Thank you for your interest in helping the Pseudomonas Genome Database with its ongoing curation of the *Pseudomonas aeruginosa* PAO1 genome.

For those not familiar with Gene Ontology annotation and the benefits that will arise from this initiative, we have prepared this short overview.

Feel free to skip to the "**Directions for adding GO annotations to spreadsheet**" if you are already comfortable with the topic.

Gene Ontology Overview

The Gene Ontology provides a controlled vocabulary (GO Terms) and associated definitions to describe genes and their products. All Gene Ontology terms fall into one of three namespaces (domains)

1) Molecular Function
2) Biological Process
3) Cellular Component

The ontology terms are structured as a "directed acyclic graph" whereby each term has a number of defined relationships between other terms in the same namespace. **Benefits of mapping genes to GO terms:**

1. The use of GO terms by collaborating databases facilitates uniform queries across them.
2. Controlled vocabularies are structured so that they can be queried at different levels.
    a. For example, you can use GO to find all the gene products in the Pseudomonas aeruginosa PAO1 genome that are involved in signal transduction (GO:0007165), or you can zoom in on all gene products with phosphorelay response regulator activity (GO:0000156).

3. You can take advantage of a wide range of tools developed for querying gene ontology terms. For example, you can do GO term enrichment (over-representation analysis) on large data sets including RNA-Seq and microarray data.
4. Depending on how much we know about a particular gene, we can assign annotations at different levels of specificity.

Directions for adding GO annotations to spreadsheet

Choose a list of genes that you would like to annotate GO terms for and obtain their locus tags (PA numbers).

1) Add the gene's locus tag to the first column of each line.

In many cases, a single gene may have multiple GO terms (e.g. some bifunctional enzymes can be assigned to multiple molecular functions and transcription factors may regulate several biological processes).

Go to www.pseudomonas.com and look at the entry for your gene. See if any GO terms are already assigned to that gene. If the term is already assigned but is missing a reference OR has an ISS or RCA evidence code, please go ahead and annotate it. This will be very valuable. If you feel that any previously added GO terms are incorrect, please add a note in column 5(E) of the spreadsheet and we will flag it for removal.

2) Add a single GO accession to the second column (e.g. GO:2000284).

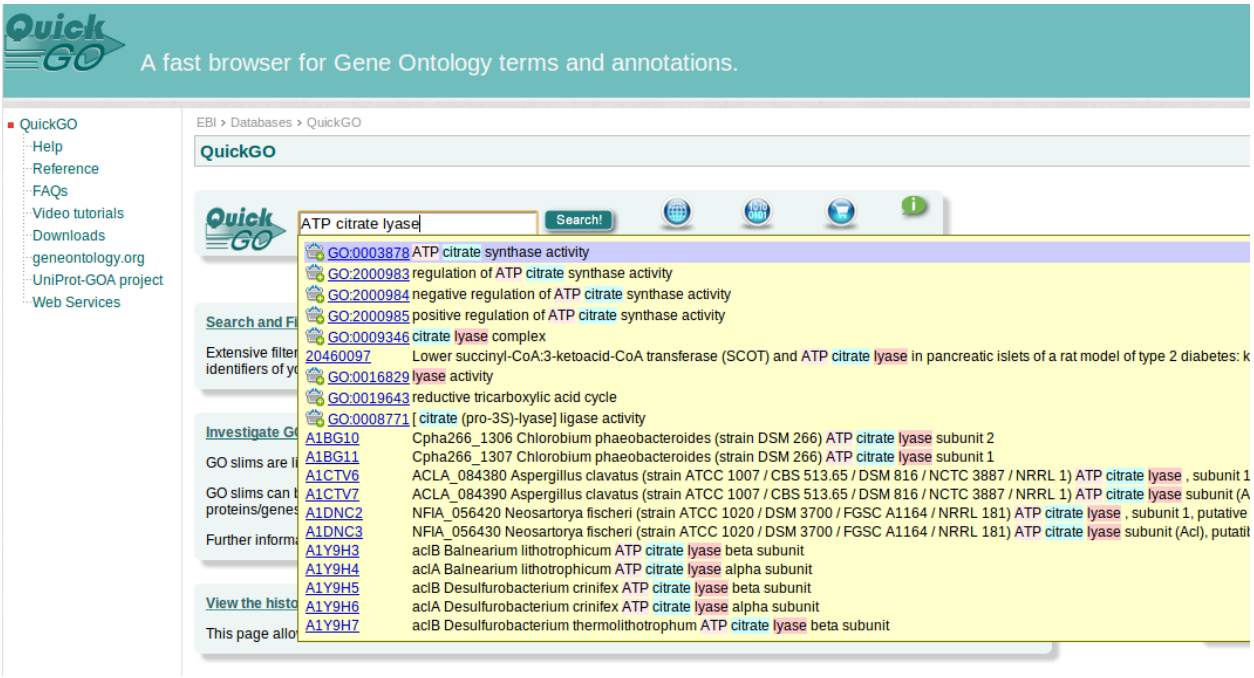
GO accessions and terms can be obtained using the QuickGO search form at the EBI website: http://www.ebi.ac.uk/QuickGO/

Important: Please be careful in distinguishing a gene's role as member of a biological process or molecular function versus **regulating** a biological process or molecular function.

For example, a gene can be a regulator of a gene involved in a cellular carbohydrate metabolic process OR the gene can itself be involved in a cellular carbohydrate metabolic process.:

regulation of cellular carbohydrate metabolic process - GO:0010675
cellular carbohydrate metabolic process - GO:0044262

For using Quick GO to determine the GO accession number, enter a function corresponding to the protein of interest (it is helpful to use keywords from the reference). A variety of hits will appear according to how well the search terms you have entered match functions, components or processes in the database.



Try to be as specific as possible in describing the protein function. GO ontologies are hierarchical and so broader functional descriptions will be captured by a specific GO term. Often, the specific function entered will not have and accession number, however most GO terms encompass synonyms of closely related or exact matches and so top hits from Quick GO should be examined for synonyms. The GO term capturing the appropriate function whether directly or as a synonym) should be used.

Example: the top GO hit for “ATP-citrate lyase” is ATP citrate lyase activity.

Therefore, GO:0003878 captures all of these alternatives as well:

Type	Synonym
exact	ATP:citrate oxaloacetate-lyase ((pro-S)-CH(2)COO(-)->acetyl-CoA) (ATP- dephosphorylating) activity
related	citric cleavage enzyme activity
exact	acetyl-CoA:oxaloacetate acetyltransferase (isomerizing; ADP-phosphorylating)
exact	ATP-citric lyase activity
exact	citrate-ATP lyase activity
exact	acetyl-CoA:oxaloacetate acetyltransferase (isomerizing; ADP- phosphorylating) activity
exact	adenosine triphosphate citrate lyase activity
related	citrate cleavage enzyme activity
related	ATP:citrate oxaloacetate-lyase [(pro-S)-CH2COO-rightacetyl-CoA] (ATP-dephosphorylating)
exact	ATP-citrate (pro-S)-lyase activity
exact	ATP-citrate (pro-S-)-lyase activity
exact	ATP citrate (pro-S)-lyase activity
related	acetyl-CoA:oxaloacetate C-acetyltransferase [(pro-S)-carboxymethyl-forming, ADP-phosphorylating]

Multiple GO annotations are desirable and they should describe different aspects of the protein's function. A single reference per GO annotation is required. The same reference can be used for multiple annotations, however a single line per annotation should be used. E.g. ATP citrate lyase could be described by all of the following: ATP citrate lyase activity, acetyl-CoA biosynthetic process, fatty acid biosynthetic process.

3) Add the three-character evidence code to column three.

Evidence codes are 3 letter terms which describe how the function of the protein has been verified/proven. Most evidence codes will be IDA (inferred from direct assay and includes assays such as enzyme assays, in vitro reconstitution, immunofluorescence, cell fractionation), IPI (inferred from physical interaction/binding assays), IMP (inferred from mutant phenotype and includes any gene mutation/knockout), or ISS (inferred from sequence or structural similarity, which includes sequence similarity [homolog of/most closely related to], recognized domains, structural similarity). If the reference refers to a homolog in an organisms other the *P. aeruginosa*, the evidence code will be ISS. More direct evidence codes are only applicable to orthologs in *P. aeruginosa*. Even other species of Pseudomonas are not close enough to warrant a direct evidence code and should be denoted as an ISS code.

Some examples are:
Inferred from Experiment (EXP)
Inferred from Direct Assay (IDA)
Inferred from Physical Interaction (IPI)
Inferred from Mutant Phenotype (IMP)

Please be as specific as possible when adding an evidence code (e.g. try not to use the EXP code). Many annotations will probably fall into the IDA or IMP evidence classification, but feel free to choose from any in the list linked to above.

For more information on assigning evidence codes, visit
http://www.geneontology.org/GO.evidence.shtml

4) Add a PubMed ID (PMID) to column 4.

Please add just one PMID per GO term. If you have two references for a particular term, please make a duplicate row (each containing different PubMed ID). Choose a reference which is recent, but moreover, describes an assay exhibiting the function of the protein. References using direct assays or describing knockout studies are preferential to sequence comparisons for providing evidence of function. Preference should be given to references in which the protein homolog under study was from Pseudomonas aeruginosa over other Pseudomonas species

strains, which are preferred over other organisms (for example E. coli).

6) Feel free to add extra comments or non-GO annotation data in the fifth column. We will use it to update the gene's annotation at a later date.

7) Not sure about what term to use OR missing a useful GO term? Contact us at pseudocap-mail@sfu.ca.

Appendix A:

Table A1: A demonstration of how to enter GO terms into the spreadsheet.

Locus Tag	GO Accession	Evidence Code	PubMed ID	Comments/Extra Annotations
PA0037	GO:2000284	IMP	1904858	
PA0289	GO:0051345	IDA	16339950	
PA0292	GO:0047632	IDA	11673419	
PA0294	GO:0051346	IMP	11673419	Please remove GO:0006560 (proline metabolic process) and GO:0006525 (arginine metabolic process).
PA0425	GO:0015238	IMP	8540696	