Integrating omics data from xylose-fermenting yeast using network dynamic modeling for bioethanol production

Lucas Miguel de Carvalho[1], Gabriela Vaz de Meirelles[1], Renan Pirolla[1,2], Leandro Vieira dos Santos[1,2], Gonçalo Amarante Guimarães Pereira[1] and Marcelo Falsarella Carazzolle[1].

(1) Laboratório de Genômica e Expressão, Instituto de Biologia-Unicamp, (2) BioCelere Agroindustrial LTDA-Campinas-SP, (3) Laboratório Dalton de Espectrometria de Massas, Instituto de Química-Unicamp.

Brazil is one of the biggest producers of ethanol in the world, a pioneer in the ethanol industry. However, the country is already facing a major limitation imposed by the first-generation ethanol production technology, which the sugarcane juice is converted by ethanol using industrial yeast Saccharomyces cerevisiae. Therefore, a new alternative has been proposed, called second generation, which is based on lignocellulosic residues of sugarcane (bagasse and straw) for ethanol production using recent methodologies for biomass desconstruction that generates soluble sugars, majority represented by glucose and xylose. One of the biggest challenges of this technology is the development of genetically modified industrial yeast that can not only produce ethanol from glucose as usual, but also from pentoses that represents 15% to 45% of the lignocellulosic material. Several works have developed xylose-fermenting yeast using different exogenous genes and genetic engineering approaches, but always resulting in very low yield and productivity mainly caused by unbalanced redox potential and metabolic bottleneck. The combination of omics data (transcriptomic, proteomic and metabolomic) and bioinformatics analysis is an essential step for a better understanding of this system. The objective of this work is to develop bioinformatics analysis for integration of omics data from xylose-fermenting yeast in different fermentation conditions. The omics data are being generated in our laboratory, complemented by public datasets and analyzed individually. After that, these data are submitted for protein-protein and metabolite-protein interaction networks using Integrated Interactome System, a web-based platform recently developed in our laboratory that gather novel identified interactions, protein and metabolite expression/concentration levels, subcellular localization and computed topological metrics, GO biological processes and KEGG pathways enrichment. Subsequently, the information is extracted from these networks and the dynamic modeling is applied to each node of the network. In first case of study, we used a public transcriptome dataset from industrial yeast S. cerevisiae during industrial-scale bioethanol production to perform protein-protein interaction network and simulations using expression profile for each gene. Basically, the bioinformatics pipeline developed until this moment converts metabolites networks from KEGG database (KGML format) to models in Petri Net (SBML format) and estimates the transition probabilities using transcriptome data, which can be dynamically modeled by the user using the software Snoopy. The software can produce Stochastic Petri Net models from regulatory and signaling networks. The results of the simulations were validated using experimental data as deletion and high gene expression in glycolytic network that alter the production of ethanol.