Abstract

The recent epidemic growth of diseases related to metabolically abnormal conditions has raised questions about how correct is our understanding of the nature of those diseases, and to what extent can they be explained exclusively in terms of genotypical causes. Additionally, the use of high-throughput methods has produced a significantly larger amount of metabolic data yet to be thoroughly analyzed. In particular, the development of computational methods to understand the metabolic similarities among different species has gained an increased interest as a way to help close the genotype-phenotype gap, i.e., how small differences at the genotypical level can produce the greatly diverse phenotypes we observe in nature. Metabolic pathway alignment is a promising approach based on the idea of establishing a correspondence between metabolic reactions, in a similar way to how sequence alignment establishes a correspondence between nucleotides or amino acids.

In this thesis, we present a new method for metabolic pathway alignment based on the similarity of enzymes, compounds, and reactions present in the pathways. A maximum similarity score is calculated by aligning reactions in both directions from one pathway to the other. We also present several applications of our method to problems of biological interest: phylogenetic reconstruction from metabolic similarity, election of model organisms metabolically similar to humans for specific diseases, and detection of conserved reactions among a set of organisms. A web server implementing the phylogenetic reconstruction functionality is also described, together with a standalone distribution of the code.

Our approach has several advantages over previous methods: we only rely on metabolic information, we can establish the relative importance of enzymes and metabolites in the global metabolic similarity, and our approach is computationally faster than graph-based methods. Results presented in this thesis show that our method outperforms previous approaches for phylogenetic reconstruction based on comparison of metabolism. Furthermore, we show how filtering of noisy data, use of complete metabolic information and fuzzy clustering results in highly accurate phylogenies. Finally, by studying the alignment produced by our method on the metabolism of several species, we were able to identify fundamental processes in different species, as well as probable misannotations of reactions in a metabolic information repository.