

Browsing and Searching Gene Ontology Resources Using AmiGO: A Tutorial

EMBRACE Workshop — Applied GO
November 7, 2007

Midori A. Harris, Ph.D, EMBL-EBI
Erika Feltrin, CRIBI, Padova
Based on a tutorial by Judith Blake, Ph.D,
Mouse Genome Informatics

Exploring GO Resources with AmiGO

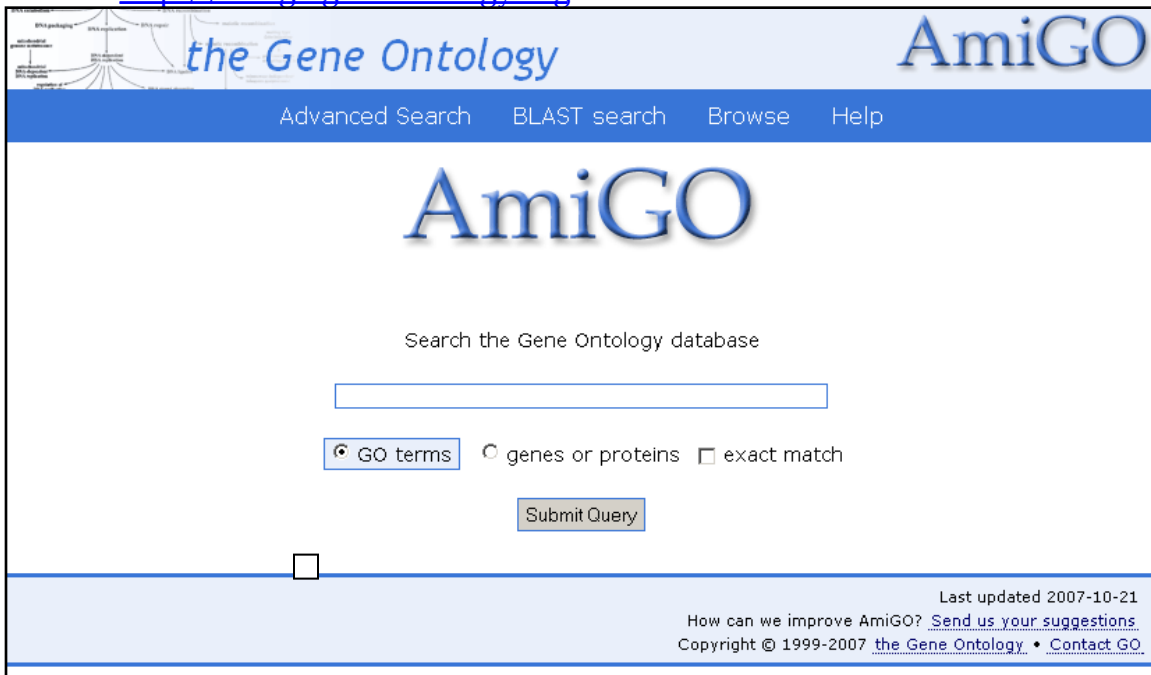
The Gene Ontology is a controlled vocabulary of terms to describe gene product characteristics in the domains of localization and function. Databases using GO terms to annotate their genes and gene products can submit their annotations to the GO consortium where they are made freely available for other users to download and utilize. In this tutorial, you will learn how to browse and search the GO and the gene product annotations made using GO terms.

This tutorial uses the AmiGO browser (<http://amigo.geneontology.org>) developed by the GO Consortium; there are also many other GO browsers, developed by outside groups, that can be used (see <http://www.geneontology.org/GO.tools.browsers.shtml>)

AmiGO Home Page – A Quick Search

The AmiGO home page provides a simple search interface.









- Go to <http://amigo.geneontology.org>:



- Enter the term 'carbohydrate metabolism' and click 'submit query'.

The query results will show all the terms that include the string 'carbohydrate metabolism'.

- Click on the first term, 'carbohydrate biosynthetic process'.

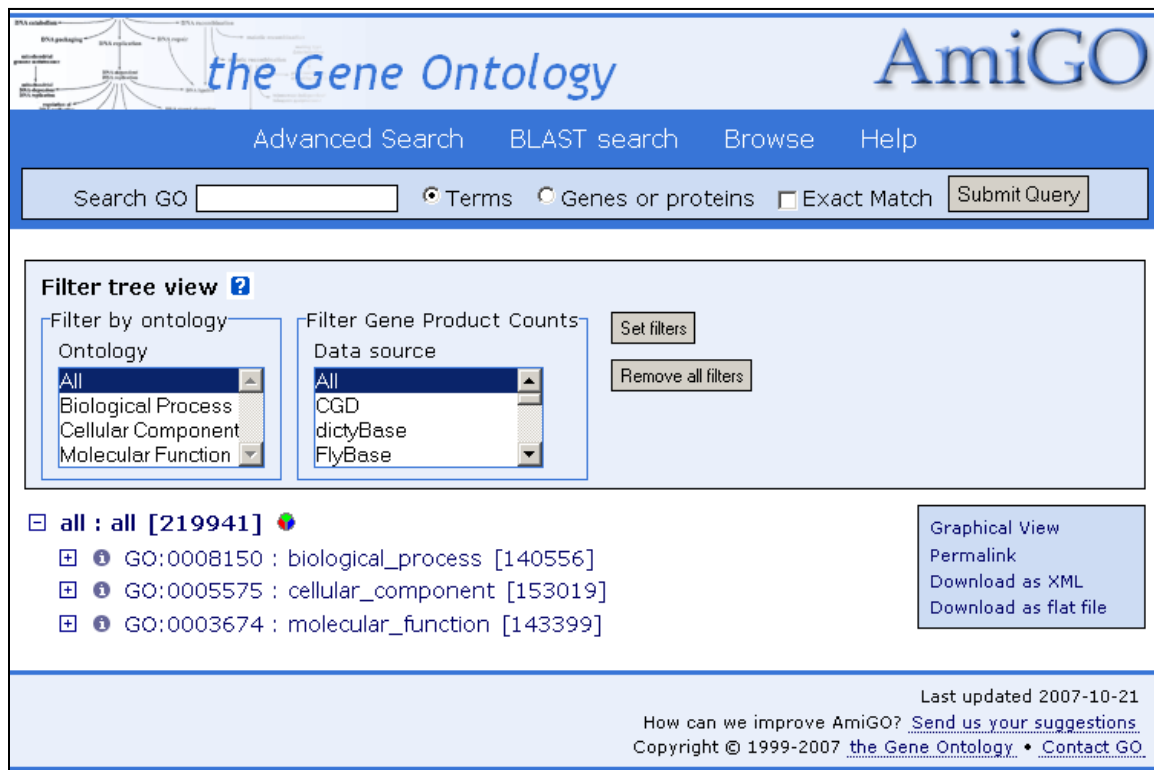
The first thing on each line can be either a ,  or a  icon.  can be used to expand a node, showing all the children of the selected term.  can be used to deselect a term, closing the node. Finally,  means that the term on that line has no children. The next thing on each line can be either a  or  icon, which represent, respectively, a part_of or is_a relationship. Following each term is a number in parentheses. This shows the total number of gene products that have been annotated to this term or to a more specific term below this in the GO tree.

Browsing and searching GO and its annotations

This part of the tutorial will show you how to browse and search the Gene Ontology and the annotations made using its terms.

- Return to AmiGO (<http://amigo.geneontology.org>); in the top toolbar, click on 'Browse'.







You will see the Gene Ontology graph with the three ontologies: **biological process**, **cellular component** and **molecular function**.



The screenshot shows the AmiGO web interface. At the top, there is a navigation bar with "Advanced Search", "BLAST search", "Browse", and "Help". Below this is a search bar with the text "Search GO" and a search button labeled "Submit Query". There are also radio buttons for "Terms" (selected), "Genes or proteins", and "Exact Match".

Below the search bar is a "Filter tree view" section. It contains two dropdown menus: "Filter by ontology" (with options: All, Biological Process, Cellular Component, Molecular Function) and "Filter Gene Product Counts" (with options: All, CGD, dictyBase, FlyBase). There are "Set filters" and "Remove all filters" buttons.

Below the filter section, the results are displayed as a list of terms with expand/collapse icons and gene product counts:

-   GO:0008150 : biological_process [140556]
-   GO:0005575 : cellular_component [153019]
-   GO:0003674 : molecular_function [143399]

On the right side of the results, there is a box with the following options: "Graphical View", "Permalink", "Download as XML", and "Download as flat file".

At the bottom of the page, there is a footer with the text: "Last updated 2007-10-21", "How can we improve AmiGO? [Send us your suggestions](#)", and "Copyright © 1999-2007 [the Gene Ontology](#) • [Contact GO](#)".

You can navigate the tree by clicking on the + and – icons at the start of each line.

- + expands a node, showing all the children of the term.
- closes the node, hiding the children.
- means that the term on that line has no children.

Each term gets one horizontal line.

- [-] ⓘ **GO:0005575 : cellular_component [153019]** 🌈
 - [+] ⓘ GO:0005623 : cell [109203]
 - [+] ⓘ GO:0044464 : cell part [109164]
 - [+] ⓘ GO:0031975 : envelope [3163]
 - [-] ⓘ **GO:0031012 : extracellular matrix [612]** 🌈
 - [+] ⓘ GO:0044420 : extracellular matrix part [301]
 - [+] ⓘ GO:0048196 : middle lamella-containing extracellular matrix [6]
 - [+] ⓘ GO:0005578 : proteinaceous extracellular matrix [553]

The ⓘ and ⓘ icons represent the relationship of the term to its parent, either “is a” or “part of” the parent term.

The GO term identifier and term name can be clicked to get a more detailed view of the term, including the definition and all genes and gene products annotated to the term.

Following the term ID and name is a number in parentheses. This is the total number of genes manually annotated to this term and its children. Electronic annotations (evidence code IEA) are not shown for two reasons: there are large numbers of these annotations, and they are deemed lower quality as they have not been checked by a human.

Terms may be followed by the 🌈 icon. Clicking this icon will bring you to a pie chart which displays the percentage of gene products annotated to each term below that selected. Clicking on a term itself will also give you the pie chart.

- **Open the nodes [biological process](#) -> [metabolic process](#) -> [catabolic process](#)**

How many annotations does [catabolic process](#) have?

Does this number include **IEA** annotations?

Click on [catabolic process](#). What relationship does [regulation of catabolic process](#) have to its parent [catabolic process](#)?

Clicking on a term name or ID opens the term detail page.

- **Click on ‘[regulation of catabolic process](#)’.**

The term detail page shows all the information available about the term: the term name and ID, any synonyms it might have, the term definition, its position in the GO structure, references to external databases, and the gene products associated

the Gene Ontology AmiGO

Advanced Search BLAST search Browse Help

Search GO Terms Genes or proteins Exact Match

regulation of catabolic process

[Term information](#) ↓ [Term lineage](#) ↓ [External references](#) ↓ [Term associations](#) →

Term Information

Accession GO:0009894

Ontology biological process

Synonyms
 exact: regulation of breakdown
 exact: regulation of catabolism
 exact: regulation of degradation

Definition Any process that modulates the frequency, rate, or extent of the chemical reactions and pathways resulting in the breakdown of substances. [source: GOC:go_curators]

Comment None

[Back to top](#)

Term Lineage

Filter tree view ?

Filter Gene Product Counts

Data source

- All
- CGD
- dictyBase
- FlyBase

Term View Options

Term ancestors Term parents, siblings and children

- ⊕ all : all [219941]
 - ⊕ **GO:0008150 : biological_process** [140556]
 - ⊕ **GO:0065007 : biological regulation** [20383]
 - ⊕ **GO:0050789 : regulation of biological process** [18539]
 - ⊕ **GO:0019222 : regulation of metabolic process** [10001]
 - ⊕ **GO:0009894 : regulation of catabolic process** [86]
 - ⊕ **GO:0008152 : metabolic process** [55902]
 - ⊕ **GO:0009056 : catabolic process** [4604]
 - ⊕ **GO:0009894 : regulation of catabolic process** [86]
 - ⊕ **GO:0019222 : regulation of metabolic process** [10001]
 - ⊕ **GO:0009894 : regulation of catabolic process** [86]

[Back to top](#)

External References

None.

with that term. Note that children of the selected term are not shown.

You can click on “Graphical view” for an alternate representation of the tree structure.

- Using the ‘search box’ at the upper left, search for the term **cytokine secretion**. It will open in a new window.

The screenshot shows the AmiGO Gene Ontology search interface. At the top, there is a navigation bar with links for 'Advanced Search', 'BLAST search', 'Browse', and 'Help'. Below this is a search bar with the text 'Search GO cytokine secretion' and a red circle around the search input field. To the right of the search bar are radio buttons for 'Terms' (selected), 'Genes or proteins', and 'Exact Match', along with a 'Submit Query' button. Below the search bar, the heading 'Term Search Results' is displayed. The main content area shows '8 results for cytokine secretion in field(s) term name, synonyms'. There is a 'Filter search results' section with a dropdown menu set to 'All' and buttons for 'Set filters' and 'Remove all filters'. Below this is a table of search results:

Term	Ontology
cytokine secretion ; GO:0050663 <small>[show def] [view associations]</small>	biological process
cytokine secretion during immune response ; GO:0002374 <small>[show def] [view associations]</small>	biological process
negative regulation of cytokine secretion ; GO:0050710 <small>[show def] [view associations]</small>	biological process
negative regulation of cytokine secretion during immune response ; GO:0002740 <small>[show def] [view associations]</small>	biological process

- Now click the top term, ‘cytokine secretion’, to open the term lineage; that is, all the terms above it in the tree are shown in this view.

What are the parents of **cytokine secretion** and what are the relationships between cytokine secretion and its parents?

For more info on the contributing databases see the Gene Ontology Consortium member list on the GO web site (<http://www.geneontology.org/GO.consortiumlist.shtml>) and the acknowledgements page (<http://www.geneontology.org/GO.acknowledgements.shtml>).

- Now click on the number next to the term ‘cytokine secretion’; in this case, on the [42]. A new page will open. Beneath the term information are the annotations, the genes or gene products assigned the selected term.

[cytokine secretion ; GO:0050663](#) [\[show def\]](#)

Qualifier	Name / Symbol	Information	Evidence	Reference	Assigned by
<input type="checkbox"/>	Lcp2 lymphocyte cytosolic protein 2	BLAST gene from <i>Mus musculus</i>	IMP With MGI:MGI:2180002	PMID:15153494	MGI
<input type="checkbox"/>	Lcp2 lymphocyte cytosolic protein 2	gene from <i>Rattus norvegicus</i>	ISS With MGI:1321402	RGD:1580654	RGD
<input type="checkbox"/>	Lyn Yamaguchi sarcoma viral (v-yes-1) oncogene homolog	gene from <i>Rattus norvegicus</i>	IMP	PMID:16177098	RGD
<input type="checkbox"/>	Ms4a2 membrane-spanning 4-domains, subfamily A, member 2	gene from <i>Rattus norvegicus</i>	IGI With RGD:621017	PMID:16177098	RGD
<input type="checkbox"/>	NOX5_HUMAN NOX5: NADPH oxidase 5	BLAST protein from <i>Homo sapiens</i>	NAS	PMID:11483596	UniProt

The first column is the gene or gene product symbol and name; clicking on the name will take you to the AmiGO gene product detail page, which shows the information held in the GO database about that gene product, including all its GO annotations and the peptide sequence (if available). By clicking on the left box and creating a check for the gene, you can then forward the FASTA protein sequence stored in the database to a BLAST search. We will do this at the end of this tutorial.

The second column is the data source that submitted the annotation (e.g. FlyBase, SGD, UniProt), and clicking on this takes you to the detail page at the source's website. The boldface term, such as '**gene**', indicates the entity (real thing) that is being annotated (associated with a GO term).

The third column is the evidence code for the association; when underlined, clicking on the evidence code brings up the source reference used to make the association.

The fourth column links to the biomedical literature from which the annotation was curated. In the case of PubMed IDs, this will link to PubMedCentral.

The final column has the name of the GOC resource group that generated this annotation.

You can choose to view the annotations to the term itself or to the term and its children.

- Go back to the [cytokine secretion](#) term detail page. From the drop-down menu on the right, choose get "All associations with terms" and click 'submit'.

How many gene products are annotated directly to **cytokine secretion**?

Note the filtering menus in the light grey box. You can also choose to filter annotations by the database that supplied them, by the evidence code used in the annotation and by species.

- Choose 'MGI' from the 'datasource' menu and click the 'Submit Query' button.

MGI stands for 'Mouse Genome Informatics', which is the *M. musculus* genome database. By filtering on this datasource, only gene associations from MGI i.e. all mouse are displayed in the gene list. This also works on the main page by using the filter boxes on the left of the interface.

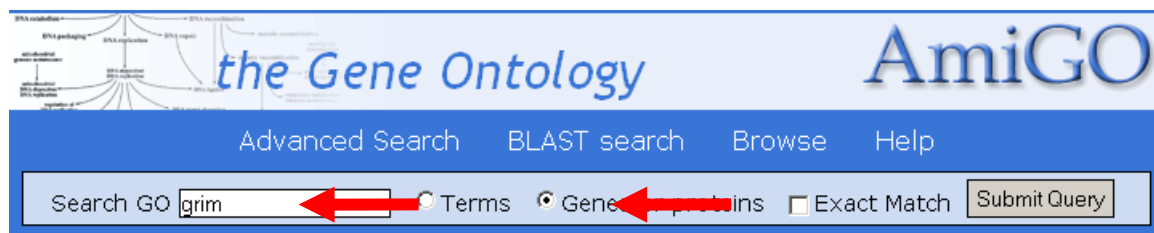
- Shut the popup window and return to the main page.

Searching with AmiGO

We have already used the search box to search by a GO term. Now we will search AmiGO using a gene product name

At the top of the page there is a search box. GO terms or associated gene products can be searched by checking the "terms" or "gene products" boxes respectively.

- Type 'grim' into the search box, and select 'Genes or proteins'.



The results list displays all gene products containing the text "grim", and the name of the external database which the gene product is from.

Why is the gene [Ndufa13](#) returned by the search with the gene symbol / name 'grim' ?

Click on the first gene 'grim' from the *Drosophila melanogaster* database. Scroll down the page to see all the gene products that are returned by this search.

Notice that one of the rows of annotation reads **'not'** nurse cell apoptosis. This means that the reference cited provides experimental information that the 'grim' gene is explicitly not involved in this process. This information is reported by scientists and reflected in the annotation when a gene product is *expected* to be involved in a particular function or process, but is shown not to be. The expectation may arise from overall sequence similarity to other genes that *do* have a particular function or are involved in a particular process.

	negative regulation of protein biosynthetic process [view associations]	biological process	TAS	PMID:12441239	FlyBase
	nurse cell apoptosis [view associations]	biological process	TAS	PMID:11139280	FlyBase
			TAS	PMID:12072176	
not	nurse cell apoptosis [view associations]	biological process	TAS	PMID:11139280	FlyBase
			TAS	PMID:12072176	
	programmed cell death [view associations]	biological process	TAS	PMID:12006672	FlyBase
			TAS	PMID:12072176	

Perform another search by gene product name. This time search for the FlyBase gene product name 'Buffy'

How many GO terms is the FlyBase gene product "Buffy" associated with?

- Do a search for your favorite gene product.

What gene did you search on?

Did you find this gene represented in AmiGO?

How many GO terms have been used to annotated this gene?

Does this gene need additional functional annotation in the GO system?

If your favorite gene product is missing critical information that you are aware of, you can contact the GO with suggestions for annotation. Go to the top menu bar in AmiGO and select the 'Help' button. Down the left menu of the Help page, click on 'HelpDesk'. Here you can contribute to the GO by sending in information about your favorite gene. Be sure to include:

Gene ID [from MOD, EntrezGene or UnitProt]
Gene Symbol or Name
Publication [PubMed ID] with experimental data
Suggestions [GO:ID and/or GO term] for GO annotation

Thank you.

The screenshot shows the AmiGO website interface. At the top, there is a navigation bar with links for 'Advanced Search', 'BLAST search', 'Browse', and 'Help'. Below this is a search bar with a 'Search GO' input field and radio buttons for 'Terms', 'Genes or proteins', and 'Exact Match', along with a 'Submit Query' button. The main content area is titled 'AmiGO Help: Gene Product Details' and contains a table of contents with links for 'Overview', 'Information', 'Term associations', and 'Filtering the results'. A sidebar on the left lists various navigation options, with 'GO helpdesk' highlighted. Two red arrows are present: one pointing to the 'Help' button in the top navigation bar and another pointing to the 'GO helpdesk' link in the sidebar.

Pie Charts

Return to the AmiGO main page (<http://amigo.geneontology.org>) and expand the term **biological process** to view its children. Choose 'MGI' from the 'datasource' menu on the left-hand side and click 'Submit Query'. Now click the pie icon that now appears next to **biological process**.

This displays the annotation totals from MGI (mouse) as a pie-chart. The terms included in the chart can be changed by expanding and collapsing nodes.

BLAST in GOdb =GOst

GOst is the Gene Ontology BLAST server, which allows you to blast a protein

sequence against all gene products that have a GO annotation. All the groups that contribute GO annotations to the AmiGO database also contribute a protein identifier for the gene product. This is used to create BLASTable database integrated with the GO annotations.

Return to the main AmiGO page and click “BLAST” (at the top of the page). Enter the UniProt accession number Q61337 into the top box and click submit.

This performs a Blast search of the protein sequence of **Q61337** against the GO database, and returns the top hits, with their GO annotations.

What is the top hit? What GO terms are associated with them?

Thank you for completing this GO tutorial using the AmiGO browser. We hope you have found this useful. Please send suggestions and comments to the GO help desk (use the link from AmiGO or email gohelp@geneontology.org). We need your help as much as we hope we can help you. Cheers!