M2 INTERNSHIP PROPOSAL

Laboratory name : Center for Interdisciplinary Research in Biology
Supervisors : Olivier Rivoire & Ivan Junier
e-mail : olivier.rivoire@college-de-france.fr
Web page : http://statbio.net
Internship location : Collège de France, 11 place Marcellin Berthelot, 75005 Paris

Generative Statistical Models of Genomes

One of the great challenges of biology is to decode the information contained in genomes. We now have the complete sequence of the genomes of many species. These genomes display both a remarkable similarity — they share orthologous genes of common evolutionary origin and comparable function — and a great diversity — they differ in their gene content as well as in the way in which genes are organized along the DNA [1].

Our goal is to understand the rules that govern the composition and organization of bacterial genomes [2]. To this end, we propose to build a statistical model based on a database of thousands of complete bacterial genomes. Given a putative genome, the model should tell us if it is a viable genome. It should also be generative and allow us to design new viable genomes or to propose modifications of current genomes that preserve their viability.

We will follow an approach that has proved to be successful in designing generative models of proteins [3]. The model will be based on the maximum-entropy approach and have a form reminiscent of Potts models in statistical physics. These types of models are defined by the statistical constraints that they enforce. One objective will be to identify the constraints that are necessary and sufficient to obtain a generative model of genomes. There will also be opportunities to make predictions to be compared with experimental data.

The project will require good skills in statistical physics or/and machine learning and an interest in biological questions. It will take place in an interdisciplinary team of physicists and biologists working theoretically and experimentally on related projects and involve a collaboration with Ivan Junier (TIMC, Grenoble).

References:

[1] E Koonin, The Logic of Chance: The Nature and Origin of Biological Evolution. 2011

[2] I Junier & O Rivoire. Conserved units of co-expression in bacterial genomes: an evolutionary insight into transcriptional regulation. (2016) PloS one, 11(5), e0155740.

[3] W. P Russ et al. An evolution-based model for designing chorismate mutase enzymes. 2020. Science, 369 (6502), 440-445.