

Saccharomyces cerevisiae Protein-Protein Interaction Network Reconstruction to Study Ethanol Tolerance

Ivan R. Wolf, Lauana Fogaça, Leonardo Nazário de Moraes, Rejane M. T.Grotto, Rafael P. Simões, Guilherme Valente

Department of Bioprocess and Biotechnology. São Paulo State University (UNESP), Botucatu

The interest to develop biofuels such as bioethanol have been increasing. The most common process to produce bioethanol is the first generation technology, in which the most widely used organism is *Saccharomyces cerevisiae*. However, the high concentration of ethanol produces toxicity to *S. cerevisiae*, which is the main limiting factor to produce this fuel. Despite lots of effort to understand this phenomenon, the ethanol tolerance on the systemic view point is poorly understood. In the present study, the highest ethanol tolerance was experimentally determined for five yeast strains (S288c, BY4741, BY4742, SEY6210 and X2180-1A), in which unsupervised learning was applied over the experimental data matrix to classify the strains as highest tolerant or lowest tolerant. For the reconstruction of protein-protein interactions (PPIs) networks, all protein sequences for those strains were obtained from the *Saccharomyces* Genome Database and all protein pairs were submitted to UNISPPI PPI prediction, in which each protein pair was defined as an edge of the network. Three scores were calculated for each edge: (I) probability of protein interaction from UNISPPI; (II) the percentage of shared sub-cellular locations in YeastGFP database; (III) the percentage of shared sub-cellular locations in ComPPI database. In order to filter out the edges, they were split into four datasets: (A) edges with scores I, II and III; (B) interactions with scores I and II; (C) edges with scores I and III; (D) edges only with score I. The Principal Component Analysis (PCA) was applied on each dataset, independently, and edges inside the limits established from first to fourth quartile over PCA coordinates were selected. The dataset D consists only of UNISPPI score, and in this case only edges with score ≥ 0.724 were selected. The procedure was efficient to reconstruct the networks since the degree distribution follow a power law, fitting the Barabási-Albert model; and they also have a scale-free topology. The average degree and average betweenness centrality were high for all nets, which express the presence of hubs. Moreover, the networks are assortatives, which means that the nodes with high degree are preferentially connected on each other, following the principle of “rich-get-richer” dynamics. Those global topological characteristics fit with the expectations of a real biological network. Furthermore, a clustering analysis over all nodes degree for all networks showed that the degree is a property closely related to the highest ethanol tolerance on yeast (clusters obey the percentages of ethanol tolerance).

Funding Support: BIOEN FAPESP and PROPe UNESP.