

# CREDIBILITY OF HORIZONTAL GENE TRANSFER MODELS

## Abstract

The evolution of gene families is a complicated process that is not always reflecting the evolution of the species they come from. These differences between genes and species evolution may be explained by using the tree reconciliation model in which differences between a gene tree and its species tree are explained by the presence of duplication and loss of genes. In studies related to the bacteria, apart from duplications and losses, horizontal gene transfer (HGT) should also be taken into account. However, the extension of reconciliation model with HGT makes the algorithms computationally more complex which significantly affects the performance. So far, several algorithms for HGT inference had been developed for simplified models that allow for the polynomial time complexity. Despite a good knowledge of computational properties of these algorithms, little is known about their applicability to real data. In particular, the study of the reliability of their results is still an open question. In this project we plan to take the following scientific objectives:

1. Development of methods for determining the credibility of scenarios with HGT .
2. Analysis of the algorithms with particular emphasis on their applicability to empirical and simulated data.
3. Applying the elaborated methods to a variety of problems concerning HGT, in particular to the problem of inference of gene-species relationships in the context of metagenomic experiments.
4. Development of effective algorithms and their implementation in the form of publicly available tools.

The starting point of our research is the algorithm for determining credibility of embedding of a gene tree into a species tree which was developed by us and will be presented in June 2016 at the AICoB conference, and published in LNCS. Our tool allows for estimation of credibility of the entire embedding basing on the credibility of the particular evolutionary events - gene duplications and speciations. In the first step we intend to expand the existing solution with a model taking HGT into account. Then, using both empirical and simulated data, we intend to investigate properties of algorithms for HGT reconstruction. Basing on our algorithms and models we want to create hierarchy of algorithms for HGT reconstruction, and then investigate the credibility of their results. The data for the experiments will be simulated with the help of our program, which will be extended with a model of HGT. The elaborated method of assessing embedding credibility will be also utilized in the context of the reconstruction of gene-species relationships in metagenomic experiments. Next, we will examine the influence of the integration of information for many gene families on the improvement of credibility of reconstructed gen-species relations.

Microorganisms are an extremely important element in the Earth's ecosystem. Within the bodies of people and animals bacteria help in digestion and decomposition of toxins, strengthen immunity, carry synthesis of vitamins and fulfill many other necessary functions. Bacteria are also important for industry. They are used for the production of antibiotics and fertilizers, for production of food requiring fermentation, as well as for wastewater purification from industrial pollution. Studies on HGT can help in better understanding of the bacteria evolution so they could of a great importance for medicine, industry and environment protection. One of extremely important issues is bacterial antibiotic resistance, which can be acquired by subsequent bacterial strains with the use of HGT mechanisms. Our algorithms could help to facilitate the analysis of bioinformatic metagenomic experimental data and increase the credibility of the obtained results.