The Gene Ontology Newsletter

Issue No. 8 February 2008

New relationship types in Biological Process Ontology

On March 25, 2008, the Gene Ontology Consortium will introduce three new relationship types -- regulates, negatively_regulates and positively_regulates -- into the Biological Process ontology. Until now, regulatory processes have been represented as part_of the processes they regulate. These part_of relationships will be replaced with the new 'regulates' relationship type. We will also add positively_regulates and negatively_regulates relationships for appropriate child terms. The regulates relationships are transitive over both the is_a and part_of relationships.

Software developers should ensure that their procedures for loading the ontologies into their resources are compatible with these changes in advance of the release date.

A test OBO 1.2 file containing these new relationships is available:
ftp://ftp.geneontology.org/pub/go/scratch/regulates_relations_examples/go_regtest_withPosNeg_noPF_noXP.obo

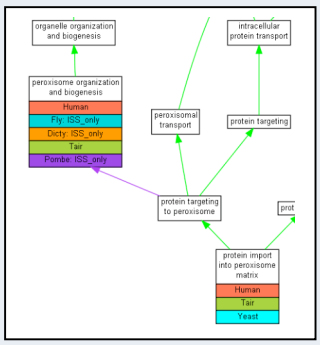
SGD gene_association file with IEA

On March 8, 2008, the *Saccharomyces* Genome Database (SGD) will amend the contents of its file that contains GO annotations to now include annotations generated using computational prediction methods. This is a major change, as the SGD file currently does not include annotations that are made using the IEA evidence code (Inferred from Electronic Annotation). These additional GO annotations include those computationally predicted by the Gene Ontology Annotation (GOA) project at the EBI, Hinxton, UK. Note that other gene association files already include IEA annotations.

The file available from the GO Consortium is called 'gene_association.sgd' and is available from the GO Consortium web page:
http://www.geneontology.org/GO.current.annotations.shtml

Genes of the quarter: Peroxins

The GO Consortium is working to annotate all model organism genes with homologs involved in human diseases. To this end, curators have recently revised and updated annotations of PEX genes in various model organisms.



From yeast to human, the biogenesis of peroxisomes requires a group of conserved "peroxin" protein factors. In humans, failure to properly develop or maintain peroxisomes leads to peroxisome biogenesis disorders (PBDs), a group of rare, genetically heterogeneous diseases characterized by severe mental retardation, renal, neuronal, and hepatic abnormalities, and death in early infancy. The genetic defects underlying PBDs all affect the import of peroxisomal proteins. The study of pex mutants and peroxisome biogenesis in model organisms has enhanced understanding of how the human orthologs function.

GO annotations have been made for 11 human PEX genes and their orthologs in 8 organisms (*M. musculus, R. norvegicus, D. melanogaster, D. discoideum, A. thaliana, C. elegans, S. pombe, and S. cerevisiae*). Annotations and the full version of the PEX10 graphic (above), as well as those for other genes, can be accessed at the GO website.

Graphic: http://www.geneontology.org/images/RefGenomeGraphs/

Annotations: http://www.geneontology.org/GO.current.annotations.shtml

New subcodes for ISS evidence code

The GO Consortium will add three new subcodes for the ISS (Inferred from Sequence or Structural Similarity) evidence code as of April 1, 2008 in order to clarify the type of sequence-based methods used as evidence in making annotations: Inferred from Sequence Alignment (ISA), Inferred from Sequence Orthology (ISO), and Inferred from Sequence Model (ISM).

ISA should be used when an annotation is based on pairwise or multiple alignments of the query protein with experimentally characterized proteins. Examples of tools that produce these types of alignments are BLAST, MUSCLE, and ClustalW.

ISO should be used when a protein is determined to be orthologous with an experimentally characterized protein from another species. Orthologous genes share a common ancestor and have arisen due to a speciation event. Orthologs are determined from phylogenetic analysis using algorithms such as maximum likelihood or nearest neighbor joining.

ISM indicates use of a statistical modeling tool to determine a protein's membership in a particular functional family, or to predict the presence of a particular sequence domain or structure. Examples of ISM evidence types are Hidden Markov Models (HMMs), tRNAscan, and transmembrane-HMM (TMHMM).

Full documentation on the new codes will be available as of April 1, 2008 at:
http://www.geneontology.org/GO.evidence.shtml

Upcoming Events

9-11 Apr 2008: Eukaryotic Genome Annotation & Analysis Course
J. Craig Venter Institute, Rockville, Maryland
http://www.jcvi.org/cms/education/professional-development/

6-10 May 2008: The Biology of Genomes
Cold Spring Harbor, New York
http://meetings.cshl.edu/meetings/genome08.shtml

13-14 May 2008: IGS Annotation Engine Workshop
Institute for Genome Sciences, Baltimore, Maryland
http://ae.igs.umaryland.edu/workshop_info.html

14-16 Jul 2008: Plant-Associated Microbe Gene Ontology (PAMGO) Training Workshop
Virginia Bioinformatics Institute, Blacksburg, Virginia
http://www.cpe.vt.edu/vbi-genome/

16-18 Jul 2008: Oomycete Bioinformatics Workshop
Virginia Bioinformatics Institute, Blacksburg, Virginia
http://www.cpe.vt.edu/vbi-genome/