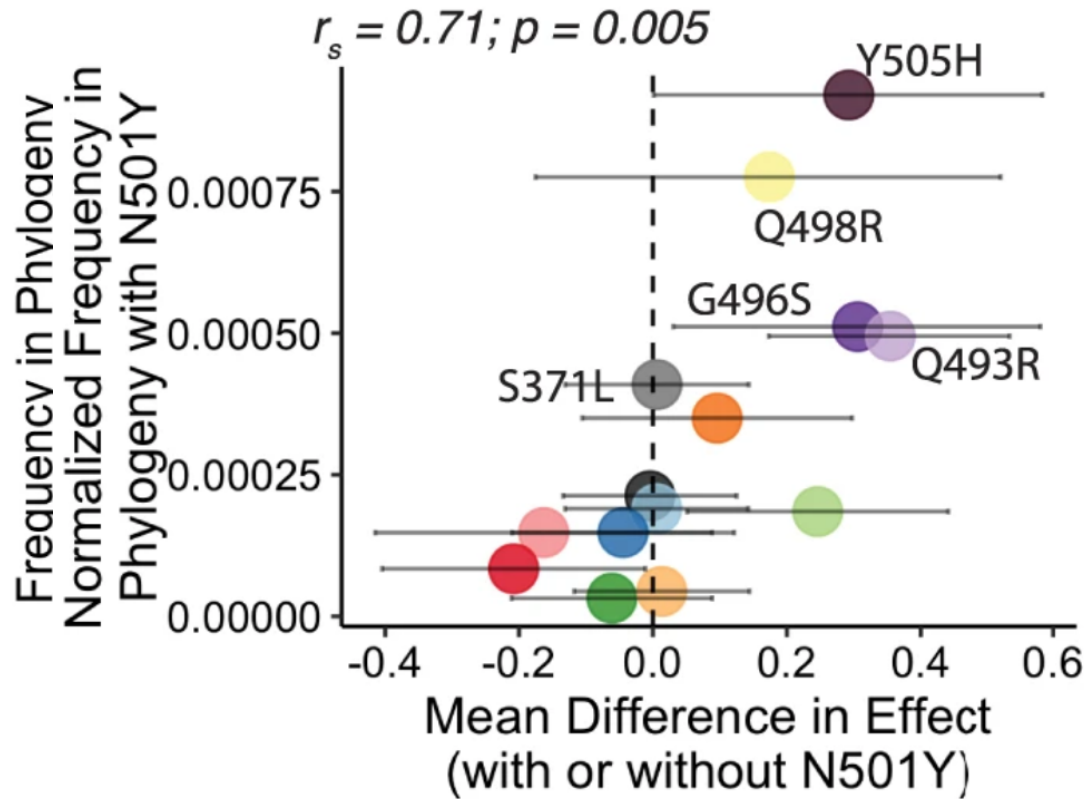
# Can we trust this ?



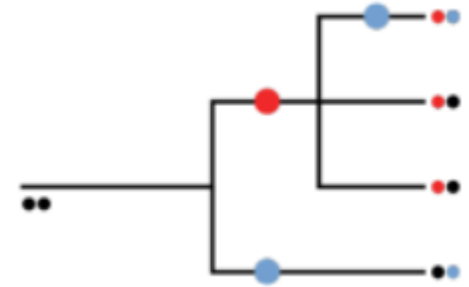
Comparison with frequencies within the phylogeny.



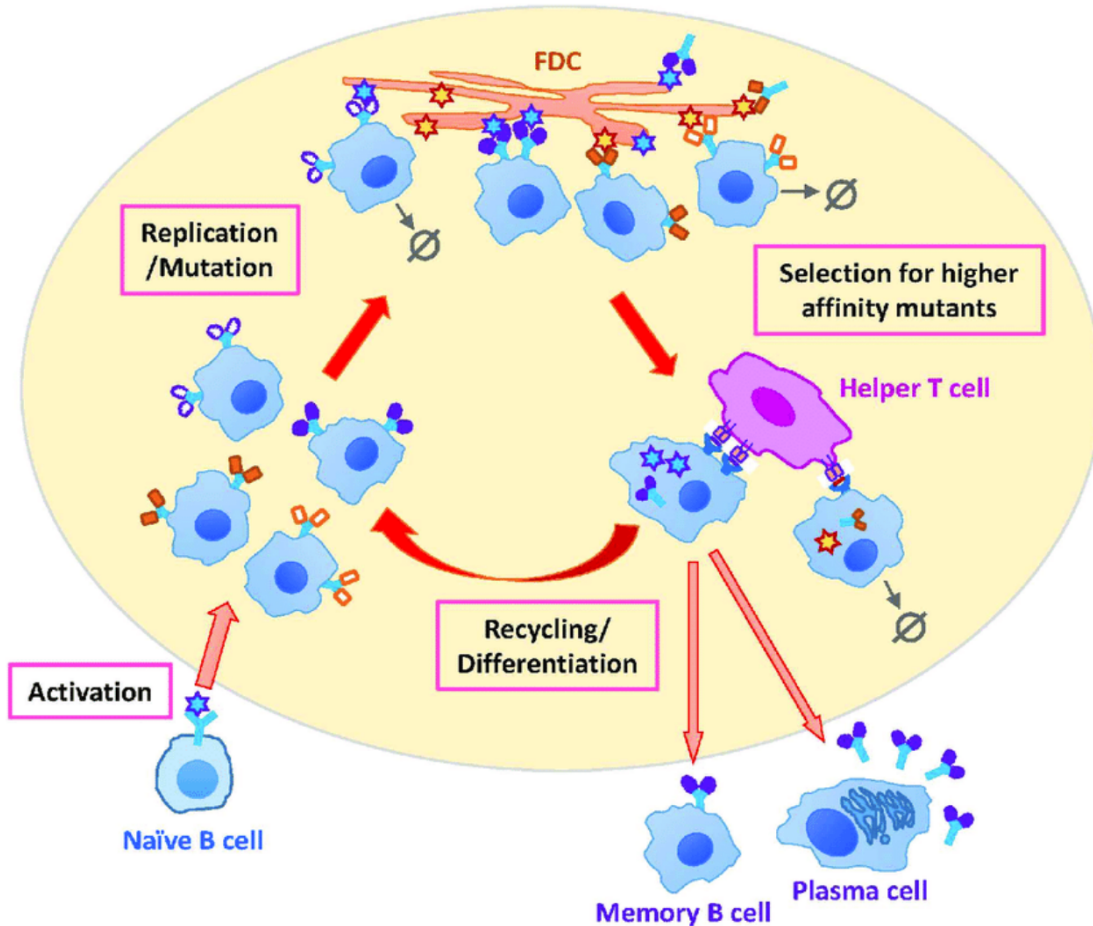
# Can we trust this ?



Comparison with frequencies within the phylogeny.



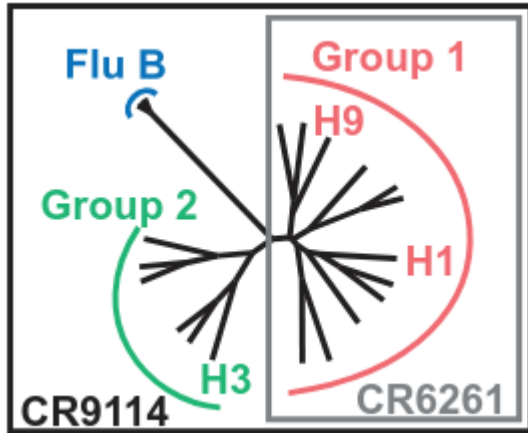
# Antibodies also evolve



Antibodies are selected in a “darwinian-like” process, affinity maturation.

Can we infer what were the selection pressure?

# Broadly neutralizing antibodies

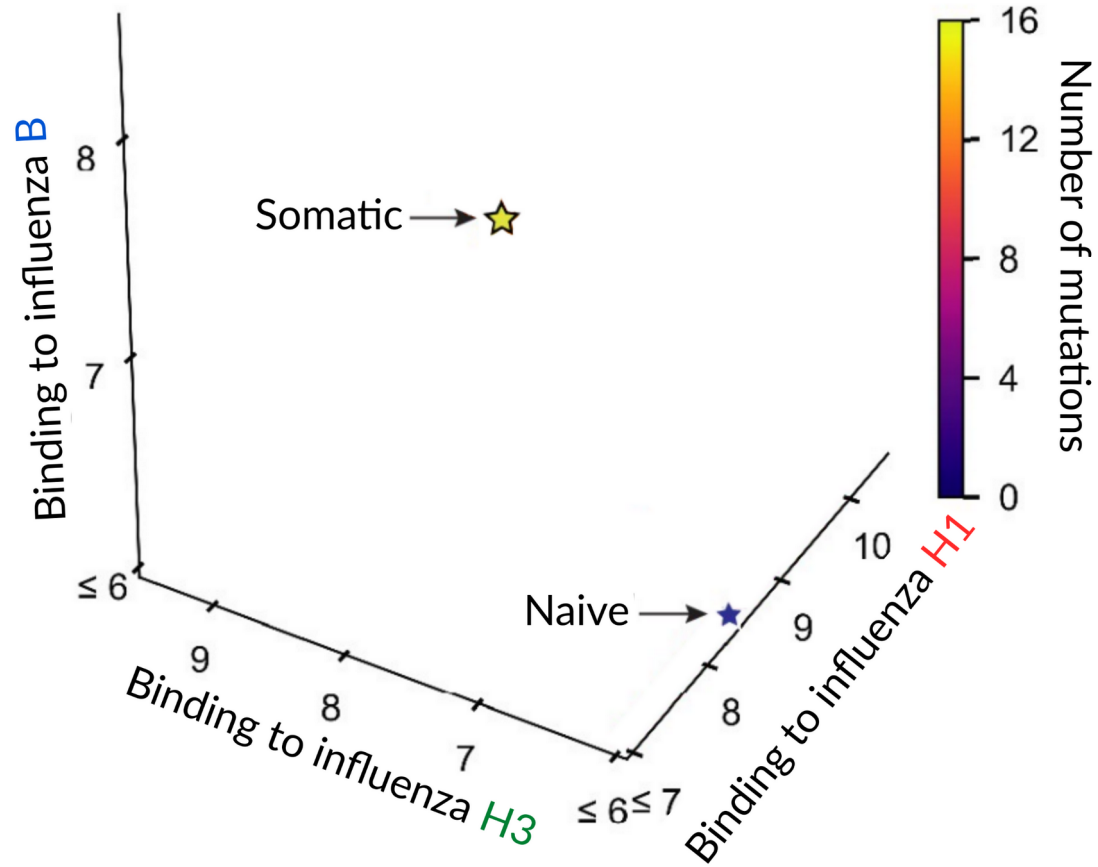


Broadly neutralizing antibodies offer protection against many variants of a virus but are rarely generated by the immune system.

How did they evolve ?

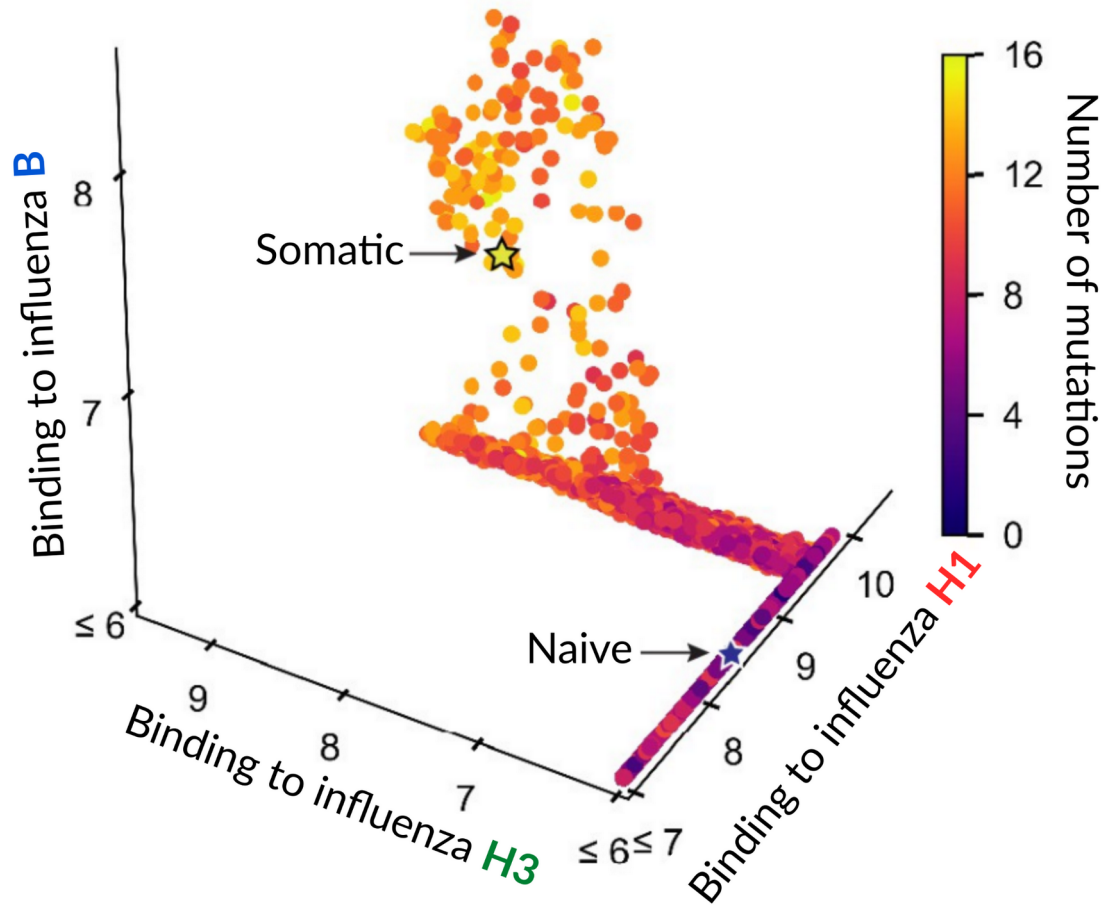
	FW1	HCDR1	FW2	HCDR2	FW3	HCDR3	FW4
CR9114 germline	QVQLVQSGAEVKKPGSSVKVSKASGGTFSSYAISWVRQAPGQGLEWMGGIIPIFGTANYAQKFQGRVTITADKSTSTAYMELSSLRSED TAVYYCARHGNYYYYYGM DVWGQGT TTVTVSS						
CR9114 somatic	.....S..... <u>SNN</u> .....D..... <u>S</u> ..... <u>STA</u> ..... <u>S</u> ..... <u>IFSN</u> ..... <u>N</u> ..... <u>T</u> ..... <u>F</u> ..... <u>S</u> .....						
		30 35 36		57 64-66	79 82-85	92 95 103	113
CR6261 germline	QVQLVQSGAEVKKPGSSVKVSKASGGTFSSYAISWVRQAPGQGLEWMGGIIPIFGTANYAQKFQGRVTITADKSTSTAYMELSSLRSED TAVYYCAKHMGYQLRETMDVWGQGT TTVTVSS						
CR6261 somatic	..... <u>E</u> ..... <u>P</u> ..... <u>R</u> ..... <u>P</u> ..... <u>TK</u> ..... <u>P</u> ..... <u>DFAG</u> ..... <u>V</u> ..... <u>M</u> ..... <u>V</u> .....						
		29 35		65 66 69	82-85 87		112.1

# Sequential exposure ?



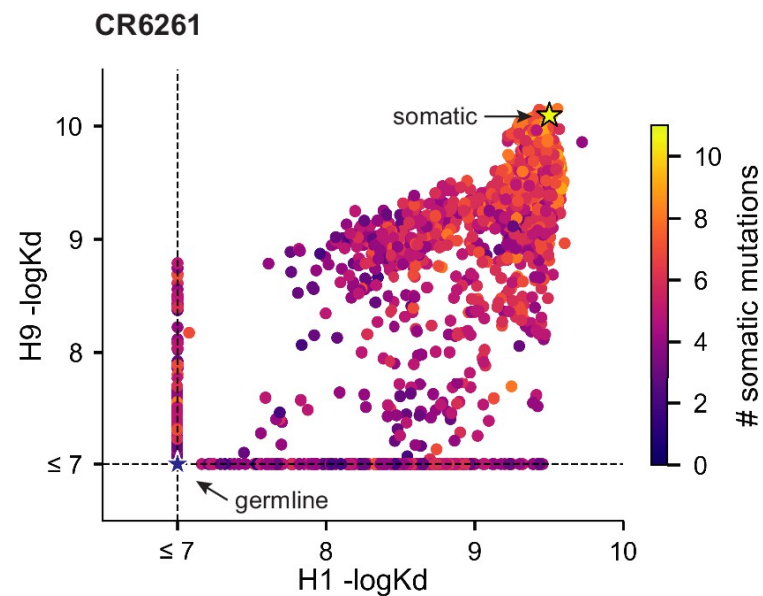
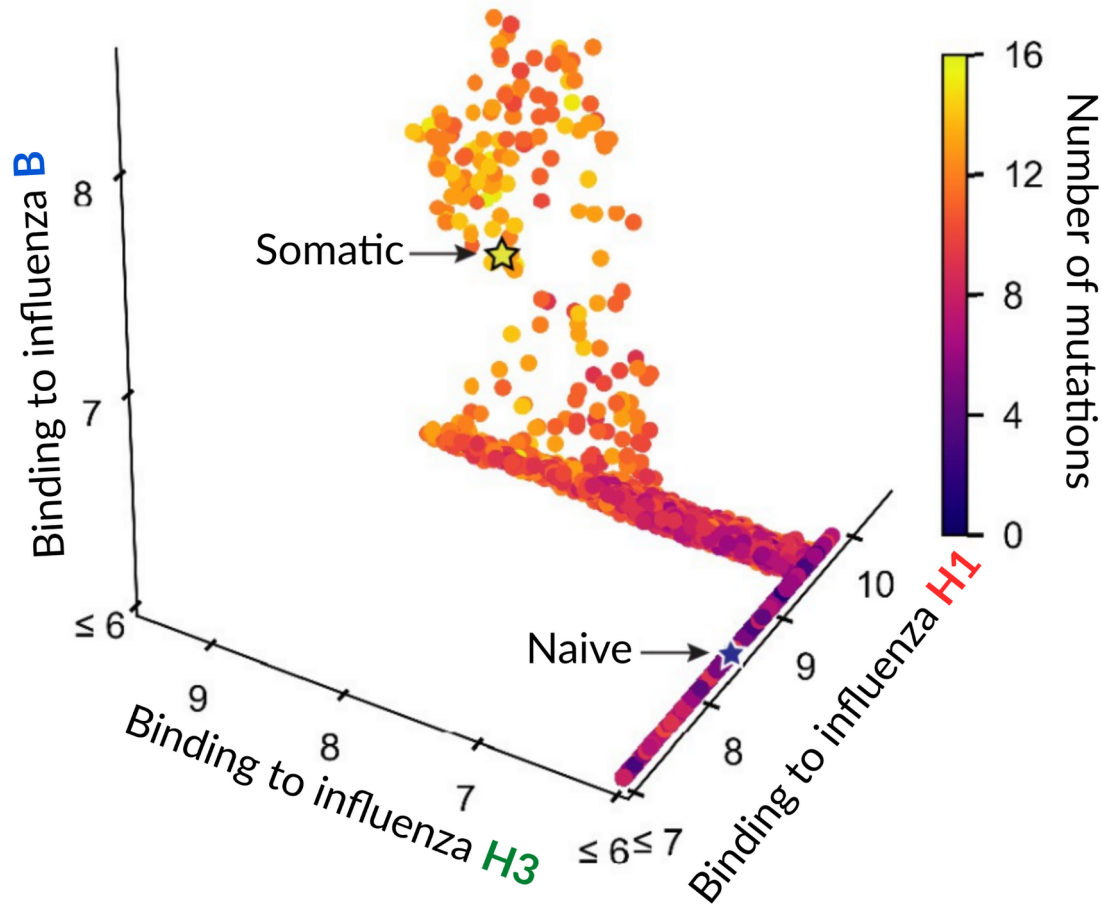
Phillips, A. M. et al. *elife*  
(2021)

# Sequential exposure ?



Phillips, A. M. et al. *elife*  
(2021)

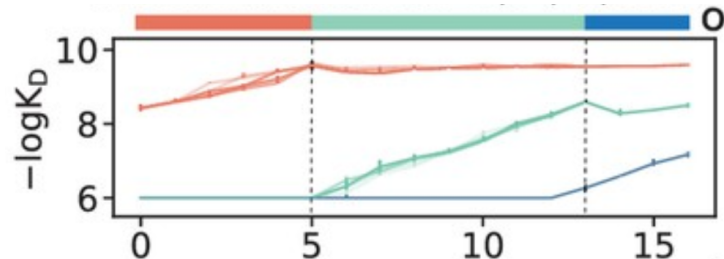
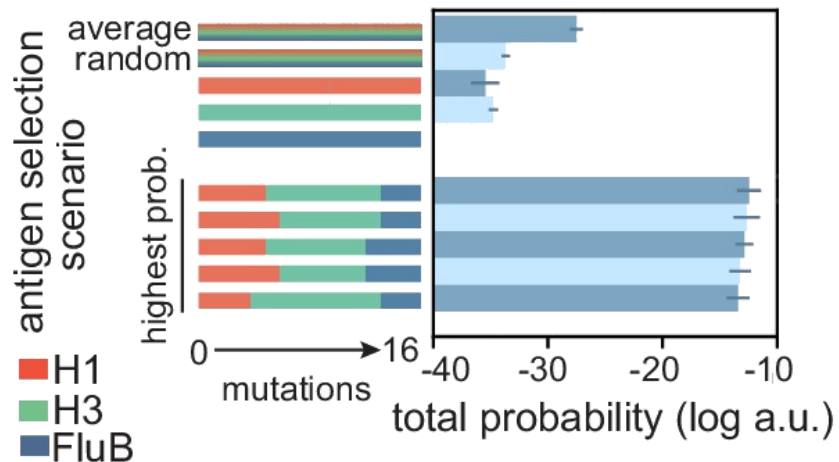
# Sequential exposure ?



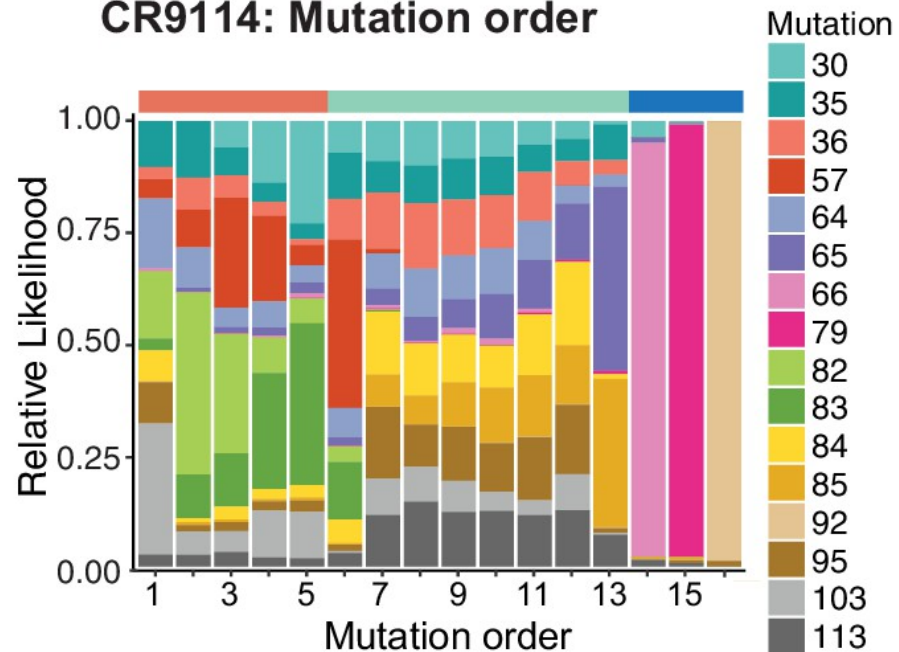
Phillips, A. M. et al. *elife* (2021)

# Predicting evolution

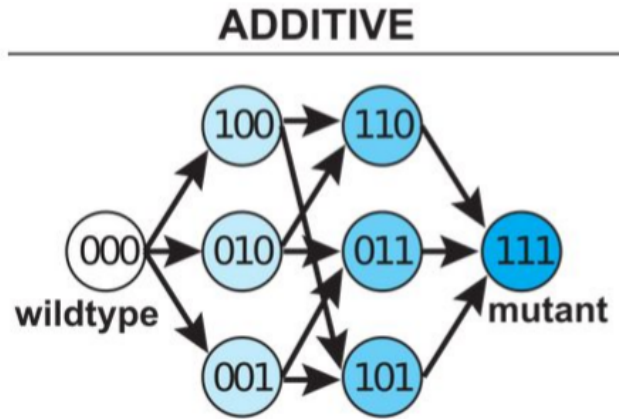
## CR9114: Scenario probability



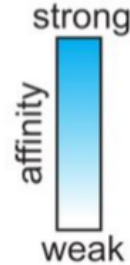
## CR9114: Mutation order



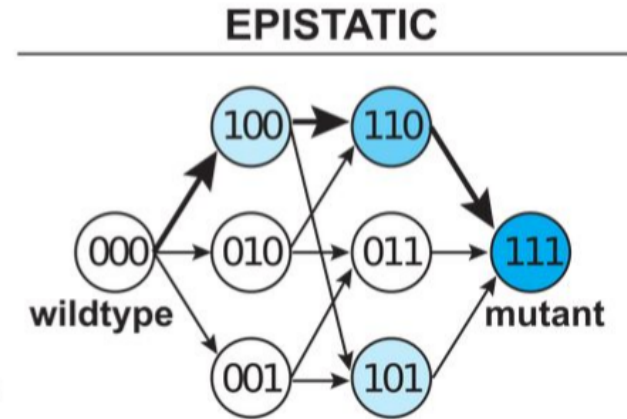
# Why does this work ?



mutation order doesn't matter



0: wildtype residue  
1: mutant residue

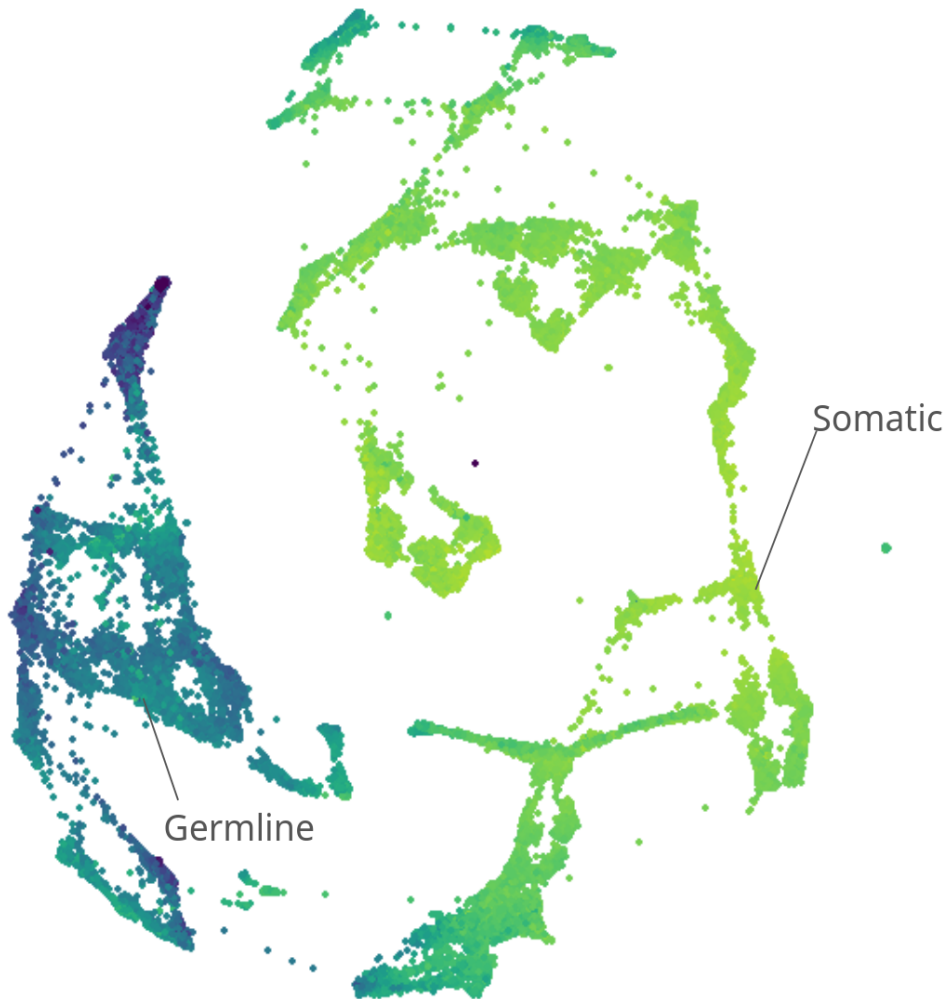


mutation order matters

We need interaction between mutations!

⇒ **Epistasis**

# Visualising Epistasis, Force-directed layout



Variants one mutation away are connected by an edge.

Nodes with similar binding affinity are drawn closer.

# Quantifying epistasis, linear regression

Least-square regression  
to infer the coefficients  
(+ regularisation)

$$f(\vec{\sigma}) = f_0 + \sum_i f_i \sigma_i + \sum_{i,j} f_{i,j} \sigma_i \sigma_j + \dots$$

- $\sigma_i = 0/1$ : Biochemical coefficients
- $\sigma_i = +1/-1$ : Statistical coefficients

# Quantifying epistasis, linear regression

Least-square regression  
to infer the coefficients  
(+ regularisation)

$$f(\vec{\sigma}) = f_0 + \sum_i f_i \sigma_i + \sum_{i,j} f_{i,j} \sigma_i \sigma_j + \dots$$

Independent of which mutations are present  
+ Match thermodynamic quantities ( $\Delta\Delta G$ )

- $\sigma_i = 0/1$ : Biochemical coefficients
- $\sigma_i = +1/-1$ : Statistical coefficients

# Quantifying epistasis, linear regression

Least-square regression  
to infer the coefficients  
(+ regularisation)

$$f(\vec{\sigma}) = f_0 + \sum_i f_i \sigma_i + \sum_{i,j} f_{i,j} \sigma_i \sigma_j + \dots$$

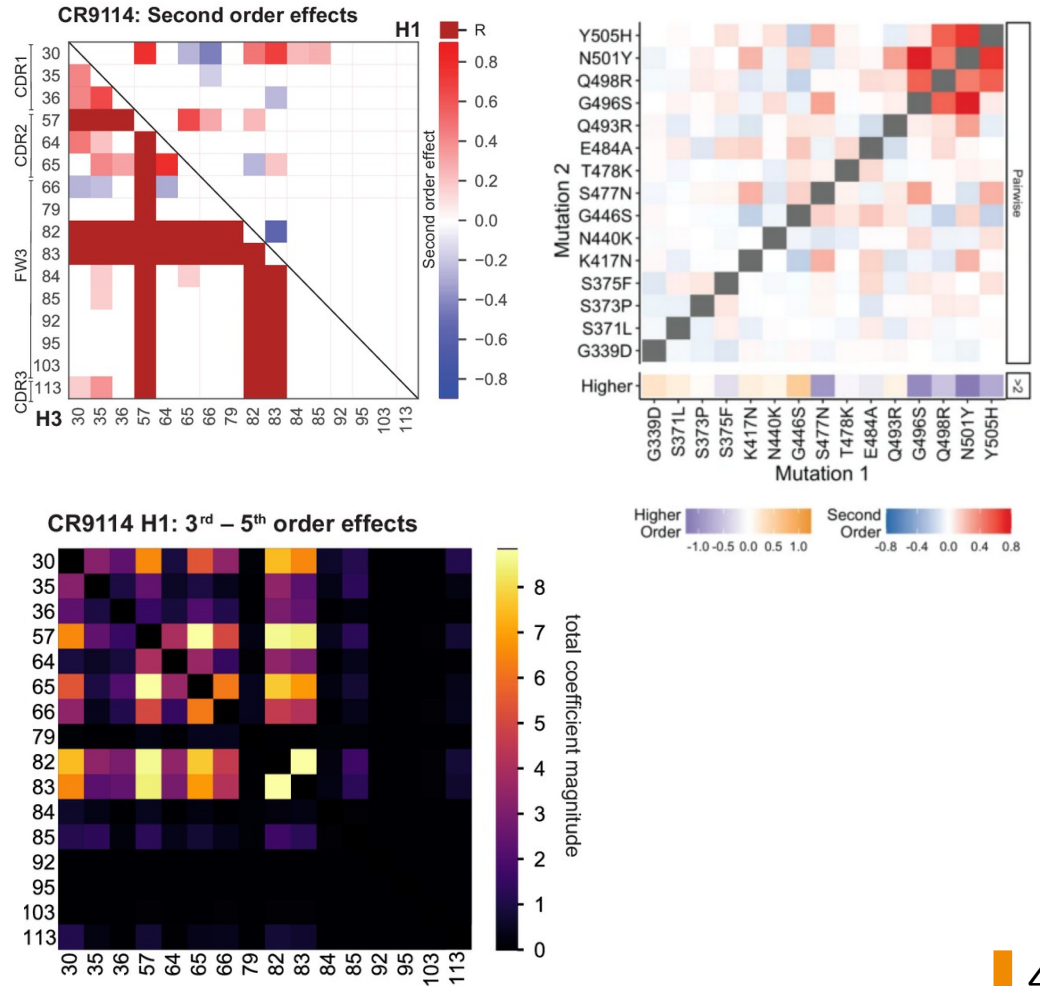
- $\sigma_i = 0/1$ : Biochemical coefficients
- $\sigma_i = +1/-1$ : Statistical coefficients

Independent of which mutations are present  
+ Match thermodynamic quantities ( $\Delta\Delta G$ )

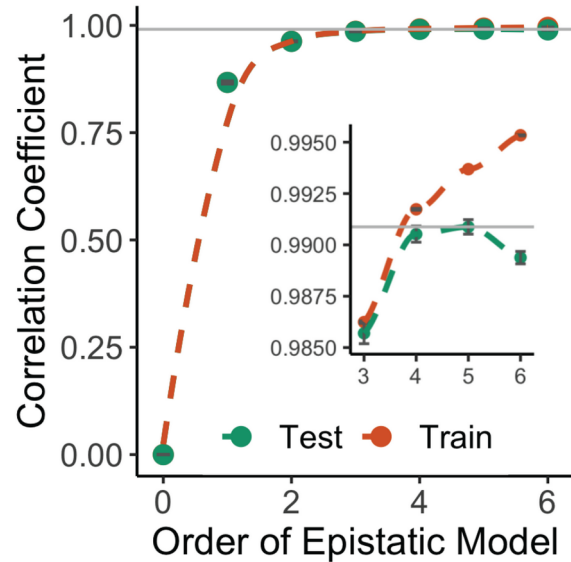
Summary statistics, equivalent to multidimensional Fourier transform

# Epistasis

We can measure all epistasis coefficients. In all landscapes, they are relatively **large**, as important as 1<sup>st</sup> order coefficients.

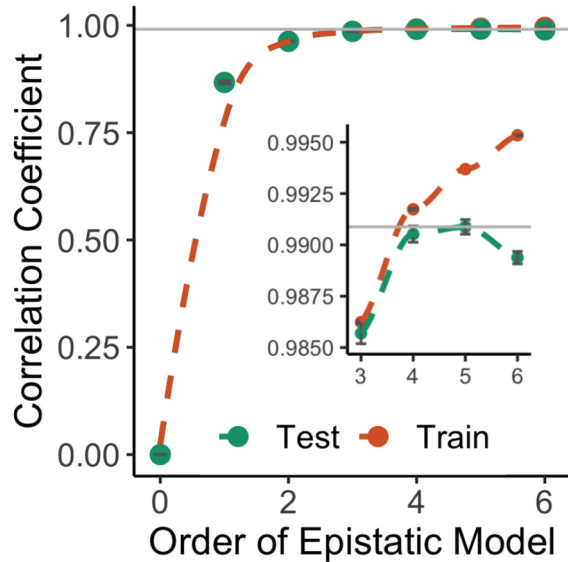


# Predicting epistasis ?



Most sequence-function predictions stop at first or second-order interactions

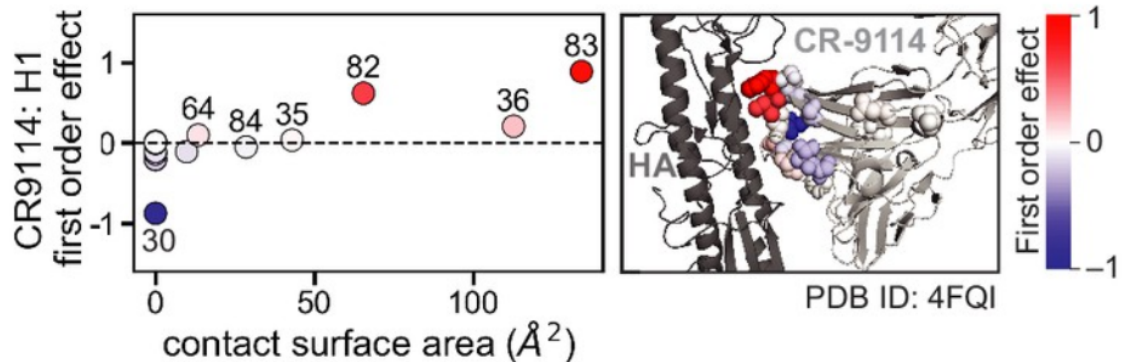
# Predicting epistasis ?



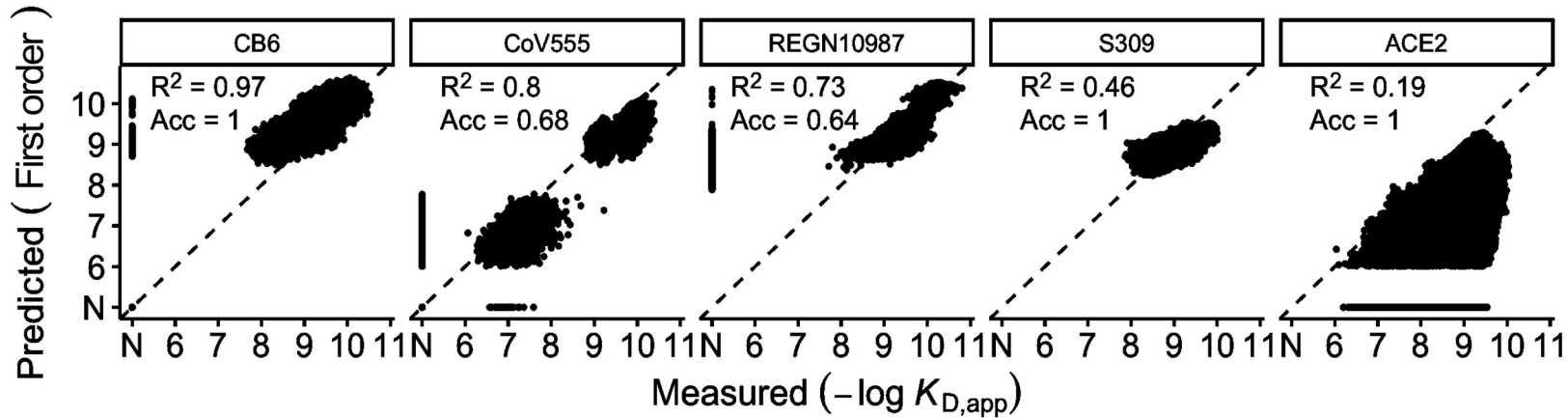
Most sequence-function predictions stop at first or second-order interactions

Structural data is not enough

First order effects vs. HA contact surface area



# Epistasis and evolution ?



Epistatic structure depends on evolutionary history ?

# Epistasis and evolution ?

Toy model, global landscape:

$$f(\vec{\sigma}) = f_0 + \sum_i f_i \sigma_i + \sum_{i,j} f_{i,j} \sigma_i \sigma_j + \dots$$

Normal, 0-centered, variance  $\sigma_{ij}^2$

# Epistasis and evolution ?

Toy model, global landscape:

$$f(\vec{\sigma}) = f_0 + \sum_i f_i \sigma_i + \sum_{i,j} f_{i,j} \sigma_i \sigma_j + \dots$$

Normal, 0-centered, variance  $\sigma_{ij}^2$

Jump probability:

$$P_{i,j} \propto \frac{1 - e^{f_i - f_j}}{1 - e^{N(f_i - f_j)}}$$



Equilibrium distribution:

$$P(f) \propto \exp(-f)$$

Satisfy detailed balance

# Epistasis and evolution ?

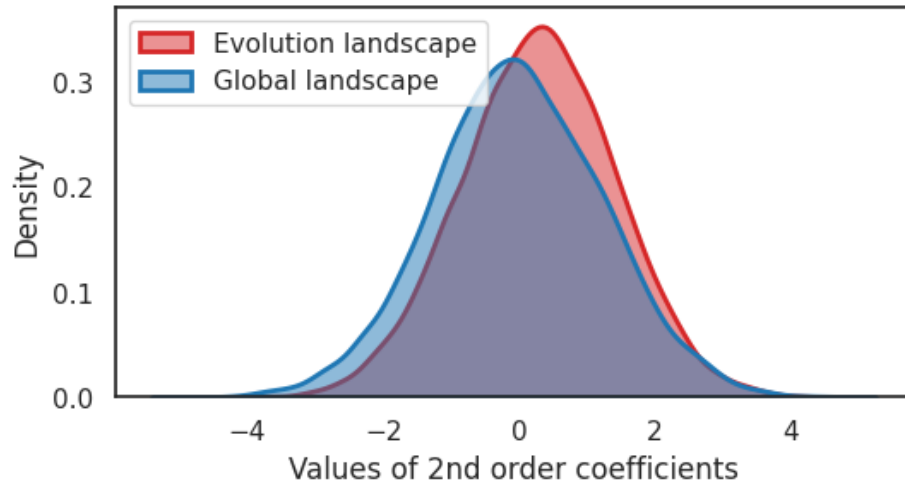
At equilibrium, the average epistatic coefficients are positive.

$$\langle f_x \rangle = \sigma_x^2$$

# Epistasis and evolution ?

At equilibrium, the average epistatic coefficients are positive.

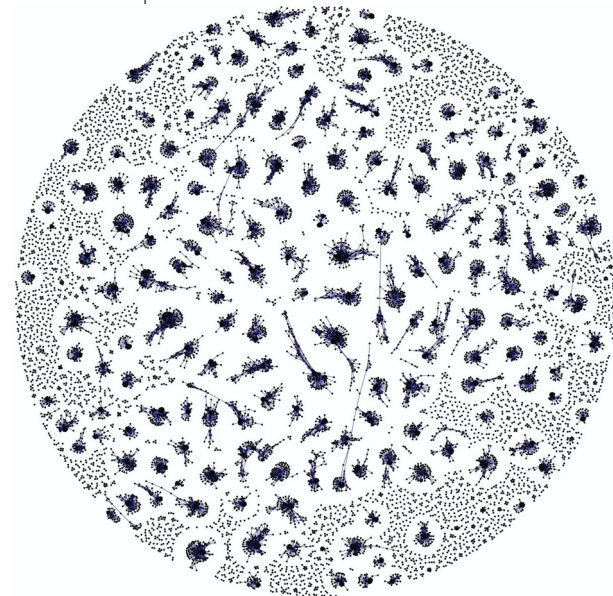
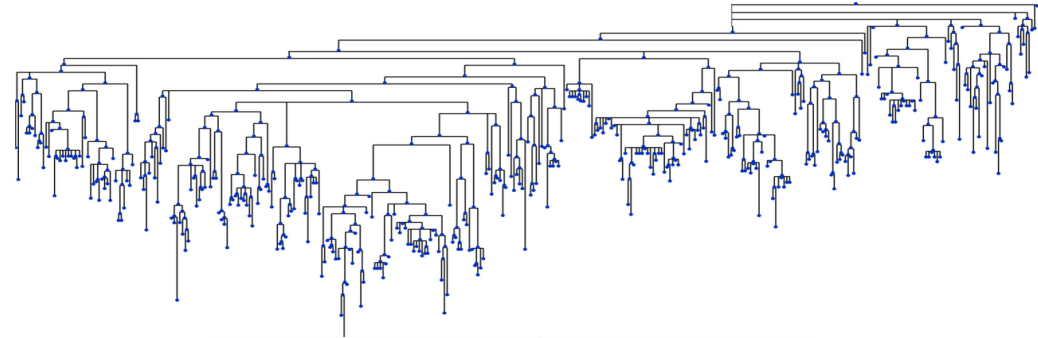
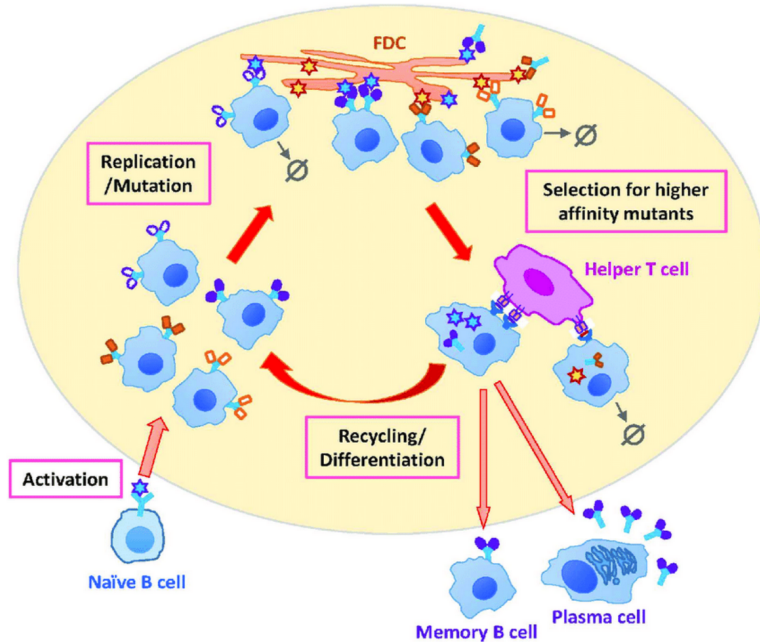
$$\langle f_x \rangle = \sigma_x^2$$



Epistasis is on average larger for evolving proteins.

1. Proteins evolution leaves traces in the epistatic network of its mutations.
2. Evolution favors interaction between mutations.
3. Proteins are complicated.

## What are the rules of antibody evolution ?



# THANKS !

Tatiana Ruiz-Bedoya

Giulio Isacchini

**Angela Phillips**

Zach Montague

Natanael Spisak

**Katherine Lawrence**

Michael Desai

**Alexandra Poret**

**Alief Moulana**

Thierry Mora

Daniel Newton

Joao Ascensao

Berk Alpay

Aleksandra Walczak

