The Molecular Biology Database Collection: 2004 update

Michael Y. Galperin*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

ABSTRACT

The Molecular Biology Database Collection is a public resource listing key databases of value to the biologist, including those featured in this issue of *Nucleic Acids Research*, and other high-quality databases. All databases included in this Collection are freely available to the public. This listing aims to serve as a convenient starting point for searching the web for reliable information on various aspects of molecular biology, biochemistry and genetics. This year's update includes 548 databases, 162 more than the previous one. The databases are organized in a hierarchical classification that should simplify finding the right database for each given task. Each database in the list comes with a recently updated brief description. The database list and the database descriptions can be accessed online at the *Nucleic Acids Research* web site http://nar.oupjournals.org/.

The great challenge in biological research today is how to turn data into knowledge. I have met people who think data is knowledge but these people are then striving for a means of turning knowledge into understanding.

Sydney Brenner. *The Scientist* 16[6]:12, March 18, 2002

COMMENTARY

The 50th anniversary of Watson and Crick's discovery of the DNA double-helix structure last year was marked by the formal completion of the Human Genome Project (1). In the feast of the ever-increasing pace of DNA sequencing, this 3-billion-letter text was unraveled barely 8 years after the completion of the first genome of a cellular life form, the 2000-fold smaller genome of *Haemophilus influenzae* strain Rd KW20 (2). The history of genome sequencing shows that the amount of accumulated DNA sequence data keeps growing at an exponential rate, nearly doubling every year. Genomes of more than a hundred organisms from all major phylogenetic lineages are already available in GenBank and sequencing of many more is currently under way. These sequence data have stimulated research in more areas of life sciences than anybody could have expected just a few years ago. They have already spawned a revolution in microbiology

and, with the progress of eukaryotic genome projects, will soon impact such areas as entomology and veterinary science. Unfortunately, a great majority of biologists, chemists and physicians still have only a very vague idea of how to use these data or even where to find them. For the last 10 years, *Nucleic Acids Research* has been devoting a special issue to the molecular biology database compilation (3), which, together with the recently launched *NAR* Web Server Issue (4), should help meet the challenge of bringing molecular biology data and computational tools to every laboratory bench and making them an integral part of every biologist's tool kit.

In order to have a real impact, molecular biology data need to be properly organized and curated. The database structure should help in improving the signal-to-noise ratio, making it easy to extract useful information. In the very beginning of the genome sequencing era, Walter Gilbert and colleagues warned of 'database explosion', stemming from the exponentially increasing amount of incoming DNA sequence and the unavoidable errors it contains (5). Luckily, this threat has not materialized so far, due to the corresponding growth in computational power and storage capacity and the strict requirements for sequence accuracy. However, having managed so far to cope with data accumulation in terms of the capacity to store sequence data, we have fared much worse in terms of our capacity to comprehend these data. Even though at least 50–70% of proteins encoded in any genome are homologous to proteins that are already in the database, every newly sequenced genome encodes hundreds or thousands of novel proteins that have never been seen before and whose very existence in the live cell, let alone function, is uncertain. Even for *Escherichia coli*, arguably the best-studied organism on this planet, almost a half of the ~4288 proteins encoded in the genome have never been studied experimentally and, at the current rate of their experimental characterization, it could take many years before this task is completed (6). For eukaryotes with their much larger genome sizes, complex gene organization, multitude of regulatory interactions and the abundance of proteins without evident enzymatic activities, the task of comprehending the genomic information is infinitely more challenging.

In a way, the proliferation of molecular biology databases can be seen as a natural response of the biological community as a whole to the challenge of staying current in this ever-increasing flow of information that faces every individual biologist. It allows one to rely on the expertise of others, typically well-known professionals in the field, to sort through

*Tel: +1 301 435 5910; Fax: +1 301 435 7794; Email: galperin@ncbi.nlm.nih.gov

the raw data and come up with a curated digest, not unlike the immensely popular mini-reviews that now show up in nearly every journal. The difference, of course, is that the databases are freely available on the web and are continuously updated, which makes each of them a live resource, rather than just a snapshot.

So what's the purpose of this compilation in the era of Google, HotBot, Overture and dozens of other search engines? Unfortunately, these engines rank web sites by popularity, not by their relevance to scientists, and are unable to discriminate between reliable and unreliable web sites. Thus, a recent Google search for 'mitochondrial myopathy' returned a huge number of links, many of them relevant, but clicking the very first of those links launched a series of new windows offering a trial subscription to a web service, cheap airline tickets, and several more items not to be named here. Even the target window was mostly devoted to the importance of treating mitochondrial myopathies with a vegetarian diet, hardly what I was looking for. In contrast, the same search of the OMIM database yielded just 38 links, all of which were relevant and provided reliable information on this family of diseases. Thus, I hope that this compilation will help bridge the 'digital divide' between those researchers who create molecular biology databases and those that would benefit most from using them but are either unaware that such databases exist or are just too busy to spend valuable time sorting through dubious web links.

Certainly, this listing is far from being complete. In order to be included, databases had to provide added value to the user and be publicly available to anyone without any need for registration or subscription. The latter requirement left out a number of useful and otherwise worthy databases, previously described in *NAR*, such as the Asthma and Allergy Gene Database (7) or BioKnowledge Library databases YPD, PombePD and WormPD (8) from Proteome Inc., currently owned by Incyte. However, exceptions were made for the databases described in this volume and for those databases that allow some limited access without registration. Naturally, the database list has grown since the last issue. This edition includes 548 databases, an increase of 162 over the last year's list (3). While most of these new databases have been created only recently, we have also added some well-known databases that were missing before, such as Colibri, FSSP (now superceded by Dali but still widely used) and GtRDB. We have also introduced a hierarchical classification of databases that should simplify searching the list. Due to the limitations of every classification, in the online version of this list, available at http://nar.oupjournals.org/, some databases appear more than once. Doing that in the print version (Table 1) would have consumed too much valuable space.

Suggestions for the inclusion of additional database resources in this Collection are encouraged and should be directed to Dr Alex Bateman at nardatabase@mrc-lmb. cam.ac.uk and to the author at galperin@ncbi.nlm.nih.gov.

REFERENCES

1. Collins,F.S., Morgan,M. and Patrinos,A. (2003) The Human Genome Project: lessons from large-scale biology. *Science*, **300**, 286–290.
2. Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J.-F., Dougherty,B.A., Merrick,J.M. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.
3. Baxevanis,A.D. (2003) The Molecular Biology Database Collection: 2003 update. *Nucleic Acids Res.*, **31**, 1–12.
4. Editorial. (2003) *Nucleic Acids Res.*, **31**, 3289.
5. Bhatia,U., Robison,K. and Gilbert,W. (1997) Dealing with database explosion: a cautionary note. *Science*, **276**, 1724–1725.
6. Koonin,E.V. and Galperin,M.Y. (2002) *Sequence–Evolution–Function. Computational Approaches in Comparative Genomics*. Kluwer Academic Publishers, Boston, MA.
7. Immervoll,T. and Wjst,M. (1999) Current status of the Asthma and Allergy Database. *Nucleic Acids Res.*, **27**, 213–214.
8. Costanzo,M.C., Crawford,M.E., Hirschman,J.E., Kranz,J.E., Olsen,P., Robertson,L.S., Skrzypek,M.S., Braun,B.R., Hopkins,K.L., Kondu,P. *et al.* (2001) YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.*, **29**, 75–79.

Table 1. Molecular Biology Database Collection[a]

Database name	Full name and/or description	URL
1. Nucleotide Sequence Databases		
1.1. International Nucleotide Sequence Database Collaboration		
GenBank	An annotated collection of all publicly available nucleotide and protein sequences	http://www.ncbi.nlm.nih.gov/
EMBL Nucleotide Sequence Database	An annotated collection of all publicly available nucleotide and protein sequences	http://www.ebi.ac.uk/embl.html
DDBJ—DNA Data Bank of Japan	An annotated collection of all publicly available nucleotide and protein sequences	http://www.ddbj.nig.ac.jp
1.2. DNA sequences: genes, motifs and regulatory sites		
1.2.1. Coding and coding DNA		
ACLAME	A classification of genetic mobile elements	http://aclame.ulb.ac.be/
CUTG	Codon usage tabulated from GenBank	http://www.kazusa.or.jp/codon/
Genetic Codes	Deviations from the standard genetic code in various organisms and organelles	http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c
HERVd	Human endogenous retrovirus database	http://herv.img.cas.cz
IMGT/LIGM-DB	Immunoglobulin, T cell receptor and MHC nucleotide sequences from human and other vertebrates	http://imgt.cines.fr/cgi-bin/IMGTlect.jv
Imprinted Gene Catalogue	Imprinted genes and parent-of-origin effects in animals	http://www.otago.ac.nz/IGC
Islander	Pathogenicity islands and prophages in bacterial genomes	http://www.indiana.edu/~islander
MICdb	Prokaryotic microsatellites	http://www.cdfd.org.in/micas
STRBase	Short tandem DNA repeats database	http://www.cstl.nist.gov/div831/strbase/
TIGR Gene Indices	Organism-specific databases of EST and gene sequences	http://www.tigr.org/tdb/tgi.shtml
Transterm	Codon usage, start and stop signals	http://uther.otago.ac.nz/Transterm.html
UniGene	Unified clusters of ESTs and full-length mRNA sequences	http://www.ncbi.nlm.nih.gov/UniGene/
UniVec	Vector sequences, adapters, linkers and primers used in DNA cloning, can be used to check for vector contamination	http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html
VectorDB	Characterization and classification of nucleic acid vectors	http://genome-www2.stanford.edu/vectordb/
Xpro	Eukaryotic protein-encoding DNA sequences, both intron-containing and intron-less genes	http://origin.bic.nus.edu.sg/xpro/
1.2.2. Gene structure, introns and exons, splice sites		
ASAP	Alternative spliced isoforms	http://www.bioinformatics.ucla.edu/ASAP
ASD	EBI's alternative splicing database project includes three databases AltSplice, AltExtron and AEdb	http://www.ebi.ac.uk/asd
ASDB	Alternative splicing database: protein products and expression patterns of alternatively-spliced genes	http://hazelton.lbl.gov/~teplitski/alt
EASED	Extended alternatively spliced EST database	http://eased.bioinf.mdc-berlin.de/
EID	Exon–intron database: introns in protein-coding genes	http://mcb.harvard.edu/gilbert/EID/
ExInt	Exon–intron structure of eukaryotic genes	http://intron.bic.nus.edu.sg/exint/exint.html
HS3D	*Homo sapiens* splice sites dataset	http://www.sci.unisannio.it/docenti/rampone/
IDB/IEDB	Intron sequence and evolution databases	http://nutmeg.bio.indiana.edu/intron/index.html
Intronerator	Introns and alternative splicing in *C.elegans* and *C.briggsae*	http://www.cse.ucsc.edu/~kent/intronerator/
SpliceDB	Canonical and non-canonical mammalian splice sites	http://genomic.sanger.ac.uk/spldb/SpliceDB.html

Table 1. *Continued*

Database name	Full name and/or description	URL
SpliceNest	A tool for visualizing splicing of genes from EST data	http://splicenest.molgen.mpg.de/
YIDB	Yeast nuclear and mitochondrial intron sequences	http://www.embl-heidelberg.DE/ExternalInfo/seraphin/yidb.html

1.2.3. Transcriptional regulator sites and transcription factors

Database name	Full name and/or description	URL
ACTIVITY	Functional DNA/RNA site activity	http://util.bionet.nsc.ru/databases/activity.html
DBTBS	*Bacillus subtilis* promoters and transcription factors	http://dbtbs.hgc.jp/
DBTSS	A database of transcriptional start sites	http://dbtss.hgc.jp/
DPInteract	Binding sites for *E.coli* DNA-binding proteins	http://arep.med.harvard.edu/dpinteract
EPD	Eukaryotic promoter database	http://www.epd.isb-sib.ch
HemoPDB	Hematopoietic promoter database: transcriptional regulation in hematopoiesis	http://bioinformatics.med.ohio-state.edu/HemoPDB
HvrBase	Primate mitochondrial DNA control region sequences	http://www.hvrbase.org/
JASPAR	PSSMs for transcription factor DNA-binding sites	http://jaspar.cgb.ki.se
PLACE	Plant *cis*-acting regulatory DNA elements	http://www.dna.affrc.go.jp/htdocs/PLACE
PlantCARE	Plant promoters and *cis*-acting regulatory elements	http://intra.psb.ugent.be:8080/PlantCARE/
PlantProm	Plant promoter sequences for RNA polymerase II	http://mendel.cs.rhul.ac.uk/
PRODORIC NET	Prokaryotic database of gene regulation networks	http://prodoric.tu-bs.de/
PromEC	*E.coli* promoters with experimentally-identified transcriptional start sites	http://bioinfo.md.huji.ac.il/marg/promec
SELEX_DB	DNA and RNA binding sites for various proteins, found by systematic evolution of ligands by exponential enrichment	http://wwwmgs.bionet.nsc.ru/mgs/systems/selex/
TESS	Transcription element search system	http://www.cbil.upenn.edu/tess
TRANSCompel	Composite regulatory elements affecting gene transcription in eukaryotes	http://www.gene-regulation.com/pub/databases.html#transcompel
TRANSFAC	Transcription factors and binding sites	http://transfac.gbf.de/TRANSFAC/index.html
TRRD	Transcription regulatory regions of eukaryotic genes	http://www.bionet.nsc.ru/trrd/

2. RNA sequence databases

Database name	Full name and/or description	URL
16S and 23S rRNA Mutation Database	16S and 23S ribosomal RNA mutations	http://ribosome.fandm.edu/
5S rRNA Database	5S rRNA sequences	http://biobases.ibch.poznan.pl/5SData/
Aptamer database	Small RNA/DNA molecules binding nucleic acids, proteins	http://aptamer.icmb.utexas.edu/
ARED	AU-rich element-containing mRNA database	http://rc.kfshrc.edu.sa/ared
Mobile group II introns	A database of group II introns, self-splicing catalytic RNAs	http://www.fp.ucalgary.ca/group2introns/
European rRNA database	All complete or nearly complete rRNA sequences	http://www.psb.ugent.be/rRNA/
GtRDB	Genomic tRNA database	http://rna.wustl.edu/GtRDB
Guide RNA Database	RNA editing in various kinetoplastid species	http://biosun.bio.tu-darmstadt.de/goringer/gRNA/gRNA.html
HIV Sequence Database	HIV RNA sequences	http://hiv-web.lanl.gov/
HyPaLib	Hybrid pattern library: structural elements in classes of RNA	http://bibiserv.techfak.uni-bielefeld.de/HyPa/
IRESdb	Internal ribosome entry site database	http://ifr31w3.toulouse.inserm.fr/IRESdatabase/
miRNA Registry	Database of microRNAs (small non-coding RNAs)	http://www.sanger.ac.uk/Software/Rfam/mirna/
NCIR	Non-canonical interactions in RNA structures	http://prion.bchs.uh.edu/bp_type/
ncRNAs Database	Non-coding RNAs with regulatory functions	http://biobases.ibch.poznan.pl/ncRNA/
PLANTncRNAs	Plant non-coding RNAs	http://www.prl.msu.edu/PLANTncRNAs
Plant snoRNA DB	snoRNA genes in plant species	http://www.scri.sari.ac.uk/plant_snoRNA/
PLMItRNA	Plant mitochondrial tRNA	http://bighost.area.ba.cnr.it/PLMItRNA/

Table 1. *Continued*

Database name	Full name and/or description	URL
PseudoBase	Database of RNA pseudoknots	http://wwwbio.leidenuniv.nl/~Batenburg/PKB.html
RDP	Ribosomal database project: rRNA sequence data	http://rdp.cme.msu.edu
Rfam	Non-coding RNA families	http://www.sanger.ac.uk/Software/Rfam/
RISCC	Ribosomal internal spacer sequence collection	http://ulises.umh.es/RISSC
RNA Modification Database	Naturally modified nucleosides in RNA	http://medlib.med.utah.edu/RNAmods/
RRNDB	rRNA operon numbers in various prokaryotes	http://rrndb.cme.msu.edu/
Small RNA Database	Small RNAs from prokaryotes and eukaryotes	http://mbcr.bcm.tmc.edu/smallRNA
SRPDB	Signal recognition particle database	http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html
Subviral RNA Database	Viroids and viroid-like RNAs	http://subviral.med.uottawa.ca/cgi-bin/home.cgi
tmRNA Website	tmRNA sequences and alignments	http://www.indiana.edu/~tmrna
tmRDB	tmRNA database	http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html
tRNA database	tRNA viewer and sequence editor	http://www.uni-bayreuth.de/departments/biochemie/trna/
UTRdb/UTRsite	5′- and 3′-UTRs of eukaryotic mRNAs	http://bighost.area.ba.cnr.it/srs6/

3. Protein sequence databases

3.1. General sequence databases

EXProt	Sequences of proteins with experimentally verified function	http://www.cmbi.kun.nl/EXProt/
NCBI Protein database	All protein sequences: translated from GenBank and imported from other protein databases	http://www.ncbi.nlm.nih.gov/entrez
PIR	Protein information resource: a collection of protein sequence databases, part of the UniProt project	http://pir.georgetown.edu/
PIR-NREF	PIR's non-redundant reference protein database	http://pir.georgetown.edu/pirwww/pirnref.shtml
PRF	Protein research foundation database of peptides: sequences, literature and unnatural amino acids	http://www.prf.or.jp/en
Swiss-Prot	Curated protein sequence database with a high level of annotation (protein function, domain structure, modifications)	http://www.expasy.org/sprot
TrEMBL	Translations of EMBL nucleotide sequence entries: computer-annotated supplement to Swiss-Prot	http://www.expasy.org/sprot
UniProt	Universal protein knowledgebase: a database of protein sequence from Swiss-Prot, TrEMBL and PIR	http://www.uniprot.org/

3.2. Protein properties

AAindex	Physicochemical properties of amino acids	http://www.genome.ad.jp/aaindex/
ProTherm	Thermodynamic data for wild-type and mutant proteins	http://gibk26.bse.kyutech.ac.jp/jouhou/Protherm/protherm.html

3.3. Protein localization and targeting

DBSubLoc	Database of protein subcellular localization	http://www.bioinfo.tsinghua.edu.cn/dbsubloc.html
MitoDrome	Nuclear-encoded mitochondrial proteins of *Drosophila*	http://bighost.area.ba.cnr.it/BIG/MitoDrome
NESbase	Nuclear export signals database	http://www.cbs.dtu.dk/databases/NESbase
NLSdb	Nuclear localization signals	http://cubic.bioc.columbia.edu/db/NLSdb/
THGS	Transmembrane helices in genome sequences	http://pranag.physics.iisc.ernet.in/thgs/
TMPDB	Experimentally characterized transmembrane topologies	http://bioinfo.si.hirosaki-u.ac.jp/~TMPDB/

3.4. Protein sequence motifs and active sites

ASC	Active sequence collection: biologically active peptides	http://bioinformatica.isa.cnr.it/ASC/
Blocks	Alignments of conserved regions in protein families	http://blocks.fhcrc.org/

Table 1. *Continued*

Database name	Full name and/or description	URL
CSA	Catalytic site atlas: enzyme active sites and catalytic residues in enzymes of known 3D structure	http://www.ebi.ac.uk/thornton-srv/databases/CSA/
COMe	Co-ordination of metals etc.: classification of bioinorganic proteins (metalloproteins and some other complex proteins)	http://www.ebi.ac.uk/come
eMOTIF	Protein sequence motif determination and searches	http://motif.stanford.edu/emotif
Metalloprotein Site Database	Metal-binding sites in metalloproteins	http://metallo.scripps.edu/
O-GlycBase	O- and C-linked glycosylation sites in proteins	http://www.cbs.dtu.dk/databases/OGLYCBASE/
PhosphoBase	Protein phosphorylation sites	http://www.cbs.dtu.dk/databases/PhosphoBase/
PROMISE	Prosthetic centers and metal ions in protein active sites	http://metallo.scripps.edu/PROMISE
PROSITE	Biologically significant protein patterns and profiles	http://www.expasy.org/prosite

3.5. Protein domain databases; protein classification

CDD	Conserved domain database: includes protein domains from Pfam, SMART and COG databases	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
CluSTr	Clusters of Swiss-Prot+TrEMBL proteins	http://www.ebi.ac.uk/clustr
Hits	A database of protein domains and motifs	http://hits.isb-sib.ch/
InterPro	Integrated resource of protein families, domains and functional sites	http://www.ebi.ac.uk/interpro
iProClass	Integrated protein classification database	http://pir.georgetown.edu/iproclass/
MetaFam	Database of protein family annotations	http://metafam.ahc.umn.edu/
PIRSF	Family/superfamily classification of whole proteins	http://pir.georgetown.edu/pirsf/
PRINTS	Hierarchical gene family fingerprints	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/
Pfam	Protein families: multiple sequence alignments and profile hidden Markov models of protein domains	http://www.sanger.ac.uk/Software/Pfam/
PIR-ALN	Curated database of protein sequence alignments	http://pir.georgetown.edu/pirwww/dbinfo/piraln.html
ProClass	Protein families defined by PIR superfamilies and PROSITE patterns	http://pir.georgetown.edu/gfserver/proclass.html
ProDom	Protein domain families	http://www.toulouse.inra.fr/prodom.html
ProtoMap	Hierarchical classification of Swiss-Prot proteins	http://protomap.cornell.edu/
ProtoNet	Hierarchical clustering of Swiss-Prot proteins	http://www.protonet.cs.huji.ac.il/
SBASE	Protein domain sequences and tools	http://www.icgeb.org/sbase
SMART	Simple modular architecture research tool: signalling, extracellular and chromatin-associated protein domains	http://smart.embl-heidelberg.de/
SUPFAM	Grouping of sequence families into superfamilies	http://pauling.mbu.iisc.ernet.in/~supfam
SYSTERS	Systematic re-searching and clustering of proteins	http://systers.molgen.mpg.de/
TIGRFAMs	TIGR protein families adapted for functional annotation	http://www.tigr.org/TIGRFAMs

3.6. Databases of individual protein families

AARSDB	Aminoacyl-tRNA synthetase database	http://rose.man.poznan.pl/aars/index.html
ABCdb	ABC transporters database	http://ir2lcb.cnrs-mrs.fr/ABCdb/
ASPD	Artificial selected proteins/peptides database	http://wwwmgs.bionet.nsc.ru/mgs/gnw/aspd/
BacTregulators	Transcriptional regulators of AraC and TetR families	http://www.bactregulators.org/
CSDBase	Cold shock domain-containing proteins	http://www.chemie.uni-marburg.de/~csdbase/
DExH/D Family Database	DEAD-box, DEAH-box and DExH-box proteins	http://www.helicase.net/dexhd/dbhome.htm
Endogenous GPCR List	G protein-coupled receptors; expression in cell lines	http://www.tumor-gene.org/GPCR/gpcr.html
ESTHER	Esterases and other alpha/beta hydrolase enzymes	http://www.ensam.inra.fr/esther
EyeSite	Families of proteins functioning in the eye	http://eyesite.cryst.bbk.ac.uk/

Table 1. *Continued*

Database name	Full name and/or description	URL
GPCRDB	G protein-coupled receptors database	http://www.gpcr.org/7tm/
Histone Database	Histone fold sequences and structures	http://research.nhgri.nih.gov/histones/
HIV Molecular Immunology Database	HIV epitopes	http://hiv-web.lanl.gov/immunology/
HIV Protease Database	HIV reverse transcriptase and protease sequences	http://hivdb.stanford.edu/
Homeobox Page	Homeobox proteins, classification and evolution	http://www.biosci.ki.se/groups/tbu/homeo.html
Homeodomain Resource	Homeodomain sequences, structures and related genetic and genomic information	http://research.nhgri.nih.gov/homeodomain
HORDE	Human olfactory receptor data exploratorium	http://bioinfo.weizmann.ac.il/HORDE/
InBase	Inteins (protein splicing elements) database: properties, sequences, bibliography	http://www.neb.com/neb/inteins.html
Kabat Database	Sequences of proteins of immunological interest	http://immuno.bme.nwu.edu/
KinG	Ser/Thr/Tyr-specific protein kinases encoded in complete genomes	http://hodgkin.mbu.iisc.ernet.in/~king
Knottins	Database of knottins—small proteins with an unusual 'disulfide through disulfide' knot	http://knottin.cbs.cnrs.fr
LGICdb	Ligand-gated ion channel subunit sequences database	http://www.pasteur.fr/recherche/banques/LGIC/LGIC.html
Lipase Engineering Database	Sequence, structure and function of lipases and esterases	http://www.led.uni-stuttgart.de/
LOX-DB	Mammalian, invertebrate, plant and fungal lipoxygenases	http://www.dkfz-heidelberg.de/spec/lox-db/
MEROPS	Database of proteolytic enzymes (peptidases)	http://www.merops.ac.uk/
MHCPEP	MHC-binding peptides	http://wehih.wehi.edu.au/mhcpep/
MPIMP	Mitochondrial protein import machinery of plants	http://millar3.biochem.uwa.edu.au/~lister/index.html
NPD	Nuclear protein database	http://npd.hgu.mrc.ac.uk/
NucleaRDB	Nuclear receptor superfamily	http://www.receptors.org/NR/
Nuclear Receptor Resource	Nuclear receptor superfamily	http://nrr.georgetown.edu/nrr/nrr.html
NUREBASE	Nuclear hormone receptors database	http://www.ens-lyon.fr/LBMC/laudet/nurebase/nurebase.html
Olfactory Receptor Database	Sequences for olfactory receptor-like molecules	http://ycmi.med.yale.edu/senselab/ordb/
ooTFD	Object-oriented transcription factors database	http://www.ifti.org/ootfd
PKR	Protein kinase resource: sequences, enzymology, genetics and molecular and structural properties	http://pkr.sdsc.edu/
PLANT-PIs	Plant protease inhibitors	http://bighost.area.ba.cnr.it/PLANT-PIs
PlantsP/PlantsT	Plant proteins involved in phosphorylation and membrane transport	http://plantsp.sdsc.edu/
Prolysis	Proteases and natural and synthetic protease inhibitors	http://delphi.phys.univ-tours.fr/Prolysis/
REBASE	Restriction enzymes and associated methylases	http://rebase.neb.com/rebase/rebase.html
Ribonuclease P Database	RNase P sequences, alignments and structures	http://www.mbio.ncsu.edu/RNaseP/home.html
RPG	Ribosomal protein gene database	http://ribosome.miyazaki-med.ac.jp/
RTKdb	Receptor tyrosine kinase sequences	http://pbil.univ-lyon1.fr/RTKdb/
S/MARt dB	Nuclear scaffold/matrix attached regions	http://smartdb.bioinf.med.uni-goettingen.de/
SDAP	Structural database of allergenic proteins and food allergens	http://fermi.utmb.edu/SDAP
SENTRA	Sensory signal transduction proteins	http://wit.mcs.anl.gov/WIT2/Sentra/HTML/sentra.html
SEVENS	7-transmembrane helix receptors (G-protein-coupled)	http://sevens.cbrc.jp/
SRPDB	Proteins of the signal recognition particles	http://bio.lundberg.gu.se/dbs/SRPDB/SRPDB.html
TrSDB	Transcription factor database	http://ibb.uab.es/trsdb
VIDA	Homologous viral protein families database	http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html
VKCDB	Voltage-gated potassium channel database	http://vkcdb.biology.ualberta.ca/
Wnt Database	Wnt proteins and phenotypes	http://www.stanford.edu/~rnusse/wntwindow.html

Table 1. *Continued*

Database name	Full name and/or description	URL
4. Structure Databases		
4.1. Small molecules		
CSD	Cambridge structural database: crystal structure information for organic and metal-organic compounds	http://www.ccdc.cam.ac.uk/prods/csd/csd.html
HIC-Up	Hetero-compound Information Centre—Uppsala	http://xray.bmc.uu.se/hicup
AANT	Amino acid–nucleotide interaction database	http://aant.icmb.utexas.edu/
Klotho	Collection and categorization of biological compounds	http://www.biocheminfo.org/klotho
LIGAND	Chemical compounds and reactions in biological pathways	http://www.genome.ad.jp/ligand/
4.2. Carbohydrates		
CCSD	Complex carbohydrate structure database (CarbBank)	http://bssv01.lancs.ac.uk/gig/pages/gag/carbbank.htm
Glycan	Carbohydrate database, part of the KEGG system	http://glycan.genome.ad.jp/
GlycoSuiteDB	N- and O-linked glycan structures and biological sources	http://www.glycosuite.com/
Monosaccharide Browser	Space filling Fischer projections of monosaccharides	http://www.jonmaber.demon.co.uk/monosaccharide
SWEET-DB	Annotated carbohydrate structure and substance information	http://www.dkfz-heidelberg.de/spec2/sweetdb/
4.3. Nucleic acid structure		
NDB	Nucleic acid-containing structures	http://ndbserver.rutgers.edu/
NTDB	Thermodynamic data for nucleic acids	http://ntdb.chem.cuhk.edu.hk/
RNABase	RNA-containing structures from PDB and NDB	http://www.rnabase.org/
SCOR	Structural classification of RNA: RNA motifs by structure, function and tertiary interactions	http://scor.lbl.gov/
4.4. Protein structure		
ArchDB	Automated classification of protein loop structures	http://gurion.imim.es/archdb
ASTRAL	Sequences of domains of known structure, selected subsets and sequence-structure correspondences	http://astral.stanford.edu/
BAliBASE	A database for comparison of multiple sequence alignments	http://www-igbmc.u-strasbg.fr/BioInfo/BAliBASE2/index.html
BioMagResBank	NMR spectroscopic data for proteins and nucleic acids	http://www.bmrb.wisc.edu/
CADB	Conformational angles in proteins database	http://cluster.physics.iisc.ernet.in/cadb/
CATH	Protein domain structures database	http://www.biochem.ucl.ac.uk/bsm/cath_new
CE	3D Protein structure alignments	http://cl.sdsc.edu/ce.html
CKAAPs DB	Structurally-similar proteins with dissimilar sequences	http://ckaap.sdsc.edu/
Dali	Protein fold classification using the Dali search engine	http://www.bioinfo.biocenter.helsinki.fi:8080/dali/
Decoys 'R' Us	Computer-generated protein conformations	http://dd.stanford.edu/
DisProt	Database of Protein Disorder: information about proteins that lack fixed 3D structure in their native states	http://divac.ist.temple.edu/disprot
DomIns	Domain insertions in known protein structures	http://stash.mrc-lmb.cam.ac.uk/DomIns
DSDBASE	Native and modeled disulfide bonds in proteins	http://www.ncbs.res.in/~faculty/mini/dsdbase/dsdbase.html
DSMM	Database of simulated molecular motions	http://projects.villa-bosch.de/dbase/dsmm/
eF-site	Electrostatic surface of Functional site: electrostatic potentials and hydrophobic properties of the active sites	http://ef-site.protein.osaka-u.ac.jp/eF-site

Table 1. *Continued*

Database name	Full name and/or description	URL
FSSP	Fold classification based on structure-structure alignment of proteins, currently maintained as Dali database	http://www.ebi.ac.uk/dali/fssp
Gene3D	Precalculated structural assignments for whole genomes	http://www.biochem.ucl.ac.uk/bsm/cath_new/Gene3D/
GTD	Genomic threading database: structural annotations of complete genomes	http://bioinf.cs.ucl.ac.uk/GTD
GTOP	Protein fold predictions from genome sequences	http://spock.genes.nig.ac.jp/~genome/
Het-PDB Navi	Hetero-atoms in protein structures	http://daisy.nagahama-i-bio.ac.jp/golab/hetpdbnavi.html
HOMSTRAD	Homologous structure alignment database: curated structure-based alignments for protein families	http://www-cryst.bioc.cam.ac.uk/homstrad
IMB Jena Image Library	Visualization and analysis of 3D biopolymer structures	http://www.imb-jena.de/IMAGE.html
IMGT/3Dstructure-DB	Sequences and 3D structures of vertebrate immunoglobulins, T cell receptors and MHC proteins	http://imgt3d.igh.cnrs.fr
ISSD	Integrated sequence-structure database	http://www.protein.bio.msu.su/issd
LPFC	Library of protein family core structures	http://www-smi.stanford.edu/projects/helix/LPFC
MMDB	NCBI's database of 3D structures, part of NCBI Entrez	http://www.ncbi.nlm.nih.gov/Structure
E-MSD	EBI's macromolecular structure database	http://www.ebi.ac.uk/msd
ModBase	Annotated comparative protein structure models	http://salilab.org/modbase
MolMovDB	Database of macromolecular movements: descriptions of protein and macromolecular motions, including movies	http://bioinfo.mbb.yale.edu/MolMovDB/
PALI	Phylogeny and alignment of homologous protein structures	http://pauling.mbu.iisc.ernet.in/~pali
PASS2	Structural motifs of protein superfamilies	http://ncbs.res.in/~faculty/mini/campass/pass.html
PepConfDB	A database of peptide conformations	http://202.41.70.49:8080/pepconfdb/index.htm
PDB	Protein structure databank: all publicly available 3D structures of proteins and nucleic acids	http://www.rcsb.org/pdb
PDB-REPRDB	Representative protein chains, based on PDB entries	http://www.cbrc.jp/pdbreprdb/
PDBsum	Summaries and analyses of PDB structures	http://www.biochem.ucl.ac.uk/bsm/pdbsum
SCOP	Structural classification of proteins	http://scop.mrc-lmb.cam.ac.uk/scop
Sloop	Classification of protein loops	http://www-cryst.bioc.cam.ac.uk/~sloop/
Structure-Superposition Database	Pairwise superposition of TIM-barrel structures	http://ssd.rbvi.ucsf.edu/
SWISS-MODEL Repository	Database of annotated 3D protein structure models	http://swissmodel.expasy.org/repository
SUPERFAMILY	Assignments of proteins to structural superfamilies	http://supfam.org/
SURFACE	Surface residues and functions annotated, compared and evaluated: a database of protein surface patches	http://cbm.bio.uniroma2.it/surface
TargetDB	Target data from worldwide structural genomics projects	http://targetdb.pdb.org/
3D-GENOMICS	Structural annotations for complete proteomes	http://www.sbg.bio.ic.ac.uk/3dgenomics
TOPS	Topology of protein structures database	http://www.tops.leeds.ac.uk

5. Genomics Databases (non-human)

5.1. Genome annotation terms, ontologies and nomenclature

Genew	Human gene nomenclature: approved gene symbols	http://www.gene.ucl.ac.uk/nomenclature
GO	Gene onthology consortium database	http://www.geneontology.org/
GOA	Gene onthology annotation project	http://www.ebi.ac.uk/GOA

Table 1. *Continued*

Database name	Full name and/or description	URL
IUBMB Nomenclature database	Nomenclature of enzymes, membrane transporters, electron transport proteins and other proteins	http://www.chem.qmul.ac.uk/iubmb
IUPAC Nomenclature database	Nomenclature of biochemical and organic compounds approved by the IUBMB-IUPAC Joint Commission	http://www.chem.qmul.ac.uk/iupac
IUPHAR-RD	The International Union of Pharmacology recommendations on receptor nomenclature and drug classification	http://www.iuphar-db.org/iuphar-rd/
PANTHER	Gene products organized by biological function	http://panther.celera.com/
SOURCE	Functional genomic resource for annotations ontologies and expression data	http://source.stanford.edu/
UMLS	Unified medical language system	http://umlsks.nlm.nih.gov/
5.1.1. Taxonomy and Identification		
ICB	*gyrB* database for identification and classification of bacteria	http://www.mbio.co.jp/icb
NCBI Taxonomy	Names and taxonomic lineages of all organisms in GenBank	http://www.ncbi.nlm.nih.gov/Taxonomy/
RIDOM	rRNA-based differentiation of medical microorganisms	http://www.ridom-rdna.de/
RDP	Ribosomal database project	http://rdp.cme.msu.edu
Tree of Life	Information on phylogeny and biodiversity	http://phylogeny.arizona.edu/tree/phylogeny.html
5.2. General genomics databases		
COG	Clusters of orthologous groups of proteins from unicellular microorganisms	http://www.ncbi.nlm.nih.gov/COG
CORG	Comparative regulatory genomics: conserved non-coding sequence blocks	http://corg.molgen.mpg.de/
DEG	Database of essential genes from bacteria and yeast	http://tubic.tju.edu.cn/deg
EBI Genomes	EBI's collection of databases for the analysis of complete and unfinished viral, pro- and eukaryotic genomes	http://www.ebi.ac.uk/genomes
EGO	Eukaryotic gene orthologs: orthologous DNA sequences in the TIGR gene indices	http://www.tigr.org/tdb/tgi/ego/
EMGlib	Enhanced microbial genomes library: completely sequenced genomes of unicellular organisms	http://pbil.univ-lyon1.fr/emglib/emglib.html
Entrez Genomes	NCBI's collection of databases for the analysis of complete and unfinished viral, pro- and eukaryotic genomes	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome
ERGOLight	Integrated biochemical data on seven bacterial genomes: publicly available portion of the ERGO database	http://www.ergo-light.com/ERGO
FusionDB	Database of bacterial and archaeal gene fusion events	http://igs-server.cnrs-mrs.fr/FusionDB
Genome information broker	DDBJ's collection of databases for the analysis of complete and unfinished viral, pro- and eukaryotic genomes	http://gib.genes.nig.ac.jp
GOLD	Genomes online database: a listing of completed and ongoing genome projects	http://www.genomesonline.org/
TIGR Microbial Database	Lists of completed and ongoing genome projects with links to complete genome sequences	http://www.tigr.org/tdb/mdb/mdbcomplete.html
HGT-DB	Putative horizontally transferred genes in prokaryotic genomes	http://www.fut.es/~debb/HGT/
KEGG	Kyoto encyclopedia of genes and genomes: integrated suite of databases on genes, proteins, and metabolic pathways	http://www.genome.ad.jp/kegg

Table 1. *Continued*

Database name	Full name and/or description	URL
MBGD	Microbial genome database for comparative analysis	http://mbgd.genome.ad.jp/
ORFanage	Database of orphan ORFs (ORFs with no homologs) in complete microbial genomes	http://www.cs.bgu.ac.il/~nomsiew/ORFans
PACRAT	Archaeal and bacterial intergenic sequence features	http://www.biosci.ohio-tate.edu/~pacrat
PEDANT	Results of an automated analysis of genomic sequences	http://pedant.gsf.de
TIGR Comprehensive Microbial Resource	Various data on complete microbial genomes: uniform annotation, properties of DNA and predicted proteins	http://www.tigr.org/CMR
TransportDB	Predicted membrane transporters in complete genomes, classified according to the TC classification system	http://www.membranetransport.org
WIT	What is there? Metabolic reconstruction for completely sequenced microbial genomes	http://wit.mcs.anl.gov/WIT2/

5.3. Organism-specific genomic databases

5.3.1. Viruses

HCVDB	The hepatitis C virus database	http://hepatitis.ibcp.fr/
HIV Drug Resistance Database	Mutations in HIV genes that confer resistance to anti-HIV drugs	http://resdb.lanl.gov/Resist_DB/default.htm
VirGen	Annotated and curated database for complete viral genome sequences	http://bioinfo.ernet.in/virgen/virgen.html

5.3.2. Prokaryotes

5.3.2.1. *Escherichia coli*

ASAP	A systematic annotation package for community analysis of *E.coli* and related genomes	https://asap.ahabs.wisc.edu/annotation/php/ASAP1.htm
CCDB	CyberCell database: *E.coli* database at U. Alberta	http://redpoll.pharmacy.ualberta.ca/CCDB
coliBase	A database for *E.coli*, *Salmonella* and *Shigella*	http://colibase.bham.ac.uk/
Colibri	*E.coli* genome database at Institut Pasteur	http://genolist.pasteur.fr/Colibri/
Essential genes in *E.coli*	First results of an *E.coli* gene deletion project	http://magpie.genome.wisc.edu/~chris/essential.html
GenoBase	*E.coli* genome database at Nara Institute	http://ecoli.aist-nara.ac.jp/
GenProtEC	*E.coli* K-12 genome and proteome database	http://genprotec.mbl.edu
PEC	Profiling of *E.coli* chromosome	http://shigen.lab.nig.ac.jp/ecoli/pec
EcoCyc	*E.coli* K-12 genes, metabolic pathways, transporters, and gene regulation	http://ecocyc.org/
EcoGene	Sequence and literature data on *E.coli* genes and proteins	http://bmb.med.miami.edu/EcoGene/EcoWeb/
RegulonDB	Transcriptional regulation and operon organization in *E.coli*	http://www.cifn.unam.mx/Computational_Genomics/regulondb/

5.3.2.2. *Bacillus subtilis*

BSORF	*Bacillus subtilis* genome database at Kyoto U.	http://bacillus.genome.ad.jp/
NRSub	Non-redundant *Bacillus subtilis* database at U. Lyon	http://pbil.univ-lyon1.fr/nrsub/nrsub.html
SubtiList	*Bacillus subtilis* genome database at Institut Pasteur	http://genolist.pasteur.fr/SubtiList/

5.3.2.3. Other bacteria

BioCyc	Pathway/genome databases for many bacteria	http://biocyc.org/
CampyDB	Database for *Campylobacter* genome analysis	http://campy.bham.ac.uk/
ClostriDB	Finished and unfinished genomes of *Clostridium* spp.	http://clostri.bham.ac.uk/
CyanoBase	Cyanobacterial genomes	http://www.kazusa.or.jp/cyano
LeptoList	*Leptospira interrogans* genome	http://bioinfo.hku.hk/LeptoList
MolliGen	Genomic data on mollicutes	http://cbi.labri.fr/outils/molligen/

Table 1. *Continued*

Database name	Full name and/or description	URL
RsGDB	*Rhodobacter sphaeroides* genome	http://www-mmg.med.uth.tmc.edu/sphaeroides
5.3.3. Unicellular eukaryotes		
5.3.3.1. Yeast		
SGD	*Saccharomyces* genome database	http://www.yeastgenome.org/
CYGD	MIPS Comprehensive yeast genome database	http://mips.gsf.de/proj/yeast
Génolevures	A comparison of *S.cerevisiae* and 14 other yeast species	http://cbi.labri.fr/Genolevures
MitoPD	Yeast mitochondrial protein database	http://bmerc-www.bu.edu/mito
SCMD	*Saccharomyces cerevisiae* morphological database: micrographs of budding yeast mutants	http://yeast.gi.k.u-tokyo.ac.jp/
SCPD	*Saccharomyces cerevisiae* promoter database	http://cgsigma.cshl.org/jian
TRIPLES	Transposon-insertion phenotypes, localization, and expression in *Saccharomyces*	http://ygac.med.yale.edu/triples/
YDPM	Yeast deletion project and mitochondria database	http://www-deletion.stanford.edu/YDPM/YDPM_index.html
Yeast Intron Database	Ares laboratory database of spliceosomal introns in *S.cerevisiae*	http://www.cse.ucsc.edu/research/compbio/yeast_introns.html
Yeast snoRNA Database	Yeast small nucleolar RNAs	http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html
yMGV	Yeast microarray global viewer	http://www.transcriptome.ens.fr/ymgv/
5.3.3.2. Other unicellular eukaryotes		
ApiEST-DB	EST sequences from various Apicomplexan parasites	http://www.cbil.upenn.edu/paradbs-servlet
CryptoDB	*Cryptosporidium parvum* genome database	http://cryptodb.org/
DictyBase	Genome information, literature and experimental resources for *Dictyostelium discoideum*	http://dictybase.org/
Full-Malaria	Full-length cDNA library from erythrocytic-stage *Plasmodium falciparum*	http://fullmal.ims.u-tokyo.ac.jp/
GeneDB	Curated database for *Trypanosoma brucei*, *Leishmania major*, *S.pombe* and other Sanger-sequenced genomes	http://www.genedb.org/
PlasmoDB	*Plasmodium* genome database	http://plasmodb.org/
TcruziDB	*Trypanosoma cruzi* genome database	http://tcruzidb.org/
ToxoDB	*Toxoplasma gondii* genome database	http://toxodb.org/
5.3.4. Plants		
5.3.4.1. General plant databases		
CropNet	Genome mapping in crop plants	http://ukcrop.net/
FLAGdb++	Integrative database about plant genomes	http://genoplante-info.infobiogen.fr/FLAGdb/
GénoPlante-Info	Plant genomic data from the Génoplante consortium	http://genoplante-info.infobiogen.fr/
GrainGenes	Molecular and phenotypic information on wheat, barley, rye, triticale and oats	http://wheat.pw.usda.gov or http://www.graingenes.org
Mendel	Database of plant EST and STS sequences annotated with gene family information	http://www.mendel.ac.uk/
PHYTOPROT	Clusters of (predicted) plant proteins	http://genoplante-info.infobiogen.fr/phytoprot
PlantGDB	Plant genome database: actively-transcribed plant genomic sequences	http://www.plantgdb.org/
Sputnik	Plant EST clustering and functional annotation	http://mips.gsf.de/proj/sputnik
TIGR plant repeat database	Classification of repetitive sequences in plant genomes	http://www.tigr.org/tdb/e2k1/plant.repeats
TropGENE DB	Genetic and genomic information about tropical crops: sugarcane, banana, cocoa	http://tropgenedb.cirad.fr/

Table 1. *Continued*

Database name	Full name and/or description	URL
5.3.4.2. *Arabidopsis thaliana*		
ARAMEMNON	*Arabidopsis thaliana* membrane proteins and transporters	http://aramemnon.botanik.uni-koeln.de/
AthaMap	Genome-wide map of putative transcription factor binding sites in *Arabidopsis thaliana*	http://www.athamap.de/
CATMA	Complete *Arabidopsis* transcriptome microarray: gene sequence tags	http://www.catma.org
FLAGdb/FST	*Arabidopsis thaliana* T-DNA transformants	http://genoplante-info.infobiogen.fr/
MAtDB	MIPS *Arabidopsis thaliana* database	http://mips.gsf.de/proj/thal/db
SeedGenes	Genes essential for *Arabidopsis* development	http://www.seedgenes.org/
TAIR	The *Arabidopsis* information resource	http://www.arabidopsis.org/
5.3.4.3. Rice		
BGI-RISe	Beijing genomics institute rice information system	http://rise.genomics.org.cn/
INE	Integrated rice genome explorer	http://rgp.dna.affrc.go.jp/giot/INE.html
IRIS	International rice information system: all rice data	http://www.iris.irri.org/
MOsDB	MIPS *Oryza sativa* database	http://mips.gsf.de/proj/rice
Oryzabase	Rice genetics and genomics	http://www.shigen.nig.ac.jp/rice/oryzabase/
RiceGAAS	Rice genome automated annotation system	http://ricegaas.dna.affrc.go.jp/
Rice PIPELINE	Unification tool for rice databases	http://cdna01.dna.affrc.go.jp/PIPE
RPD	Rice proteome database	http://gene64.dna.affrc.go.jp/RPD/
5.3.4.4. Other plants		
MaizeGDB	Maize genetics and genomics database, a successor to MaizeDB and ZmDB databases	http://www.maizegdb.org/
MGI	*Medicago* genome initiative: ESTs, gene expression and proteomic data	http://xgi.ncgr.org/mgi
MtDB	*Medicago trunculata* genome	http://www.medicago.org/MtDB
SGMD	Soybean genomics and microarray database	http://psi081.ba.ars.usda.gov/SGMD/default.htm
5.3.5. Fungi		
CADRE	Central *Aspergillus* data repository	http://www.cadre.man.ac.uk/
COGEME	Phytopathogenic fungi and oomycete EST database	http://cogeme.ex.ac.uk
MagnaportheDB	*Magnaporthe grisea* integrated physical/genetic map	http://www.fungalgenomics.ncsu.edu/Projects/mgdatabase/int.htm
MNCDB	MIPS *Neurospora crassa* database	http://mips.gsf.de/proj/neurospora/
Phytophthora Genome Consortium Database	ESTs from *Phytophthora infestans* and *P.sojae*	https://xgi.ncgr.org/pgc
5.3.6. Invertebrates		
5.3.6.1. *Caenorhabditis elegans*		
C.elegans Project	Genome sequencing data at the Sanger Institute	http://www.sanger.ac.uk/Projects/C_elegans
Intronerator	Introns and alternative splicing in *C.elegans* and *C.briggsae*	http://www.cse.ucsc.edu/~kent/intronerator/
RNAiDB	RNAi phenotypic analysis of *C.elegans* genes	http://www.rnai.org/
WILMA	*C.elegans* annotation database	http://www.came.sbg.ac.at/wilma/
WorfDB	*C.elegans* ORFeome	http://worfdb.dfci.harvard.edu/
WormBase	Data repository for *C.elegans* and *C.briggsae*: curated genome annotation, genetic and physical maps, pathways	http://www.wormbase.org/
5.3.6.2. *Drosophila melanogaster*		
FlyBase	*Drosophila* sequences and genomic information	http://flybase.bio.indiana.edu/
GadFly	Genome annotation database of *Drosophila*	http://www.fruitfly.org
FlyBrain	Database of the *Drosophila* nervous system	http://flybrain.neurobio.arizona.edu
FlyTrap	*Drosophila* transgenic lines created using an intron protein trap strategy	http://flytrap.med.yale.edu/

Table 1. *Continued*

Database name	Full name and/or description	URL
InterActive Fly	*Drosophila* genes and their roles in development	http://sdb.bio.purdue.edu/fly/aimain/1aahome.htm
Drosophila microarray centre	Data and tools for *Drosophila* gene expression studies	http://www.flyarrays.com/fruitfly
5.3.6.3. Other invertebrates		
AppaDB	A database on the nematode *Pristionchus pacificus*	http://appadb.eb.tuebingen.mpg.de
CnidBase	Cnidarian evolution and gene expression database	http://cnidbase.bu.edu/
Nematode.net	Parasitic nematode sequencing project	http://nematode.net/
NEMBASE	Nematode sequence and functional data database	http://www.nematodes.org

6. Metabolic Enzymes and Pathways; Signaling Pathways

6.1. Enzymes and Enzyme Nomenclature

ENZYME	Enzyme nomenclature and properties	http://www.expasy.org/enzyme
BRENDA	Enzyme names and properties: sequence, structure, specificity, stability, reaction parameters, isolation data	http://www.brenda.uni-koeln.de
IntEnz	Integrated enzyme database and enzyme nomenclature	http://www.ebi.ac.uk/intenz
Enzyme Nomenclature	IUBMB Nomenclature Committee recommendations	http://www.chem.qmw.ac.uk/iubmb/enzyme

6.2. Metabolic Pathways

KEGG	Kyoto encyclopedia of genes and genomes: metabolic and regulatory pathways encoded in complete genomes	http://www.genome.ad.jp/kegg
MetaCyc	Metabolic pathways and enzymes from various organisms	http://metacyc.org
PathDB	Biochemical pathways, compounds and metabolism	http://www.ncgr.org/pathdb
UM-BBD	University of Minnesota biocatalysis and biodegradation database: microbial catabolism and biotransformations	http://umbbd.ahc.umn.edu/
WIT2	Integrated system for functional curation and development of metabolic models	http://wit.mcs.anl.gov/WIT2/

6.3. Intermolecular Interactions and Signaling Pathways

aMAZE	A system for the annotation, management and analysis of biochemical and signaling pathway networks	http://www.amaze.ulb.ac.be/
BIND	Biomolecular interaction network database	http://www.bind.ca
BioCarta	Online maps of metabolic and signaling pathways	http://www.biocarta.com/genes/allPathways.asp
BRITE	Biomolecular relations in information transmission and expression, part of the KEGG system	http://www.genome.ad.jp/brite
DIP	Database of interacting proteins: experimentally determined protein–protein interactions	http://dip.doe-mbi.ucla.edu
DRC	Database of ribosomal crosslinks	http://www.mpimg-berlin-dahlem.mpg.de/~ag_ribo/ag_brimacombe/drc
GeneNet	Database on gene network components	http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet
IntAct project	Protein–protein interaction data	http://www.ebi.ac.uk/intact
InterDom	Putative protein domain interactions	http://interdom.lit.org.sg
JenPep	Functional and quantitative thermodynamic data on peptide binding to immunological biomacromolecules	http://www.jenner.ac.uk/Jenpep2
MPID	MHC—peptide interaction database	http://surya.bic.nus.edu.sg/mpid
ROSPath	Reactive oxygen species (ROS) signaling pathway	http://rospath.ewha.ac.kr
STCDB	Signal transductions classification database	http://www.techfak.uni-bielefeld.de/~mchen/STCDB

Table 1. *Continued*

Database name	Full name and/or description	URL
STRING	Predicted functional associations between proteins	http://www.bork.embl-heidelberg.de/STRING
TRANSPATH	Gene regulatory networks and microarray analysis	http://www.biobase.de/pages/products/databases.html

7. Human and other Vertebrate Genomes

7.1. Mitochondrial Genes and Proteins

AMmtDB	Metazoan mitochondrial genes	http://bighost.area.ba.cnr.it/mitochondriome
GOBASE	Organelle genome database	http://megasun.bch.umontreal.ca/gobase/gobase.html
MitoDat	Mitochondrial proteins (predominantly human)	http://www-lecb.ncifcrf.gov/mitoDat/
MitoMap	Human mitochondrial genome	http://www.mitomap.org/
MitoNuc	Nuclear genes coding for mitochondrial proteins	http://bio-www.ba.cnr.it:8000/BioWWW/#MitoNuc
MITOP2	Mitochondrial proteins, genes and diseases	http://ihg.gsf.de/mitop2/
MitoProteome	Mitochondrial protein sequences encoded by mitochondrial and nuclear genes	http://www.mitoproteome.org
OGRe	Complete mitochondrial genome sequences for 200 metazoan species	http://www.bioinf.man.ac.uk/ogre

7.2. Model organisms, comparative genomics

ACeDB	*C.elegans*, *S.pombe*, and human sequences and genomic information	http://www.acedb.org/
AllGenes	Human and mouse gene, transcript and protein annotation	http://www.allgenes.org/
ArkDB	Genome databases for farm and other animals	http://www.thearkdb.org/
Cre Transgenic Database	Cre transgenic mouse lines with links to publications	http://www.mshri.on.ca/nagy/
DRESH	Human cDNA clones homologous to *Drosophila* mutant genes	http://www.tigem.it/LOCAL/drosophila/dros.html
Ensembl	Annotated information on eukaryotic genomes	http://www.ensembl.org/
FANTOM	Functional annotation of mouse full-length cDNA clones	http://fantom2.gsc.riken.go.jp
FREP	Functional repeats in mouse cDNAs	http://facts.gsc.riken.go.jp/FREP/
GenetPig	Genes controlling economic traits in pig	http://www.infobiogen.fr/services/Genetpig
IPD-MHC Database	Non-human major histocompatibility complex sequences	http://www.ebi.ac.uk/ipd/mhc
KOG	Eukaryotic orthologous groups of proteins	http://www.ncbi.nlm.nih.gov/COG/new/shokog.cgi
LocusLink	Curated sequences and descriptions of genetic loci	http://www.ncbi.nlm.nih.gov/LocusLink
Mouse Genome Database	Mouse genome database	http://www.informatics.jax.org/
Mouse SAGE	SAGE libraries from various mouse tissues and cell lines	http://mouse.biomed.cas.cz/sage
Mouse Targeted Mutations	Information on transgenic animals and targeted mutations	http://tbase.jax.org/
MTID	Mouse transposon insertion database	http://mouse.ccgb.umn.edu/transposon/
PEDE	Pig EST data explorer: full-length cDNA libraries and ESTs	http://pede.gene.staff.or.jp/
Rat Genome Database	Rat genetic and genomic data	http://rgd.mcw.edu/
TIGR Gene Indices	Organism-specific databases of EST and gene sequences	http://www.tigr.org/tdb/tgi.shtml
UniGene	Unified clusters of ESTs and full-length mRNA sequences	http://www.ncbi.nlm.nih.gov/UniGene/
UniSTS	Unified non-redundant view of sequence tagged sites with marker and mapping data from a variety of resources	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists
ZFIN	Genetic, genomic and developmental data from zebrafish	http://zfin.org/

Table 1. *Continued*

Database name	Full name and/or description	URL
7.3. Human genome databases, maps and viewers		
Ensembl	Annotated information on eukaryotic genomes	http://www.ensembl.org/
AluGene	Complete *Alu* map in the human genome	http://alugene.tau.ac.il/
CroW 21	Human chromosome 21 database	http://bioinfo.weizmann.ac.il/crow21/
G3-RH	Stanford G3 and TNG radiation hybrid maps	http://www-shgc.stanford.edu/RH/
GB4-RH	Genebridge4 human radiation hybrid maps	http://www.sanger.ac.uk/Software/RHserver/RHserver.shtml
GDB	Human genes and genomic maps	http://www.gdb.org/
GenAtlas	Human genes, markers and phenotypes	http://www.citi2.fr/GENATLAS/
GeneCards	Integrated database of human genes, maps, proteins and diseases	http://bioinfo.weizmann.ac.il/cards/
GeneLoc	Gene location database (formerly UDB—Unified database for human genome mapping)	http://genecards.weizmann.ac.il/geneloc/
GeneNest	Gene indices of human, mouse, zebrafish, etc.	http://genenest.molgen.mpg.de/
GenMapDB	Mapped human BAC clones	http://genomics.med.upenn.edu/genmapdb
Gene Resource Locator	Alignment of ESTs with finished human sequence	http://grl.gi.k.u-tokyo.ac.jp/
HOWDY	Human organized whole genome database	http://www-alis.tokyo.jst.go.jp/HOWDY/
HuGeMap	Human genome genetic and physical map data	http://www.infobiogen.fr/services/Hugemap
Human BAC Ends Database	Non-redundant human BAC end sequences	http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_intro.html
IXDB	Physical maps of human chromosome X	http://ixdb.mpimg-berlin-dahlem.mpg.de/
NCBI RefSeq	Non-redundant DNA and protein sequence collection	http://www.ncbi.nlm.nih.gov/RefSeq/
ParaDB	Paralogy mapping in human genomes	http://abi.marseille.inserm.fr/paradb/
RHdb	Radiation hybrid map data	http://www.ebi.ac.uk/RHdb
STACK	Sequence tag alignment and consensus knowledgebase	http://www.sanbi.ac.za/Dbases.html
UCSC Genome Browser	Genome assemblies and annotation	http://genome.ucsc.edu/
7.4. Human proteins		
HPMR	Human plasma membrane receptome: protein sequences, literature, and expression database	http://receptome.stanford.edu/
HPRD	Human protein reference database: domain architecture, post-translational modifications, and disease association	http://www.hprd.org
HUNT	Human novel transcripts: annotated full-length cDNAs	http://www.hri.co.jp/HUNT
HUGE	Human unidentified gene-encoded large (>50 kDa) protein and cDNA sequences	http://www.kazusa.or.jp/huge
LIFEdb	Localization, interaction and functional assays of human proteins	http://www.dkfz.de/LIFEdb
trome, trEST and trGEN	Databases of predicted human protein sequences	ftp://ftp.isrec.isb-sib.ch/pub/databases/
8. Human Genes and Diseases		
8.1. General Databases		
Genetics Home Reference	A general guide on human hereditary diseases	http://ghr.nlm.nih.gov/
Homophila	*Drosophila* homologs of human disease genes	http://homophila.sdsc.edu/
IMGT	International immunogenetics information system: immunoglobulins, T cell receptors, MHC and RPI	http://imgt.cines.fr/
Mutation Spectra Database	Mutations in viral, bacterial, yeast and mammalian genes	http://info.med.yale.edu/mutbase/
OMIA	Online Mendelian inheritance in animals: a catalog of animal genetic and genomic disorders	http://www.angis.org.au/omia

Table 1. *Continued*

Database name	Full name and/or description	URL
OMIM	Online Mendelian inheritance in man: a catalog of human genetic and genomic disorders	http://www.ncbi.nlm.nih.gov/Omim/
ORFDB	Collection of ORFs that are sold by Invitrogen	http://orf.invitrogen.com/
PathBase	European mutant mice pathology database: histopathology photomicrographs and macroscopic images	http://www.pathbase.net/
PMD	Compilation of protein mutant data	http://pmd.ddbj.nig.ac.jp/
8.2. Human Mutations Databases		
8.2.1. General polymorphism databases		
ALFRED	Allele frequencies and DNA polymorphisms	http://alfred.med.yale.edu/
BayGenomics	Genes relevant to cardiovascular and pulmonary disease	http://baygenomics.ucsf.edu/
dbSNP	Database of single nucleotide polymorphisms	http://www.ncbi.nlm.nih.gov/SNP/
FIMM	Functional molecular immunology data	http://sdmc.krdl.org.sg:8080/fimm/
HGVS Databases	A compilation of human mutation databases	http://www.hgvs.org/
HGVbase	Human genome variation database: curated human polymorphisms	http://hgvbase.cgb.ki.se/
HGMD	Human gene mutation database	http://www.hgmd.org/
IPD	Immuno polymorphism database: data on human killer-cell Ig-like receptors and human platelet antigens	http://www.ebi.ac.uk/ipd
JSNP	Japanese SNP database	http://snp.ims.u-tokyo.ac.jp/
rSNP Guide	SNPs in regulatory gene regions	http://util.bionet.nsc.ru/databases/rsnp.html
SNP Consortium database	SNP Consortium data	http://snp.cshl.org/
TopoSNP	Topographic database of non-synonymous SNPs	http://gila.bioengr.uic.edu/snp/toposnp
8.2.2. Cancer		
Atlas of Genetics and Cytogenetics in Oncology and Haematology	Cancer related genes, chromosomal abnormalities in oncology and haematology, and cancer-prone diseases	http://www.infobiogen.fr/services/chromcancer/
CGED	Cancer gene expression database	http://love2.aist-nara.ac.jp/CGED
Database of Germline p53 Mutations	Mutations in human tumor and cell line p53 gene	http://www.lf2.cuni.cz/win/projects/germline_mut_p53.htm
IARC TP53 Database	Human TP53 somatic and germline mutations	http://www.iarc.fr/p53/
MTB	Mouse tumor biology database: mouse tumor types, genes, classification, incidence, pathology	http://tumor.informatics.jax.org/
Oral Cancer Gene Database	Cellular and molecular data for genes involved in oral cancer	http://www.tumor-gene.org/Oral/oral.html
RB1 Gene Mutation Database	Mutations in the human retinoblastoma (RB1) gene	http://www.d-lohmann.de/Rb/
RTCGD	Mouse retroviral tagged cancer gene database	http://rtcgd.ncifcrf.gov/
SNP500Cancer	Re-sequenced SNPs from 102 reference samples	http://snp500cancer.nci.nih.gov
SV40 Large T-Antigen Mutant Database	Mutations in SV40 large tumor antigen gene	http://bigdaddy.bio.pitt.edu/SV40/
Tumor Gene Family Databases	Cellular, molecular and biological data about genes involved in various cancers	http://www.tumor-gene.org/tgdf.html
8.2.3. Gene-, system- or disease-specific		
ALPSbase	Autoimmune lymphoproliferative syndrome database	http://research.nhgri.nih.gov/alps/
Androgen Receptor Gene Mutations Database	Mutations in the androgen receptor gene	http://www.mcgill.ca/androgendb/
BTKbase	Mutation registry for X-linked agammaglobulinemia	http://bioinf.uta.fi/BTKbase/
CASRDB	Calcium-sensing receptor database: CASR mutations causing hypercalcemia and/or hyperparathyroidism	http://www.casrdb.mcgill.ca/

Table 1. *Continued*

Database name	Full name and/or description	URL
Cytokine Gene Polymorphism in Human Disease	Cytokine gene polymorphism literature database	http://bris.ac.uk/pathandmicro/services/GAI/cytokine4.htm
Collagen Mutation Database	Human type I and type III collagen gene mutations	http://www.le.ac.uk/genetics/collagen/
ERGDB	Estrogen responsive genes database	http://sdmc.lit.org.sg/ergdb/cgi-bin/explore.pl
FUNPEP	Low-complexity peptides capable of forming amyloid plaque	http://www.cmbi.kun.nl/swift/FUNPEP/gergo/
GOLD.db	Genomics of lipid-associated disorders database	http://gold.tugraz.at
tGRAP	Mutants of G-protein coupled receptors of family A	http://tinygrap.uit.no/GRAP/
HaemB	Factor IX gene mutations, insertions and deletions	http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html
HbVar	Human hemoglobin variants and thalassemias	http://globin.cse.psu.edu/globin/hbvar
Human p53/hprt, rodent lacI/lacZ databases	Mutations at the human p53 and hprt genes; rodent transgenic lacI and lacZ mutations	http://www.ibiblio.org/dnam/mainpage.html
Human PAX2 Allelic Variant Database	Mutations in human PAX2 gene	http://pax2.hgu.mrc.ac.uk/
Human PAX6 Allelic Variant Database	Mutations in human PAX6 gene	http://pax6.hgu.mrc.ac.uk/
IL2Rgbase	X-linked severe combined immunodeficiency mutations	http://research.nhgri.nih.gov/scid/
IMGT/Gene-DB	Vertebrate immunoglobulin and T cell receptor genes	http://imgt.cines.fr/cgi-bin/GENElect.jv
IMGT/HLA	Polymorphism of human MHC and related genes	http://www.ebi.ac.uk/imgt/hla/
INFEVERS	Hereditary inflammatory disorder and familial mediterranean fever mutation data	http://fmf.igh.cnrs.fr/infevers
KinMutBase	Disease-causing protein kinase mutations	http://www.uta.fi/imt/bioinfo/KinMutBase/
Lowe Syndrome Mutation Database	Phosphatidylinositol-4,5-bisphosphate 5-phosphatase mutations causing Lowe oculocerebrorenal syndrome	http://research.nhgri.nih.gov/lowe/
NCL Mutation Database	Polymorphisms in neuronal ceroid lipofuscinoses genes	http://www.ucl.ac.uk/ncl/
PAHdb	Mutations at the phenylalanine hydroxylase locus	http://www.pahdb.mcgill.ca/
PGDB	Prostate and prostatic diseases gene database	http://www.ucsf.edu/PGDB
PHEXdb	PHEX mutations causing X-linked hypophosphatemia	http://www.phexdb.mcgill.ca/
PTCH1 Mutation Database	Mutations and SNPs found in PTCH1 gene	http://www.cybergene.se/PTCH/ptchbase.html

9. Microarray Data and other Gene Expression Databases

Database name	Full name and/or description	URL
ArrayExpress	Public collection of microarray gene expression data	http://www.ebi.ac.uk/arrayexpress
Axeldb	Gene expression in *Xenopus laevis*	http://www.dkfz-heidelberg.de/abt0135/axeldb.htm
BodyMap	Human and mouse gene expression data	http://bodymap.ims.u-tokyo.ac.jp/
BGED	Brain gene expression database	http://love2.aist-nara.ac.jp/BGED
CleanEx	Expression reference database, linking heterogeneous expression data to facilitate cross-dataset comparisons	http://www.cleanex.isb-sib.ch/
EICO DB	Expression-based imprint candidate organiser: a database for discovery of novel imprinted genes	http://fantom2.gsc.riken.jp/EICODB/
emap Atlas	Edinburgh mouse atlas: a digital atlas of mouse embryo development and spatially-mapped gene expression	http://genex.hgu.mrc.ac.uk/
EPConDB	Endocrine pancreas consortium database	http://www.cbil.upenn.edu/EPConDB
EpoDB	Genes expressed during human erythropoiesis	http://www.cbil.upenn.edu/EpoDB/
FlyView	*Drosophila* development and genetics	http://pbio07.uni-muenster.de/
GeneAnnot	Revised and improved annotation of Affymetrix human gene probe sets	http://genecards.weizmann.ac.il/geneannot/
GeneNote	Human genes expression profiles in healthy tissues	http://genecards.weizmann.ac.il/genenote/

Table 1. *Continued*

Database name	Full name and/or description	URL
GenePaint	Gene expression patterns in the mouse	http://www.genepaint.org/Frameset.html
GeneTrap	Expression patterns in an embryonic stem library of gene trap insertions	http://www.cmhd.ca/sub/genetrap.asp
GermOnline	Expression data relevant for the mitotic and meiotic cell cycle and gametogenesis in yeast and higher eukaryotes	http://www.germonline.org/
GXD	Mouse gene expression database	http://www.informatics.jax.org/menus/expression_menu.shtml
HemBase	Genes transcribed in differentiating human erythroid cells	http://hembase.niddk.nih.gov/
HugeIndex	Expression levels of human genes in normal tissues	http://hugeindex.org/
Interferon Stimulated Gene Database	Genes induced by treatment with interferons	http://www.lerner.ccf.org/labs/williams/xchip-html.cgi
Kidney Development Database	Kidney development and gene expression	http://golgi.ana.ed.ac.uk/kidhome.html
MAGEST	Ascidian (*Halocynthia roretzi*) gene expression patterns	http://www.genome.ad.jp/magest
MEPD	Medaka (freshwater fish *Oryzias latipes*) gene expression pattern database	http://medaka.dsp.jst.go.jp/MEPD
MethDB	DNA methylation data, patterns and profiles	http://www.methdb.de/
NASCarrays	Nottingham *Arabidopsis* Stock Centre microarray database	http://affymetrix.arabidopsis.info
NetAffx	Public Affymetrix probesets and annotations	http://www.affymetrix.com/
PEDB	Prostate expression database: ESTs from prostate tissue and cell type-specific cDNA libraries	http://www.pedb.org/
PEPR	Public expression profiling resource: expression profiles in a variety of diseases and conditions	http://microarray.cnmcresearch.org/pgadatatable.asp
RECODE	Genes using programmed translational recoding in their expression	http://recode.genetics.utah.edu/
RefExA	Reference database for human gene expression analysis	http://www.lsbm.org/db/index_e.html
Stanford Microarray Database	Raw and normalized data from microarray experiments	http://genome-www.stanford.edu/microarray
Tooth Development Database	Gene expression in dental tissue	http://bite-it.helsinki.fi/

10. Proteomics Resources

Database name	Full name and/or description	URL
GelBank	2D gel electrophoresis patterns of proteins from complete microbial genomes	http://gelbank.anl.gov/
PEP	Predictions for entire proteomes: summarized analyses of protein sequences	http://cubic.bioc.columbia.edu/pep/
Proteome Analysis Database	Functional classification of proteins in whole genomes	http://www.ebi.ac.uk/proteome/
RESID	Pre-, co- and post-translational protein modifications	http://www-nbrf.georgetown.edu/pirwww/dbinfo/resid.html
SWISS-2DPAGE	Annotated 2D gel electrophoresis database	http://www.expasy.org/ch2d/

11. Other Molecular Biology Databases

11.1. Drugs and drug design

Database name	Full name and/or description	URL
ANTIMIC	Database of natural antimicrobial peptides	http://research.i2r.a-star.edu.sg/Templar/DB/ANTIMIC/
APD	Antimicrobial peptide database	http://aps.unmc.edu/AP/main.php
BSD	Biodegradative strain database: microorganisms that can degrade aromatic and other organic compounds	http://bsd.cme.msu.edu/
DART	Drug adverse reaction target database	http://xin.cz3.nus.edu.sg/group/drt/dart.asp
Peptaibol	Peptaibol (antibiotic peptide) sequences	http://www.cryst.bbk.ac.uk/peptaibol/welcome.html
Pharmacogenomics and Pharmacogenetics Knowledge Base	Variation in drug response based on human variation	http://www.pharmgkb.org/
TTD	Therapeutic target database	http://xin.cz3.nus.edu.sg/group/cjttd/ttd.asp

Table 1. *Continued*

Database name	Full name and/or description	URL
11.2. Probes		
IMGT/PRIMER-DB	Immunogenetics oligonucleotide primer database	http://imgt3d.igh.cnrs.fr/PrimerDB/Query_PrDB.pl
MPDB	Information on synthetic oligonucleotides proven useful as primers or probes	http://www.biotech.ist.unige.it/interlab/mpdb.html
probeBase	rRNA-targeted oligonucleotide probe sequences, DNA microarray layouts and associated information	http://www.microbial-ecology.net/probebase
RTPrimerDB	Real-time PCR primer and probe sequences	http://medgen31.ugent.be/primerdatabase/index.php
VirOligo	Virus-specific oligonucleotides for PCR and hybridization	http://viroligo.okstate.edu/
11.3. Unclassified databases		
PubMed	Citations and abstracts of biomedical literature	http://pubmed.gov/
BioImage	Database of multidimensional biological images	http://www.bioimage.org/

[a]Category assignments of many databases are inherently subjective (e.g. MITOP could easily fit into 'yeast', 'mitochondria', 'protein targeting' and even 'comparative genomics'). Database coordinators are therefore encouraged to contact the author with suggestions regarding the category structure and requests to re-assign their databases to a different category.