

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

The European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database constitutes Europe's primary nucleotide sequence resource. DNA and RNA sequences have been deposited from individual researchers, genome sequencing projects and patent applications.

<http://www.ddbj.nig.ac.jp/>

The DNA Data Bank of Japan (DDBJ) is the sole DNA databank in Japan. It is officially certified to collect DNA sequences from researchers and to issue the internationally recognized accession number for data submission. GenBank(R), EMBL, and DDBJ all share their sequence information.

<http://ndbserver.rutgers.edu/NDB/ndb.html>

The Nucleic Acid Database Project (NDB) assembles and distributes structural information about nucleic acids.

A.3 GENOMES

<http://igweb.integratedgenomics.com/GOLD/>

The Genomes Online Database (GOLD) is a resource for information regarding complete and ongoing genome projects around the world.

<http://www.ncbi.nlm.nih.gov/Entrez/Genome/org.html>

The National Center for Biotechnology Information (NCBI) maintains a searchable database of all known completed genome projects.

<http://www.tigr.org/tdb/>

The Institute for Genomic Research (TIGR) databases are a collection of curated databases containing DNA and protein sequence, gene expression, cellular role, protein family, and taxonomic data for microbes, plants, and humans. Anonymous FTP access to sequence data is also provided.

A.4 EXPRESSED SEQUENCE TAGS (ESTs)

<http://www.ncbi.nlm.nih.gov/dbEST/index.html>

dbEST is a division of GenBank that contains sequence data and other information on “single-pass” cDNA sequences, or expressed sequence tags, for a number of organisms. A brief account of the history of human ESTs in GenBank is available at the site.

<http://www.tigr.org/tdb/tgi.shtml>

TIGR Gene Indices Database integrates data from international EST sequencing and gene research projects. The gene indices are an analysis of the transcribed sequences represented in the world’s public EST data.

A.5 SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

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

In collaboration with the National Human Genome Research Institute, the National Center for Biotechnology Information has established the dbSNP database to serve as a central repository for both single-base nucleotide substitutions and short deletion and insertion polymorphisms.

A.6 RNA STRUCTURES

<http://rdp.cme.msu.edu/html/>

The Ribosomal Database Project (RDP) provides ribosome-related data services to the scientific community, including online data analysis, rRNA-derived phylogenetic trees, and aligned and annotated rRNA sequences.

<http://www.rna.icmb.utexas.edu/>

The Comparative RNA Web (CRW) is a repository for data on the four major types of comparative information and systems available for ribosomal RNAs (5S, 16S, and 23S rRNA), transfer RNA (tRNA), and two of the catalytic intron RNAs (group I and group II), including: (1) current comparative structure models; (2) nucleotide frequency and conservation information; (3) sequence and structure data; and (4) data access systems.

A.7 PROTEINS

<http://www-nbrf.georgetown.edu/pir/>

The Protein Information Resource—International Protein Sequence Database (PIR-PSD) is a comprehensive, nonredundant, expertly annotated, fully classified, and extensively cross-referenced protein sequence database in the public domain. The PIR-PSD, iProClass, and other PIR auxiliary databases integrate sequences, functional, and structural information to support genomics and proteomics research.

<http://www.expasy.ch/sprot/>

SWISS-PROT is a curated protein sequence database that strives to provide a high level of annotations (e.g., description of the function of a protein, its domains structure, post-translational modifications, variants), a minimal level of redundancy, and high level of integration with other databases.

<http://www.prf.or.jp/en/dbi.html>

The Protein Research Foundation (PRF) in Japan has a collection of protein sequences and synthetic peptides as well as a literature search tool.

<http://www.rcsb.org/pdb/index.html>

The Protein Data Bank (PDB) is the single worldwide repository for the processing and distribution of three-dimensional biological macromolecular structure data (for proteins, RNA, and DNA).

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

The EMBL Macromolecular Structure Database is the European Project for the collection, management, and distribution of data about macromolecular structures, including proteins.

http://www.biochem.ucl.ac.uk/bsm/cath_new/index.html

CATH is a novel hierarchical classification of protein domain structures, which clusters proteins at four major levels: class (C), architecture (A), topology (T), and homologous superfamily (H).

A.8 METABOLIC PATHWAYS

<http://www.genome.ad.jp/kegg/kegg2.html>

Kyoto Encyclopedia of Genes and Genomes (KEGG) is an effort to computerize current knowledge of molecular and cellular biology in terms of the information pathways that consist of interacting molecules or genes and provide links from the gene catalogs produced by genome sequencing projects. The KEGG project is undertaken at the Bioinformatics Center, Institute for Chemical Research, Kyoto University, with support from the Ministry of Education, Culture, Sports, Science, and Technology and the Japan Society for the Promotion of Science.

<http://ecocyc.pangeasystems.com/ecocyc/ecocyc.html>

EcoCyc is a bioinformatics database that describes the genome and the biochemical machinery of the bacterium *Escherichia coli*. The long-term goal of the project is to describe the molecular catalog of the *E. coli* cell, as well as the functions of each of its molecular parts, to facilitate a system-level understanding of *E. coli*.

A.9 EDUCATIONAL RESOURCES

<http://www.sequenceanalysis.com/>

A particularly useful site offering a guide for researchers interested in the analysis of nucleotide or amino acid sequences by using online bioinformatics tools and databases.

<http://www.umass.edu/microbio/rasmol/edsites.htm>

The site contains useful introductory chapters to a variety of structure-based problems in the biological sciences.

A.10 SOFTWARE

Each of the following Web sites contains lists of links and descriptions for useful software that can be downloaded and applied to a variety of problems in the biological sciences.

<http://www.sanger.ac.uk/Software/>

http://www.scsb.utmb.edu/sb_on_net.html

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

<http://www.isrec.isb-sib.ch/software/software.html>

<http://bioinfo.weizmann.ac.il/mb/software.html>