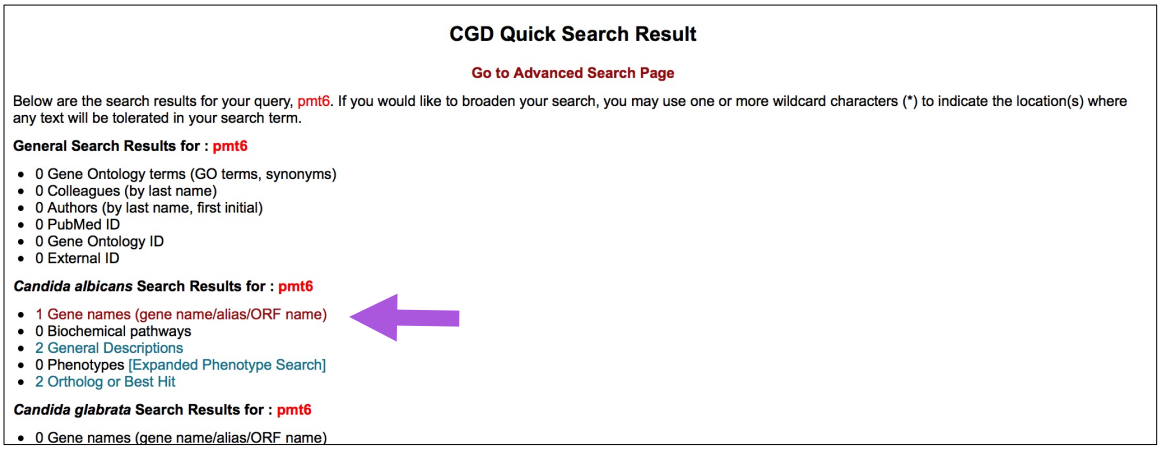
GO Enrichment, Phenotype Data at CGD

The Gene Ontology (GO) provides a common language to describe aspects of a gene product's biology. GO Terms are standardized phrases, arranged in a hierarchy, that describe a gene product's **molecular function** ("protein kinase activity"), **biological process** ("gluconeogenesis"), and **cellular component** ("cytoplasm"). Together, molecular function, biological process, and cellular component are the three ontologies of GO that describe a gene product's function, the processes that function is involved in, and the location where the function is performed.

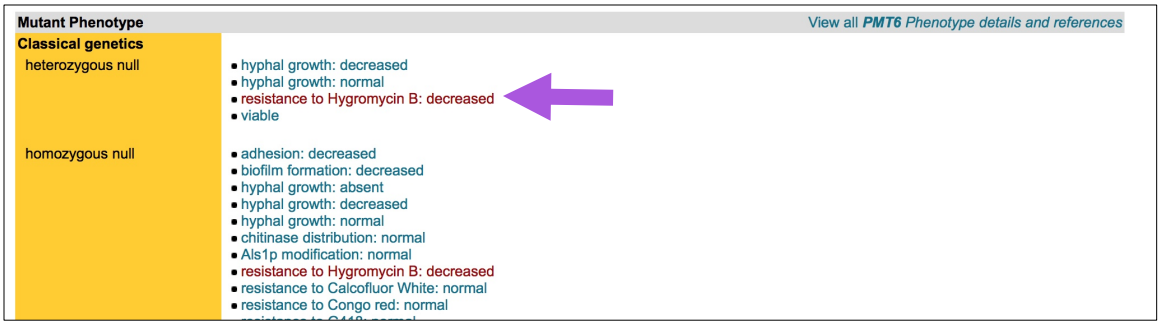
GO Term Finder takes a list of genes and identifies what GO terms are significant for the list. It is a powerful way to interpret the results of omics experiments or any situation where determining common functions and roles are important. For example, GO Term Finder can take a list of upregulated genes from a microarray experiment and determine what biological processes are significant for the set of genes, providing an idea of what processes are being upregulated in the cell.

In this exercise, we will attempt to uncover what processes are important for hygromycin B tolerance in *C. albicans.* To do so, we will use the CGD GO Term Finder to find shared biological processes for a set of genes whose mutation lowers resistance to hygromycin B.

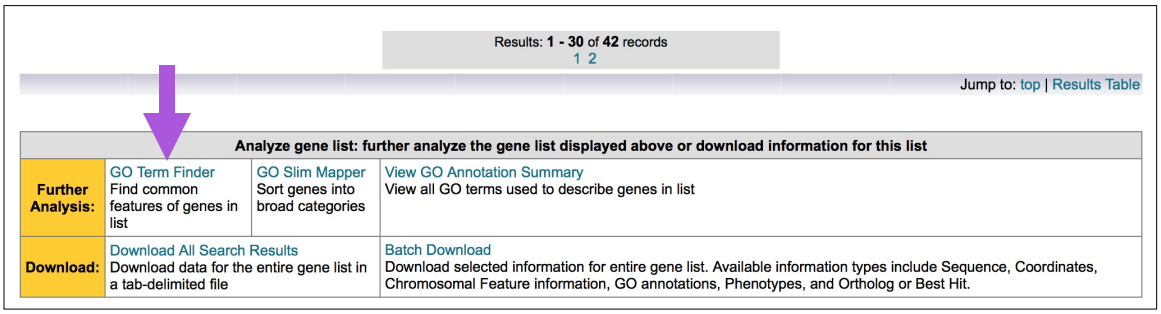
- From the CGD home page (www.candidagenome.org), go to the Locus Summary page for the hygromycin B-sensitivity gene PMT6. Enter **PMT6** into the **search our site** box and click **GO**. On the next page, under ***Candida albicans* Search Results**, click on hyperlinked **1 Gene names (gene name/alias/ORF name)**.



- From the PMT6 Locus Summary page, find other genes involved in hygromycin B sensitivity: scroll down to the **Mutant Phenotype** section and click on **resistance to Hygromycin B: decreased**

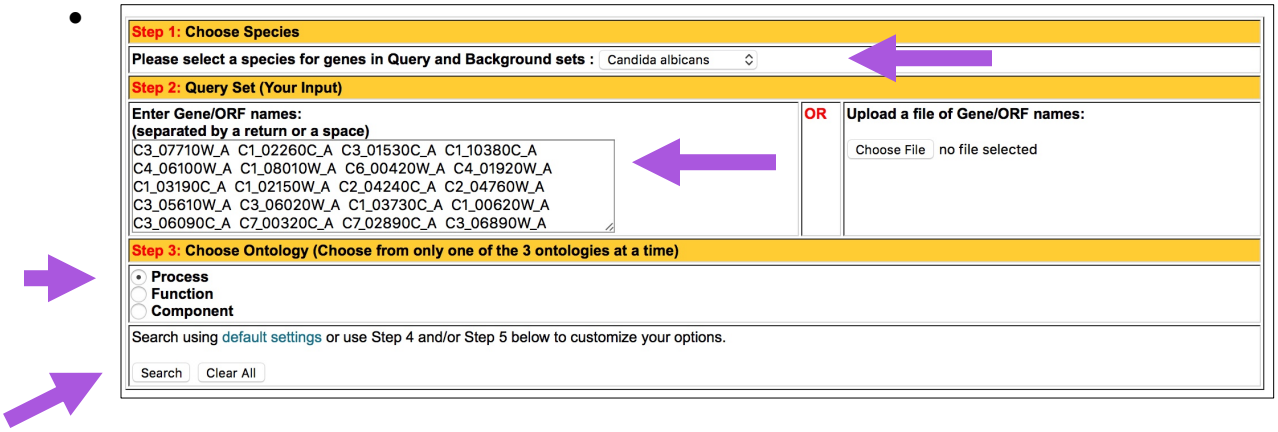


- On the **Phenotype Search Results** page, click on **Jump to: Analyze Gene List** above the table on the right (or simply scroll down to the bottom of the page). Click on **GO Term Finder** link.



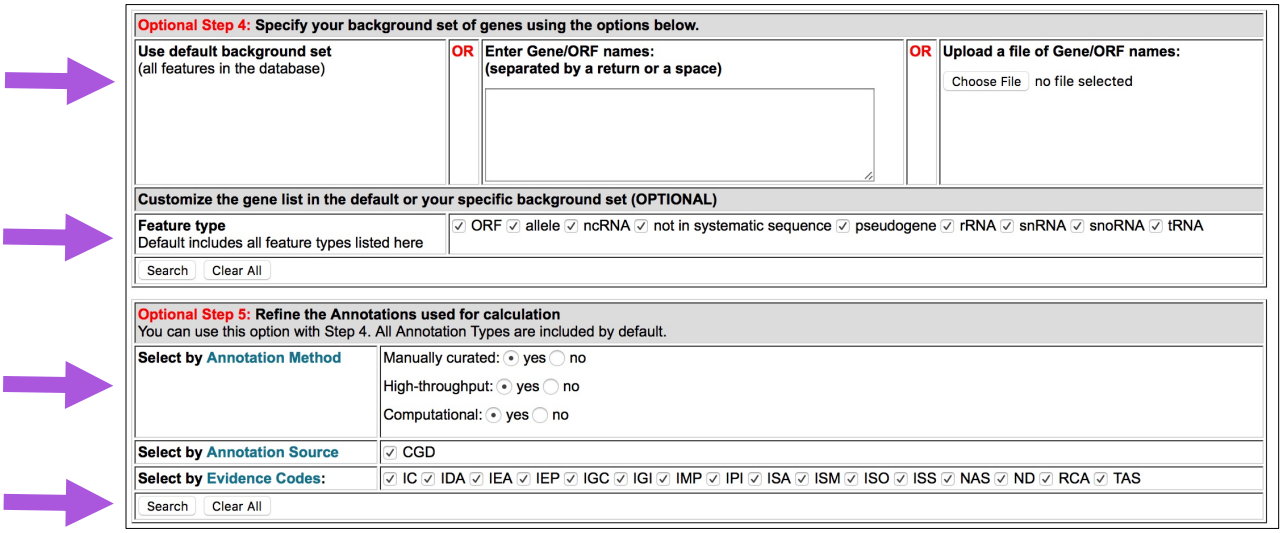
  - With your own list of genes, you can access GO Term Finder from any CGD page by opening **GO** menu in the banner on top and clicking on **GO Term Finder**. Or you use this URL: http://www.candidagenome.org/cgi-bin/GO/goTermFinder

- The **CGD Gene Ontology Term Finder** has five steps (two optional) to specify your query. First, make sure that **Candida albicans** is selected as your species.
- Your input genes should be already entered. Alternatively, copy and paste your own list of genes into the text box (note: the more genes processed, the longer it takes).Choose **Process** as the ontology. Click the **Search** button to use the default settings.
-

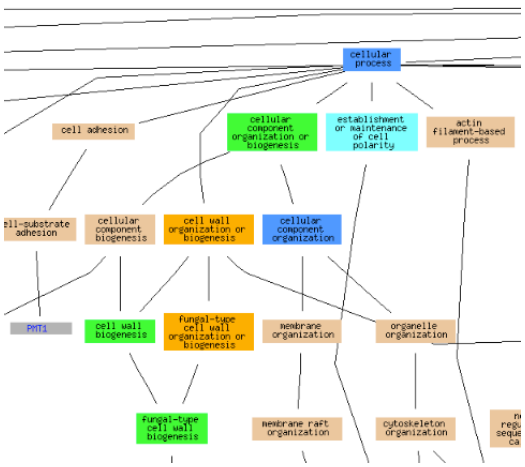


You can further customize your query in the next steps down the page:

- Optional Step 4 allows submitting a custom background set; use default set, all *C. albicans* genes in CGD
- Step 4 also allows restricting the search to specific feature types; use default settings
- Optional Step 5 allows selection of annotation methods, sources and evidence; leave all options checked

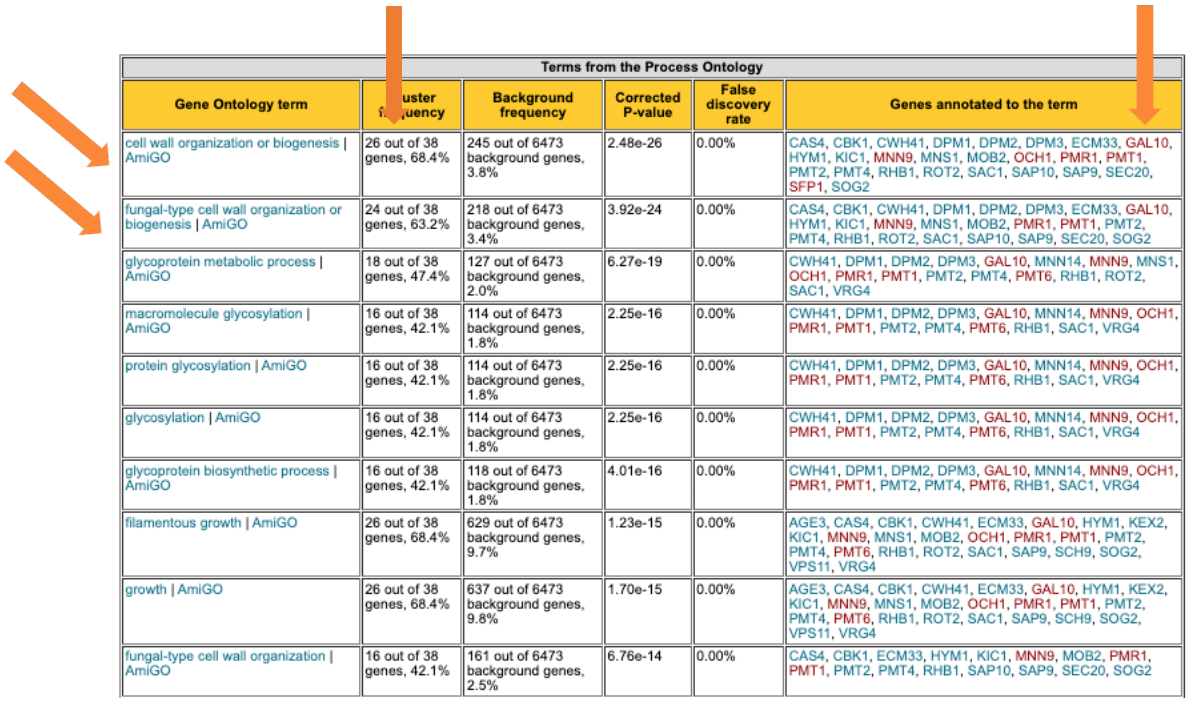


- Click **Search**. The input is checked and any genes that are not recognized as valid for the selected *Candida* species are rejected; click on **Proceed** in the following window.
- The results page displays the significant shared GO terms (or their parents) in both graphic and table form, within the set of genes associated with hygromycin B sensitivity entered on the previous page:



The graph shows the GO tree that includes terms used directly or indirectly in annotations for the genes in your list. The terms are color-coded to indicate their statistical significance (p-value score). Genes associated with the GO terms are shown in gray boxes, with links to their respective Locus Summary pages

- The table below the graph lists each significant GO term, the number of times the GO term is used to annotate genes in the list, and the number of times that the term is used to annotate genes in the background set (all genes in *C. albicans* genome)



Terms from the Process Ontology

Gene Ontology term	Cluster frequency	Background frequency	Corrected P-value	False discovery rate	Genes annotated to the term
cell wall organization or biogenesis | AmiGO	26 out of 38 genes, 68.4%	245 out of 6473 background genes, 3.8%	2.48e-26	0.00%	CAS4, CBK1, CWH41, DPM1, DPM2, DPM3, ECM33, GAL10, HYM1, KIC1, MNN9, MNS1, MOB2, OCH1, PMR1, PMT1, PMT2, PMT4, RHB1, ROT2, SAC1, SAP10, SAP9, SEC20, SFP1, SOG2
fungal-type cell wall organization or biogenesis | AmiGO	24 out of 38 genes, 63.2%	218 out of 6473 background genes, 3.4%	3.92e-24	0.00%	CAS4, CBK1, CWH41, DPM1, DPM2, DPM3, ECM33, GAL10, HYM1, KIC1, MNN9, MNS1, MOB2, PMR1, PMT1, PMT2, PMT4, RHB1, ROT2, SAC1, SAP10, SAP9, SEC20, SOG2
glycoprotein metabolic process | AmiGO	18 out of 38 genes, 47.4%	127 out of 6473 background genes, 2.0%	6.27e-19	0.00%	CWH41, DPM1, DPM2, DPM3, GAL10, MNN14, MNN9, MNS1, OCH1, PMR1, PMT1, PMT2, PMT4, PMT6, RHB1, ROT2, SAC1, VRG4
macromolecule glycosylation | AmiGO	16 out of 38 genes, 42.1%	114 out of 6473 background genes, 1.8%	2.25e-16	0.00%	CWH41, DPM1, DPM2, DPM3, GAL10, MNN14, MNN9, OCH1, PMR1, PMT1, PMT2, PMT4, PMT6, RHB1, SAC1, VRG4
protein glycosylation | AmiGO	16 out of 38 genes, 42.1%	114 out of 6473 background genes, 1.8%	2.25e-16	0.00%	CWH41, DPM1, DPM2, DPM3, GAL10, MNN14, MNN9, OCH1, PMR1, PMT1, PMT2, PMT4, PMT6, RHB1, SAC1, VRG4
glycosylation | AmiGO	16 out of 38 genes, 42.1%	114 out of 6473 background genes, 1.8%	2.25e-16	0.00%	CWH41, DPM1, DPM2, DPM3, GAL10, MNN14, MNN9, OCH1, PMR1, PMT1, PMT2, PMT4, PMT6, RHB1, SAC1, VRG4
glycoprotein biosynthetic process | AmiGO	16 out of 38 genes, 42.1%	118 out of 6473 background genes, 1.8%	4.01e-16	0.00%	CWH41, DPM1, DPM2, DPM3, GAL10, MNN14, MNN9, OCH1, PMR1, PMT1, PMT2, PMT4, PMT6, RHB1, SAC1, VRG4
filamentous growth | AmiGO	26 out of 38 genes, 68.4%	629 out of 6473 background genes, 9.7%	1.23e-15	0.00%	AGE3, CAS4, CBK1, CWH41, ECM33, GAL10, HYM1, KEX2, KIC1, MNN9, MNS1, MOB2, OCH1, PMR1, PMT1, PMT2, PMT4, PMT6, RHB1, ROT2, SAC1, SAP9, SCH9, SOG2, VPS11, VRG4
growth | AmiGO	26 out of 38 genes, 68.4%	637 out of 6473 background genes, 9.8%	1.70e-15	0.00%	AGE3, CAS4, CBK1, CWH41, ECM33, GAL10, HYM1, KEX2, KIC1, MNN9, MNS1, MOB2, OCH1, PMR1, PMT1, PMT2, PMT4, PMT6, RHB1, ROT2, SAC1, SAP9, SCH9, SOG2, VPS11, VRG4
fungal-type cell wall organization | AmiGO	16 out of 38 genes, 42.1%	161 out of 6473 background genes, 2.5%	6.76e-14	0.00%	CAS4, CBK1, ECM33, HYM1, KIC1, MNN9, MOB2, PMR1, PMT1, PMT2, PMT4, RHB1, SAP10, SAP9, SEC20, SOG2

- Additional columns list the p-value, the false discovery rate (FDR), and a list of all the genes annotated, either directly or indirectly, to the term. FDR is an estimate of the percent chance that a particular GO term might actually be a false positive. It represents the fraction of the nodes with p-values as good or better than the node with this FDR that would be expected to be false positives.

- Explore the table. Based on the results, what biological processes are important for resisting the antibiotic action of hygromycin B in *C. albicans* cells?