

# Query tools at NCBI and EBI

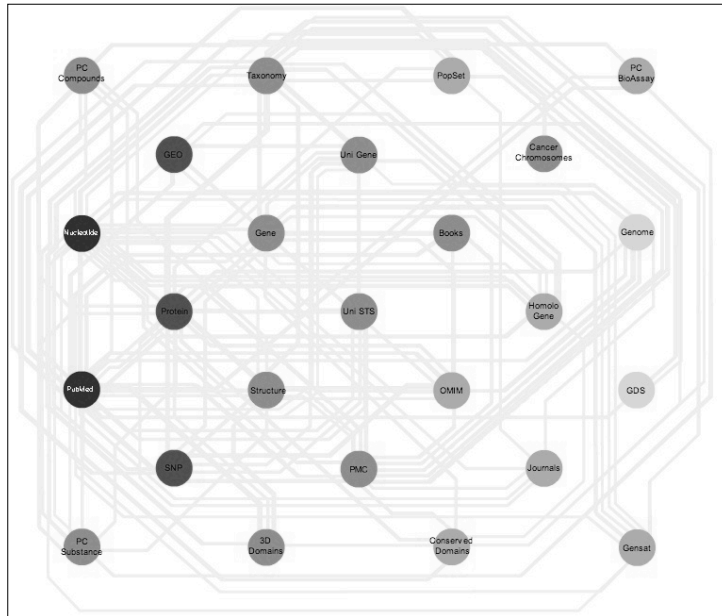
## Entrez; EB-eye

In the broadest sense, a DATABASE is anything that stores data.



A database is generally a collection of RECORDS, each of which contains one or more FIELD.

Aside from the ability to store data, a database also provides a way for other computer programs to quickly retrieve and update desired pieces of data.



## NCBI - Entrez

Entrez is an integrated search and retrieval system that provides access to all databases simultaneously with a single query string and user interface.

Nearly all search boxes that appear on the NCBI site access the Entrez system.

One of the most powerful features of the Entrez system is the integration of the data so that relationships between records can easily be explored.

There are two types of relationships established in Entrez:

**HARD LINKS**

**NEIGHBORS**

**HARD LINKS:** reciprocal links between databases established from the records themselves.

For example, molecular records such as sequences or structures are linked to the PubMed citations where these data were reported. The literature citation has the reciprocal links back to the molecular databases. PubMed citations are linked to the full-text article in PubMed Central. Databases such as UniGene or Gene that gather together many records from different sources have links back to their source records.

**NEIGHBORS:** are established automatically for many Entrez databases by computing on the data in the records.

For example, related sequences are identified through similarity searches with the BLAST algorithm; related PubMed citations are determined by an algorithm that compares information rich words and phrases in the abstracts.

The Entrez search interface features powerful options for constructing precise searches and managing results.

Options include:

popular configurable Limits and preset filters to help focus on specific kinds of results

Advanced Search interface that facilitates constructing more sophisticated queries.

Other useful Entrez features include Search History with access to recent results and a Clipboard where search results can be saved temporarily.

Not only this interconnectivity enhances navigation and allows search results to be quickly focused or expanded.

## NCBI - Entrez

### GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (GAPD)

The protein catalyzes an important energy-yielding step in carbohydrate metabolism, the reversible oxidative phosphorylation of glyceraldehyde-3-phosphate in the presence of inorganic phosphate and nicotinamide adenine dinucleotide (NAD).

<http://www.ncbi.nlm.nih.gov>

rete

## **Boolean operators**

**AND:** To 'AND' two search terms together instructs Entrez to find all documents that contain **BOTH** terms

**OR:** To 'OR' two search terms together instructs Entrez to find all documents that contain **EITHER** term.

**NOT:** To 'NOT' two search terms together instructs Entrez to find all documents that contain search term 1 **BUT NOT** search term 2.

Entrez processes all Boolean operators in a left-to-right sequence.

The order in which Entrez processes a search statement can be changed by enclosing individual concepts in parentheses.

The terms inside the parentheses are processed first as a unit and then incorporated into the overall strategy.

# EXERCISE

Search for **GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (GAPD)** entries in:

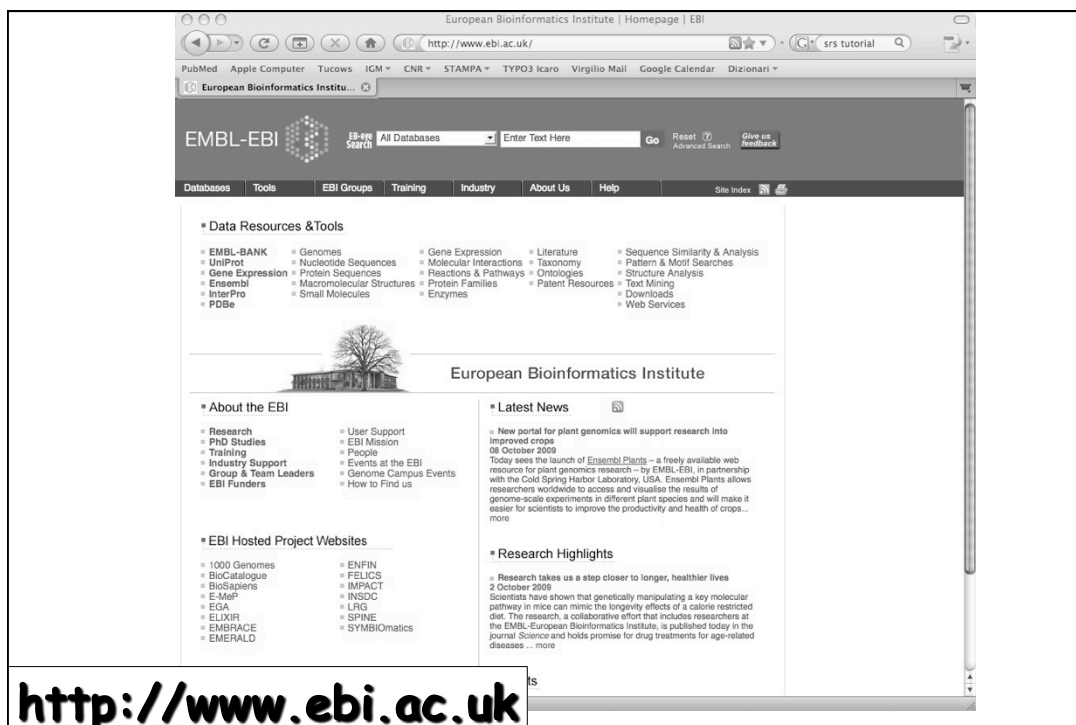
1. ALL ORGANISMS
2. PRIMATES
3. HUMANS
4. PRIMATES EXCLUDING HOMO SAPIENS

# EXERCISE

SEARCH FOR ARTICLES ON

1. *gapd AND Drosophila melanogaster*
2. *gapd NOT human*
3. *gapd AND (human OR Drosophila melanogaster)*

# EBI - EB-eye

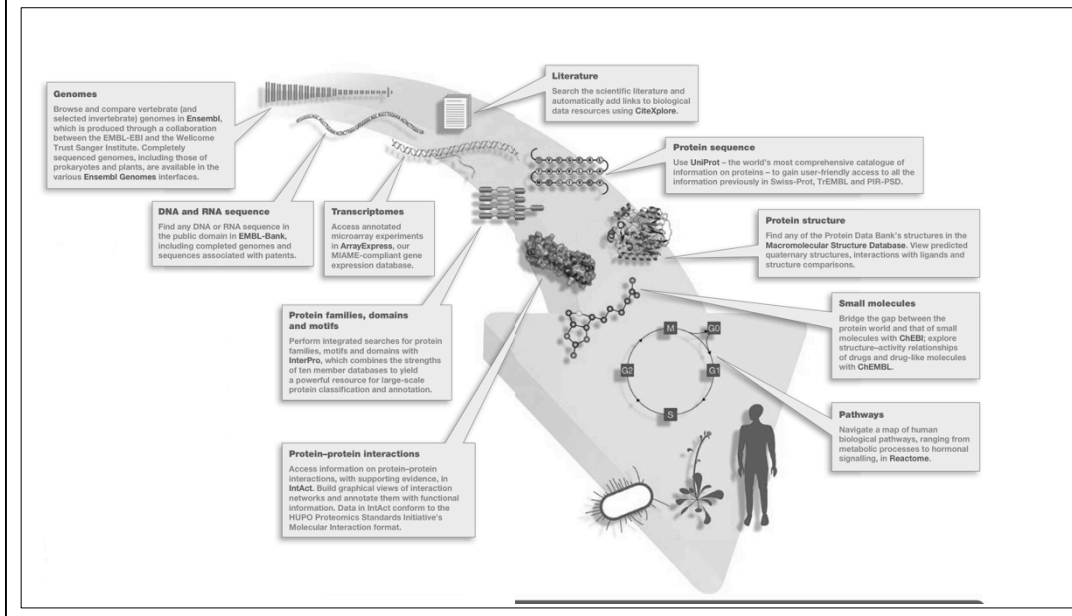


The screenshot shows the homepage of the European Bioinformatics Institute (EBI). At the top, there is a navigation bar with links to PubMed, Apple Computer, Tucows, ICM, CNR, STAMPA, TYPO3 Icaro, Virgilio Mail, Google Calendar, and Dizionari. Below this is the EMBL-EBI logo and a search bar with the text "EB-eye Search" and "All Databases". The main content area is divided into several sections:

- Data Resources & Tools**: A grid of links including EMBL-BANK, UniProt, Gene Expression, Ensembl, InterPro, PDBe, Genomes, Nucleotide Sequences, Protein Sequences, Macromolecular Structures, Small Molecules, Gene Expression, Molecular Interactions, Reactions & Pathways, Protein Families, Enzymes, Literature, Taxonomy, Ontologies, Patent Resources, Sequence Similarity & Analysis, Pattern & Motif Searches, Structure Analysis, Text Mining, Downloads, and Web Services.
- About the EBI**: Links to Research, PhD Studies, Training, Industry Support, Group & Team Leaders, EBI Funders, User Support, EBI Mission, People, Events at the EBI, Genome Campus Events, and How to Find us.
- EBI Hosted Project Websites**: Links to 1000 Genomes, BioCatalogue, BioSisipiens, E-MaP, EGA, ELIXIR, EMBRACE, EMERALD, ENFIN, FELICS, IMPACT, INSDC, LRJ, SPINE, and SYMBIOmatics.
- Latest News**: A news item dated 08 October 2009 about the launch of *Ensembl Plants*, a freely available web resource for plant genomics research.
- Research Highlights**: A news item dated 2 October 2009 about research showing that genetically manipulating a key molecular pathway in rice can mimic the longevity effects of a calorie restricted diet.

At the bottom of the page, the URL <http://www.ebi.ac.uk> is displayed in a large, bold font.

# EBI European Biotechnology Institute



## EB-eye Search

EB-eye allows the user to search globally across all EBI databases (or individually in selected resources by using an Advance search) in a very simple and semantically uniform manner.

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

Link rete



Search for *gapd* in All the EBI

Expand all Collapse all

Genomes	111	Molecular Interactions	13
Nucleotide Sequences	1,008	Reactions & Pathways	2
Protein Sequences	104	Protein Families	0
Macromolecular Structures	1	Enzymes	0
Small molecules	0	Literature	163
Gene Expression	8	Ontologies	0
		EBI Web Site	0

Refine your search:

Search for *gapd* in All the EBI

with the following keywords  Refine

Terms of Use EBI Funding Contact EBI © European Bioinformatics Institute 2010. EBI is an Outstation of the European Molecular Biology Laboratory.

**AND** - (default) meaning that term1 AND term2 must exist in the searched documents.

**OR** - meaning that either term1 OR term2 must exist.

**NOT** - meaning that term must not be present in any of the displayed documents (e.g. excludes documents containing the term1).

+ '+term1' - The document must contain the term1.

- '-term1' - Prohibit operator: The document must not contain term1.

'\*' - as in 'gluta\*' (glutacin, glutamate, glutamic, etc.)

'?' - as in 'b?nd' (bind, bond, band, etc.)

EMBL-EBI Search All Databases  Go [Reset](#) [Advanced Search](#) [Give us feedback](#)

Databases Tools EBI Groups Training Industry About Us Help Site Index

**Search for *gapd* AND *homo sapiens* in All the EBI**

<input type="button" value="Genomes"/> 18	<input type="button" value="Molecular Interactions"/>
<input type="button" value="Nucleotide Sequences"/> 73	<input type="button" value="Reactions &amp; Pathways"/>
<input type="button" value="Protein Sequences"/> 4	<input type="button" value="Protein Families"/>
<input type="button" value="Macromolecular Structures"/> 1	<input type="button" value="Enzymes"/>
<input type="button" value="Small molecules"/> 0	<input type="button" value="Literature"/>
<input type="button" value="Gene Expression"/> 1	<input type="button" value="Ontologies"/>
	<input type="button" value="EBI Web Site"/>

**Refine your search:**

Search for *gapd* AND *homo sapiens* in All the EBI

with the following keywords

EMBL-EBI Search All Databases  Go [Reset](#) [Advanced Search](#) [Give us feedback](#)

Databases Tools EBI Groups Training Industry About Us Help Site Index

**Search for *gapd* AND *homo sapiens* in All the EBI**

<input type="button" value="Genomes"/> 18	<input type="button" value="Molecular Interactions"/> 0
<ul style="list-style-type: none"> <li>DGVa <input type="text" value="0"/></li> <li><i>Database of Genomic Variants Archive</i></li> <li>EGA <input type="text" value="0"/></li> <li><i>The European Genome-phenome Archive</i></li> <li>HGNC <input type="text" value="0"/></li> <li><i>HUGO Gene Nomenclature Committee</i></li> <li>LRG <input type="text" value="0"/></li> <li><i>A stable genomic reference framework for describing sequence variations</i></li> <li>Ensembl <input type="text" value="18"/></li> <li><i>Selected eukaryotic genomes</i></li> <li>Ensembl Genomes <input type="text" value="0"/></li> <li><i>Non-vertebrate genomes of high scientific interest</i></li> <li>Integr8 <input type="text" value="0"/></li> <li><i>Completed genomes and proteomes</i></li> </ul>	<ul style="list-style-type: none"> <li><input type="button" value="Reactions &amp; Pathways"/> 0</li> <li><input type="button" value="Protein Families"/> 0</li> <li><input type="button" value="Enzymes"/> 0</li> <li><input type="button" value="Literature"/> 0</li> <li><input type="button" value="Ontologies"/> 0</li> <li><input type="button" value="EBI Web Site"/> 0</li> </ul>
<input type="button" value="Nucleotide Sequences"/> 73	
<ul style="list-style-type: none"> <li>INSDC Project <input type="text" value="0"/></li> <li><i>INSDC Project records from the European Nucleotide Archive</i></li> <li>ASTD <input type="text" value="0"/></li> <li><i>Database of alternative splice events and transcripts of genes from human, mouse and rat.</i></li> <li>EMBL-Bank <input type="text" value="44"/></li> <li><i>Europe's primary nucleotide sequence resource</i></li> <li>EMBL-Bank (Coding Sequence) <input type="text" value="29"/></li> <li><i>Coding Sequences in EMBL-Bank</i></li> <li>Sequence Read Archive (SRA) <input type="text" value="0"/></li> <li><i>Next generation sequencing raw data repository from the European Nucleotide Archive</i></li> </ul>	
<input type="button" value="Protein Sequences"/> 4	

# **Query tools**

## **SRS**

**SRS6 at EBI (<http://srs.ebi.ac.uk/>)**

## **Sequence Retrieval System**

**SRS is a multi-database retrieval system developed by Thure Etzold and commercialized by LION bioscience AG. It is free for academic use.**

**SRS is available on many sites and can be implemented on a server in our own labs.**

**SRS interface is more powerful but also more complex than Entrez.**

## What can we do with SRS?

- perform simple and complex queries across one or several databases
- view your results in different formats
- create your own views for your results
- save results to file or to a browser
- launch applications on results
- link results to different databases

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

**SRS**

Start a Permanent Project

**Tips**

- Want to know more about using SRS?
- go to the [Help Center](#) for online searchable help.
- look in our [SRS@EBI FAQ](#) for answers to commonly asked questions
- Linking to SRS?
- Please read our [Linking to SRS](#) guide for important information regarding linking to our SRS server.
- Public SRS servers worldwide

**Quick Text Search** Search Tips

Find : Nucleotides matching : Enter Text Here

Searches Databases: EMBL Nucleotides

**News and Announcements** Search Tips

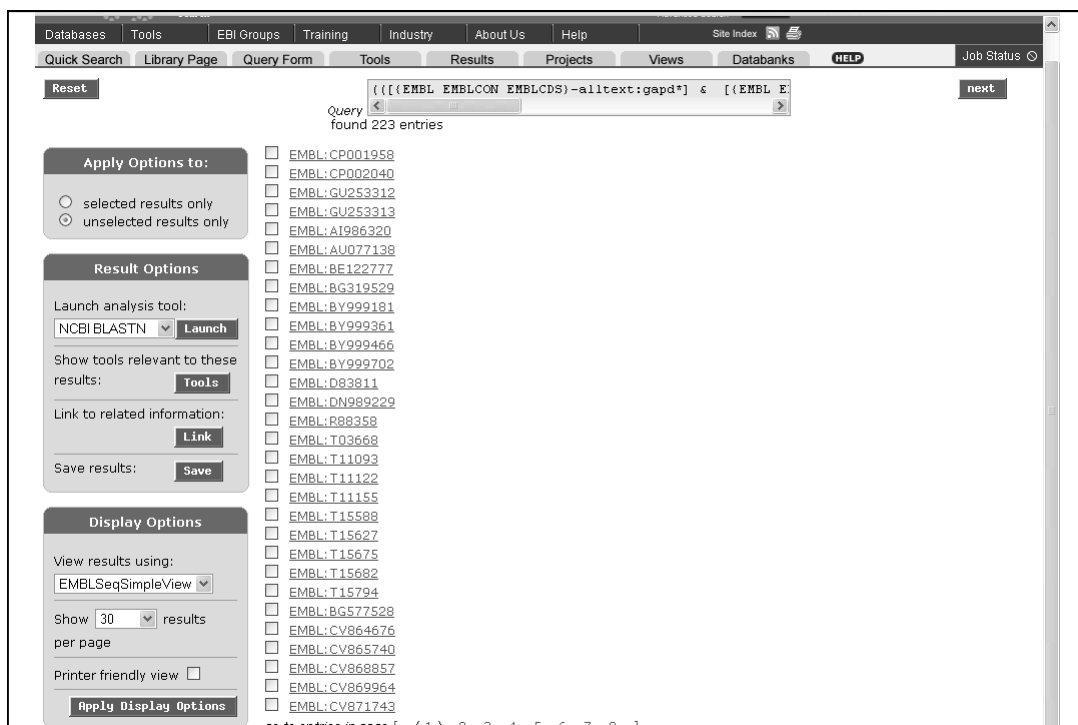
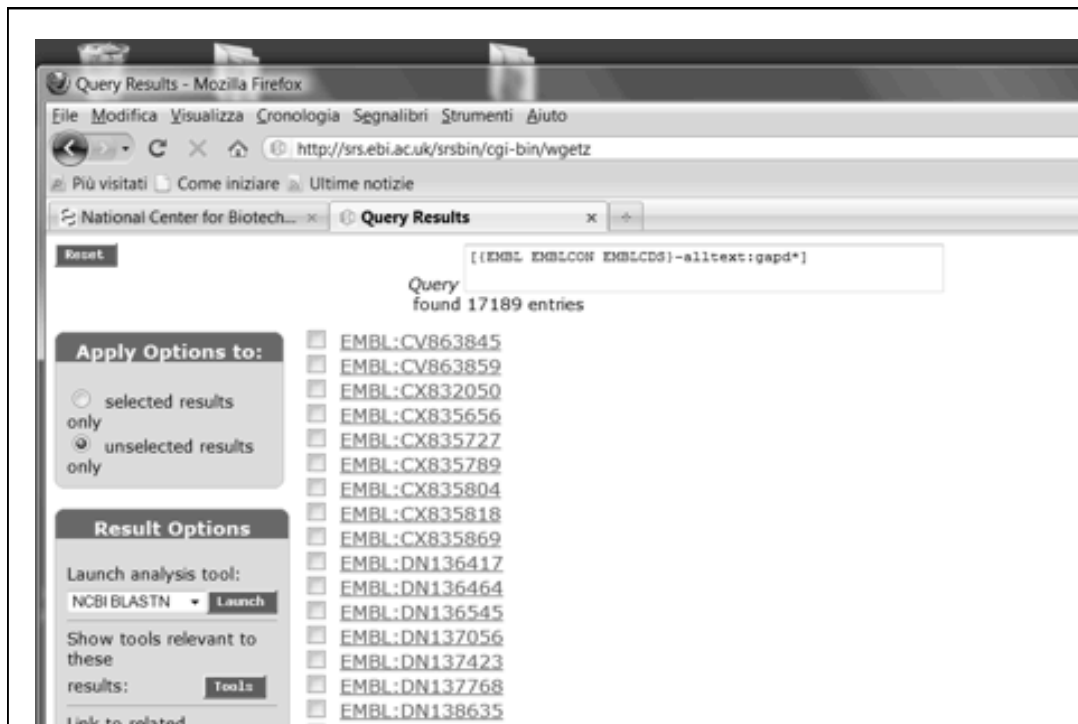
**Important announcements:**

- 07.09.09 EMBL Release 101 is now on-line ([release notes](#), [data notes](#)).
- 21.08.09 **Please Note:** Sunday 23rd August 09:30-10:00 AM. Some services will be unavailable due to essential maintenance on Sunday 23rd August. Apologies for the inconvenience caused.
- 10.06.09 EMBL Release 100 is now on-line ([release notes](#), [data notes](#)).
- 05.03.09 EMBL Release 99 is now on-line ([release notes](#), [data notes](#)).
- 17.02.09 The MEDLINE databanks (MEDLINE, MEDLINE2009 and MEDLINE99) now have additional views to provide citation information in formats compatible with popular citation management software. In addition to the existing MEDLINE XML

**List Search** Search Tips

Paste in a list of sequence ID's. The list must be of the format DATABASE:ID. e.g. EMBL:AB046566 or UNIPROT:104K\_THEAN. For more details see the [wiki](#).

Ensure each entry is on a single line and that the database(s) exists on this server. Multiple databases can be searched simultaneously. There is a maximum limit of 500 ID's.



**Where to study**

**NCBI Entrez Help**

**EBI EB-eye Help**

**EBI SRS Help Center**