

Contents

Nucleotide Sequence Databases	1
Protein Sequence Databases	1
Database Searching by Sequence Similarity	2
Human Genome Databases	2
Databases of other Organisms	3
Genome-wide Analysis	3
Protein Domains: Databases and Search Tools	3
Motif and Pattern Search in Sequences	3
Protein 3D Structure	4
Phylogeny & Taxonomy	4
Gene Prediction	4
Gene Expression Databases	5
Gene Regulation	5
Metabolic, Gene Regulatory & Signal Transduction Network Databases	5
Systems Biology	6
Other Databases (Annotations, Ontologies, Consortia, etc.)	6
Miscellaneous Tools	7

Another great list that I have yet to compile into this list is:

http://en.wikipedia.org/wiki/List_of_biological_databases

Nucleotide Sequence Databases

- [NCBI](#) - National Center for Biotechnology Information
- [EBI](#) - European Bioinformatics Institute
- [DDBJ](#) - DNA Data Bank of Japan

Protein Sequence Databases

- [SWISS-PROT & TrEMBL](#) - Protein sequence database and computer annotated supplement

- [UniProt](#) - UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.
- [PIR](#) - Protein Information Resource
- [MIPS](#) - Munich Information centre for Protein Sequences
- [HUPO](#) - HUman Proteome Organization

Database Searching by Sequence Similarity

- [BLAST @ NCBI](#)
- [PSI-BLAST @ NCBI](#)
- [FASTA @ EBI](#)
- [BLAT](#)

Sequence Alignment

- [USC Sequence Alignment Server](#) - align 2 sequences with all possible varieties of dynamic programming
- [T-COFFEE](#) - multiple sequence alignment
- [ClustalW @ EBI](#) - multiple sequence alignment
- [MSA 2.1](#) - optimal multiple sequence alignment using the Carrillo-Lipman method
- [BOXSHADE](#) - pretty printing and shading of multiple alignments
- [Splign](#) - Splign is a utility for computing cDNA-to-Genomic, or spliced sequence alignments. At the heart of the program is a global alignment algorithm that specifically accounts for introns and splice signals.
- [Spidey](#) - an mRNA-to-genomic alignment program
- [SIM4](#) - a program to align cDNA and genomic DNA
- [Wise2](#) - align a protein or profile HMM against genomic sequence to predict a gene structure, and related tools
- [PipMaker](#) - computes alignments of similar regions in two (long) DNA sequences
- [VISTA](#) - align + detect conserved regions in long genomic sequences
- [myGodzilla](#) - align a sequence to its ortholog in the human genome

Human Genome Databases

- [Draft Human Genome @ NCBI](#)
- [Draft Human Genome @ UCSC](#)
- [Ensembl](#) - automatically annotated human genome.
- [GDB](#) - Genome Database
- [Mammalian Gene Collection](#) - full-length (open reading frame) sequences for human and mouse
- [STACK](#) - Sequence Tag Alignment and Consensus Knowledgebase
- [GeneCards](#) - human genes, proteins and diseases

Databases of other Organisms

- [GOLD](#) - Genomes OnLine Database, information on complete and ongoing genome projects
- [TIGR Comprehensive Microbial Resource](#)
- [TIGR Microbial Database](#)
- [The Proteome Databases](#) - yeast, worm, & human, good annotation
- [Saccharomyces Genome Database](#)
- [WormBase](#) - C. elegans
- [FlyBase](#)
- [Berkeley Drosophila Genome Project](#)
- [Mouse Genome Informatics](#)
- [The Arabidopsis Information Resource](#)
- [ZFIN](#) - Zebrafish Information Network
- [DictyBase](#) - Dictyostelium discoideum
- [EcoGene](#) - E. coli
- [HIV sequence database](#)

Genome-wide Analysis

- [MBGD](#) - comparative analysis of completely sequenced microbial genomes
- [COGs](#) - phylogenetic classification of orthologous proteins from complete genomes
- [STRING](#) - detect whether a given query gene occurs repeatedly with certain other genes in potential operons
- [Pedant](#) - automatic whole genome annotation
- [GeneCensus](#) - various whole genome comparisons

Protein Domains: Databases and Search Tools

- [InterPro](#) - integration of Pfam, PRINTS, PROSITE, SWISS-PROT + TrEMBL
- [PROSITE](#) - database of protein families and domains
- [Pfam](#) - alignments and hidden Markov models covering many common protein domains
- [SMART](#) - analysis of domains in proteins
- [ProDom](#) - protein domain database
- [PRINTS Database](#) - groups of conserved motifs used to characterise protein families
- [Blocks](#) - multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins
- [Protein Domain Profile Analysis @ BMERC](#) - search a library of profiles with a protein sequence
- [TIGRFAMs](#) - yet more protein families based on Hidden Markov Models

Motif and Pattern Search in Sequences

- [Gibbs Motif Sampler](#) - identification of conserved motifs in DNA or protein sequences

- [AlignACE Homepage](#) - gene regulatory motif finding
- [MEME](#) - motif discovery and search in protein and DNA sequences
- [SAM](#) - tools for creating and using Hidden Markov Models
- [Pratt](#) - discover patterns in unaligned protein sequences
- [Motivated Proteins](#) - a web facility for exploring small hydrogen-bonded motifs

Protein 3D Structure

- [PDB](#) - protein 3D structure database
- [RasMol / Protein Explorer](#) - molecule 3D structure viewers
- [SCOP](#) - Structural Classification Of Proteins
- [UCL BSM CATH classification](#)
- [The DALI Domain Database](#)
- [FSSP](#) - fold classification based on structure-structure alignment of proteins
- [SWISS-MODEL](#) - homology modeling server
- [Structure Prediction Meta-server](#)
- [K2](#) - protein structure alignment
- [DALI](#) - 3D structure alignment server
- [DSSP](#) - defines secondary structure and solvent exposure from 3D coordinates
- [HSSP Database](#) - Homology-derived Secondary Structure of Proteins
- [PredictProtein & PHD](#) - predict secondary structure, solvent accessibility, transmembrane helices, and other stuff
- [Jpred2](#) - protein secondary structure prediction
- [PSIpred \(& MEMSAT & GenTHREADER\)](#) - protein secondary structure prediction (& transmembrane helix prediction & tertiary structure prediction by threading)

Phylogeny & Taxonomy

- [The Tree of Life](#)
- [Species 2000](#) - index of the world's known species
- [TreeBASE](#) - a database of phylogenetic knowledge
- [PHYLIP](#) - package of programs for inferring phylogenies
- [TreeView](#) - user friendly tree displaying for Macs & Windows

Gene Prediction

- [Genscan](#) - eukaryotes
- [GeneMark](#)
- [Genie](#) - eukaryotes
- [GLIMMER](#) - prokaryotes
- [tRNAscan - SE 1.1](#) - search for tRNA genes in genomic sequence
- [GFF \(General Feature Format\) Specification](#) - a standard format for genomic sequence annotation

Gene Expression Databases

- [HuGE](#) - database of human gene expression using arrays
- [ExpressDB](#) - yeast and E. coli RNA expression data
- [SAGE @ NCBI](#) - Serial Analysis of Gene Expression
- [Stanford Microarray Database](#)
- [Gene Expression Omnibus \(GEO\)](#)

Gene Regulation

- [TRAFAC](#) - For identifying conserved and shared cis regulatory elements between a pair of genes.
- [CisMols](#) - For identifying conserved and shared cis regulatory elements between a set of co-expressed genes.
- [TRANSFAC](#) - database of eukaryotic cis-acting regulatory DNA elements and trans-acting factors
- [EPD](#) - eukaryotic promoter database
- [DBTSS](#) - DataBase of Transcriptional Start Sites (human)
- [SCPD](#) - Saccharomyces cerevisiae promoter database
- [DCPD](#) - Drosophila Core Promoter Database
- [RegulonDB](#) - a database on transcriptional regulation in E. coli
- [DPInteract](#) - protein binding sites on E. coli DNA
- [PromoterInspector](#) - prediction of promoter regions in mammalian genomic sequences
- [MatInspector](#) - search for transcription factor binding sites
- [Cister](#) - cis-element cluster finder
- [Gene regulatory Tools](#)
- [microRNA.org: microRNA Targets & Expression Profiles](#)
- [miRBase](#)
- [TarBase](#) Provides a means of searching through a comprehensive set of experimentally supported microRNA targets in at least 8 organisms
- [microRNA resource](#) A gateway to all types of information about microRNAs, including articles, products, news, events, and other websites

Metabolic, Gene Regulatory & Signal Transduction Network Databases

- [KEGG](#) - Kyoto Encyclopedia of Genes and Genomes
- [BioCarta](#)
- [DAVID](#) - Database for Annotation, Visualization and Integrated Discovery - A useful server to for annotating microarray and other genetic data.
- [stke](#) - Signal Transduction Knowledge Environment
- [BIND](#) - Biomolecular Interaction Network Database
- [EcoCyc](#)
- [WIT](#)
- [PathGuide](#) A very useful collection of resources dealing primarily with pathways

- [SPAD](#) - Signaling Pathway Database
- [CSNDB](#) - Cell Signalling Networks Database
- [PathDB](#)
- [Transpath](#)
- [DIP](#) - Database of Interacting Proteins
- [PFBP](#) - Protein Function and Biochemical Networks
- [Alliance for Cellular Signalling](#)

Systems Biology

- **Gene List Annotation Tools**
 - [DAVID](#) - Database for Annotation, Visualization and Integrated Discovery - A useful server to for annotating microarray and other genetic data.
 - [MSigDB](#) - Molecular Signatures Database
 - [ToppGene Suite](#) Gene list functional enrichment and candidate gene prioritization
 - [Panther](#) - Protein ANalysis THrough Evolutionary Relationships
 - [L2L](#)
 - [Babelomics \(FatiGO+\)](#)
 - [OntoExpress](#)

Other Databases (Annotations, Ontologies, Consortia, etc.)

- [Entrez Gene](#) - Gene provides a unified query environment for genes defined by sequence and/or in NCBI's Map Viewer. You can query on names, symbols, accessions, publications, GO terms, chromosome numbers, E.C. numbers, and many other attributes associated with genes and the products they encode.
- [Cancer Genome Anatomy Project](#)
- [HUGO's Human Gene Nomenclature](#)
- [Gene Ontology Consortium](#) - a controlled vocabulary of eukaryotic gene roles
- [Open Biological Ontologies](#) an umbrella web address for well-structured controlled vocabularies for shared use across different biological domains.
- [ACUTS](#) - compilation of Ancient Conserved UnTranslated Sequences
- [UTR database](#)
- [ENZYME](#) - enzyme nomenclature database
- [BRENDA](#) - enzyme database
- [TC-DB](#) - comprehensive classification of membrane transport proteins
- [The SNP Consortium](#)
- [HGBASE](#) - database of sequence variations in the human genome
- [MethDB](#) - DNA methylation database
- [SpliceDB](#) - canonical and non-canonical splice site sequences in mammalian genes
- [SpliceOme](#) - database of intron-exon boundaries
- [InBase](#) - intein database
- [The I.M.A.G.E. Consortium](#)

- [The Kabat Database of Sequences of Proteins of Immunological Interest](#)
- [Nelson Lab: Cytochrome C](#)
- [REBASE](#) - restriction enzyme database
- [Chemfinder.com](#) - molecule database
- [Genomics Institute of the Novartis Research Foundation](#)
- [Mouse SNPs Database](#)- 670,000+ SNP records, 8.0+ million allele calls. Allele tables are provided by investigators or retrieved from public sources. All SNPs are mapped to NCBI Mouse Genome build 33 (C57BL/6J assembly). Most are linked to NCBI dbSNP build 123.
- [MetaBase](#) is a user contributed database of databases, listing all the biological databases currently available on the internet.
- [Bio-computing.org](#) Bioinformatics, Databases and Software for Medicine.

Miscellaneous Tools

- [NCBI Genome Workbench](#) - NCBI Genome Workbench is an integrated application for viewing and analyzing sequence data. With Genome Workbench, you can view data in publically available sequence databases at NCBI, and mix this data with your own private data.
- [Repeatmasker](#) - mask repetitive elements in DNA sequences
- [Tandem Repeats Finder](#)
- [Vienna RNA Package](#) - RNA secondary structure prediction
- [mfold \(1\)](#) - RNA secondary structure prediction
- [mfold \(2\)](#) - RNA secondary structure prediction
- [EST parser](#) - find alternative polyadenylation sites in mRNAs, using ESTs
- [UTR-extender](#) - extends missing ends of an mRNA using EST and genome sequence data
- [CpG Islands](#) - predict CpG islands
- [NetStart](#) - prediction of translation start sites in vertebrate and A.thaliana sequences
- [ATGpr](#) - prediction of translation start sites in cDNA sequences
- [SignalP](#) - secretory signal peptide prediction
- [PSORT](#) - prediction of protein sorting signals and transmembrane helices
- [CBS Prediction Servers](#) - prediction of protein subcellular localization and various sites in protein and nucleotide sequences
- [Compute pI/Mw Tool](#)
- [Translate Tool](#)
- [Reverse complement nucleotide sequences](#)
- [Melting](#) - calculate melting temperature for nucleic acid duplexes
- [bend.it](#) - calculate curvature and bendability of a DNA sequence
- [webcutter](#) - detect restriction enzyme cutting sites in DNA sequences
- [Primer3](#) - pick primers from a DNA sequence
- [Probability Distribution Calculators](#) - normal, chi square, t, F, etc.