 Nucleic Acids Research, 2000, Vol. 28, No. 1 1–7

The Molecular Biology Database Collection: an online compilation of relevant database resources

Andreas D. Baxevanis*

Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Building 49, Room 4A-22, Bethesda, MD 20892-4470, USA

Received October 26, 1999; Accepted October 27, 1999

ABSTRACT

The Molecular Biology Database Collection represents an effort geared at making molecular biology database resources more accessible to biologists. This online resource, available at http://www.oup.co.uk/nar/Volume_28/Issue_01/html/gkd115_gml.html, is intended to serve as a searchable, up-to-date, centralized jumping-off point to individual Web sites. An emphasis has also been placed on including databases where new value is added to the underlying data by virtue of curation, new data connections, or other innovative approaches.

With systematic sequencing efforts on the human and other model organisms in full swing, a tremendous flood of sequence data is being produced at breakneck speed. Considering the Human Genome Project alone, it is currently estimated that over 2 million bases are deposited into GenBank each day. This growth will only accelerate in the future, given the new goal of having a working draft covering 90% of the genome by the spring of 2000 and the complete sequence in-hand in 2002 (1). When considering all organismal sequencing efforts together, it immediately becomes obvious that there must be concrete and coherent plans in place to both warehouse and access these data, data which will become more and more important as we enter into an era of sequence-based biology.

The most recent five-year plan for the Human Genome Project (1) recognizes the importance of the connection between the *in silico* world and the bench, and has put forth recommendations regarding bioinformatics and computational biology, particularly with respect to databases and the development of analytical tools. Databases that grow out of the human and other systematic sequencing efforts should not be just storehouses for thousands of bases or amino acids; rather, they need to make logical connections to other types of information that are available, such as phenotypic or expression data, as these types of connections serve to increase the intrinsic value of the raw sequence data, allowing for biological discovery. Of particular importance will be the curation and annotation of specialized databases, which will help to ensure data quality as well as foster the ability to access and analyze sequence data.

Many such specialized databases have emerged over the years, and this journal has devoted its first issue over the last several years to documenting the availability and features of these databases. With the current issue, *Nucleic Acids Research* will be facilitating access to these databases online through a new entity called the Molecular Biology Database Collection. By doing so, it is hoped that individual investigators will be able to more easily find and use specialized databases that are appropriate to their scientific needs.

The initial set of databases included in the online compilation, available at http://www.oup.co.uk/nar/Volume_28/Issue_01/html/gkd115_gml.html, is shown in Table 1. The list includes databases for which there are full-length descriptions in the current issue, as well as other databases which represent a valuable resource to the biological community. Also included are databases collected by Christian Burks as part of the 1999 Database Issue (2). Efforts have been made to reduce redundancy in the list with respect to content wherever possible. An emphasis has also been placed on including databases where new value is added to the underlying data by virtue of curation, new data connections, or other innovative approaches. While Table 1 attempts to classify these databases by type, it is often true that these databases are more global in nature, providing multiple types of information. Links to these databases will be updated regularly. Suggestions for the inclusion of additional database resources in this online compendium are encouraged and may be directed to the author (andy@nhgri.nih.gov).

REFERENCES

1. Collins,F.S., Patrinos,A., Jordan,E., Chakravarti,A., Gesteland,R., Walters,L. and members of the DOE and NIH Planning Groups (1998) *Science*, **282**, 682–689.
2. Burks,C. (1999) *Nucleic Acids Res.*, **27**, 1–9.

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

Table 1. Molecular Biology Database Collection

Major Sequence Repositories		
GenBank	http://www.ncbi.nlm.nih.gov/Web/Genbank/	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
DNA Data Bank of Japan (DDBJ)	http://www.ddbj.nig.ac.jp	All known nucleotide and protein sequences; International Nucleotide Sequence Database Collaboration
Genome Sequence Database (GSDB)	http://www.ncgr.org/gsdb	All known nucleotide and protein sequences
TIGR Gene Indices	http://www.tigr.org/tdb/tdb.html	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 21 complete genomes
XREFdb	http://www.ncbi.nlm.nih.gov/XREFdb/	Cross-referencing of model organism genetics with mammalian phenotypes
Gene Expression		
ASDB	http://cbcg.nersc.gov/asdb	Protein products and expression patterns of alternatively-spliced genes
Axeldb	http://www.dkfz-heidelberg.de/abt0135/axeldb.htm	Gene expression in *Xenopus*
BodyMap	http://bodymap.ims.u-tokyo.ac.jp	Human and mouse gene expression data
EpoDB	http://www.cbil.upenn.edu/epodb	Genes expressed in vertebrate RBC
FlyView	http://pbio07.uni-muenster.de/	*Drosophila* development and genetics
Gene Expression Database (GXD)	http://www.informatics.jax.org/gxdindex.html	Mouse gene expression and genomics
Kidney Development Database	http://www.ana.ed.ac.uk/anatomy/database/kidbase/kidhome.html	Kidney development and gene expression
MAGEST	http://star.scl.kyoto-u.ac.jp/magest/	Ascidian (*Halocynthia roretzi*) gene expression patterns
Mouse Atlas and Gene Expression Database	http://genex.hgu.mrc.ac.uk	Spatially-mapped gene expression data
PEDB	http://chroma.mbt.washington.edu/PEDB/	Normal and aberrant prostate gene expression
Tooth Development Database	http://honeybee.helsinki.fi/toothexp/toothdev.htm	Gene expression in dental tissue
TRIPLES	http://ycmi.med.yale.edu/ygac/triples.htm	TRansposon-Insertion Phenotypes, Localization and Expression in *Saccharomyces*
Gene Identification and Structure		
Ares Lab Intron Site	http://www.cse.ucsc.edu/research/compbio/yeast_introns.html	Yeast spliceosomal introns
COMPEL	http://compel.bionet.nsc.ru/FunSite.html	Composite regulatory elements
CUTG	http://www.kazusa.or.jp/codon/	Codon usage tables
EID	http://mcb.harvard.edu/gilbert/EID/	Protein-coding, intron-containing genes
EPD	http://www.epd.isb-sib.ch	Eukaryotic POL II promoters
ExInt	http://intron.bic.nus.edu.sg/exint/exint.html	Exon–intron structure of eukaryotic genes
IDB/IEDB	http://nutmeg.bio.indiana.edu/intron/index.html	Intron sequence and evolution
PLACE	http://www.dna.affrc.go.jp/htdocs/PLACE	Plant *cis*-acting regulatory elements
PlantCARE	http://sphinx.rug.ac.be:8080/PlantCARE	Plant *cis*-acting regulatory elements
TransTerm	http://biochem.otago.ac.nz:800/Transterm/homepage.html	Codon usage, start and stop signals
TRRD	http://wwwmgs.bionet.nsc.ru/mgs/dbases/trrd4/	Regulatory regions of eukaryotic genes
YIDB	http://www.EMBL-Heidelberg.DE/ExternalInfo/seraphin/yidb.html	Yeast nuclear and mitochondrial intron sequences
Genetic Maps		
GeneMap ’99	http://www.ncbi.nlm.nih.gov/genemap/	International Radiation Mapping Consortium human gene map
G3-RH	http://www-shgc.stanford.edu/RH/	Stanford G3 and TNG radiation hybrid maps

Table 1. *Continued*

GB4-RH	http://www.sanger.ac.uk/RHserver/RHserver.shtml	Genebridge4 (GB4) human radiation hybrid maps
GDB	http://www.gdb.org	Human genes and genomic maps
DRESH	http://www.tigem.it/LOCAL/drosophila/dros.html	Human cDNA clones homologous to *Drosophila* mutant genes
GenAtlas	http://www.citi2.fr/GENATLAS/	Human genes, markers and phenotypes
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
Radiation Hybrid Database	http://www.ebi.ac.uk/RHdb	Radiation hybrid map data
Genomic Databases		
ACeDB	http://www.sanger.ac.uk/Software/Acedb/	*Caenorhabditis elegans*, *Schizosaccharomyces pombe* and human sequences and genomic information
FlyBase	http://www.fruitfly.org	*Drosophila* sequences and genomic information
Mouse Genome Database (MGD)	http://www.informatics.jax/org	Mouse genetics and genomics
Saccharomyces Genome Database (SGD)	http://genome-www.stanford.edu/Saccharomyces	*Saccharomyces cerevisiae* genome
AMmtDB	http://bio-www.ba.cnr.it:8000/BioWWW/#AMMTDB	Metazoan mitochondrial DNA sequences
Arabidopsis Database (AtDB)	http://genome-www.stanford.edu/Arabidopsis	*Arabidopsis thaliana* genome
CropNet	http://synteny.nott.ac.uk	Genome mapping in crop plants
CyanoBase	http://www.kazusa.or.jp/cyano/mutants	*Synechocystis* sp. genome
EcoGene	http://bmb.med.miami.edu/EcoGene/EcoWeb	*Escherichia coli* K-12 sequences
EMGlib	http://pbil.univ-lyon1.fr/emglib/emglib.html	Completely sequenced bacterial genomes and the yeast genome
GOBASE	http://megasun.bch.umontreal.ca/gobase/gobase.html	Organelle genome database
HIV Sequence Database	http://hiv-web.lanl.gov/	HIV RNA sequences
Human BAC Ends Database	http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_intro.html	Non-redundant human BAC end sequences
INE	http://www.dna.affrc.go.jp:82/giot/INE.html	Rice genetic and physical maps and sequence data
Mendel Database	http://jiio6.jic.bbsrc.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
MITONUC/MITOALN	http://bio-www.ba.cnr.it:8000/srs6/	Nuclear genes coding for mitochondrial proteins
MITOP	http://websvr.mips.biochem.mpg.de/proj/medgen/mitop	Mitochondrial proteins, genes and diseases
Munich Information Center for Protein Sequences (MIPS)	http://www.mips.biochem.mpg.de	Protein and genomic sequences
NRSub	http://pbil.univ-lyon1.fr/nrsub/nrsub.html	*Bacillus subtilis* genome
Phytophthora Genome Initiative Database	http://www.ncgr.org/pgi	Oomycete sequences and genetic maps
RsGDB	http://utmmg.med.uth.tmc.edu/sphaeroides	*Rhodobacter sphaeroides* genome
TIGR Microbial Database	http://www.tigr.org/tdb/mdb/mdb.html	Microbial genomes and chromosomes
ZFIN	http://zfish.uoregon.edu/ZFIN/	Zebrafish genetics and development; mutant and wild-type lines
ZmDB	http://zmdb.iastate.edu/	Maize genome database
Intermolecular Interactions		
Database of Ribosomal Crosslinks (DRC)	http://www.mpimg-berlin-dahlem.mpg.de/~ag_ribo/ag_brimacombe/drc/	Ribosomal crosslinking data
DIP	http://dip.doe-mbi.ucla.edu/	Catalog of protein–protein interactions
DPInteract	http://arep.med.harvard.edu/dpinteract/	Binding sites for *Escherichia coli* DNA-binding proteins
Metabolic Pathways and Cellular Regulation		
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.ad.jp/kegg	Metabolic and regulatory pathways
EcoCyc	http://ecocyc.pangeasystems.com/ecocyc	*Escherichia coli* K-12 genome, gene products and metabolic pathways

Table 1. *Continued*

ENZYME	http://www.expasy.ch/enzyme/	Enzyme nomenclature
EpoDB	http://cbil.humgen.upenn.edu/epodb	Genes expressed during human erythropoiesis
FlyNets	http://gifts.univ-mrs.fr/FlyNets/FlyNets_home_page.html	*Drosophila melanogaster* molecular interactions
Klotho	http://www.ibc.wustl.edu/klotho/	Collection and categorization of biological compounds
LIGAND	http://www.genome.ad.jp/dbget/ligand.html	Enzymatic ligands, substrates and reactions
RegulonDB	http://www.cifn.unam.mx/Computational_Biology/regulondb/	*Escherichia coli* pathways and regulation
UM-BBD	http://www.labmed.umn.edu/umbbd/	Microbial biocatalytic reactions and biodegradation pathways primarily for xenobiotic, chemical compounds
WIT2	http://wit.mcs.anl.gov/WIT2/	Integrated system for functional curation and development of metabolic models
Mutation Databases		
Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/Omim/	Catalog of human genetic and genomic disorders
ALFRED	http://fondue.med.yale.edu/db2/	Allele frequencies and DNA polymorphisms
Androgen Receptor Gene Mutations Database	http://www.mcgill.ca/androgendb/	Mutations in the androgen receptor gene
Asthma and Allergy Database	http://cooke.gsf.de	Genetics of allergy and asthma, including linkage studies and mutation data
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 Hematology	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
Cytokine Gene Polymorphism Database	http://www.pam.bris.ac.uk/services/GAI/cytokine4.htm	Cytokine gene polymorphisms, *in vitro* expression and disease-association studies
Database of Germline p53 Mutations	http://www.lf2.cuni.cz/homepage.html	Mutations in human tumor and cell line p53 gene
dbSNP	http://www.ncbi.nlm.nih.gov/dbSNP	Single nucleotidepolymorphisms
GRAP Mutant Databases	http://tinyGRAP.uit.no/GRAP/	Mutants of family A G-Protein Coupled Receptors (GRAP)
Haemophilia B Mutation Database	http://www.umds.ac.uk/molgen/haemBdatabase.htm	Point mutations, short additions and deletions in the Factor IX gene
HAMSTeRS	http://europium.mrc.rpms.ac.uk/usr/WWW/WebPages/main.dir/main.htm	Hemophilia A mutation database
HGBASE	http://hgbase.interactiva.de/	Intragenic sequence polymorphisms
HIV-RT	http://hivdb.stanford.edu/hiv/	HIV reverse transcriptase and protease sequence variation
Human Gene Mutation Database (HMGD)	http://uwcm.web.cf.ac.uk/uwcm/mg/hgmd0.html	Known (published) gene lesions responsible for human inherited disease
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 and Type III Collagen Mutation Database	http://www.le.ac.uk/genetics/collagen/	Human type I and type III collagen gene mutations
HvrBase	http://www.eva.mpg.de/hvrbase	Primate mtDNA control region sequences
iARC p53 Database	http://www.iarc.fr/p53/	Missense mutations and small deletions in human p53 reported in peer-reviewed literature.
KinMutBase	http://www.uta.fi/imt/bioinfo/KinMutBase/	Disease-causing protein kinase mutations
KMDB	http://mutview.dmb.med.keio.ac.jp/mutview3/kmeyedb/index.html	Mutations in human eye disease genes
MmtDB	http://www.ba.cnr.it/~areamt08/MmtDBWWW.htm	Mutations and polymorphisms in metazoan mitochondrial DNA sequences
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
p53 Databases	http://metalab.unc.edu/dnam/mainpage.html	Human p53 and hprt mutations; transgenic lacZ and transgenic/bacterial lacI mutations
PAHdb	http://www.mcgill.ca/pahdb/	Mutations at the phenylalanine hydroxylase locus
PMD	http://pmd.ddbj.nig.ac.jp/	Compilation of protein mutant data
RB1 Gene Mutation Database	http://home.kamp.net/home/dr.lohmann/	Mutations in the human retinoblastoma (RB1) gene
Ribosomal RNA Mutational Database	http://ribosome.FandM.edu/	16S and 23S ribosomal RNA mutation database

Table 1. *Continued*

SV40 Large T-Antigen Mutant Database	http://bigdaddy.bio.pitt.edu/SV40/	Mutations in SV40 large tumor antigen gene
Pathology		
FIMM	http://sdmc.krdl.org.sg:8080/fimm/	Functional molecular immunology data (diseases, antigens, peptides and HLA binding sites
Mouse Tumor Biology Database (MTB)	http://tumor.informatics.jax.org	Mouse tumor names, classification, incidence, pathology, genetic factors
PEDB	http://www.mbt.washington.edu/PEDB/	Sequences from prostate tissue and cell type-specific cDNA libraries
Protein Databases		
AARSDB	http://rose.man.poznan.pl/aars/index.html	Aminoacyl-tRNA synthetase sequences
DAtA	http://luggagefast.Stanford.EDU/group/arabprotein/	Annotated coding sequences from *Arabidopsis*
DExH/D Family Database	http://www.columbia.edu/~ej67/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 α/β hydrolase enzymes and relatives
FUNPEP	http://swift.embl-heidelberg.de/FUNPEP/	Low-complexity or compositionally-biased protein sequences
GenProtEC	http://dbase.mbl.edu/genprotec/start	*Escherichia coli* genes, gene products and homologs
GPCRDB	http://swift.embl-heidelberg.de/7tm/	G protein-coupled receptors
Histone Sequence 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://copan.bioz.unibas.ch/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://www.ebi.ac.uk/imgt/hla/	Immunoglobulin, T cell receptor and MHC sequences
InBase	http://www.neb.com/neb/inteins.html	Intervening protein sequences (inteins) and motifs
Kabat Database	http://immuno.bme.nwu.edu/	Sequences of proteins of immunological interest
LGIC	http://www.pasteur.fr/recherche/banques/LGIC/LGIC.html	Ligand-gated ion channel sequences, alignments and phylogeny
Membrane Protein Database	http://biophys.bio.tuat.ac.jp/ohshima/database/	Membrane protein sequences, transmembrane regions and structures
MEROPS	http://www.bi.bbsrc.ac.uk/Merops/Merops.htm	Peptidase sequences and structures
MHCPEP	http://wehih.wehi.edu.au/mhcpep/	MHC-binding peptides
NRR	http://nrr.georgetown.edu/nrr/NRR.html	Steroid and thyroid hormone receptor superfamily
Olfactory Receptor Database	http://ycmi.med.yale.edu/senselab/ordb/	Sequences for olfactory receptor-like molecules
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://delphi.phys.univ-tours.fr/Prolysis/	Protein kinase sequences, enzymology, genetics, and molecular and structural properties
PPMdb	http://sphinx.rug.ac.be:8080/ppmdb/index.html	*Arabidopsis* 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://www.nbrf.georgetown.edu/pir/	Non-redundant protein sequence database
Receptor Database (RDP)	http://impact.nihs.go.jp/RDB.html	Receptor protein sequences
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/	Sensory signal transduction proteins
SWISS-PROT/TrEMBL	http://www.expasy.ch/sprot	Curated protein sequences
TRANSFAC	http://transfac.gbf.de/TRANSFAC/index.html	Transcription factors and binding sites
Wnt Database	http://www.stanford.edu/~rnusse/wntwindow.html	Wnt proteins and phenotypes
Protein Sequence Motifs		
BLOCKS	http://www.blocks.fhcrc.org	Protein sequence motifs and alignments

Table 1. *Continued*

PROSITE	http://www.expasy.ch/prosite/	Biologically-significant protein patterns and profiles
Pfam	http://www.sanger.ac.uk/Software/Pfam/	Multiple sequence alignments and hidden Markov models of common protein domains
O-GLYCBASE	http://www.cbs.dtu.dk/databases/OGLYCBASE/	Glycoproteins and O-linked glycosylation sites
PIR-ALN	http://www-nbrf.georgetown.edu/pirwww/dbinfo/piraln.html	Protein sequence alignments
PRINTS	http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/printscontents.html	Protein squence motifs and signatures
ProClass	http://pir.georgetown.edu/gfserver/proclass.html	Families defined by PROSITE patterns and PIR superfamilies
ProDom	http://www.toulouse.inra.fr/prodom.html	Protein domain families
ProtoMap	http://www.protomap.cs.huji.ac.il/	Automated hierarchical classification of SWISS-PROT proteins
SBASE	http://www2.icgeb.trieste.it/~sbasesrv/	Annotated protein domain sequences
SMART	http://coot.embl-Heidelberg.de/SMART/	Signalling domain sequences
SYSTERS	http://www.dkfz-heidelberg.de/tbi/services/cluster/systersform	Protein clusters
Proteome Resources		
AAindex	http://www.genome.ad.jp/dbget/	Physicochemical properties of peptides
REBASE	http://rebase.neb.com/rebase/rebase.html	Restriction enzymes and associated methylases
SWISS-2DPAGE	http://www.expasy.ch/ch2d/	2D-PAGE images and reference maps
Yeast Proteome Database (YPD)	http://www.proteome.com/YPDhome.html	*Saccharomyces cerevisiae* proteome
Retrieval Systems and Database Structure		
KEYnet	http://www.ba.cnr.it/keynet.html	Keywords extracted from EMBL and GenBank
Virgil	http://www.infobiogen.fr/services/virgil	Database interconnectivity
RNA Sequences		
5S Ribosomal RNA Databank	http://www.rose.man.poznan.pl/5SData/5SRNA.html	5S rRNA sequences
ACTIVITY	http://wwwmgs.bionet.nsc.ru/systems/Activity/	Functional DNA/RNA site sequences
Collection of mRNA-like non-coding RNAs	http://www.man.poznan.pl/5SData/ncRNA/	Non-protein-coding RNA transcripts
Database on the Structure of Large Subunit Ribosomal RNA	http://rrna.uia.ac.be/	Alignment of large subunit ribosomal RNA sequences
Database on the Structure of Small Subunit Ribosomal RNA	http://rrna.uia.ac.be/ssu	Alignment of small subunit ribosomal RNA sequences
Guide RNA Database	http://www.biochem.mpg.de/~goeringe/	Guide RNA sequences
Intronerator	http://www.cse.ucsc.edu/~kent/intronerator/	RNA splicing and gene structure in *Caenorhabditis elegans*
Non-canonical Base Pair Database	http://prion.bchs.uh.edu/bp_type/	RNA structures containing rare base pairs
PLMItRNA	http://bio-www.ba.cnr.it:8000/srs/	Plant mitochondrial tRNAs and tRNA genes
Pseudobase	http://wwwbio.leidenuniv.nl/~Batenburg/PKB.html	Information on RNA pseudoknots
Ribosomal Database Project (RDP)	http://www.cme.msu.edu/RDP	rRNA sequences, alignments, and phylogenies
RNA Modification Database	http://medlib.med.utah.edu/RNAmods	Naturally modified nucleosides in RNA
SELEX_DB	http://wwwmgs.bionet.nsc.ru/mgs/systems/selex/	Selected DNA/RNA functional site sequences
Small RNA Database	http://mbcr.bcm.tmc.edu/smallRNA/smallrna.html	Direct sequencing of small RNA sequences
SRPDB	http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html	Signal recognition particle RNA, protein, and receptor sequences
tmRDB	http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html	tmRNA (10Sa RNA) sequences
tmRNA Website	http://sunflower.bio.indiana.edu/~kwilliam/tmRNA/home.html	tmRNA (10Sa RNA) sequences
tRNA Sequences	http://www.uni-bayreuth.de/departments/biochemie/trna/	tRNA and tRNA gene sequences
UTRdb	http://bigarea.area.ba.cnr.it:8000/EmbIT/UTRHome/	5′ and 3′ UTRs of eukaryotic mRNAs
Viroid and Viroid-Like RNA Database	http://www.callisto.si.usherb.ca/~jpperra	Viroid and viroid-like RNA and vHDV sequences
Yeast snoRNA Database	http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html	Yeast small nucleolar RNAs

Table 1. *Continued*

Structure		
PDB	http://www.rcsb.org/pdb/	Structure data determined by X-ray crystallography and NMR
CATH	http://www.biochem.ucl.ac.uk/bsm/cath/	Hierarchical classification of protein domain structures
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/	Familial and structural protein relationships
ASTRAL	http://astral.stanford.edu	Analysis of protein structures and their sequences
BioImage	http://www-embl.bioimage.org	Searchable database of multi-dimensional biological images
BioMagResBank	http://www.bmrb.wisc.edu/	NMR spectroscopic data from proteins, peptides and nucleic acids
CSD	http://www.ccdc.cam.ac.uk/prods/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
HIC-Up	http://alpha2.bmc.uu.se/hicup/	Structures of small molecules ('hetero-compounds')
HSSP	http://www.sander.ebi.ac.uk/hssp/	Structural families and alignments; structurally-conserved regions and domain architecture
IMB Jena Image Library	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 three-dimensional structures, linked to NCBI Entrez system
MODBASE	http://guitar.rockefeller.edu/modbase/	Comparative protein structure models
NDB	http://ndbserver.rutgers.edu/NDB/ndb.html	Nucleic acid-containing structures
PDB-REPRDB	http://www.rwcp.or.jp/papia	Representative protein chains, based on PDB entries
PRESAGE	http://presage.stanford.edu/	Protein structures with experimental and predictive annotations
Protein Motions Database	http://hyper.stanford.edu/~mbg/ftp/ProtMotDB/ProtMotDB.all.html	Motions of protein loops, domains and subunits
ProTherm	http://www.rtc.riken.go.jp/protherm.html	Thermodynamic data for wild-type and mutant proteins
RESID	http://www-nbrf.georgetown.edu/pirwww/dbinfo/resid.html	Protein structure modifications
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		
CarbBank	http://128.192.9.29/carbbank/	Complex carbohydrate/polysaccharide sequences
DBcat	http://www.infobiogen.fr/services/dbcat	Catalog of databases
DrugDB	http://pharminfo.com/drugdb/db_mnu.html	Pharmacologically-active compounds; generic and trade names
HOX-PRO	http://spirov.iephb.nw.ru/hox_pro/hox-pro00.html	Clustering of homeobox genes
LocusLink/RefSeq	http://www.ncbi.nlm.nih.gov/LocusLink	Curated sequence and descriptive information about genetic loci
Molecular Probe Database	http://srs.ebi.ac.uk/	Synthetic oligonucleotides, probes and PCR primers
MPDB	http://www.biotech.ist.unige.it/interlab/mpdb.html	Information on synthetic oligonucleotides
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
Tree of Life	http://phylogeny.arizona.edu/tree/phylogeny.html	Information on phylogeny and biodiversity
Vectordb	http://vectordb.atcg.com/	Characterization and classification of nucleic acid vectors