

Table S1 Some important bioinformatics resources and their applications

Institutions		
Name	URL	Description
European Bioinformatics Institute (EBI)	www.ebi.ac.uk	European Centre for research and services in bioinformatics, part of EMBL (European Molecular Biology Laboratory)
National Center for Biotechnology Information	www.ncbi.nih.gov	Centre for bioinformatics research and services in the United States, part of the National Institutes of Health.
Primary Databases		
Name	URL	Description
International Nucleotide Sequence Database Collaboration (INSDC)	www.insdc.org	Collaboration that maintains archival repository of nucleotide sequence data
EMBL Nucleotide Sequence Database	www.ebi.ac.uk/embl	European partner in the INSDC
Genbank	www.ncbi.nih.gov/Genbank	American partner in the INSDC
DNA Database of Japan (DDBJ)		Japanese partner in the INSDC
UniProt Knowledgebase (UniProtKB)	www.uniprot.org	Central access point for protein-related information maintained by the UniProt consortium
Swiss-Prot		Manually curated portion of the UniProtKB, commonly regarded as a gold standard for functional annotation
Gene	www.ncbi.nih.gov/query.fcgi?db=gene	Database of known genes maintained by NCBI. Provides reference system for NCBI annotation efforts.
RefSeq	www.ncbi.nih.gov/RefSeq	Database of reference DNA and protein sequences maintained by NCBI. Closely linked to Gene.
Ensembl	www.ensembl.org	Metazoan genome annotation database
Saccharomyces Genome Database (SGD)	www.yeastgenome.org	<i>Saccharomyces</i> genome database
Mouse Genome Informatics (MGI)	www.informatics.jax.org	Mouse genome database
FlyBase	flybase.bio.indiana.edu	<i>Drosophila</i> genome database
WormBase	www.wormbase.org	<i>Caenorhabditis</i> genome database
GOA	www.ebi.ac.uk/goa	Database of functional annotation, part-curated, part derived from InterPro analysis
IntAct	www.ebi.ac.uk	Database of protein interactions
BioCyc	www.biocyc.org	Collection of pathway and genome databases
World-wide Protein Data Bank (wwPDB)	www.wwpdb.org	Repository for protein structure information
ArrayExpress	www.ebi.ac.uk/arrayexpress	European repository for microarray data
Gene Expression Omnibus (GEO)	www.ncbi.nlm.nih.gov/geo	American repository for microarray data
MEDLINE	www.nlm.nih.gov/pubs/factsheets/jsel.html	Database of information on biomedical literature

PubMed	www.ncbi.nih.gov/entrez/query/static/overview.html	Expanded version of MEDLINE, searchable through Entrez at NCBI
Sequence Analysis		
Name	URL	Description
Smith-Waterman (SW)	Programs running most sequence analysis algorithms are available from many bioinformatics service providers.	Dynamic programming algorithm for sequence analysis, widely considered good, but slow. Various hardware-optimised approximations to SW also exist.
Basic Local Alignment Search Tool (BLAST)		High speed algorithm for sequence comparison.
FAST-AII (FASTA)		Another widely used sequence comparison algorithm
InterProScan	www.ebi.ac.uk/InterProScan	Domain analysis tool incorporating algorithms and annotations from 16 consortium members. Results for known sequences stored in InterPro database.
TMHMM	www.cbs.dtu.dk/services/TMHMM-2.0	Prediction of transmembrane helices in proteins
SWISS-MODEL	swissmodel.expasy.org	Automated protein modelling server
PSIPRED	bioinf.cs.ucl.ac.uk/psipred	Protein structure prediction server, offers access to the GenTHREADER program
Warehouses and web portals		
EBI SRS server	srs.ebi.ac.uk	Primary data warehouse/query engine maintained by EBI
NCBI Entrez server	www.ncbi.nih.gov/database/index.html	Primary data warehouse/query engine maintained by NCBI
Integr8	www.ebi.ac.uk/integr8	Provides structured overview of genes and genomes
iHOP	www.ihop-net.org/UniPub/iHOP	Portal for assessing automatically retrieved literature linked to genes and proteins
Generic database schemas		
GMOD	www.gmod.org	Schema maintained by several model organism databases
GUS	www.gusdb.org	Schema for functional genomics data
Integration tools		
BioMart	www.biomart.org	Data warehousing system designed for easy local install, interface generation and distributed querying
Taverna	taverna.sourceforge.net	Workflow manager for bioinformatics. Part of the MyGRID project.
MyGRID	mygrid.org.uk	Project developing a suite of tools for distributed computing in bioinformatics