

The Molecular Biology Database Collection: 2002 update

Andreas D. Baxevanis*

Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Building 50, Room 5222, Bethesda, MD 20892-8002, USA

Received October 9, 2001; Accepted November 20, 2001

ABSTRACT

The Molecular Biology Database Collection is an online resource listing key databases of value to the biological community. This Collection is intended to bring fellow scientists' attention to high-quality databases that are available throughout the world, rather than just be a lengthy listing of all available databases. As such, this up-to-date listing is intended to serve as the initial point from which to find specialized databases that may be of use in biological research. The databases included in this Collection provide new value to the underlying data by virtue of curation, new data connections or other innovative approaches. Short, searchable summaries and updates for each of the databases included in the Collection are available through the *Nucleic Acids Research* Web site at <http://nar.oupjournals.org>.

One of the most significant scientific events in the year 2001 was the publication of the initial sequence and analysis of the human genome resulting from both public (1) and private sector (2) efforts. With these publications, we have entered into a new era for modern biology, one where the majority of biological and biomedical research being conducted will use sequence data as its basic underpinning. Having such a rich source of information will prove invaluable for basic researchers whose findings will, in time, lead to improved strategies for the diagnosis, treatment and prevention of diseases having a genetic basis. In short, the stage has been set for genetic medicine having a prominent role in the delivery of healthcare in the future (3).

A number of significant insights have already been made into the secrets hidden within the 3 billion bases that comprise the human genome (1). There is marked variation in the distribution of features such as genes, transposable elements, GC content, CpG islands and recombination rate; this uneven distribution may provide important clues about the functions of these features and how they may be involved in regulation. There is a preferential retention of *Alu* elements in GC-rich regions, correlating them (in a loose sense) with actively-transcribed genes. These elements may actually turn out to not be just 'junk DNA', instead providing a tangible benefit to their human hosts. In general, repetitive elements may not have a direct function *per se*, but may influence chromosome

structure. Probably the most telling finding is that the total number of genes in the human genome is only in the order of 30 000 to 35 000. Previously, numbers in the 80 000 range (and as high as 140 000) had been put forward. While the new estimate in the number of genes gives the human about twice that seen in *Caenorhabditis elegans* or in *Drosophila*, the genes themselves have a more complex structure. This big down-estimate in the number of genes immediately brings into question the one gene-one protein hypothesis: we are now finding more and more examples of alternative splicing generating a larger number of protein products (consistent with a more complex gene structure), as well as cases where identical proteins can be used for different functions, depending on their compartmentalization (4).

While the near-completion of human genome sequencing marks a significant milestone, there are many other sequence-based efforts currently underway that will have just as much impact on the scientific and medical community. The most eagerly-anticipated model organism map is that of the mouse. The most recent physical map released on the Ensembl web site (<http://mouse.ensembl.org>, September 2001) provides an estimated 95% coverage of the mouse genome, with 15 694 genes confirmed over 361 Mb. To the issue of human health, single nucleotide polymorphisms (SNPs) continue to be identified at a breakneck pace. Over 1 million SNPs have already been identified, and a random sampling chosen for validation shows that 95% of these are indeed both polymorphic and unique (<http://snp.cshl.org/data/>). SNP alleles can be used as genetic markers, and often, the SNP itself is the variant that causes or contributes to the risk of developing a particular genetic disorder. To increase the power of using SNPs as markers for human disease, efforts are currently under way to develop a haplotype map, where 'blocks' of SNPs (rather than individual SNPs) could be used to find chromosomal regions associated with disease.

The sequence data that has been generated by these and other systematic sequencing projects can be browsed and downloaded from a variety of Web sites, with the major portals being located at NCBI (<http://www.ncbi.nlm.nih.gov>), Ensembl (<http://www.ensembl.org>) and UCSC (<http://genome.cse.ucsc.edu>). The problem that many investigators encounter, however, is that these larger databases often do not contain specialized information that would be of interest to specific groups within the scientific community. Many such databases have emerged to fill the void, and these databases often provide not just sequence-based information, but data such as phenotypes,

*Tel: +1 301 496 8570; Fax: +1 301 402 6858; Email: andy@nhgri.nih.gov

experimental conditions, strain crosses and map features, data that might not fit neatly onto a large physical map of a genome. Most importantly, data in these smaller databases tend to be curated by experts in a particular speciality and are often experimentally-verified, meaning that they represent the best state of knowledge in that particular area. The savvy user will, therefore, make use of both types of databases in their experimental planning and design. This journal has devoted its first issue over the last several years to documenting the availability and features of these specialized databases in order to better serve its readership and to promote the use of these resources in the design and analysis of experiments. These reviewed databases are collectively listed in the Molecular Biology Database Collection.

The databases included in the current version of the Collection are shown in Table 1. This year, the total number of databases listed is 335, up from 281 the year before. Several new databases have been added to the Collection, while others that are no longer actively curated or no longer available have been removed. These databases all distinguish themselves by their approach to presenting the underlying data—for example, by adding new value to the underlying data by virtue of curation, by providing new types of data connections or by implementing other innovative approaches that facilitate biological discovery. The individual entries are classified by type, but the reader should recognize that the distinctions between these classes are often arbitrary, and that many of these databases provide more than one type of information to the user.

In addition to the list presented in this paper, an electronic version of the Database Issue and Collection can be accessed online and is freely available to everyone, regardless of subscription status, at <http://nar.oupjournals.org>. While the list contains the databases described in the papers comprising the

current issue, it should be immediately apparent to the reader that there are simply not enough pages in this journal to accommodate full-length, printed descriptions of all of the 335 databases featured here. To address this, the online version of the Collection now includes short summaries of many of the databases, the summaries having been provided directly by the investigators responsible for the individual databases. We have also asked contributors to point out new features of their databases in the *Recent Developments* section of their entry. It is hoped that this approach will provide the reader with an additional source of information that will facilitate finding and selecting the sources of data that would be of most value in addressing a specific biological problem. Contributors will be encouraged to keep their entries up-to-date.

Suggestions for the inclusion of additional database resources in this collection are encouraged and may be directed to the author (andy@nhgri.nih.gov).

ACKNOWLEDGEMENT

I wish to thank Yi-Chi Barash for designing the new Web-based submission tool for this Collection, as well as for her technical support.

REFERENCES

1. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
2. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
3. Collins, F.S. and McKusick, V.A. (2001) Implications of the Human Genome Project for medical science. *J. Am. Med. Assoc.*, **285**, 540–544.
4. Jeffery, C.J. (1999) Moonlighting proteins. *Trends Biochem Sci.*, **24**, 8–11.

Table 1. Molecular Biology Database Collection

Major Public Sequence Repositories		
DNA Data Bank of Japan (DDBJ)	http://www.ddbj.nig.ac.jp	All known nucleotide and protein sequences; International Nucleotide Sequence Database Collaboration
EMBL Nucleotide Sequence Database	http://www.ebi.ac.uk/embl.html	All known nucleotide and protein sequences; International Nucleotide Sequence Database Collaboration
GenBank	http://www.ncbi.nlm.nih.gov/	All known nucleotide and protein sequences; International Nucleotide Sequence Database Collaboration
Ensembl	http://www.ensembl.org	Annotated human genome sequence data
STACK	http://www.sanbi.ac.za/Dbases.html	Non-redundant, gene-oriented clusters
TIGR Gene Indices	http://www.tigr.org/tdb/tgi.shtml	Non-redundant, gene-oriented clusters
UniGene	http://www.ncbi.nlm.nih.gov/UniGene/	Non-redundant, gene-oriented clusters
Comparative Genomics		
Clusters of Orthologous Groups (COG)	http://www.ncbi.nlm.nih.gov/COG	Phylogenetic classification of proteins from 44 complete genomes
Comparative Genometrics	http://www.unil.ch/igbm/genomics/genometrics.html	Biometric comparisons of whole genomes
euGenes	http://iubio.bio.indiana.edu:89/	Common summary of gene and genomic information from eukaryotic databases
Genome Information Broker	http://gib.genes.nig.ac.jp	Comparative analysis of completed microbial genomes
Gramene	http://www.gramene.org	Comparative genome analysis in the grasses
Homophila	http://homophila.sdsc.edu	Relationship of human disease genes to genes in <i>Drosophila</i>
XREFdb	http://www.ncbi.nlm.nih.gov/XREFdb/	Cross-referencing of model organism genetics with mammalian phenotypes
Gene Expression		
ASDB	http://cbcg.lbl.gov/asdb	Protein products and expression patterns of alternatively-spliced genes
Axeldb	http://www.dkfz-heidelberg.de/abt0135/axeldb.htm	Gene expression in <i>Xenopus</i>
BodyMap	http://bodymap.ims.u-tokyo.ac.jp/	Human and mouse gene expression data
EPConDB	http://www.cbil.upenn.edu/EPConDB	Endocrine pancreas consortium database
FlyView	http://pbio07.uni-muenster.de/	<i>Drosophila</i> development and genetics
Gene Expression Database (GXD)	http://www.informatics.jax.org/menu/expression_menu.shtml	Mouse gene expression and genomics
Gene Expression Omnibus (GEO)	http://www.ncbi.nlm.nih.gov/geo	Gene expression and hybridization array data repository
HugeIndex	http://www.hugeindex.org	mRNA expression levels of human genes in normal tissues
Interferon Stimulated Gene Database	http://www.lerner.ccf.org/labs/williams/xchip-html.cgi	Genes induced by treatment with interferons
Kidney Development Database	http://golgi.ana.ed.ac.uk/kidhome.html	Kidney development and gene expression
MAGEST	http://www.genome.ad.jp/magest	Ascidian (<i>Halocynthia roretzi</i>) gene expression patterns
MethDB	http://www.methdb.de	DNA methylation data, patterns, and profiles
Mouse Atlas and Gene Expression Database	http://genex.hgu.mrc.ac.uk	Spatially-mapped gene expression data
READ	http://read.gsc.riken.go.jp/READ/	RIKEN expression array database
RECODE	http://recode.genetics.utah.edu	Genes using programmed translational recoding in their expression
Stanford Microarray Database	http://genome-www.stanford.edu/microarray	Raw and normalized data from microarray experiments
Tooth Development Database	http://bite-it.helsinki.fi/	Gene expression in dental tissue

Table 1. Continued

TRIPLES	http://ygac.med.yale.edu	Transposon-insertion phenotypes, localization and expression in <i>Saccharomyces</i>
yMGV	http://www.transcriptome.ens.fr/ymgv/	Yeast microarray data and mining tools
Gene Identification and Structure		
AllGenes	http://www.allgenes.org	Human and mouse gene index integrating gene, transcript and protein annotation
Ares Lab Intron Site	http://www.cse.ucsc.edu/research/compbio/yeast_introns.html	Yeast spliceosomal introns
AsMamDB	http://166.111.30.65/ASMAMDB.html	Alternatively-spliced mammalian genes
COMPEL	http://compel.bionet.nsc.ru/	Composite regulatory elements
CUTG	http://www.kazusa.or.jp/codon/	Codon usage tables
DBTBS	http://elmo.ims.u-tokyo.ac.jp/dbtbs/	<i>Bacillus subtilis</i> binding factors and promoters
DBTSS	http://elmo.ims.u-tokyo.ac.jp/dbtss/	Transcriptional start sites
EID	http://mcb.harvard.edu/gilbert/EID/	Protein-coding, intron-containing genes
EPD	http://www.epd.isb-sib.ch/	Eukaryotic POL II promoters with experimentally-determined transcription start sites
ExInt	http://intron.bic.nus.edu.sg/exint/exint.html	Exon-intron structure of eukaryotic genes
FUGOID	http://wnt.cc.utexas.edu/~ifmr530/introndata/main.htm	Functional and structural information on organellar introns
Gene Resource Locator	http://grl.gi.k.u-tokyo.ac.jp	Alignment of ESTs with finished human sequence
HS3D	http://www.sci.unisannio.it/docenti/rampone/	Human exon, intron and splice regions
HUNT	http://www.hri.co.jp/HUNT	Annotated human full-length cDNA sequences
HvrBase	http://www.hvrbase.org	Primate mtDNA control region sequences
IDB/IEDB	http://nutmeg.bio.indiana.edu/intron/index.html	Intron sequence and evolution
PALSdb	http://palsdb.ym.edu.tw	Putative alternative splice sites
PLACE	http://www.dna.affrc.go.jp/htdocs/PLACE	Plant <i>cis</i> -acting regulatory elements
PlantCARE	http://sphinx.rug.ac.be:8080/PlantCARE/	Plant <i>cis</i> -acting regulatory elements
PromEC	http://bioinfo.md.huji.ac.il/marg/promec	<i>Escherichia coli</i> mRNA promoters with experimentally-identified transcriptional start sites
RRNDB	http://rrndb.cme.msu.edu	Variation in prokaryotic ribosomal RNA operons
RSDB	http://rsdb.csie.ncu.edu.tw	Repetitive elements from completed genomes
rSNP Guide	http://www.mgs.bionet.nsc.ru/mgs/systems/rsnp/	Single nucleotide polymorphisms in regulatory gene regions
SpliceDB	http://genomic.sanger.ac.uk/spldb/SpliceDB.html	Canonical and non-canonical mammalian splice sites
STRBase	http://www.cstl.nist.gov/div831/strbase/	Short tandem DNA repeats
TransCOMPEL	http://compel.bionet.nsc.ru/FunSite/CompelPatternSearch.html	Transcriptional regulatory elements in eukaryotic genes
Transterm	http://uther.otago.ac.nz/Transterm.html	Codon usage, start and stop signals
TRRD	http://www.mgs.bionet.nsc.ru/mgs/dbases/trrd4/	Transcription regulatory regions of eukaryotic genes
VIDA	http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html	Virus genome open reading frames
WormBase	http://www.wormbase.org	Guide to <i>C.elegans</i> biology
YIDB	http://www.embl-heidelberg.de/ExternalInfo/seraphin/yidb.html	Yeast nuclear and mitochondrial intron sequences
Genetic and Physical Maps		
DRESH	http://www.tigem.it/LOCAL/drosophila/dros.html	Human cDNA clones homologous to <i>Drosophila</i> mutant genes
G3-RH	http://www-shgc.stanford.edu/RH/	Stanford G3 and TNG radiation hybrid maps
GB4-RH	http://www.sanger.ac.uk/Software/RHserver/RHserver.shtml	Genebridge4(GB4) human radiation hybrid maps
GenAtlas	http://www.citi2.fr/GENATLAS/	Human genes, markers and phenotypes

Table 1. Continued

GeneMap '99	http://www.ncbi.nlm.nih.gov/genemap/	International Radiation Mapping Consortium human gene map
GenMapDB	http://genomics.med.upenn.edu/genmapdb	Mapped human BAC clones
HuGeMap	http://www.infobiogen.fr/services/Hugemap	Human genome genetic and physical map data
IXDB	http://ixdb.mpimg-berlin-dahlem.mpg.de	Physical maps of human chromosome X
RHdb	http://www.ebi.ac.uk/RHdb	Radiation hybrid map data
Genomic Databases		
ACeDB	http://www.acedb.org/	<i>C.elegans</i> , <i>Schizosaccharomyces pombe</i> and human sequences and genomic information
AMmtDB	http://bighost.area.ba.cnr.it/mitochondriome/	Metazoan mitochondrial genes
<i>Arabidopsis</i> Information Resource (TAIR)	http://www.arabidopsis.org/	<i>Arabidopsis thaliana</i> genome
ArkDB	http://www.thearkdb.org/	Genome databases for farm and other animals
Celera Discovery System	http://www.celera.com/genomics/academic/	Integrated, web-based discovery platform
Comprehensive Microbial Resource	http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl	Completed microbial genomes
CropNet	http://ukcrop.net/	Genome mapping in crop plants
CyanoBase	http://www.kazusa.or.jp/cyano/	<i>Synechocystis</i> sp. genome
<i>Dictyostelium</i> Genome Sequencing Project	http://dictygenome.bcm.tmc.edu	<i>Dictyostelium</i> genome resources
EcoGene	http://bmb.med.miami.edu/EcoGene/EcoWeb/	<i>E.coli</i> K-12 sequences
EMGlib	http://pbil.univ-lyon1.fr/emglib/emglib.html	Completely-sequenced prokaryotic genomes
FANTOM2	http://fantom.gsc.riken.go.jp/fantom2/doc/	RIKEN Mouse Gene Encyclopedia Project (functional annotation of mouse cDNA clones)
FlyBase	http://www.fruitfly.org	<i>Drosophila</i> sequences and genomic information
Full-Malaria	http://fullmal.ims.u-tokyo.ac.jp	Full-length cDNA library from erythrocytic-stage <i>Plasmodium falciparum</i>
Genew: Human Gene Nomenclature Database	http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl	Approved symbols for all human genes
GOBASE	http://megasun.bch.umontreal.ca/gobase	Organelle genome database
GOLD	http://igweb.integratedgenomics.com/GOLD/	Information regarding complete and ongoing genome projects
HERV	http://herv.img.cas.cz/	Human endogenous retroviruses
HIV Sequence Database	http://hiv-web.lanl.gov/	HIV RNA sequences
HOWDY	http://gdb.tokyo.jst.go.jp/HOWDY	Integrated human genome information parsed from primary sources
Human BAC Ends Database	http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_intro.html	Non-redundant human BAC end sequences
ICB	http://www.mbio.co.jp/icb	Identification and classification of bacterial protein-coding regions
INE	http://rgp.dna.affrc.go.jp/giot/INE.html	Rice genome analysis and sequencing
MagnaportheDB	http://www.cals.ncsu.edu/fungal_genomics/mgdatabase/int.htm	Integrated physical and genetic maps for the rice blast fungus <i>Magnaporthe grisea</i>
MatDB	http://mips.gsf.de/proj/thal/db/	<i>Arabidopsis</i> Genome Initiative data
Medicago Genome Initiative (MGI)	https://xgi.ncgr.org/mgi	Model legume <i>Medicago</i> ESTs, gene expression and proteomic data
Mendel Database	http://www.mendel.ac.uk/	Database of plant EST and STS sequences annotated with gene family information
MitBASE	http://www3.ebi.ac.uk/Research/Mitbase/mitbase.pl	Mitochondrial genomes, intra-species variants and mutants
MitoDat	http://www-lecb.ncifcrf.gov/mitoDat/	Mitochondrial proteins (predominantly human)
MITOMAP	http://www.gen.emory.edu/mitomap.html	Human mitochondrial genome

Table 1. Continued

MitoNuc/MitoAln	http://bighost.area.ba.cnr.it/srs6bin/wgetz?-page+LibInfo+-lib+MITONUC	Nuclear genes coding for mitochondrial proteins
MITOP	http://www.mips.biochem.mpg.de/proj/medgen/mitop/	Mitochondrial proteins, genes and diseases
Mouse Genome Database (MGD)	http://www.informatics.jax.org	Mouse genetics, genomics, alleles and phenotypes
MIPS	http://www.mips.biochem.mpg.de/	Protein and genomic sequences
NRSub	http://pbil.univ-lyon1.fr/nrsub/nrsub.html	<i>B. subtilis</i> genome
Oryzabase	http://www.shigen.nig.ac.jp/rice/oryzabase/	Rice genetics and genomics
Phytophthora Genome Consortium Database	https://xgi.ncgr.org/pgc	ESTs from <i>Phytophthora infestans</i> and <i>Phytophthora sojae</i>
PlasmoDB	http://PlasmoDB.org	<i>Plasmodium</i> genome
Proteome BioKnowledge Library	http://www.proteome.com	Model organism, pathogen and mammalian proteomes
Rat Genome Database	http://rgd.mcw.edu	Rat genetic and genomic data
RiceGAAS	http://RiceGaas.dna.affrc.go.jp/	Rice genome sequence and predicted gene structure
RsGDB	http://www-mm.g.med.uth.tmc.edu/sphaeroides	<i>Rhodobacter sphaeroides</i> genome
Saccharomyces Genome Database (SGD)	http://genome-www.stanford.edu/Saccharomyces	<i>Saccharomyces cerevisiae</i> genome
SubtiList	http://genolist.pasteur.fr/SubtiList/	<i>B. subtilis</i> 168 genome
TIGR Microbial Database	http://www.tigr.org/tdb/mdb/mdbcomplete.html	Microbial genomes and chromosomes
Wanda	http://www.evolutionsbiologie.uni-konstanz.de/Wanda/	Duplicated fish genes
WILMA	http://www.came.sbg.ac.at/wilma/	<i>C. elegans</i> annotation
ZFIN	http://zfin.org/	Genetic, genomic and developmental data from zebrafish
ZmDB	http://zmdb.iastate.edu/	Maize genome database
Intermolecular Interactions		
BIND	http://bind.ca	Molecular interactions, complexes and pathways
Database of Interacting Proteins	http://dip.doe-mbi.ucla.edu	Experimentally-determined protein-protein interactions
Database of Ribosomal Crosslinks (DRC)	http://www.mpimg-berlin-dahlem.mpg.de/~ag_ribo/ag_brimacombe/drc/	Ribosomal crosslinking data
DPIInteract	http://arep.med.harvard.edu/dpinteract/	Binding sites for <i>E. coli</i> DNA-binding proteins
MHC-Peptide Interaction Database	http://surya.bic.nus.edu.sg/mpid	Class I and Class II MHC-peptide complexes
Metabolic Pathways and Cellular Regulation		
EcoCyc	http://ecocyc.org/	<i>E. coli</i> K-12 genome, metabolic pathways, transporters and gene regulation
ENZYME	http://www.expasy.ch/enzyme/	Enzyme nomenclature
EpoDB	http://www.cbil.upenn.edu/EpoDB/	Genes expressed during human erythropoiesis
GeneNet	http://wwwmgs.bionet.nsc.ru/mgs/systems/genenet/	Formalized descriptions of the structure and functional organization of gene networks
Klotho	http://www.ibr.wustl.edu/klotho/	Collection and categorization of biological compounds
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.ad.jp/kegg	Metabolic and regulatory pathways
LIGAND	http://www.genome.ad.jp/ligand/	Chemical compounds and reactions in biological pathways
MetaCyc	http://ecocyc.org/	Metabolic pathways and enzymes from various organisms
PathDB	http://www.ncgr.org/pathdb	Biochemical pathways, compounds and metabolism
RegulonDB	http://www.cifn.unam.mx/regulondb/	<i>E. coli</i> transcriptional regulation and operon organization
UM-BBD	http://umbbd.ahc.umn.edu/	Microbial biocatalytic reactions and biodegradation pathways
WIT2	http://wit.mcs.anl.gov/WIT2/	Integrated system for functional curation and development of metabolic models

Table 1. Continued

Mutation Databases		
ALFRED	http://alfred.med.yale.edu/alfred/	Allele frequencies and DNA polymorphisms
Androgen Receptor Gene Mutations Database	http://www.mcgill.ca/androgendb/	Mutations in the androgen receptor gene
Asthma Gene Database	http://cooke.gsf.de/asthmagen/main.cfm	Linkage and mutation studies on the genetics of asthma and allergy
Atlas of Genetics and Cytogenetics in Oncology and Haematology	http://www.infobiogen.fr/services/chromcancer/	Chromosomal abnormalities in cancer
BTKbase	http://www.uta.fi/laitokset/imt/bioinfo/BTKbase/	Mutation registry for X-linked agammaglobulinemia
CASRDB	http://data.mch.mcgill.ca/casrdb/	CASR mutations causing FHH, NSHPT and ADH
Cytokine Gene Polymorphism Database	http://www.bris.ac.uk/pathandmicroservices/GAI/cytokine4.htm	Cytokine gene polymorphisms, <i>in vitro</i> expression and disease-association studies
Database of Germline p53 Mutations	http://www.lf2.cuni.cz/win/projects/germline_mut_p53.htm	Mutations in human p53
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/	Single nucleotide polymorphisms
DT40	http://genetics.hpi.uni-hamburg.de/dt40.html	Knockout mutants in chicken DT40 B-cells
FLAGdb/FST	http://genoplante-info.infobiogen.fr	<i>Arabidopsis thaliana</i> T-DNA transformants
GRAP Mutant Databases	http://tinyGRAP.uit.no/GRAP/	Mutants of family A G-Protein Coupled Receptors (GRAP)
jSNP	http://snp.ims.u-tokyo.ac.jp	SNPs in the Japanese population
Haemophila B Mutation Database	http://www.umds.ac.uk/molgen/haemBdatabase.htm	Point mutations, short additions and deletions in the Factor IX gene
HGVbase	http://hgvbase.cgb.ki.se	Curated human polymorphisms
HIV-RT	http://hivdb.stanford.edu/hiv/	HIV reverse transcriptase and protease sequence variation
Human Gene Mutation Database (HGMD)	http://www.hgmd.org	Known (published) gene lesions underlying human inherited disease
Human p53, human hpert, rodent lacI and rodent lacZ databases	http://metalab.unc.edu/dnam/mainpage.html	Mutations in human p53 and hpert; rodent transgenic lacI and lacZ mutations
Human PAX2 Allelic Variant Database	http://www.hgu.mrc.ac.uk/Softdata/PAX2/	Mutations in human PAX2 gene
Human PAX6 Allelic Variant Database	http://www.hgu.mrc.ac.uk/Softdata/PAX6/	Mutations in human PAX6 gene
Human Type I/III Collagen Mutation Database	http://www.le.ac.uk/genetics/collagen/	Human type I and type III collagen gene mutations
iARC p53 Database	http://www.iarc.fr/p53/	Compilation of TP53 gene mutations
KinMutBase	http://www.uta.fi/imt/bioinfo/KinMutBase/	Disease-causing protein kinase mutations
KMDB	http://131.113.190.126/mutview3/mutview/index_eye.html	Mutations in human eye disease genes
Mutation Spectra Database	http://info.med.yale.edu/mutbase/	Mutations in viral, bacterial, yeast and mammalian genes
NCL Mutations	http://www.ucl.ac.uk/ncl/	Mutations and polymorphisms in neuronal ceroid lipofuscinoses (NCL) genes
Online Mendelian Inheritance in Man	http://www.ncbi.nlm.nih.gov/Omim/	Human genetic and genomic disorders
PAHdb	http://data.mch.mcgill.ca/pahdb_new/	Mutations at the phenylalanine hydroxylase locus
PHEXdb	http://data.mch.mcgill.ca/phexdb	Mutations in PHEX gene causing X-linked hypophosphatemia
PMD	http://pmd.ddbj.nig.ac.jp/	Compilation of protein mutant data
PTCH1 Mutation Database	http://www.cybergene.se/PTCH/ptchbase.html	Mutations and SNPs found in PTCH1
RB1 Gene Mutation Database	http://www.d-lohmann.de/Rb/	Mutations in the human retinoblastoma gene
SV40 Large T-Antigen Mutant Database	http://bigdaddy.bio.pitt.edu/SV40/	Mutations in SV40 large tumor antigen gene
Pathology		
AngioDB	http://angiodb.snu.ac.kr	Angiogenesis and angiogenesis-related molecules
FIMM	http://sdmc.krdl.org.sg:8080/fimm/	Functional molecular immunology data
HCFForum	http://hcforum.imag.fr/welcome_eng.html	Human cytogenetics database

Table 1. Continued

IDR	http://www.uta.fi/imt/bioinfo/idr/	Immunodeficiency mutations
Mouse Tumor Biology Database (MTB)	http://tumor.informatics.jax.org	Mouse tumor names, classification, incidence, pathology, genetic factors
Oral Cancer Gene Database	http://www.tumor-gene.org/Oral/oral.html	Cellular, molecular and biological data for genes involved in oral cancer
PEDB	http://www.pedb.org/	Sequences from prostate tissue and cell type-specific cDNA libraries
Tumor Gene Family Databases (TGDBs)	http://www.tumor-gene.org/tgdf.html	Cellular, molecular and biological data about genes involved in various cancers
Protein Databases		
AARSDB	http://rose.man.poznan.pl/aars/index.html	Aminoacyl-tRNA synthetase sequences
ABCdb	http://ir2lcb.cnrs-mrs.fr/ABCdb/	ABC transporters
AraC/XylS database	http://www.AraC-XylS.org	AraC/XylS family of positive regulators in bacteria
ASPD	http://www.mgs.bionet.nsc.ru/mgs/gnw/aspd	Artificial proteins and peptides
BRENDA	http://www.brenda.uni-koeln.de/	Extensive functional data on enzymes
CSDBase	http://www.chemie.uni-marburg.de/~csdbase	Cold shock domain-containing proteins
Data	http://luggagefast.Stanford.EDU/group/arabprotein/	Annotated coding sequences from <i>Arabidopsis</i>
DExH/D Family Database	http://www.helicase.net/dexhd/dbhome.htm	DEAD-box, DEAH-box and DExH-box proteins
Endogenous GPCR List	http://www.biomedcomp.com/GPCR.html	G protein-coupled receptors; expression in cell lines
ESTHER	http://www.ensam.inra.fr/cholinesterase/	Esterases and alpha/beta hydrolase enzymes and relatives
EXProt	http://www.cmbi.nl/exprot	Proteins with experimentally-verified function
FUNPEP	http://picsou.cmbi.kun.nl:8080/	Low-complexity or compositionally-biased protein sequences
GenProtEC	http://genprotec.mbl.edu	<i>E.coli</i> K-12 genome, gene products and homologs
GPCRDB	http://www.gpcr.org/7tm/	G protein-coupled receptors
Histone Database	http://genome.nhgri.nih.gov/histones	Histone and histone fold sequences and structures
HIV Molecular Immunology Database	http://hiv-web.lanl.gov/immunology/	HIV epitopes
Homeobox Page	http://www.biosci.ki.se/groups/tbu/homeo.html	Information relevant to homeobox proteins, classification and evolution
Homeodomain Resource	http://genome.nhgri.nih.gov/homeodomain	Homeodomain sequences, structures and related genetic and genomic information
HUGE	http://www.kazusa.or.jp/huge/	Large (>50 kDa) human proteins and cDNA sequences
IMGT	http://imgt.cines.fr	Immunoglobulin, T cell receptor and MHC sequences from human and other vertebrates
IMGT/HLA	http://www.ebi.ac.uk/imgt/hla/	Human MHC sequences
InBase	http://www.neb.com/neb/inteins.html	All known inteins (protein splicing elements): properties, sequences, bibliography
Kabat Database	http://immuno.bme.nwu.edu/	Sequences of proteins of immunological interest
LGICdb	http://www.pasteur.fr/recherche/banques/LGIC/LGIC.html	Ligand-gated ion channel subunit sequences
MEROPS	http://www.merops.ac.uk	Proteolytic enzymes (proteases/peptidases)
MetaFam	http://metafam.ahc.umn.edu/	Integrated protein family information
Metalloprotein Database and Browser	http://metallo.scripps.edu/	Metal-binding sites in metalloproteins
MHCBN	http://www.imtech.res.in/raghava/mhcbn/	MHC-binding and non-binding peptides
MHCPEP	http://wehih.wehi.edu.au/mhcpep/	MHC-binding peptides
Nuclear Receptor Resource	http://nrr.georgetown.edu/nrr/nrr.html	Nuclear receptor superfamily
NUREBASE	http://www.ens-lyon.fr/LBMC/laudet/nurebase.html	Nuclear hormone receptors
Olfactory Receptor Database	http://ycmi.med.yale.edu/senselab/ordb/	Sequences for olfactory receptor-like molecules

Table 1. Continued

ooTFD	http://www.ifti.org/	Transcription factors and gene expression
Peptaibol	http://www.cryst.bbk.ac.uk/peptaibol/welcome.html	Peptaibol (antibiotic peptide) sequences
PhosphoBase	http://www.cbs.dtu.dk/databases/PhosphoBase/	Protein phosphorylation sites
PKR	http://pkr.sdsc.edu	Protein kinase sequences, enzymology, genetics, molecular/structural properties
PLANT-PIs	http://bighost.area.ba.cnr.it/PLANT-PIs/	Plant protease inhibitors
PlantsP	http://PlantsP.sdsc.edu	Plant protein kinases and phosphatases
PPMdb	http://sphinx.rug.ac.be:8080/ppmdb/	<i>Arabidopsis</i> plasma membrane protein sequence and expression data
Prolysis	http://delphi.phys.univ-tours.fr/Prolysis/	Proteases and natural and synthetic protease inhibitors
PROMISE	http://bioinf.leeds.ac.uk/promise/	Prosthetic centers and metal ions in protein active sites
Protein Information Resource (PIR)	http://pir.georgetown.edu	Comprehensive, annotated, non-redundant protein sequence database
Ribonuclease P Database	http://www.mbio.ncsu.edu/RNaseP/home.html	RNase P sequences, alignments and structures
SENTRA	http://wit.mcs.anl.gov/WIT2/Sentra/HTML/sentra.html	Sensory signal transduction proteins
S/MARt db	http://transfac.gbf.de/SMARTDB/	Scaffold/matrix attached regions
SWISS-PROT/TrEMBL	http://www.expasy.ch/sprot	Curated protein sequences
TIGRFAMs	http://www.tigr.org/TIGRFAMs	Protein family resource for the functional identification of proteins
TRANSFAC	http://transfac.gbf.de/TRANSFAC/	Transcription factors and binding sites
trEST, trGEN, Hits	http://hits.isb-sib.ch	Hypothetical protein sequences; precompiled list of predicted domains/signatures
VIDA	http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html	Homologous viral protein families
Wnt Database	http://www.stanford.edu/~rnusse/wntwindow.html	Wnt proteins and phenotypes
Protein Sequence Motifs		
BLOCKS	http://blocks.fhrc.org	Multiple alignments of conserved regions of protein families
CDD	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml	Alignment models for conserved protein domains
CluSTR	http://www.ebi.ac.uk/clustr/	Automatic classification of SWISS-PROT+TrEMBL proteins
eMOTIF	http://motif.stanford.edu/emotif	Protein sequence motif determination and searches
InterPro	http://www.ebi.ac.uk/interpro/	Integrated documentation resource for protein families, domains and sites
iPROCLASS	http://pir.georgetown.edu/iproclass/	Annotated protein classification database with structure and function information
O-GLYCBASE	http://www.cbs.dtu.dk/databases/OGLYCBASE/	Glycoproteins and O-linked glycosylation sites
Pfam	http://www.sanger.ac.uk/Software/Pfam/	Multiple sequence alignments and hidden Markov models of common protein domains
PIR-ALN	http://pir.georgetown.edu/pirwww/dbinfo/piraln.html	Protein sequence alignments
PRINTS	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/	Hierarchical gene family fingerprints
ProClass	http://pir.georgetown.edu/gfserver/proclass.html	Protein families defined by PIR superfamilies and PROSITE patterns
ProDom	http://www.toulouse.inra.fr/prodom.html	Protein domain families
PROSITE	http://www.expasy.org/prosite	Biologically-significant protein patterns and profiles
ProtoMap	http://protomap.cornell.edu	Automated hierarchical classification of SWISS-PROT proteins
SBASE	http://www.icgeb.trieste.it/sbase	Annotated protein domain sequences
SMART	http://smart.embl-heidelberg.de	Simple Modular Architecture Research Tool

Table 1. Continued

SUPFAM	http://pauling.mbu.iisc.ernet.in/~supfam	Sequence families correlated to structure
SYSTEMS, GeneNest, SpliceNest	http://cmb.molgen.mpg.de	Integrated database of protein families, EST clusters and their genomic positions
Proteome Resources		
Aaindex	http://www.genome.ad.jp/dbget/	Physicochemical properties of peptides
GELBANK	http://gelbank.anl.gov	2D-gel electrophoresis patterns from completed genomes
Human Proteome Survey Database	http://www.proteome.com/services	Detailed information on human, mouse and rat proteomes
Predictome	http://predictome.bu.edu	Putative functional links between proteins
Proteome Analysis Database	http://www.ebi.ac.uk/proteome/	Online application of InterPro and cluSTr for the functional classification of proteins in whole genomes
REBASE	http://rebase.neb.com/rebase/rebase.html	Restriction enzymes and associated methylases
SWISS-2DPAGE	http://www.expasy.ch/ch2d/	Annotated two-dimensional polyacrylamide gel electrophoresis database
YPL	http://fstgal12.tu-graz.ac.at:7777/pls/al12/ypl.htm	Yeast protein localization as determined by GFP-tagging and confocal microscopy
Retrieval Systems and Database Structure		
KEYnet	http://www.ba.cnr.it/keynet.html	Hierarchical list of gene and protein names for data retrieval
TESS	http://www.cbil.upenn.edu/tess	Transcription element search system
Virgil	http://www.infobiogen.fr/services/virgil	Database interconnectivity
RNA Sequences		
16S and 23S rRNA Mutation Database	http://ribosome.fandm.edu	16S and 23S ribosomal RNA mutations
5S rRNA Database	http://biobases.ibch.poznan.pl/5SData/	5S rRNA sequences
ACTIVITY	http://www.mgs.bionet.nsc.ru/mgs/systems/activity/	Functional DNA/RNA site activity
ARED	http://rc.kfshrc.edu.sa/ared	AU-rich element-containing mRNAs
Collection of mRNA-like Noncoding RNAs	http://biobases.ibch.poznan.pl/ncRNA/	Non-protein-coding RNA transcripts
European Large Subunit rRNA Database	http://rrna.uia.ac.be/lisu/index.html	Alignment of large subunit ribosomal RNA sequences with secondary structure information
European Small Subunit rRNA Database	http://rrna.uia.ac.be/ssu/index.html	Alignment of small subunit ribosomal RNA sequences with secondary structure information
Guide RNA Database	http://biosun.bio.tu-darmstadt.de/goringer/gRNA/gRNA.html	Guide RNA sequences
HyPaLib	http://bibiserv.techfak.uni-bielefeld.de/HyPa/	Structural elements characteristic for classes of RNA
Intronerator	http://www.cse.ucsc.edu/~kent/intronerator/	RNA splicing and gene structure in <i>C.elegans</i> ; alignments of <i>C.briggsae</i> and <i>C.elegans</i> genomic sequences
Non-Canonical Interactions in RNA	http://prion.bchs.uh.edu/bp_type/	Non-standard base-base interactions in known RNA structures
PLANTncRNAs	http://www.prl.msu.edu/PLANTncRNAs/	Plant non-protein coding RNAs with relevant gene expression information
PLMItrRNA	http://bigarea.area.ba.cnr.it:8000/PLMItrRNA/	Mitochondrial tRNA genes and molecules in photosynthetic eukaryotes
PseudoBase	http://www.bio.leidenuniv.nl/~Batenburg/PKB.html	Structural, functional and sequence data related to RNA pseudoknots
Ribosomal Database Project (RDP-II)	http://rdp.cme.msu.edu	rRNA sequence data, alignments and phylogenies
RISCC	http://ulises.umh.es/RISCC	Ribosomal 16S–23S RNA gene spacer regions
RNA Modification Database	http://medlib.med.utah.edu/RNAmods/	Naturally-modified nucleosides in RNA
SELEX_DB	http://www.mgs.bionet.nsc.ru/mgs/systems/selex/	Selected DNA/RNA functional site sequences
Small RNA Database	http://mbcr.bcm.tmc.edu/smallRNA	Direct sequencing of small RNA sequences from prokaryotes and eukaryotes

Table 1. Continued

SRPDB	http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html	Signal recognition particle RNA, SRP protein, and SRP receptor sequences and alignments
tmRDB	http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html	tmRNA (10Sa RNA) sequences and alignments
tmRNA	http://www.indiana.edu/~tmrna	tmRNA sequences, foldings and alignments
tRNA Sequences	http://www.uni-bayreuth.de/departments/biochemie/trna/	tRNA and tRNA gene sequences
UTRdb/UTRsite	http://bighost.area.ba.cnr.it/srs6/	5'- and 3'-UTRs of eukaryotic mRNAs and relevant functional patterns
Viroids and viroid-like RNAs	http://nt.ars-grin.gov/subviral/	Viroids and viroid-like RNAs
Yeast snoRNA Database	http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html	Yeast small nucleolar RNAs
Structure		
ASTRAL	http://astral.stanford.edu/	Sequences of domains of known structure, selected subsets and sequence-structure correspondences
BioImage	http://www-embl.bioimage.org/	Searchable database of multidimensional biological images
BioMagResBank	http://www.bmrb.wisc.edu/	NMR spectroscopic data from proteins, peptides and nucleic acids
CATH	http://www.biochem.ucl.ac.uk/bsm/cath/	Hierarchical classification of protein domain structures
CE	http://cl.sdsc.edu/ce.html	Computation and review of 3D alignments
CKAAPs DB	http://ckaaps.sdsc.edu/ckaap/ckaap.home	Structurally-similar proteins with dissimilar sequences
CSD	http://www.ccdc.cam.ac.uk/prods/csd/csd.html	Crystal structure information for organic and metal organic compounds
Database of Macromolecular Movements	http://bioinfo.mbb.yale.edu/MolMovDB/	Descriptions of protein and macromolecular motions, including movies
Decoys 'R' Us	http://dd.stanford.edu/	Computer-generated protein conformations based on sequence data
DSDBASE	http://www.ncbs.res.in/~faculty/mini/dsdbase/dsdbase.html	Native and modeled disulfide bonds in proteins
GTOP	http://spock.genes.nig.ac.jp/~genome/gtop-j.html	Protein structures predicted from genome sequences
HIC-Up	http://alpha2.bmc.uu.se/hicup/	Structures of small molecules
HSSP	http://www.sander.ebi.ac.uk/hssp/	Structural families and alignments; structurally-conserved regions and domain architecture
IMB Jena Image Library of Biological Macromolecules	http://www.imb-jena.de/IMAGE.html	Visualization and analysis of three-dimensional biopolymer structures
ISSD	http://www.protein.bio.msu.su/issd/	Integrated sequence and structural information
LPFC	http://www-smi.stanford.edu/projects/helix/LPFC/	Library of protein family core structures
MMDB	http://www.ncbi.nlm.nih.gov/Structure/	All experimentally-determined three-dimensional structures, linked to NCBI Entrez
ModBase	http://guitar.rockefeller.edu/modbase	Annotated comparative protein structure models
NDB	http://ndbserver.rutgers.edu/NDB/ndb.html	Nucleic acid-containing structures
NTDB	http://ntdb.chem.cuhk.edu.hk	Thermodynamic data for nucleic acids
PALI	http://pauling.mbu.iisc.ernet.in/~pali	Phylogeny and alignment of homologous protein structures
PASS2	http://www.ncbs.res.in/~faculty/mini/campass/pass.html	Protein structural superfamilies
PDB	http://www.rcsb.org/pdb/	Structure data determined by X-ray crystallography and NMR
PDB-REPRDB	http://www.cbrc.jp/papia/	Representative protein chains, based on PDB entries
PDBsum	http://www.biochem.ucl.ac.uk/bsm/pdbsum	Summaries and analyses of PDB structures
PRESAGE	http://presage.berkeley.edu/	Protein structures with experimental and predictive annotations

Table 1. Continued

ProTherm	http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html	Thermodynamic data for wild-type and mutant proteins
RESID	http://www-nbrf.georgetown.edu/pirwww/dbinfo/resid.html	Protein structure modifications
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop	Familial and structural protein relationships
SCOR	http://scor.lbl.gov	RNA structural relationships
Sloop	http://www-cryst.bioc.cam.ac.uk/~sloop/	Classification of protein loops
SUPERFAMILY	http://stash.mrc-lmb.cam.ac.uk/SUPERFAMILY/	Assignments of proteins to structural superfamilies
Transgenics		
Cre Transgenic Database	http://www.mshri.on.ca/nagy/cre.htm	Cre transgenic mouse lines
Transgenic/Targeted Mutation Database	http://tbase.jax.org/	Information on transgenic animals and targeted mutations
Varied Biomedical Content		
BaliBASE	http://www-igbmc.u-strasbg.fr/BioInfo/BaliBASE2/index.html	Benchmark database for comparison of multiple sequence alignments
Dbcats	http://www.infobiogen.fr/services/dbcats/	Catalog of databases
DrugDB	http://www.chem.ac.ru/Chemistry/Databases/DRUGDBPH.en.html	Pharmacologically-active compounds; generic and trade names
Global Image Database	http://www.gwer.ch/qv/gid/gid.htm	Annotated biological images
GlycoSuiteDB	http://www.glycosuite.com	<i>N</i> - and <i>O</i> -linked glycan structures and biological source information
HOX-PRO	http://www.mssm.edu/molbio/hoxpro/new/hox-pro00.html	Clustering of homeobox genes
Imprinted Genes and Parent-of-Origin Effects	http://www.otago.ac.nz/IGC	Imprinted genes and parent-of-origin effects in animals
LocusLink/RefSeq	http://www.ncbi.nlm.nih.gov/LocusLink/	Curated reference sequence standards for genes, transcripts and proteins
MPDB	http://www.biotech.ist.unige.it/interlab/mpdb.html	Information on synthetic oligonucleotides proven useful as primers or probes
NCBI Taxonomy Browser	http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html	Names of all organisms that are represented in the genetic databases with at least one nucleotide or protein sequence
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/	MEDLINE and Pre-MEDLINE citations
PharmGKB	http://pharmgkb.org	Variation in drug response based on human variation
RIDOM	http://www.ridom.de/	rRNA (16S and ITS) sequence-based identification of medical microorganisms
SWEET-DB	http://www.dkfz-heidelberg.de/spec2/	Annotated carbohydrate structure and substance information
Therapeutic Target Database	http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp	Therapeutic protein and nucleic acid targets, metabolic pathway and drug information
Tree of Life	http://phylogeny.arizona.edu/tree/phylogeny.html	Information on phylogeny and biodiversity
Vectordb	http://www.atcg.com/vectordb/	Characterization and classification of nucleic acid vectors
VirOligo	http://virologo.okstate.edu	Virus-specific oligonucleotides for PCR and hybridization