BIOINFORMATICS APPLICATIONS NOTE Vol. 25 no. 2 2009, pages 288–289
doi:10.1093/bioinformatics/btn615

Databases and ontologies

AmiGO: online access to ontology and annotation data

Seth Carbon[1,*,†], Amelia Ireland[2,†], Christopher J. Mungall[1], ShengQiang Shu[3], Brad Marshall[1], Suzanna Lewis[1], the AmiGO Hub[‡] and the Web Presence Working Group[‡]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [2]GO Editorial Office, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK and [3]Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA

Received on October 2, 2008; revised on November 20, 2008; accepted on November 21, 2008
Advance Access publication November 25, 2008
Associate Editor: Martin Bishop

ABSTRACT

AmiGO is a web application that allows users to query, browse and visualize ontologies and related gene product annotation (association) data. AmiGO can be used online at the Gene Ontology (GO) website to access the data provided by the GO Consortium[1]; it can also be downloaded and installed to browse local ontologies and annotations.[2] AmiGO is free open source software developed and maintained by the GO Consortium.

Availability: http://amigo.geneontology.org
Download: http:// sourceforge.net/projects/geneontology/
Contact: sjcarbon@berkeleybop.org

1 INTRODUCTION

The Gene Ontology [GO (http://www.geneontology.org); Gene Ontology Consortium, 2000] project develops structured controlled vocabularies, or ontologies, to describe fundamental characteristics of genes and their products ('gene products') in a species-independent manner. Members of the GO Consortium submit annotations made using these ontologies to the GO database for integration and dissemination. To give broad access to this resource, the GO Consortium has developed AmiGO, a web-based application that allows users to search, sort, analyze, visualize and download data of interest. Along with providing details of the ontologies, gene products and annotations, AmiGO features a BLAST (Altschul *et al*., 1990) search, Term Enrichment and GO Slimmer tools, the GO Online SQL Environment and a user help guide.

2 CORE AMIGO FUNCTIONALITY

2.1 Term information

Term data are split into two types: details of the term itself and information on the gene products annotated to that term or its children. Clicking on a term name in AmiGO will bring up the term details page, which displays ontology information, including the term accession, synonyms, definition, comments and mappings to similar or equivalent terms in other databases. The term lineage is displayed in a tree browser, allowing users to view and explore the areas of the ontology graph surrounding the term.

The term annotations page, which can be accessed via the associations link next to a term name, lists gene products annotated to (i.e. associated with) the GO term and its children in the ontology. Each annotation includes the evidence used to make the assertion, commonly a paper reference and a code representing the evidence type (e.g. IMP, Inferred from Mutant Phenotype; TAS, Traceable Author Statement; IC, Inferred by Curator). Annotations can be filtered by the properties of the gene products, such as species, source database or gene product type (e.g. protein, gene, transcript); by evidence code; and by whether the gene products are annotated directly to the term or to a child of the term. Links are provided to download the data in RDF-XML or the GO Consortium's gene association file format. Gene products can be selected and used for other operations, such as retrieval of annotation data, or as input for the Term Enrichment or GO Slimmer tools.

2.2 Gene product information

As with term-related data, AmiGO provides gene product information and a list of the terms to which a gene product is annotated. Gene product details are supplied by the source database and include the symbol, full name, synonyms and database ID; if a sequence is available, this will be displayed along with information parsed from the FASTA headers. The terms to which the gene product is annotated are shown on a separate page; these data can be filtered to display only annotations with certain evidence codes and/or terms from chosen ontologies. Download links are provided and terms can be selected for further actions, such as viewing in the context of the GO tree or for use with the GO Slimmer.

2.3 Searching

Every page in AmiGO offers a simple search box through which users can query the GO database for GO terms or gene products. AmiGO returns search results ordered by how closely the result matches the original query; results can also be sorted by other parameters, such as accession in term searches, or gene symbol when querying for gene products. Users can fine-tune their results

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
‡Gene Ontology Consortium, go-help@geneontology.org.
[1]Current GO database annotation information can be found at: http://www.geneontology.org/GO.database.db_spp.shtml
[2]Detailed installation information can be found at: http://wiki.geneontology.org/index.php/AmiGO_Manual:_Installation

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

by setting term- or gene product-related filters, or confine the search to certain parameters of the term or gene product. Each search result has links to both the details and annotations of the term or gene product.

2.4 Browsing

An alternative to searching for terms and gene products is to explore the GO ontology using the browse option. Starting at the top of the ontology tree, users can expand term nodes to view their children and navigate down through the hierarchy. If a term appears in multiple locations in the tree, all occurrences are displayed, thus giving users an idea of how terms are related. The display shows the number of gene products annotated to each term, including annotations to its descendants; filters can be used to generate the number for selected databases or species. There are several options for viewing and saving tree information. Users can visualize the ontology as rendered by GraphViz (http://www.graphviz.org), or download it as a GraphViz dot file, as RDF-XML, or in OBO (http://www.geneontology.org/GO.format.obo-1_2.shtml) format. Additionally, AmiGO shows the distribution of annotations to a term and its children in both graphical and tabular formats via the bar chart viewer.

3 OTHER FUNCTIONALITY

3.1 BLAST

AmiGO offers a BLAST search against the gene products in the GO database. Users can enter a UniProtKB ID, upload a sequence or click on a BLAST link within AmiGO to perform the search. The results are returned in a summary table, with links to the detail and annotation pages for each gene product, along with the raw BLAST output. Gene products discovered by BLAST can be selected for further analysis.

3.2 Term enrichment

The GO Term Enrichment tool determines whether the observed level of annotation for a group of genes is significant in the context of a background set, typically all the genes in the genome. This can be useful for discovering relationships between genes, e.g. when analyzing gene clusters from microarray expression data. The Term Enrichment tool makes use of the GO-TermFinder Perl module (Boyle *et al*., 2004), and offers a number of input, filtering and output options, including a visualization component to display significantly overrepresented terms in the context of the GO tree.

3.3 GO Slimmer

The function of the GO Slimmer tool is to remap granular, specific annotations up to a user-specified set of high-level terms. This subset of terms, referred to as a 'GO slim', provides a useful overview of a dataset and facilitates the reporting and analysis of large result sets, such as GO annotations to a genome or microarray expression data. Like the Term Enrichment tool, GO Slimmer also offers a wide array of input, filtering and output options.

3.4 GO Online SQL Environment (GOOSE)

This web utility gives users the ability to perform SQL queries directly on the GO database. GOOSE includes many sample queries to aid novice users and allows results to be retrieved as a web page or as tab-delimited text.

4 ARCHITECTURE

AmiGO is a server-based Perl application that retrieves information from a database and processes it to produce web pages. It employs the go-perl (http://search.cpan.org/∼cmungall/go-perl) and go-db-perl (http://search.cpan.org/∼cmungall/go-db-perl) libraries to access and structure database information; this gives AmiGO the flexibility to be used with a number of different ontology formats and database schemas. To ensure its utility as a web-based application, developers emphasize making AmiGO standards compliant and accessible for all users. AmiGO is open source software and is freely available to distribute and modify.

5 FUTURE DEVELOPMENT

AmiGO is under active development to keep pace with growth in the volume and variety of data available. The evolution of AmiGO is directed by the changing needs of the user base, new initiatives within the GO project and the possibilities presented by new web technologies. Current areas of focus include integration of GO Reference Genome (http://www.geneontology.org/GO.refgenome.shtml) data, development of new ontology visualizations and exploration of community-driven annotation through wiki-style software. User feedback is welcomed by the AmiGO developers, and can be submitted via the AmiGO feature request (http://sourceforge.net/tracker/?group_id=36855&atid=494390) or bug (http://sourceforge.net/tracker/?group_id=36855&atid=908269) trackers.

ACKNOWLEDGEMENTS

The AmiGO Hub Group is composed of A.I., Jane Lomax (EBI, Hinxton, UK); S.C., Chris Mungall (BBOP, LBNL, Berkeley, CA, USA); Benjamin Hitz, Rama Balakrishnan (SGD, Stanford University, Stanford, CA, USA); Mary Dolan (MGI, The Jackson Laboratory, Bar Harbor, ME, USA). The Web Presence Working Group is composed of the AmiGO Hub Group; Valerie Wood (GeneDB, Wellcome Trust Sanger Institute, Hinxton, UK); Eurie Hong (SGD, Stanford University, Stanford, CA, USA); Pascale Gaudet (dictyBase, Northwestern University, Chicago, IL, USA). The authors gratefully acknowledge the support and advice of our colleagues in the GO Consortium.

Funding: National Human Genome Research Institute (P41 grant 5P41HG002273-08 to Gene Ontology Consortium).

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al*. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Boyle,E.I. *et al*. (2004) GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, **20**, 3710–3715.

Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genet*., **25**, 25–29.

The submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC02-05CH11231. Accordingly, the U.S. Government retains a nonexclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.