THÈSE

Soutenance de thèse d'Hugo Menet

Multi-scale phylogenetic approaches for the evolution of the holobiont

Jury :

Lars Arvestad, Professeur, Université de Stockholm, Suède, Rapporteur

Catherine Matias, DR CNRS, Rapporteure

Gergely Szöllősi, Chercheur, Université Loránd Eötvös, Budapest, Hongrie, Rapporteur

Sabine Peres, Professeure, Lyon 1, Examinatrice

Eric Tannier, DRINRIA, Directeur de thèse

Vincent Daubin, DRCNRS, Co-directeur de thèse

Abstract :

Biological systems like holobionts are made of entities at many scales (macro-organisms, micro-organisms, genes...), which are, on the one hand, bound to a common history because they all function together and depend on the others, and on the other hand, driven by their individual interests. The evolution of such a system is approached by phylogenetic reconciliation, which describes the coevolution of two different levels, genes and species, or hosts and symbionts, for example. The limit to two levels has confined the use of reconciliation to either molecular studies on genes and species trees or ecological studies on host-symbiont associations. The holobiont concept is an occasion to gather all these scales by modeling multi-level inter-dependencies. In this thesis, we explore and extend reconciliation to model such multi-level systems

Phylogenetic reconciliation is a phylogenetic method that arose at the interplay of two communities, the coevolution of host and symbiont, and the comparison of gene and species trees. Lately, despite this initial development, these two communities tend not to interact much, even if they have a lot to learn from each other. We review the development of these methods, take a generic approach, and highlight the new advances that propose more integrative models, reaching out for multi-level reconciliation.

In recent years, new methods have been proposed to integrate the evolution of

species, gene and gene domain, or geography, host and symbiont, but none have yet investigated the levels at the heart of the holobiont: host, symbiont and genes, and none in a probabilistic setting and with horizontal transfers. We reimplemented ALE, a probabilistic DTL reconciliation software, and extended it to consider the reconciliation of three levels: host, symbiont, and gene. This new probabilistic model of the evolution of three nested levels allows gene transfer, host switch, gene duplication, symbiont diversification inside a host, and gene or symbiont loss. Given three phylogenetic trees, we devise a Monte Carlo algorithm able to infer joint scenarios and compute their likelihood in the model, accounting for gene transfer rates' dependence on host symbiont reconciliation as well as the impact of ghost lineages on these rates. As in ALE, we use amalgamation to take into account uncertainty in the gene trees, but also to infer the symbiont tree using universal unicopy genes as a topology distribution for the symbiont tree. This method was evaluated using a simulated dataset on which we showed its capacity to distinguish models of 2-level and 3-level coevolution using the computed likelihood. It is able on aphids/enterobacteria systems to retrieve transfers better than the host unaware method.

With potentially an exponential number of most parsimonious solutions, reconciliation output can be hard to interpret, notably when considering multiple sampled scenarios or multiple gene families, moreover when we want to look at multi-level systems. Few graphical software exists, and none are generic and can use RecPhyloXML, a common format endorsed by the gene species community. We propose Thirdkind, a software we developed, able to output a graphical display of a reconciliation scenario as an SVG file. It is easy to use and install. It can handle the embedment of three trees that is the output of our 3-level reconciliation framework and can resume the evolution of multiple gene families or sampled scenarios in a single figure by aggregating redundant transfers.

A fascinating example of complex coevolution history is the relationship between Helicobacter pylori and its human host. Helicobacter pylori is a bacterial pathogen that is believed to have followed its human host during its ancestral migrations, during the colonization of Africa, Asia, Europe, Oceania, and the Americas. The bacterial strains are structured in populations whose geographical repartition is mostly congruent with that of their host. One of the significant discrepancies is the European population, which seems to result from introgression between two ancestral populations, one related to modern African and the other to modern Asian. These hypotheses rely on Bayesian models of SNPs attribution to populations, for whole genomes, or a small subset of genes via Multi Locus Sequence Typing. We took a more phylogeny-focused approach using a dataset constructed in the team, with a phylogeny for 120 strains, comprising the ancestral strain found in Ötzi, dated to 5 kilo years ago, and 1034 gene trees. We applied reconciliation to gene trees and population trees to better understand the mixed origins of the genes in the European population. This new approach, which relies on matching certain leaves of the gene trees (here the European ones) uniformly to all the leaves of the upper tree and then looking at the posterior probability of matching, could be easily

transposed to other problems. We also used our 3-level reconciliation framework to test different population trees