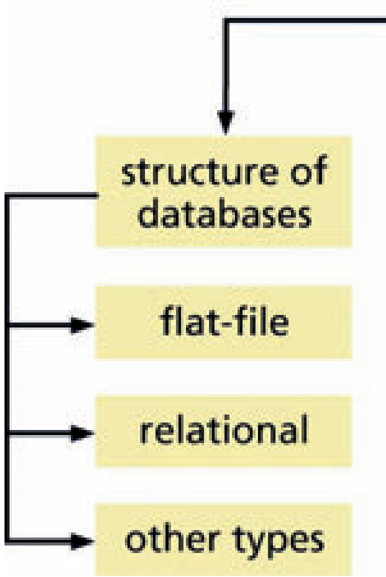
Biological Databases

1. http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.0010034
2. http://en.wikipedia.org/wiki/Biological_databases
3. http://www.uniprot.org/help/about
4. Some basic bioinformatics databases, see course page

Databases



DEALING WITH DATABASES

structure of databases

flat-file

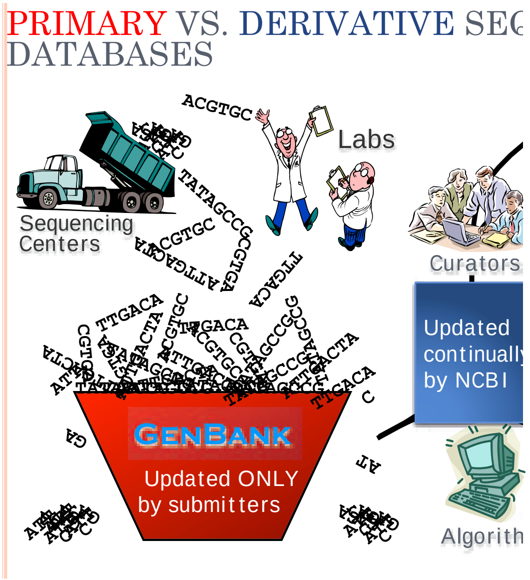
relational

other types

What is a database

- **Structured** collection of information.
- Consists of basic units called records or entries.
- Each record consists of fields, which hold **pre-defined** data related to the record.
- For example, a protein database would have protein entries as records and protein properties as fields (e.g., name of protein, length, amino-acid sequence)

Types of databases


PRIMARY VS. DERIVATIVE SEQ
DATABASES

Primary Databases

Original submissions by experimentalists

Content controlled by the **submitter**

Examples: GenBank, Trace, SRA, SNP, GEO

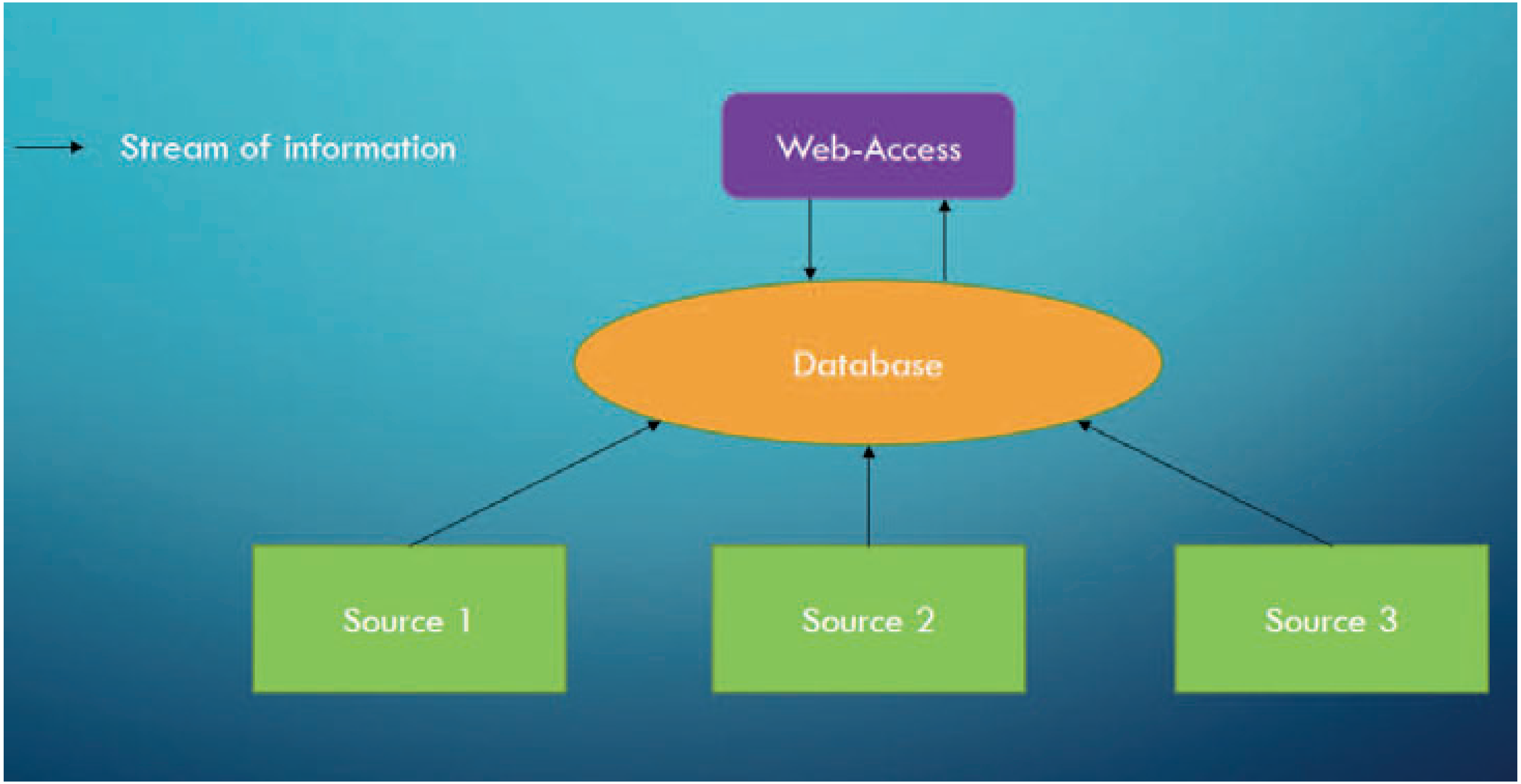
Derivative Databases

Derived from primary data

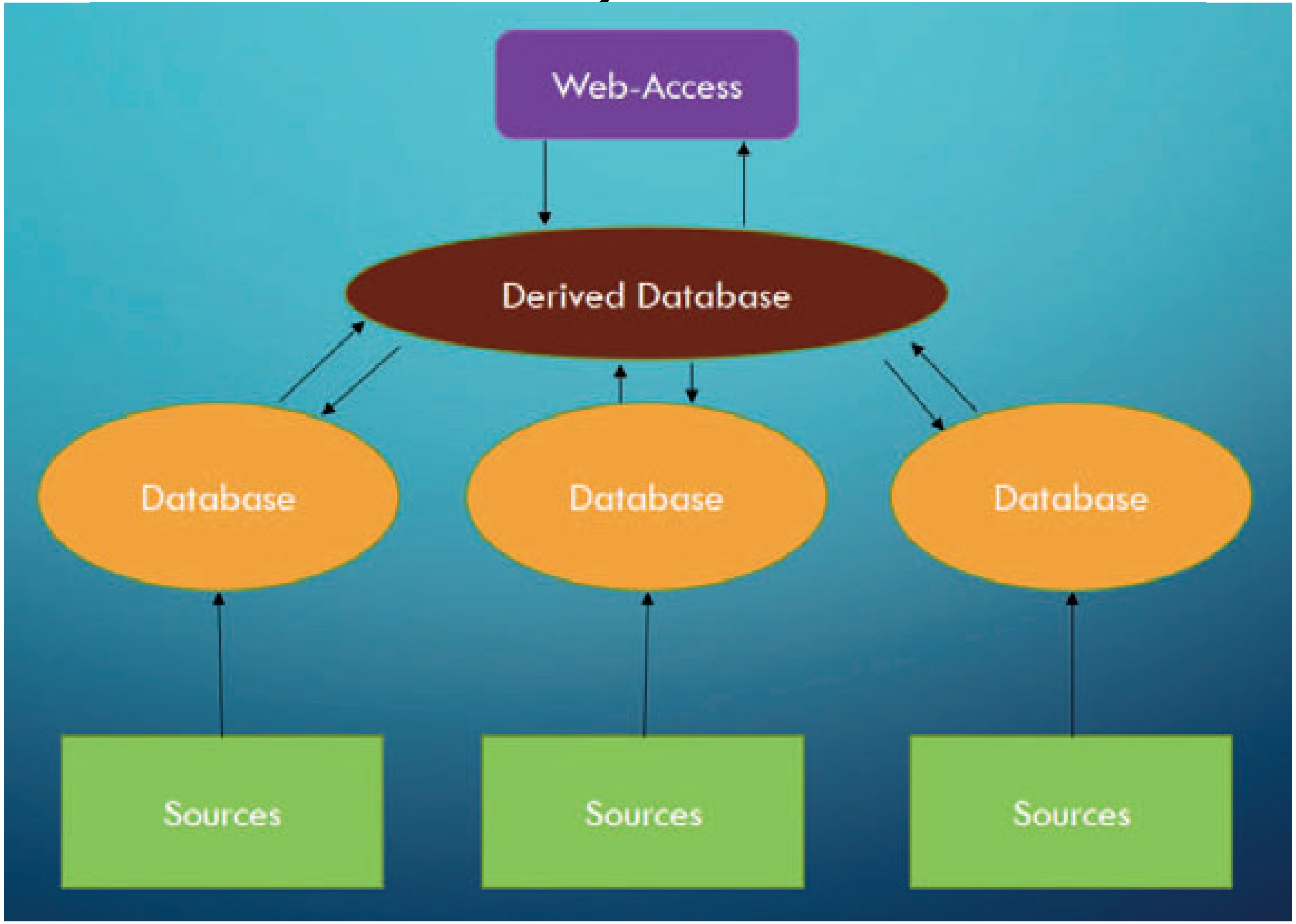
Content controlled by **third party** (NCBI)

Examples: NCBI Protein, Refseq, TPA, RefSNP, GEO datasets, UniGene, Homologene, Structure, Conserved Domain

Primary databases



Secondary databases



Web-Access

Derived Database

Database

Database

Database

Sources

Sources

Sources

A flat-file database

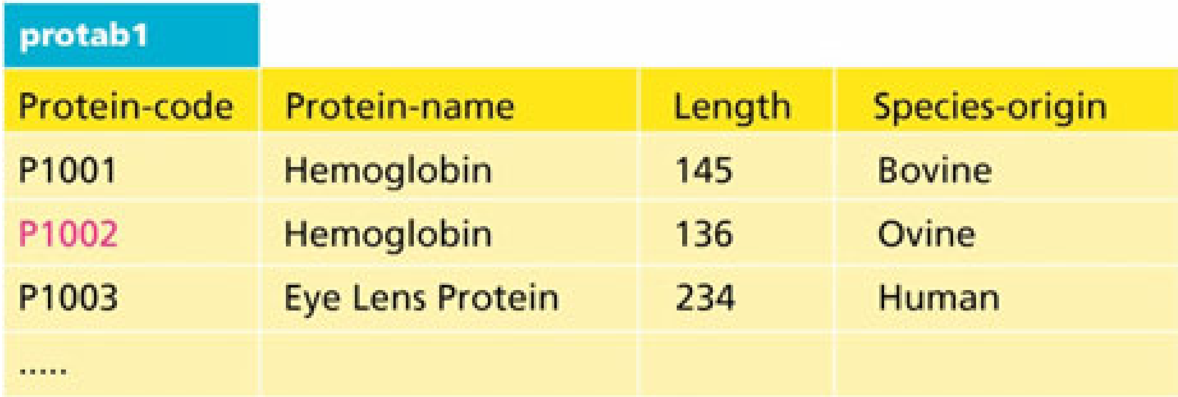
(A)

NAME	TELEPHONE	ADDRESS
S. Claus	0203 450	The North Pole, Lapland
M. Mouse	0202 453	Disneyworld, Florida
A. Moonman	0104 459	Craterland, The Moon

(B) GenBank Flat-File Format

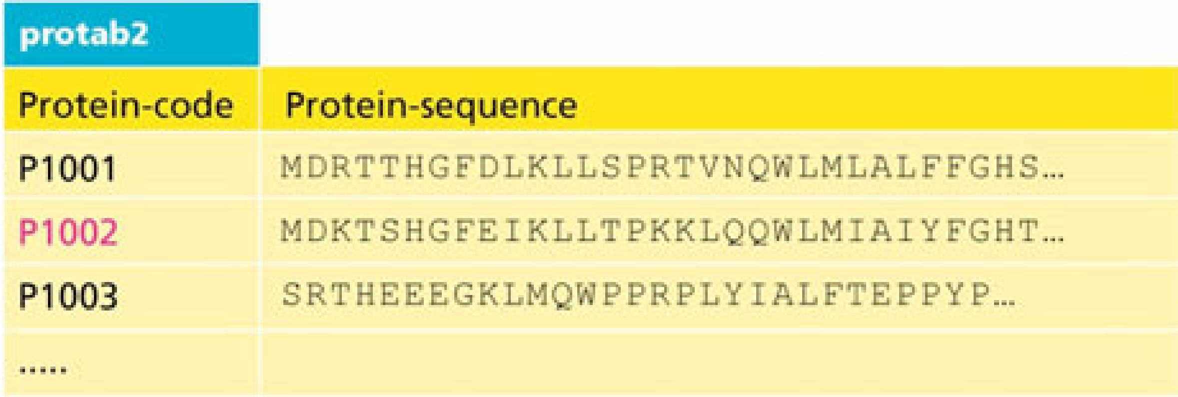
```
LOCUS        SCU49845     5028 bp    DNA
DEFINITION   Saccharomyces cerevisiae TCP1-beta gene, partial cds, and
              Axl2p
              (AXL2) and Rev7p (REV7) genes, complete cds.
ACCESSION    U49845
VERSION      U49845.1  GI:1293613
KEYWORDS     .
SOURCE       Saccharomyces cerevisiae (baker's yeast)
  ORGANISM   Saccharomyces cerevisiae
             Eukaryota; Fungi; Ascomycota; Saccharomycotina;
            Saccharomycetes;
             Saccharomycetales; Saccharomycetaceae; Saccharomyces.
```

Relational databases



protab1

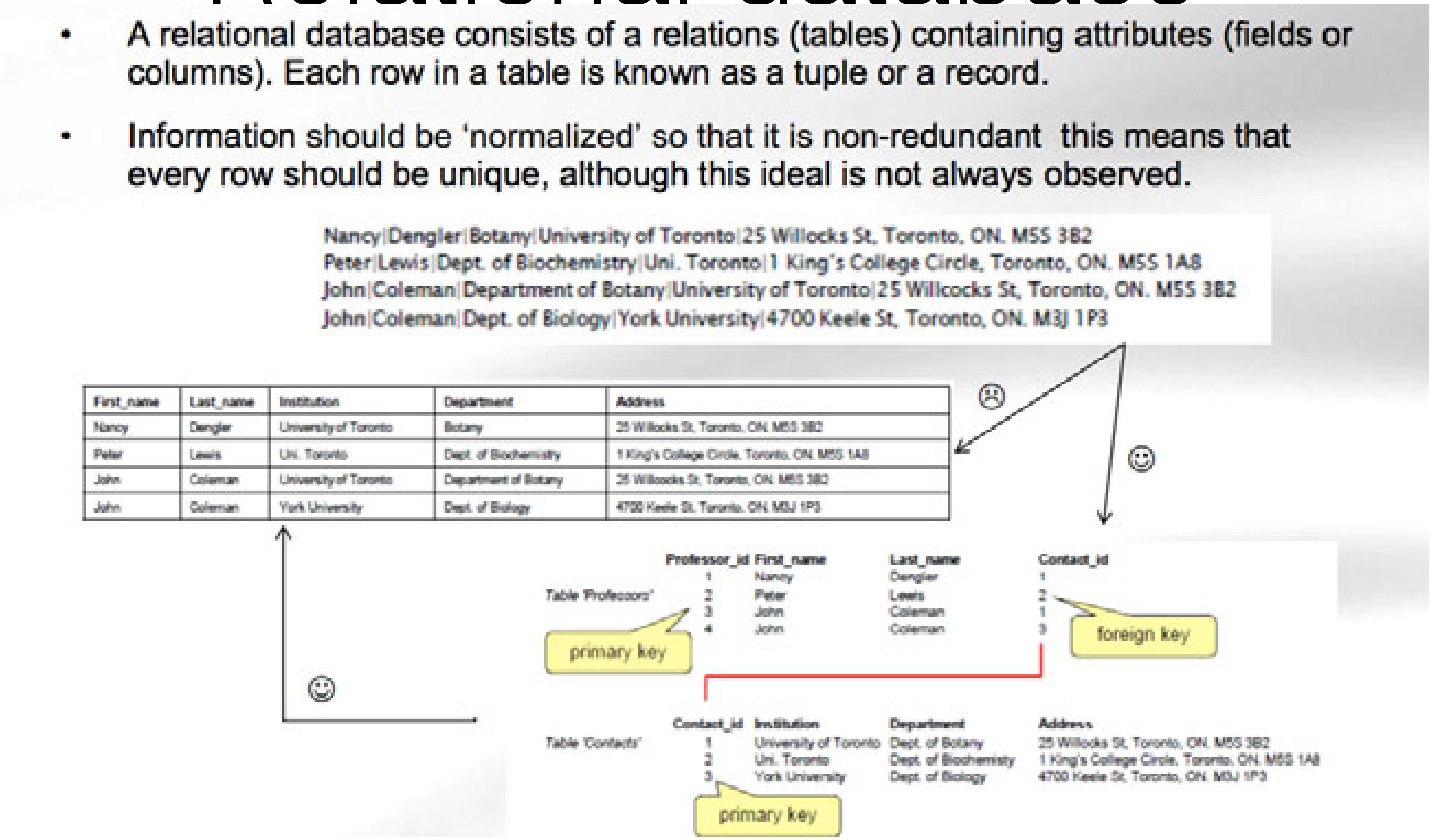
Protein-code	Protein-name	Length	Species-origin
P1001	Hemoglobin	145	Bovine
P1002	Hemoglobin	136	Ovine
P1003	Eye Lens Protein	234	Human
.....			



protab2

Protein-code	Protein-sequence
P1001	MDRTTHGFDLKLLSPRTVNQWLMLALFFGHS…
P1002	MDKTSHGFEIKLLTPKKLQQWLMIAIYFGHT…
P1003	SRTHEEEGKLMQWPPRPLYIALFTEPPYP…
.....	

Relational database



- A relational database consists of a relations (tables) containing attributes (fields or columns). Each row in a table is known as a tuple or a record.
- Information should be 'normalized' so that it is non-redundant this means that every row should be unique, although this ideal is not always observed.

```
Nancy|Dengler|Botany|University of Toronto|25 Willocks St, Toronto, ON. M5S 3B2
Peter|Lewis|Dept. of Biochemistry|Uni. Toronto|1 King's College Circle, Toronto, ON. M5S 1A8
John|Coleman|Department of Botany|University of Toronto|25 Willcocks St, Toronto, ON. M5S 3B2
John|Coleman|Dept. of Biology|York University|4700 Keele St, Toronto, ON. M3J 1P3
```

First_name	Last_name	Institution	Department	Address
Nancy	Dengler	University of Toronto	Botany	25 Willocks St, Toronto, ON M5S 3B2
Peter	Lewis	Uni. Toronto	Dept. of Biochemistry	1 King's College Circle, Toronto, ON M5S 1A8
John	Coleman	University of Toronto	Department of Botany	25 Willcocks St, Toronto, ON M5S 3B2
John	Coleman	York University	Dept. of Biology	4700 Keele St, Toronto, ON M3J 1P3

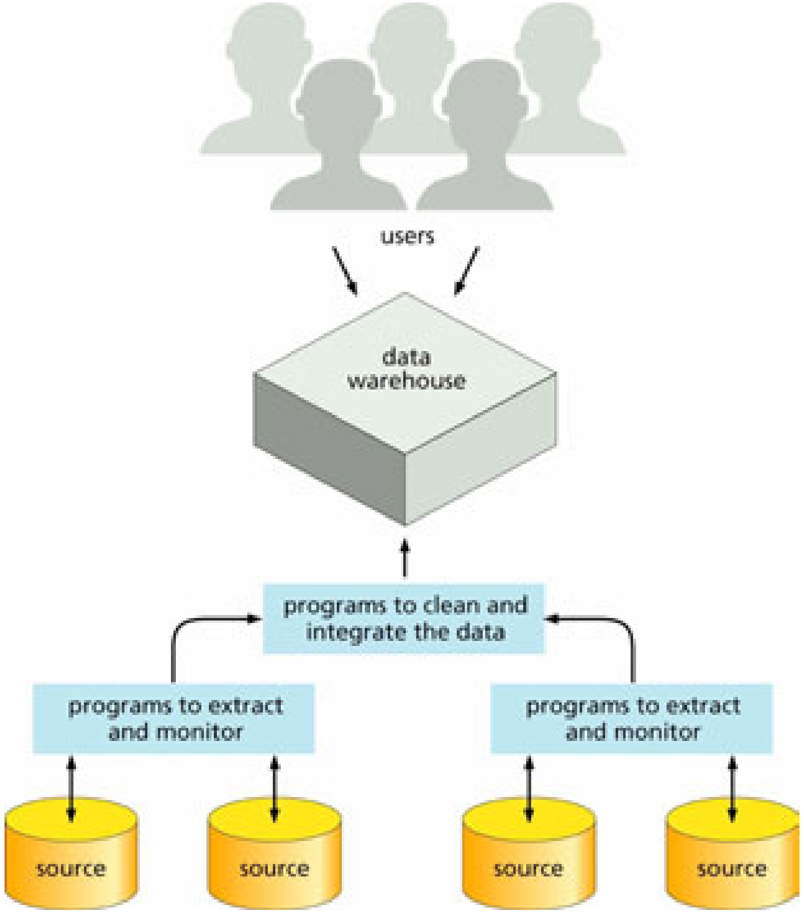
Table 'Professors'

Professor_id	First_name	Last_name	Contact_id
1	Nancy	Dengler	1
2	Peter	Lewis	2
3	John	Coleman	1
4	John	Coleman	3

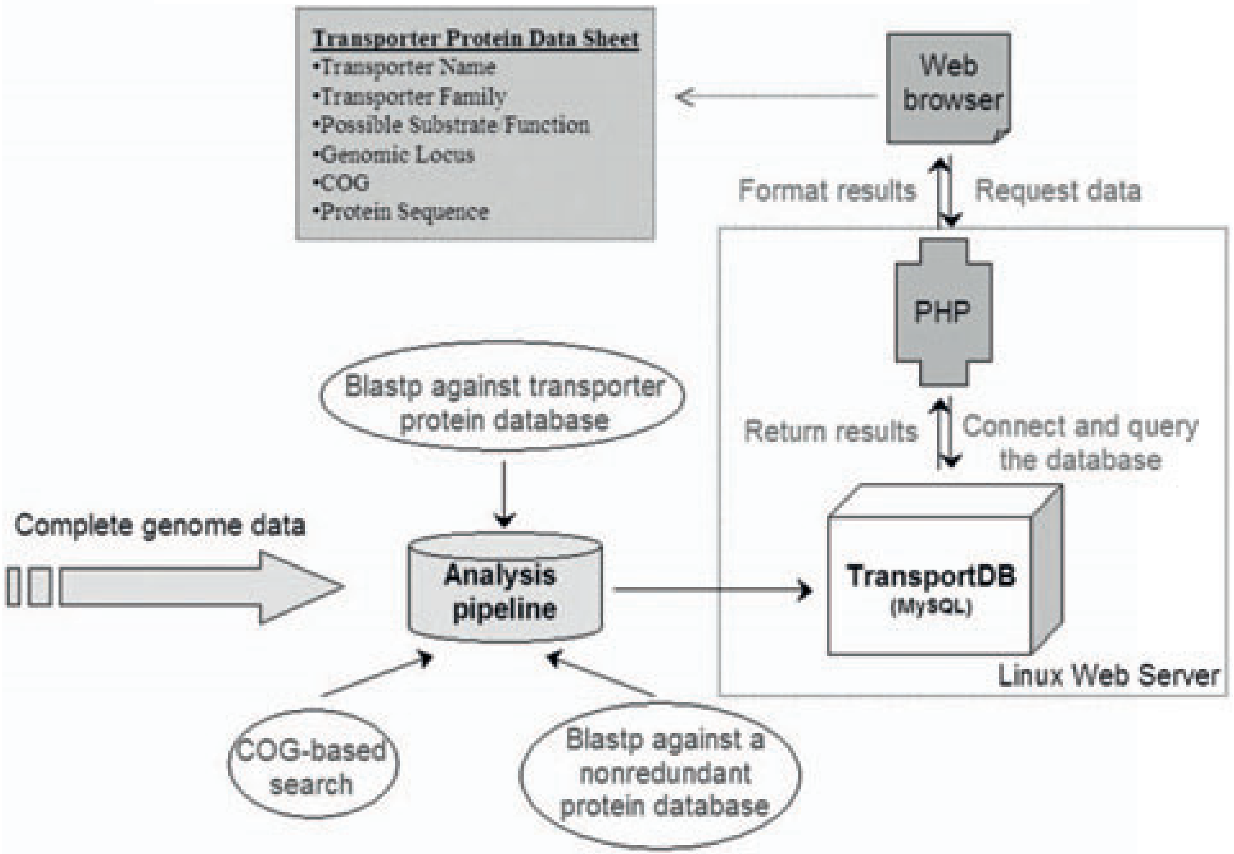
Table 'Contacts'

Contact_id	Institution	Department	Address
1	University of Toronto	Dept. of Botany	25 Willocks St, Toronto, ON. M5S 3B2
2	Uni. Toronto	Dept. of Biochemistry	1 King's College Circle, Toronto, ON. M5S 1A8
3	York University	Dept. of Biology	4700 Keele St, Toronto, ON. M3J 1P3

Data warehouse



Data warehouse



Types of databases

DNA & RNA
genes, genomes & variation

Gene expression
RNA, protein & metabolite expression

Proteins
sequences, families & motifs

Structures
Molecular & cellular structures

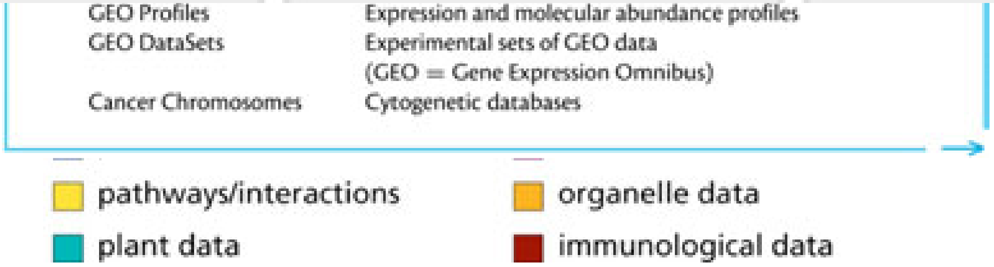
Systems
reactions, interactions & pathways

Chemical biology
chemogenomics & metabolomics

Ontologies
taxonomies & controlled vocabularies

Literature
Scientific publications & patents

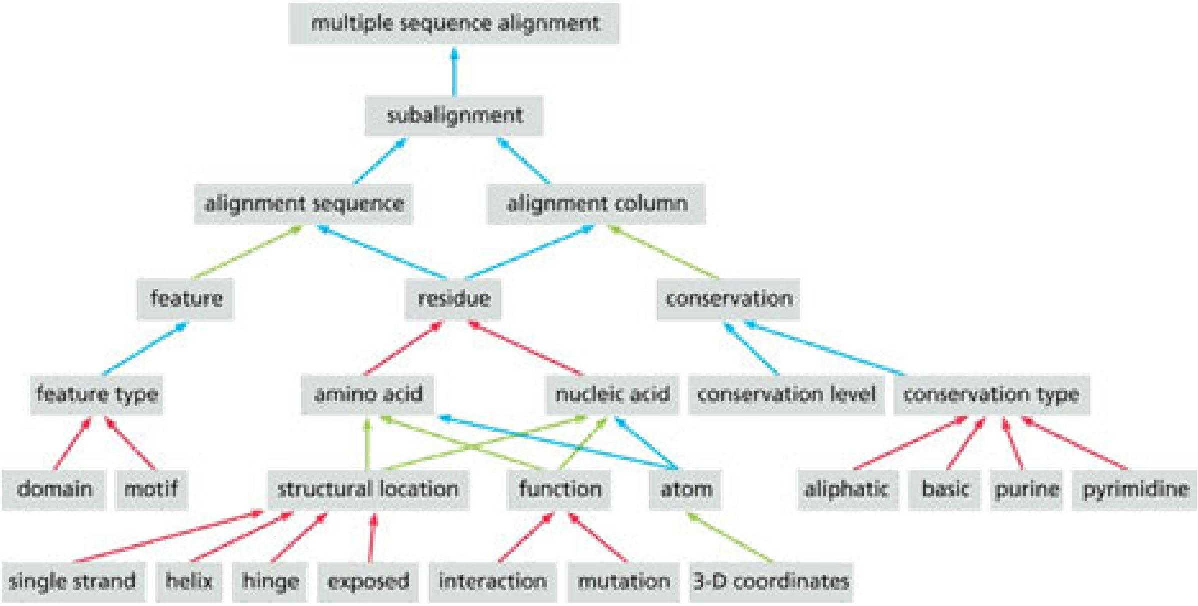
Other software
cross-domain tools & resources



GEO Profiles	Expression and molecular abundance profiles
GEO DataSets	Experimental sets of GEO data (GEO = Gene Expression Omnibus)
Cancer Chromosomes	Cytogenetic databases

🟨 pathways/interactions
🟧 organelle data
🟦 plant data
🟥 immunological data

MAO ontology



NCBI

DDBJ
DNA Data Bank of Japan

EMBL-EBI

Database providers

- The National Center for Biotechnology Information (NCBI) offers data banks, databases and tools at the USA
- The European Bioinformatics Institute (EBI) does a similar function in Europe
- GenomeNet gathers several databases from Japan

Data quality

- How are things entered
- What are the evidence ?
- How new is the data.
- Can the data be secret ?
- Redundant or non-redundant ?

Database providers

- The National Center for Biotechnology Information (NCBI) offers data banks, databases and tools at the USA
- The European Bioinformatics Institute (EBI) does a similar function in Europe
- GenomeNet gathers several databases from Japan

NCBI

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | NCBI News

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How Tos: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

NCBI YouTube channel

Learn how to get the most out of NCBI tools and databases with video tutorials on the NCBI YouTube Channel.

NCBI Announcements

Milestone: NCBI annotates 150th eukaryotic genome

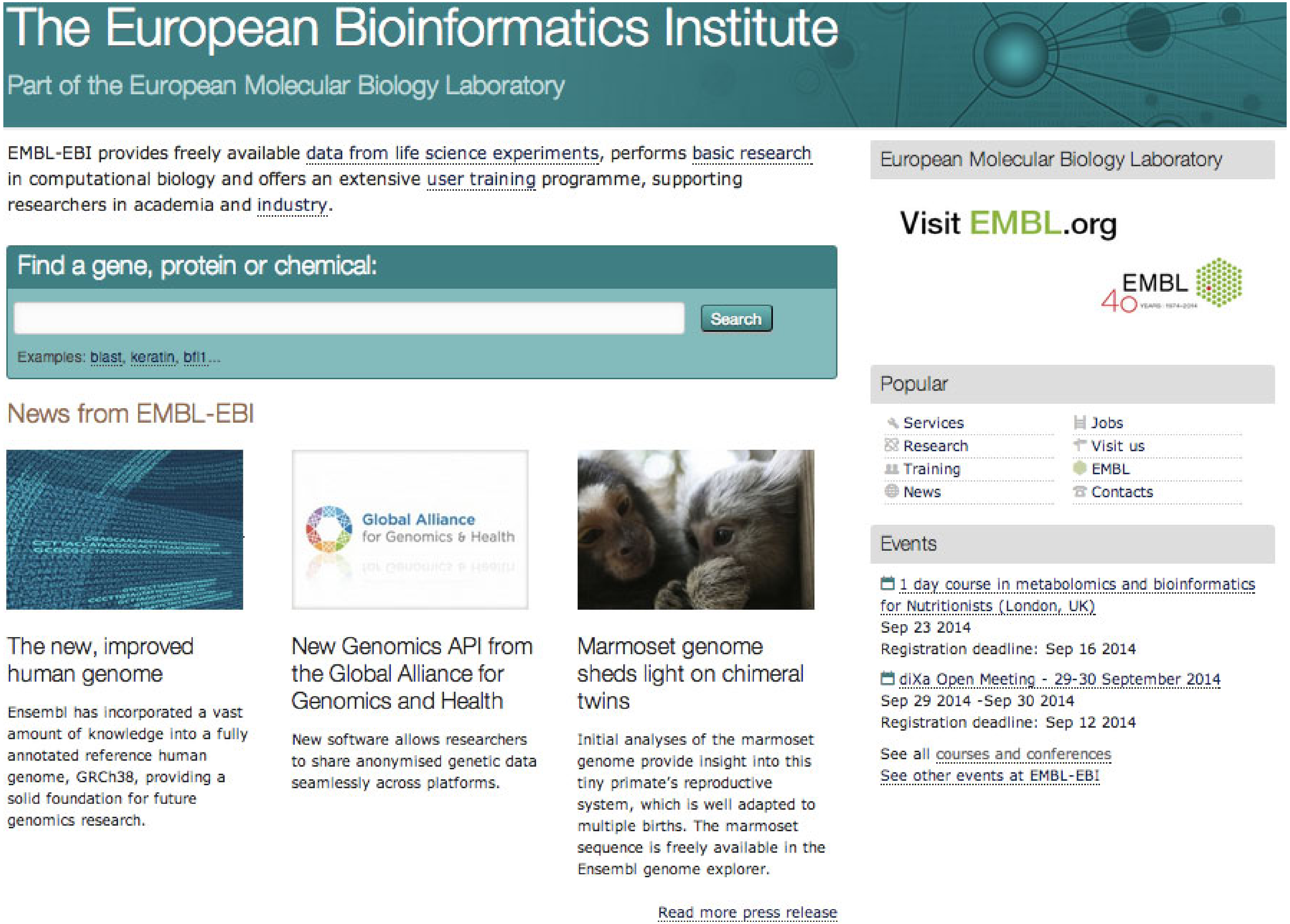
Aug 28, 2014

NCBI has now completed the genome

The new NCBI Genomes FTP site is here!

Aug 28, 2014

NCBI has released a major revision of the



The European Bioinformatics Institute

Part of the European Molecular Biology Laboratory

EMBL-EBI provides freely available data from life science experiments, performs basic research in computational biology and offers an extensive user training programme, supporting researchers in academia and industry.

Find a gene, protein or chemical:

Examples: blast, keratin, bfl1...

News from EMBL-EBI

The new, improved human genome

Ensembl has incorporated a vast amount of knowledge into a fully annotated reference human genome, GRCh38, providing a solid foundation for future genomics research.

New Genomics API from the Global Alliance for Genomics and Health

New software allows researchers to share anonymised genetic data seamlessly across platforms.

Marmoset genome sheds light on chimeral twins

Initial analyses of the marmoset genome provide insight into this tiny primate's reproductive system, which is well adapted to multiple births. The marmoset sequence is freely available in the Ensembl genome explorer.

Read more press release

European Molecular Biology Laboratory

Visit EMBL.org

Popular

Events

1 day course in metabolomics and bioinformatics for Nutritionists (London, UK)
Sep 23 2014
Registration deadline: Sep 16 2014

diXa Open Meeting - 29-30 September 2014
Sep 29 2014 -Sep 30 2014
Registration deadline: Sep 12 2014

See all courses and conferences
See other events at EMBL-EBI

GenomeNet
- About GenomeNet
- Release notes
- Acknowledgments

DBGET
- Overview
- DB release info

KEGG

varDB

Community DBs

Bioinformatics tools
- Other tools

FTP

Feedback

GenomeNet Database Resources

DBGET: Integrated Database Retrieval System
- DBGET search
- LinkDB search SPARQL endpoint available *New!*

KEGG: Kyoto Encyclopedia of Genes and Genomes
- KEGG2 - Table of contents
- KEGG PATHWAY - Systems information: pathways
- KEGG BRITE - Systems information: ontologies
- KEGG Organisms - Organism-specific entry points
- KEGG GENES - Genomic information
- KEGG LIGAND - Chemical information
- KEGG MEDICUS - Health information

Reaction Ontology: Reaction classifications

varDB: Antigenic variation database

Community Databases
- CYORF - Cyanobacteria annotation database
- BSORF - Bacillus subtilis genome database
- EXPRESSION - Gene expression profile database

GenomeNet Bioinformatics Tools

Sequence Analysis
- BLAST / FASTA - Sequence similarity search
  - *Ribosomal databases are availalbe Updated!*
- MOTIF - Sequence motif search
- CLUSTALW / MAFFT / PRRN - Multiple alignment

Genome Analysis
- OC Viewer - KEGG ortholog clusters *Updated!*
  - REST service is available
- KAAS - KEGG automatic annotation server

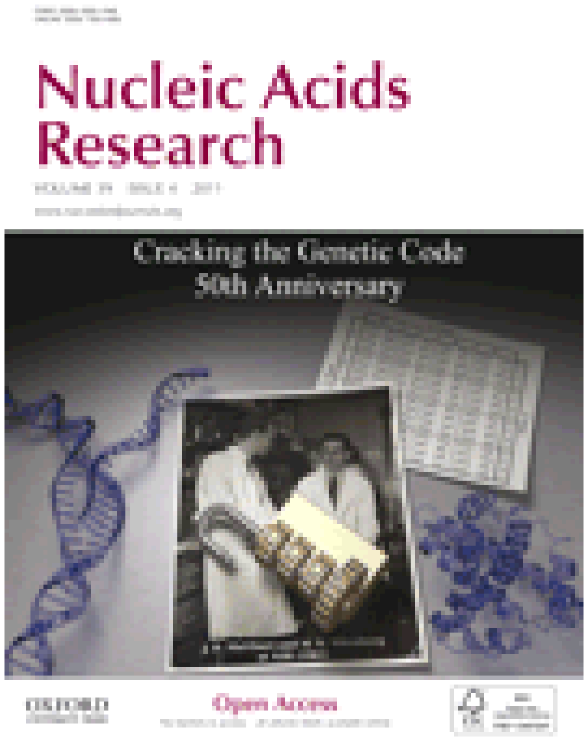
BRITE Functional Hierarchies
- KEGG Pathway Maps
- KEGG Organisms

Genome Catalogs
- Complete (KEGG GENES)
- Draft (KEGG DGENES)
- Meta (KEGG MGENES)
- Viral (KEGG VGENES)
- EST (EGENES)

Ortholog Clusters (OC)

Database links

NAR database issue

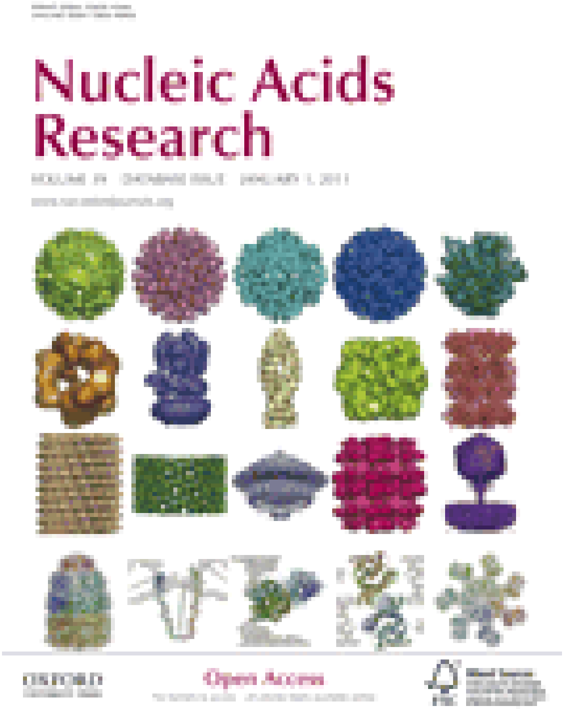


Nucleic Acids Research

Cracking the Genetic Code
50th Anniversary

OXFORD

Open Access



Nucleic Acids Research

OXFORD

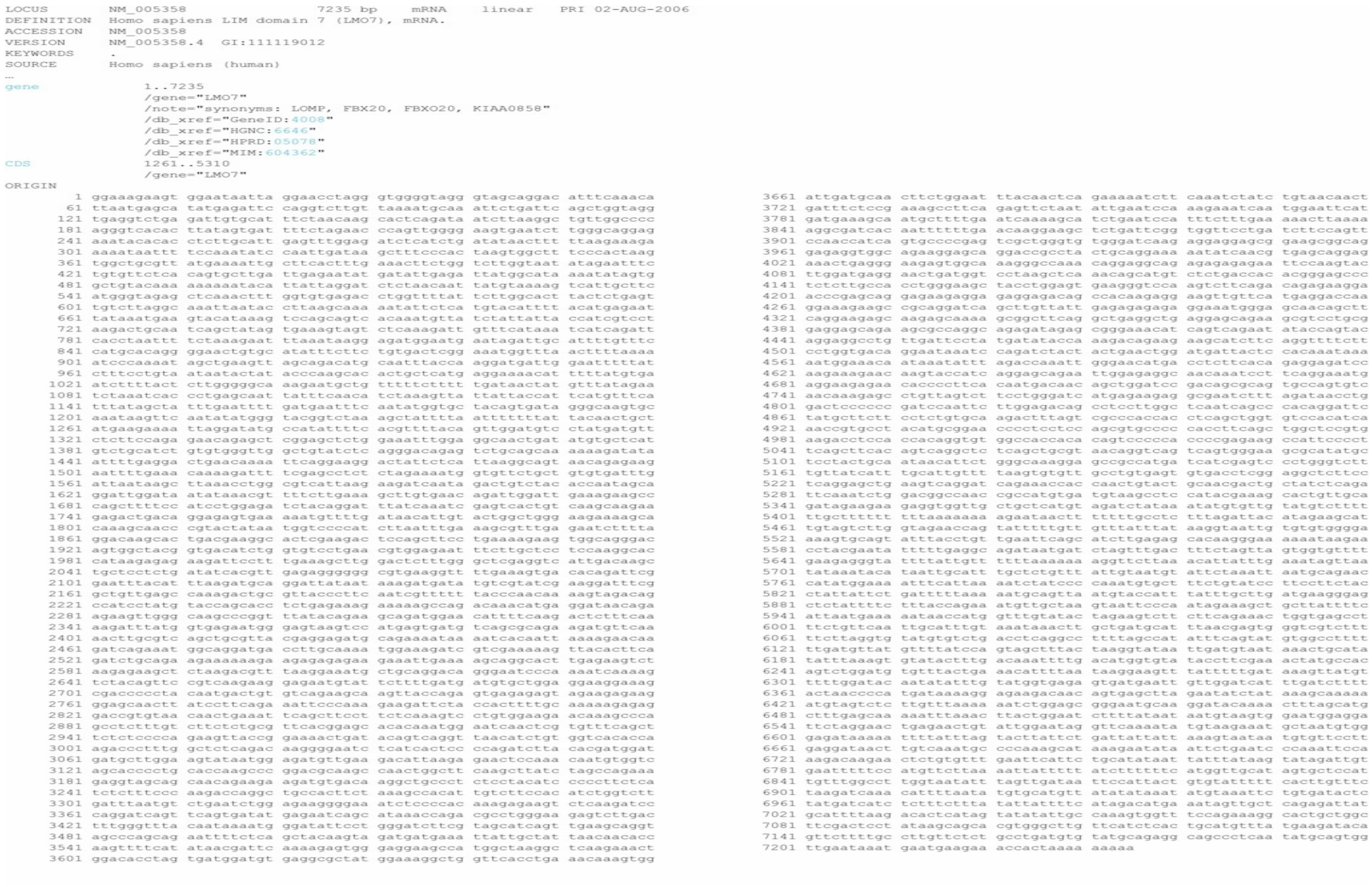
Open Access

Nucleotide databases

- International nucleotide sequence database collaborations
  - Genbank
  - EMBL
  - DDBJ

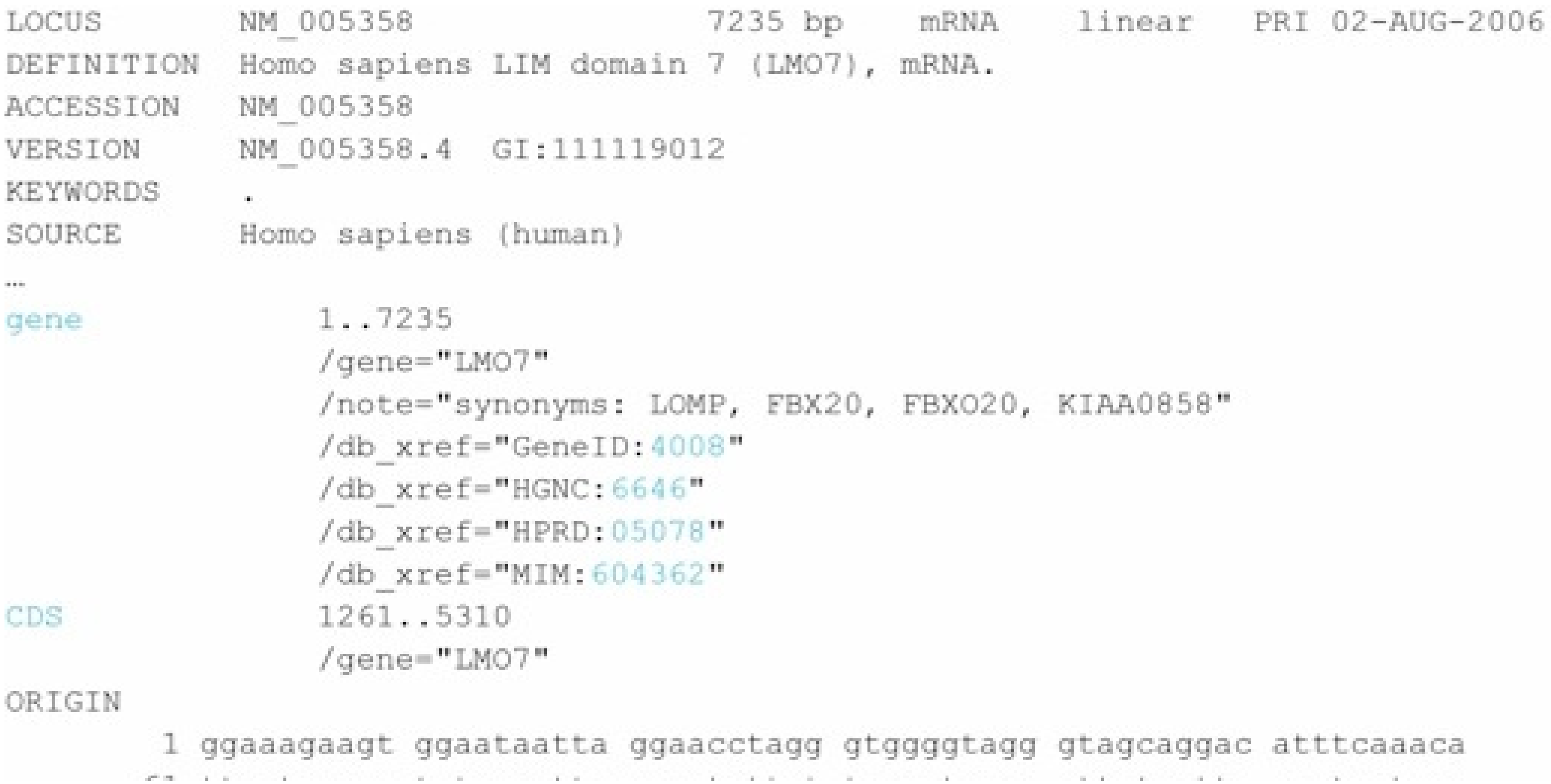
The nucleotide sequence databases are data repositories, accepting nucleic acid sequence data from the scientific community and making it freely available. The databases strive for completeness, with the aim of recording every publicly known nucleic acid sequence. These data are heterogenous, they vary with respect to the source of the material (e.g. genomic versus cDNA), the intended quality (e.g. finished versus single pass sequences), the extent of sequence annotation and the intended completeness of the sequence relative to its biological target (e.g. complete versus partial coverage of a gene or a genome). The nucleotide databases are distributed free of charge over the internet.

Genbank entry



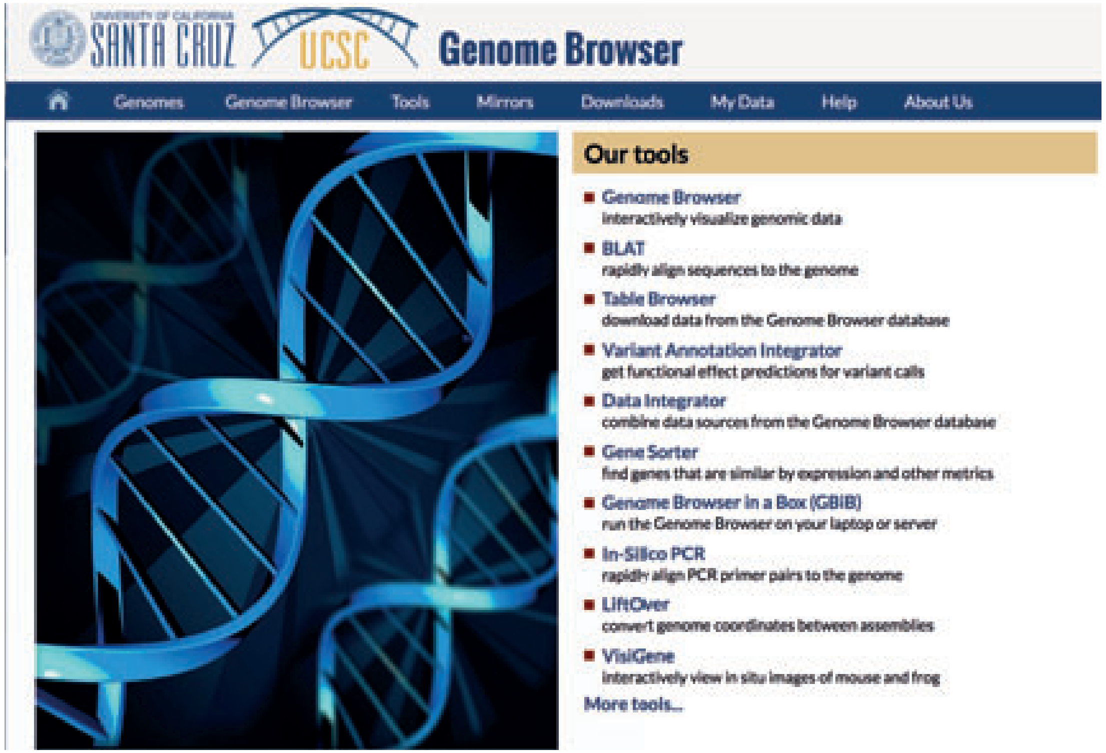
```
LOCUS       NM_005358               7235 bp    mRNA    linear   PRI 02-AUG-2006
DEFINITION  Homo sapiens LIM domain 7 (LMO7), mRNA.
ACCESSION   NM_005358
VERSION     NM_005358.4  GI:111119012
KEYWORDS    .
SOURCE      Homo sapiens (human)
...
gene            1..7235
                /gene="LMO7"
                /note="synonyms: LOMP, FBX20, FBXO20, KIAA0858"
                /db_xref="GeneID:4008"
                /db_xref="HGNC:6646"
                /db_xref="HPRD:05078"
                /db_xref="MIM:604362"
CDS             1261..5310
                /gene="LMO7"
ORIGIN
        1 ggaaagaagt ggaataatta ggaacctagg gtggggtagg gtagcaggac atttcaaaca
       61 ttaatgagca tatgagattc caggtcttgt taaaatgcaa attctgattc agctggtagg
      121 tgaggtctga gattgtgcat ttctaacaag cactcagata atcttaaggc tgttggcccc
      181 agggtcacac ttatagtgat tttctagaac ccagttgggg aagtgaatct tgggcaggag
      241 aaatacacac ctcttgcatt gagtttggag atctcatctg atataacttt ttaagaaaga
      301 aaaataattt tccaaatatc caattgataa gctttcccac taagtggctt tcccactaag
      361 tggctgcgtt atgaaaattg cttcactttg aaacttctgg tcttggtaat atagaatttc
      421 tgtgttctca cagtgcttga ttgagaatat gatattgaga ttatggcata aaatatagtg
      481 gctgtacaaa aaaaaataca ttattaggat ctctaacaat tatgtaaaag tcattgcttc
      541 atgggtagag ctcaaacttt ggtgtgagac ctggttttat tcttggcact tactctgagt
      601 tgtcttaggc aaattaatac cttaagcaaa aatattctca tgtacatttt acatgagaat
      661 tataaatgaa gtacataaag tccagcagtc acaaatgtta tctattatta ccatcgtcct
      721 aagactgcaa tcagctatag tgaaagtagt ctcaaagatt gtttcataaa tcatcagatt
      781 cacctaattt tctaaagaat ttaaataagg agatggaatg aatagattgc attttgtttc
      841 catgcacagg ggaactgtgc atatttcttc tgtgactcgg aaatggttta acttttaaaa
      901 atcccaaaat agctgaagtt agcagacatg caatttacca aggatgattg gaatttttat
      961 ctttcctgta ataatactat acccaagcac actgctcatg aggaaaacat ttttatgtga
     1021 atcttttact cttgggggca aagaatgctg tttttctttt tgataactat gtttatagaa
     1081 tctaaatcac cctgagcaat tatttcaaca tctaaagtta ttattaccat tcatgtttca
     1141 tttatagcta tttgaatttt gatgaatttc aatatggtgc tacagtgata gggcaagtgc
     1201 aaataagttc aatatatggg tacggtctaa agctatttta atttttttat tacaactgct
     1261 atgaagaaaa ttaggatatg ccatattttc acgttttaca gttggatgtc ctatgatgtt
     1321 ctcttccaga gaacagagct cggagctctg gaaatttgga ggcaactgat atgtgctcat
     1381 gtctgcatct gtgtgggttg gctgtatctc agggacagag tctgcagcaa aaaagatata
     1441 atttgagga ctgaacaaaa ttcaggaagg actattctca ttaaggcagt aacagagaag
     1501 aattttgaaa caaaagattt tcgagcctct ctagaaaatg gtgttctgct gtgtgatttg
     1561 attaataagc ttaaacctgg cgtcattaag aagatcaata gactgtctac accaatagca
     1621 ggattggata atataaacgt tttcttgaaa gcttgtgaac agattggatt gaaagaagcc
     1681 cagcttttcc atcctggaga tctacaggat ttatcaaatc gagtcactgt caagcaagaa
     1741 gagactgaca ggagagtgaa aaatgttttg ataacattgt actggctggg aagaaaagca
     1801 caaagcaacc cgtactataa tggtccccat cttaatttga aagcgtttga gaatctttta
     1861 ggacaagcac tgacgaaggc actcgaagac tccagcttcc tgaaaagaag tggcagggac
     1921 agtggctacg gtgacatctg gtgtcctgaa cgtggagaat ttcttgctcc tccaaggcac
     1981 cataagagag aagattcctt tgaaagcttg gactctttgg gctcgaggtc attgacaagc
     2041 tgctcctctg atatcacgtt gagagggggg cgtgaaggtt ttgaaagtga cacagattcg
     2101 gaatttacat ttaagatgca ggattataat aaagatgata tgtcgtatcg aaggatttcg
     2161 gctgttgagc caaagactgc gttacccttc aatcgttttt tacccaacaa aagtagacag
     2221 ccatcctatg taccagcacc tctgagaaag aaaaagccag acaaacatga ggataacaga
     2281 agaagttggg caagcccggt ttatacagaa gcagatggaa cattttcaag actctttcaa
     2341 aagatttatg gtgagaatgg gagtaagtcc atgagtgatg tcagcgcaga agatgttcaa
     2401 aacttgcgtc agctgcgtta cgaggagatg cagaaaataa aatcacaatt aaaagaacaa
     2461 gatcagaaat ggcaggatga ccttgcaaaa tggaaagatc gtcgaaaaag ttacacttca
     2521 gatctgcaga agaaaaaaga agagagagaa gaaattgaaa agcaggcact tgagaagtct
     2581 aagagaagct ctaagacgtt taaggaaatg ctgcaggaca gggaatccca aaatcaaaag
     2641 tctacagttc cgtcaagaag gagaatgtat tcttttgatg atgtgctgga ggaaggaaag
     2701 cgacccccta caatgactgt gtcagaagca agttaccaga gtgagagagt agaagagaag
     2761 ggagcaactt atccttcaga aattcccaaa gaagattcta ccacttttgc aaaaagagag
     2821 gaccgtgtaa caactgaaat tcagcttcct tctcaaagtc ctgtggaaga acaaagccca
     2881 gcctctttgt cttctctgcg ttcacggagc acacaaatgg aatcaactcg tgtttcagct
     2941 tctctcccca gaagttaccg gaaaactgat acagtcaggt taacatctgt ggtcacacca
     3001 agaccctttg gctctcagac aaggggaatc tcatcactcc ccagatctta cacgatggat
     3061 gatgcttgga agtataatgg agatgttgaa gacattaaga gaactccaaa caatgtggtc
     3121 agcaccccctg caccaagccc ggacgcaagc caactggctt caagcttatc tagccagaaa
     3181 gaggtagcag caacagaaga agatgtgaca aggctgccct ctcctacatc cccccttctca
     3241 tctctttccc aagaccaggc tgccacttct aaagccacat tgtcttccac atctggtctt
     3301 gatttaatgt ctgaatctgg agaaggggaa atctccccac aaagagaagt ctcaagatcc
     3361 caggatcagt tcagtgatat gagaatcagc ataaaccaga cgcctgggaa gagtcttgac
     3421 tttgggttta caataaaatg ggatattcct gggatcttcg tagcatcagt tgaagcaggt
     3481 agcccagcag aattttctca gctacaagta gatgatgaaa ttattgctat taacaacacc
     3541 aagttttcat ataacgattc aaaagagtgg gaggaagcca tggctaaggc tcaagaaact
     3601 ggacacctag tgatggatgt gaggcgctat ggaaaggctg gttcacctga aacaaagtgg
     3661 attgatgcaa cttctggaat ttacaactca gaaaaatctt caaatctatc tgtaacaact
     3721 gatttctccg aaagccttca gagttctaat attgaatcca aagaaatcaa tggaattcat
     3781 gatgaaagca atgcttttga atcaaaagca tctgaatcca tttctttgaa aaacttaaaa
     3841 aggcgatcac aatttttga acaaggaagc tctgattcgg tggttcctga tcttccagtt
     3901 ccaaccatca gtgccccgag tcgctgggtg tgggatcaag aggaggagcg gaagcggcag
     3961 gagaggtggc agaaggagca ggaccgccta ctgcaggaaa aatatcaacg tgagcaggag
     4021 aaactgaggg aagagtggca aagggccaaa caggaggcag agagagagaa ttccaagtac
     4081 ttggatgagg aactgatggt cctaagctca aacagcatgt ctctgaccac acgggagccc
     4141 tctcttgcca cctgggaagc tacctggagt gaagggtcca agtcttcaga cagagaagga
     4201 acccgagcag gagaagagga gaggagacag ccacaagagg aagttgttca tgaggaccaa
     4261 ggaaagaagc cgcaggatca gcttgttatt gagagagaga ggaaatggga gcaacagctt
     4321 caggaagagc aagagcaaaa gcggcttcag gctgaggctg aggagcagaa gcgtcctgcg
     4381 gaggagcaga agcgccaggc agagatagag cgggaaacat cagtcagaat ataccagtac
     4441 aggaggcctg ttgattccta tgatatacca aagacagaag aagcatcttc aggtttctt
     4501 cctggtgaca ggaataaatc cagatctact actgaactgg atgattactc cacaaataaa
     4561 aatggaaaca ataaatattt agaccaaatt gggaacatga cctcttcaca gaggagatcc
     4621 aagaaagaac aagtaccatc aggagcagaa ttggagaggc aacaaatcct tcaggaaatg
     4681 aggaagagaa cacccttca caatgacaac agctggatcc gacagcgcag tgccagtgtc
     4741 aacaaagagc ctgttagtct tcctgggatc atgagaagag gcgaatcttt agataacctg
     4801 gactccccc gatccaattc ttggagacag cctccttggc tcaatcagcc cacaggattc
     4861 tatgcttctt cctctgtgca agactttagt cgcccaccac ctcagctggt gtccacatca
     4921 aaccgtgcct acatgcggaa cccctcctcc agcgtgcccc caccttcagc tggctccgtg
     4981 aagacctcca ccacaggtgt ggccaccaca cagtccccca ccccgagaag ccattcccct
     5041 tcagcttcac agtcaggctc tcagctgcgt aacaggtcag tcagtgggaa gcgcatatgc
     5101 tcctactgca ataacattct gggcaaagga gccgccatga tcatcgagtc cctgggtctt
     5161 tgttatcatt tgcattgttt taagtgtgtt gcctgtgagt gtgacctcgg aggctcttcc
     5221 tcaggagctg aagtcaggat cagaaaccac caactgtact gcaacgactg ctatctcaga
     5281 ttcaaatctg gacggccaac cgccatgtga tgtaagcctc catacgaaag cactgttgca
     5341 gatagaagaa gaggtggttg ctgctcatgt agatctataa atatgtgttg tatgtctttt
     5401 ttgcttttt tttaaaaaaa agaataactt tttttgcctc tttagattac atagaagcat
     5461 tgtagtcttg gtagaaccag tatttttgtt gtttatttat aaggtaattg tgtgtgggga
     5521 aaagtgcagt atttacctgt tgaattcagc atcttgagag cacaagggaa aaaataagaa
     5581 cctacgaata tttttgaggc agataatgat ctagtttgac tttctagtta gtggtgtttt
     5641 gaagagggta ttttattgtt ttttaaaaaa aggttcttaa acattatttg aaatagttaa
     5701 tataaataca taattgcatt tgctctgttt attgtaatgt attctaaatt aatgcagaac
     5761 catatggaaa atttcattaa aatctatccc caaatgtgct ttctgtatcc ttccttctac
     5821 ctattattct gatttttaaa aatgcagtta atgtaccatt tatttgcttg atgaagggag
     5881 ctctattttc tttaccagaa atgttgctaa gtaattccca atagaaagct gcttattttc
     5941 attaatgaaa aataaccatg gtttgtatac tagaagtctt cttcagaaac tggtgagcct
     6001 ttctgttcaa ttgcatttgt aaataaactt gctgatgcat ttaacgagtg ggtcgtcttt
     6061 ttcttaggtg tatgtgtctg acctcaggcc ttttagccat atttcagtat gtggcctttt
     6121 ttgatgttat gttttatcca gtagctttac taaggtataa ttgatgtaat aaactgcata
     6181 tatttaaagt gtatactttg acaaattttg acatggtgta taccttcgaa actatgccac
     6241 agtctggatg tgtttactga aacattttaa taaggaagtt tatttttgat aaagttatgt
     6301 ttttggatac aatatatttg tatggtgaga gtgatgaatt gttggatcat ttgatctttt
     6361 actaacccca tgataaaagg agaagacaac agtgagctta gaatatctat aaagcaaaaa
     6421 atgtagtctc ttgtttaaaa aatctggagc gggaatgcaa ggatacaaaa ctttagcatg
     6481 ctttgagcaa aaatttaaac ttactggaat cttttataat aatgtaagtg gaatggagga
     6541 ttctaggaac tgagaactgt attggaatag gttcaaaata tgtaagaaat gctaatgtgg
     6601 gagataaaaa ttttatttag tacttattct gattattatt aaagtaataa tgtgttcctt
     6661 gaggataact tgtcaaatgc cccaaagcat aaagaatata attctgaatc ccaaattcca
     6721 aagacaagaa ctctgtgttt gaattcattc tgcatataat tatttataag tatagattgt
     6781 gaattttcc atgttcttaa aattattttt atcttttttc atggttgcat agtgctccat
     6841 tgtttggcct tggtaatatt tagttgataa ttccattact gtgtattttt cacttgtttc
     6901 taagatcaaa cattttaata tgtgcatgtt atatataaat atgtaaattc tgtgatactc
     6961 tatgatcatc tctttcttta tattattttc atagacatga aatagttgct cagagattat
     7021 gcattttaag acactcatag tatatattgc caaagtggtt tccagaaagg cactgctggc
     7081 ttcgactcct ataagcagca cgtgggcttg ttcatctcac tgcatgttta tgaagataca
     7141 gttcttttgc cttgttctct gcctgatgtg tatgcagagg cagccctcaa tatgcagtgg
     7201 ttgaataaat gaatgaagaa accactaaaa aaaaa
```

Genbank details



```
LOCUS       NM_005358              7235 bp    mRNA    linear   PRI 02-AUG-2006
DEFINITION  Homo sapiens LIM domain 7 (LMO7), mRNA.
ACCESSION   NM_005358
VERSION     NM_005358.4  GI:111119012
KEYWORDS    .
SOURCE      Homo sapiens (human)
…
gene            1..7235
                /gene="LMO7"
                /note="synonyms: LOMP, FBX20, FBXO20, KIAA0858"
                /db_xref="GeneID:4008"
                /db_xref="HGNC:6646"
                /db_xref="HPRD:05078"
                /db_xref="MIM:604362"
CDS             1261..5310
                /gene="LMO7"
ORIGIN
        1 ggaaagaagt ggaataatta ggaacctagg gtggggtagg gtagcaggac atttcaaaca
```

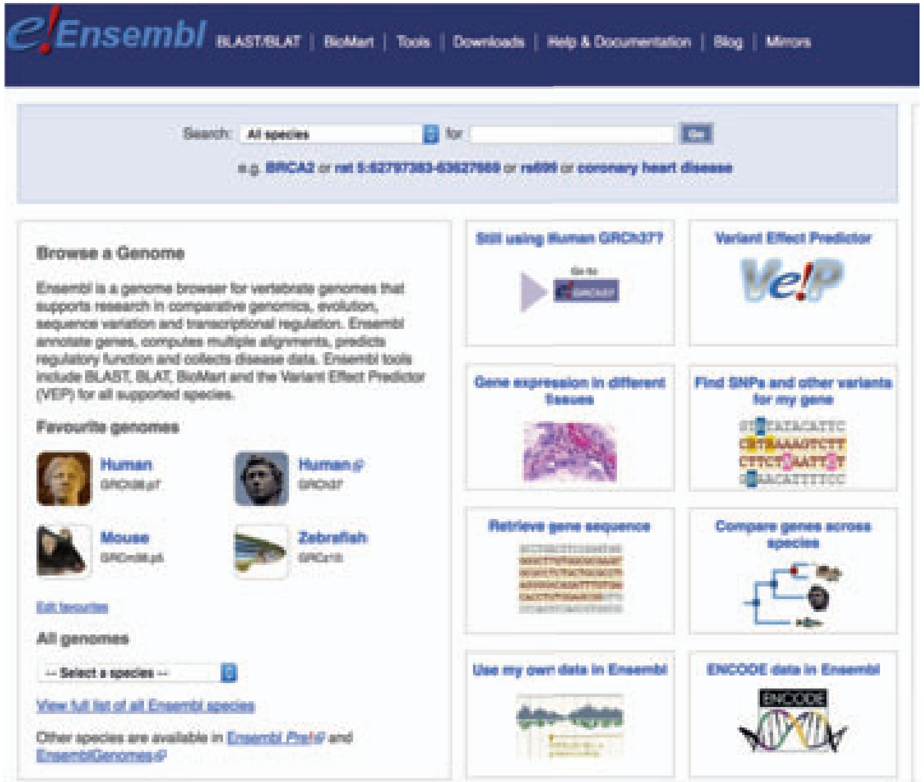
Genome specific databases



Our tools

- **Genome Browser**
  interactively visualize genomic data
- **BLAT**
  rapidly align sequences to the genome
- **Table Browser**
  download data from the Genome Browser database
- **Variant Annotation Integrator**
  get functional effect predictions for variant calls
- **Data Integrator**
  combine data sources from the Genome Browser database
- **Gene Sorter**
  find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBiB)**
  run the Genome Browser on your laptop or server
- **In-Silico PCR**
  rapidly align PCR primer pairs to the genome
- **LiftOver**
  convert genome coordinates between assemblies
- **VisiGene**
  interactively view in situ images of mouse and frog

More tools...



Search: All species for

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Favourite genomes

Human GRCh38.p7
Human GRCh37
Mouse GRCm38.p4
Zebrafish GRCz10

Edit favourites

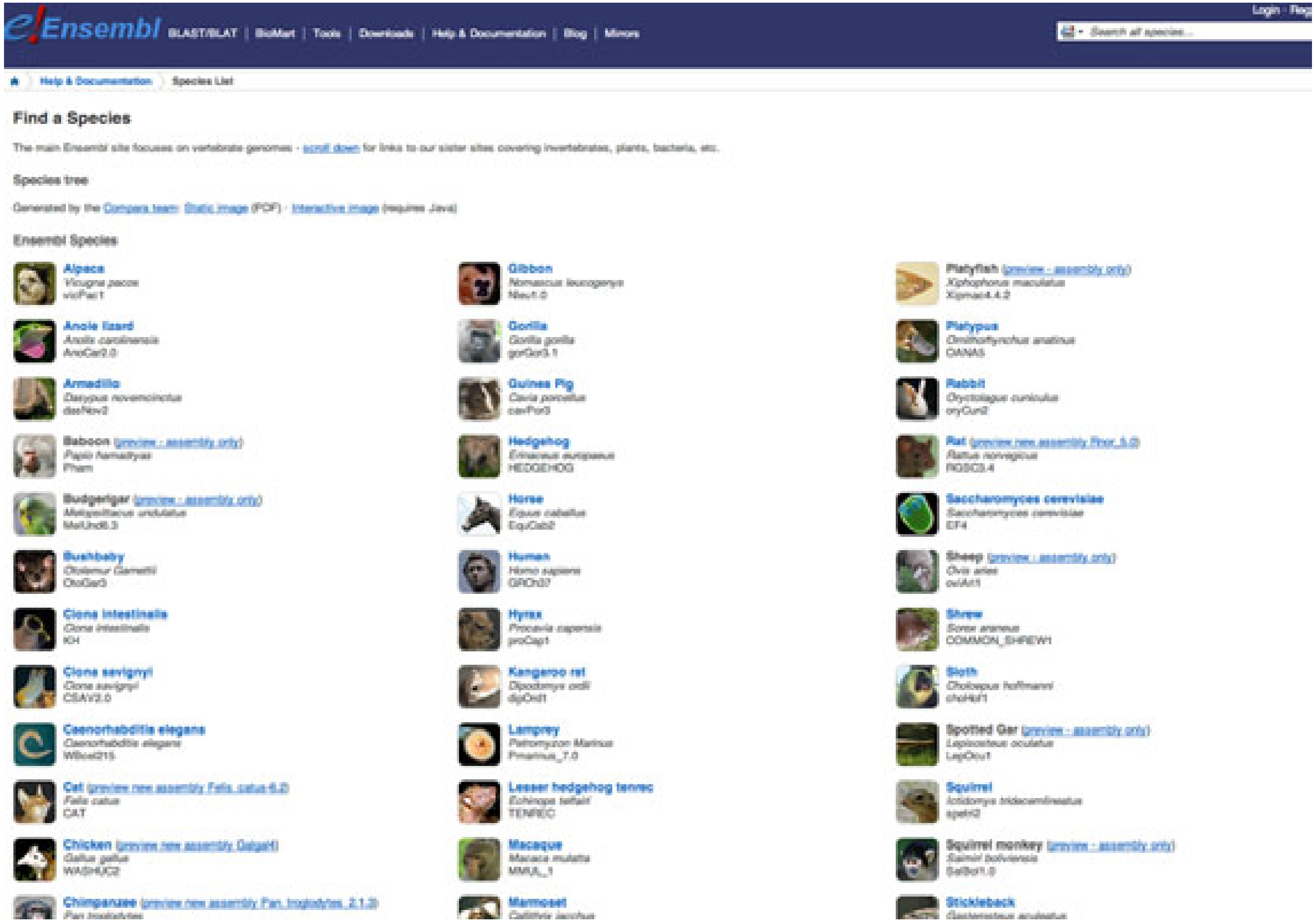
All genomes

-- Select a species --

View full list of all Ensembl species

Other species are available in Ensembl Pre! and EnsemblGenomes

Ensembl



Find a Species

The main Ensembl site focuses on vertebrate genomes - scroll down for links to our sister sites covering invertebrates, plants, bacteria, etc.

Species tree

Generated by the Compara team: Static image (PDF) · Interactive image (requires Java)

Ensembl Species

Alpaca
Vicugna pacos
vicPac1

Anole lizard
Anolis carolinensis
AnoCar2.0

Armadillo
Dasypus novemcinctus
dasNov2

Baboon (preview - assembly only)
Papio hamadryas
Pham

Budgerigar (preview - assembly only)
Melopsittacus undulatus
MelUnd6.3

Bushbaby
Otolemur Garnetti
OtoGar3

Ciona intestinalis
Ciona intestinalis
KH

Ciona savignyi
Ciona savignyi
CSAV2.0

Caenorhabditis elegans
Caenorhabditis elegans
WBcel215

Cat (preview new assembly Felis_catus-6.2)
Felis catus
CAT

Chicken (preview new assembly Galgal4)
Gallus gallus
WASHUC2

Chimpanzee (preview new assembly Pan_troglodytes_2.1.3)
Pan troglodytes

Gibbon
Nomascus leucogenys
Nleu1.0

Gorilla
Gorilla gorilla
gorGor3.1

Guinea Pig
Cavia porcellus
cavPor3

Hedgehog
Erinaceus europaeus
HEDGEHOG

Horse
Equus caballus
EquCab2

Human
Homo sapiens
GRCh37

Hyrax
Procavia capensis
proCap1

Kangaroo rat
Dipodomys ordii
dipOrd1

Lamprey
Petromyzon Marinus
Pmarinus_7.0

Lesser hedgehog tenrec
Echinops telfairi
TENREC

Macaque
Macaca mulatta
MMUL_1

Marmoset
Callithrix jacchus

Platyfish (preview - assembly only)
Xiphophorus maculatus
Xipmac4.4.2

Platypus
Ornithorhynchus anatinus
OANA5

Rabbit
Oryctolagus cuniculus
oryCun2

Rat (preview new assembly Rnor_5.0)
Rattus norvegicus
RGSC3.4

Saccharomyces cerevisiae
Saccharomyces cerevisiae
EF4

Sheep (preview - assembly only)
Ovis aries
oviAri1

Shrew
Sorex araneus
COMMON_SHREW1

Sloth
Choloepus hoffmanni
choHof1

Spotted Gar (preview - assembly only)
Lepisosteus oculatus
LepOcu1

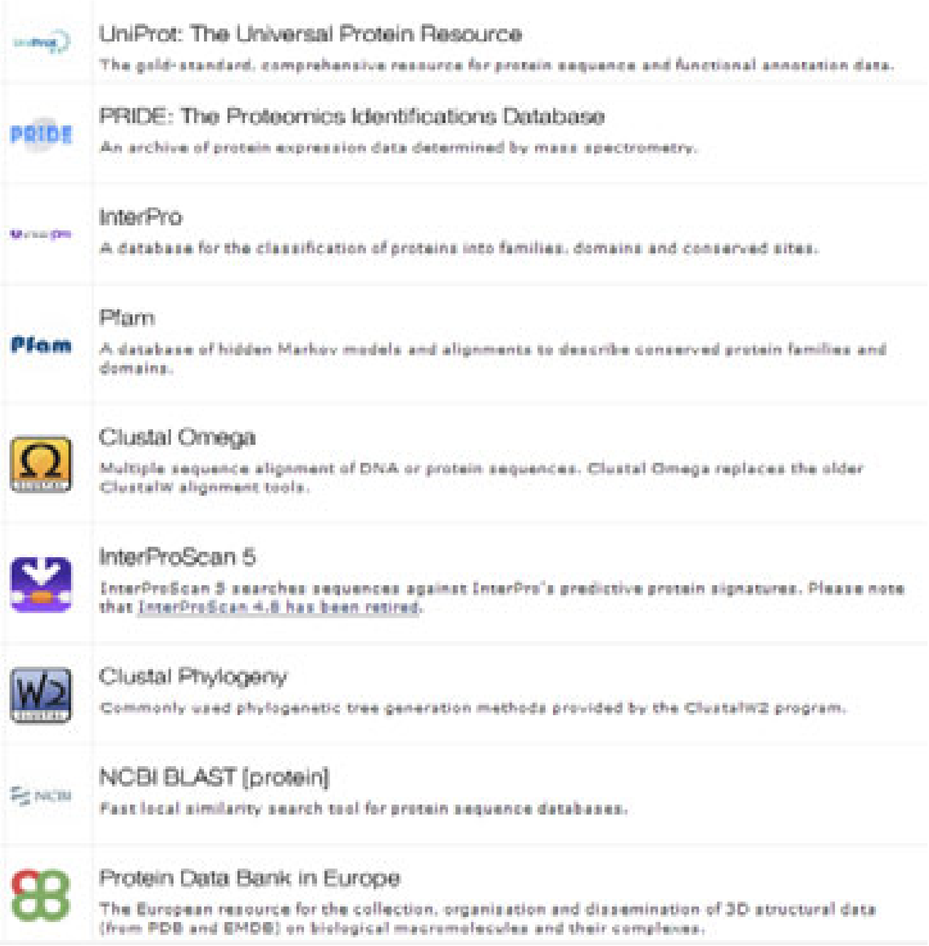
Squirrel
Ictidomys tridecemlineatus
spetri2

Squirrel monkey (preview - assembly only)
Saimiri boliviensis
SaiBol1.0

Stickleback
Gasterosteus aculeatus

Protein Databases

- Sequences are in Uniprot
- Structures are in PDB
- Enzyme classifications EC
- Protein families: Pfam, Interpro etc



UniProt: The Universal Protein Resource
The gold-standard, comprehensive resource for protein sequence and functional annotation data.

PRIDE: The Proteomics Identifications Database
An archive of protein expression data determined by mass spectrometry.

InterPro
A database for the classification of proteins into families, domains and conserved sites.

Pfam
A database of hidden Markov models and alignments to describe conserved protein families and domains.

Clustal Omega
Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

InterProScan 5
InterProScan 5 searches sequences against InterPro's predictive protein signatures. Please note that InterProScan 4.8 has been retired.

Clustal Phylogeny
Commonly used phylogenetic tree generation methods provided by the ClustalW2 program.

NCBI BLAST [protein]
Fast local similarity search tool for protein sequence databases.

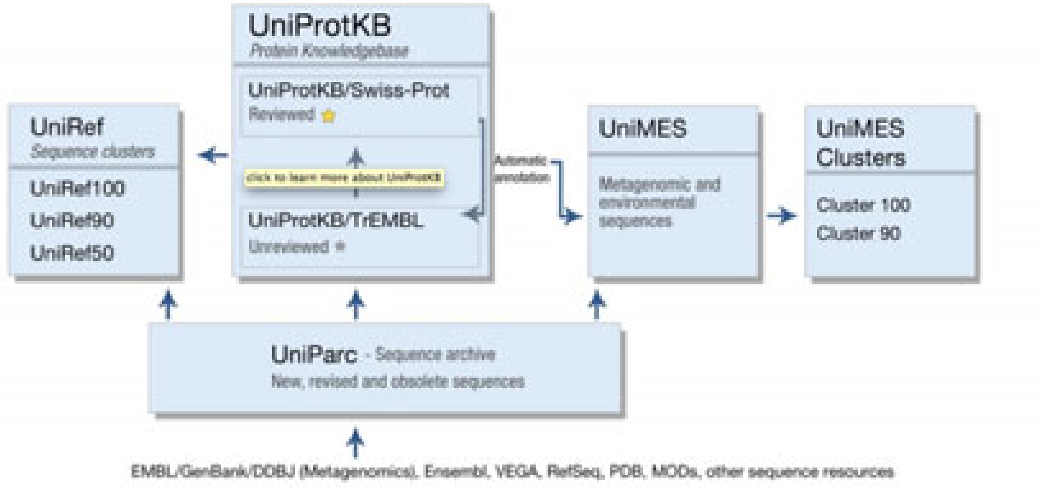
Protein Data Bank in Europe
The European resource for the collection, organisation and dissemination of 3D structural data (from PDB and EMDB) on biological macromolecules and their complexes.

Uniprot

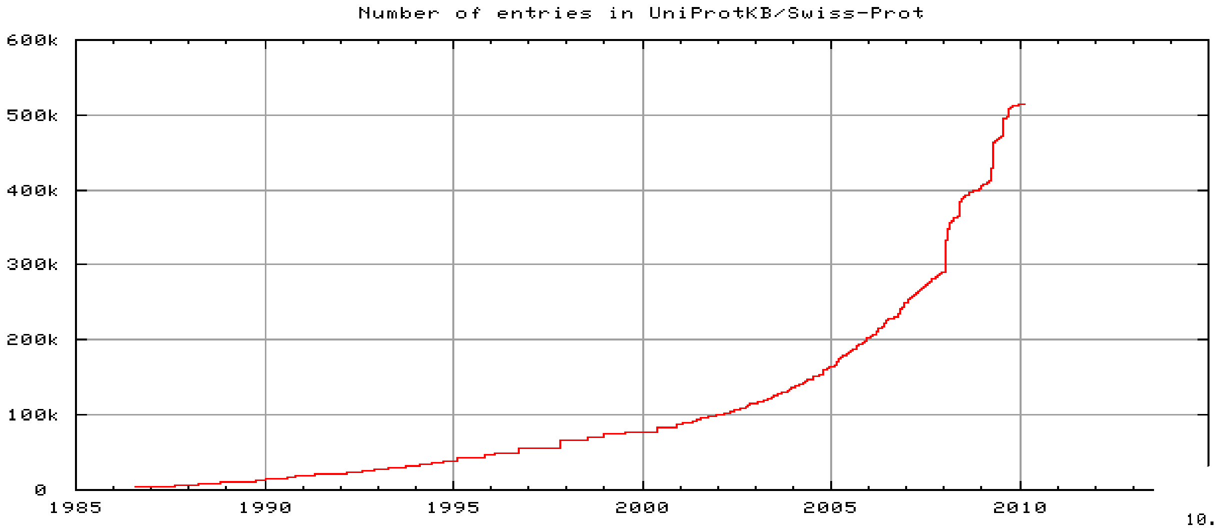
UniProtKB
Protein knowledgebase, consists of two sections:

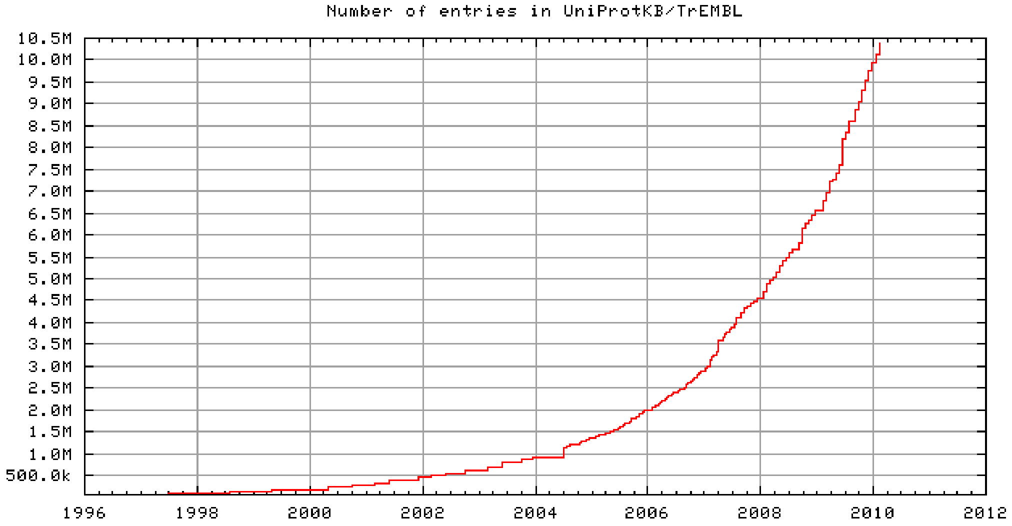
1. Swiss-Prot, which is manually annotated and reviewed.
2. TrEMBL, which is automatically annotated and is**not** reviewed.

Includes complete and reference proteome sets.
UniRef
Sequence clusters, used to speed up sequence similarity searches.
UniParc
Sequence archive, used to keep track of sequences and their identifiers.
Supporting data
Literature citations, taxonomy, keywords, subcellular locations, cross-referenced databases and more.

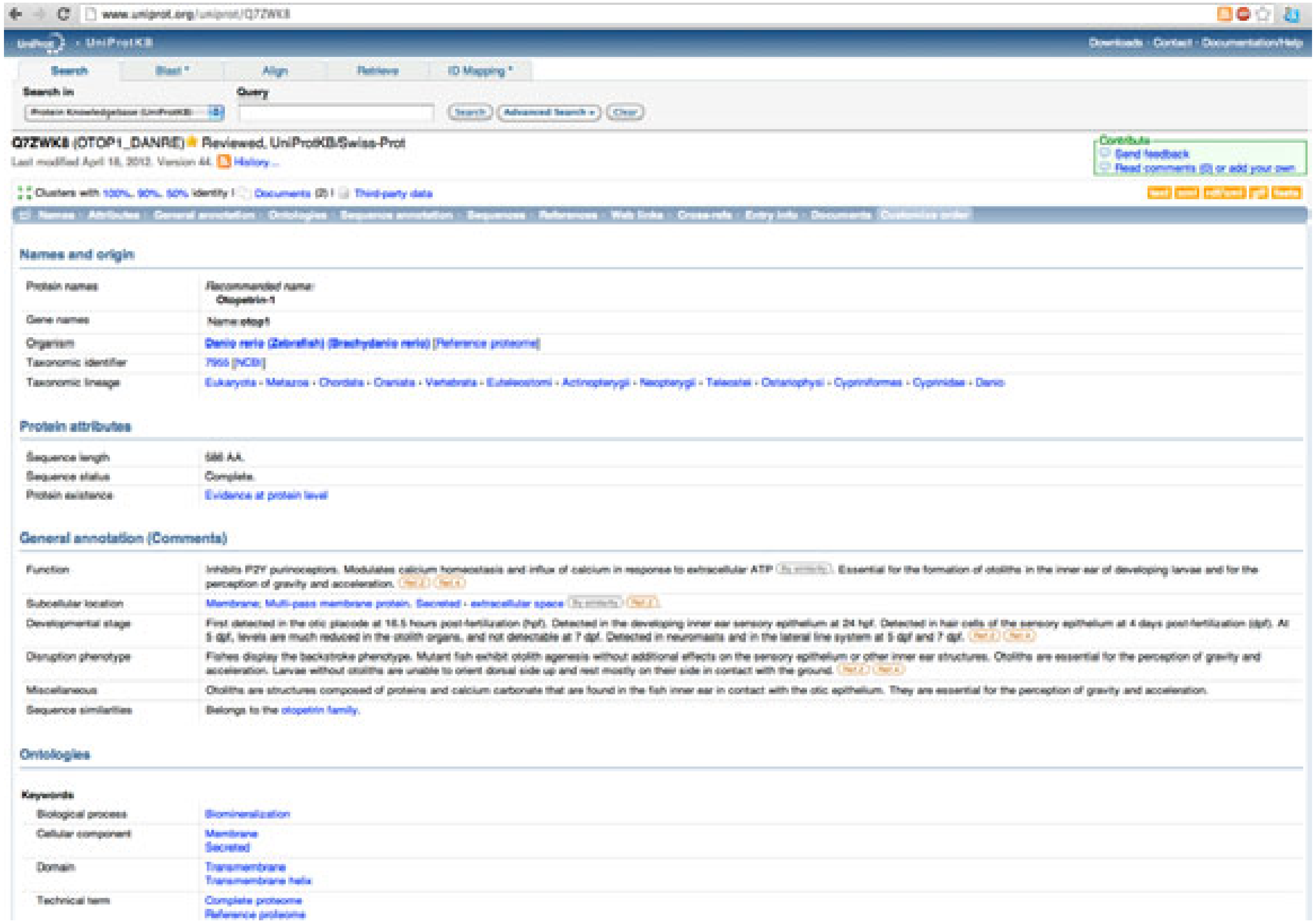


Growth of Protein sequences





One uniprot entry



Q7ZWK8 (OTOP1_DANRE) ★ Reviewed, UniProtKB/Swiss-Prot

Last modified April 18, 2012. Version 44. History...

Clusters with 100%, 90%, 50% identity | Documents (2) | Third-party data

Names and origin

Protein names	*Recommended name:* **Otopetrin-1**
Gene names	Name: **otop1**
Organism	Danio rerio (Zebrafish) (Brachydanio rerio) [Reference proteome]
Taxonomic identifier	7955 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Actinopterygii › Neopterygii › Teleostei › Ostariophysi › Cypriniformes › Cyprinidae › Danio

Protein attributes

Sequence length	586 AA.
Sequence status	Complete.
Protein existence	Evidence at protein level

General annotation (Comments)

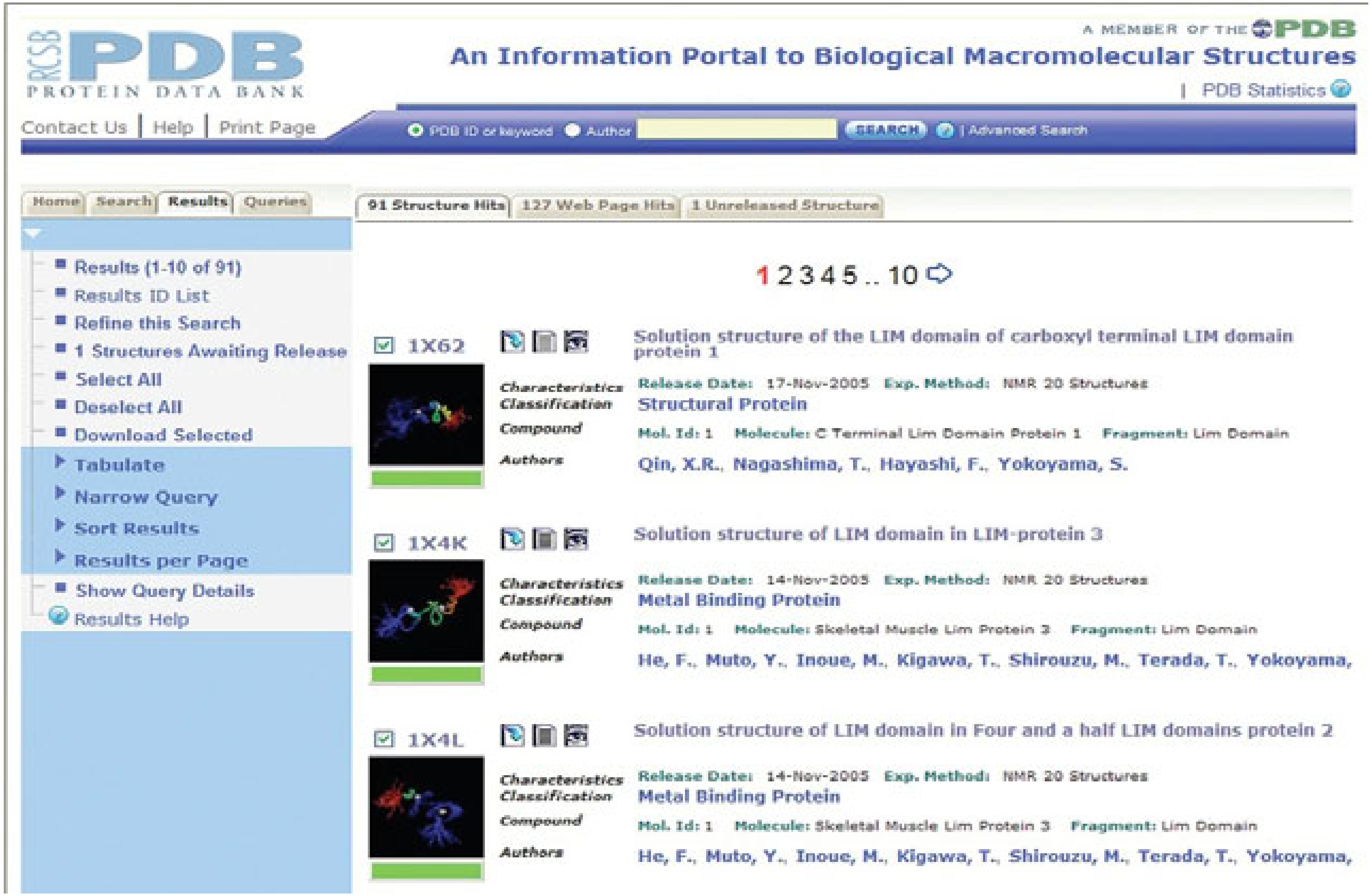
Function	Inhibits P2Y purinoceptors. Modulates calcium homeostasis and influx of calcium in response to extracellular ATP By similarity. Essential for the formation of otoliths in the inner ear of developing larvae and for the perception of gravity and acceleration. Ref.2 Ref.4
Subcellular location	Membrane; Multi-pass membrane protein. Secreted › extracellular space By similarity Ref.2.
Developmental stage	First detected in the otic placode at 18.5 hours post-fertilization (hpf). Detected in the developing inner ear sensory epithelium at 24 hpf. Detected in hair cells of the sensory epithelium at 4 days post-fertilization (dpf). At 5 dpf, levels are much reduced in the otolith organs, and not detectable at 7 dpf. Detected in neuromasts and in the lateral line system at 5 dpf and 7 dpf. Ref.2 Ref.4
Disruption phenotype	Fishes display the backstroke phenotype. Mutant fish exhibit otolith agenesis without additional effects on the sensory epithelium or other inner ear structures. Otoliths are essential for the perception of gravity and acceleration. Larvae without otoliths are unable to orient dorsal side up and rest mostly on their side in contact with the ground. Ref.2 Ref.4
Miscellaneous	Otoliths are structures composed of proteins and calcium carbonate that are found in the fish inner ear in contact with the otic epithelium. They are essential for the perception of gravity and acceleration.
Sequence similarities	Belongs to the otopetrin family.

Ontologies

Keywords

Biological process	Biomineralization
Cellular component	Membrane Secreted
Domain	Transmembrane Transmembrane helix
Technical term	Complete proteome Reference proteome

PDB



☑ **1X62**

Solution structure of the LIM domain of carboxyl terminal LIM domain protein 1

Characteristics **Release Date:** 17-Nov-2005 **Exp. Method:** NMR 20 Structures

Classification **Structural Protein**

Compound **Mol. Id:** 1 **Molecule:** C Terminal Lim Domain Protein 1 **Fragment:** Lim Domain

Authors **Qin, X.R., Nagashima, T., Hayashi, F., Yokoyama, S.**

☑ **1X4K**

Solution structure of LIM domain in LIM-protein 3

Characteristics **Release Date:** 14-Nov-2005 **Exp. Method:** NMR 20 Structures

Classification **Metal Binding Protein**

Compound **Mol. Id:** 1 **Molecule:** Skeletal Muscle Lim Protein 3 **Fragment:** Lim Domain

Authors **He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama,**

☑ **1X4L**

Solution structure of LIM domain in Four and a half LIM domains protein 2

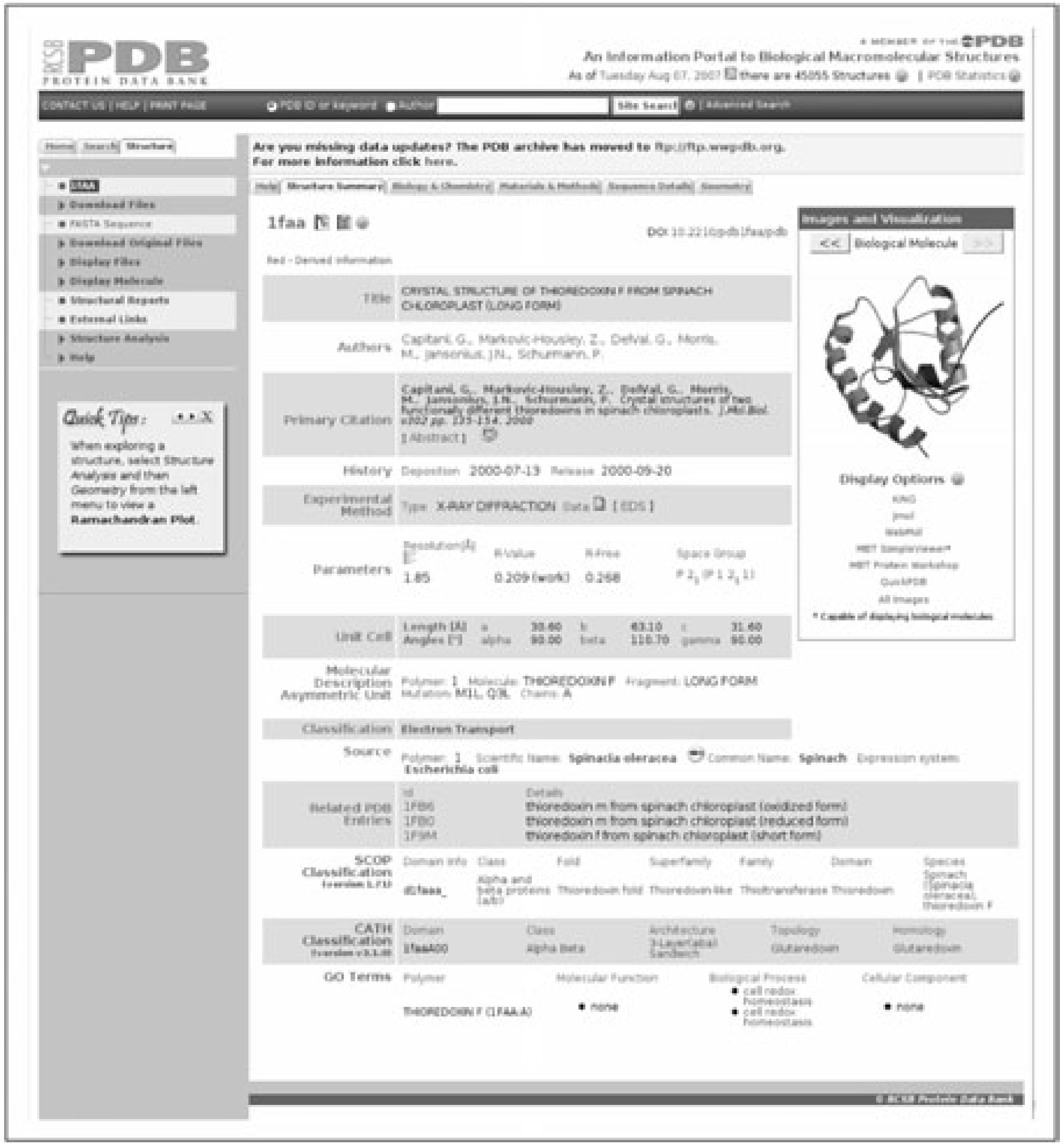
Characteristics **Release Date:** 14-Nov-2005 **Exp. Method:** NMR 20 Structures

Classification **Metal Binding Protein**

Compound **Mol. Id:** 1 **Molecule:** Skeletal Muscle Lim Protein 3 **Fragment:** Lim Domain

Authors **He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama,**

PDB



Are you missing data updates? The PDB archive has moved to ftp://ftp.wwpdb.org.
For more information click here.

Quick Tips:

When exploring a structure, select Structure Analysis and then Geometry from the left menu to view a **Ramachandran Plot**.

1faa

DOI: 10.2210/pdb1faa/pdb

Red - Derived information

Title	CRYSTAL STRUCTURE OF THIOREDOXIN F FROM SPINACH CHLOROPLAST (LONG FORM)
Authors	Capitani, G., Markovic-Housley, Z., DelVal, G., Morris, M., Jansonius, J.N., Schurmann, P.
Primary Citation	Capitani, G., Markovic-Housley, Z., DelVal, G., Morris, M., Jansonius, J.N., Schurmann, P. Crystal structures of two functionally different thioredoxins in spinach chloroplasts. J.Mol.Biol. v302 pp. 135-154, 2000 [Abstract]
History	Deposition 2000-07-13 Release 2000-09-20
Experimental Method	Type X-RAY DIFFRACTION Data [EDS]
Parameters	Resolution[Å] 1.85 · R-Value 0.209 (work) · R-Free 0.268 · Space Group $P\ 2_1$ $(P\ 1\ 2_1\ 1)$
Unit Cell	Length [Å] a 30.60 b 63.10 c 31.60 Angles [°] alpha 90.00 beta 110.70 gamma 90.00
Molecular Description Asymmetric Unit	Polymer: 1 Molecule: THIOREDOXIN F Fragment: LONG FORM Mutation: M1L, Q3L Chains: A
Classification	**Electron Transport**
Source	Polymer: 1 Scientific Name: **Spinacia oleracea** Common Name: **Spinach** Expression system: **Escherichia coli**

Related PDB Entries

Id	Details
1FB6	thioredoxin m from spinach chloroplast (oxidized form)
1FB0	thioredoxin m from spinach chloroplast (reduced form)
1F9M	thioredoxin f from spinach chloroplast (short form)

SCOP Classification (version 1.73)

Domain info	Class	Fold	Superfamily	Family	Domain	Species
d1faa_	Alpha and beta proteins (a/b)	Thioredoxin fold	Thioredoxin-like	Thioltransferase	Thioredoxin	Spinach (Spinacia oleracea), thioredoxin F

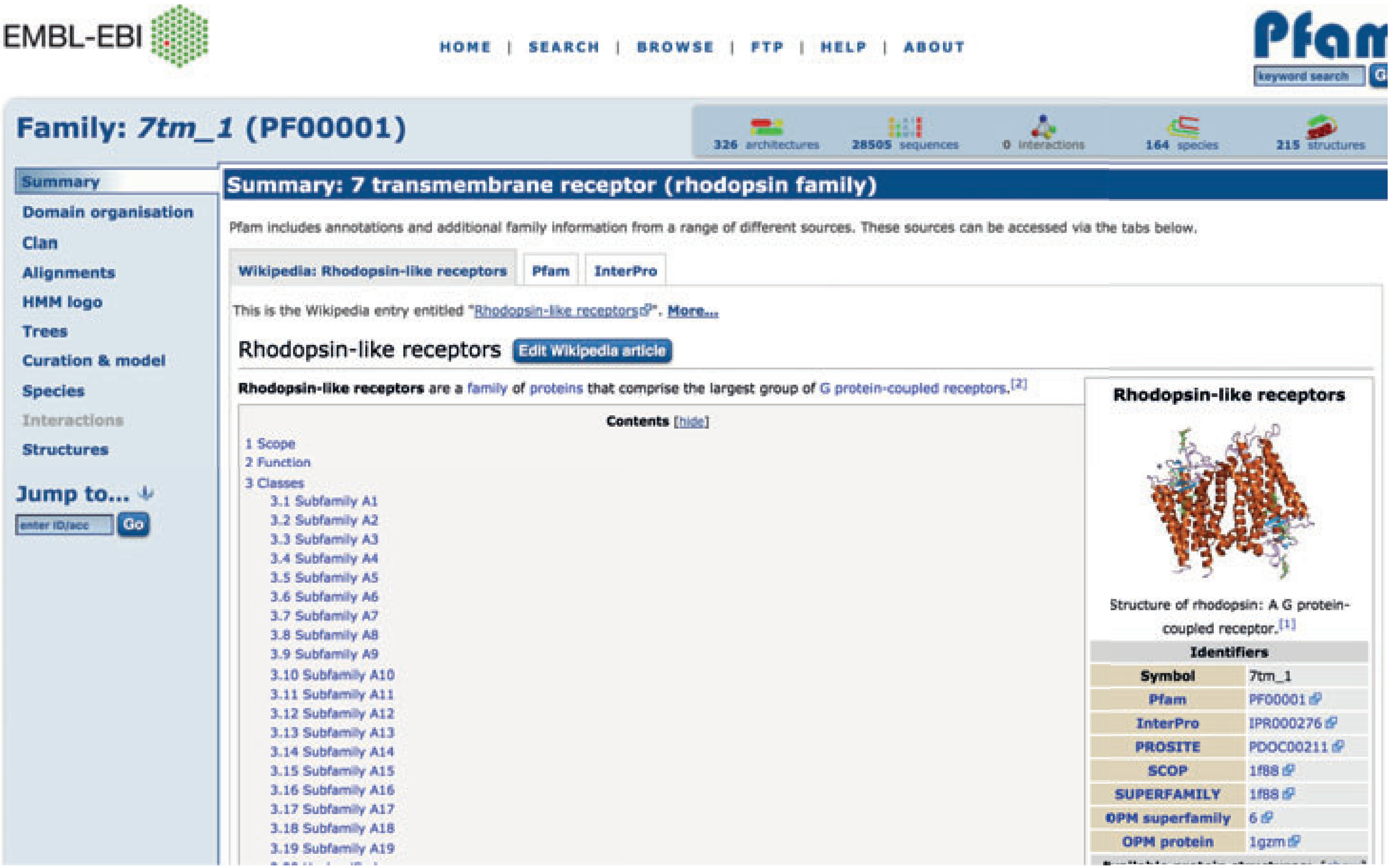
CATH Classification (version v3.1.0)

Domain	Class	Architecture	Topology	Homology
1faaA00	Alpha Beta	3-Layer(aba) Sandwich	Glutaredoxin	Glutaredoxin

GO Terms

Polymer	Molecular Function	Biological Process	Cellular Component
THIOREDOXIN F (1FAA:A)	none	cell redox homeostasis cell redox homeostasis	none

Pfam



Family: *7tm_1* (PF00001)

Summary: 7 transmembrane receptor (rhodopsin family)

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

This is the Wikipedia entry entitled "Rhodopsin-like receptors". **More...**

Rhodopsin-like receptors

Rhodopsin-like receptors are a family of proteins that comprise the largest group of G protein-coupled receptors.[2]

Contents [hide]

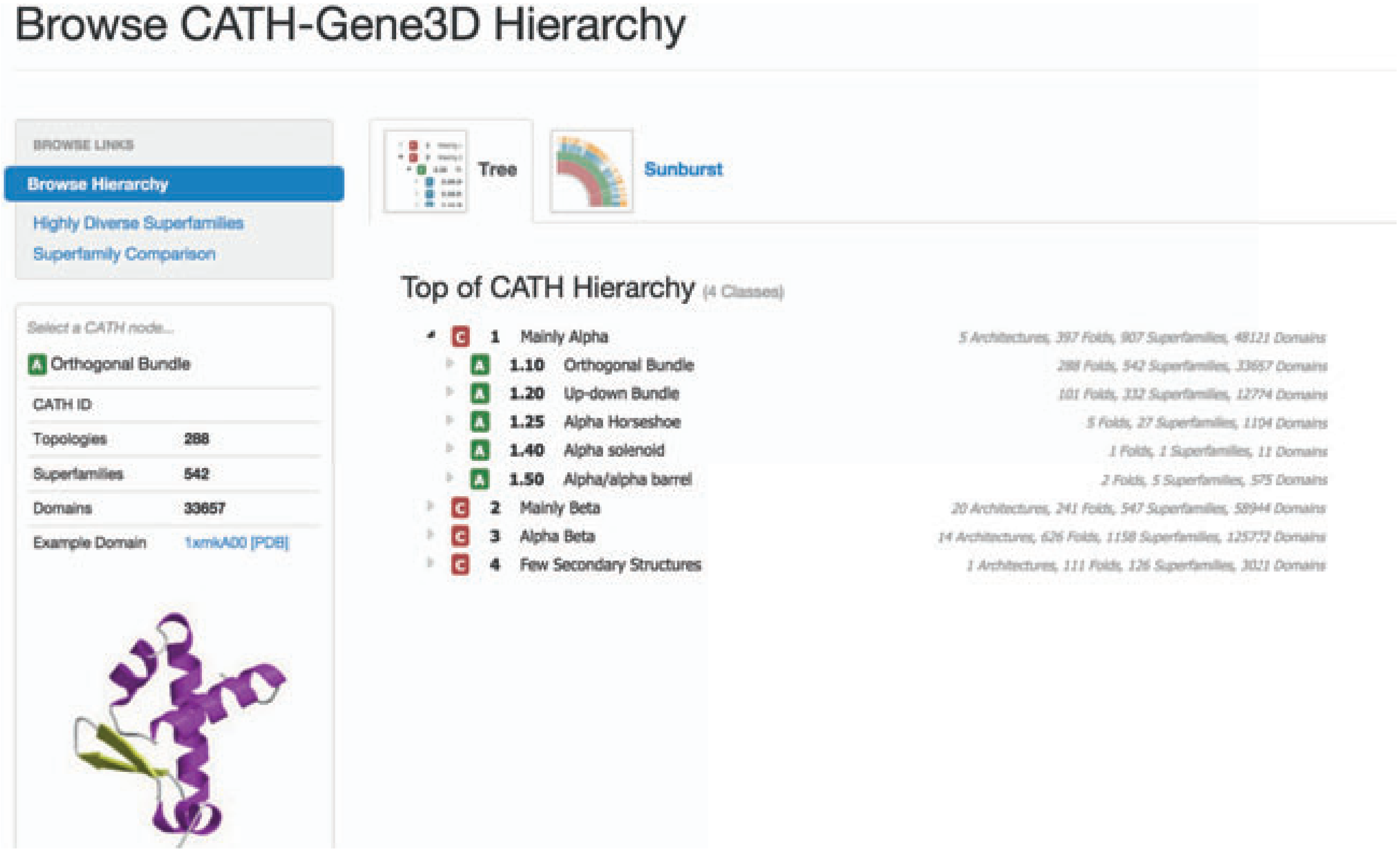
1 Scope
2 Function
3 Classes
    3.1 Subfamily A1
    3.2 Subfamily A2
    3.3 Subfamily A3
    3.4 Subfamily A4
    3.5 Subfamily A5
    3.6 Subfamily A6
    3.7 Subfamily A7
    3.8 Subfamily A8
    3.9 Subfamily A9
    3.10 Subfamily A10
    3.11 Subfamily A11
    3.12 Subfamily A12
    3.13 Subfamily A13
    3.14 Subfamily A14
    3.15 Subfamily A15
    3.16 Subfamily A16
    3.17 Subfamily A17
    3.18 Subfamily A18
    3.19 Subfamily A19

Rhodopsin-like receptors

Structure of rhodopsin: A G protein-coupled receptor.[1]

Identifiers	
Symbol	7tm_1
Pfam	PF00001
InterPro	IPR000276
PROSITE	PDOC00211
SCOP	1f88
SUPERFAMILY	1f88
OPM superfamily	6
OPM protein	1gzm

CATH (and related databases such as SCOP, ECOD)



Browse CATH-Gene3D Hierarchy

BROWSE LINKS

- **Browse Hierarchy**
- Highly Diverse Superfamilies
- Superfamily Comparison

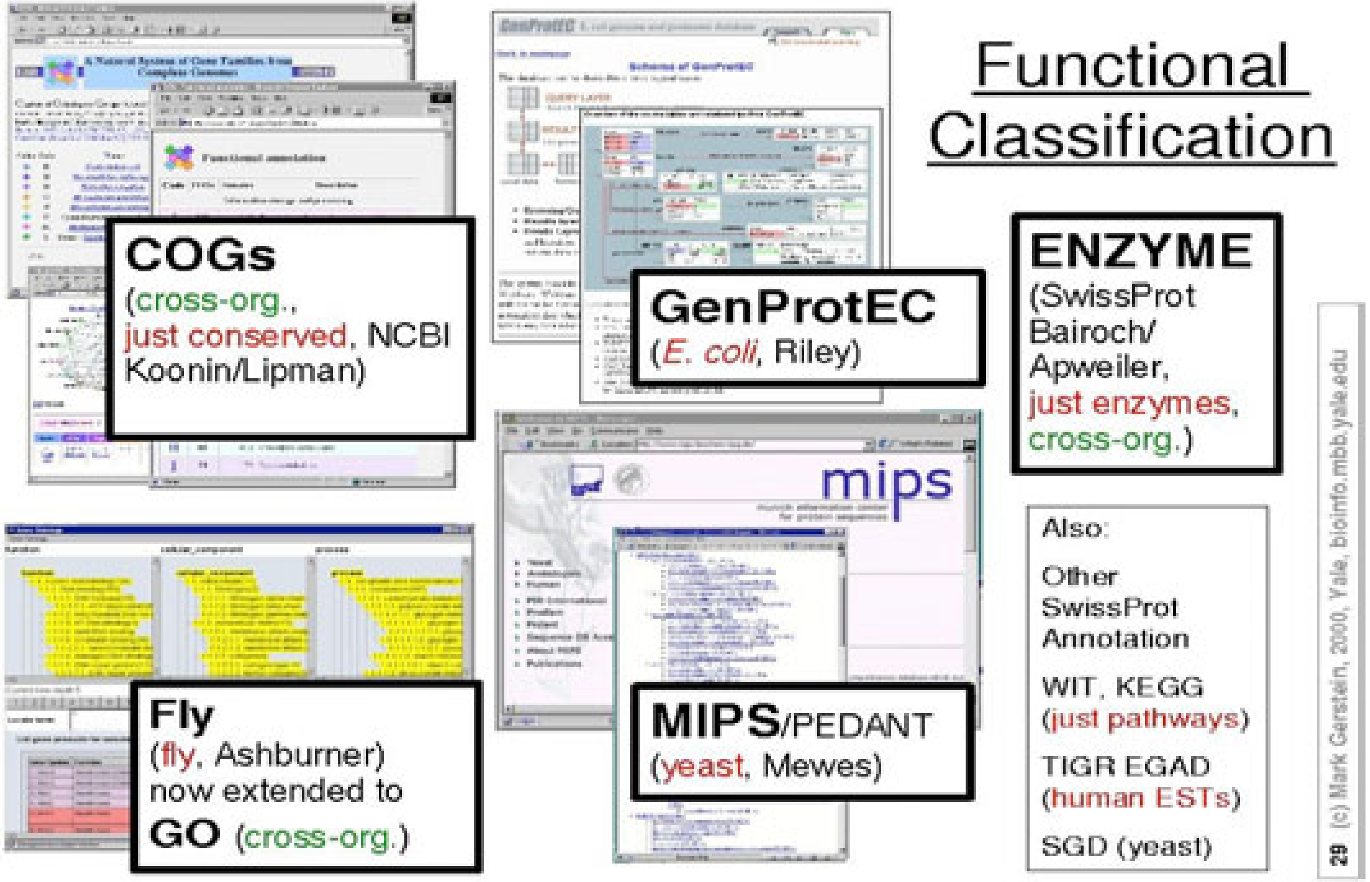
Select a CATH node...

A Orthogonal Bundle

CATH ID	
Topologies	288
Superfamilies	542
Domains	33657
Example Domain	1xmkA00 [PDB]

Top of CATH Hierarchy (4 Classes)

C	1	Mainly Alpha	5 Architectures, 397 Folds, 907 Superfamilies, 48121 Domains
A	1.10	Orthogonal Bundle	288 Folds, 542 Superfamilies, 33657 Domains
A	1.20	Up-down Bundle	101 Folds, 332 Superfamilies, 12774 Domains
A	1.25	Alpha Horseshoe	5 Folds, 27 Superfamilies, 1104 Domains
A	1.40	Alpha solenoid	1 Folds, 1 Superfamilies, 11 Domains
A	1.50	Alpha/alpha barrel	2 Folds, 5 Superfamilies, 575 Domains
C	2	Mainly Beta	20 Architectures, 241 Folds, 547 Superfamilies, 58944 Domains
C	3	Alpha Beta	14 Architectures, 626 Folds, 1158 Superfamilies, 125772 Domains
C	4	Few Secondary Structures	1 Architectures, 111 Folds, 126 Superfamilies, 3021 Domains



Functional Classification

COGs
(cross-org.,
just conserved, NCBI
Koonin/Lipman)

GenProtEC
(*E. coli*, Riley)

ENZYME
(SwissProt
Bairoch/
Apweiler,
just enzymes,
cross-org.)

Fly
(fly, Ashburner)
now extended to
GO (cross-org.)

MIPS/PEDANT
(yeast, Mewes)

Also:

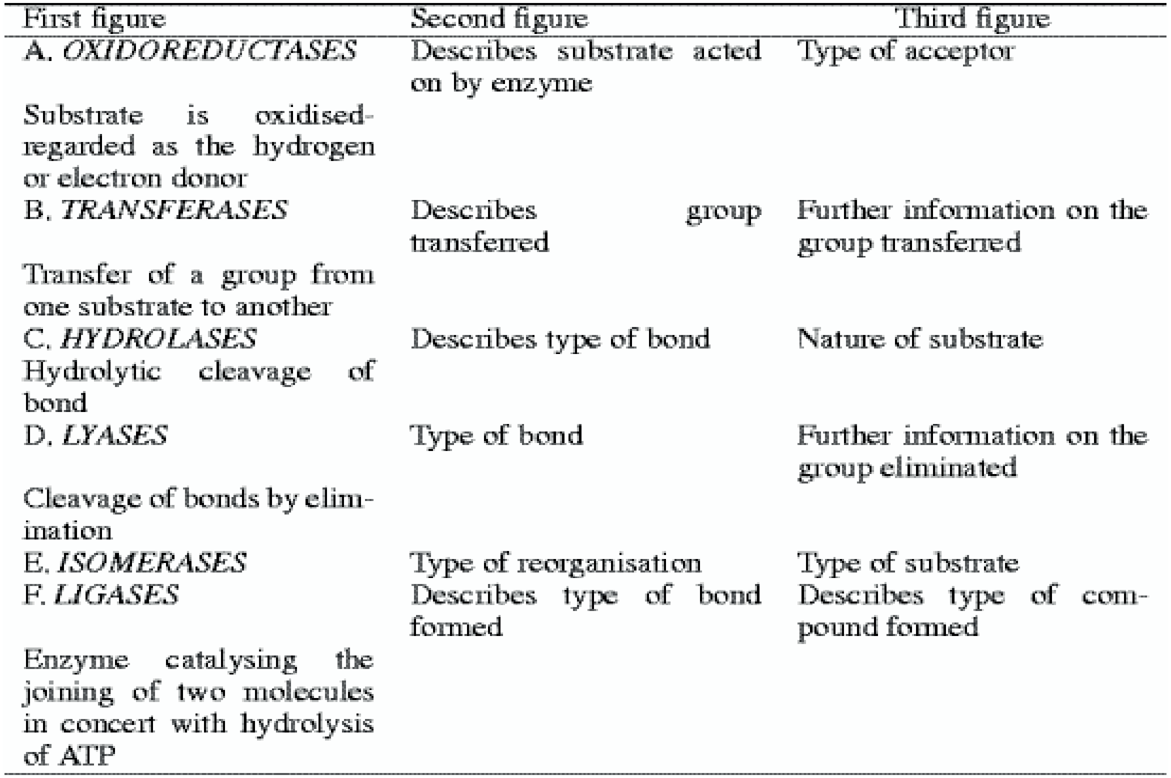
Other
SwissProt
Annotation

WIT, KEGG
(just pathways)

TIGR EGAD
(human ESTs)

SGD (yeast)

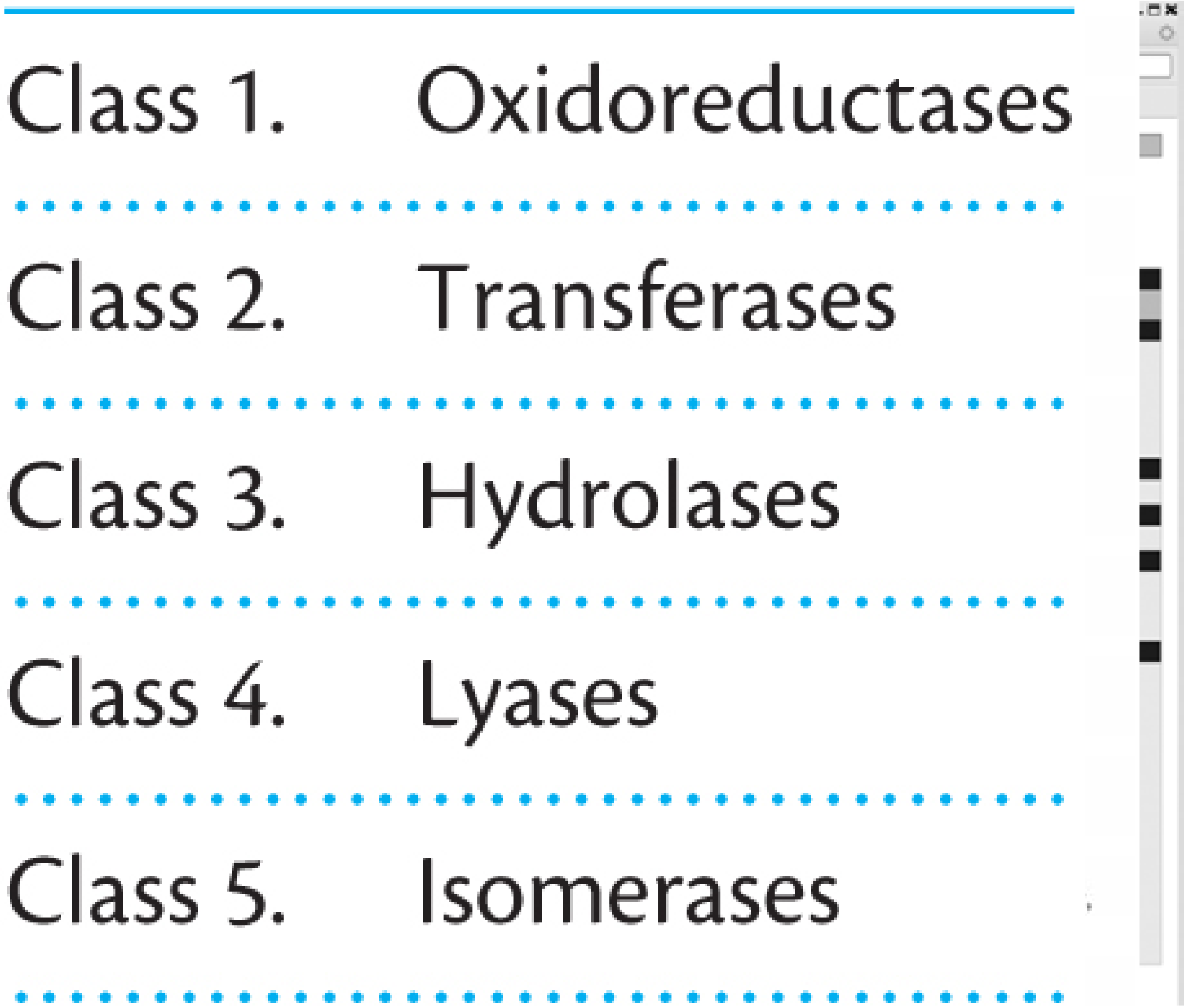
Functional classifications



First figure	Second figure	Third figure
A. *OXIDOREDUCTASES* Substrate is oxidised-regarded as the hydrogen or electron donor	Describes substrate acted on by enzyme	Type of acceptor
B. *TRANSFERASES* Transfer of a group from one substrate to another	Describes group transferred	Further information on the group transferred
C. *HYDROLASES* Hydrolytic cleavage of bond	Describes type of bond	Nature of substrate
D. *LYASES* Cleavage of bonds by elimination	Type of bond	Further information on the group eliminated
E. *ISOMERASES*	Type of reorganisation	Type of substrate
F. *LIGASES* Enzyme catalysing the joining of two molecules in concert with hydrolysis of ATP	Describes type of bond formed	Describes type of compound formed

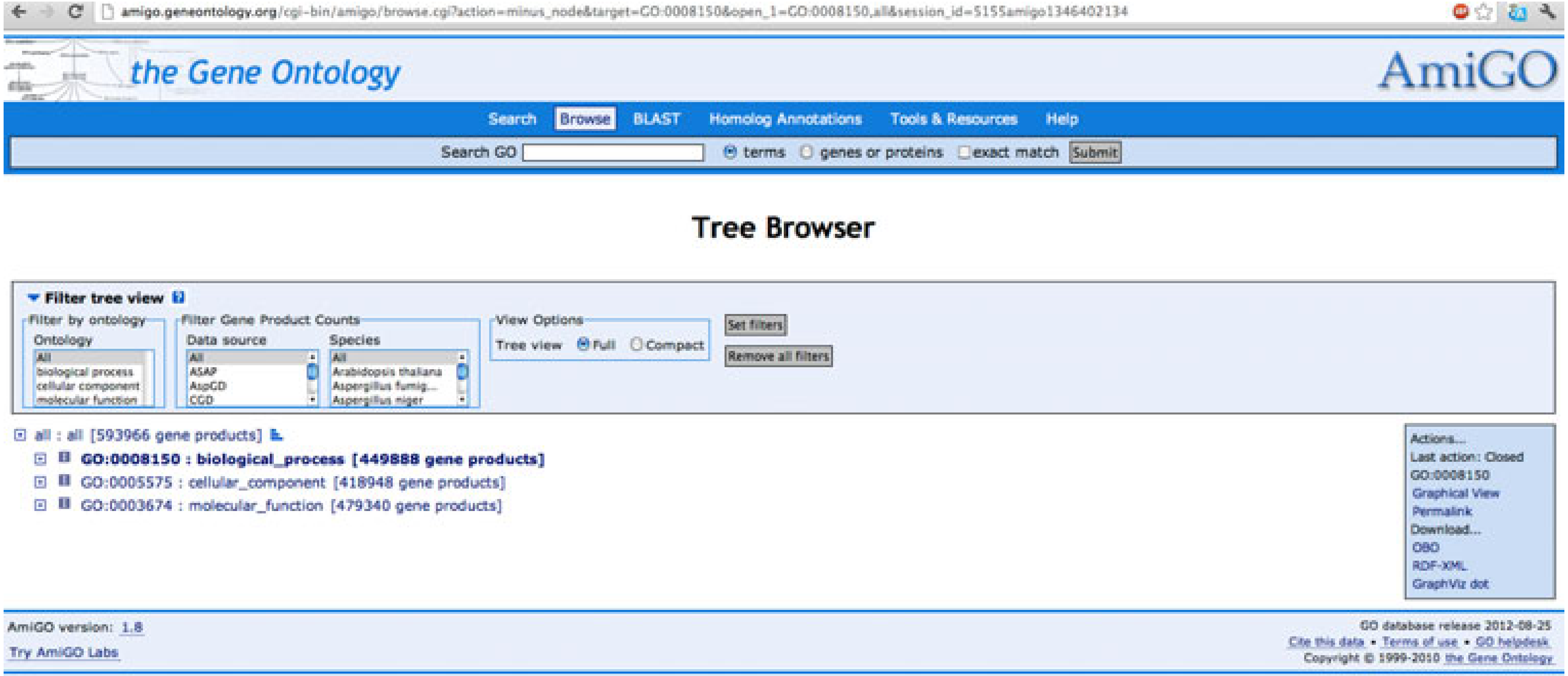
EC classifications

Enzyme Classifications



Class 1.	Oxidoreductases
Class 2.	Transferases
Class 3.	Hydrolases
Class 4.	Lyases
Class 5.	Isomerases

Function classifications: GO



amigo.geneontology.org/cgi-bin/amigo/browse.cgi?action=minus_node&target=GO:0008150&open_1=GO:0008150,all&session_id=5155amigo1346402134

Search GO ○ terms ○ genes or proteins ☐ exact match

Tree Browser

Filter tree view

Filter by ontology
Ontology: All, biological process, cellular component, molecular function

Filter Gene Product Counts
Data source: All, ASAP, AspGD, CGD
Species: All, Arabidopsis thaliana, Aspergillus fumig..., Aspergillus niger

View Options
Tree view ● Full ○ Compact

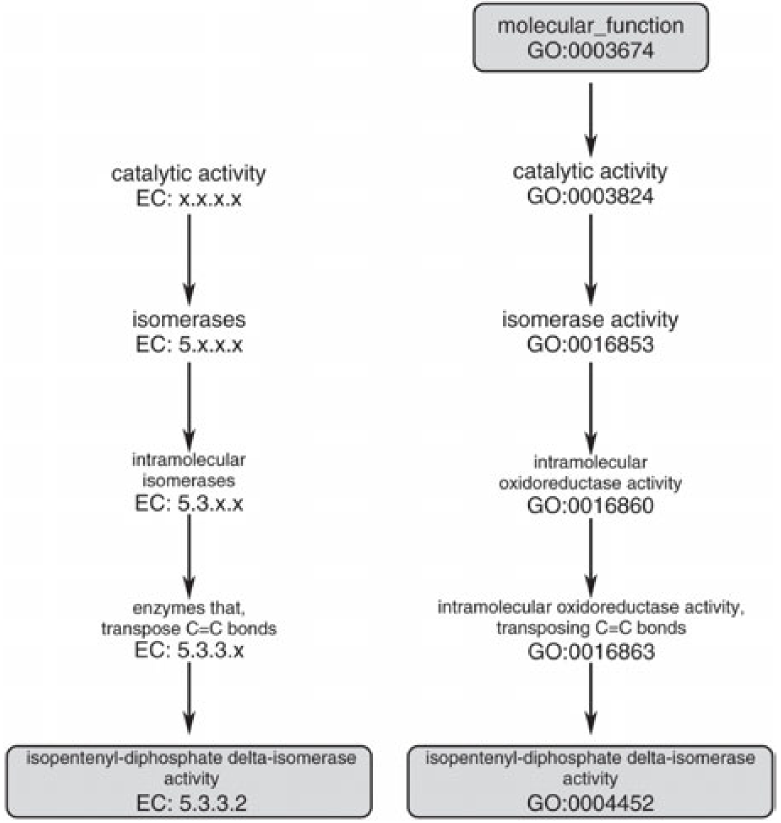
- all : all [593966 gene products]
  - **GO:0008150 : biological_process [449888 gene products]**
  - GO:0005575 : cellular_component [418948 gene products]
  - GO:0003674 : molecular_function [479340 gene products]

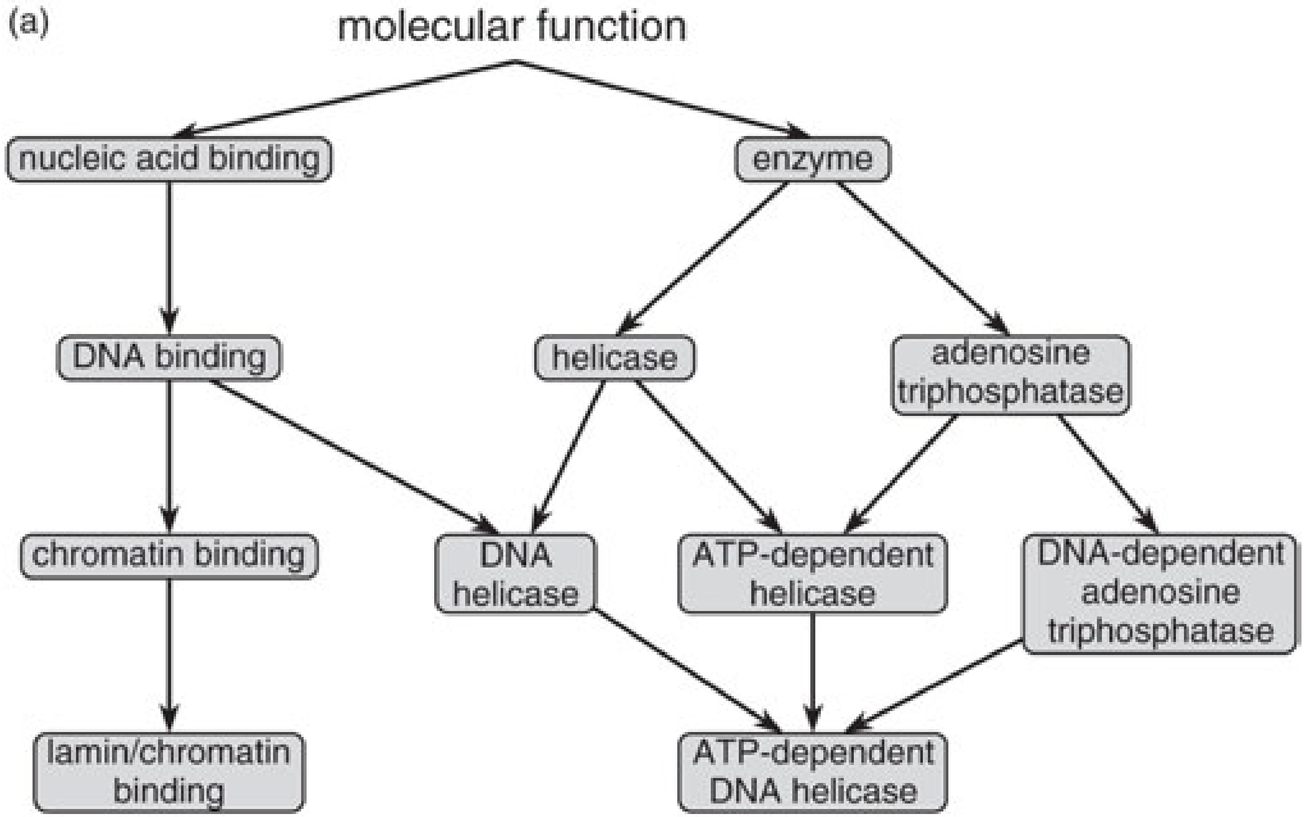
Actions...
Last action: Closed GO:0008150
Graphical View
Permalink
Download...
OBO
RDF-XML
GraphViz dot

AmiGO version: 1.8
Try AmiGO Labs

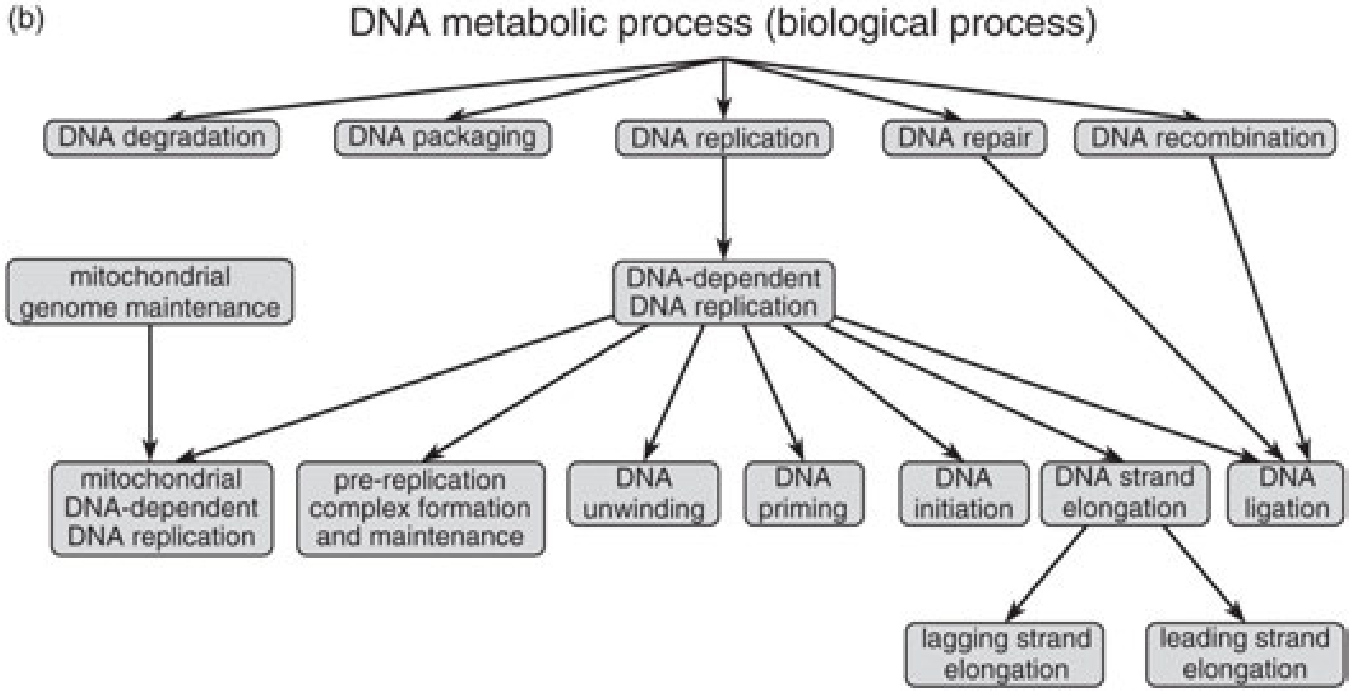
GO database release 2012-08-25
Cite this data • Terms of use • GO helpdesk
Copyright © 1999-2010 the Gene Ontology

GO - molecular function

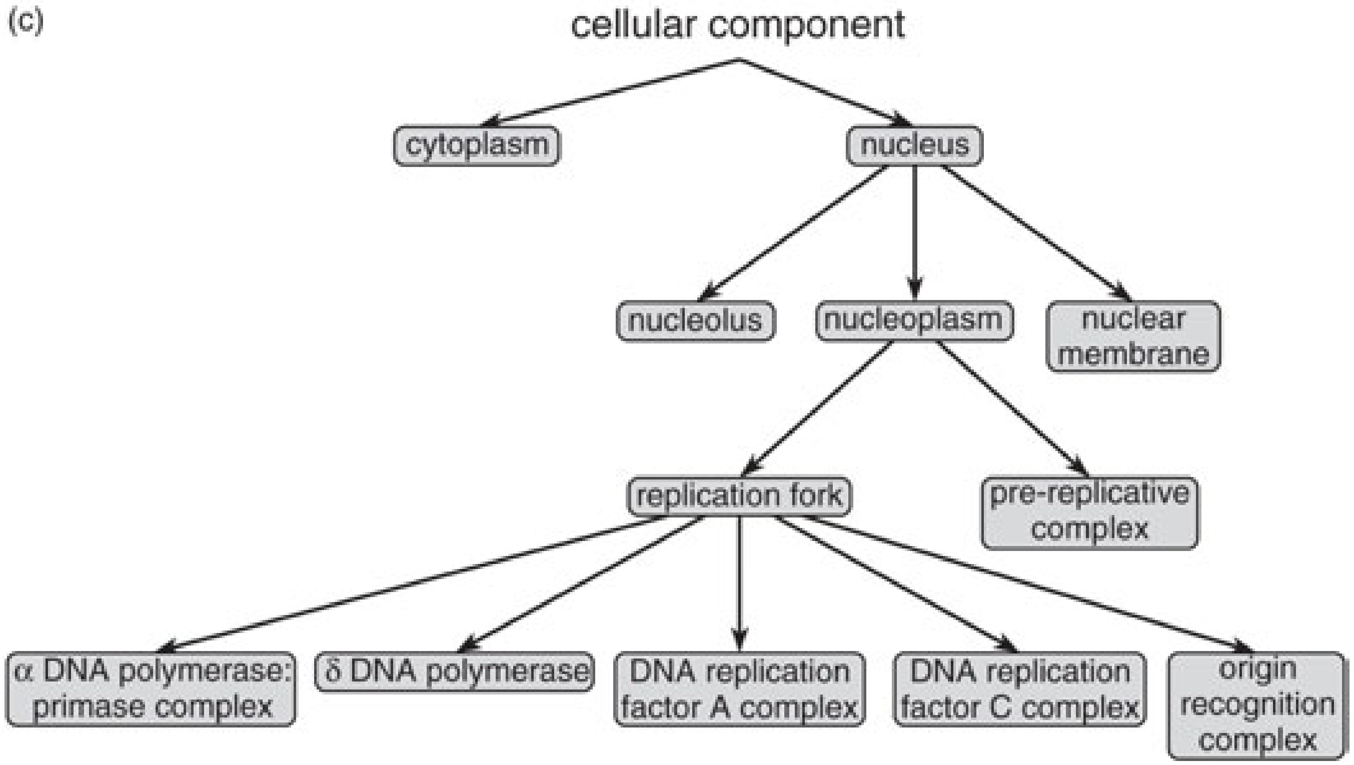




GO metabolic process



GO cellular component



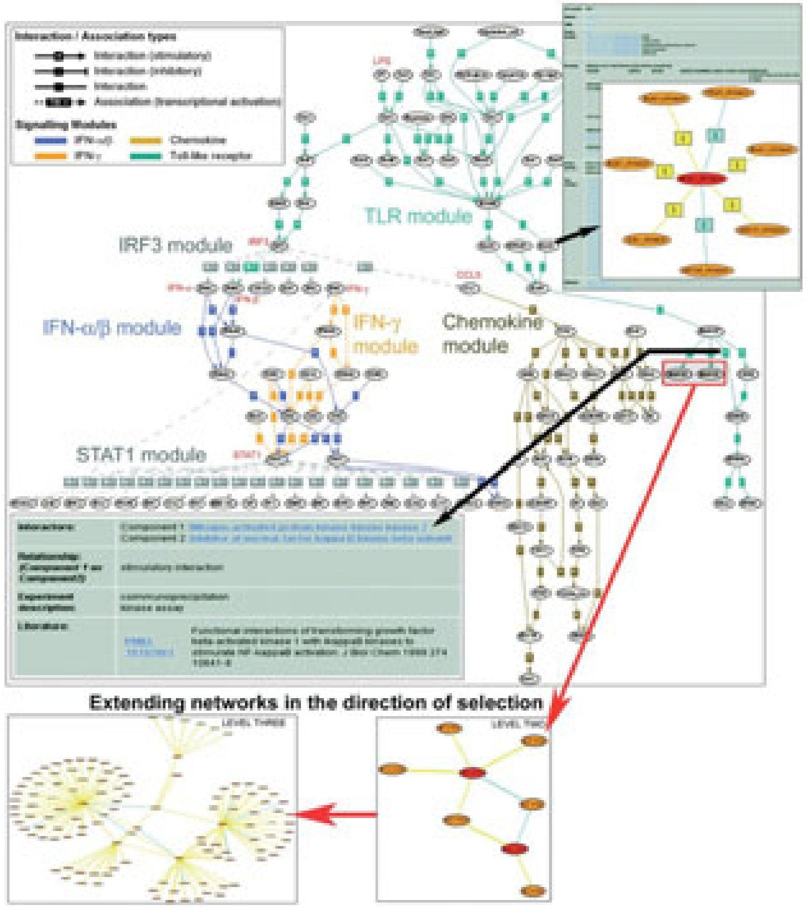
Expression databases

- Gene Expression, GeneExpressAtlas
  - total experiments 3476
  - total genes 692307
- Protein Expression: Pride
  - 25,816 Experiments
  - 11,042,722 Identified Proteins

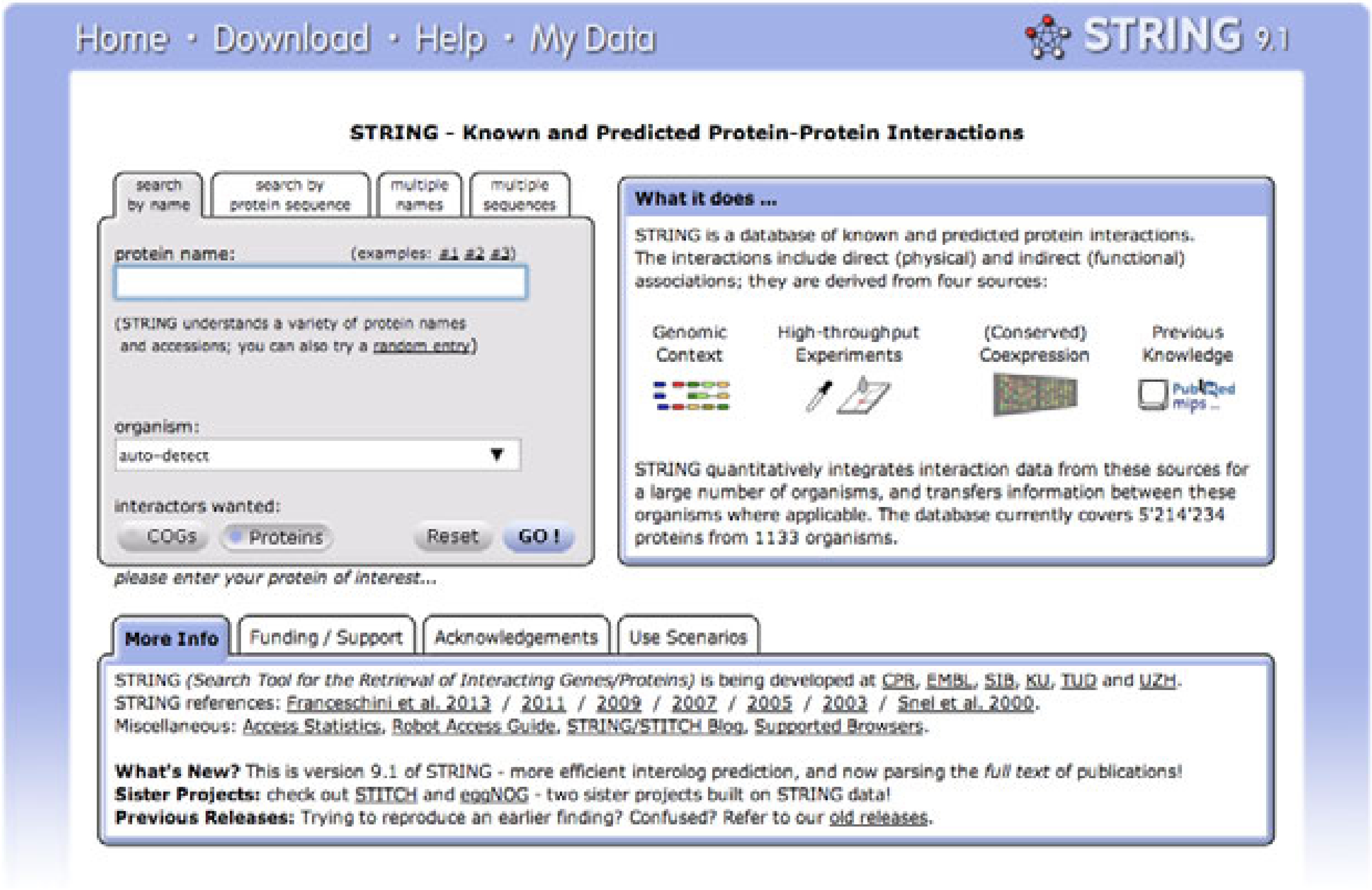
Molecular Interactions

- Intact:

Interacting maps



String Database



STRING - Known and Predicted Protein-Protein Interactions

(STRING understands a variety of protein names and accessions; you can also try a random entry)

please enter your protein of interest...

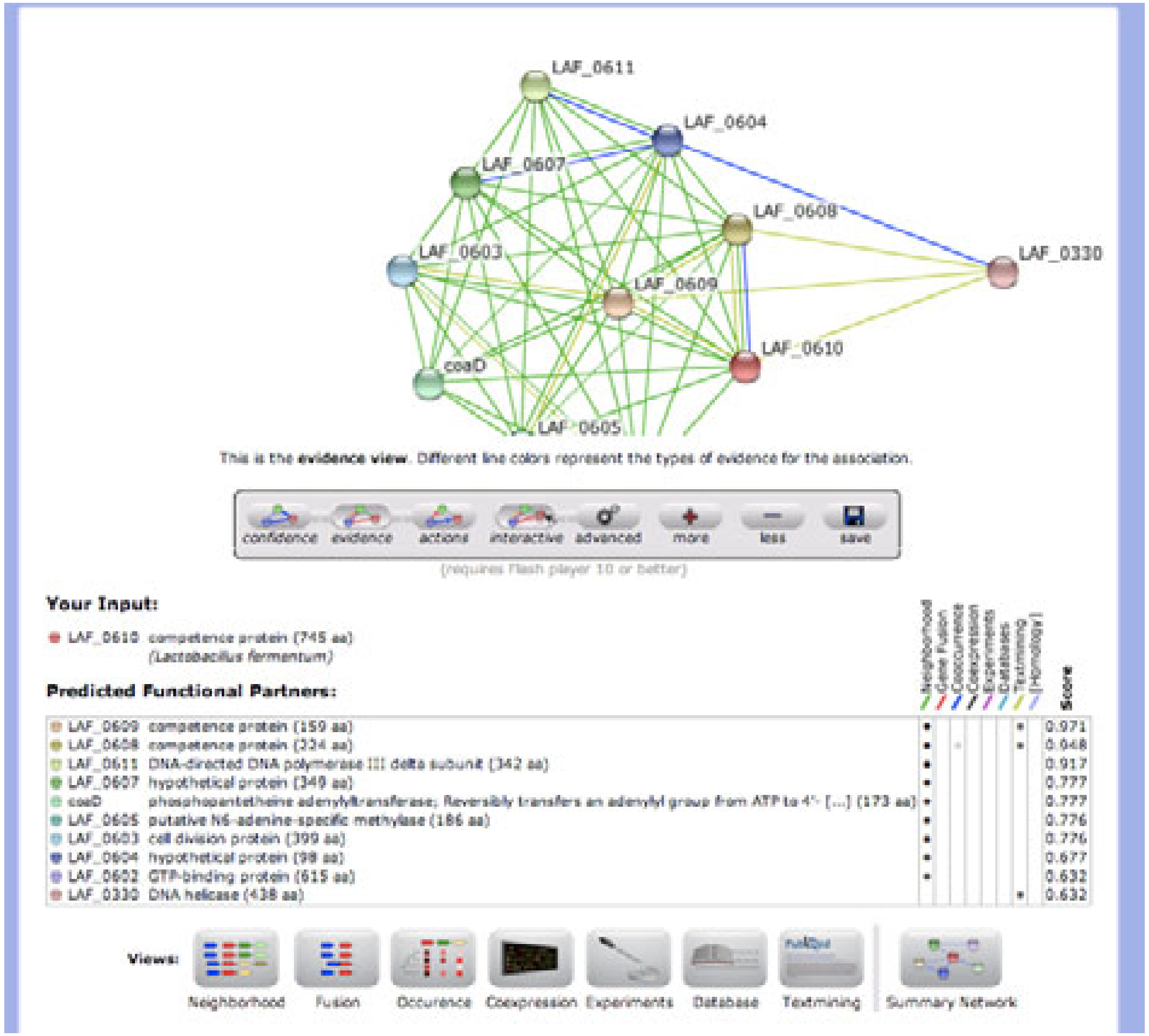
What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

STRING *(Search Tool for the Retrieval of Interacting Genes/Proteins)* is being developed at CPR, EMBL, SIB, KU, TUD and UZH.
STRING references: Franceschini et al. 2013 / 2011 / 2009 / 2007 / 2005 / 2003 / Snel et al. 2000.
Miscellaneous: Access Statistics, Robot Access Guide, STRING/STITCH Blog, Supported Browsers.

What's New? This is version 9.1 of STRING - more efficient interolog prediction, and now parsing the *full text* of publications!
Sister Projects: check out STITCH and eggNOG - two sister projects built on STRING data!
Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our old releases.



This is the **evidence view**. Different line colors represent the types of evidence for the association.

(requires Flash player 10 or better)

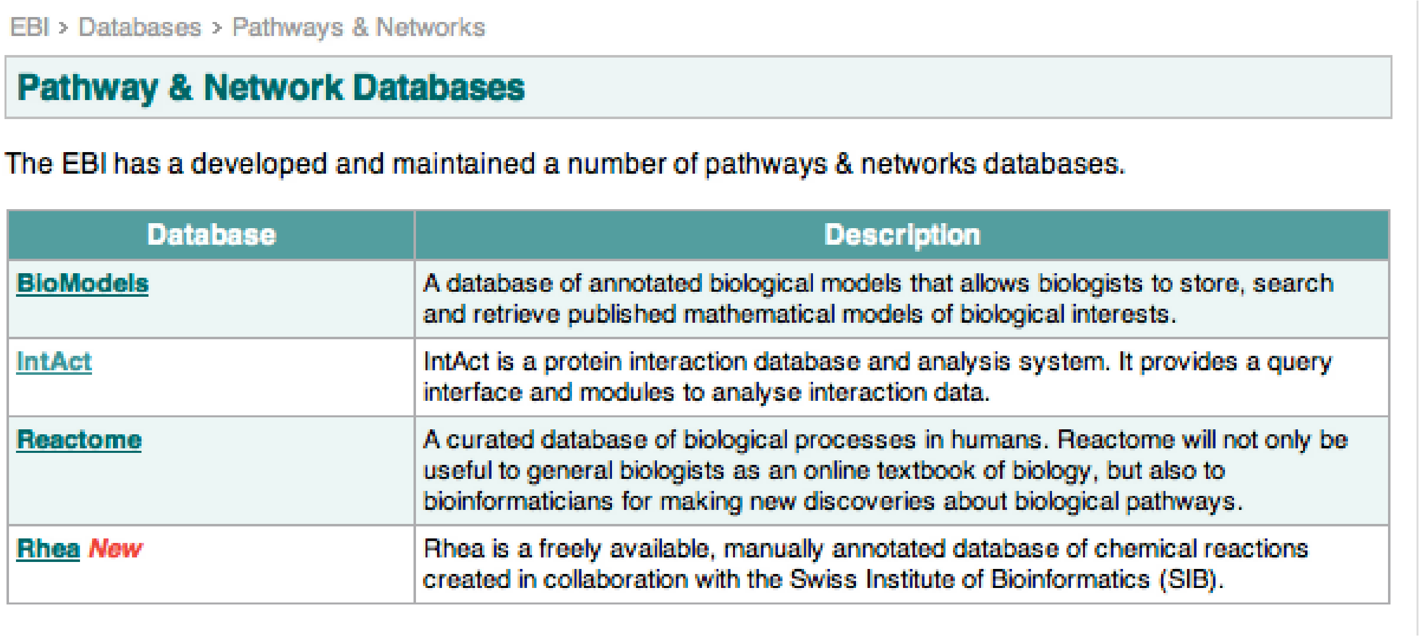
Your Input:

● LAF_0610 competence protein (745 aa)
(Lactobacillus fermentum)

Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
LAF_0609	competence protein (159 aa)	•						•		0.971
LAF_0608	competence protein (334 aa)	•		·				•		0.948
LAF_0611	DNA-directed DNA polymerase III delta subunit (342 aa)	•								0.917
LAF_0607	hypothetical protein (349 aa)	•								0.777
coaD	phosphopantetheine adenylyltransferase; Reversibly transfers an adenylyl group from ATP to 4'- [...] (173 aa)	•								0.777
LAF_0605	putative N6-adenine-specific methylase (186 aa)	•								0.776
LAF_0603	cell division protein (399 aa)	•								0.776
LAF_0604	hypothetical protein (98 aa)	•								0.677
LAF_0602	GTP-binding protein (615 aa)	•								0.632
LAF_0330	DNA helicase (438 aa)							•		0.632

Reactions and Pathways



EBI > Databases > Pathways & Networks

Pathway & Network Databases

The EBI has a developed and maintained a number of pathways & networks databases.

Database	Description
BioModels	A database of annotated biological models that allows biologists to store, search and retrieve published mathematical models of biological interests.
IntAct	IntAct is a protein interaction database and analysis system. It provides a query interface and modules to analyse interaction data.
Reactome	A curated database of biological processes in humans. Reactome will not only be useful to general biologists as an online textbook of biology, but also to bioinformaticians for making new discoveries about biological pathways.
Rhea ***New***	Rhea is a freely available, manually annotated database of chemical reactions created in collaboration with the Swiss Institute of Bioinformatics (SIB).

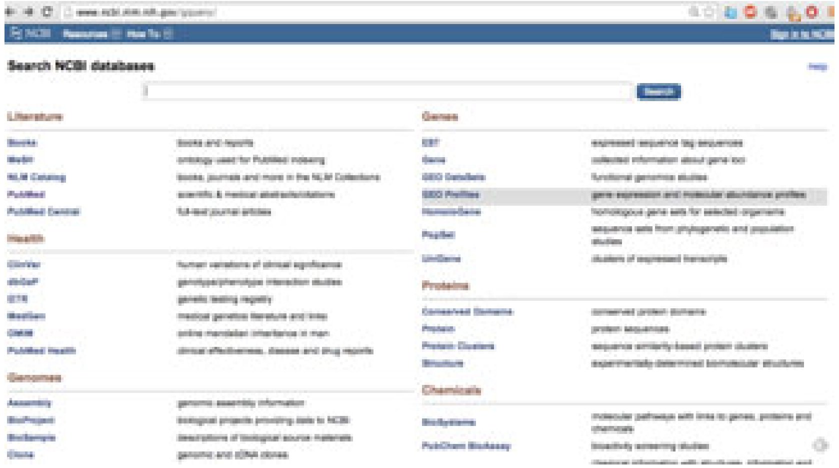
Gateways

- NCBI - Entrez -Pubmed
- EBI
- SRS
- BioMart
- ExPASy

Google

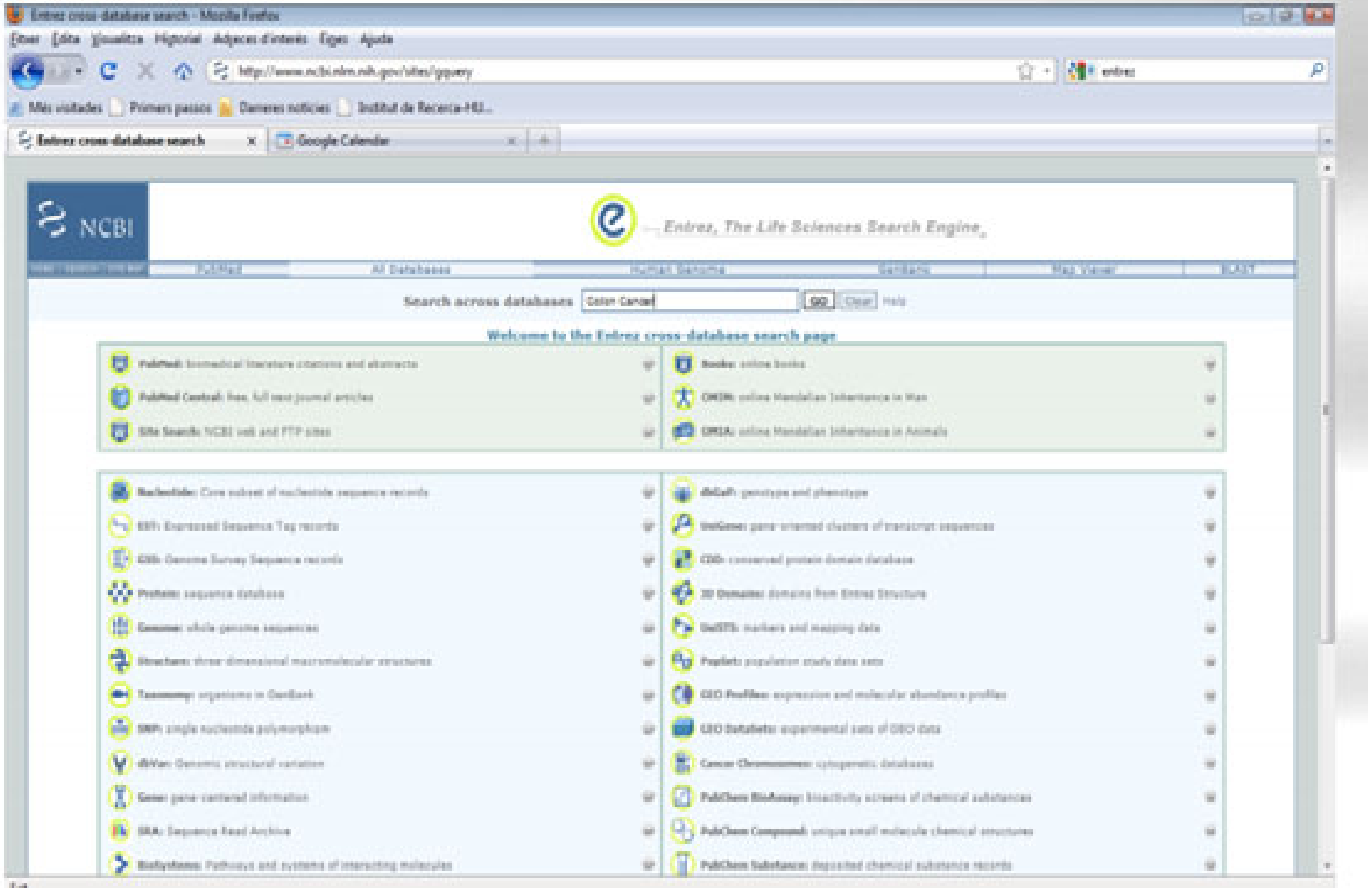
Entrez

- Entrez searches the following databases:
- PubMed: biomedical literature citations and abstracts, including Medline - articles from entral and other full-text resources are provided for articles from the 1990s.
- PubMed Central: free, full-text journal articles
- Site Search: NCBI web and FTP web sites
- Books: online books
- Online Mendelian Inheritance in Man (OMIM)
- Online Mendelian Inheritance in Animals (OMIA)
- Nucleotide: sequence database (GenBank)
- Protein: sequence database
- Genome: whole genome sequences and mapping
- Structure: three-dimensional macromolecular structures
- Taxonomy: organisms in GenBank Taxonomy
- SNP: single nucleotide polymorphism
- Gene: gene-centered information



From: Alex Sánchez

Cross search in entrez



Search across databases: Colon Cancer

Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts

PubMed Central: free, full text journal articles

Site Search: NCBI web and FTP sites

Books: online books

OMIM: online Mendelian Inheritance in Man

OMIA: online Mendelian Inheritance in Animals

Nucleotide: Core subset of nucleotide sequence records

EST: Expressed Sequence Tag records

GSS: Genome Survey Sequence records

Protein: sequence database

Genome: whole genome sequences

Structure: three-dimensional macromolecular structures

Taxonomy: organisms in GenBank

SNP: single nucleotide polymorphism

dbVar: Genomic structural variation

Gene: gene-centered information

SRA: Sequence Read Archive

BioSystems: Pathways and systems of interacting molecules

dbGaP: genotype and phenotype

UniGene: gene-oriented clusters of transcript sequences

CDD: conserved protein domain database

3D Domains: domains from Entrez Structure

UniSTS: markers and mapping data

PopSet: population study data sets

GEO Profiles: expression and molecular abundance profiles

GEO DataSets: experimental sets of GEO data

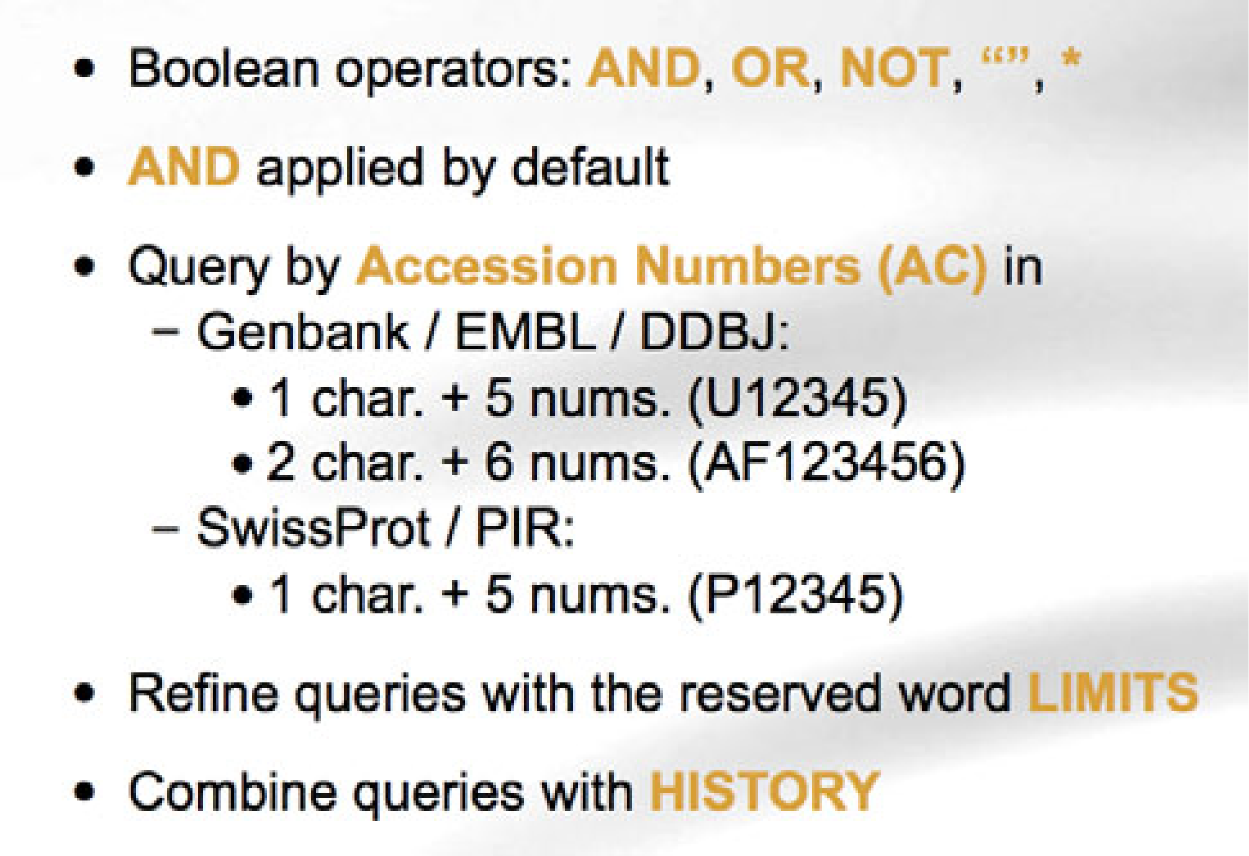
Cancer Chromosomes: cytogenetic databases

PubChem BioAssay: bioactivity screens of chemical substances

PubChem Compound: unique small molecule chemical structures

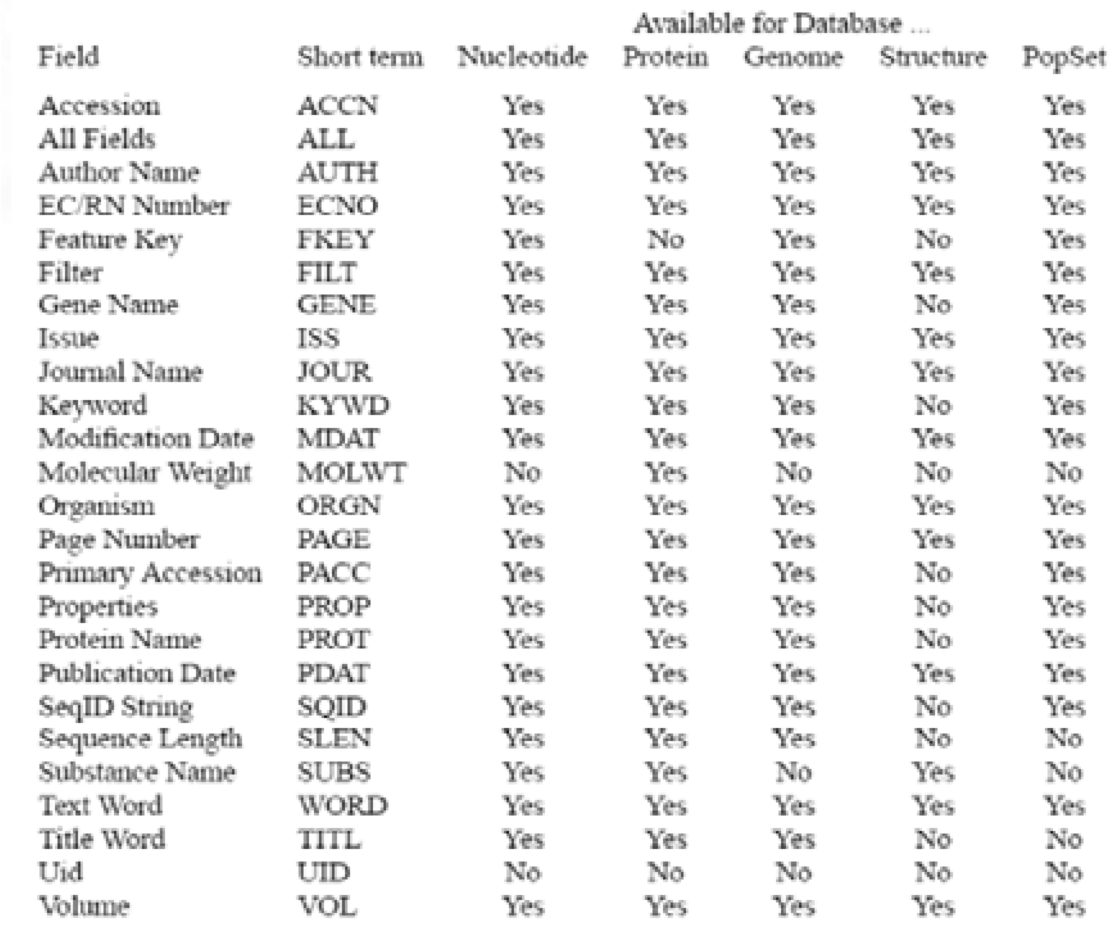
PubChem Substance: deposited chemical substance records

Searching entrez



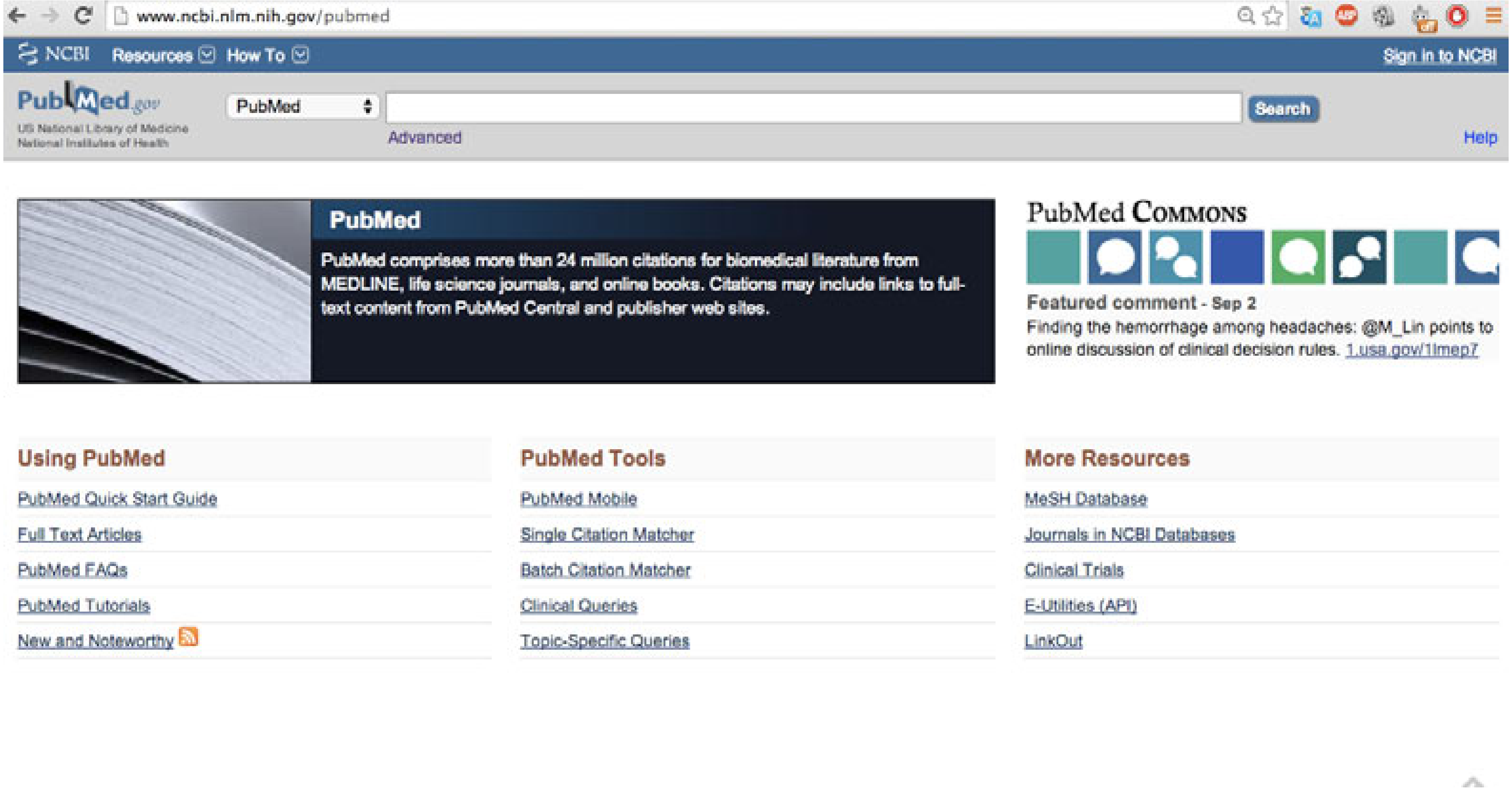
- Boolean operators: **AND**, **OR**, **NOT**, **“”**, *****
- **AND** applied by default
- Query by **Accession Numbers (AC)** in
  - Genbank / EMBL / DDBJ:
    - 1 char. + 5 nums. (U12345)
    - 2 char. + 6 nums. (AF123456)
  - SwissProt / PIR:
    - 1 char. + 5 nums. (P12345)
- Refine queries with the reserved word **LIMITS**
- Combine queries with **HISTORY**

Search fields



		Available for Database ...				
Field	Short term	Nucleotide	Protein	Genome	Structure	PopSet
Accession	ACCN	Yes	Yes	Yes	Yes	Yes
All Fields	ALL	Yes	Yes	Yes	Yes	Yes
Author Name	AUTH	Yes	Yes	Yes	Yes	Yes
EC/RN Number	ECNO	Yes	Yes	Yes	Yes	Yes
Feature Key	FKEY	Yes	No	Yes	No	Yes
Filter	FILT	Yes	Yes	Yes	Yes	Yes
Gene Name	GENE	Yes	Yes	Yes	No	Yes
Issue	ISS	Yes	Yes	Yes	Yes	Yes
Journal Name	JOUR	Yes	Yes	Yes	Yes	Yes
Keyword	KYWD	Yes	Yes	Yes	No	Yes
Modification Date	MDAT	Yes	Yes	Yes	Yes	Yes
Molecular Weight	MOLWT	No	Yes	No	No	No
Organism	ORGN	Yes	Yes	Yes	Yes	Yes
Page Number	PAGE	Yes	Yes	Yes	Yes	Yes
Primary Accession	PACC	Yes	Yes	Yes	No	Yes
Properties	PROP	Yes	Yes	Yes	No	Yes
Protein Name	PROT	Yes	Yes	Yes	No	Yes
Publication Date	PDAT	Yes	Yes	Yes	Yes	Yes
SeqID String	SQID	Yes	Yes	Yes	No	Yes
Sequence Length	SLEN	Yes	Yes	Yes	No	No
Substance Name	SUBS	Yes	Yes	No	Yes	No
Text Word	WORD	Yes	Yes	Yes	Yes	Yes
Title Word	TITL	Yes	Yes	Yes	No	No
Uid	UID	No	No	No	No	No
Volume	VOL	Yes	Yes	Yes	Yes	Yes

PubMed



PubMed

PubMed comprises more than 24 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

PubMed Commons

Featured comment - Sep 2

Finding the hemorrhage among headaches: @M_Lin points to online discussion of clinical decision rules. 1.usa.gov/1lmep7

Using PubMed

- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities (API)
- LinkOut