GO Content Meeting Nov 15-16 2005
The Institute for Genomic Research, Rockville, MD

Attendees

Alison Deckhut Augustine (NIAID)
Judith Blake (Mouse Genome Informatics)
Bernard de Bono (EMBL-EBI)
Shane Burgess (Mississippi State University)
Jennifer Clark (EMBL-EBI)
Russell Collins (University of Cambridge)
Alan Collmer (Cornell University)
Candace Collmer (Wells College/Cornell University)
Lindsay Cowell (Duke University)
Alexander Diehl (Mouse Genome Informatics)
Michelle Gwinn Giglio (The Institute for Genomic Research)
Linda Hannick (The Institute for Genomic Research)
Midori Harris (EMBL-EBI)
David Hill (Mouse Genome Informatics)
Amelia Ireland (EMBL-EBI)
Jamie Lee (U.T. Southwestern Medical Center)
Jane Lomax (EMBL-EBI)
Chris Mungall (HHMI)
Richard H. Scheuermann (U.T. Southwestern Medical Center)
Alex Sette (La Jolla Institute for Allergy & Immunology)

Contents

Changes in the Representation of Immunology in the Gene Ontology

Summary of Discussion of Changes in the Representation of Immunology in the Gene Ontology:
Alexander Diehl of Mouse Genome Informatics, The Jackson Laboratory, presented an overview of suggested term additions and changes, and reorganization of the biological_process ontology structure in the areas of immunology and response and detection. Prior to the meeting, Alex and Jamie Lee of the UT Southwestern Medical School, had prepared a series of DAGs (directed-acyclic graphs) showing revised portions biological_process ontology, and written a proposal outlining the suggested changes, which

participants in the meeting had been invited to review beforehand. These DAGs are available at SourceForge in entry 1301814: https://sourceforge.net/tracker/?func=detail&aid=1301814&group_id=36855&atid=440764

Alex outlined a set of changes to high level terms in the process ontology. These changes included changing the relationship of GO:0006955 immune response to GO:0006952 defense response to be sibling terms under a new term "physiological response to stimulus," to reflect the fact that not all immune responses are defense responses (tolerance induction to peripheral antigens, for instance) and the relationship of GO:0006954 inflammatory response to GO:0006955 immune response to be sibling terms under GO:0050874 organismal physiological process, to reflect the fact that inflammatory responses have immune and non-immune components. In the new scheme, GO:0045087 innate immune response and GO:0006954 inflammatory response are also direct children of a new term "physiological defense response," to reflect that these processes are types of defense response. These suggestions met with approval by the participants.

Alex outlined other changed and new DAGs showing immune system processes that occur prior to or outside the context of an immune response, such as early steps in T cell and B cell differentiation. Following discussion at the meeting, it was decided to create a high level term "immune system process" to be a parent of GO:0006955 immune response and related immune system processes such as "tolerance induction" and "immune system development."

Additional discussion at the meeting centered on definitions for "immune system process" and GO:0006955 immune response. After consideration of several alternate definitions, a consensus was reached on one that could encompass the breadth of these processes in animals and plants (see below for definitions).

Participants at the meeting indicated broad consensus with the lower level changes, additions, and reorganization of subprocesses of the immune response, inflammatory response, and other areas that Alex and Jamie had worked on prior to meeting, so that these can be implemented.

Further discussion focused on the so-called response and detection terms. A recent change to the GO was that detection terms are now considered subtypes (is-a) of their matching response terms, to reflect that detection processes are essentially induced responses in particular signaling molecules. While many participants at the meeting agreed with this change, there were some dissenters, and discussion was curtailed without complete consensus, and Amelia Ireland has requested additional discussion of the issue following the meeting. The new detection and response terms proposed by Alex were by and large accepted, with some minor changes to be made prior to implementation.

See the files immunology-introduction.ppt, immunology-new-defs-and-structure.ppt and big-picture.pdf for more detail.

Definitions decided upon at the meeting for selected terms:

immune system process
An organismal physiological process involved in the development or functioning of the

immune system, an organismal system for calibrated responses to potential internal or invasive threats.

immune response
Any immune system process that functions in the calibrated response by an organism to a potential internal or invasive threat.

In these two preceding definitions, the use of the word "calibrated" allows for tolerance and regulation, "potential" allows for not knowing whether something is really a threat or not ahead of time, and "internal or invasive" covers tumors and microbes, as well as individual proteins or other immunogenic molecules of self or non-self origin that in particular contexts within the body induce immune responses or tolerance induction, centrally or peripherally.

tolerance induction
An immune system process that results in a selective unresponsiveness to a specific antigen by an organism.

antigen processing and presentation
The process by which antigen-presenting cells express antigen (protein/peptide/lipid) on their cell surface in a form recognizable by lymphocytes.

immune system development
The process whose specific outcome is the progression of an organismal system whose objective is to provide calibrated responses by an organism to a potential internal or invasive threat, over time, from its formation to the mature structure. A system is a regularly interacting or interdependent group of organs or tissues that work together to carry out a given biological process.

Action Items for improving the representation of immunology in the GO

Action item 1. Revise a few selected term names per discussion at the GO Content Meeting – Alexander Diehl

Action item 2. Write up definitions for other proposed new terms remain to be written – Alexander Diehl and Jamie Lee.

Action item 3. Assign GO IDs to new terms and prepare OBO formatted file of new terms and definitions – Alexander Diehl

Action item 4. Prepare summary document describing the changes and listing all new terms, changed terms, and terms to be obsoleted – CONTACT _Con-415ED46E1 \c \s \l Alexander Diehl

Action item 5. Post summary document and finalized DAGs including definitions on SourceForge and seek comments – Alexander Diehl

Action item 6. Implement changes in the GO using the OBO file -- Jane Lomax and Amelia Ireland

Dual taxon proposal

Amelia gave a presentation [see file dual-taxon-annotation.ppt] proposing a mechanism for capturing the taxon of the interacting species when the process is an interaction between organisms, crucial for the increasing numbers of groups annotating parasites and host pathogen interactions.

Currently, in the taxon column (column 13) of the gene association files, the taxon id refers to the species that encodes the gene or gene product. The idea was to use a pipe in this column, and the taxon after the pipe would be that of the interacting species. In theory, this sytem can already be used in column 13, but so far no groups have put it in to practice, and the guidance as to how it can be used is misleading.

The wording for how the column should be used is proposed to be as follows:

“For cardinality 1, the ID of the species encoding the gene product.

For cardinality 2, to be used only in conjunction with terms that have the term 'interaction between organisms' as an ancestor. The first taxon id should be that of the organism encoding the gene or gene product, and the taxon id after the pipe should be that of the other organism in the interaction.

This field is mandatory, cardinality 1, 2; for cardinality 2 use a pipe to separate entries (e.g. taxon:1|taxon:1000)”

A biological example of how this could be used is in viral infection, where a virus infects a host via a natural portal such as the nose:

entry into host through natural portals taxon:1111|taxon:9999

where taxon:1111 is that of the virus, and taxon:9999 is that of the species being infected. Where the symbiont is able to perform the same interaction with several different species, the broadest taxon id should be used e.g. if a virus infects different mammalian species, the second taxon id should be that of all mammalia.

There was some discussion of this proposal by the group, and the consensus was that it was a good idea, and could be adopted immediately.

Action item: Publicize the use of dual taxon to annotations groups. Add to GO documentation.

Self And Non-Self

See also the file self-non-self.ppt and the SourceForge item 'Self and non-self processes (for content meeting)', SF:1351322

Jane and Amelia gave a presentation on processes involving more than one organism and how these should be represented in the ontology. Biological processes can be categorised by considering how many organisms are involved; some processes only involve a single organism, some involve two or more organisms, and some may be single or multi-organism. The problem with the current representation in the ontologies is that there are processes involving two organisms which have parentage under terms which are implicitly single organism. For example:

cell death
[i] programmed cell death
[i] cell killing of cells of other organism

This would lump together gene products involved in killing other organisms (e.g. toxins or venom) with those involved in apoptosis. As parent terms must be broader than their children, cell death (and its parent terms) therefore means the death of cells either in the organism being annotated *or* in another organism; in this case, it would be unclear whether the gene products were suicidal or homicidal!

The vast majority of terms make no reference to whether the process involves a single or multiple organisms, and most users assume the process involves a single organism, and that when they are annotating their gene product, the process that they are annotating it to occurs in the organism that they are annotating. Having terms such as 'cell killing of cells of other organism' under 'cell death' makes sense lexically, but either causes true path violations or violates assumptions about the meaning of terms, depending on which way you view it.

To solve these problems, the proposal was to make explicit whether a process involves one or more than one organism. Two solutions were proposed; the first was an annotation-based method, wherein GO terms would not be explicitly single or multi-organism; this detail would be captured in the annotation, either by using the dual taxon annotation method or by adding a new 'self' qualifier to denote endogenous processes. The alternative was an ontology-based solution, where ontology terms would explicitly state whether they referred to a single or multi-organism process.

There was some discussion on this topic following the presentation, but no consensus was reached as there was not universal comprehension of the problem being presented.

Interaction Between Organisms : Changes and Improvements

See also the file interactions.ppt

Problems to be addressed:

Problem I: Lack of Counterpart Terms

The terms under 'symbiosis, mutualism through parasitism ; GO:0044403' are biased towards processes involving a host. There should be counterpart terms for the same process in the other organism, and for those cases where there is no distinguishable 'host'.

Problem II: The 'Non-host' Problem

The term 'symbiotic interaction with other non-host organism ; GO:0043298' uses 'non-host' to describe a member of a symbiotic pairing where it is not clear . This brings into question what we mean by 'host', how we should refer to the other organism in the symbiosis, and what nomenclature should be used for cases where there is no clear 'host'.

Problem III: The 'Non-host' Problem II

Terms such as 'cytolysis of cells of another, non-host, organism ; GO:0001901' use 'non-host' to describe a second organism outside the context of a symbiotic relationship. This is confusing as it contradicts the usage in 'symbiotic interaction with other non-host organism ; GO:0043298'. It was also unclear from the term string that terms such as 'cytolysis of cells of another, non-host, organism ; GO:0001901' were supposed to exclude processes occurring during symbiosis.

Problem IV: Problems with Viruses

The processes under 'viral life cycle ; GO:0016032' have been used by all manner of non-viral organisms, despite being defined as terms describing viral processes.

Discussion of Interaction Between Organisms

There was some discussion on the topic of what to call the organisms involved in a symbiosis, and how to name terms so that it would be clear whether a term referred to symbiotic or nonsymbiotic process. The group agreed that the term 'host' be used for the larger organism in a symbiosis, and 'symbiont' be used for the smaller organism. Where there was no obvious size distinction between the organisms, the wording 'other organism' would be used, and the term appended with the phrase 'during symbiotic interaction' to make it clear that the process was symbiotic. The structure of symbiotic processes will now be:

acquisition of nutrients from other organism during symbiotic interaction
[i] acquisition of nutrients from host
[i] acquitision of nutrients from symbiont

As 'host' and 'symbiont' imply a symbiotic relationship, these terms do not need to be appended with 'during symbiotic interaction'.

For processes involving another organism which are not symbiotic in nature, it was suggested that the wording 'another organism' be employed, rather than the confusing string 'another non-host organism'. Rather than creating terms to specifically represent nonsymbiotic processes, the group agreed that these processes could be annotated by using a generic term, of which the symbiotic processes would be children. For example:

acquisition of nutrients from another organism
[i] acquisition of nutrients from other organism during symbiotic interaction
---[i] acquisition of nutrients from host
---[i] acquitision of nutrients from symbiont

Organisms obtaining nutrients from another organism where there was no symbiotic relationship between the two would use the term 'acquisition of nutrients from another organism'.

It was agreed that all existing terms under 'symbiosis, mutualism through parasitism ; GO:0044403' should be examined and where appropriate, counterpart terms added for 'symbiont' and 'other organism' processes.

Action item 1. Add counterpart terms to represent 'symbiont' and 'other organism' processes.
Action item 2. Rename 'non-host' terms according to new schema.

The discussion then moved to the terms involving viruses. These terms were initially created in two separate batches in consultation with virus experts, and are all housed under the term 'viral life cycle'. Since many of the processes involved are not unique to viruses, there are incidences in the ontology where there exists a process and a virus-specific child beneath it. It was suggested that the virus terms be broadened out so that they could be used by any organism. As the virus terms have also been used by annotators covering the host species, new terms may be required to cover the host involvement in the interaction.

Action item 3. Examine virus terms and remove virus specificity. Add terms to represent host involvement if required.