

GO Consortium Meeting – Pasadena, CA – April 8-9, 2005

[Next Meeting: Berkeley organizing – March 25-29, 2006 (to be confirmed)]

Group Participant List:

SGD (Rama Balakrishnan, Mike Cherry, Karen Christie)
TAIR (Tanya Berardini, Sue Rhee)
MGI (Judy Blake, Alexander Diehl, Mary Dolan, Harold Drabkin, David Hill, Li Ni)
ZFIN (Doug Howe)
RGD (Simon Twigger, Jennifer Smith)
dictyBase (Rex Chisholm, Pascale Gaudet, Karen Pilcher)
GO Editorial Office (Midori Harris, Jane Lomax, Amelia Ireland, Jen Clark)
GOA (Evelyn Camon)
Wormbase (Igor Antoshechkin, Carol Bastiani, Wen Chen, Eimear Kenny, Ranjana Kishore, Raymond Lee, Hans-Michael Muller, Erich Schwarz, Paul Sternberg, Kimberly Van Auken)
BioBase (Incyte) (not present)
Gramene (not present)
IRIS (not present)
BDGP (Suzanna Lewis, John Day-Richter, Chris Mungall, Shenqiang Shu)
TIGR (Michelle Gwinn-Giglio)
FlyBase (Michael Ashburner, Russ Collins)
S. pombe/Sanger (Val Wood)
Pathogen/Sanger (not present)
SO (not present)
Reactome (not present)
TGD (Mike Cherry)

TABLE OF CONTENTS

<u>INTRODUCTION AND WELCOME</u>	2
<u>Report from GO External Advisory Board (Judy Blake)</u>	2
<u>ONTOLOGY ISSUES</u>	3
<u>Regulator vs. regulation</u>	3
<u>"Structural constituent" function terms</u>	4
<u>Behavioral Response terms</u>	5
<u>Issue about oocyte growth from the Development interest group</u>	6
<u>Chemoattractant/chemorepellant</u>	8
<u>ANNOTATION ISSUES</u>	9
<u>Redundancy in the annotations (Mike Cherry)</u>	9
<u>Online annotation form</u>	9
<u>Allow genes to be entered in the 'with' column with colocalizes with</u>	10
<u>Use of 'with' column for entering species</u>	10
<u>"Views" for particular organism sets</u>	11
<u>Pure hypothetical proteins</u>	11

<u>Inter-annotator consistency</u>	12
<u>Evaluation of electronic annotations</u>	12
<u>Obsolescence and moving annotations</u>	12
<u>Supporting new annotation groups</u>	13
<u>RESOURCE ISSUES</u>	14
<u>Advances in representations and use of cross-products (Chris Mungall)</u>	14
<u>OBO-Edit (John Day-Richter)</u>	15
<u>Advisory Board and other GO notes (Mike Cherry)</u>	15
<u>MOBY namespace issues (Chris Mungall)</u>	16
<u>BRAINSTORMING SESSION</u>	17
<u>GO Home Page</u>	17
<u>Home Page prototypes (Amelia Ireland)</u>	18
<u>AmiGO: searching and visualization</u>	19
<u>GO survey</u>	20
<u>NEXT MEETINGS</u>	20
<u>Annotation Camp</u>	20
<u>Ontology Development Meeting</u>	21
<u>Grant Meeting</u>	21
<u>GO Consortium Meeting</u>	21
<u>GO Users Meeting</u>	21
<u>ACTION ITEMS</u>	21
<u>Summary of Action items from last meeting, not yet done [Chicago]</u>	21
<u>Summary of Action items from this meeting [Pasadena]</u>	22
<u>Review of all Action items from last meeting [Chicago]</u>	23

INTRODUCTION AND WELCOME

Report from GO External Advisory Board (Judy Blake):

Topics that were discussed include annotation coverage and numbers of annotations from the different databases (David Hill). Annotation consistency was a concern and we need to explicitly describe how the Consortium addresses this issue (Judy Blake). Progress on ontology development, including the newly instituted 'ontology workshop' groups, was also discussed (Jane Lomax), and the development of new tools to deal with complexity and integrating other ontologies with GO was presented (Chris Mungall).

The progress and impact of GO were presented by Mike Cherry. The metrics included the number of hits to the geneontology.org pages and the number of publications found in PubMed by searching for "Gene Ontology." In 2004, there were 120 publications of this type, and so far in 2005 there are 90. These are impressive numbers compared to 0 in 1999, especially considering PubMed searches only title and abstract, indicating that GO was integral to the study.

The goals for the grant were also discussed (due February 1, 2006). Because funding from the NIH may not increase very much, yet we have ideas for expanding the scope of GO activities, the Advisory Board suggested looking for other possible sources for

funding. One possibility is to look at what agencies are funding the research publications that cite the Gene Ontology.

Overall, the response from the committee was very positive and they had many suggestions for what to do next. They encourage the Consortium to focus on the core mission and understanding the users and the community.

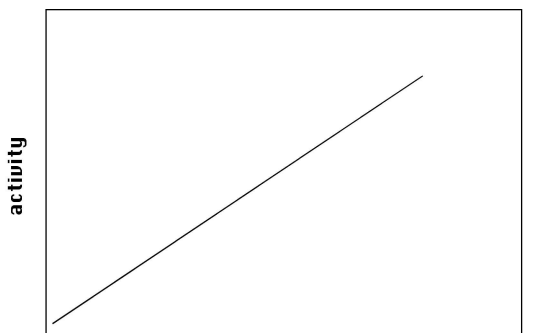
ONTOLOGY ISSUES

Regulator vs. regulation:

There is redundancy in regulation terms in the process and function ontologies. For example, we have 'regulation of enzyme activity' (process) and 'enzyme regulator activity' (function). Can we move all of this to function or process? There is a proposal to add the new relationship type 'regulates' that would potentially take care of this issue, allowing annotation to anonymous classes of regulation: inhibition, activation, down-regulation, up-regulation. (David Hill, Harold Drabkin)

Discussion:

- Is the ontology redundant in function/process? Should all regulation be process? Many of the regulation terms are not defined. Many regulation terms should probably not be function, they should be process. A difficult example is the regulatory subunit of a kinase. Should this be annotated to 'enzyme regulator activity' and also contributes_to 'kinase activity'? You sometimes need to annotate to a function and a process. It is not possible to transfer everything to either function or process. For example, a gene product can be involved in transport (process) but does not have transporter activity (function).
- There is a question of what exactly is a function? And what exactly is a process? In the GO documentation, a process is defined as a collection of functions, although a process could potentially be only one function. There are different views on what exactly constitutes a function. One view is that to have an activity must have a direct inhibitor, and can be thought of as a dose-dependent. Example:



Concentration of gene_product having the activity

- The opposite view is that everything has a function, even a brick. A brick has no inhibitor; you cannot plot concentration of brick vs. what a brick does.
- Making the decision to annotate to function or process is often based on how much information is available.
- Another issue is regulators and regulated gene products. Gene products that act as regulators are also regulated. The `is_a` and the `part_of` relationships don't really capture the interactions here. One proposal to help clarify this is with a new relationship type 'regulates.' Regulation terms would have different relationship types in different nodes, for example:

```
-'regulation'
--is_a 'regulation of cell differentiation'
-'cell differentiation'
--regulates 'regulation of cell differentiation'
```

- Overall, people are in favor of the new relationship type 'regulates.' One criticism is that 'is_a' and 'part_of' are very concrete whereas 'regulates' is vague. Also, there were concerns that the ontology would look different if the "regulates" information is found in the relationship rather than in the term name. This objection is unfounded: the ontology and terms will look exactly the same; only the relationship types will change. Another problem is the degrees and kinds of regulation. Some children of positive regulation terms include activation and maintenance, and children of negative regulation are down-regulation and inhibition.

Action item: Look at how the relationship type 'regulates' and positive/negative regulation will affect the ontologies and the annotation files. [Chris Mungall]

Action item: Look at current annotations to 'enzyme regulation' type terms to see what has been used. [Jane Lomax]

Action item: New relationship type 'regulates'. [John Day-Richter]

"Structural constituent" function terms:

[See SourceForge item 1113374]

Some children of 'structural molecule activity' are functions but include cellular component information. These terms include 'extracellular matrix structural constituent,' 'extracellular matrix constituent conferring elasticity,' 'extracellular matrix constituent,' 'lubricant activity', etc.

Discussion:

- Non-catalytic functions belong in the function ontology. These 'structural constituent' functions are such terms. After examination, people agreed that they all looked like legitimate functions. The problem seems to be the difficulty of

describing the activity that we are trying to represent, and nobody was able to suggest better phrasing.

- 'Structural molecule activity' (GO:005198) is a legitimate function, although it could be reworded to just 'structural molecule' as it is not exactly an activity. The child terms are problematic, however, because they incorporate cellular component or anatomical terms. A few of the existing terms could be renamed to remove the reference to the extracellular matrix, thereby generating legitimate children of 'structural molecule':

extracellular matrix constituent conferring elasticity ; GO:0030023
change to: structural molecule constituent conferring elasticity

extracellular matrix constituent, lubricant activity ; GO:0030197
change to: structural molecule constituent, lubricant activity

matrix constituent conferring compression resistance ; GO:0030021
change to: structural molecule constituent conferring compression resistance

matrix constituent conferring compression tensile strength ; GO:0030020
change to: structural molecule constituent conferring tensile strength

- The decision was to postpone this until we have the tools to deal with these terms that will allow us to do the decomposition and annotations.

Behavioral Response terms:

There is currently no link between the behavior terms and the 'response to' terms, which means there is quite a lot of redundancy. Do we want to keep these terms? We've talked before about how these terms shouldn't be applied to higher animals. For lower animals, should they perhaps be combined with 'response to' terms?

Discussion:

- Some people felt that behavior is a type of response. After discussion it was agreed that they are different things. For example, 'cell motility' is a behavior, and not a response. 'Chemotaxis' is a response. Some responses are not behaviors, such as 'response to pathogen,' 'inflammatory response,' 'response to UV,' etc.
- Definitions: 'Behavior' is: "The specific actions or reactions of an organism in response to external or internal stimuli. Patterned activity of a whole organism in a manner dependent upon some combination of that organism's internal state and external conditions." 'Response to stimulus' is: "A change in state or activity of a cell or organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of the perception of a stimulus." The distinction is whether the whole organism or just certain cells of the organism are responding. Also, some behaviors are independent of stimulus, whereas all responses are dependent on stimulus.

Action item: Document the fact that we will systematically put both parents on behavioral response terms where there are terms in both nodes. The terms currently in the ontology have correct parentage. [Jane Lomax]

Issue about oocyte growth from the Development interest group:

[See SourceForge item 1170007]

Chris found a missing relationship using OBOL and wanted to know why we rejected it.

[GO:0001555] oocyte growth

is_a

[GO:0007281] germ cell development

We would like to decide whether it is best to have a process term 'development' or to have a collector term 'developmental process' instead. The issue is a philosophical/semantic/logical one.

Currently the term is called 'development' but the relationships to child terms are such as would fit the term 'developmental process'. This sets up 'development' as a collector term rather than a process term in spite of its name. The definition of development states that it is:

Definition: Biological processes specifically aimed at the progression of an organism over time from an initial condition (e.g. a zygote, or a young adult) to a later condition (e.g. a multicellular animal or an aged adult).

Comments: Note that this term was initially 'developmental process' and was renamed.

The concept of 'developmental process' was spotted by OBOL as inconsistent since at this high level 'growth' is an is_a child of 'development', whilst the lower level terms 'x growth' terms are part_of 'x development'. For example 'oocyte growth' would be considered to be part_of 'germ cell development'. Therefore in creating a standard dag structure for the development node it is impossible to state a consistent relationship between 'x growth' and 'x development' without considering the high level terms to be simple collector terms that are exempt from the rules.

Likewise, is the actin cytoskeleton a type of cytoskeleton, or is it a part of THE cytoskeleton?

So the question here is:

1) Should the term 'development' be changed to 'developmental process' and documented as a collector term that is exempt from the rules of the standard dag structure applying to all the more specific terms?

or

2) Should the term 'development' be redefined so that it is truly a process term and then all its child term changed to be part_of, in keeping with the lower level terms?

Discussion:

- From the ontology point of view, it is problematic to have growth R' development and oocyte growth R germ cell development, where R is not equal to R'.
- 'Oocyte' may have a different role relative to 'growth' than 'germ cell' does to 'development,' or development may mean a different thing in both contexts (developmental process vs. development as a whole). If this is the case it should be clearly specified in the definition (and preferably reflected in the name).
- An oocyte is_a germ cell (according to CL) and growth is_a development (according to GO). Is it because oocyte growth doesn't actually refer to developmental growth, but rather to an increase in diameter that is not necessarily related to growth in the developmental sense?

Could we have two kinds of growth?

1. 'Developmental growth.'
2. 'Growth' that occurs when development is not taking place.

- When we get to the higher nodes, the distinction between is_a and part_of gets a bit fuzzy and is dependent on how we interpret the definition of something like development. If we interpret it as the **collection** of processes, then growth is an is_a because it is a developmental process. If we consider development the entirety of processes, then growth would be a part_of because it is a part of the entirety of development.
- This cannot be the case in GO because if 'developmental process' is a collector term then its is_a children must also be collector terms. An analogy is a stamp collection. A stamp can be part_of a stamp collection, but it is not an is_a child to a stamp collection since only another stamp collection can be a type of stamp collection. A stamp cannot be a type of stamp collection.
- Conclusion: The term 'development' cannot be a collector term, and that it must be a process term.
- Next discussion point: Since development is certainly now a process term then if 'growth' was kept as a child of 'development' then it must have the part_of relationship.
- This would be a problem for other terms since 'development' of all individuals of all species is not really a single process. For example, 'plant development' is not really a part of the grand overarching process of development of all species. 'Plant development' is a type of development.
- Perhaps the development term could be considered as 'organism development' so that the graph would appear as follows:

```
[i]organism development
---[i]plant development
---[p]growth
---[p]organ development
```

- Dictyostelium undergoes development as an aggregation of cells rather than as a single organism so 'organism development' would not work. How about 'entity development' rather than 'organism development' since that would accommodate Dictyostelium?
- An alternative suggestion is to move growth outside development, as a direct child of 'biological_process.' This would also fix the problem that some growth events (like certain instances of cell growth), are not developmental events. Developmental growth events would have both 'development' and 'growth' as parents.

Current structure:

```
[i]development
---[i] growth
-----[i]oocyte growth
---[i]oocyte development
-----[p]oocyte growth
```

With new structure:

```
[i]growth
---[i]cell growth
-----[i]oocyte growth
[i]development
---[i]oocyte development
-----[p]oocyte growth
```

- Relationships are now okay with OBOL.

Action item: Move growth outside development, as a direct child of 'biological_process.'

Chemoattractant/chemorepellant:

[See SourceForge item 1052249]

At the last meeting, it was decided that 'chemoattractant activity' should be restored. However, there were still objections as to whether this was a legitimate function term. The problem is that chemo[attract/repell]ant definition would involve binding to a specific receptor, setting off a signaling cascade that induces positive or negative chemotaxis. This definition invokes the function of 'receptor binding' and the process of 'induction of positive/negative chemotaxis'. Chemo[attract/repell]ants can be defined without referencing chemotaxis, which is a process.

A way to represent these molecules is with new receptor binding terms - 'chemo[attract/repell]ant receptor binding' - in combination with the existing process terms 'induction of [positive/negative] chemotaxis'. Both sets of terms can have chemo[attract/repell]ant as narrower-than synonyms.

Discussion:

- The argument against this was that there is no way to clearly annotate a chemoattractant that would unequivocally distinguish the chemoattractant itself from a protein inside the cell that would bind the chemoattractant receptor and trigger signaling events.
- Resolution: Functions will be conceptually divided into two classes, those that involve reactions and those that do not. A number of other terms that were obsoleted based on the fact that they were functions for which activities could not be assigned will be restored.

Action item: A list of function terms to be restored that includes 'chemoattractant' and 'chemorepellent' activities will be circulated. [Amelia Ireland]

ANNOTATION ISSUES

Redundancy in the annotations (Mike Cherry):

To address the presence of redundancy in the annotations, we will be filtering to remove redundant annotations. Everyone will still submit association files, but the GO database will now include a directory that contains filtered files.

Some of the requirements for filtering:

- Every row in the file must be correct. If it's not, it will not be included. You will be notified if it's incorrect.
- If the GO ID is a secondary ID, it will be replaced with a primary ID.
- If the GO term is obsolete, it will be removed.
- If an IEA is more than 1 year old, it will be removed.

In addition, each species/taxon ID will have only one "authority" MOD, and all other annotations for that species will be filtered. If another source wants their annotations to appear in this filtered file, they will coordinate with the MOD to have their annotations included in the set submitted from the MOD. This will not be implemented for a month or two.

Action item: The information regarding the requirements for implementation of the filtering of annotations will be sent to the GOC members. [Mike Cherry]

Online annotation form:

Should we allow bench scientists to submit annotations? Many researchers have annotations that would be useful to have. If we allow submissions from non-Consortium/non-database groups, how should we do it? What tools should we use? Do the annotations need to be reviewed by curators?

Discussion:

- We should definitely allow researchers to submit their annotations. These submissions could be sent to GOA, who will send them to an appropriate MOD. The database can submit that data in their association file, citing the source of the annotation.
- If there are groups, such as the chicken group (Fiona McCarthy), that come to us for mentoring. We need to talk further about training of other groups.
- Many tools for annotation exist; only TAIR and TIGR have tools for user submission.

Action item: Suzi wants to hear from everyone about what tools they have for annotation. Send to Shu at Berkeley. [Everyone]

Action item: Ask GMOD if there are tools available for user submission of annotations.

Allow genes to be entered in the 'with' column with colocalizes with:

There is a proposal to use the 'with' field to describe the qualifier when using 'colocalizes_with.' An example in which MGI would like to use it is when a paper localizes a gene product to the lysosome by colocalization with Lamp1, a marker for lysosomes. (David Hill)

GeneX | GO: lysosome | IDA | colocalizes_with | Lamp1

Discussion:

- Should we use this column in a new way? The 'with' column refers to the evidence code, not to the qualifier. Perhaps we are muddying the use of this column by using it in all different ways. The use of the colocalizes_with qualifier is already confusing for annotators, and this is adding another layer of complexity.
- Additional information is not necessarily being added if the db object in the 'with' column is already annotated.
- Part of the rationale for doing this is that we are not 100% confident that the gene product is actually in the lysosome. However, this is the way that many highly studied proteins were originally localized, through a colocalization experiment.
- The idea of adding a new column to the association file was suggested. See also discussion on species in the 'with' column.

Action item: Add a SourceForge item for the issue of using the 'with' column when using colocalizes_with. [David Hill]

Use of 'with' column for entering species:

TAIR would also like to use the 'with' column in a new way, adding taxon IDs of bacteria for terms like 'response to bacteria' and 'response to fungi.' The use case we have in mind

is capturing which organism was used in an experiment where plants are subjected to bacterial attack and then respond. The pathogen discussion group is quite enthusiastic about this.

Discussion:

- If this is implemented, you can no longer do ISS with another database object in the 'with' column. Entering gene products and tax IDs are two totally different concepts.
- Another important question is whether this is in the scope of GO.
- MGI has a detailed notes field separate from GO from which they can also retrieve data and where they keep this type of information.
- Using the 'with' column is not necessarily the best way to do it. Perhaps this should be a new column in the annotation file if people want to capture taxon IDs and use the 'with' column to describe the qualifier.

Action item: Write up a proposal on how to capture information such as taxon IDs (describing the GO term) and proteins that colocalize_ with other proteins (describing the qualifier). [John Day-Richter, Chris Mungall]

"Views" for particular organism sets:

The prokaryote and plant groups are using a subset of GO terms that are specific to their organisms. The resolution is to just leave it; it is not being maintained and will not be updated unless the file is required for specific purposes.

Discussion:

- We should not encourage people to use only a subset of GO. Some groups do in fact filter out portions of the ontologies to which they cannot annotate (for example, removing 'chloroplast' for animal annotation). However, this is for annotation purposes and is not intended for viewing the ontologies.
- The groups that are using these "views" are using them because they have resource issues and a tight time constraint. People argue that 7,000 terms is not that much better than 18,000 terms.

Pure hypothetical proteins:

What should annotators do about purely hypothetical proteins with no sequence similarity to anything and have no possible GO terms? The former guidelines were that these genes would not get annotated at all; should we continue doing that, or annotate them to "unknown", with the RCA evidence code? (Michelle Gwinn-Giglio)

The consensus is that this is not an appropriate use of RCA. Function/Process/Component "unknown" should be used here with the ND evidence code. GO is not for determining whether a gene is real or not. Moreover, there have been significant improvements in the quality of the genome assemblies since the GO project started.

Inter-annotator consistency:

We need to ensure consistency between annotators and document the way each database addresses annotation consistency. This is important for the grant renewal. (Mike Cherry)

Discussion:

- Comparison of curators (MGI vs. UniProt mouse annotation) showed that there are many discrepancies between curators (approximately 50% of the time), however, all curators are over 90% accurate. This means that the vast majority of annotations are correct but they are also incomplete.
- At SGD, curators pair up monthly to review papers and discuss the possible GO annotations. Measures like this and attendance at the GO annotation camp help ensure consistency. However, because of the difference in resources (i.e., number of curators), each database needs to set their own standards and document their protocols for annotation.

Action item: Add information about inter-annotator checks in the README files and in the progress reports. [Everyone]

Action item: Inter-annotator consistency will be discussed further at the GO annotation camp.

Evaluation of electronic annotations:

Several users have been interested in the reliability of electronic annotations, and there is only one publication, still in press, that addresses this question (Camon, E.B. et al. 2005. BMC Bioinformatics 6 (Suppl 1):S17). What has been done to assess electronic annotations? Can we say something on the website to answer this kind of question? (Jen Clark)

Discussion:

- There have been a few studies for different tools. TargetP is 80-90% correct in its predictions but does not work well for plants. EC2go and InterPro2GO are quite good and have 91-100% precision.
- This is something we should definitely look into and post on the GO website.

Action item: Write up a summary of the reliability of electronic annotations and reference the paper that talks about this. Add this to the FAQs. [Jen Clark]

Obsolescence and moving annotations:

At the last meeting, we agreed that terms do not need to be obsoleted if the definition change was meant only to improve the wording and did not change the way annotators use the term. There are cases, like the new definition of morphogenesis and development, where only a few annotations need to be changed. The question was whether it was okay to keep the term and ask the annotators to verify their annotations and change them where appropriate.

Discussion:

- The argument for doing this is that it is simpler.
- The group generally argued against that: if the definition changed sufficiently such that annotators need to verify or change their annotations, then the term needs a new ID. This is important because not only do the GO Consortium members need to be aware of this; every GO user needs to be alerted to the change, and the only way to do it is by obsoleting the term.
- The GO editorial office will go back to obsoleting or merging if the annotators need to check their annotations. They will only redefine without obsoleting if there is no possibility of a mistake. In the case of the morphogenesis terms, this means that 'x morphogenesis' terms defined as 'development' will need to be merged into the corresponding 'x development term' and a new 'x morphogenesis' term made.

Action item: Document what happens when a term/definition changes enough such that annotators need to modify annotations, using development/morphogenesis as an example. [Tanya Berardini, Jen Clark]

Supporting new annotation groups:

- Tutorials at meetings have so far been quite effective. The PAG meeting generated a lot of interest in GO. We should be on the lookout for opportunities to hold tutorials at specialty meetings, particularly in areas where the ontology needs to be developed.
- Direct mentoring is also very effective. Fiona McCarthy (chicken annotation) visited MGI for two weeks and generated a full set of electronic annotations. This tutorial turned into much more than just GO topics; much time went into general management of an MOD.
- The Annotation Camp last year was very successful, and registration for the second camp looks to be about 40+ annotators. These people will be split into smaller groups for discussion. The meeting is June 1-4 at Stanford.
- Genomes for infectious organisms are a big focus for the NIH right now. We should try to send a "trainer" to the NIAID Bioinformatics Resource Meeting.

Action item: Work with the BRC centers (NIAID) to provide any support they want. Try to tie in a tutorial with one of their group meetings. [Judy Blake]

RESOURCE ISSUES

Advances in representations and use of cross-products (Chris Mungall):

Currently there are many cross-products in the GO. Many of them we would like to keep but many are redundant. (For example, cysteine metabolism and sulfur amino acid biosynthesis. The reasons for this are historical; GO preceded the chemical ontology ChEBI.) To successfully create cross-products there must be consistency amongst the different ontologies. Chris has been looking at the cell ontology (CL) and the cross-products that can be made with GO. There are around 800 cell types in GO, where as the CL has 700 terms, 300 of which do not have matches in GO.

Differences between CL and GO:

- **Definitions:** Example: T cell vs. T lymphocyte. The GO definitions are not consistent with the CL develops_from relationship type.
- **Inconsistent structure:** Example: hemocyte, plasmocyte, lamellocyte.
- **Granularity:** The CL sometimes has less detail (e.g., retinal cone cell vs. photoreceptor cell). More commonly, the GO has less detail (e.g. neuroblast, cell proliferation). The question is, though, do we need to create more GO terms only if we annotate to them?
- **Naming style:** Differences in hyphenation (B cell), suffix (CL: neuron, GO: neuron cell).
- **Missing synonyms.**
- **Spelling:** Example: oesinophil.
- **Different relationships:** Some terms are part_of in one ontology and is_a in another ontology.

OBOL parses terms to find OBO terms embedded inside the GO terms and finds inconsistencies between GO and OBO. The new approach is to remove dependency on text analysis and augment the GO to integrate information from other ontologies. Chris showed a few examples of how the ontology file would look with the CL terms integrated. The relationship types need better names: currently they are intersection_of and has_output.

```
GO: 30183B-cell differentiation ; GO: 30183
```

```
intersection_of cell differentiation ; GO: 30154  
intersection_of B lymphocyte ; CL: 236
```

Why should we do this?

- Makes GO more computable: easier to maintain ontologies and consistency between ontologies.
- Integrates ontologies such that you can query across ontologies.
- Automates inconsistency detection: batch mode, dynamically from OBO-Edit.
- Explores new browsing and navigation paradigms. For example, when looking at the DAG for 'lymphocyte differentiation,' show 'cell differentiation' (GO) and 'lymphocyte' (CL) in adjacent windows.

- A prototype for browsing cross-products was generated using MGI's structured notes fields that show cell types, anatomical structures, cell lines in which the gene product was expressed, and developmental stages for the GO annotations.
- Filters dynamically: allows us to thin out complex parts of the ontology.

Proposed plan:

- Integrate GO with CL to begin with. The next logical ontology to integrate after CL is ChEBI. Eventually integrate anatomy ontologies but these will be the hardest.
- Create cross-product information using OBOL and integrate it with GO. Add the cross-product information to the GO terms but don't do cross-products for everything.
- Synchronize the ontologies.
- Iterate and evaluate.
- Look at cross-products of process and process.

OBO-Edit (John Day-Richter):

- OBO-Edit has a new plug-in that allows generation of cross-products. To create a cross-product, first load two or more ontologies and, using the plug-in, select a core term and drag and drop the terms to be crossed into the cross-product window. Next, drag and drop a relationship type into the <type> box, add a property (for example, 'has_output'), and create a rule for generating the cross product, (for example, (\$1 \$2 = term 1, space, term2). When the cross is executed, the new term(s) are created with new ID(s). The term name can later be edited, as well as any operation you would do on a 'normal' term. Clicking the "+" on the right side of the plug-in window allows you to add an unlimited number of terms.
- OBO-Edit has new editing modes: for example, merging terms, when you drag a term, a menu appears to let you chose what to do with the term. There are also new keystrokes for functions using letters (M = merge).
- OBO-Edit has a major improvement in speed; all due to displaying fewer paths in the DAG viewer.
- OBO-Edit has "instances": instantiated versions of the classes we have been creating. Allows instance browsers, for example different sequence.
- OBO1.2 file format changes the way synonyms are handled. One can specify synonym type. This new feature makes the OBO-Edit-generated files incompatible with DAG Edit. One can open OBO files with OBO-Edit, but should not commit them in cvs.

Advisory Board and other GO notes (Mike Cherry):

- Web usage for GO web site (filtered for robots, images): GOC home page had close to 1 million hits within the past year; GO database had over 4 million hits.
- More than 5000 different IP addresses.
- Most users from USA, Europe, then Asia.
- Publications: 0 papers in 1999; 169 papers mentioned "Gene Ontology" in 2004; 98 as of April 9, 2005. This is particularly impressive because PubMed searches only titles and abstracts, indicating that GO is an important aspect of the study.
- Citations: the two most cited GO Consortium papers have been cited over 1000 times, which is extremely high.
- Number of links reported by Google: >7200 pages link to geneontology.org.
- Links from NCBI = approx. 46,000; links from EBI = approx. 14,000.
- AmiGO will be run from Stanford as of May 1, 2005. The database is built weekly.
- Datafiles download site will change slightly (web address).

MOBY namespace issues (Chris Mungall):

To synchronize our data with MOBY, we must fix two things in our association file. First, we need to be consistent with the authority in column 1. Sometimes Flybase is Flybase, but sometimes it is FB. Similarly, SGD is sometimes SGD, at other times, SGDID. Second, we need to modify our global IDs to represent the namespace. For example, the NCBI IDs for gi and PubMed ID are indistinguishable. One way to fix this is with the use of LSIDs (Life Science Identifier, <http://lsid.sourceforge.net/>). With an LSID, this ID:

PMID: 12571434

becomes this LSID:

`urn:lsid:ncbi.nlm.nih.gov:pubmed:12571434`

where 'urn' stands for Uniform Resource Name.

Discussion:

- Using the LSID allows you to represent more information. It is also a way to help unify IDs globally, not just within GO groups. If the LSID is going to become a life sciences standard, perhaps we should move towards this ID. The LSIDs don't necessarily need to be shown on the web pages but can be used for storing IDs.
- One of the problems with the LSID is that you cannot automatically generate a URL from it; you need a resolver. Also, some long-standing IDs are not amenable to LSIDs, such as EC numbers.
- For the time being, it is sufficient that everyone is aware of this issue. MOBY is not yet ready for us.
- Whatever the solution, it needs to satisfy GO and the community of users, so we need to get a list of our requirements. Simon Twigger will be at the MOBY meeting in May, so he can talk to Mark Wilkinson about some of the ways to solve this ID problem.

Action item: Talk to Mark Wilkinson about LSIDs and come up with a proposal for how we might make changes to our IDs. [Chris Mungall]

BRAINSTORMING SESSION

GO Home Page:

This brainstorming session was to get ideas for improving the GO website.

- We should consult with biologists and web designers for help on this.
- Reactome has a very nice front page. Was designed by professionals.
- You still need to know who the users are.
- A survey of colleagues who are not involved with GO, see what they think.
- Less text, more graphics.
- Look at some of the other homepages and see what we like and don't like. Reactome, Google, etc. There are many examples of how **not** to do it.
- Send around your favorite 5 home pages. Doesn't need to be bioinformatics. Needs to have similar content types.
- Done this at TAIR. Worked with a professional, fed them ideas about needs for biologists, and they came up with 3 different options.
- <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html> talks about evaluating webpages.
- <http://webpagesthatsuck.com/>
- What would Amazon do? We're trying to sell our product. Amazon is successful and uses certain principles. We should try to use these principles.
- What is your community? Website should be for community. Market research: what is community and what are their needs? A lot of this will flow from that. Lots of documentation on the website but no one reads it.
- In paper, there are 10 points... if you are text miner, do this... etc.
- Look at what parts of the website are used most and focus on those.
- Do online survey of users.
- If somebody comes to a page, goes to another page, and jumps back, can we look at that?
- How are the web pages used?
- We need to target biologists. They have no idea what it is about. Need to encourage use by biologists.
- How can we better serve?
- Look at failed queries. Figure out what people are looking for but can't find.
- Of the sites linking to GO, can we look at where they are coming from and who is linking the most?
- Objections to the current page are that the front page is kind of random. They are important things but not well arranged.
- Things have changed over time. What we did before is not the same so we need to adapt.
- "The goal of gene ontology is to do controlled vocabulary"...so now what?

- Menus on left are very hard to navigate.
- Searching is hard for documentation...always get hits from email archive.
- The kinds of questions we ask are different from other people.
- We are more persistent. We know things exist and work harder to find stuff. Other people don't know it's there and will give up.
- We need to clarify who the documentation is intended for: new users, advanced users, etc. Not much distinction between annotation and editing the ontologies.
- We need a direct way to get to annotation parts and ontology development.

Home Page prototypes (Amelia Ireland):

1. **New index page:** mission statement needs to be in 10th grade style. Needs to be shorter and catchier.
 Search box: should be right at the top. Can search by gene/protein name or by gene/protein description.
 Should set up AmiGO so that you don't have to click any options for term/name.
 If you are biologist and type in your gene name, you won't get it.
 Section for popular links: this could come from usage stats. Links could change from time to time. Maybe "What would you like to do?" on home page.
 Sections: for annotators, for biologists, for programmers, etc.
 News bulletins: put news on separate page.
 Recent ontology changes: more news.
 Site info: stats for website, grant information.
 Left menu: open menus, home, downloads, GO tools, documentation, get involved, curator guides, about GO, terms of use, contact GO, site map. can be opened up.
 Annotator/curator: confusing to people. Ontology editing, annotations...what to use? What makes sense to biologists?
 Search boxes: are they ever used on pages other than the home page? Many say yes.
 Search boxes should stay on every page.
2. **Quickstart guide:** Don't want to read through whole manual, just go and do it.
 Four main categories. Get a gene with annotations. Find gene products associated with process, function, component. General introduction to GO.
 Three types of people: annotators, ontology editors, consumers. Need to aim at consumers.
 Simple start – search, short statement, and quickstart.
 Usage stats: basis for quickstart.
3. **Simple page:** quickstart, downloads, tools, get involved, contact. (People like this one.)
4. **Pictures:** small icons next to each broad topic. (People really like this one.)
5. **Separate by users:** New users + Advanced users. (People like the previous one better.)

6. **Separate by main topics:** tools, downloads, etc.
7. **I want to:** know more about what GO is... see the def of a GO term... others...

Discussion:

- Don't want to come to a page and see text. Too long.
- At ZFIN, we have a process where we sit down as a group. What does the webpage have to do? What is the goal? Do paper mock-ups of what it should look like. What can we get rid of and what do we need? Take it across the street to people in the labs and ask them to use it. Ask them: What do you expect when you click here? What is important to you? What is missing? Commonly: always have search box at top or at the right. Doug will send some criteria to the GO list.
- Need LESS text, even if it is high school level.
- Frequently cited paper: "a tool for the unification of biology" – possible tag line for webpage.
- What we need to focus on is the process. Who will be responsible? How will we get feedback? Who will take time to work on this and report back to everyone?
- The reason to have this discussion now is that all of you need to feel empowered. Everyone's ideas are important. We're not designing this for ourselves, it's for others. Who are we trying to reach?
- Can we send out a message to GO list and ask what do you need from the GO page? What would make it easier to use?
- We can't be limited by that. There are many more users than there are on the Email list.
- Maybe the survey would be good because it will reach more users.
- Have a link at the bottom of pages that says, "Did you find what you are looking for?"
- We should keep track of failed searches, and use that information to assess what users are looking for.
- Access to the database is limited. The static pages are only supporting material.

Action item: To help in the new design of the GO home page, send examples of web pages that we like, and pages that we do not like. [Everyone]

AmiGO: searching and visualization:

This brainstorming session was to get ideas for improving the AmiGO browser.

- Current behavior seems non-intuitive to some people: when you click on a term, a new page opens.
- Perhaps the search can retrieve a list of terms rather than a DAG.
- Synonyms are not displayed prominently enough. Many users search on a synonym and get to a page that has different information.
- The search should integrate annotations and ontologies better. The SGD search tool, for example, searches gene names, gene products, GO terms, etc.

- A new functionality would be the ability to search all ontologies in OBO. With the cross-products this will happen anyway. How will cross products affect the speed of AmiGO? We may need to think about new ways to display the data. Chris Mungall presented some ideas addressing that issue in his talk about cross-products.
- Everyone agreed that the ontologies must be updated more frequently. Currently updated every month; at Stanford it is currently updated every week. Is that enough? Should we have daily updates? Annotations can be updated less frequently.
- Should the different databases look more similar? One way to start would be to use the same ontology browser. What would it take for everyone to use the same browser? (People agree that ontology data need to be up-to-date.)
- AmiGO and GOC "look and feel" could be more similar; currently it's not clear they are the same thing or part of the same thing.
- Ideas of tools for annotators: term suggestion tool (for example, Incyte/Biobase GO pairs); the Wormbase tool OrthoGO lists annotations from InParanoid calculated orthologs; also, a way to transfer GO terms into GO annotation tool.
- GOST-type tool for ESTs; the problem is about the number of sequences a user would submit.
- Some databases have their own GO browsers; we need to know what improvements they have made over AmiGO.

GO survey:

To better understand who the GO users are, what information they are looking for when they come to the GO/AmiGO web pages, and how to better target the users we are trying to reach, it was suggested that we do surveys when we go to meetings. Suggested questions include:

- Have you heard of GO?
- What do you think it is?
- Why do you use GO?
- Has it worked for you?
- What do you do? (arrays, etc.)

There may be two surveys, one for beginners and one for more advanced users.

Action item: Write up a survey that we can take to meetings. Try to minimize effort by putting it online with easy analysis. Put the link in your MOD newsletter. [Mike Cherry, Everyone]

NEXT MEETINGS

Annotation Camp: June 1-4, 2005, Stanford.

Ontology Development Meeting: October/November, 2005, TIGR. Focus on immunology with transport side. We have \$15,000 support for this meeting from anonymous sources.

Grant Meeting: November (possibly November 17-18, 2005), Banbury? Washington, DC? Meeting of the PIs and others.

GO Consortium Meeting: Possibly March 25-29, 2006, St. Croix. Berkeley will organize. Also plan to do outreach in Puerto Rico.

GO Users Meeting: September 14-15, 2005. Integrated with MGED 8 meeting in Bergen, Norway.

ACTION ITEMS

Summary of Action items from last meeting, not yet done [Chicago]:

1. **Action item:** Jen will get the formatted Excel spreadsheet for user submission of GO annotations from TAIR and will put it on the website to help new annotators. *Jen has this from TAIR; not on website yet.*
2. **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. *Don't know; Incyte (now called BioBase) not present.*
3. **Action item:** Each database must submit a set of 10 papers and accompanying GO annotations (see SourceForge item #1047963 for details). *3 or 4 databases have done this.*
4. **Action item:** Each database must submit a README file describing annotation strategy to accompany its gene association file. *Some have done this; Mike will warn people.*
5. **Action item:** The Editorial Office will broaden the definition of molecular function to include complex functions. *Draft written by Midori; will include non-activity functions.*
6. **Action item:** Jen will ask Mike for information about the new checking script and will document the annotation checks on the GO website. *In progress?*
7. **Action item:** SGD personnel will continue with and finish the installation of AmiGO at Stanford. *Almost done.*

8. **Action item:** The Editorial Office will document the status of GO Associate and invite interested groups to join. *In progress; haven't decided on description of GO Associate.*
9. **Action item:** Chris will proceed with using OBOL to make computed definitions of cell differentiation and maintain them in a cell type ontology. *In progress.*

Summary of Action items from this meeting [Pasadena]:

1. **Action item:** Look at how the relationship type 'regulates' and positive/negative regulation will affect the ontologies and the annotation files. [Chris Mungall]
2. **Action item:** Look at current annotations to 'enzyme regulation' type terms to see what has been used. [Jane Lomax]
3. **Action item:** Add new relationship type 'regulates'. [John Day-Richter]
4. **Action item:** Document the fact that we will systematically put both parents on behavioral response terms where there are terms in both nodes. For example, behavioral response to ether. The terms currently in the ontology have correct parentage. [Jane Lomax]
5. **Action item:** Move growth outside development, as a direct child of 'biological_process.' [GO Office]
6. **Action item:** A list of function terms to be restored that includes 'chemoattractant' and 'chemorepellent' activities will be circulated. [Amelia Ireland]
7. **Action item:** The information regarding the requirements for implementation of the filtering of annotations will be sent to the GOC members. [Mike Cherry]
8. **Action item:** Suzi wants to hear from everyone what tools they have for annotation. Send to Shu at Berkeley. [Everyone]
9. **Action item:** Ask GMOD if there are tools available for user submission of annotations.
10. **Action item:** Add a SourceForge item for the issue of using the 'with' column when using colocalizes_with. [David Hill]
11. **Action item:** Write up a proposal on how to capture information such as taxon IDs (describing the GO term) and proteins that colocalize_with other proteins (describing the qualifier). [John Day-Richter, Chris Mungall]
12. **Action item:** Add information about inter-annotator consistency in the README files and in the progress reports. [Everyone]

13. **Action item:** Inter-annotator consistency will be discussed further at the GO annotation camp.
14. **Action item:** Write up a summary of the reliability of electronic annotations and reference the paper that talks about this. Add this to the FAQs. [Jen Clark]
15. **Action item:** Document what happens when a term/definition changes enough such that annotators need to modify annotations, using development/morphogenesis as an example. [Tanya Berardini, Jen Clark]
16. **Action item:** Work with the BRC centers (NIAID) to provide any support they want. Try to tie in a tutorial with one of their group meetings. [Judy Blake]
17. **Action item:** In the bibliography section of the GO web site, make an option to sort by year or by topic. [GO Office]
18. **Action item:** Talk to Mark Wilkinson about LSIDs and come up with a proposal for how we might make changes to our IDs. [Chris Mungall]
19. **Action item:** To help in the new design of the GO home page, send examples of web pages that we like, and pages that we do not like. [Everyone]
20. **Action item:** Write up a survey that we can take to meetings. Try to minimize effort by putting it online with easy analysis. Put the link in your MOD newsletter. [Mike Cherry, Everyone]
21. **Action item:** Investigate running GO Users Meetings as satellites of other conferences (not necessarily GO Consortium Meetings), both because of timing and to reach different/broader audiences. Contact the organizers of the MGED meeting (Bergen, Norway) to set up an adjacent Users Meeting. [Midori Harris]

Review of all Action items from last meeting [Chicago]:

1. **Action item:** Jen will update the “GO People” section of the GO website. ***DONE.***
2. **Action item:** Midori will look into whether the format of SourceForge emails can be customized. *Can't be customized. Unresolvable.*
3. **Action item:** Jen will get the formatted Excel spreadsheet for user submission of GO annotations from TAIR and will put it on the website to help new annotators. *Jen has this from TAIR; not on website yet.*
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. *Don't know; Incyte (now called BioBase) not present.*
5. **Action item:** Each database must submit a set of 10 papers and accompanying GO annotations (see SourceForge item #1047963 for details). *3 or 4 databases have done this.*
 6. **Action item:** Each database must submit a README file describing annotation strategy to accompany its gene association file. *Some have done this; Mike will warn people.*
 7. **Action item:** Michael Ashburner will make the evidence code hierarchy available at the OBO site. ***DONE.***
 8. **Action item:** The Editorial Office will incorporate all relevant parts of the Annotation Camp report into online GO annotation documentation. ***DONE.***
 9. **Action item:** The Editorial Office will finalize the new symbiosis/pathogenesis terms and incorporate them into the process ontology. ***DONE.***
 10. **Action item:** Jane will finalize the new metabolism terms and incorporate them into the process ontology. ***DONE.***
 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. ***DONE.***
 12. **Action item:** Amelia will proceed to revise the cell cycle node along the lines already established. *In progress.*
 13. **Action item:** Amelia will reinstate terms 'chemoattractant/chemorepellant activity'. *No one responded to SourceForge item; terms will be reinstated.*
 14. **Action item:** Jane will add 'ATP-binding cassette transporter' as a narrower-than synonym of 'ATPase activity, coupled to transmembrane movement of substances'. ***DONE.***
 15. **Action item:** The Editorial Office will broaden the definition of molecular function to include complex functions. *Draft written by Midori; will include non-activity functions.*
 16. **Action item:** The Editorial Office will add and document the new evidence code, RCA (reviewed computational analysis). ***DONE.***
 17. **Action item:** Mike Cherry will give each database a list of the errors in its gene association file. ***DONE.***
 18. **Action item:** Jen will ask Mike for information about the new checking script and

will document the annotation checks on the GO website. *In progress?*

19. **Action item:** SGD personnel will continue with and finish the installation of AmiGO at Stanford. *Almost done.*
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. *Amelia presented some prototypes; in progress.*
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. ***DONE.***
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. ***DONE.***
23. **Action item:** The Editorial Office will document the status of GO Associate and invite interested groups to join. *In progress; haven't decided on description of GO Associate.*
24. **Action item:** Chris will proceed with using OBOL to make computed definitions of cell differentiation and maintain them in a cell type ontology. *In progress.*
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. ***DONE.***