

Human Protein–protein Interaction Prediction by A Novel Sequence-based Co-evolution Method: Co-evolutionary Divergence

Shin-Sheng Yuan

Institute of Statistical Science, Academia Sinica, Taiwan

Abstract

Motivation: Protein–protein interaction (PPI) plays an important role in understanding gene functions, and many computational PPI prediction methods have been proposed in recent years. Despite the extensive efforts, PPI prediction still has much room to improve. Sequence-based co-evolution methods include the substitution rate method and the mirror tree method, which compare sequence substitution rates and topological similarity of phylogenetic trees, respectively.

Although they have been used to predict PPI in species with small genomes like *Escherichia coli*, such methods have not been tested in large scale proteome like *Homo sapiens*.

Result: In this study, we propose a novel sequence-based co-evolution method, co-evolutionary divergence (CD), for human PPI prediction. Built on the basic assumption that protein pairs with similar substitution rates are likely to interact with each other, the CD method converts the evolutionary information from 14 species of vertebrates into likelihood ratios and combined them together to infer PPI. We showed that the CD method outperformed the mirror tree method in three independent human PPI datasets by a large margin. With the arrival of more species genome information generated by next generation sequencing, the performance of the CD method can be further improved.