

## **Internship proposed by S. Cocco (LPENS)**

### **Data-driven modelling of fast evolution in response to environmental changes in bacterial proteins**

The need to evolve in response to environmental changes has led living organisms to acquire powerful diversification strategies. Among these, Diversity-Generating Retroelements (DGRs) are natural directed mutagenesis systems capable of efficiently exploring sequence space. DGRs have recently been identified and characterized in populations of bacteria from the gut microbiota [1], allowing them to adapt to and colonize the human gut, particularly by diversifying ligand-binding proteins, including pilus-associated adhesins.

In this internship project, we propose to study the extent to which a generic protein, such as those described in [1], can undergo mutations with a DGR system. Our aim is to quantitatively characterize how this system is capable of generating a wide variety of protein variants with appropriate functional and structural properties.

This objective will require us to:

- establish new models to score protein sequences using unsupervised machine-learning techniques [2] that we will train on available sequence databases [1];
- study the dynamics of mutations with a population dynamics framework.

We may potentially extend the study to the ecological community level in the gut, considering interactions between different species to describe various environmental pressures, driven by both negative interactions (e.g., resource competition) and positive interactions (e.g., horizontal gene transfer) within the bacterial community, as experimentally investigated in [1].

Results will be compared to existing data [1] and, if time permits, will be used to generate new DGR template that could be tested by the experimental group of D. Bikard (Institut Pasteur) [3].

Our scope is therefore to advance the understanding of a fundamental biological mechanism, through machine learning, theoretical modelling and simulations, and to exploit it to design testable evolutionary dynamics.

The internship will be supervised in collaboration with Leo Régnier.

## Bibliography

- [1] [Targeted protein evolution in the gut microbiome by diversity-generating retroelements](https://doi.org/10.1126/science.adv2111). B. R. Macadangdang et al.. Science (2025) <https://doi.org/10.1126/science.adv2111>
- [2] [Learning protein constitutive motifs from sequence data](https://doi.org/10.1101/2025.03.24.644984) J. Tubiana, S. Cocco, R. Monasson. [eLife 2019;8:e39397 \(2019\)](https://doi.org/10.1101/2025.03.24.644984).
- [3] [Harnessing Diversity Generating Retroelements for \*in vivo\* targeted hyper-mutagenesis](https://doi.org/10.1101/2025.03.24.644984). R. Laurenceau et al. Biorxiv doi: <https://doi.org/10.1101/2025.03.24.644984>