

COMPUTATIONAL GENOMICS: PROBLEMS, ALGORITHMS AND MODELS

Abstract

The scientific objective of this project is to develop new methods and tools for the study of evolution of biological entities such as genes, genomes and molecular sequences. The project consists of three interdisciplinary topics, whose objectives are addressing algorithms, mathematical theory, computations and applications in particular in metagenomics and cancer research. Our solutions will be developed and implemented in a close international cooperation between mathematicians, computer scientists, biologists and clinicians.

The following list presents three objectives for our proposal:

- Development of models, algorithms and tools for the inference of gene-genome relationships for metagenomic datasets with special focus on horizontal gene transfer (HGT).
- Discovery of advanced mathematical properties of tree reconciliation based cost functions in the context of matching between genes and genomes.
- Development of models, algorithms and tools for the reconstruction of evolutionary trees of B-cells with applications to leukemia cancer data.

Basic research in this project corresponds to mathematics and computer science with a special focus on topics in computational biology and medicine. In particular, we plan to develop novel algorithms, that will expand our knowledge on the possibilities to process molecular data of complex nature. In addition, we expect new mathematical results on functions that are often applied to compare genes and their corresponding genomes. Such functions will be also studied in the topic related to metagenomics, where the relation between a gene and a genome is a component of a standard question (e.g., what genomes are present in a metagenomic sample?). A central notion in this project is the notion of an evolutionary tree, that can represent relationships between biological units such as genes, genomes or species. In this project we will also propose novel algorithms for the reconstruction of the evolutionary tree of B-cells for DNA sequences from a single individual (a patient). The nature of such sequences is special, therefore there is a need for developing solutions that will take into account their specificity. Our algorithms will be tested on real and simulated data. In particular we plan tests on DNA sequences from leukemia cancer and on the sequences from metagenomic samples.

Topics proposed in this proposal are not only in the field of interest of computational biology and bioinformatics researchers, but also address important problems in personalized medicine related to leukemia cancer research. Among many theoretical elements of basic research nature in the area of computer science and mathematics, we emphasize the potential of our methods and tools, that will be developed on the basis of our algorithms. Novel methods and algorithms proposed by our team will significantly improve the possibility of analysis of large complex datasets from metagenomics, genomics and medicine.