List of biological databases

From Wikipedia, the free encyclopedia

Biological databases are stores of biological information.[1]

Contents

Primary nucleotide sequence databases

International Nucleotide Sequence Database (INSD) consists of the following databases.

1. DNA Data Bank of Japan (National Institute of Genetics)

2. EMBL (European Bioinformatics Institute)
3. GenBank (National Center for Biotechnology Information)

The three databases, DDBJ (Japan), GenBank (USA) and European Nucleotide Archive (Europe), are repositories for nucleotide sequence data from all organisms. All three databases accept nucleotide sequence submissions, and then exchange new and updated data on a daily basis to achieve optimal synchronisation between them. These three databases are primary databases, as they house original sequence data.

Meta databases

These databases of databases collect data from different sources and make them available in new and more convenient form, or with an emphasis on a particular disease or organism.

1. BioGraph (http://biograph.be) (University of Antwerp, Vlaams Instituut voor Biotechnologie) A knowledge discovery service based on the integration of more than 20 heterogeneous databases
2. Bioinformatic Harvester[1] (http://harvester.fzk.de) (Karlsruhe Institute of Technology) - Integrating 26 major protein/gene resources.
3. Neuroscience Information Framework[2] (http://neuinfo.org/mynif/search.php?q=*) (University of California, San Diego) - Integrates hundreds of neuroscience relevant resources, many are listed below.
4. ConsensusPathDB - A molecular functional interaction database, integrating information from 12 other databases.
5. Entrez[3] (http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi) (National Center for Biotechnology Information)
6. Enzyme Portal (http://www.ebi.ac.uk/enzymeportal) Integrates enzyme information such as small-molecule chemistry, biochemical pathways and drug compounds. (European Bioinformatics Institute)
7. euGenes (http://eugenes.org) (Indiana University)
8. GeneCards (http://www.genecards.org) (Weizmann Inst.)
9. MetaBase[4] (http://BioDatabase.Org) (KOBIC) - A user contributed database of biological databases.
10. mGen (http://www.cyber-indian.com/bioperl/index.html) containing four of the world biggest databases GenBank, Refseq, EMBL and DDBJ - easy and simple program friendly gene extraction
11. MOPED (https://www.proteinspire.org/MOPED) (Seattle Children's Research Institute) - A multi-omics expression profiling database providing integrated proteomics and transcriptomics data from human, mouse, worm, and yeast.
12. PathogenPortal (http://pathogenportal.org/portal/portal/PathPort/Home) A repository linking to the Bioinformatics Resource Centers (BRCs) sponsored by the National Institute of Allergy and Infectious Diseases (NIAID)
13. SOURCE (http://smd.stanford.edu/cgi-bin/source/sourceSearch) (Stanford University) encapsulates the genetics and molecular biology of genes from the genomes of *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* into easy to navigate GeneReports
14. iRefIndex (http://irefindex.org): provides an index of protein interactions available in a number of primary

interaction databases including BIND, BioGRID, CORUM, DIP, HPRD, InnateDB, IntAct, MatrixDB, MINT, MPact, MPIDB, MPPI and OPHID.

15. Pathway Commons (http://www.pathwaycommons.org) (Memorial Sloan-Kettering Cancer Center and University of Toronto)

16. Nowomics (http://nowomics.com) Tracks changes in several biological databases, users 'follow' genes and keywords to see a news feed of new data and papers.

Genome databases

These databases collect genome sequences, annotate and analyze them, and provide public access. Some add curation of experimental literature to improve computed annotations. These databases may hold many species genomes, or a single model organism genome.

1. Bioinformatic Harvester

2. Gene Disease Database

3. SNPedia

4. CAMERA (http://camera.calit2.net/index.php/) Resource for microbial genomics and metagenomics

5. Corn (http://www.maizegdb.org/), the Maize Genetics and Genomics Database

6. EcoCyc (http://ecocyc.org) a database that describes the genome and the biochemical machinery of the model organism *E. coli K-12*

7. Ensembl provides automatic annotation databases for human, mouse, other vertebrate and eukaryote genomes.

8. Ensembl Genomes provides genome-scale data for bacteria, protists, fungi, plants and invertebrate metazoa, through a unified set of interactive and programmatic interfaces (using the Ensembl software platform).

9. PATRIC (http://patricbrc.vbi.vt.edu/), the PathoSystems Resource Integration Center

10. Flybase, genome of the model organism Drosophila melanogaster

11. MGI Mouse Genome (http://www.informatics.jax.org) (Jackson Lab.)

12. JGI Genomes (http://genome.jgi.doe.gov/) of the DOE-Joint Genome Institute provides databases of many eukaryote and microbial genomes.

13. National Microbial Pathogen Data Resource (http://www.nmpdr.org/). A manually curated database of annotated genome data for the pathogens Campylobacter, Chlamydia, Chlamydophila, Haemophilus, Listeria, Mycoplasma, Neisseria, Staphylococcus, Streptococcus, Treponema, Ureaplasma, and Vibrio.

14. RegulonDB (http://regulondb.ccg.unam.mx/) RegulonDB is a model of the complex regulation of transcription initiation or regulatory network of the cell E. coli K-12.

15. Repbase (http://www.girinst.org/repbase/) Repbase is the most commonly used database for repetitive elements (transposons).

16. Saccharomyces Genome Database, genome of the yeast model organism.

17. Viral Bioinformatics Resource Center (http://troy.bioc.uvic.ca/) Curated database containing annotated

genome data for eleven virus families.

18. The SEED (http://seed-viewer.theseed.org/) platform for microbial genome analysis includes all complete microbial genomes, and most partial genomes. The platform is used to annotate microbial genomes using subsystems.

19. Xenbase, genome of the model organism Xenopus tropicalis and Xenopus laevis

20. Wormbase, genome of the model organism Caenorhabditis elegans

21. Zebrafish Information Network, genome of this fish model organism.

22. TAIR (http://arabidopsis.org/), The Arabidopsis Information Resource.

23. UCSC Malaria Genome Browser, genome of malaria causing species (*Plasmodium falciparum* and others)

24. RGD (http://rgd.mcw.edu) Rat Genome Database: Genomic and phenotype data for Rattus norvegicus

25. [5] (http://integrall.bio.ua.pt) INTEGRALL: Database dedicated to integrons, bacterial genetic elements involved in the antibiotic resistance

26. Fourmidable ant genome database (http://www.antgenomes.org) provides ant genome blast (http://www.antgenomes.org/blast) search and sequence download (http://www.antgenomes.org/downloads).

27. VectorBase (http://www.vectorbase.org/) The NIAID Bioinformatics Resource Center for Invertebrate Vectors of Human Pathogens

28. (http://ezgenome.ezbiocloud.net/)EzGenome, comprehensive information about manually curated genome projects of prokaryotes (archaea and bacteria) [2]

29. Banana Genome Hub (http://banana-genome.cirad.fr), The Banana Genome database.

Protein sequence databases

1. UniProt Universal Pesource (EBI, Swiss Institute of Bioinformatics, PIR)

2. Protein Information Resource (Georgetown University Medical Center (GUMC))

3. Swiss-Prot Protein Knowledgebase (Swiss Institute of Bioinformatics)

4. PEDANT (http://pedant.gsf.de) Protein Extraction, Description and ANalysis Tool (Forschungszentrum f. Umwelt & Gesundheit)

5. PROSITE Database of Protein Families and Domains

6. Database of Interacting Proteins (Univ. of California)

7. Pfam Protein families database of alignments and HMMs (Sanger Institute)

8. PRINTS a compendium of protein fingerprints from (Manchester University)

9. ProDom (http://protein.foulouse.inra.fr/prodom/current/html/home.php) Comprehensive set of Protein Domain Families (INRA/CNRS)

10. SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) Server for signal peptide prediction (including cleavage site prediction), based on artificial neural networks and HMMs

11. SUPERFAMILY Library of HMMs representing superfamilies and database of (superfamily and family) annotations for all completely sequenced organisms

12. Annotation Clearing House (http://clearinghouse.nmpdr.org/aclh.cgi) a project from the National Microbial Pathogen Data Resource

13. InterPro Classifies proteins into families and predicts the presence of domains and sites.

Proteomics databases

1. Proteomics Identifications Database (PRIDE) (http://www.ebi.ac.uk/pride) A public repository for proteomics data, containing protein and peptide identifications and their associated supporting evidence as well as details of post-translational modifications. (European Bioinformatics Institute)

2. MitoMiner (http://mitominer.mrc-mbu.cam.ac.uk) - A mitochondrial proteomics database integrating large-scale experimental datasets from mass spectrometry and GFP studies for 12 species. (MRC Mitochondrial Biology Unit)

3. GelMap (http://gelmap.de) - A public database of proteins identified on 2D gels (University of Hanover Proteomics Department)

Protein structure databases

Protein Data Bank (PDB) comprising:

- Protein DataBank in Europe (http://www.pdbe.org) (PDBe)
- ProteinDatabank in Japan (http://www.pdbj.org) (PDBj)
- Research Collaboratory for Structural Bioinformatics (http://www.rcsb.org/pdb/) (RCSB)

Secondary databases

1. SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/) Structural Classification of Proteins
2. CATH (http://www.cathdb.info/) Protein Structure Classification
3. PDBsum (http://www.ebi.ac.uk/pdbsum/)

For more protein structure databases, see also Protein structure database

Protein model databases

1. Swiss-model[6] (http://swissmodel.expasy.org//SWISS-MODEL.html) Server and Repository for Protein Structure Models

2. ModBase[7] (http://salilab.org/modbase) Database of Comparative Protein Structure Models (Sali Lab, UCSF)

3. Protein Model Portal[8] (http://proteinmodelportal.org) (PMP) Meta database that combines several databases of protein structure models (Biozentrum, Basel, Switzerland)

RNA databases

1. LncRNAWiki [9] (http://lncrna.big.ac.cn), a wiki-based database for community curation of human long non-coding RNAs
2. Rfam [10] (http://rfam.sanger.ac.uk), a database of RNA families
3. miRBase [11] (http://www.mirbase.org), the microRNA database
4. snoRNAdb (http://lowelab.ucsc.edu/snoRNAdb/), a database of snoRNAs
5. lncRNAdb (http://www.lncrnadb.org/), a database of lncRNAs
6. MONOCLdb (http://www.monocldb.org/) The MOuse NOnCode Lung database: Annotations and expression profiles of mouse long non-coding RNAs (lncRNAs) involved in Influenza and SARS-CoV infections.
7. piRNAbank (http://pirnabank.ibab.ac.in/), a database of piRNAs
8. GtRNAdb (http://gtrnadb.ucsc.edu/), a database of genomic tRNAs
9. SILVA (http://www.arb-silva.de/), a database of ribosomal RNAs
10. RDP (http://rdp.cme.msu.edu/), the Ribosomal Database Project
11. tmRDB (http://www.ag.auburn.edu/mirror/tmRDB/), a database of tmRNAs
12. SRPDB (http://rth.dk/resources/rnp/SRPDB/), a database of signal recognition particle RNAs
13. yeast snoRNA database (http://people.biochem.umass.edu/fournierlab/snornadb/main.php)
14. Sno/scaRNAbase (http://bioinfo.fudan.edu.cn/snoRNAbase.nsf), a database of snoRNA and scaRNAs
15. snoRNA-LBME-db (http://www-snorna.biotoul.fr/), a snoRNA database

Carbohydrate structure databases

1. EuroCarbDB[12] (http://www.ebi.ac.uk/eurocarb), A repository for both carbohydrate sequences/structures and experimental data.

Protein-protein and other molecular interactions

1. BIND Biomolecular Interaction Network Database (http://www.bind.ca/)
2. BioGRID [13] (http://www.thebiogrid.org) A General Repository for Interaction Datasets (Samuel Lunenfeld Research Institute)
3. CCSB Interactome (http://interactome.dfci.harvard.edu/)
4. DIP Database of Interacting Proteins (http://dip.doe-mbi.ucla.edu/)
5. IntAct molecular interaction database (http://ebi.ac.uk/intact/): a central, standards-compliant repository of

molecular interactions, including protein–protein, protein–small molecule and protein–nucleic acid interactions.

6. NetPro (http://www.molecularconnections.com/home/en/home/resources/case-studies/alzheimer-disease-netpro/)
7. STRING: STRING is a database of known and predicted protein-protein interactions. (http://string.embl.de) (EMBL)
8. The Cell Collective (http://thecellcollective.org)
9. MINT: Molecular INTeraction database (http://mint.bio.uniroma2.it/)
10. iRefIndex (http://irefindex.org): provides an index of protein interactions available in a number of primary interaction databases including BIND, BioGRID, CORUM, DIP, HPRD, InnateDB, IntAct, MatrixDB, MINT, MPact, MPIDB, MPPI and OPHID.
11. RNA-binding protein database
12. BioLiP: Protein-ligand binding database (http://zhanglab.ccmb.med.umich.edu/BioLiP/)

Signal transduction pathway databases

1. Cancer Cell Map (http://cancer.cellmap.org)
2. Netpath - A curated resource of signal transduction pathways in humans
3. NCI-Nature Pathway Interaction Database
4. Reactome - Navigable map of human biological pathways, ranging from metabolic processes to hormonal signalling.
5. SignaLink Database (http://signalink.org/)
6. WikiPathways
7. The Cell Collective (http://thecellcollective.org)
8. Literature-curated human signaling network, the largest human signaling network database (http://www.cancer-systemsbiology.org/datasoftware.htm)

Metabolic pathway and Protein Function databases

1. BioCyc Database Collection including EcoCyc and MetaCyc
2. BRENDA The Comprehensive Enzyme Information System, including FRENDA, AMENDA, DRENDA, and KENDA, [14] (http://www.brenda-enzymes.org/)
3. KEGG PATHWAY Database[15] (http://www.genome.ad.jp/kegg/pathway.html) (Univ. of Kyoto)
4. MANET database [16] (http://www.manet.uiuc.edu/) (University of Illinois)
5. Metabolights (http://www.ebi.ac.uk/metabolights) Metabolomics experiments and derived information: metabolite structures, reference spectra, biological roles, locations and concentrations. (European Bioinformatics Institute)

6. MetaNetX (http://metanetx.org) Automated Model Construction and Genome Annotation for Large-Scale Metabolic Networks
7. Reactome[17] (http://www.reactome.org) Navigable map of human biological pathways, ranging from metabolic processes to hormonal signalling. (Cold Spring Harbor Laboratory, European Bioinformatics Institute, Gene Ontology Consortium)
8. Small Molecule Pathway Database (SMPDB) (http://www.smpdb.ca)

Microarray databases

1. ArrayExpress (http://www.ebi.ac.uk/arrayexpress) (European Bioinformatics Institute)
2. Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) (National Center for Biotechnology Information)
3. GPX (http://www.gti.ed.ac.uk/GPX)(Scottish Centre for Genomic Technology and Informatics)
4. maxd (http://www.bioinf.man.ac.uk/microarray/maxd/index.html) (Univ. of Manchester)
5. Stanford Microarray Database (SMD) (http://smd.stanford.edu/) (Stanford University)
6. Genevestigator - Expression Search Engine (http://www.genevestigator.com/gv/) (Nebion AG)
7. Bgee (http://bgee.unil.ch) Bgee is a database to retrieve and compare gene expression patterns between species. It contains wild-type and manually curated microarray experiments only.

Exosomal databases

- ExoCarta

Mathematical model databases

1. Biomodels Database: published mathematical models describing biological processes.
2. CellML (http://www.cellml.org/models)
3. The Cell Collective (http://thecellcollective.org): build and simulate large-scale models in real-time and in a highly collaborative fashion

PCR and quantitative PCR primer databases

1. PathoOligoDB: A free QPCR oligo database for pathogens (http://www.pathooligodb.com/)
2. RTPrimerDB - a public primers and probes database for real-time PCR reactions (http://medgen.ugent.be/rtprimerdb/index.php/)

Phenotype databases

1. PhenCode (http://www.bx.psu.edu/phencode/) linking human mutations with phenotype

2. PhenomicDB (http://www.phenomicdb.de/) multi-organism database linking genotype to phenotype

3. PHI-base Pathogen-host interaction database. It links gene information to phenotypic information from microbial pathogens on their hosts. Information is manually curated from peer reviewed literature.

4. RGD (http://rgd.mcw.edu) Rat Genome Database: Genomic and phenotype data for Rattus norvegicus

5. Planform (http://planform.daniel-lobo.com/): planarian formalized-experiments database, linking surgical, genetic, and pharmacological perturbations to morphological phenotypic outcomes from published planarian regeneration experiments.

6. Limbform (http://limbform.daniel-lobo.com/): limb formalized-experiments database, linking surgical, genetic, and pharmacological perturbations to morphological phenotypic outcomes from published multi-organism limb regeneration experiments.

Specialized databases

- Antibody Central (http://antibody-central.com) Antibody information database and search resource.
- AntibodyRegistry.org (http://antibodyregistry.org) assigns unique identifiers used to track antibody reagents in published literature.
- Bgee (http://bgee.unil.ch) Bgee is a database to retrieve and compare gene expression patterns between species.
- BIOMOVIE (http://biomovie.ethz.ch) (ETH Zurich) movies related to biology and biotechnology
- BioNumbers (http://bionumbers.hms.harvard.edu/) a database of useful biological numbers
- Barcode of Life Data Systems, a database of DNA barcodes
- CGAP Cancer Genes (http://cgap.nci.nih.gov/Genes/GeneFinder) (National Cancer Institute)
- Clone Registry Clone Collections (http://www.ncbi.nlm.nih.gov/genome/clone) (National Center for Biotechnology Information)
- Connectivity map (http://www.broad.mit.edu/cmap/) Transcriptional expression data and correlation tools for drugs
- CTD (http://ctdbase.org/) The Comparative Toxicogenomics Database describes chemical-gene-disease interactions
- DBGET H.sapiens (http://www.genome.ad.jp/dbget-bin/www_bfind?h.sapiens) (Univ. of Kyoto)
- DiProDB A database to collect and analyse thermodynamic, structural and other dinucleotide properties.
- Drug2Gene (http://www.drug2gene.com/) Provides integrated information for identified and reported relations between genes/proteins and drugs/compounds
- Dryad a repository of data underlying scientific publications in the basic and applied biosciences.
- Edinburgh Mouse Atlas
- FunSecKB (http://proteomics.ysu.edu/secretomes/fungi.php) The fungal secretome knowledgebase.
- FunSecKB2 (http://proteomics.ysu.edu/secretomes/fungi2/index.php) The fungal secretome and subcellular

proteome knowledgebase (version 2)

- GreenPhylDB (http://www.greenphyl.org) (A phylogenomic database for plant comparative genomics)
- GDB Hum. Genome Db (http://www.gdb.org/gdb) (Human Genome Organisation)
- HGMD disease-causing mutations (http://www.hgmd.cf.ac.uk/) (HGMD Human Gene Mutation Database)
- HUGO (http://www.gene.ucl.ac.uk/nomenclature) (Official Human Genome Database: HUGO Gene Nomenclature Committee)
- HvrBase++ (http://www.hvrbase.org/) Human and primate mitochondrial DNA
- INTERFEROME (http://www.interferome.org/) The Database of Interferon Regulated Genes
- List with SNP-Databases (http://hgvbase.cgb.ki.se/databases.htm)
- MetazSecKB (http://proteomics.ysu.edu/secretomes/animal/index.php) The metazoa [human/animal] secretome and subcellular proteome knowledgebase
- Minimotif Miner -Database of short contiguous functional peptide motifs
- NCBI-UniGene (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene) (National Center for Biotechnology Information)
- Oncogenomic databases A compilation of databases that serve for cancer research.
- OMIM Inherited Diseases (http://www.ncbi.nlm.nih.gov/Omim) (Online Mendelian Inheritance in Man)
- OrthoMaM (http://www.orthomam.univ-montp2.fr/) (A database of Orthologous Mammalian Markers)
- OrthoMCL (http://orthomcl.org) Ortholog Groups of Protein Sequences from Multiple Genomes including Archaea, Bacteria and Eukaryotes.
- p53 (http://p53.bii.a-star.edu.sg) The p53 Knowledgebase
- PASD (http://proteomics.ysu.edu/altsplice/) The plant alternative splicing database
- PlantSecKB (http://proteomics.ysu.edu/secretomes/plant/index.php) The plant secretome and subcullular proteome knowledgebase
- Plasma Proteome Database (http://www.plasmaproteomedatabase.org/) Human plasma proteins along with their isoforms
- SABIO-RK (http://sabiork.h-its.org/) SABIO-RK is a curated database that contains information about biochemical reactions, their kinetic rate equations with parameters and experimental conditions.
- SciClyc (http://www.SciClyc.com) An Open-access database to shared antibodies, cell cultures, and documents for biomedical research.
- Selectome (http://selectome.unil.ch) Selectome is a database of positive selection based on a rigorous branch-site specific likelihood test. Positive selection is detected using CODEML on all branches of animal gene trees.
- SHMPD (http://shmpd.bii.a-star.edu.sg) The Singapore Human Mutation and Polymorphism Database
- SNPSTR database (http://www.sbg.bio.ic.ac.uk/~ino/cgi-bin/SNPSTRdatabase.html) A database of SNPSTRs - compound genetic markers consisting of a microsatellite (STR) and one tightly linked SNP - in human, mouse, rat, dog and chicken.

- TDR Targets A chemogenomics database focused on drug discovery in tropical diseases.
- TRANSFAC A database about eukaryotic transcription factors, their genomic binding sites and DNA-binding profiles.
- TreeBASE (http://www.treebase.org/) An open-access database of phylogenetic trees and the data behind them
- Treefam (http://www.treefam.org/) TreeFam (Tree families database) is a database of phylogenetic trees of animal genes
- XTractor (http://www.xtractor.in) Discovering Newer Scientific Relations Across PubMed Abstracts. A tool to obtain manually annotated relationships for Proteins, Diseases, Drugs and Biological Processes as they get published in PubMed.

Taxonomic databases

1. Catalogue of Life source databases (http://www.catalogueoflife.org/info/databases)
2. Encyclopedia of Life (http://eol.org/)
3. Integrated Taxonomic Information System (http://www.itis.gov/)
4. EzTaxon-e, database for the identification of prokaryotes based on 16S ribosomal RNA gene sequences

Wiki-style databases

1. CHDwiki (http://homes.esat.kuleuven.be/~bioiuser/chdwiki/)
2. EcoliWiki (http://ecoliwiki.net/colipedia)
3. Gene Wiki
4. GyDB (http://gydb.uv.es/index.php/Main_Page)
5. NeuroLex (http://neurolex.org)
6. OpenWetWare (http://openwetware.org/)
7. PDBWiki (http://pdbwiki.org/)
8. Proteopedia (http://www.proteopedia.org/)
9. RiceWiki (http://ricewiki.big.ac.cn)
10. LncRNAWiki (http://lncrna.big.ac.cn)
11. Topsan (http://www.topsan.org/)
12. WikiGenes (http://www.wikigenes.org/)
13. WikiPathways (http://www.wikipathways.org/)
14. WikiProfessional
15. YTPdb (http://homes.esat.kuleuven.be/~sbrohee/ytpdb/)

Metabolomic Databases

1. MetaboLights (http://www.ebi.ac.uk/metabolights/)
2. Human Metabolome Database (HMDB) (http://www.hmdb.ca)
3. Yeast Metabolome Database (YMDB) (http://www.ymdb.ca)
4. E. coli Metabolome Database (ECMDB) (http://www.ecmdb.ca)
5. DrugBank (http://www.drugbank.ca)
6. ChEBI (https://www.ebi.ac.uk/chebi/)
7. BioMagResBank (http://www.bmrb.wisc.edu/)
8. Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de/)
9. MassBank (http://www.massbank.eu/MassBank/)

Unsorted

- FINDbase (the Frequency of INherited Disorders database)
- RIKEN integrated database of mammals

References

1. ^ Wren JD, Bateman A (2008). "Databases, data tombs and dust in the wind.". *Bioinformatics* **24** (19): 2127–8. doi:10.1093/bioinformatics/btn464 (http://dx.doi.org/10.1093%2Fbioinformatics%2Fbtn464). PMID 18819940 (https://www.ncbi.nlm.nih.gov/pubmed/18819940).
2. ^ http://ezgenome.ezbiocloud.net/

Retrieved from "http://en.wikipedia.org/w/index.php?title=List_of_biological_databases&oldid=640408760"

Categories: Biological databases │ Bioinformatics