

GO meeting - Tucson, Arizona - Feb. 2-3, 2002

Introductions.

Progress reports.

GO central (Midori Harris).

- Adding terms, reorganizing ontology.
- Add Jane Lomax as editor.
- Gene product search and destroy, GO-slim etc still awaiting action.
- MOBY effort (Mark Wilkinson's brainchild). Provide registry for sequence retrieval, annotation of where genes are go terms from different sources. Biomoby.org site for more info. Get MOBY white paper from Midori.

Flybase (Michael Ashburner).

- Becky Foulger lot of clean up.
- Better representation of data. Release 3 of Dm sequence.
- Reannotation of sequence will be available soon.
- MA and SL have had a collaboration with Celera Proteomics to compare their assignment of GO terms to Drosophila proteins (made with their PANTHER) system to those made by FlyBase. This is now being prepared for publication.

SGD (Karen Christie and Mike Cherry).

- EC definitions downloaded EC datafile function ontology file to clean up. All EC definitions MUST be checked before loading to be sure they don't overwrite.
- Total number of 1500 GO terms that get a definition as a result of this EC to GO effort.
- Anonymous CVS server is running.
- Website reorganized since Chicago meeting. Check updates on people page of GO web site.

MGI. (Harold Drabkin).

- Two full-time curators.
- 23% increase in annotated genes since October. 45% increase in swissprot assignments.
- Increase in hand annotation-triage process picks papers that are GO-friendly.
- Backpopulating entries through interactions with rat database.
- Developing new web interface.
- Since GO is now in the database searches will be much better. In addition tracking of literature will be improved.

TAIR (Lukas Mueller).

- Manual go curation is being done by all curators.
- Annotations to 10k genes.
- Linked to metacyc. Aracyc database established. Used metacyc to GO. Generated 1612 annotations.

- Secondary metabolism not yet well represented in GO. TAIR will add many new terms in this area.
- TAIR has developed anatomy (1000) and development (120 terms) ontologies.
- Created tool called pubsearch. Links literature, gene info and go annotation. Contains abstracts and full text files.

WormBase. Paul Sternberg.

- Focused on large RNAi screens (1100 genes assigned to about 2000 go terms). Should produce a large number of new curated terms.
- Life stages DAG for embryo and post embryonic stages and adults currently circulating through the community .
- Anatomy DAG (Raymond Lee). Cell lineage relationship included in DAG.

DictyBase (Rex Chisholm, Tricia Dyck).

- 7200 cDNAs
- chromosome 2 complete.
- Currently have about 2000 genes with annotations.

PSU (Matt Berriman)

- Malaria. Genome shrunk. 412 genes annotated.
- Tryps bruc. Chr 1. 400 genes
- Leishmania. Interested in using GO
- Entamoeba. Matt will present.
- Life cycle using cycle function of DAG-Edit.

GRAMENE. (Pankaj Jaiswal)

- 8000 gene product in SP-tREMBL about 50% have GO associations.
- Rice should be available for next meeting.
- 300 plant related terms.
- Manual curation of 4000 proteins in next few years.

GKB (Beth Nickerson).

- Waiting to hire.

EBI (Evelyn Camon)

- GO release in works
- Keyword to GO mapping improved.
- Single GO curator
- Receptor database at EBI
- Interaction database (INTACT)

TIGR (Michelle Gwinn)

- 10000 microbial genes annotated to go
- Arabidopsis

Compugene (Han Xie)

- annotating protein new version in about 2 months. Based on 1M proteins, 90% annotated in some way.
- Used for oligodesign, and in commercial database
- Algorithm development with GO clustering to generate primitive ontology from literature;

AstraZeneca (Courtland Yockney)

- GO has an official home within AZ. Gene association files reconciled with annotation of internal databases.
- Feedback—good: used for microarray data analysis. Increasing the number of things that can be prioritized for understanding arrays.
- Need immunology, physiology and hematology areas added. GPCR has added value. Use function ontology as an organizing principle.

Incyte. (Lisa Matthews).

- Incorporation of GO for YPD (6200 genes—46000 go terms)
- Mycopath (1800 genes with 7900 GO terms).
- Improving tracking.

October Action Item updates (see list of action items at end).

1. mouse
- 2.
- 3.
4. no progress
5. on agenda for later
6. standardization complete
7. added ND evidence code
8. add date field completed
9. taxon ID done
10. change submission requirement
11. update data directory—done
12. action items for John R.
13. DAG edit user guide. In process
14. transition to DB in progress
15. continued need for DBA
16. UK mirror in the works.

Software and Database

- DAG-Edit
 - Moved to sourceforge site.
 - DBxrefs now more automated
 - Type filtering
 - Find improved
 - Plugins launched in background

- Reduced size flat file format
- Dangling parent references now working—can link out to other references
- Cycles are supported
- Macros can be saved
- History plugin restored
- Term change tracker plugin added
- Can associate pictures with terms. Plugin.
- Future directions. Database beta test. Postgres problems.
 - Gene products viewer
 - Create servlets
 - Move DAML/OIL compatability
 - Interactive database mode
 - Need to track obsolete terms somehow.
- Database beta trial to continue
 - Will continue until a week has passed without problems. First priority to get save to work. Then get load to work. Then all editing will be on database.
- SourceForge repository reviewed. <http://sourceforge.net/projects/geneontology>
 - Bugs should be submitted via tracker at sourceforge
 - Requests for term changes needs to be made through sourceforege
- Fasta/sequence status. Most sequences are available.
- Database schema. Changed to match DAG-edit. Currently running two systems, MYSql and Postgres
- AmiGO update
 - Added peptide sequences derived from sp.
 - Icons that go back to sources added
 - Active link to ISS to show sequences.
- FAQ-O-matic.
 - Chris will implement a basic faq and develop a system for allowing it to be updated by consortium members.

Content.

- Presence of non-coding RNAs in GO. But need ways of representing genes of various classes. MA proposed an ontology for sequence features--SO (sequence ontology).
- Removal of gene products will continue with MH, MA and RLC. They will be replaced with more appropriate descriptors.
- Remove cyclin as we have an appropriate replacement, but assure that synonyms are present to allow searches.
- GO-slim still on the agenda for future.

GOBO.

- MA proposed adding GOBO to the GO repository. The group approved.
- Ontologies must be orthogonal

- Ontologies owned by developers will be responsible for maintaining them.

Cross product Generation.

- Discussed the production of cross products for anatomy agreed to wait until next time to discuss once tools are available.

Annotation issues.

- Have been using broad evidence codes. Discussed the need to expand evidence codes to more precisely capture the evidence supporting use of GO terms.
 - Use of more detailed codes would be optional
 - JR suggested an ontology of evidence codes. The group agreed.
 - Important to be sure that ISS codes trace back to a high quality annotation, not just IEA.
 - Agreed to make finer grained evidence codes available.
 - Discussed what codes we should use. (see list—get electronic version from Midori). Agreed that each must have definitions to be agreed upon in the period before the next meeting.
 - Where possible suffixes should be used consistently.
- David Hill raised the issue of cardinality induced by mutant phenotype evidence code. Midori suggested simply making a note that the field may be used this way in this case.
- Matt Berriman asked what citation should be used when there is no database. The suggestion was to point to a URL.
- Discussed situations where two terms from separate ontologies are often linked. Considered the possibility of developing tools to provide curators with options. This broadened to a discussion of tools for annotation including Talisman, Pubsearch from TAIR and sequence based tools. Sharing of these tools via GMOD was encouraged.
- Use of NOT was discussed. This is to be used in cases where something explicitly says something is not true.

Content issues:

- Discussed pathogenesis. How should we handle cases when this is a function of a pathogen. Agreed to add pathogenesis as a term with synonym of virulence.

Ontology Structure and Representation.

- The need to begin to develop more varied relationship types was discussed. It was agreed that this was needed, but could be delayed. MA presented a list of some possible relationship types. He agreed to share his list with the group for consideration and possible expansion.

Next meeting. 12-14 May at CSH?? Hinxton in the fall coordinate with ontology meeting (mid November). [Later changed to September, immediately after Genome Informatics]

Action Items.

- Submit electronic annotation methods and tools to Suzi et al.
- Suzi will generate a report on progress since the last meeting. To speed up organism reports.
- JR action item: spell checking method
- JR action item: Other people submit rules and use to check ontology
- Brad: grey out obsolete terms in AmiGO
- GO requests for terms through sourceforge.
- Add links to GO web site to submit or track requests.
- Brad: Amigo. ISS add links out.
- Individual groups should look at GO.xrf_abbs to check that URLs are correct
- Brad: Add documentation for format of GO.xrf_abbs.
- Chris will set up GO FAQ.
- We will establish a process for updating the FAQ using a distributed.
- MA will send out SO (sequence ontology)
- JR will develop plug in for producing cross-products
- Send GO slims that have used and published on to repository.
- MH and MA will circulate expanded evidence code vocabulary with definitions for review.
- Suzi et al will develop tools for annotation. Talisman tool.
- Pubsearch from TAIR should reside on GMOD.
- Beth will pursue May meeting at CSH, with Michelle Gwinn from TIGR as a backup.

Action Items from October 2001 Meeting (this is just the list, no details or results)

1. Get temporal and anatomical CVs into CVS from participating databases
2. Tool to create cross-product terms
3. Post new biological process that incorporates updating developmental processes
4. Come up with initial default GO slim
5. Ontological formalisms (for Feb meeting)
6. Standardization of database abbreviations
7. New evidence codes
8. Add date field to the entire gene association file
9. Update Documentation to explain new use of the TaxonID column
10. Change requirement for submission of sequence information for gene products
11. Update directory structure on GO web site
12. Action items for John Richter
13. User's guide for DAG-Edit
14. Transition into using database as primary repository
15. Continue to consider the need for a DBA
16. UK mirror