Exploring Gene Ontology relationships in Enzyme Families: An application to Polysaccharide Lyases

The webtool GRYFUN (http://xldb.fc.ul.pt/biotools/gryfun/) allows visualization, filtering and subsequent analysis of Gene Ontology (GO) functional annotation profiles, that is, the collection of GO terms annotated to a given protein family (or set). GO comprehends three orthogonal ontologies – biological process, molecular function and cellular component. These ontologies are structured as Directed Acyclic Graphs (DAGs) and in GRYFUN's visualizations sub-graphs of them are displayed. These DAGs are comprised of nodes and edges and whereas each node represents a GO term, the connecting directed edges represent the type of relationship between terms. A protein can be annotated, for example, with the term 'carbon-oxygen lyase activity' which is connected through a directed edge from its parent term 'lyase activity'. Thus, that protein, also inherits this term and all the ancestors until the DAG's root. GRYFUN nodes are differentially colored to distinguish between direct and inherited term annotations. The edges' thickness is proportional to the number of proteins annotated with a given term. Thus, GRYFUN allows, at quick glance, to grasp both the functional broadness and depth of GO annotation within a given protein set (see Figure 1). Moreover, the tool enables the user to create subsets of proteins associated to terms by clicking on their respective nodes on the DAG. These subsets can then be further processed, for example, by semantic similarity computation, sequence alignment methods or they can be just exported to file for use in any other relevant application.

Polysaccharide Lyases (PLs) are among the enzyme classes described on the CAZy database (www.cazy.org), a structural-based family classification system. In CAZy, PLs are those enzymes that cleave uronic acid-containing polysaccharide chains via a ß-elimination according to general syn- and anti-elimination mechanisms, and generate saccharides containing unsaturated hexenuronic acid residues at the non-reducing end. With 22 families, PLs constitute a small family set, but has the widest experimental characterization coverage within CAZy. The usage of the GO nomenclature is not widespread in the Glycogenomics community. In an effort to resume family information, and ultimately to identify both the potential of the GO annotation and the possible shortcomings of the existing vocabulary a GO analysis of all PL families was performed with original family summaries being displayed with GRYFUN.