Poster B-27

Development of Computational Tools for a Dynamic Association between Genomic and Proteomic Data and Metabolic Pathways

Authors:
Renato Milani (*Department of Biochemistry - Unicamp*)
Itaraju J. B. Brum (*Department of Biochemistry - Unicamp*)
Anderson R. Martins (*Alellyx Applied Genomics*)
Eduardo Galembeck (*Department of Biochemistry - Unicamp*)

Short Abstract: This project aimed to build tools for an easier association of proteomics and genomics information with metabolic pathways. GenPath and ECPath, unlike the currently available tools, offer the possibility to customize the data source analyzed, free from the pre-defined maps based on data present in metabolic pathways databases.

Long Abstract:
The large amount of information arising from genome sequencing projects has led to development of approaches highlighting the interaction networks between genes and proteins. This work presents a novel approach to viewing metabolic pathways in a visual and dynamic way, allowing not only to perform intra-specific analysis when a single organism data is available, but also to make multidimensional comparisons, using a user-defined pathway among different organisms or among different environmental situations to which the same species can be exposed. In this context, we aimed to build tools for an easier association of proteomics and genomics information with metabolic pathways. GenPath and ECPath, unlike the currently available tools, offer the possibility to customize the data source and the metabolic pathways to be analyzed, free from the pre-defined maps based on pre-defined data presented currently in metabolic pathways databases. GenPath is a Perl written tool that uses nucleotide sequence as main input, BLASTing the input sequences to compare and identify them. ECPath is written in PHP and accepts EC numbers as main input. Both have a Web interface for data submission. The data used to rebuild the metabolic maps were extracted from NCBI and KEGG databases and incorporated into the tools. Scripts were built to constantly update the databases used. In order to assess the reliability of the associations and reconstructions made by GenPath and ECPath, a standard-pathway based on glycolysis from four major metabolic pathways databanks (KEGG, ExPASy, MetaCyc and PUMA2) was stored in a MySQL database. Around 30 enzymes were selected to compose the bank because of their relation to a glycolysis map from any of the four databanks. SQL queries can provide the output needed for comparison with the tools output. The enzymes in this database are the expected ones to be pointed out by the tools when querying glycolysis or, conversely, glycolysis should be the output when these enzymes (EC numbers or gene sequences) are the input. GenPath proved to be a useful tool for gene annotation, classifying correctly submitted gene sequences. Results showed AGD2 and ALD1, Arabidopsis thaliana proteins not incorporated into the NCBI databank used in the tool at the time of the test, correctly to be transaminases. ECPath relates protein expression profiles to involved metabolic pathways. Test results showed 8 enzymes to be related to glycolysis in Xylella fastidiosa. Both tools are useful for researchers interested in studying

gene and protein function, performing comparative analysis involving large amounts of data and annotation of sequences with unknown function, all by associating enzymes and genes with metabolic pathways, and offering data outputs with a graphic representation of the metabolic maps. The described tools can be found at http://www.proteome.ibi.unicamp.br/tools/genpath/ and http://www.proteome.ibi.unicamp.br/tools/ECPath/.