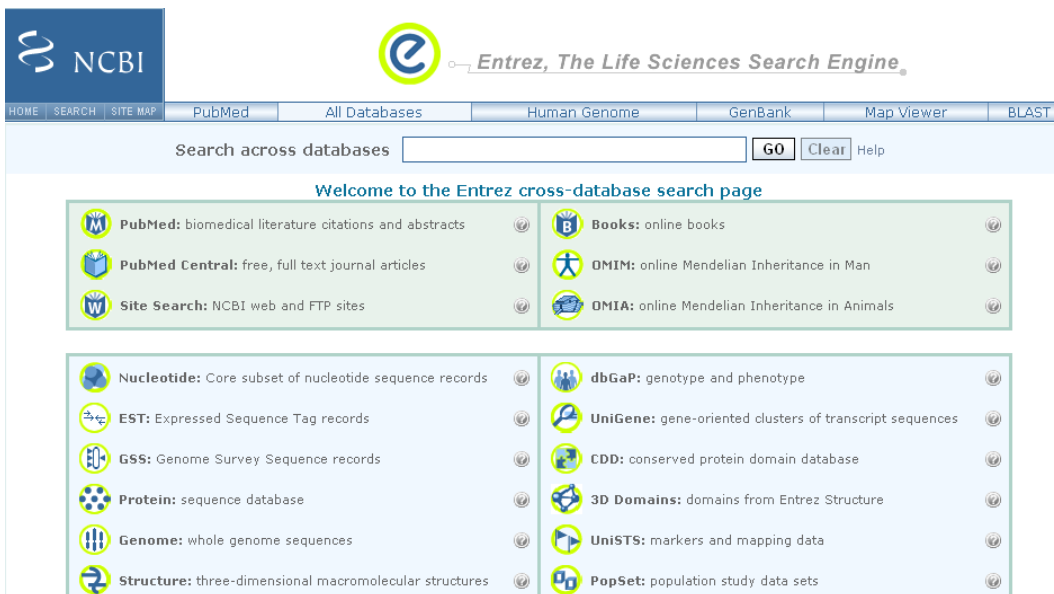


Info Hubs Central Access Points to Bioinformatics Databases



















1. [National Center For Biotechnology Information - Entrez](#)
2. [European Bioinformatics Institute](#)
3. [Nucleic Acids Research](#) - The Molecular Biology Database Collection
4. [Gene Ontology](#)

National Center for Biotechnology Information - Entrez <http://www.ncbi.nlm.nih.gov/Entrez>

The National Center for Biotechnology Information (NCBI) is a division of the National Institutes of Health. It develops and provides information resources for the international study of molecular biology and genetics data. The majority of NCBI resources are integrated within a universal search interface called Entrez so that the data search and retrieval system is consistent for all of the different data domains. Each data domain can be searched individually or they can be all simultaneously searched using the Entrez Global Search. The Global Search Portal allows the user to retrieve an approximate idea of how many records are found in each database for a topic of interest.



The screenshot shows the NCBI Entrez search engine homepage. At the top left is the NCBI logo. To its right is the Entrez logo and the tagline "Entrez, The Life Sciences Search Engine". Below this is a navigation bar with links for HOME, SEARCH, SITE MAP, PubMed, All Databases, Human Genome, GenBank, Map Viewer, and BLAST. A search bar is present with a "GO" button and a "Clear" button. The main content area is titled "Welcome to the Entrez cross-database search page" and displays a grid of database icons and descriptions:

 PubMed: biomedical literature citations and abstracts	 Books: online books
 PubMed Central: free, full text journal articles	 OMIM: online Mendelian Inheritance in Man
 Site Search: NCBI web and FTP sites	 OMIA: online Mendelian Inheritance in Animals
 Nucleotide: Core subset of nucleotide sequence records	 dbGaP: genotype and phenotype
 EST: Expressed Sequence Tag records	 UniGene: gene-oriented clusters of transcript sequences
 GSS: Genome Survey Sequence records	 CDD: conserved protein domain database
 Protein: sequence database	 3D Domains: domains from Entrez Structure
 Genome: whole genome sequences	 UniSTS: markers and mapping data
 Structure: three-dimensional macromolecular structures	 PopSet: population study data sets

Also note that each of the databases has a “limits” tab, but just as the field tags vary for specific databases, so do the limit options. The limits can be extremely helpful in restrict search results to certain types of information. Simply remember that they vary greatly between databases.

Additional Features

1. Clipboard: The clipboard allows users to save citations for up to eight hours and to email, print, or send groups of citations to a MyNCBI account. Select citations by clicking the small check box to the left of each citation. Then use the “send to” drop-down menu to send the items to the clipboard. Click on the "Clipboard" tab to see the citations saved on the clipboard.

2. MyNCBI: This tool that allows users to organize citations, save searches, and store automatic search filters. Send citations from the clipboard or the main search page by using the “send to” drop-down menu and selecting “collections.” This will open up MyNCBI. Save searches by using the “save search” link to the right of the search box in each database. Filters are a way of automatically grouping search results. Click on the “search filters” link in MyNCBI to set up filters.

European Bioinformatics Institute

<http://www.ebi.ac.uk/>

The European Bioinformatics Institute is the European counterpart to the U.S. National Center for Biotechnology Information (NCBI). The EBI maintains and makes available databases and information services relevant to molecular biology, as well as carrying out research in bioinformatics and computational molecular biology.

Key Databases:

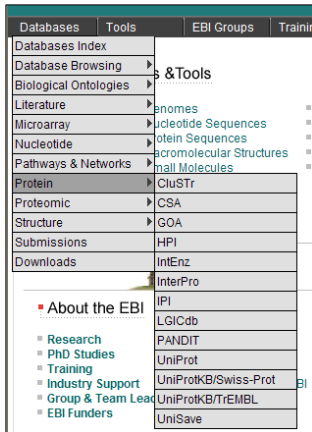
EBI provides access to a wide spectrum of databases and software tools focusing on biologic ontologies, proteomics, molecular structure, and microarrays. A selection of prominent EBI resources is listed below.

- **UniProtKB/Swiss-Prot database** – an annotated protein sequence database
- **EMBL nucleotide database** - a nucleotide sequence archive that shares data daily with the NCBI's GenBank and Japan's DDBJ.
- **InterPro** – contains searchable information regarding protein families, domains, and sites that help users identify relationships between proteins and determine their function.

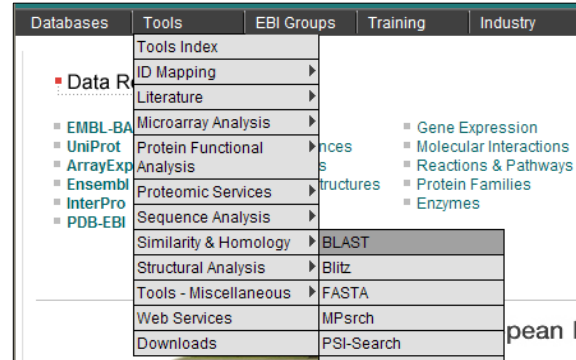
The screenshot shows the EMBL-EBI website homepage. At the top, there is a navigation bar with the EMBL-EBI logo, a search bar, and a 'GO' button. Below the navigation bar, there are several sections: 'Data Resources & Tools' with links to EMBL BANK, UniProt, ArrayExpress, Ensembl, InterPro, and POB-EBI; 'About the EBI' with links to Research, PhD Studies, Training, Industry Support, Group & Team Leaders, EBI Funders, User Support, EBI Mission, People, Events at the EBI, and How to Find us; 'Training' with a link to 'Hands-on training at EBI (EMBRACE) Workshop: Understanding protein structures'; and 'Research Highlights' with a link to 'New bacteria survive environmental change'. The website is designed with a clean, professional layout and a color scheme of green and white.

- **ArrayExpress** – a database of gene expression data
- **IntAct** – a free, open source database system that contains analysis tools for protein interaction data.
- **ClustalW program** - free software program used to create multiple sequence alignments.

Browsing the EBI:



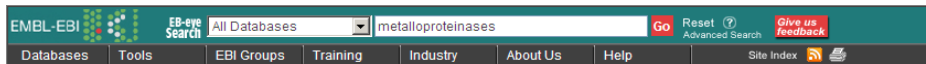
Users can browse the EBI by looking through various dropdown lists of databases organized by broad categories such as protein, microarray, structure and other areas.



The Tools dropdown also allows users to browse available software applications for genetic data analysis.

Searching EBI:

Perform a broadcast search of all EBI content and databases by using the “EB-eye search box”.



Results will be displayed in categories that can be expanded to reveal results from individual databases.

Search options available:

- View results summary of database subsets
- Refine search strategy with new keywords

Results summary	
Protein Sequences	246
> PRIDE	0
> UniProtKB	188
> UniRef	58
> UniRef100	24
> UniRef50	21
> UniRef30	13
> UniParc	0

Searching Hints for EBI Global Search:

- Global Search allows use of **Boolean Operators** [i.e. and, or, not, +, -], **nesting techniques** [(metalloproteinase and mouse) not human], and phrase searching [“tay sachs”]
- **Term modifiers** (? and *) allow users to truncate the end of search terms or replace 1 letter in a search term (i.e. b?nd = bond, bind..., gluta* = glutacin, glutamate...)
- **Domain searching** allows the user to limit search to an individual database [i.e. desc:(+insulin +receptor) in UniProt]
- **Field searching** not available in the Global search option, but can be used in individual databases if field options are known. Examples of common field options are: description:glutathione, name:clip, id:farc_valco, and authors:birney.

Nucleic Acids Research: The Molecular Biology Database Collection

2008 Issue: <http://www.oxfordjournals.org/nar/database/search/>

What it is:

- Every January, Nucleic Acids Research (NAR) releases the “database issue.”
- The current and past issues and additional content describe >1000 databases in molecular biology.
- NAR is open access and freely available without a subscription. The reviewed databases are also freely available resources, although some require registration or have some additional content in subscription-only areas.
- Each entry provides a description of the purpose and content of a molecular biology database and a link to access the database.

How to Get There:

- Although you can get to Nucleic Acids Research through the Digital Library, the quickest way is to go straight to <http://www.oxfordjournals.org/nar/database/search/>
- If you do go through the Digital Library: search for NAR and click on resulting link -> click current Database Issue link in right sidebar -> scroll down to the first of the Articles and click Database Summaries -> click on Search Summary Papers.

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > [Database Summary Paper Search](#)

Search NAR Database Summary Papers

Search Term

Search In Title Author Affiliation Paper References All

[▶ Compilation Paper](#)
[▶ Category List](#)
[▶ Alphabetical List](#)
[▶ Category/Paper List](#)
[▶ Search Summary Papers](#)

[▶ Compilation Paper](#)
[▶ Category List](#)
[▶ Alphabetical List](#)
[▶ Category/Paper List](#)
[▶ Search Summary Papers](#)

Basic Search:

- Enter keywords in the Search Term box; there is no advanced search option.
- You likely won't need to change the selection from "All."
- Your results will include the database titles, one-line descriptions, and links to both more detailed summaries and the databases themselves.

Browsing:

- You can browse alphabetically or by category; the category browse is particularly useful for viewing a list of databases within a particular topic.
- Click on any category to see more specific categories or a list of databases fitting the category.

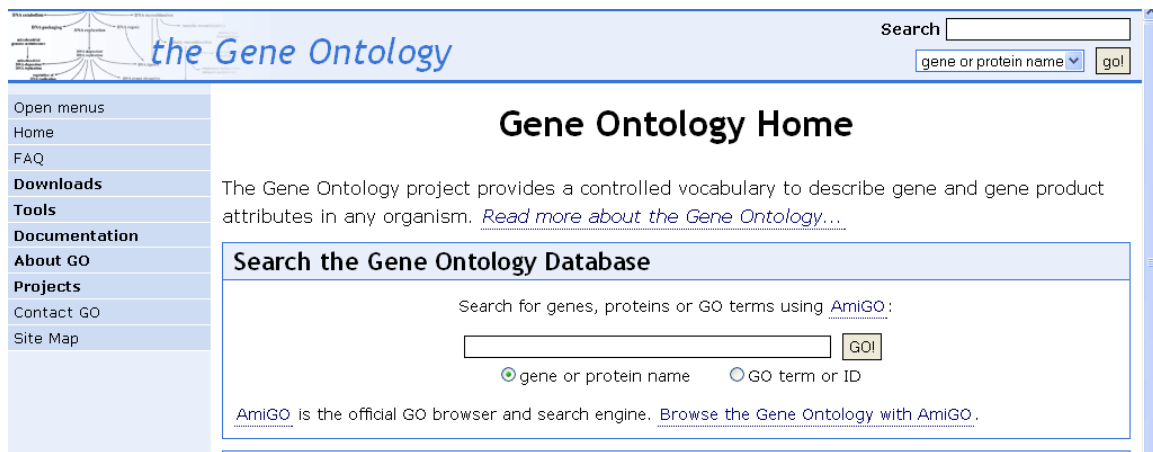
NAR Database Categories List

[Nucleotide Sequence Databases](#)
[RNA sequence databases](#)
[Protein sequence databases](#)
[Structure Databases](#)
[Genomics Databases \(non-vertebrate\)](#)
[Metabolic and Signaling Pathways](#)
[Human and other Vertebrate Genomes](#)
[Human Genes and Diseases](#)
[Microarray Data and other Gene Expression Databases](#)
[Proteomics Resources](#)
[Other Molecular Biology Databases](#)
[Organelle databases](#)
[Plant databases](#)
[Immunological databases](#)

Gene Ontology Consortium

<http://www.geneontology.com>

The goal of the Gene Ontology™ Consortium is to produce a dynamic controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing. Multiple genome projects and organizations are collaborative partners. This vocabulary constitutes a standard for gene annotation and allows searching for genes by protein function, biochemical process, or cellular location. It consists of a hierarchical tree structure of terms. The primary search engine is AmiGO which allows text searching by gene or protein name or GO term or identifier number.



the Gene Ontology

Search

gene or protein name

Gene Ontology Home

The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism. [Read more about the Gene Ontology...](#)

Search the Gene Ontology Database

Search for genes, proteins or GO terms using [AmiGO](#):

gene or protein name GO term or ID

[AmiGO](#) is the official GO browser and search engine. [Browse the Gene Ontology with AmiGO](#).

AmiGO also provides a browse function that can be used to view the tree structure of the terminology in a graphical form. Use the "+" and "-" symbols to expand or collapse sections of the tree. The filter options allow the user to view only certain sections of the tree or to limit by contributing data source or species.

Tree Browser

The screenshot shows the Tree Browser interface with the following components:

- Filter tree view** (expanded):
 - Filter by ontology:** All, biological process, cellular component, molecular function.
 - Filter Gene Product Counts:**
 - Data source:** All, CGD, dictyBase, EcoCyc.
 - Species:** All, Anaplasma phagocy..., Arabidopsis thaliana, Bacillus anthraci...
 - View Options:** Tree view, Full, Compact. Buttons: Set filters, Remove all filters.
- Tree Structure:**
 - all : all [251544 gene products]
 - GO:0008150 : biological_process [165780 gene products]
 - GO:0022610 : biological adhesion [1569 gene products]
 - GO:0051825 : adhesion to other organism during symbiotic interaction [94 gene products]**
 - GO:0044406 : adhesion to host [94 gene products]
 - GO:0051856 : adhesion to symbiont [0 gene products]
 - GO:0007155 : cell adhesion [1527 gene products]
 - GO:0022608 : multicellular organism adhesion [14 gene products]
 - GO:0065007 : biological regulation [31913 gene products]
 - GO:0001906 : cell killing [223 gene products]
 - GO:0009987 : cellular process [78835 gene products]
 - GO:0032502 : developmental process [20636 gene products]
- Actions...** (right sidebar): Last action: Opened, GO:0051825, Graphical View, Permalink, Download..., OBO, RDF-XML, GraphViz dot.

Clicking on the "# of gene products" link by any GO term will display a list of genes to which that term has been applied by the collaborating organizations. This is not a complete list of all gene to which the term *may* be applied, as new genes are being regularly discovered and the lists are limited to the information supplied by GO's collaborators.

One of the most valuable uses of GO terms is as search terms for the many outside databases which are now incorporating GO terms into their gene annotation records. For example, the NCBI's Entrez Gene database can be field searched with GO terms in the following format: "term"[GO]. This allows the user to search, for example, for all genes that encode proteins that perform a specific function, belong to a specific biological process, or are located in a specific cellular location.

Getting Additional Help:

- Request training:
 - Jennifer Lyon at Jennifer.lyon@vanderbilt.edu or x6-5787
 - Rachel Walden at rachel.walden@vanderbilt.edu or x6-3116
- View other available training opportunities (including PubMed, NCBI databases, ENSEMBL, and other genetics-focused resources):
<http://www.mc.vanderbilt.edu/biolib/services/training.html>
- Get assistance with your research-related questions via the BioSearchDoc request tool: <http://www.mc.vanderbilt.edu/vumcdiglib/biosearchdoc/index.html> (VUnet ID & password required)