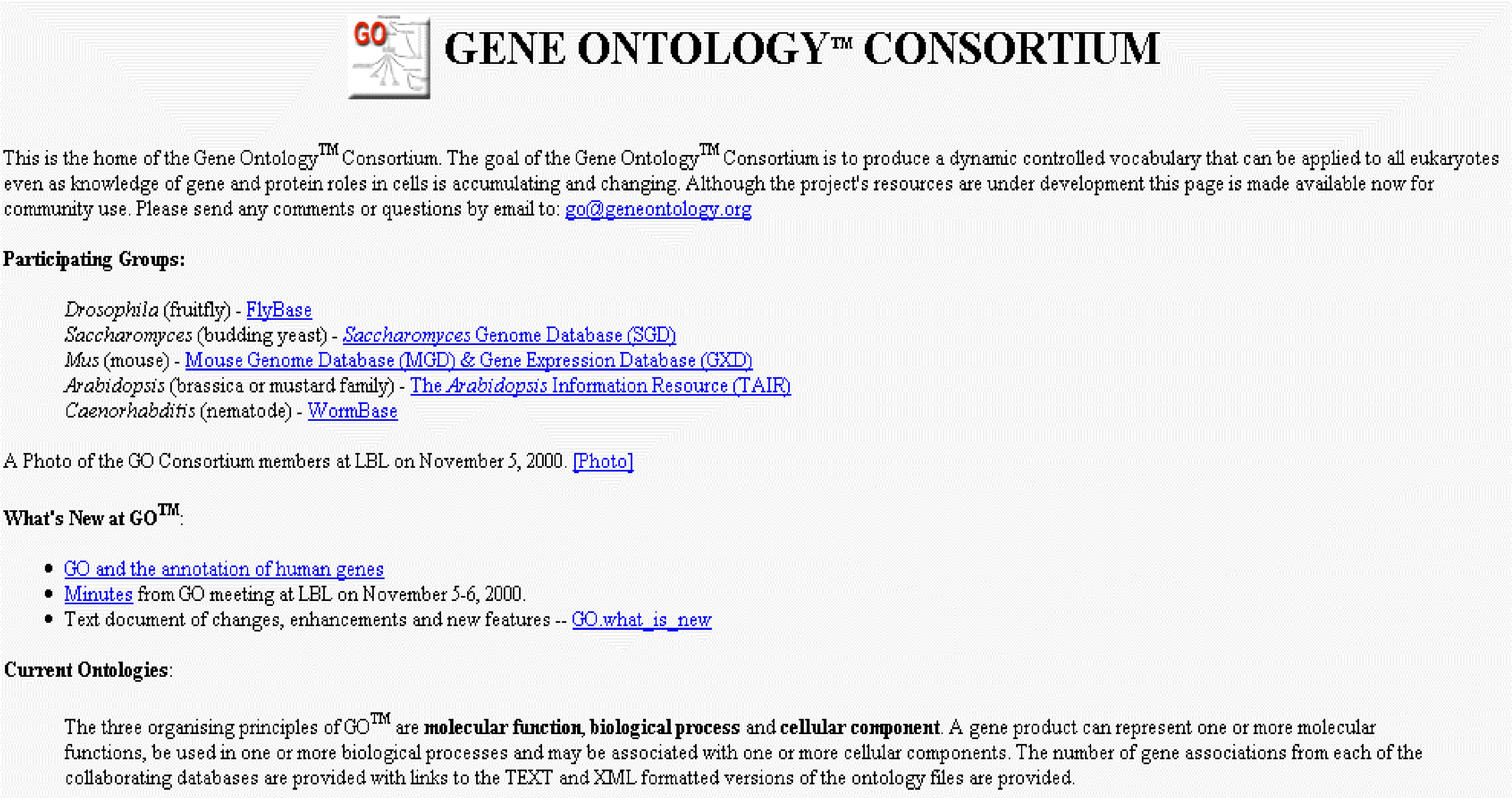
MAPPING OF SEQUENCES TO GENE ONTOLOGY

GO consortium



GENE ONTOLOGY™ CONSORTIUM

This is the home of the Gene Ontology^{TM} Consortium. The goal of the Gene Ontology^{TM} Consortium is to produce a dynamic controlled vocabulary that can be applied to all eukaryotes even as knowledge of gene and protein roles in cells is accumulating and changing. Although the project's resources are under development this page is made available now for community use. Please send any comments or questions by email to: go@geneontology.org

Participating Groups:

Drosophila (fruitfly) - FlyBase
Saccharomyces (budding yeast) - *Saccharomyces* Genome Database (SGD)
Mus (mouse) - Mouse Genome Database (MGD) & Gene Expression Database (GXD)
Arabidopsis (brassica or mustard family) - The *Arabidopsis* Information Resource (TAIR)
Caenorhabditis (nematode) - WormBase

A Photo of the GO Consortium members at LBL on November 5, 2000. [Photo]

What's New at GO^{TM}:

- GO and the annotation of human genes
- Minutes from GO meeting at LBL on November 5-6, 2000.
- Text document of changes, enhancements and new features -- GO.what_is_new

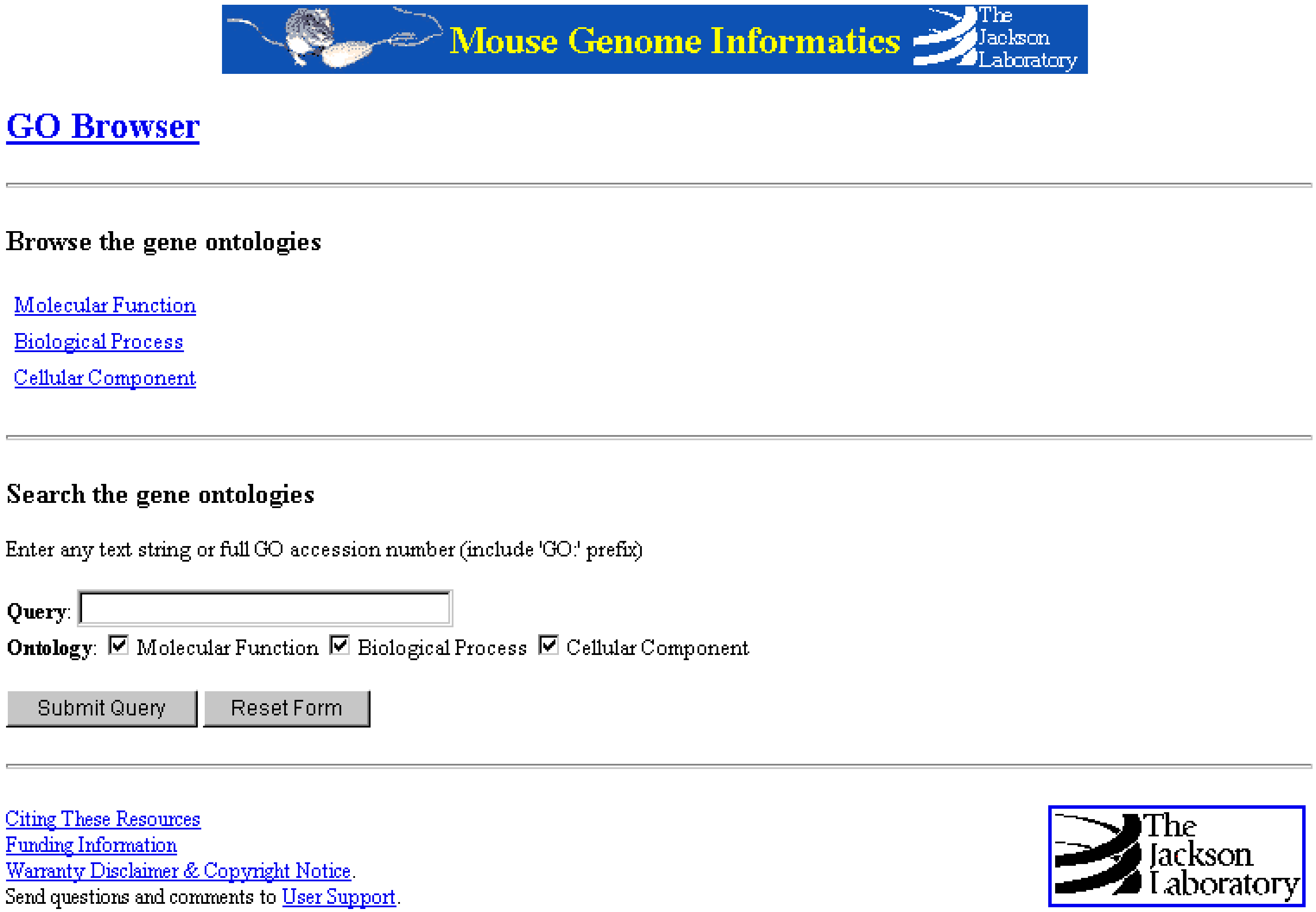
Current Ontologies:

The three organising principles of GO^{TM} are **molecular function, biological process** and **cellular component**. A gene product can represent one or more molecular functions, be used in one or more biological processes and may be associated with one or more cellular components. The number of gene associations from each of the collaborating databases are provided with links to the TEXT and XML formatted versions of the ontology files are provided.

Current Ontologies

- **Molecular function**: tasks performed by gene product
- **Biological process**: broad biological goals accomplished by ordered assemblies of molecular functions
- **Cellular component**: subcellular structures, locations and macromolecular complexes

MGD GO browser



GO Browser

Browse the gene ontologies

Molecular Function
Biological Process
Cellular Component

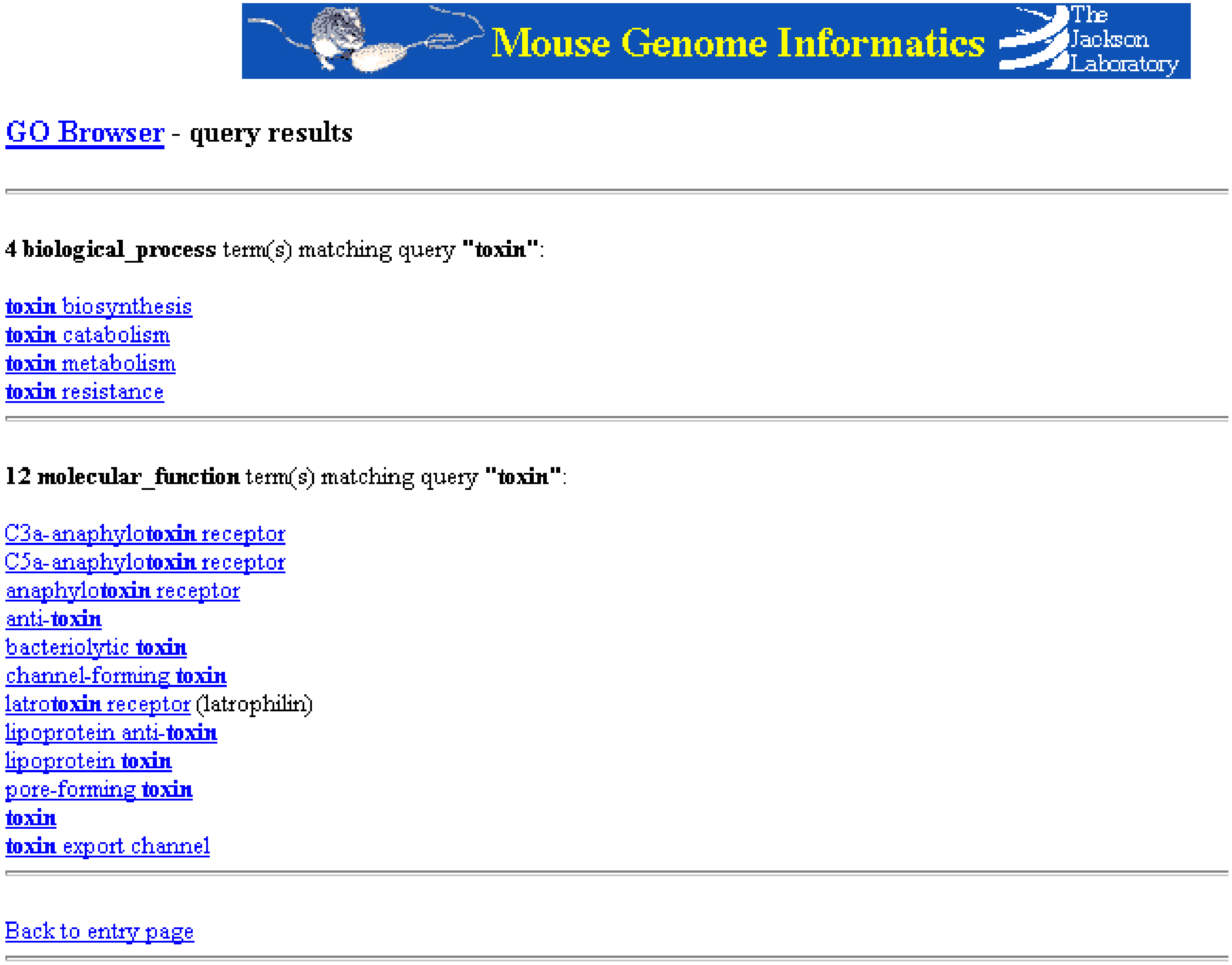
Search the gene ontologies

Enter any text string or full GO accession number (include 'GO:' prefix)

Query:
Ontology: ☑ Molecular Function ☑ Biological Process ☑ Cellular Component

Citing These Resources
Funding Information
Warranty Disclaimer & Copyright Notice.
Send questions and comments to User Support.

Search result for toxin



GO Browser - query results

---

4 biological_process term(s) matching query **"toxin"**:

toxin biosynthesis
toxin catabolism
toxin metabolism
toxin resistance

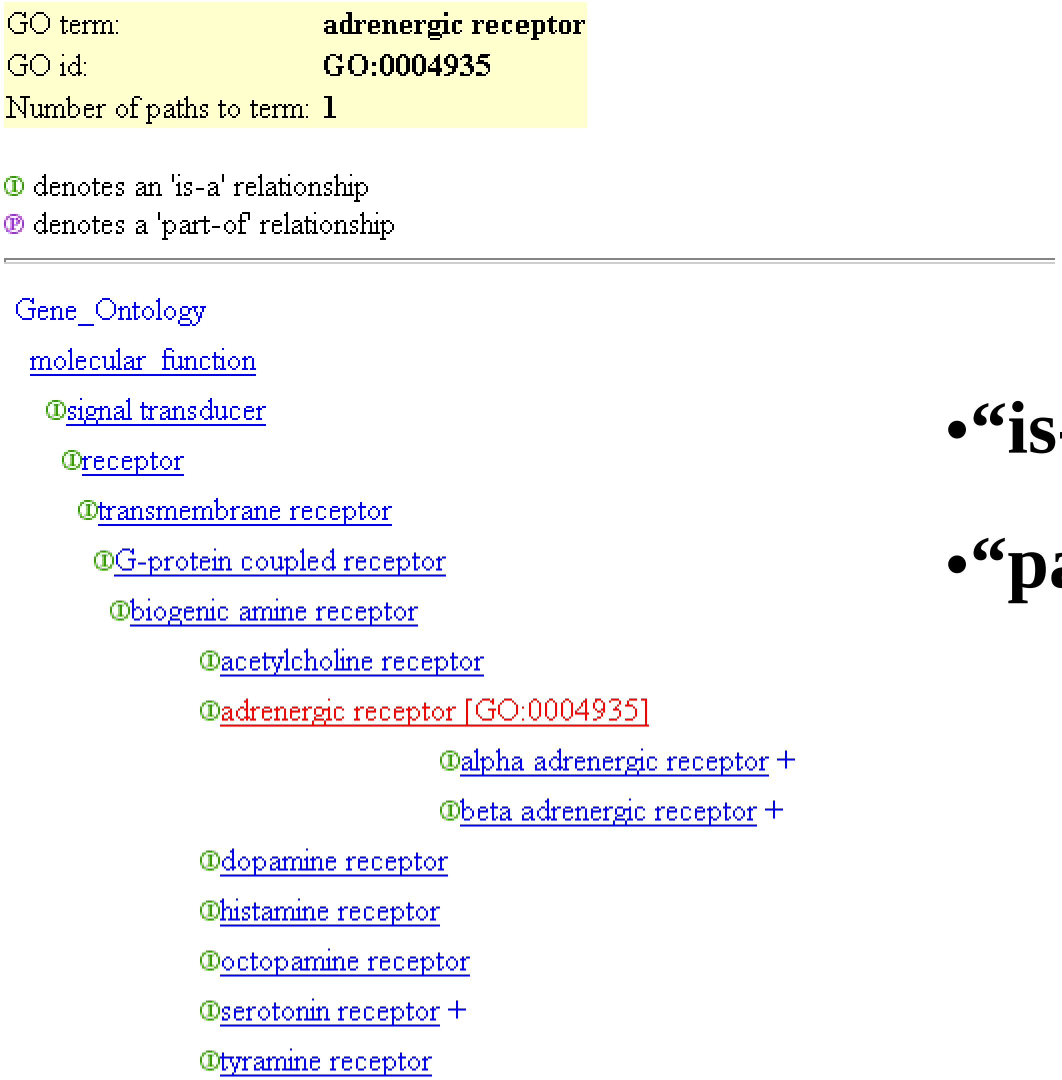
---

12 molecular_function term(s) matching query **"toxin"**:

C3a-anaphylo**toxin** receptor
C5a-anaphylo**toxin** receptor
anaphylo**toxin** receptor
anti-**toxin**
bacteriolytic **toxin**
channel-forming **toxin**
latro**toxin** receptor (latrophilin)
lipoprotein anti-**toxin**
lipoprotein **toxin**
pore-forming **toxin**
toxin
toxin export channel

Back to entry page

Relationships in GO



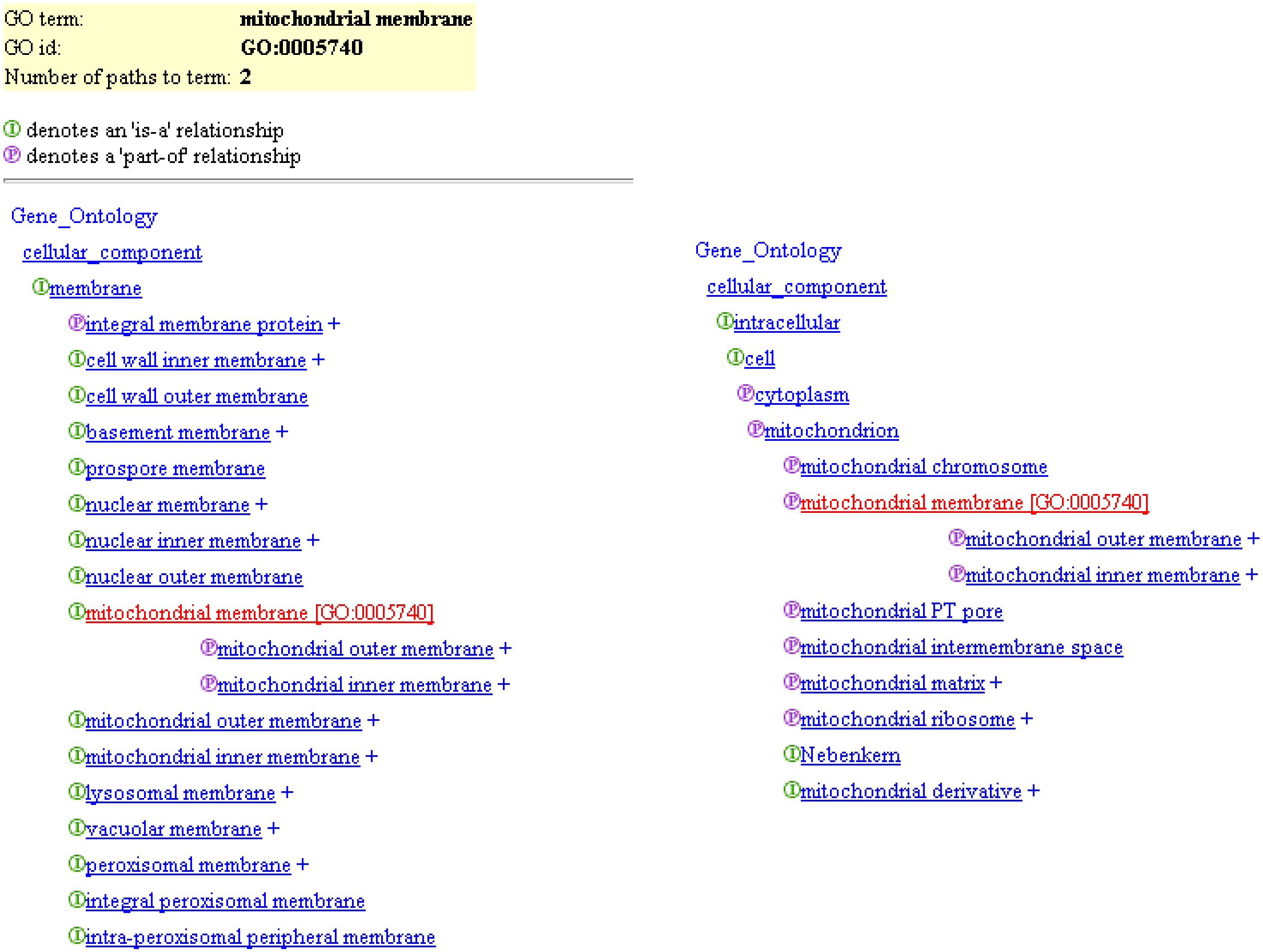
GO term:	**adrenergic receptor**
GO id:	**GO:0004935**
Number of paths to term:	**1**

Ⓘ denotes an 'is-a' relationship
Ⓟ denotes a 'part-of' relationship

- Gene_Ontology
  - molecular_function
    - Ⓘsignal transducer
      - Ⓘreceptor
        - Ⓘtransmembrane receptor
          - ⒾG-protein coupled receptor
            - Ⓘbiogenic amine receptor
              - Ⓘacetylcholine receptor
              - Ⓘadrenergic receptor [GO:0004935]
                - Ⓘalpha adrenergic receptor +
                - Ⓘbeta adrenergic receptor +
              - Ⓘdopamine receptor
              - Ⓘhistamine receptor
              - Ⓘoctopamine receptor
              - Ⓘserotonin receptor +
              - Ⓘtyramine receptor

- “is-a”
- “part of”

GO paths to terms



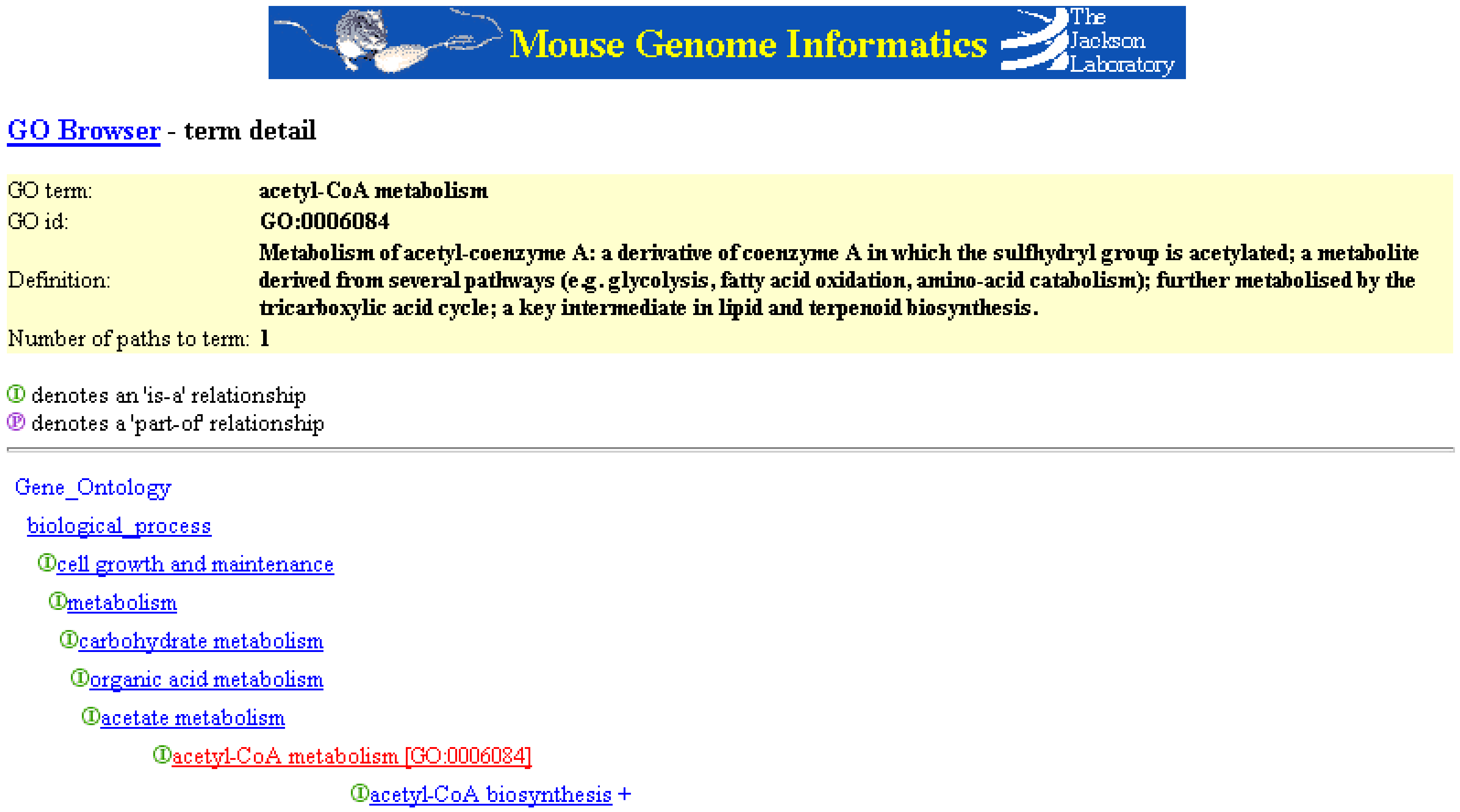
GO term:	**mitochondrial membrane**
GO id:	**GO:0005740**
Number of paths to term:	**2**

Ⓘ denotes an 'is-a' relationship
Ⓟ denotes a 'part-of' relationship

- Gene_Ontology
  - cellular_component
    - Ⓘmembrane
      - Ⓟintegral membrane protein +
      - Ⓘcell wall inner membrane +
      - Ⓘcell wall outer membrane
      - Ⓘbasement membrane +
      - Ⓘprospore membrane
      - Ⓘnuclear membrane +
      - Ⓘnuclear inner membrane +
      - Ⓘnuclear outer membrane
      - Ⓘmitochondrial membrane [GO:0005740]
        - Ⓟmitochondrial outer membrane +
        - Ⓟmitochondrial inner membrane +
      - Ⓘmitochondrial outer membrane +
      - Ⓘmitochondrial inner membrane +
      - Ⓘlysosomal membrane +
      - Ⓘvacuolar membrane +
      - Ⓘperoxisomal membrane +
      - Ⓘintegral peroxisomal membrane
      - Ⓘintra-peroxisomal peripheral membrane

- Gene_Ontology
  - cellular_component
    - Ⓘintracellular
      - Ⓘcell
        - Ⓟcytoplasm
          - Ⓟmitochondrion
            - Ⓟmitochondrial chromosome
            - Ⓟmitochondrial membrane [GO:0005740]
              - Ⓟmitochondrial outer membrane +
              - Ⓟmitochondrial inner membrane +
            - Ⓟmitochondrial PT pore
            - Ⓟmitochondrial intermembrane space
            - Ⓟmitochondrial matrix +
            - Ⓟmitochondrial ribosome +
            - ⒾNebenkern
            - Ⓘmitochondrial derivative +

GO definitions



GO Browser - term detail

GO term:	**acetyl-CoA metabolism**
GO id:	**GO:0006084**
Definition:	**Metabolism of acetyl-coenzyme A: a derivative of coenzyme A in which the sulfhydryl group is acetylated; a metabolite derived from several pathways (e.g. glycolysis, fatty acid oxidation, amino-acid catabolism); further metabolised by the tricarboxylic acid cycle; a key intermediate in lipid and terpenoid biosynthesis.**
Number of paths to term:	**1**

Ⓘ denotes an 'is-a' relationship
Ⓟ denotes a 'part-of' relationship

Gene_Ontology
 biological_process
  Ⓘcell growth and maintenance
   Ⓘmetabolism
    Ⓘcarbohydrate metabolism
     Ⓘorganic acid metabolism
      Ⓘacetate metabolism
       Ⓘacetyl-CoA metabolism [GO:0006084]
        Ⓘacetyl-CoA biosynthesis +

Why the interest in GO?

- Universal ontology
- Functional classification scheme with many different levels in a DAG
- Widespread interest from scientific community
- Already mappings to SP keywords and gene products- annotation of fly, mouse and yeast

Current Mappings to GO

- Consortium mappings -MGD, SGD, FlyBase
- Other DB -TAIR, Pombe
- Swiss-Prot keywords
- EC numbers
- InterPro entries
- Medline ID
- Commercial companies -CompuGen, Proteome

MAPPING OF INTERPRO TO GO

- PFD Protein folding and degradation
  - -PFDc chaperone
  - -PFDp protease/endopeptidase
  - -PFDi protease inhibitor

- TRS Transport and secretion
  - -TRSt transport-subtrates
  - -TRSi transport-ions
  - -TRSs secretion
  - -TRSr carrier proteins

- CYS Cytoskeletal/structural
  - -CYSc cytoskeletal
  - -CYSs structural
  - -CYSv virus coat/capsid protein

- STD Signal trandsuction & kinases
  - -STDk signal transduction kinases
  - -STDp signal transduction phosphatases
  - -STDr signal transduction RR
  - -STDs signal transduction sensors
  - -STDc cell signalling

- DRM DNA/RNA metabolism
  - -DRMr DNA repair & recombination
  - -DRMp DNA replication
  - -DRMm DNA/RNA modification
  - -DRMt transcription/translation
  - -DRMb ribosomal protein

- CGD Cell cycle, growth, death
  - -CGDc cell cycle & division
  - -CGDg cell growth & development
  - -CGDd cell death

- MET Metabolism
  - -METs general substrate metabolism
  - -METa amino acid metabolism
  - -METn nucleic acid metabolism
  - -METm metal binding proteins
  - -METe electron transfer

- DRG DNA/RNA binding- regulation

- PRG Protein-binding & other regulation
  - -PRGg GPCRs
  - -PRGo other regulation

- OTH Other functions
  - -OTHm cell motility
  - -OTHt transposition
  - -OTHh hormones
  - -OTHa cell adhesion
  - -OTHo miscellaneous functions

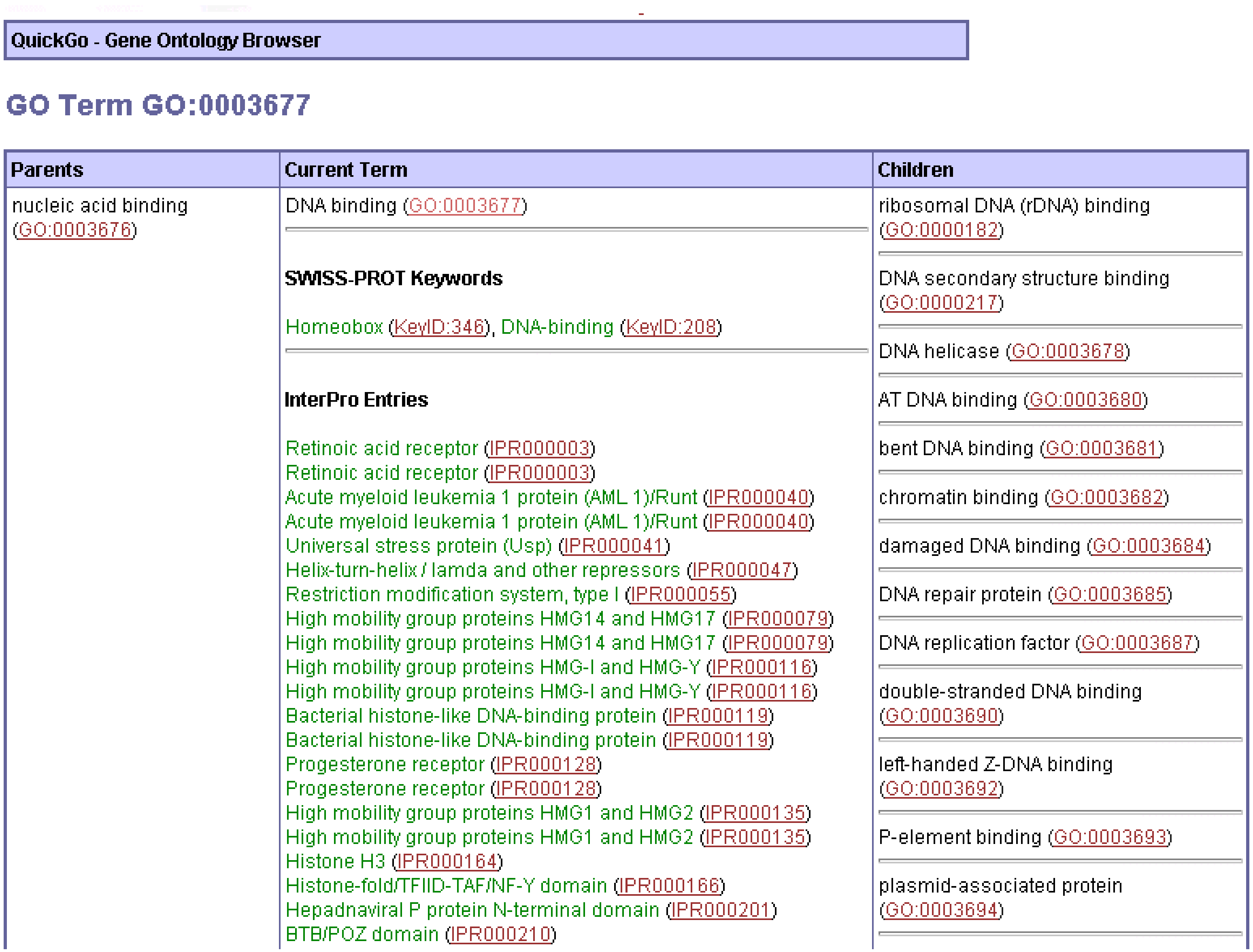
- DIT Defense/Immunity protein/Toxin

- UNK Unclassified/unknown function

- (DIS Disease-related)

```
IPR000977|FUN GO:0003910, FUN GO:0005524, PRO GO:0006281, PRO GO:0006310, PRO GO:0006260
IPR001037|No GO term, viral integrase
IPR001044|FUN GO:0004519, FUN GO:0003697, PRO GO:0006289, COM GO:0005634
IPR001126|FUN GO:0003685, PRO GO:0006281
IPR001161|FUN GO:0004003, FUN GO:0003677, FUN GO:0005524, PRO GO:0006289, COM GO:0005634
IPR001162|FUN GO:0004518, FUN GO:0003677, PRO GO:0006289
IPR001238|FUN GO:0003685, FUN GO:0003697, FUN GO:0005524, PRO GO:0006281
IPR001357|No GO terms, unspecific
IPR001405|FUN GO:0003685, PRO GO:0006281
IPR001497|FUN GO:0003908, PRO GO:0006281
IPR001502|FUN GO:0004519, PRO GO:0006281
IPR001532|FUN GO:0004518, FUN GO:0003677, PRO GO:0006281
IPR001553|FUN GO:0008094, FUN GO:0003685, FUN GO:0003677, FUN GO:0005524, PRO GO:0006281,
IPR001668|FUN GO:0003677, PRO GO:0006310, COM GO:0005727
IPR001679|FUN GO:0003911, PRO GO:0006281, PRO GO:0006260
IPR001719|FUN GO:0004519, FUN GO:0003677, PRO GO:0006281
IPR001822|FUN GO:0000150, GO:0006310
IPR001943|FUN GO:0004518, FUN GO:0003677, PRO GO:0006289
IPR001945|FUN GO:0004003, FUN GO:0003677, FUN GO:0005524, PRO GO:0006289, COM GO:0005634
IPR002043|FUN GO:0004844, PRO GO:0006281
IPR002081|FUN GO:0003913
IPR002099|FUN GO:0003685, PRO GO:0006298
IPR002176|FUN GO:0004520, PRO GO:0006281, PRO GO:0006310
IPR002625|FUN GO:0003685, FUN GO:0003677, FUN GO:0005524, PRO GO:0006298
IPR002706|FUN GO:0003685, PRO GO:0000012, COM GO:0005634
IPR002863|FUN GO:0003685, FUN GO:0003677, FUN GO:0005524, PRO GO:0006298
IPR003011|FUN GO:0004527, FUN GO:0003685, PRO GO:0006281
IPR003021|FUN GO:0004527, FUN GO:0003685, PRO GO:0006281, COM GO:0005634
IPR003027|FUN GO:0008296, FUN GO:0003685, PRO GO:0006281, COM GO:0005634

DNA binding [GO:0003677]
single-stranded DNA binding [GO:0003697]
alkylbase DNA glycosidase [GO:0004036]
DNA repair protein [GO:0003685]
DNA repair enzyme [GO:0003686]
formamidopyrimidine-DNA glycosylase [GO:0003907]
deoxyribodipyrimidine photolyase [GO:0003904]
methylated-DNA--[protein]-cysteine S-methyltransferase [GO:0003908]
```



QuickGo - Gene Ontology Browser

GO Term GO:0003677

Parents	Current Term	Children
nucleic acid binding (GO:0003676)	DNA binding (GO:0003677) **SWISS-PROT Keywords** Homeobox (KeyID:346), DNA-binding (KeyID:208) **InterPro Entries** Retinoic acid receptor (IPR000003) Retinoic acid receptor (IPR000003) Acute myeloid leukemia 1 protein (AML 1)/Runt (IPR000040) Acute myeloid leukemia 1 protein (AML 1)/Runt (IPR000040) Universal stress protein (Usp) (IPR000041) Helix-turn-helix / lamda and other repressors (IPR000047) Restriction modification system, type I (IPR000055) High mobility group proteins HMG14 and HMG17 (IPR000079) High mobility group proteins HMG14 and HMG17 (IPR000079) High mobility group proteins HMG-I and HMG-Y (IPR000116) High mobility group proteins HMG-I and HMG-Y (IPR000116) Bacterial histone-like DNA-binding protein (IPR000119) Bacterial histone-like DNA-binding protein (IPR000119) Progesterone receptor (IPR000128) Progesterone receptor (IPR000128) High mobility group proteins HMG1 and HMG2 (IPR000135) High mobility group proteins HMG1 and HMG2 (IPR000135) Histone H3 (IPR000164) Histone-fold/TFIID-TAF/NF-Y domain (IPR000166) Hepadnaviral P protein N-terminal domain (IPR000201) BTB/POZ domain (IPR000210)	ribosomal DNA (rDNA) binding (GO:0000182) DNA secondary structure binding (GO:0000217) DNA helicase (GO:0003678) AT DNA binding (GO:0003680) bent DNA binding (GO:0003681) chromatin binding (GO:0003682) damaged DNA binding (GO:0003684) DNA repair protein (GO:0003685) DNA replication factor (GO:0003687) double-stranded DNA binding (GO:0003690) left-handed Z-DNA binding (GO:0003692) P-element binding (GO:0003693) plasmid-associated protein (GO:0003694)

InterPro

QuickGo - Gene Ontology Browser

GO Terms for SWISS-PROT Keywords

SWISS-PROT Keywords	GO Terms
Vision	vision (GO:0007601)

Search by Keyword

To search for GO terms associated with given SWISS-PROT keywords, select the keywords in the list below and hit the search button.

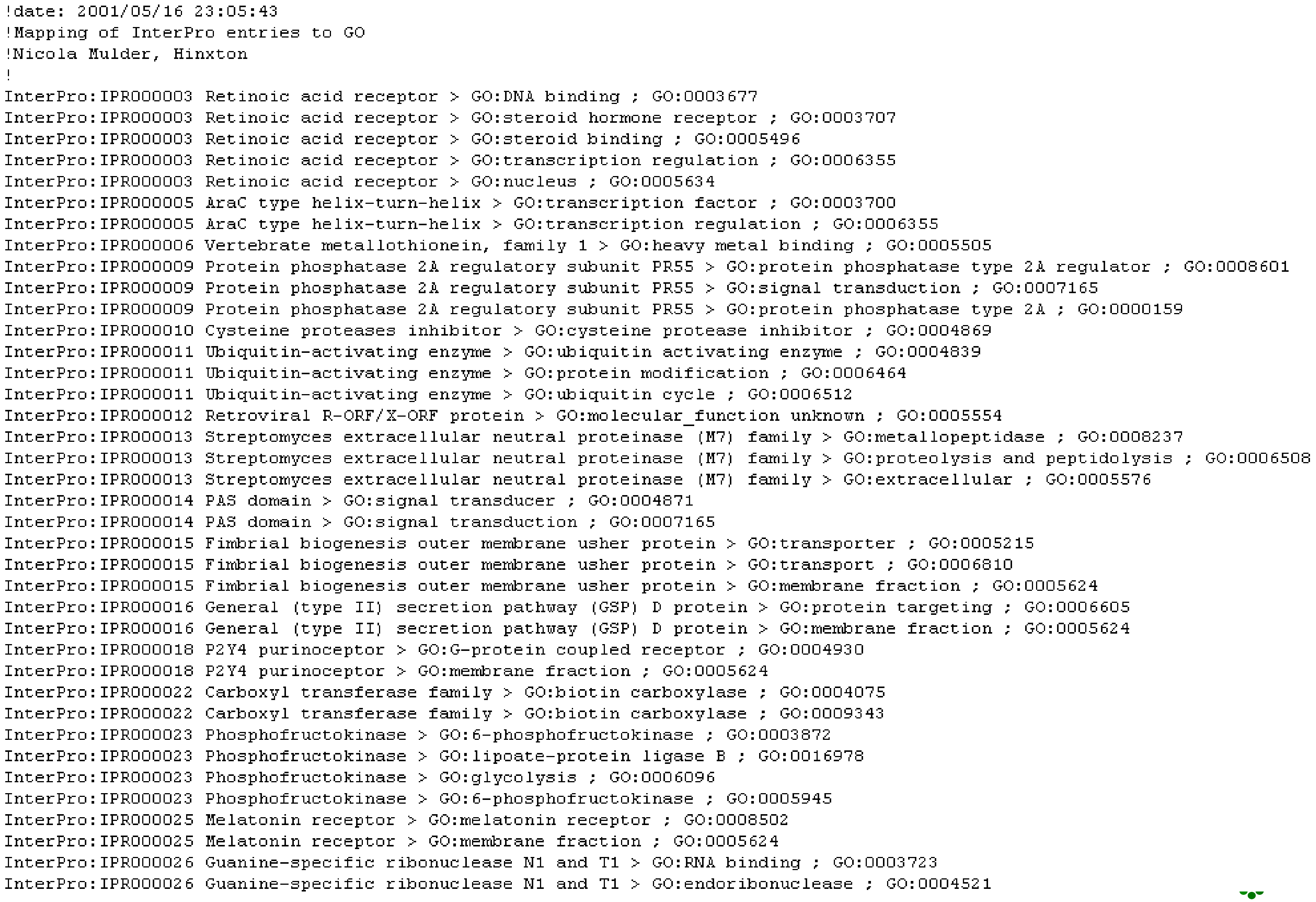
InterPro

QuickGo - Gene Ontology Browser

GO Term GO:0000182

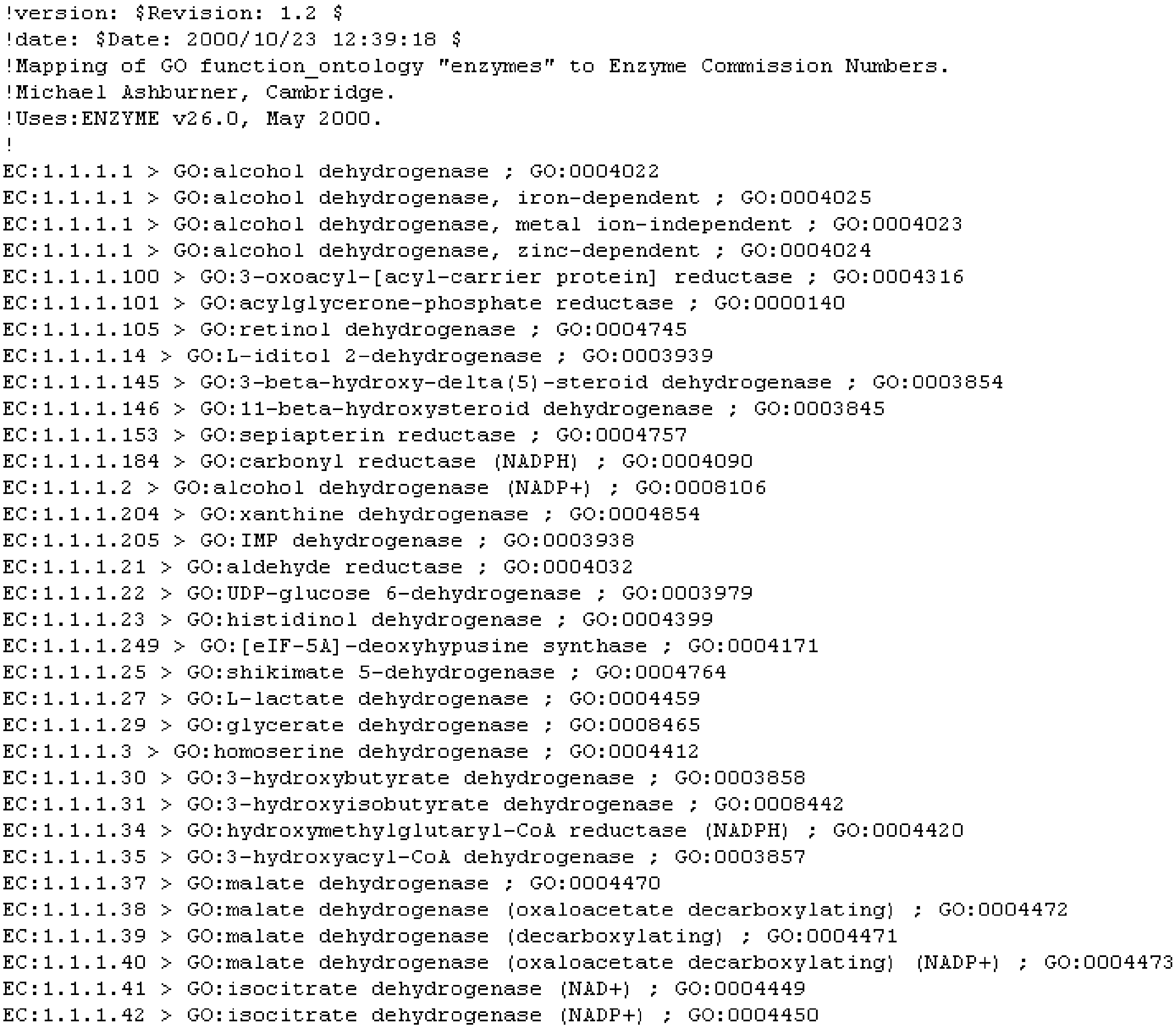
Parents	Current Term	Children
DNA binding (GO:0003677)	ribosomal DNA (rDNA) binding (GO:0000182) **No associated keywords in SWISS-PROT** **No associated entried in InterPro**	No Children

InterPro-to-GO



```
!date: 2001/05/16 23:05:43
!Mapping of InterPro entries to GO
!Nicola Mulder, Hinxton
!
InterPro:IPR000003 Retinoic acid receptor > GO:DNA binding ; GO:0003677
InterPro:IPR000003 Retinoic acid receptor > GO:steroid hormone receptor ; GO:0003707
InterPro:IPR000003 Retinoic acid receptor > GO:steroid binding ; GO:0005496
InterPro:IPR000003 Retinoic acid receptor > GO:transcription regulation ; GO:0006355
InterPro:IPR000003 Retinoic acid receptor > GO:nucleus ; GO:0005634
InterPro:IPR000005 AraC type helix-turn-helix > GO:transcription factor ; GO:0003700
InterPro:IPR000005 AraC type helix-turn-helix > GO:transcription regulation ; GO:0006355
InterPro:IPR000006 Vertebrate metallothionein, family 1 > GO:heavy metal binding ; GO:0005505
InterPro:IPR000009 Protein phosphatase 2A regulatory subunit PR55 > GO:protein phosphatase type 2A regulator ; GO:0008601
InterPro:IPR000009 Protein phosphatase 2A regulatory subunit PR55 > GO:signal transduction ; GO:0007165
InterPro:IPR000009 Protein phosphatase 2A regulatory subunit PR55 > GO:protein phosphatase type 2A ; GO:0000159
InterPro:IPR000010 Cysteine proteases inhibitor > GO:cysteine protease inhibitor ; GO:0004869
InterPro:IPR000011 Ubiquitin-activating enzyme > GO:ubiquitin activating enzyme ; GO:0004839
InterPro:IPR000011 Ubiquitin-activating enzyme > GO:protein modification ; GO:0006464
InterPro:IPR000011 Ubiquitin-activating enzyme > GO:ubiquitin cycle ; GO:0006512
InterPro:IPR000012 Retroviral R-ORF/X-ORF protein > GO:molecular_function unknown ; GO:0005554
InterPro:IPR000013 Streptomyces extracellular neutral proteinase (M7) family > GO:metallopeptidase ; GO:0008237
InterPro:IPR000013 Streptomyces extracellular neutral proteinase (M7) family > GO:proteolysis and peptidolysis ; GO:0006508
InterPro:IPR000013 Streptomyces extracellular neutral proteinase (M7) family > GO:extracellular ; GO:0005576
InterPro:IPR000014 PAS domain > GO:signal transducer ; GO:0004871
InterPro:IPR000014 PAS domain > GO:signal transduction ; GO:0007165
InterPro:IPR000015 Fimbrial biogenesis outer membrane usher protein > GO:transporter ; GO:0005215
InterPro:IPR000015 Fimbrial biogenesis outer membrane usher protein > GO:transport ; GO:0006810
InterPro:IPR000015 Fimbrial biogenesis outer membrane usher protein > GO:membrane fraction ; GO:0005624
InterPro:IPR000016 General (type II) secretion pathway (GSP) D protein > GO:protein targeting ; GO:0006605
InterPro:IPR000016 General (type II) secretion pathway (GSP) D protein > GO:membrane fraction ; GO:0005624
InterPro:IPR000018 P2Y4 purinoceptor > GO:G-protein coupled receptor ; GO:0004930
InterPro:IPR000018 P2Y4 purinoceptor > GO:membrane fraction ; GO:0005624
InterPro:IPR000022 Carboxyl transferase family > GO:biotin carboxylase ; GO:0004075
InterPro:IPR000022 Carboxyl transferase family > GO:biotin carboxylase ; GO:0009343
InterPro:IPR000023 Phosphofructokinase > GO:6-phosphofructokinase ; GO:0003872
InterPro:IPR000023 Phosphofructokinase > GO:lipoate-protein ligase B ; GO:0016978
InterPro:IPR000023 Phosphofructokinase > GO:glycolysis ; GO:0006096
InterPro:IPR000023 Phosphofructokinase > GO:6-phosphofructokinase ; GO:0005945
InterPro:IPR000025 Melatonin receptor > GO:melatonin receptor ; GO:0008502
InterPro:IPR000025 Melatonin receptor > GO:membrane fraction ; GO:0005624
InterPro:IPR000026 Guanine-specific ribonuclease N1 and T1 > GO:RNA binding ; GO:0003723
InterPro:IPR000026 Guanine-specific ribonuclease N1 and T1 > GO:endoribonuclease ; GO:0004521
```

EC number-to-GO



```
!version: $Revision: 1.2 $
!date: $Date: 2000/10/23 12:39:18 $
!Mapping of GO function_ontology "enzymes" to Enzyme Commission Numbers.
!Michael Ashburner, Cambridge.
!Uses:ENZYME v26.0, May 2000.
!
EC:1.1.1.1 > GO:alcohol dehydrogenase ; GO:0004022
EC:1.1.1.1 > GO:alcohol dehydrogenase, iron-dependent ; GO:0004025
EC:1.1.1.1 > GO:alcohol dehydrogenase, metal ion-independent ; GO:0004023
EC:1.1.1.1 > GO:alcohol dehydrogenase, zinc-dependent ; GO:0004024
EC:1.1.1.100 > GO:3-oxoacyl-[acyl-carrier protein] reductase ; GO:0004316
EC:1.1.1.101 > GO:acylglycerone-phosphate reductase ; GO:0000140
EC:1.1.1.105 > GO:retinol dehydrogenase ; GO:0004745
EC:1.1.1.14 > GO:L-iditol 2-dehydrogenase ; GO:0003939
EC:1.1.1.145 > GO:3-beta-hydroxy-delta(5)-steroid dehydrogenase ; GO:0003854
EC:1.1.1.146 > GO:11-beta-hydroxysteroid dehydrogenase ; GO:0003845
EC:1.1.1.153 > GO:sepiapterin reductase ; GO:0004757
EC:1.1.1.184 > GO:carbonyl reductase (NADPH) ; GO:0004090
EC:1.1.1.2 > GO:alcohol dehydrogenase (NADP+) ; GO:0008106
EC:1.1.1.204 > GO:xanthine dehydrogenase ; GO:0004854
EC:1.1.1.205 > GO:IMP dehydrogenase ; GO:0003938
EC:1.1.1.21 > GO:aldehyde reductase ; GO:0004032
EC:1.1.1.22 > GO:UDP-glucose 6-dehydrogenase ; GO:0003979
EC:1.1.1.23 > GO:histidinol dehydrogenase ; GO:0004399
EC:1.1.1.249 > GO:[eIF-5A]-deoxyhypusine synthase ; GO:0004171
EC:1.1.1.25 > GO:shikimate 5-dehydrogenase ; GO:0004764
EC:1.1.1.27 > GO:L-lactate dehydrogenase ; GO:0004459
EC:1.1.1.29 > GO:glycerate dehydrogenase ; GO:0008465
EC:1.1.1.3 > GO:homoserine dehydrogenase ; GO:0004412
EC:1.1.1.30 > GO:3-hydroxybutyrate dehydrogenase ; GO:0003858
EC:1.1.1.31 > GO:3-hydroxyisobutyrate dehydrogenase ; GO:0008442
EC:1.1.1.34 > GO:hydroxymethylglutaryl-CoA reductase (NADPH) ; GO:0004420
EC:1.1.1.35 > GO:3-hydroxyacyl-CoA dehydrogenase ; GO:0003857
EC:1.1.1.37 > GO:malate dehydrogenase ; GO:0004470
EC:1.1.1.38 > GO:malate dehydrogenase (oxaloacetate decarboxylating) ; GO:0004472
EC:1.1.1.39 > GO:malate dehydrogenase (decarboxylating) ; GO:0004471
EC:1.1.1.40 > GO:malate dehydrogenase (oxaloacetate decarboxylating) (NADP+) ; GO:0004473
EC:1.1.1.41 > GO:isocitrate dehydrogenase (NAD+) ; GO:0004449
EC:1.1.1.42 > GO:isocitrate dehydrogenase (NADP+) ; GO:0004450
```

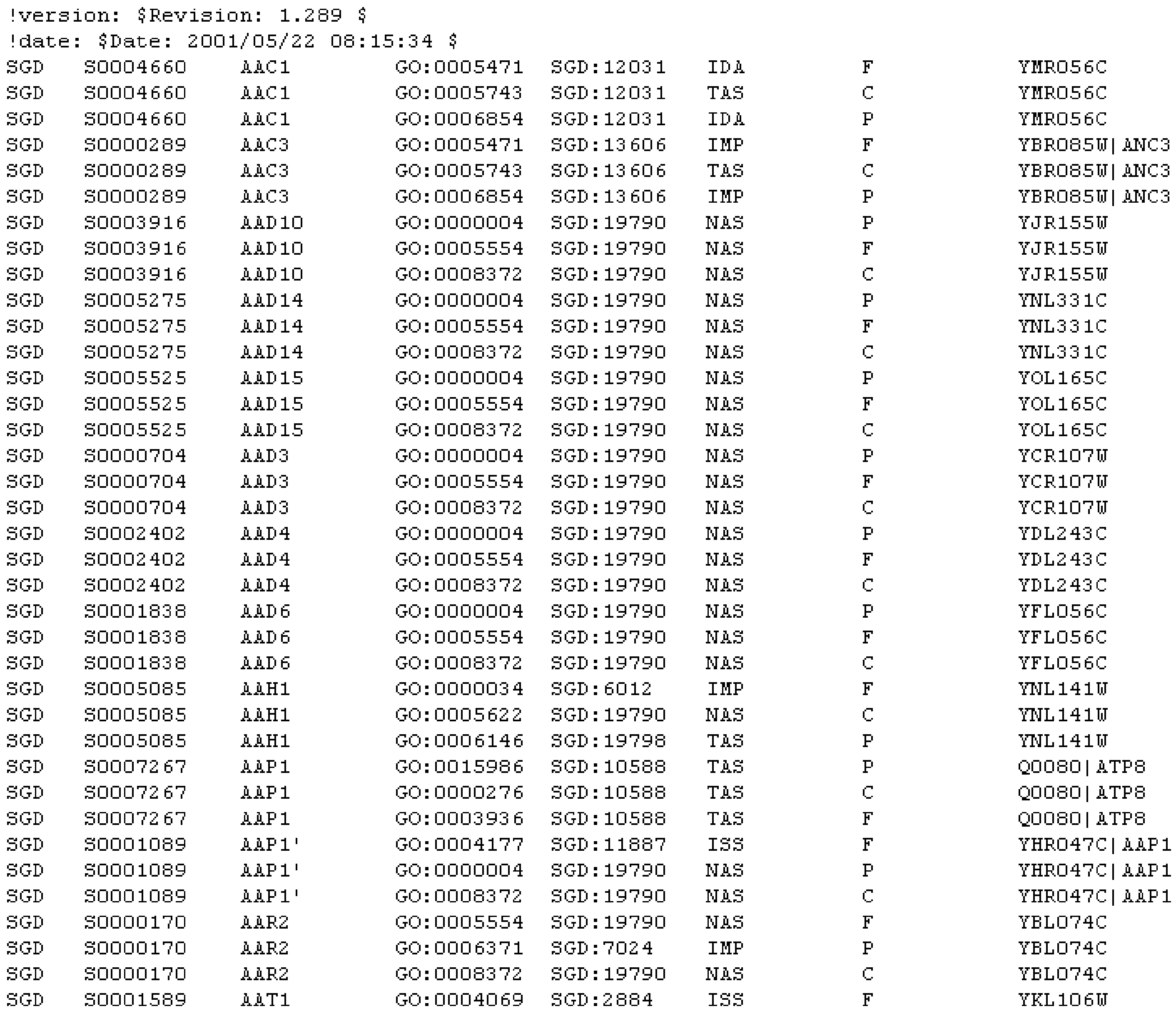
EMBL
European

SP keyword-to-GO

```
!version: $Revision: 1.8 $
!date: $Date: 2001/04/18 18:14:37 $
!Mapping of SwissProt KEYWORDS to GO terms.
!David Hill, Jackson Labs.
!Uses:SwissProt KEYWORDS.
!
SP_KW:Acetylcholine receptor inhibitor > GO:synaptic transmission, cholinergic ; GO:0007271
SP_KW:Actin-binding > GO:actin binding ; GO:0003779
SP_KW:Acute phase > GO:acute-phase response ; GO:0006953
SP_KW:ADP-ribosylation > GO:protein ADP-ribosylation ; GO:0006471
SP_KW:Albumin > GO:extracellular space ; GO:0005615
SP_KW:Alkylation > GO:protein modification ; GO:0006464
SP_KW:Allergen > GO:defense (immune) response ; GO:0006952
SP_KW:Allosteric enzyme > GO:enzyme ; GO:0003824
SP_KW:Alpha-amylase inhibitor > GO:enzyme inhibitor ; GO:0004866
SP_KW:Amino-acid biosynthesis > GO:amino-acid metabolism ; GO:0006520
SP_KW:Amino-acid transport > GO:amino-acid transport ; GO:0006865
SP_KW:Aminoacyl-tRNA synthetase > GO:amino-acid activation ; GO:0006418
SP_KW:Aminopeptidase > GO:aminopeptidase ; GO:0004177
SP_KW:Amyloid > GO:amyloid protein ; GO:0005208
SP_KW:Anion exchange > GO:inorganic anion exchanger ; GO:0005452
SP_KW:Annexin > GO:calcium binding ; GO:0005509
SP_KW:Anti-oncogene > GO:cell growth and maintenance ; GO:0008151
SP_KW:Antibiotic > GO:xenobiotic metabolism ; GO:0006807
SP_KW:Antibiotic biosynthesis > GO:xenobiotic metabolism ; GO:0006807
SP_KW:Antibiotic resistance > GO:xenobiotic metabolism ; GO:0006807
SP_KW:Antiviral > GO:defense (immune) response ; GO:0006952
SP_KW:Apoptosis > GO:apoptosis ; GO:0006915
SP_KW:Arabinose catabolism > GO:monosaccharide metabolism ; GO:0005996
SP_KW:Arginine biosynthesis > GO:arginine biosynthesis ; GO:0006526
SP_KW:Arginine metabolism > GO:arginine metabolism ; GO:0006525
SP_KW:Aromatic amino acid biosynthesis > GO:amino-acid metabolism ; GO:0006520
SP_KW:Aromatic hydrocarbons catabolism > GO:xenobiotic metabolism ; GO:0006807
SP_KW:Ascorbate biosynthesis > GO:water-soluble vitamin metabolism ; GO:0006767
SP_KW:Asparagine biosynthesis > GO:asparagine biosynthesis ; GO:0006529
SP_KW:Aspartic protease inhibitor > GO:proteinase inhibitor ; GO:0004866
SP_KW:Aspartyl protease > GO:aspartic-type endopeptidase ; GO:0004190
SP_KW:ATP synthesis > GO:ATP synthesis ; GO:0006758
```

EMBL
European

SGD-to-GO



```
!version: $Revision: 1.289 $
!date: $Date: 2001/05/22 08:15:34 $
SGD     S0004660     AAC1         GO:0005471  SGD:12031    IDA          F            YMR056C
SGD     S0004660     AAC1         GO:0005743  SGD:12031    TAS          C            YMR056C
SGD     S0004660     AAC1         GO:0006854  SGD:12031    IDA          P            YMR056C
SGD     S0000289     AAC3         GO:0005471  SGD:13606    IMP          F            YBR085W|ANC3
SGD     S0000289     AAC3         GO:0005743  SGD:13606    TAS          C            YBR085W|ANC3
SGD     S0000289     AAC3         GO:0006854  SGD:13606    IMP          P            YBR085W|ANC3
SGD     S0003916     AAD10        GO:0000004  SGD:19790    NAS          P            YJR155W
SGD     S0003916     AAD10        GO:0005554  SGD:19790    NAS          F            YJR155W
SGD     S0003916     AAD10        GO:0008372  SGD:19790    NAS          C            YJR155W
SGD     S0005275     AAD14        GO:0000004  SGD:19790    NAS          P            YNL331C
SGD     S0005275     AAD14        GO:0005554  SGD:19790    NAS          F            YNL331C
SGD     S0005275     AAD14        GO:0008372  SGD:19790    NAS          C            YNL331C
SGD     S0005525     AAD15        GO:0000004  SGD:19790    NAS          P            YOL165C
SGD     S0005525     AAD15        GO:0005554  SGD:19790    NAS          F            YOL165C
SGD     S0005525     AAD15        GO:0008372  SGD:19790    NAS          C            YOL165C
SGD     S0000704     AAD3         GO:0000004  SGD:19790    NAS          P            YCR107W
SGD     S0000704     AAD3         GO:0005554  SGD:19790    NAS          F            YCR107W
SGD     S0000704     AAD3         GO:0008372  SGD:19790    NAS          C            YCR107W
SGD     S0002402     AAD4         GO:0000004  SGD:19790    NAS          P            YDL243C
SGD     S0002402     AAD4         GO:0005554  SGD:19790    NAS          F            YDL243C
SGD     S0002402     AAD4         GO:0008372  SGD:19790    NAS          C            YDL243C
SGD     S0001838     AAD6         GO:0000004  SGD:19790    NAS          P            YFL056C
SGD     S0001838     AAD6         GO:0005554  SGD:19790    NAS          F            YFL056C
SGD     S0001838     AAD6         GO:0008372  SGD:19790    NAS          C            YFL056C
SGD     S0005085     AAH1         GO:0000034  SGD:6012     IMP          F            YNL141W
SGD     S0005085     AAH1         GO:0005622  SGD:19790    NAS          C            YNL141W
SGD     S0005085     AAH1         GO:0006146  SGD:19798    TAS          P            YNL141W
SGD     S0007267     AAP1         GO:0015986  SGD:10588    TAS          P            Q0080|ATP8
SGD     S0007267     AAP1         GO:0000276  SGD:10588    TAS          C            Q0080|ATP8
SGD     S0007267     AAP1         GO:0003936  SGD:10588    TAS          F            Q0080|ATP8
SGD     S0001089     AAP1'        GO:0004177  SGD:11887    ISS          F            YHR047C|AAP1
SGD     S0001089     AAP1'        GO:0000004  SGD:19790    NAS          P            YHR047C|AAP1
SGD     S0001089     AAP1'        GO:0008372  SGD:19790    NAS          C            YHR047C|AAP1
SGD     S0000170     AAR2         GO:0005554  SGD:19790    NAS          F            YBL074C
SGD     S0000170     AAR2         GO:0006371  SGD:7024     IMP          P            YBL074C
SGD     S0000170     AAR2         GO:0008372  SGD:19790    NAS          C            YBL074C
SGD     S0001589     AAT1         GO:0004069  SGD:2884     ISS          F            YKL106W
```

EMBL
European

Current status

Method	**mapping**	**proteins**
David KW	212775	123840
IPR true	383303	124840
EC no. DE	22567	16999
MGD	59734	4934
FB	5938	2439
SGD	6809	1281
Total	691126	191174 (49.1%)

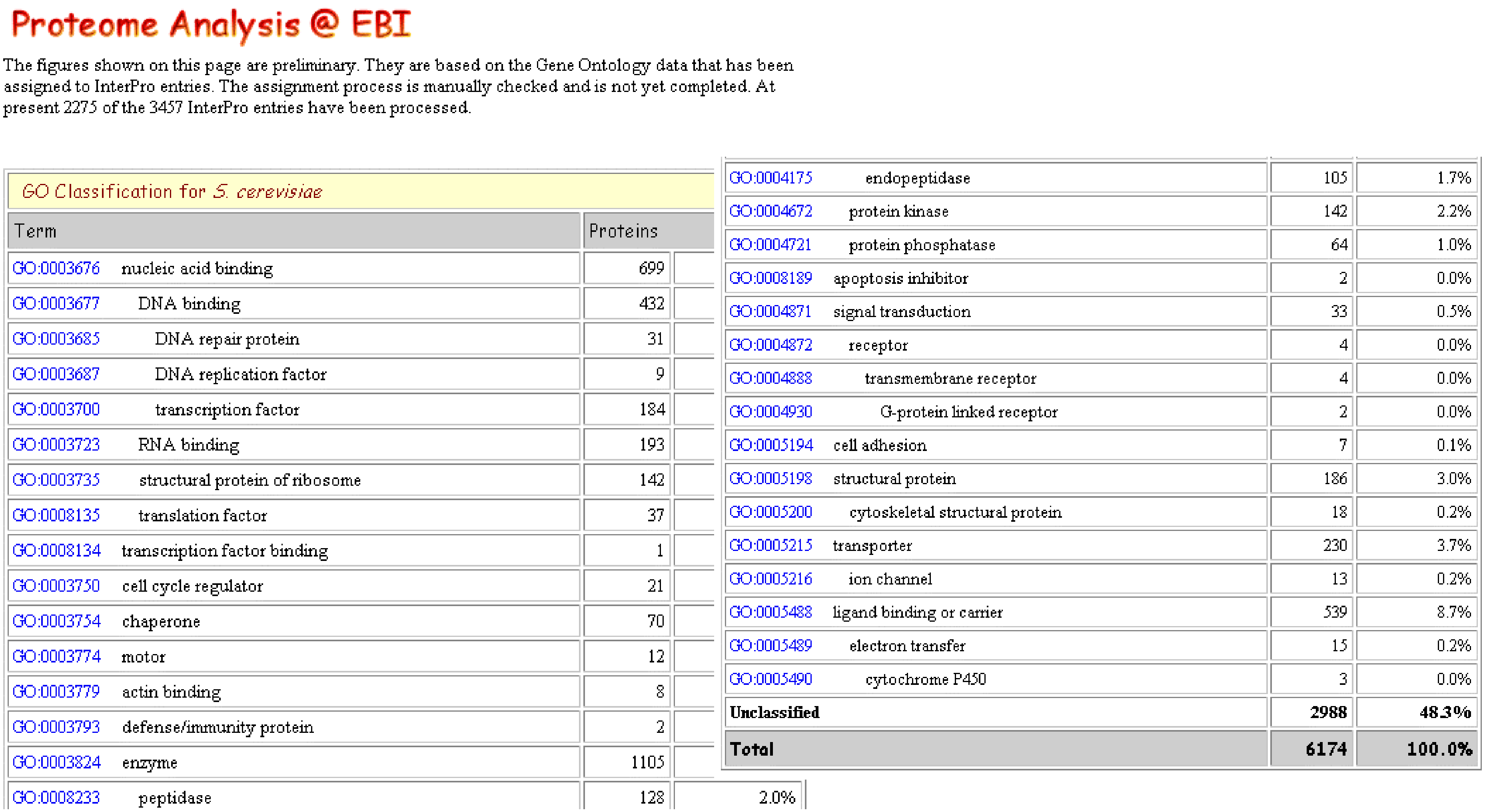
QUALITY OF ASSIGNMENTS

- Full assessment and comparison not yet done
- Manual annotation is best -especially if Medline number attached (biochemical evidence)
- InterPro useful, assuming protein hit is true and should hit all signatures in an entry
- EC numbers good, but need mapping of protein to these, so may be extra step
- SWISS-PROT keywords fine, but automatic and has some incorrect assignments

USING MAPPINGS TO ASSIGN GO TERMS

- Need compiled list of protein acc (all protein databases) and GO terms with evidence -link to BLAST search results
- Have GO term assignment linked to InterProScan, in the meantime, link hits to GO via mapping file
- Use EC number mappings if your protein hits an enzyme

Applications

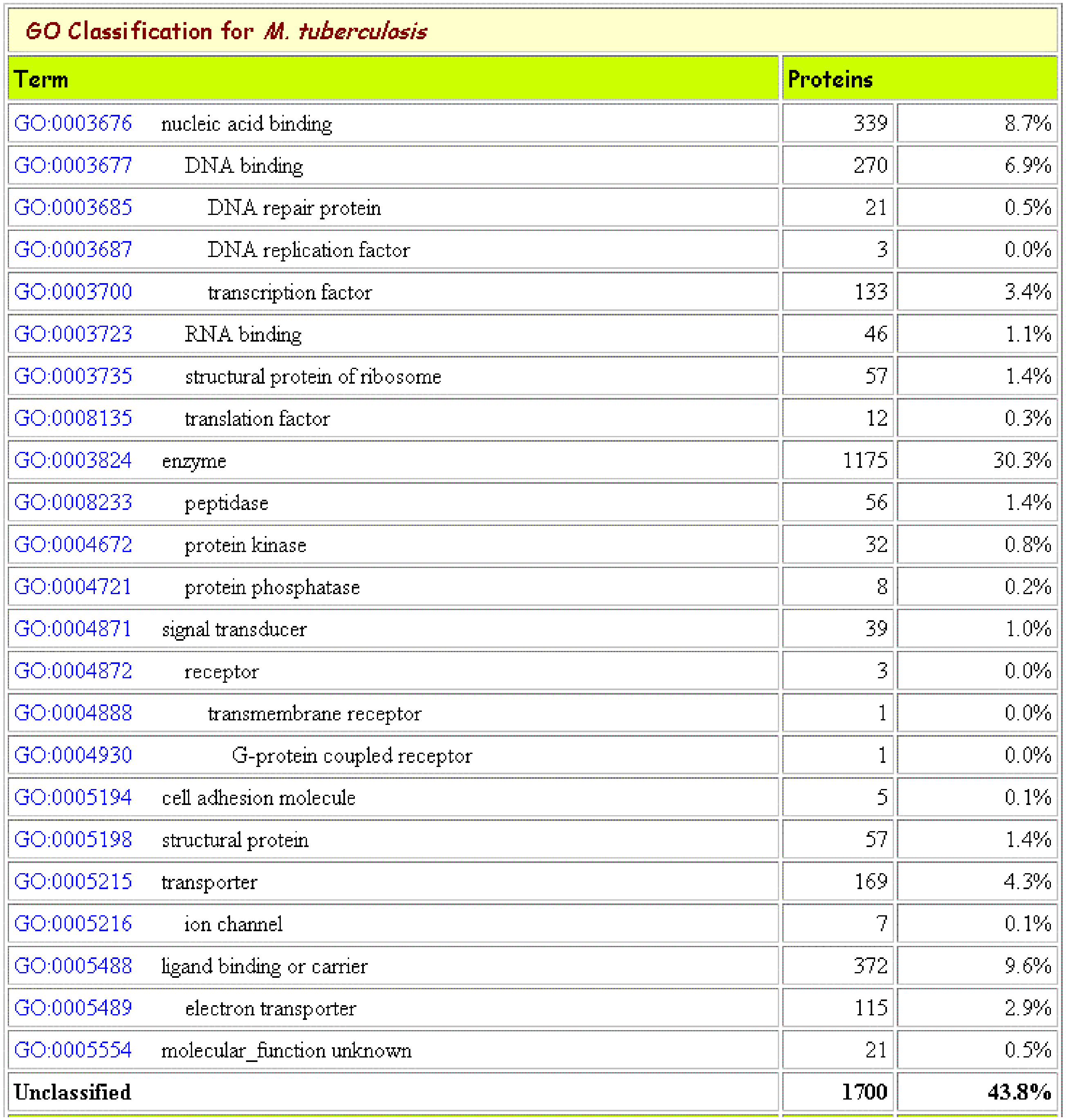


Proteome Analysis @ EBI

The figures shown on this page are preliminary. They are based on the Gene Ontology data that has been assigned to InterPro entries. The assignment process is manually checked and is not yet completed. At present 2275 of the 3457 InterPro entries have been processed.

GO Classification for *S. cerevisiae*

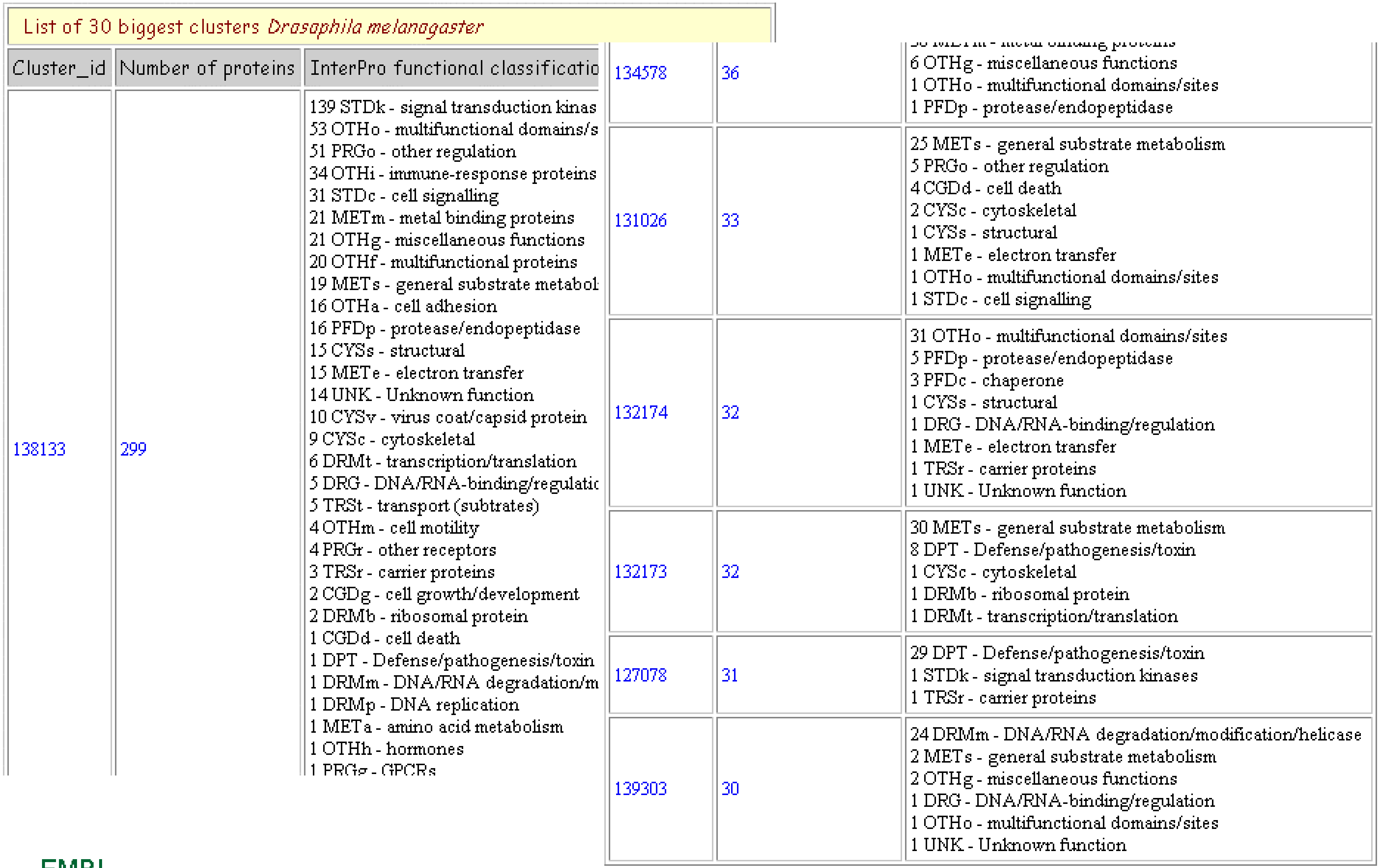
Term		Proteins	
GO:0003676	nucleic acid binding	699	
GO:0003677	DNA binding	432	
GO:0003685	DNA repair protein	31	
GO:0003687	DNA replication factor	9	
GO:0003700	transcription factor	184	
GO:0003723	RNA binding	193	
GO:0003735	structural protein of ribosome	142	
GO:0008135	translation factor	37	
GO:0008134	transcription factor binding	1	
GO:0003750	cell cycle regulator	21	
GO:0003754	chaperone	70	
GO:0003774	motor	12	
GO:0003779	actin binding	8	
GO:0003793	defense/immunity protein	2	
GO:0003824	enzyme	1105	
GO:0008233	peptidase	128	2.0%
GO:0004175	endopeptidase	105	1.7%
GO:0004672	protein kinase	142	2.2%
GO:0004721	protein phosphatase	64	1.0%
GO:0008189	apoptosis inhibitor	2	0.0%
GO:0004871	signal transduction	33	0.5%
GO:0004872	receptor	4	0.0%
GO:0004888	transmembrane receptor	4	0.0%
GO:0004930	G-protein linked receptor	2	0.0%
GO:0005194	cell adhesion	7	0.1%
GO:0005198	structural protein	186	3.0%
GO:0005200	cytoskeletal structural protein	18	0.2%
GO:0005215	transporter	230	3.7%
GO:0005216	ion channel	13	0.2%
GO:0005488	ligand binding or carrier	539	8.7%
GO:0005489	electron transfer	15	0.2%
GO:0005490	cytochrome P450	3	0.0%
Unclassified		**2988**	**48.3%**
Total		**6174**	**100.0%**



GO Classification for *M. tuberculasis*

Term		Proteins	
GO:0003676	nucleic acid binding	339	8.7%
GO:0003677	DNA binding	270	6.9%
GO:0003685	DNA repair protein	21	0.5%
GO:0003687	DNA replication factor	3	0.0%
GO:0003700	transcription factor	133	3.4%
GO:0003723	RNA binding	46	1.1%
GO:0003735	structural protein of ribosome	57	1.4%
GO:0008135	translation factor	12	0.3%
GO:0003824	enzyme	1175	30.3%
GO:0008233	peptidase	56	1.4%
GO:0004672	protein kinase	32	0.8%
GO:0004721	protein phosphatase	8	0.2%
GO:0004871	signal transducer	39	1.0%
GO:0004872	receptor	3	0.0%
GO:0004888	transmembrane receptor	1	0.0%
GO:0004930	G-protein coupled receptor	1	0.0%
GO:0005194	cell adhesion molecule	5	0.1%
GO:0005198	structural protein	57	1.4%
GO:0005215	transporter	169	4.3%
GO:0005216	ion channel	7	0.1%
GO:0005488	ligand binding or carrier	372	9.6%
GO:0005489	electron transporter	115	2.9%
GO:0005554	molecular_function unknown	21	0.5%
Unclassified		**1700**	**43.8%**

Proteome Analysis @ EBI

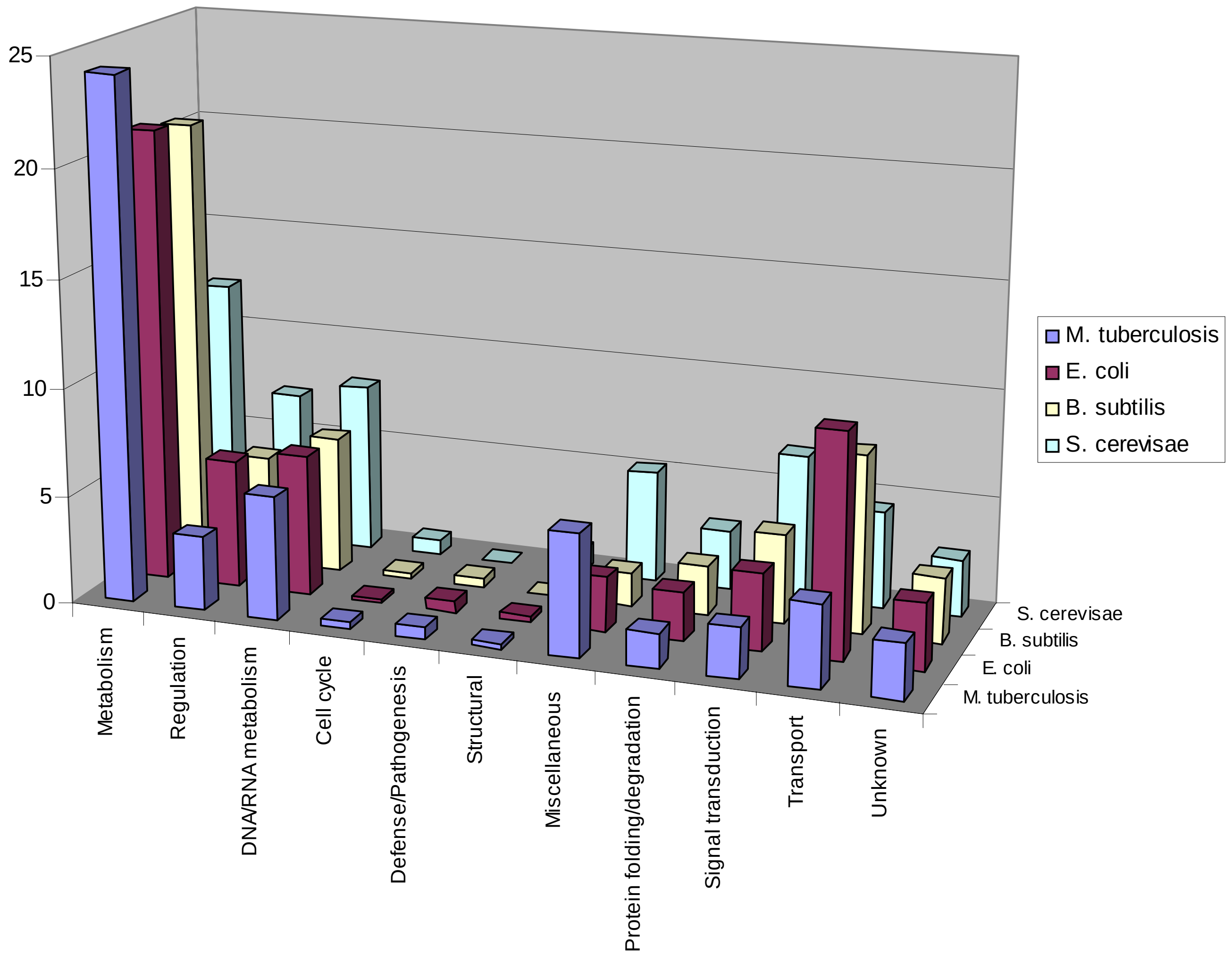


List of 30 biggest clusters *Drasaphila melanagaster*

Cluster_id	Number of proteins	InterPro functional classificatio
138133	299	139 STDk - signal transduction kinas 53 OTHo - multifunctional domains/s 51 PRGo - other regulation 34 OTHi - immune-response proteins 31 STDc - cell signalling 21 METm - metal binding proteins 21 OTHg - miscellaneous functions 20 OTHf - multifunctional proteins 19 METs - general substrate metabol 16 OTHa - cell adhesion 16 PFDp - protease/endopeptidase 15 CYSs - structural 15 METe - electron transfer 14 UNK - Unknown function 10 CYSv - virus coat/capsid protein 9 CYSc - cytoskeletal 6 DRMt - transcription/translation 5 DRG - DNA/RNA-binding/regulatic 5 TRSt - transport (subtrates) 4 OTHm - cell motility 4 PRGr - other receptors 3 TRSr - carrier proteins 2 CGDg - cell growth/development 2 DRMb - ribosomal protein 1 CGDd - cell death 1 DPT - Defense/pathogenesis/toxin 1 DRMm - DNA/RNA degradation/m 1 DRMp - DNA replication 1 METa - amino acid metabolism 1 OTHh - hormones 1 PRGg - GPCRs
134578	36	METm - metal binding proteins 6 OTHg - miscellaneous functions 1 OTHo - multifunctional domains/sites 1 PFDp - protease/endopeptidase
131026	33	25 METs - general substrate metabolism 5 PRGo - other regulation 4 CGDd - cell death 2 CYSc - cytoskeletal 1 CYSs - structural 1 METe - electron transfer 1 OTHo - multifunctional domains/sites 1 STDc - cell signalling
132174	32	31 OTHo - multifunctional domains/sites 5 PFDp - protease/endopeptidase 3 PFDc - chaperone 1 CYSs - structural 1 DRG - DNA/RNA-binding/regulation 1 METe - electron transfer 1 TRSr - carrier proteins 1 UNK - Unknown function
132173	32	30 METs - general substrate metabolism 8 DPT - Defense/pathogenesis/toxin 1 CYSc - cytoskeletal 1 DRMb - ribosomal protein 1 DRMt - transcription/translation
127078	31	29 DPT - Defense/pathogenesis/toxin 1 STDk - signal transduction kinases 1 TRSr - carrier proteins
139303	30	24 DRMm - DNA/RNA degradation/modification/helicase 2 METs - general substrate metabolism 2 OTHg - miscellaneous functions 1 DRG - DNA/RNA-binding/regulation 1 OTHo - multifunctional domains/sites 1 UNK - Unknown function

EMBL
European Bioinformatics Institute

Distribution of protein functions



GO-SLIM

A list of processes covering most general cellular processes

- UNKNOWN
- METABOLISM
- CELL GROWTH & METABOLISM
- PROTEIN METABOLISM AND MODIFICATION
- DEVELOPMENTAL PROCESSES
- NUCLEIC ACID METABOLISM
- CELL COMMUNICATION
- TRANSCRIPTION
- SIGNAL TRANSDUCTION
- TRANSPORT
- INTRACELLULAR PROTEIN TRAFFICKING
- STRESS RESPONSE
- DEATH

URLs

- http://www.informatics.jax.org/go/
- http://genome-www.stanford.edu/GO/
- http://www.ebi.ac.uk/interpro/QuickGo
- go@geneontology.org
- gofriends-request@geneontology.org : subscribe gofriends [your username]@[your mail server]