Internship proposal

Comparing ancient and recent coevolution in proteins

Supervisors

Guillaume Achaz (Muséum d'Histoire Naturelle & Collège de France)
Olivier Rivoire (Collège de France).

Location

Collège de France (Statistical Biology team) & Muséum (Atelier de Bioinformatique), Paris.

Subject

Evolution is often described as an irreversible process, but the extent to which current evolution differs from past evolution is rarely quantified. The goal of the internship will be to address this question by analyzing coevolution in protein sequences at different time scales.

In proteins, coevolution refers to the non-independence of substitutions of amino acids. It generically arises from adaptive and physical constraints. Several computational methods have been developed to infer coevolution from available datasets of protein sequences. Broadly speaking, these methods fall in two categories: (1) methods that analyze correlations between amino acids by treating the sequences as statistically independent samples; (2) methods that infer the co-occurrence of substitutions by reconstructing the evolutionary history (phylogeny). These approaches have very distinct ranges of applications: the first class of methods is applicable only to sequences that are sufficiently dissimilar while the second is conversely applicable only to sequences that are sufficiently similar. Each class of methods thus describes coevolution at a very different evolutionary time scale.

The internship will consist in comparing results from these different approaches and in confronting them to structural and functional informations, as well as to experimental measurements of the phenotypic effect of mutations. From the comparison, we will characterize the differences between recent and ancient evolution, and identify the time scales at which these differences are significant.

The internship requires basic skills in programming (e.g. python or matlab) and statistics but no prior knowledge on coevolution or proteins. The work will involve concepts and methods from bioinformatics, statistical physics and evolutionary biology, which the student should be open to learn during the internship. The work can possibly be followed by a PhD thesis.

References

- A Behdenna, A Lambert, J Pothier & G Achaz. Testing for independence between evolutionary processes. Systematic Biology (2016), 65(5):812-23
- L Ferretti, D Weinreich, B Schmiegelt, Y Kobayashi, A Yamaushi, F Tajima & G Achaz. Measuring epistasis in fitness landscapes: the correlation of fitness effects of mutations. Journal of Theoretical Biology (2016), 396:132-43
- O. Rivoire, K. Reynolds, R. Ranganathan. Evolution-based functional decomposition of proteins. PLoS Comput Biol (2016), 12:e1004817
- O. Rivoire. Elements of coevolution in biological sequences. Phys. Rev. Lett. (2013), 110:178102

Contacts

guillaume.achaz@mnhn.fr

olivier.rivoire@college-de-france.fr