THE SUCEST-FUN DATABASE FOR GENOME AND REGULATORY NETWORKS STUDY

Nishiyama-Jr,M.Y.[1,2], Ferreira, J.E.[1] , Souza, G.M.[2]
[1] IQ-USP – Instituto de Química – Universidade de São Paulo-Brazil
[2] IME-USP – Instituto de Matemática e Estatística – Universidade de Sao Paulo -Brazil

The study of regulatory networks brings important issues regarding the acquisition of biological information. Integrating gene expression data with metabolic pathways, physiological and technological information has been challenging but is necessary if one wants to discover novel gene functions and eventually manipulate pathways to increase yield and other traits of interest. Sugarcane cultivars are complex hybrids resulting from crosses among several Saccharum species. Conventional variety improvement may be limited by the narrow pool of suitable genes and a low efficiency in generating new resistant or useful varieties. Thus, molecular genetics is seen as a promising tool to assist in the process of developing improved varieties. The SUCEST-FUN Project Database (http://sucest-fun.org) is been developed to provide a uniform conceptual schema to minimize heterogeneous data representation, suitable semantics and query translation techniques. This approach requires the challenging task of harmonizing meaning around a common set of semantics, which would then enable semantic interoperability across many applications. This platform aims to associate function to sugarcane genes using a variety of tools, in particular those that enable the study of the sugarcane transcriptome. The transcriptome have been annotated and curated, based on NCBI, Swissprot, PFAM, KEGG and Gene Ontology, resulting in the protein kinase and phosphatase catalogues, the classification of Cell Wall related proteins and in the categorization of transcription factors for the grasses (GRASSIUS). To study of gene expression was customized three microarray platforms (SUCAST, SUCAMET, CaneRegnet) summarizing 16,786 genes, around 200 experiments and more than 70,000 differentially expressed genes, that has been integrated with the metabolic pathways and available physiological measures (19 experiments).Through the analysis of different cultivars, genes associated with sucrose content, yield, lignin and drought have been identified. Currently, tools are being developed to determine signaling and regulatory networks in grasses, sequence the sugarcane genome and identify sugarcane promoters.