# Biologically Meaningful Inference of Phylogenetic Networks

Popular Scientific Abstract

Without doubts networks are present everywhere. Internet is a network, commute system is a network, cells create networks, human interactions are networks, and so on.

Everything is a Network

The same applies to biology. In particular various types of relationships between sequences, genes, genomes or species can be represented in the form of a phylogenetic network.
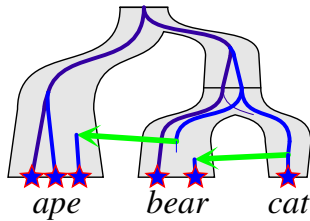
Gene-species network for three species and six genes (blue stars) with two horizontal gene transfers (green arrows).
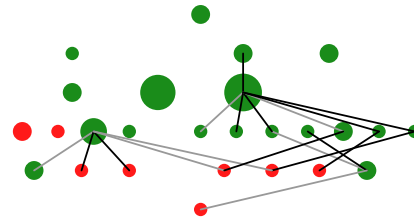
An example of a network of highly similar DNA sequences (nodes) obtained from human white blood cells (B-cells).

This project investigates the development of novel biological models and efficient algorithms to identify and study phylogenetic networks. To successfully address this multidisciplinary endeavour, we have assembled an international team of experts in biology, medicine, mathematics, and computer science.

In this project, we propose to study and analyse a novel generic way to compare a phylogenetic tree with phylogenetic network. Using this general tree-network comparison framework, we propose to solve several classical problems in computational biology including phylogenetic network inference, and to develop tools for solving these problems. Techniques to be employed will include standard mathematical proof techniques, graph theory, rule-based, genetic algorithms, standard algorithmic design paradigms, run-time complexity analysis and so on. For the data processing and validation of results, we will develop phylogenetic and simulation pipelines. In these uniform environments, we will prepare a collection of empirical and simulated benchmark datasets for testing our tools. Validating our results will be conducted in close collaboration with biological and medical experts.

The proposed framework for comparing trees and phylogenetic networks provides to practitioners new tools to study complex evolutionary relationships between sequences, genes, genomes and species using biologically meaningful models. Since relationships in genomic datasets do not often follow tree-like patterns, these solutions will significantly improve the possibility of analysis and help researchers to understand better the underlying biological processes.