GO Consortium Meeting – Chicago, IL - October 15-16, 2004
[Next Meeting: Pasadena, CA – WormBase organizing – May 2005]

Group Participant List

SGD (Rama Balakrishnan, Mike Cherry, Maria Costanzo, Mayank Thanawala)
TAIR (Suparna Mundodi)
MGI (Judy Blake, Harold Drabkin, David Hill, Mary Dolan)
ZFIN (Doug Howe)
RGD (Victoria Petri, Mary Shimoyama, Simon Twigger)
dictyBase (Rex Chisholm, Petra Fey, Pascale Gaudet, Warren Kibbe, Karen Pilcher, Sohel Merchant)
EBI-Ontology Group (Midori Harris, Jane Lomax, Jen Clark, Amelia Ireland)
GOA (Daniel Barrell, Emily Dimmer)
Wormbase (Ranjana Kishore)
Incyte (Renee White)
Gramene (not present)
IRIS (not present)
DB Group (Chris Mungall)
TIGR (Michelle Gwinn-Giglio)
FlyBase (Michael Ashburner, Rebecca Foulger)
S. pombe/Sanger (Arnaud Kerhornou)
Pathogen/Sanger (Arnaud Kerhornou)
SO (Karen Eilbeck)
Reactome (not present)
TGD (Mike Cherry, Nick Stover, Mayank Thanawala)

A. Report from GO External Advisory Board Meeting (Judy Blake)

The advisory board made three key points:

1) We should put a heavy emphasis on production and annotation. If we move to a more formal ontological structure, it must not affect the usability of GO for annotation. We need to better support users and annotation efforts. We should consider separating GO production and management from research and development.

Several who were at the meeting felt that the reasons why we should move to a more formal ontological structure might not have been conveyed adequately to the board.

2) Metrics: we need to track our progress and usage in concrete ways in order to justify our funding.

3) Usability: we need to support naïve users, whose first or only exposure to GO is often on the gene page of a MOD, as well as first-time annotators who need to know how to get started and where to find support.

Discussion:

- it's critical for us to set goals and objectives for the coming year, in order to prepare for submission of our next grant in ~1 year
- we've had conflicting requests and suggestions from computer scientists vs. biologists, and we've perhaps put disproportionate energy into interactions with the computer scientist community. We need to think about our core community, remembering that our funding is from NIH, which is looking for practical applications that facilitate biomedical research.
- it might be helpful to seek separate funding for production and for research
- in addition to our users who want to annotate genes, we also have users who want to see large numbers of annotated genes. We need to think about how the most complete annotation of genes could be achieved. As more genomes are sequenced, people will be wanting to use GO even when they don't have a MOD. NHGRI is funding genome sequences without funding annotation. We need to think about how we could capture the limited annotation that may come out of those efforts, while ensuring that new annotation is of good quality. Could a grant support GO annotation in a non-species-specific way, for example by funding a GO outreach person to assist annotators?

B. Annotation Reports and Issues

1. Reports from Consortium Groups

> GO Editorial Office
> SGD
> MGI
> FlyBase
> TAIR
> GOA
> RGD
> dictyBase
> WormBase
> TIGR
> Incyte

(Any discussions that occurred after the reports are noted here but the reports themselves are not recapitulated.)

Some new members of the group were introduced in this section.

> **Action item:** Jen will update the "GO People" section of the GO website.

GO Editorial Office:
In response to a question about how they prioritize addressing SourceForge items, Midori

said that the highest priority items are:
- small things that can be done quickly
- issues that are of interest to many groups
- adding terms about previously unrepresented areas of biology, which may have repercussions throughout the ontology

The comment was made that the format of emails about SourceForge items is not easy to read.

Action item: Midori will look into whether the format of SourceForge emails can be customized.

Judy requested that those participating in email discussions that are not entered into SourceForge or forwarded to the GO list, should periodically write a summary of the discussion and and send it to the group, so that all are aware of the discussion.

Jen has documented the development interest group discussion and (since the Consortium meeting) has written a template for documentation of interest group discussions which she will make available on the web.

TAIR:

TAIR allows user submission of GO annotation; this led to a discussion of whether we can provide a generic spreadsheet for GO annotation by inexperienced users. How can we encourage the scientific community to do GO annotations? TAIR has some GO annotators who have volunteered because they want to explore careers in bioinformatics. When community members submit individual annotations, their names are displayed on TAIR web pages; this may be an incentive.

Action item: Jen will get the formatted Excel spreadsheet for user submission of GO annotations from TAIR, will modify it to be applicable to any organism, and will put it on the website to help new annotators.

dictyBase:

Pascale has constructed a "cross-product" ontology Dicty development ontology (presented at GO Users meeting) that contains terms created by combining GO process terms with terms from a Dicty anatomy ontology. The presentation file will be available for reference at the GO website, and Jen will provide a link to it from the development interest group documentation.

Judy made the point that it might be useful to display to users any previous GO annotations (deleted or obsoleted) for a gene product.

Incyte:

Renee described a quality control method that involves assembling a list of terms that are frequently used together ("GO pairs"; e.g., 'protein kinase activity' and 'protein amino acid phosphorylation'), and checking whether all gene products annotated to one are also annotated to the other. Judy asked whether this or any other quality control method could be shared with the public efforts. Mike asked whether Incyte's Pfam-to-GO mappings are different from the public ones; Renee answered that they may be, since Incyte started with Interpro and have done hand curation of the mappings.

> **Action item:** Renee will look into whether Incyte's quality control methods or Pfam-to-GO mappings may be shared with the public GO efforts. If permission is obtained, Jen will put them on the GO website.

2. Report from GO Annotation Camp (Mike Cherry)

Mike went through each point of the report; lack of comments indicates assent by the group. Discussions about specific items are noted here. Numbering of items corresponds to numbering in the Annotation Camp report.

1) Curation examples. Submission of annotation examples is a requirement for GOC members. Collection of annotation examples is now in the SourceForge Annotation Issues tracker (item #1047963), or examples may be sent to Midori.

> **Action item:** Each database must submit a set of 10 papers and accompanying GO annotations (see SourceForge item #1047963 for details).

2) README file. All gene association files must be accompanied by a README file summarizing the current annotation strategy: how genes are prioritized for GO annotation; whether multiple annotations to the same term, derived from different papers, are included; and any other annotation methods that may differ between MODs.

> **Action item:** Each database must submit a README file describing annotation strategy to accompany its gene association file.

11) Component terms and IEP. The annotation campers had decided that component annotations should never be supported by IEP evidence. Midori asked whether 'never' was too strong a qualifier. It was agreed that there may be legitimate exceptions to this rule.

On a related topic, Pascale asked whether localizations inferred by localizing a GFP fusion protein should have the 'colocalizes with' qualifier. The consensus was that in general they shouldn't, since those experiments are designed to indicate localization of the wild-type protein. However, there may be other evidence that affects confidence in the results (e.g., if a fusion protein is localized to the lysosome/vacuole), and this is an area where curator judgement must be exercised on a case-by-case basis.

14) Choosing the appropriate level for GO annotation. Delete 'using IGI' from last sentence; it's not relevant to the example.

17) Points to remember for suggesting new terms. Judy pointed out that use of gene product names in terms is still under discussion. Item 2 should be changed to 'Avoid using gene product names in new term names'.

Jane suggested adding the point that an informative name should be used for the SourceForge entry.

Item 5: Suparna asked whether companion terms should really be added if not needed immediately for annotation. Midori said this is subject to curator judgement, and Mike pointed out that it's helpful to have a discussion of companion terms in SourceForge for future reference even if they aren't immediately incorporated into the ontology.

This led to a discussion of the issue of bare leaf nodes: how many GO terms are not used to annotate genes?

Chris looked up the current statistics:
18,000 terms exist, including obsoletes
8,700 have a gene(s) attached
10,000 have genes attached OR have child terms with genes attached

20) Policy on curation of every paper available. The annotation campers had decided that it was ideal to annotate using each available paper: the number of independent annotations can provide a measure of confidence in the assignment. Michael felt that this redundancy could interfere with computation, particularly if it's appled inconsistently; it needs to be documented in the gene association README file. Several ideas were proposed for filtering annotations in order to create a subset to use in data analysis. A file could be created containing a reference subset of annotations, or particular annotations for each gene could be tagged in the gene association file. SGD currently provides a file containing a single GO term of each aspect for each gene, but the consensus was that a representative GO annotation set for higher organisms would need more than a single annotation per gene. Chris suggested that a tool could provide users with custom-generated gene association files, restricted by number of annotations per gene, type of evidence code, etc.

22) What to put into the DB_Object_Type column? It should be added to this section that MGD and ZFin use allele identifiers for IMP annotations.

24) Expanding GO evidence codes. Michael's evidence code hierarchy was discussed. The consensus was that this hierarchy should be available on the OBO site.

      Action item: Make the evidence code hierarchy available at the OBO site.

Summary discussion:

Michael pointed out that parts of the Annotation Camp report should be incorporated into the online annotation documentation.

Action item: Incorporate all relevant parts of the Annotation Camp report into online GO annotation documentation.

The question was raised as to whether there should be future annotation camps. There was enthusiasm for continuing them, although perhaps making them only 2-3 days long rather than a week. This would help ensure annotation consistency between MODs. They could be attached to another meeting such as the BioCurator meeting. MGI already does GO workshops, aimed at new users, at larger meetings. We should perhaps design a short tutorial/workshop to present to new GO annotators.

3. GO Annotation Topics

a) Annotation consistency between groups

Future Annotation Camps should address this.

b) Curated Reference sets for data analysis groups

This was discussed previously for item 20 of the Annotation Camp report, 'Policy on curation of every paper available'. We returned to the issue. There are many different ways to do this, ranging from merely stripping out multiple identical annotations, to providing users with a selected annotation set. It was felt that it's potentially dangerous to create a selected set of annotations, and we explored possible ways to do it. Michael suggested that if a gene is annotated to both a term and to its parent term, we could remove the annotation to the parent. However, annotations to parent terms may have better evidence than annotations to child terms, and taking away annotations to parents may remove knowledge: e.g., a protein found in both nucleus and nucleolus would lose its nuclear localization term. Multiple functions or localizations of a single gene product could also be lost.

Judy suggested that we table this discussion for next meeting while individual groups think further about the issue; providing a tool for users to generate custom subsets could obviate the need for us to make a decision on this.

c) NAS vs. TAS

There are differences between databases on the use of NAS. For a statement in a paper describing a direct experiment and referring to "data not shown", Flybase would typically assign NAS evidence, while MGI would assign IDA. Suparna explained that at

TAIR, if the paper explicitly states that 'Data not shown', NAS evidence would be used. But if a paper is describing a direct experiment and based on the direct experiment, the author is making a statement on gene function or process, they do TAS annotation and use an evidence description to support the type of evidence that it is based on.

There was a discussion of whether curators should interpret data shown in figures. The general consensus was that it's acceptable to curate something that's obvious in a figure although not explicitly stated, but it's not acceptable to look at data and come to different conclusions than the authors did. However , TAIR curators have occasionally not curated a conclusion(or have tried to contact an author about it) if it doesn't appear to be supported by the data.

d) IDs permissible with IPI

IDs used with IPI should be protein identifiers - Swiss-Prot, Trembl, RefSeq – and MOD identifiers should not be used if they refer to a gene rather than a protein. The ID used should lead to a FASTA file of protein sequence. For IGI evidence, gene Ids should be used. This will be difficult for UniProt annotators, but it should also be very rare.

C. Ontology Content Reports and Issues

1. GO Editorial Office Report (Midori Harris)

See above (B1, Reports from Consortium Groups)

2. Reports from GO Content Meeting at Carnegie Sept 19-20, 2004

a) PAMGO (Michelle Gwinn-Giglio)

Problems with current pathogenesis terms:
- pathogenesis is a 'victim-centric' term – expresses the point of view of the host
- many terms missing
- need to capture symbiotic interactions
- not always obvious whether organism is pathogenic, sometimes a relationship can be either pathogenic or symbiotic

The PAMGO proposal would add a general list of host/hostee interactions, and would remove 'pathogenesis'. There was general acceptance of the proposal, but some strong opinions that pathogenesis terms should be retained and symbiosis terms added. After more research, Michelle came to the conclusion that symbiosis is any kind of relationship in which two organisms live intimately together, and pathogenesis is an instance of this.

There are currently three trees that represent these various points of view:

1. contains only general terms (PAMGO)
2. very complex, with separate general, symbiosis, and pathogenesis subtrees
3. symbiosis tree with pathogenesis subtree

The PAMGO group favors #1; Michelle favors #3, because it allows for better representation of biofilms and symbiotic relationships not involving hosts.

The defense response part of this area is still under discussion; there is even less consensus between plant and animal researchers.

David commented that in constructing development trees, issues like this were resolved by adding general terms that both camps agreed on, and then adding child terms acceptable to each group.
The general consensus at this meeting is that #3 is best and represents a large improvement over current terms; it can always be refined in the future. The final plan will be circulated and will be implemented unless there's a serious problem.

      Action item: Finalize the new symbiosis/pathogenesis terms and incorporate them into the process ontology.

b) *Metabolism (Jane Lomax)*

Currently, all metabolism terms have 'physiological process' parentage but not 'cellular physiological process' parentage; this is a problem for annotation of gene products of unicellular organisms. For example, carbohydrate catabolism occurs, with different mechanisms, at both the cellular and organismal levels, but the GO term 'carbohydrate catabolism' does not have 'cellular physiological process' parentage. The proposed solution to this problem is to create new children of metabolism: organismal metabolism, cellular metabolism, primary metabolism. Some transport terms will be included, where it is an integral part of metabolism – for example, in plants some substances have to be transported between tissues for certain types of metabolism.

Consensus: this is straightforward and should be implemented.

      Action item: Finalize the new metabolism terms and incorporate them into the process ontology.

c) *'Regulates' Relationship Type (Midori Harris)*

Many would like to add a relationship type for 'regulates'. This would have consequences

for tools and displays, and would need detailed analysis of the scope of the project.

The consensus was that this should be looked into. Chris said that converting AmiGO and DAG-Edit to deal with this relationship type would be straightforward. We should let GO users know about this soon so they have time to change their tools.

      Action item: Look into the ramifications of adding the new relationship type 'regulates'. Midori will announce this upcoming change to the GO-friends mailing list.

d) Cell cycle (Amelia Ireland)

Currently, there are separate terms for each phase of mitosis with children describing parts of each phase. This was based on the S. cerevisiae cell cycle and is problematic for organisms with different cell cycles.

The proposed solution includes redefining cell cycle phases as processes, and adding more specific terms for cell cycle events.

Consensus: the high-level terms are fine, and the child terms may need work.

      Action item: Proceed to revise the cell cycle node along the lines already established.

3. Lessons learned from content meetings (Midori Harris)

- choose topics carefully: topics should be important to many groups and not easily resolvable by email
- have GO curators acting as liaisons to outside experts
- distribute materials in advance so people are prepared

4. Other Content Proposals for Discussion

a) removal of terms

This discussion centered on the tension between "ontological purity" and "scruffy necessity". Some members of the group feel that too many terms are obsoleted, too quickly; this creates a lot of re-curation work and may adversely affect the way users look at our project. The mandate of the GO editorial group is ontological purity, but perhaps this needs to be re-evaluated. If a widely used terms describing well-known gene products (e.g., cytochrome P450) disappear, that does not serve our user community.

Everyone agreed that increased use of synonyms could help this situation. Term names

can be made more precise, while commonly used, imprecise terms could be synonyms. Synonyms should be used liberally, and we should improve our use and display of them in various tools.

Some were concerned that 'vague' terms that represent extremely important concepts for biologists (transcription factor activity, chaperone activity, G-protein coupled receptor) have been obsoleted or may be slated for obsoletion. The argument was made that these terms do have clear meanings for scientists, and they represent special cases that need to be preserved. Retaining them may even lead to blurring of the line between function and process, but we should permit this for these special cases.

Chris proposed a less severe form of obsoletion, where a term is deprecated but not immediately removed. David suggested that if if all annotations to a term are automatically transferrable to its new equivalent, then the obsoletion should not have happened in the first place; the definition of the original term should have been improved, or the term should have been merged with another.

We discussed the specific case of 'chaperone activity', which has been obsoleted. Rama explained that the word 'chaperone' is used to mean three separate activities: transporting something; unfolded protein binding; and unfolded protein binding and re-folding activity. David suggested creating a lexical grouping term to be the parent of all three of these activities. Amelia thought that this would not make sense and would be analogous to creating a term 'factor activity' to group all 'factors'. Judy suggested that we could include lexical grouping terms in the GO, but tag them in some way to mark them as 'impure'. However, some thought this solution was simplistic, and others thought there would be no point in tagging a term if people would go on using it as before. Amelia suggested that we could create special terms crossing the process/function line, e.g., transcription factor could be a child of process: transcription and function: DNA binding activity. Chris observed that tagging would address the problem of vague terms, while cross-parentage addresses the problem of precise, complex terms.

Summary: We need to use synonyms more aggressively and liberally. We can't achieve purity, so we need to explore options for alternative solutions when they become necessary. We need to write up examples for ways to deal with exceptional terms. There was no consensus on whether lexical grouping terms or cross-aspect terms are a good idea. The group recognizes that there are exceptions that need special attention rather than immediate obsoletion. The definition may need reworking, or we may need to implement special solutions.

b)chemoattractant activity

Rex questioned the obsoletion of 'chemoattractant activity' because he felt that this was a definable function and furthermore, that these molecules have no other function. Amelia said that the major problem with this term was with its definition, since "attracting motile cells" is not a function. Rex argued that the function is more than simple receptor

binding, and Harold observed that while measuring chemoattractant activity requires observing a process, it's still a function.

The consensus was that we will reinstate the term, with a better definition. Rex proposed the definitions of chemoattractant/ chemorepellant activity: "Provides a signal to induce positive/negative directional cell movement". The definition and placement in the ontology of the analogous term 'pheromone activity' may provide an example.

   Action item: Reinstate terms 'chemoattractant/chemorepellant activity'.

c) ABC transporters

Michelle explained that the old term names were ambiguous and implied particular gene products, which didn't work for bacteria, where several functions reside in separate gene products. The parent term, 'ATPase activity, coupled to transmembrane movement of substances', accurately described the molecular function of all its child terms.

One issue with this obsoletion was the sheer number of annotations involved. This has already led to a new procedure for alerting people to obsoletions.

Harold was concerned that the obsoletion will lead to loss of information about ATP binding, and that this connection (ATPases bind ATP) should be intrinsic to the ontology rather than accomplished by concurrent annotation. The problem with this is that we would need to create substrate binding terms for all enzyme-substrate pairs, and we have already made an explicit decision not to do this.

After much discussion, a consensus emerged that this relationship shouldn't be built into the ontology but should be in a curator check: curators should consider annotating with both terms. There is already a note associated with 'ATPase activity, coupled to transmembrane movement of substances' that says curators should consider also annotating to 'ATP binding'. The narrower-than synonym 'ATP-binding cassette transporter' should also be added to 'ATPase activity, coupled to transmembrane movement of substances'.

Judy observed that this is a good example of an issue that needs to be resolved face-to-face.

   Action item: Add 'ATP-binding cassette transporter' as a narrower-than synonym of 'ATPase activity, coupled to transmembrane movement of substances'.

d) Definition of molecular function

Harold pointed out that currently molecular function is defined as an 'elemental activity'. The definition needs to be broadened to include complex functions. There was consensus

that this should be done, and Jane agreed to do it.

Action item: Broaden definition of molecular function to include complex functions.

e) RCA evidence code

A new evidence code, RCA (reviewed computational analysis) was proposed to refer to computational analyses that are reviewed and published, and that don't rely on sequence comparison. Examples of this type of study are PMID:14566057 and PMID:12826619. The major reasons that a new evidence code is needed are:

1) The confidence level differs between a "typical" TAS (eg. statement in a review, where the review cites other research papers showing evidence for the annotation; generally high confidence) vs. using TAS for computational analyses (generally lower confidence then a statement in a review or introduction of a paper).

2) Computational biologists who use GO annotations in their methods would often like to eliminate annotations based on other computational methods, to reduce the circular argument/proliferation of errors problem. They would not be able to do this if we use TAS.

3) IEA would not be appropriate for these papers because IEA implies the absence of curator input. Thus, IEA, we think, is generally a different type of evidence (and generally lower confidence) than RCA.

The consensus was that we will add this evidence code.

Action item: Add and document new evidence code, RCA (reviewed computational analysis).

D. GO Database and Tools Report (Mike Cherry)

1. **GO Database**

Soon, all gene association files will be analyzed to remove errors. This filtering process will take the input file and create a new file in which the following corrections have been made:

- remove IEAs older than a year
- remove annotations to obsolete terms
- make sure GOIDs are valid
- change secondary IDs to primary

- standardize headers

The processed files would be used to generate AmiGO. If this processing were done today, only SGD's and ZFin's files would not have changes; other groups would have ~10-150 changes per file.

We discussed the issue of whether annotations with IEA evidence, older than one year, should be removed. The consensus was that they should. If groups believe that their IEA annotations are still current, they can review them and update the date yearly so they will be retained.

Chris asked whether users might want to see gene products annotated to obsolete terms – perhaps these should not be removed. There were differing opinions on whether annotations to obsoletes should be retained. On the one hand, they are more informative than the absence of annotation, at least to users who understand what obsoletion means. On the other hand, removing obsoletes would drive annotation, forcing re-annotation of obsoletes to be the highest priority. The consensus was that the process should remove annotations to obsolete terms than are older than a certain (undecided) age. Renee suggested that when a term is obsoleted, it could be replaced with a parent term as a temporary placeholder until manual re-curation can be done.

Mike will give each DB a report of the errors in their files, and after some period of time, we will enforce the filtering.

How would we want to handle one-time efforts? They will go out of date because they don't keep up with ontology changes; however, the checking script could eliminate obsolete terms. These files could be kept in a separate directory, but there was concern that users might not find them. There was no consensus on how to handle this issue.

    Action item: Mike Cherry to give each database a list of the errors in its gene association file.

    Action item: Jen will ask Mike for information about the new checking script and will document the annotation checks on the GO website.

    2. **AmiGO**

Mayank has been installing AmiGO at Stanford; AmiGO and GOST are running; will start to switch over from Berkeley to Stanford soon. Mike will send test URLs to Consortium members; when all are satisfied, AmiGO users will be redirected to Stanford. The first step will be to simply replicate what's done at Berkeley; changes will be implemented later.

    Action item: Continue with and finish the installation of AmiGO at Stanford.

3. DAG-Edit

Midori has found some minor bugs in the DAG-Edit 1.419 beta version and has emailed John. The final version should be available soon.

4. Web pages

Jen has re-done the tools page. The Advisory Board has suggested that the first page should be simpler; Jen and others will look into that.

Action item: Jen will design a simple front page for the GO website that is friendlier for biologists and other newcomers. It should include links to explanatory pages for new users and new annotators. She will also check that the links to SourceForge are working correctly.

5. User Stats

AmiGO web pages are being hit 18,000 times per week. Usage has increased.

Action item: Jen will fix the AmiGO search box on the front page of the GO website so that either terms or annotations can be searched. Mike Cherry will send her information on how to do this.

6. Documentation

No report.

E. GO User Support (Michael Ashburner)

1. Legacy annotation sets and what to do with them

See section D1 (GO Database) above.

2. Outreach to new groups

Several ideas were proposed:

- we could have GOC staff member(s) dedicated to assisting annotation efforts. Center for Bio-Ontologies?
- we could try to work with sequencing centers (JGI, Broad Institute)
- we could try to convince program officers of funding agencies of the need to fund functional annotation along with genome sequencing
- we could use large meetings such as PAG, ISMB, ASM as opportunities to hold workshops and inform people about GO
- we could advertise the GO Users meeting as a place to learn about GO as well as to present its uses. Part of the meeting could be devoted to a GO tutorial.
- we should seek out genome databases with which we are not in contact: Bombyx, Xenopus, Maize

Jen will be presenting a GO annotation tutorial at the upcoming PAG (Plant and Animal Genome) conference (January 2005). We should make sure that as many databases as possible know about this.

Action item: Jen will try to make contact with as many genome databases as possible to make sure they're aware of the tutorial at the PAG meeting.

3. **Requests to join GO Consortium, GO Associates idea**

So far, we have accepted into the Consortium groups that work with us on ontology development and return to us a gene association file. But an increasing number of groups would like to join, and we don't want the GOC meetings to grow to an unworkable size. We could limit the number of people attending from each group, but no one really liked this idea. CGD, HGNC, MetaCyc, GermOnline, and a toxicogenomics database have asked to join the GOC; except for CGD and HGNC, these are not organism-specific genome databases.

The consensus was that we should establish the status of GO Associate. Associates will contribute annotation files and participate in a GO meeting (equivalent to the current users meeting, broadened to include an educational component); we may invite specific associates to a GOC meeting as they become educated and actively involved in the project.

Action item: Document the status of GO Associate and invite interested groups to join.

F. SO/GO Development Reports

1. **OBOL (Chris Mungall)**

Chris has found 248 missing relationships. They are listed at http://www.fruitfly.org/~cjm/obol. The editorial office reviews them and adds some but not all. OBOL can be used behind the scenes to suggest new relationships or to create new definitions.

Action item: We will proceed with using OBOL to make computed definitions of cell differentiation and maintain them in a cell type ontology.

Discussion:
GO contains implicit orthogonal ontologies, e.g., a chemical ontology. How do we decide which cross-product terms should exist in GO vs. in separate cross-product ontology? Michael suggested that if a term is needed for annotation, then it should be instantiated in GO. But there is concern about 'bloat' making it difficult for annotators to find terms. David suggested creating a separate namespace for extremely specific terms, e.g., mouse development terms. Another suggestion would be to incorporate all the separate ontologies into GO but provide tools for users to filter out terms not relevant to them.

2. **SO Content Meeting reports plus SO Development report (Karen Eilbeck)**

SO Content Meetings were held at Berkeley Aug 22-23, 2004 and at Hinxton, Sept 22, 2004. The structure of SO was changed drastically after the last meeting. The revised ontology, 'so-meeting.obo', is available for comments. MODs are starting to use SO. SOFA content is frozen for 1 year, until next May.

3. **OBO (Michael Ashburner)**

We are getting many requests to add ontologies to OBO, however we can't add contradictory ontologies, and the quality of some ontologies may be variable. Michael proposes splitting the OBO site into subdirectories. The OBO core directory would contain ontologies being worked on by GOC members or which are needed by GOC members for making cross-products with GO; entries must be approved by the GOC. Another directory would contain all other ontologies.

Action item: Distribute the ontologies at the OBO site into two subdirectories, one containing GO Consortium-approved ontologies in active use by consortium members, and the other for any other ontologies.

G. Plan Meetings and Assess Past Meetings

The question was raised as to whether we really need to read database reports at the GOC meeting. At this meeting, this consumed a morning and there was relatively little discussion. The consensus was that we will not have each group talk about each report in future meeting (although reports must still be provided). Any special issues or new developments requiring discussion should be submitted as agenda items.

WormBase (Caltech) offered to host the next meeting. May and September were discussed as possible times, but it was decided that a year from now would be too long an interval, so May was agreed upon. The suggestion was made to have back-to-back GO Users and GOC meetings, 1 1/2 days each.

H. Summary of Action Items from this meeting

1. **Action item:** Jen will update the "GO People" section of the GO website.

2. **Action item:** Midori will look into whether the format of SourceForge emails can be customized.

3. **Action item:** Jen will get the formatted Excel spreadsheet for user submission of GO annotations from TAIR, will modify it to be applicable to any organism, and will put it on the website to help new annotators.

4. **Action item:** Renee will look into whether Incyte's quality control methods or Pfam-to-GO mappings may be shared with the public GO efforts. If permission is obtained, Jen will put them on the GO website.

5. **Action item:** Each database must submit a set of 10 papers and accompanying GO annotations (see SourceForge item #1047963 for details).

6. **Action item:** Each database must submit a README file describing annotation strategy to accompany its gene association file.

7. **Action item:** Michael Ashburner will make the evidence code hierarchy available at the OBO site.

8. **Action item:** The Editorial Office will incorporate all relevant parts of the Annotation Camp report into online GO annotation documentation.

9. **Action item:** The Editorial Office will finalize the new symbiosis/pathogenesis terms and incorporate them into the process ontology.

10. **Action item:** Jane will finalize the new metabolism terms and incorporate them into the process ontology.

11. **Action item:** Chris and the Editorial Office will look into the ramifications of

adding the new relationship type 'regulates'. Midori will announce this upcoming change to the GO-friends mailing list.

12. **Action item:** Amelia will proceed to revise the cell cycle node along the lines already established.

13. **Action item:** Amelia will reinstate terms 'chemoattractant/chemorepellant activity'.

14. **Action item:** Jane will add 'ATP-binding cassette transporter' as a narrower-than synonym of 'ATPase activity, coupled to transmembrane movement of substances'.

15. **Action item:** The Editorial Office will broaden the definition of molecular function to include complex functions.

16. **Action item:** The Editorial Office will add and document the new evidence code, RCA (reviewed computational analysis).

17. **Action item:** Mike Cherry will give each database a list of the errors in its gene association file.

18. **Action item:** Jen will ask Mike for information about the new checking script and will document the annotation checks on the GO website.

19. **Action item:** SGD personnel will continue with and finish the installation of AmiGO at Stanford.

20. **Action item:** Jen will design a simple front page for the GO website that is friendlier for biologists and other newcomers. It should include links to explanatory pages for new users and new annotators. She will also check that the links to SourceForge are working correctly.

21. **Action item:** Jen will fix the AmiGO search box on the front page of the GO website so that either terms or annotations can be searched. Mike Cherry will send her information on how to do this.

22. **Action item:** Jen will try to make contact with as many genome databases as possible to make sure they're aware of the tutorial at the PAG meeting.

23. **Action item:** The Editorial Office will document the status of GO Associate and invite interested groups to join.

24. **Action item:** Chris will proceed with using OBOL to make computed definitions of cell differentiation and maintain them in a cell type ontology.

25. **Action item:** Amelia will distribute the ontologies at the OBO site into two subdirectories, one containing GO Consortium-approved ontologies in active use by consortium members, and the other for any other ontologies.

I. Review of Action Items from last meeting [Stanford]

1. Action item: Eurie and Michael will strive to provide a definition for 'transcription factor activity'. *A definition has been considered but not yet incorporated into the ontology*.

2. Action item: We will try to set up a pilot project that has a web page "indexing" key point discussions in the GO email archives. *Jen has worked on a prototype and hasn't gotten much feedback yet. Judy pointed out that it's important to record key discussions so we don't revisit the same issues over and over again. Jen will work out a way to do it, then people who are actively involved in each discussion could pick the key points to index*.

3. Action item: We will add a new qualifier for "Colocalizes with" that is appropriate for indicating that the gene product has been found in the vicinity of a structure. ***DONE***.

4. Action item: Jen will update the documentation for Component rules with discussion of this qualifier and its use. ***DONE***.

5. Action item: Brad and Mike will look into whether it is possible to keep a Google search of the email archive separate from the general Google search of the GO web pages. *This is not possible*.

6. Action item: groups to investigate if large files, compressing files, will pose any problems at their own site. *No problems identified*.

7. Action item: Mike Cherry to look into how best to interact with WormBookIII to embed GO terms in on-line version of the book. Need to consider 'glossary' approach and how to maintain currency. *Mike has talked to Lisa Gerard at Wormbase; this is in progress*.

8. Action item: 'sensu' terms will have a mixture of English phrase and Latin genera, along with the taxon ID. The definition of any sensu term would include the point that it is not totally restricted to a particular grouping. ***DONE***.

9. Action item: Proposal: by end of month will post OBO files, curator trials-- use OBO for two weeks to work out bugs, then general switch to OBO as master; More general document for users. Announcement on site. ***DONE***.

10. Action item: We will try to set up a pilot project that has a web page "indexing" key

point discussions in the GO email archives. *Duplicate of item #2 above*.

11. Action Item: "Not" column will be renamed "Qualifier". When it has any other value other than NOT or NULL, it should be used for annotations for components of a complex only. This will allow reason across membership in a component to infer function. Should be checks for complex entries. Subunit will have annotation to a particular subunit activity, if known, or to "contributes_to" and that gene product must also be annotated as a component of complex.
e.g., specific example eIF2; has three subunits (alpha, beta, gamma); one binds GTP; one binds RNA. But the whole complex binds the ribosome (needs all three); so all three get "contributes to" ribosome binding, and one gets GTP binding, the other gets RNA binding. AND all three are annotated to EIF2 complex. ***DONE***.

12. Action Item: In column #12 of the gene_association file, "complex" will be allowed as a type of "DB_OBJECT". ***DONE***. *Reactome may have used it; no one else has*.

13: Action Item: Concepts relating to the use of complex functions (e.g. receptor tyrosine kinase) will be added to the documentation. ***DONE***.

14. Action Item: Add Joel Richardson's tool to the tools page. Note language change from Python to Java. (Jen) ***DONE***.

15. Action Item: Document OBO flat file format advantages for annotators (There are none.) ***DONE***.

16. Action Item: Write documentation for the process and component ontologies along the same lines as the function documentation that had already been written. (Jen) *Process ontology documentation is in progress; component ontology documentation has not been started*.

17. Action Item: Add documentation to remind people that the definition is there to clarify the meaning of the term name if there is any ambiguity. This is to be added to the general documentation as well as to the documentation for each ontology. (Jen) ***DONE***.

18. Action Item: See if there is an easy way to add the date that a definition was made. *Not done. There was agreement that we should do this, but we need to work out the specifics (which date, how to display it). Midori will write some specifications and ask for comments, then will add it as a feature request to SourceForge*.

19. Action Item: General improvements to GO website. *Since the January 2004 Consortium meeting, the following additions and changes have been made to the website: new meetings page added; OBO file format documented; instructions on accessing cvs by ssh added; function ontology documentation added; obsoletion standard operating procedure added; information on the annotation checking script added; page added for acknowledgements of outside experts; evidence code summary table added; menu added to give easier access to SourceForge; documentation added about frequency of updates*

of GO downloads and mapping files; 'mailto:' links added to site for people in interest groups and databases; tools developer's page added; more detail added to annotation guide; style of website encoded in cascading style sheets and footer fixed.