2018-00645 - Algorithmic investigation of the dynamics of species interactions [PhD Campaign]

Level of qualifications required: Graduate degree or equivalent
Function: PhD Position

About the research centre or Inria department
Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context
The Ph.D. will be carried out in the ERABLE project-team whose head is Marie-France Sagot.

The main research topics of the team include computational biology, systems biology, design and analysis of algorithms, evolution and network modelling, species interactions.

The Ph.D. will be co-advised by Marie-France Sagot (DR Inria / HDR), Mario Figueiredo (Full Professor, Instituto Superior Tecnico and University of Lisbon, and also member of the Inria Associated Team Compasso, http://team.inria.fr/erable/en/projects/inria-associated-team-compasso/ ), and Blerina Sinaimeri (CR Inria in the ERABLE Team). Visits to Mario Figueiredo in Lisbon, Portugal, will be made possible using the funds available in particular from Compasso.

Web-page of ERABLE: http://team.inria.fr/erable/

Assignment
Ph.D. research project description:
Cophylogeny allows to reconstruct the coevolutionary history of groups of species (e.g. hosts and symbionts) that interact over a long period of time and in ecologically close environments based on their phylogenetic information. The most used method in
cophylogeny studies is phylogenetic tree reconciliation [1, 2, 3, 4].

Algorithmic efficiency is a crucial problem in all reconciliation methods. The main issue concerns the complexity of the data in terms of input and output. On one hand, in many cases the phylogenetic trees used as input can be very large and this may represent a problem for the existing cophylogeny algorithms. On the other hand, the current cophylogeny model leads in many cases to a number of optimal solutions that is unrealistically large and thus difficult to be analysed. Such model is also incomplete in terms of the macro-evolutionary events that are possible. The main goal of this PhD is then to design refined models that are closer to reality, and new algorithms that will be able at the same time to work with such refined models and tackle the complexity of the problem in relation to both input and output.

Main activities
Three main issues will be addressed.

The first one is related to the complexity of the output: even when the model is refined in order to account for events that are not yet considered in the majority of the existing algorithms, we expect that in general there will be a number of optimal solutions that is unrealistically large [5], making it practically impossible to analyse each one of them separately. To address this issue, we plan to define equivalence relationships for grouping the optimal solutions of a cophylogeny reconciliation problem and design efficient algorithms for enumerating one single representative per class. As this represents a highly challenging problem, a different approach that can be pursued in parallel is to reduce the number of optimal solutions by including more constraints on the problem, such as for instance coming from geographical information, as these may reduce the number of optimal solutions.

The second issue will then be to extend the cophylogeny reconstruction models and methods to make them biologically more realistic, in particular by allowing a symbiont to be associated to more than one host as is often the case in nature.

Finally, the third issue will be to turn the focus on the complexity of the input. Indeed, most often, the phylogenetic trees used as input can be large (in the
case of the Wolbachia-arthropod dataset [6,7] for instance, these have approximately 700 leaves) and also may contain uncertainties related to, e.g., incomplete or erroneous data. The objective in this case will be to design and implement algorithms for efficiently dealing with large and incomplete and/or erroneous datasets. In the first case, parameterised or approximation approaches will be explored. In the second, we will in particular concentrate attention on the problem of polytomies, i.e. of internal vertices of the phylogenetic trees that have more than two children, and more specifically on the case of the so-called hard-polytomies where a common ancestral population speciates into multiple lineages. To the best of our knowledge, the hard polytomies are considered only in gene-species systems where reconciliation algorithms allowing for non binary species trees have been proposed [8].

Bibliographic references:
Applicants should already have or are expected to shortly obtain a M.Sc. degree in computer science, computational biology or some other relevant discipline. The ideal candidate should have a solid knowledge on the design and analysis of algorithms.

An experience already in evolution would also be very much appreciated. The candidate should demonstrate good programming skills (e.g. Java/C). Previous experience with solvers such as Cplex or similar would be considered a plus.

**Benefits package**
- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

**Remuneration**

Monthly salary after taxes: around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).