The Molecular Biology Database Collection: 2003 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

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 jumping-off point from which to find specialized databases that may be of use in advancing 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 this Collection are available through the *Nucleic Acids Research* Web site at http://nar.oupjournals.org.

COMMENTARY

The biological community will mark the completion of the Human Genome Project's major goal in April 2003: complete, high-accuracy sequencing of the human genome (1). This remarkable achievement, often compared to landing a man on the moon, lays the groundwork for a fundamental shift in how biological and biomedical research will be performed in the future. The free, widespread availability of a wide variety of data beyond human genome sequence—sequence variation data, model organism sequence data, expression data and proteomic data, to name a few—will provide a fertile playground for biologists in all disciplines to better-design and interpret their laboratory and clinical experiments, hopefully accelerating the pace of biological discovery.

Even though human sequencing is not yet 'complete' as a whole, sequencing has been completed on six human chromosomes as of the time of this writing (6, 7, 20, 21, 22 and Y). Along with the data available from numerous completed model genomes, the major public databases contain a phenomenal amount of sequence data. Currently, GenBank contains >17 billion nucleotide bases, representing >14 million sequences in 100 000 species. While the opportunities that this massive data set presents is mind-boggling, it also presents a problem in that the inexperienced user will either not know how to approach the data space or not know how to make best use of the data available to them. This problem will only continue to compound as GenBank continues its exponential rate of growth, with doubling rates on the order of 14 months or less. With the recent announcement of plans to sequence 'high-priority' model organisms by the National Human Genome Research Institute (NHGRI), it becomes more and more obvious that all biologists will need to avail themselves of the basic tools with which to navigate this large 'sequence space', as well as specialized databases that provide potentially easier access to subsets of the data.

Despite the large amount of publicity surrounding the Human Genome Project, a recent survey conducted on behalf of the Wellcome Trust indicates that only half of biomedical researchers using genome databases are familiar with the tools that can be used to actually access the data. For example, only 11% of those surveyed used the European Bioinformatics Institute's Ensembl Web site regularly, with 24% using it occasionally. Half of the remaining users had never even heard of Ensembl or its Web site. This low level of usage has led the Wellcome Trust to establish an advertising campaign aimed at increasing the public awareness of the availability of free tools such as Ensembl for searching human genome sequence data. Anecdotally, there is a similar lack of awareness or familiarity with the tools available through the University of California, Santa Cruz (UCSC) and, quite surprisingly, the National Center for Biotechnology Information (NCBI) at the National Institutes of Health, even though many biologists visit the NCBI Web site frequently. In response to this low level of awareness of the tools freely available to biologists, Wolfsberg *et al.* (2), developed a 'user's guide' to the human genome, intended to provide an elementary, hands-on guide for browsing and analyzing data produced by the International Human Genome Sequencing Consortium and other systematic sequencing efforts. The guide provides step-by-step instructions and strategies for using many of the most commonly-used tools for sequence-based discovery. NCBI, Ensembl and UCSC are all also in the process of developing (or have already released) similar, online guides for using the tools available on their respective Web sites.

*To whom correspondence should be addressed. Tel: +1 3014968570; Fax: +1 3014802634; Email: andy@nhgri.nih.gov

While educational efforts such as this help to address the need for rational ways to approach mining genomic data, additional efforts in the form of providing curated views of the data in specialized databases have been taking place for many years now. These efforts afford tremendous value to the biological researcher since they, in essence, reduce the massive 'sequence space' to specific, tractable areas of inquiry and, by doing so, allow for the inclusion of many more types of data than are found in the larger data repositories. These databases often provide not just sequence-based information, but additional data such as gene expression, macromolecular interactions, or biological pathway information, data that might not fit neatly onto a large physical map of a genome. Most importantly, data in these smaller, specialized databases tends to be curated by experts in a particular specialty and are often experimentally-verified, meaning that they represent the best state of knowledge in that particular area. 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, to promote the use of these resources in the design and analysis of experiments and to encourage the continued development of these resources. These reviewed databases are collectively listed in the Molecular Biology Database Collection.

The databases listed in this Collection distinguish themselves by their approach to presenting the underlying data—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 issue to accommodate full-length, printed descriptions of all of the databases making up the Collection. To address this, the online version of the Collection provides short summaries of many of the databases, the summaries having been provided directly by the investigators responsible for the individual databases. Contributors have been asked 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 are 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).

ACKNOWLEDGEMENTS

REFERENCES

1. Collins,F.S., Patrinos,A., Jordan,E., Chakravarti,A., Gestetland,R., Walters,L. and Members of the DOE and NIH Planning Groups (1998) New goals for the US Human Genome Project: 1998–2003. *Science*, **282**, 682–689.
2. Wolfsberg,T.G., Wetterstrand,K.A., Guyer,M.S., Collins,F.S. and Baxevanis,A.D. (2002) A User's Guide to the Human Genome. *Nature Genet.*, **32**(suppl.), 1–79.

Table 1. Molecular Biology Database Collection

Major 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
NCBI Reference Sequence Project	http://www.ncbi.nlm.nih.gov/RefSeq/	Non-redundant collection of naturally-occurring biological molecules
Ensembl	http://www.ensembl.org/	Annotated information on eukaryotic genomes
UCSC Genome Browser	http://genome.ucsc.edu/	Genome assemblies and annotation
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 43 complete genomes
CORG	http://corg.molgen.mpg.de	Conserved non-coding sequence blocks
Homophila	http://homophila.sdsc.edu	Relationship of human disease genes to genes in *Drosophila*
MBGD	http://mbgd.genome.ad.jp	Microbial genome database for comparative genomic analysis
ParaDB	http://abi.marseille.inserm.fr/paradb/	Paralogy mapping in human genomes
XREFdb	http://www.ncbi.nlm.nih.gov/XREFdb/	Cross-referencing of model organism genetics with mammalian phenotypes

Gene Expression

ArrayExpress	http://www.ebi.ac.uk/arrayexpress	Public collection of microarray gene expression data
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
EPConDB	http://www.cbil.upenn.edu/EPConDB	Endocrine pancreas consortium database
FlyView	http://pbio07.uni-muenster.de/	*Drosophila* development and genetics
Gene Expression Database (GXD)	http://www.informatics.jax.org/menus/expression_menu.shtml	Mouse gene expression and genomics
HugeIndex	http://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 (*Halocynthia roretzi*) gene expression patterns
MEPD	http://medaka.dsp.jst.go.jp/MEPD	Gene expression data from the small freshwater fish Medaka (*Oryzias latipes*)
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
MTID	http://mouse.ccgb.umn.edu/transposon/	*Sleeping beauty* transposon insertions in mice
NetAffx	http://www.affymetrix.com	Public Affymetrix probesets and annotations
RECODE expression	http://recode.genetics.utah.edu	Genes using programmed translational recoding in their expression
SeedGenes	http://www.seedgenes.org	Genes essential for *Arabidopsis* development
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
TRANSPATH	http://www.biobase.de/pages/products/databases.html	Gene regulatory networks and microarray analysis
TRIPLES	http://ygac.med.yale.edu	TRansposon-insertion phenotypes, localization, and expression in *Saccharomyces*

Table 1. *Continued*

Gene Identification and Structure

AllGenes	http://www.allgenes.org	Human and mouse gene index integrating gene, transcript and protein annotation
Ares Lab Yeast Intron Database	http://www.cse.ucsc.edu/research/compbio/ yeast_introns.htmlyeast_introns.html	Splicesomal introns in *Saccharomyces cerevisiae*
ASAP	http://www.bioinformatics.ucla.edu/ASAP	Alternative spliced isoforms
CUTG	http://www.kazusa.or.jp/codon/	Codon usage tables
DBTBS	http://elmo.ims.u-tokyo.ac.jp/dbtbs/	*Bacillus subtilis* binding factors and promoters
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
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
MICdb	http://www.cdfd.org.in/micas	Prokaryotic microsatellites
PACRAT	http://www.biosci.ohio-state.edu/∼pacrat	Archaeal and bacterial intergenic sequence features
PLACE	http://www.dna.affrc.go.jp/htdocs/PLACE	Plant *cis*-acting regulatory elements
PlantCARE	http://oberon.rug.ac.be:8080/PlantCARE/	Plant *cis*-acting regulatory elements
PlantProm	http://mendel.cs.rhul.ac.uk/	Proximal promoter sequences for RNA polymerase II
PromEC	http://bioinfo.md.huji.ac.il/marg/promec	*Escherichia coli* mRNA promoters with experimentally-identified transcriptional start sites
RRNDB	http://rrndb.cme.msu.edu	Variation in prokaryotic ribosomal RNA operons
rSNP Guide	http://util.bionet.nsc.ru/databases/rsnp.html	Single nucleotide polymorphisms in regulatory gene regions
RTPrimerDB	http://www.realtimeprimerdatabase.ht.st/	Validated PCR primer and probe sequence records
SNP Consortium database	http://snp.cshl.org	SNP Consortium data
SpliceDB	http://genomic.sanger.ac.uk/spldb/SpliceDB.html	Canonical and non-canonical mammalian splice sites
Sputnik	http://mips.gsf.de/proj/sputnik	Functional annotation of clustered plant ESTs
STRBase	http://www.cstl.nist.gov/div831/strbase/	Short tandem DNA repeats
TRANSCompel	http://www.gene-regulation.com/pub/ databases.html#transcompel	Composite regulatory elements
Transterm	http://uther.otago.ac.nz/Transterm.html	Codon usage, start and stop signals
TRRD	http://www.bionet.nsc.ru/trrd/	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 *Caenorhabditis elegans* 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 *Drosophila* 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
GDB	http://www.gdb.org	Human genes and genomic maps
GenAtlas	http://www.citi2.fr/GENATLAS/	Human genes, markers and phenotypes
GeneMap '99	http://www.ncbi.nlm.nih.gov/genemap/	International Radiation Mapping Consortium human gene map
Genetpig	http://www.infobiogen.fr/services/Genetpig	Comparative mapping in pig (*Sus scrofa*)
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
The Unified Database (UDB)	http://bioinfo.weizmann.ac.il/udb/	Integrated human maps

Table 1. *Continued*

Genomic Databases

ACeDB information	http://www.acedb.org/	*Caenorhabditis elegans,* *Schizosaccharomyces pombe*, and human sequences and genomic information
AMmtDB	http://bighost.area.ba.cnr.it/mitochondriome	Metazoan mitochondrial genes
ArkDB	http://www.thearkdb.org/	Genome databases for farm and other animals
ASAP	https://asap.ahabs.wisc.edu/ annotation/php/ASAP1.htm	Systematic annotation package for community-based annotation and analysis of genomes
BSD	http://bsd.cme.msu.edu	Comparative data on known biodegradative organisms
CATMA	http://www.catma.org	*Arabidopsis* gene sequence tags (GSTs)
CnidBase	http://www.cnidome.bu.edu/	Cnidarian evolutionary genomics and gene expression
Comprehensive Microbial Resource	http://www.tigr.org/tigr-scripts/CMR2/ CMRHomePage.spl	Completed microbial genomes
CropNet	http://ukcrop.net/	Genome mapping in crop plants
CroW 21	http://bioinfo.weizmann.ac.il/crow21/	Human chromosome 21 database
CyanoBase	http://www.kazusa.or.jp/cyano/	*Synechocystis* sp. genome
EcoGene	http://bmb.med.miami.edu/EcoGene/EcoWeb/	*E. coli* K-12 sequences
EMGlib	http://pbil.univ-lyon1.fr/emglib/emglib.html	Completely-sequenced prokaryotic genomes
ERGO	http://ergo.integratedgenomics.com/ERGO	Integrated biological data from genomic, biochemical, expression, and genetic experiments, and from the literature
FlyBase	http://flybase.bio.indiana.edu/	*Drosophila* sequences and genomic information
Full-Malaria	http://fullmal.ims.u-tokyo.ac.jp	Full-length cDNA library from erythrocytic-stage *Plasmodium falciparum*
GeneCards	http://bioinfo.weizmann.ac.il/cards/	Integrated database of human genes, maps, proteins and diseases
Genew	http://www.gene.ucl.ac.uk/cgi-bin/ nomenclature/searchgenes.pl	Approved symbols for all human genes
GOBASE	http://megasun.bch.umontreal.ca/gobase/gobase.html	Organelle genome database
GOLD	http://igweb.integratedgenomics.com/GOLD/	Information regarding complete and ongoing genome projects
GénoPlante-Info	http://genoplante-info.infobiogen.fr	Plant genomic data derived from the Génoplante consortium
GrainGenes	http://www.graingenes.org	Genomic database for small-grain crops
HGT-DB	http://www.fut.es/~debb/HGT/	Putative horizontally-transferred genes in prokaryotic genomes
HIV Sequence Database	http://hiv-web.lanl.gov/	HIV RNA sequences
HOWDY	http://www-alis.tokyo.jst.go.jp/HOWDY/	Integrated human genomic information
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	Prokaryotic protein-coding gene data
INE	http://rgp.dna.affrc.go.jp/giot/INE.html	Integrated database for rice genome analysis and sequencing
IRIS	http://www.iris.irri.org	Rice germplasm geneology and field data; rice structural and functional genomics and proteomics
Medicago Genome Initiative (MGI)	http://xgi.ncgr.org/mgi	Model legume Medicago ESTs, gene expression and proteomic data
Mendel Database family	http://www.mendel.ac.uk/	Database of plant EST and STS sequences annotated with gene family information
MIPS	http://www.mips.biochem.mpg.de/	Protein and genomic sequences
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/ BioWWW/#MitoNuc	Nuclear genes coding for mitochondrial proteins
MITOP	http://www.mips.biochem.mpg.de/ proj/medgen/mitop/	Mitochondrial proteins, genes and diseases
MOsDB	http://mips.gsf.de/proj/rice	*Oryza sativa* genome
Mouse Genome Database (MGD)	http://www.informatics.jax.org	Mouse genetics, genomics, alleles and phenotypes
MtDB	http://www.medicago.org/MtDB	*Medicago trunculata* genome
NRSub	http://pbil.univ-lyon1.fr/nrsub/nrsub.html	*B. subtilis* genome

Table 1. *Continued*

OGRe	http://www.bioinf.man.ac.uk/ogre	Complete mitochondrial genome sequences for 200 metazoan species
Oryzabase	http://www.shigen.nig.ac.jp/rice/oryzabase/	Rice genetics and genomics
PEDANT genome database	http://pedant.gsf.de	Automated analysis of genomic sequences
Phytophthora Genome Consortium Database	https://xgi.ncgr.org/pgc	ESTs from *Phytophthora infestans* and *Phytophthora sojae*
PlantGDB	http://zmdb.iastate.edu/PlantGDB/	Actively-transcribed plant genomic sequences
PlasmoDB	http://PlasmoDB.org	Plasmodium 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
RsGDB	http://www-mmg.med.uth.tmc.edu/sphaeroides	*Rhodobacter sphaeroides* genome
RTPrimerDB	http://www.realtimeprimerdatabase.ht.st	Real-time PCR primer and probe sequences
Saccharomyces Genome Database	http://genome-www.stanford.edu/Saccharomyces/	*Saccharomyces cerevisiae* genome
SOURCE	http://source.stanford.edu	Functional genomic resource for annotations ontologies, and expression data
SubtiList	http://genolist.pasteur.fr/SubtiList/	*Bacillus subtilis* 168 genome
The Arabidopsis Information Resource (TAIR)	http://www.arabidopsis.org/	*Arabidopsis thaliana* genome
TIGR Microbial Database	http://www.tigr.org/tdb/mdb/mdbcomplete.html	Microbial genomes and chromosomes
TIGR Rice Genome Annotation Resource	http://www.tigr.org/tdb/e2k1/osa1/	Rice sequence, BAC/PAC clones and related mapping data
ToxoDB: The Toxoplasma gondii Genome Database	http://ToxoDB.org	Apicomplexan parasite *Toxoplasma gondii* genome
WILMA	http://www.came.sbg.ac.at/wilma/	*Caenorhabditis elegans* annotation
WorfDB	http://worfdb.dfci.harvard.edu	*Caenorhabditis elegans* ORFeome
WormBase	http://www.wormbase.org/	Genomic data on *C. elegans* and related nematodes
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 (DIP)	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
DPInteract	http://arep.med.harvard.edu/dpinteract/	Binding sites for *E. coli* DNA-binding proteins
InterDom	http://InterDom.lit.org.sg	Putative protein domain interactions
JenPep	http://www.jenner.ac.uk/Jenpep2	Functional and quantitative thermodynamic data on peptide binding to immunological biomacromolecules
KDBI	http://xin.cz3.nus.edu.sg/group/kdbi.asp	Kinetic data on biomolecular interactions
MHC—Peptide Interaction Database	http://surya.bic.nus.edu.sg/mpid	Class I and Class II MHC-peptide complexes
STRING	http://www.bork.embl-heidelberg.de/STRING/	Predicted functional associations between proteins
Metabolic Pathways and Cellular Regulation		
EcoCyc	http://ecocyc.org/	*Escherichia coli* 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
Klotho	http://www.ibc.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
The University of Minnesota Biocatalysis Biodegradation Database	http://umbbd.ahc.umn.edu/	Curated information on microbial catabolism and related biotransformations
PathDB	http://www.ncgr.org/pathdb	Biochemical pathways, compounds and metabolism
PRODORIC	http://prodoric.tu-bs.de	Prokaryotic database of gene regulation and regulatory networks

Table 1. *Continued*

RegulonDB	http://www.cifn.unam.mx/Computational_Genomics/regulondb/	*Escherichia coli* 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 metabolic models
Mutation Databases		
ALFRED	http://alfred.med.yale.edu	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 oncology and haematology
BTKbase	http://bioinf.uta.fi/BTKbase/	Mutation registry for X-linked agammaglobulinemia
CASRDB	http://data.mch.mcgill.ca/casrdb/	CASR mutations causing FHH, NSHPT and ADH
Database of Germline p53 Mutations	http://www.lf2.cuni.cz/win/projects/germline_mut_p53.htm	Mutations in human tumor and cell line p53 gene
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/	Single nucleotide polymorphisms
FLAGdb/FST	http://genoplante-info.infobiogen.fr	*Arabidopsis thaliana* T-DNA transformants
GRAP Mutant Databases	http://tinyGRAP.uit.no/GRAP/	Mutants of family A G-Protein Coupled Receptors (GRAP)
Haemophila B Mutation Database IX	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/hprt, rodent lacI/lacZ databases	http://www.ibiblio.org/dnam/mainpage.html	Mutations at the human p53 and hprt genes; 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 and III Collagen Mutation Database	http://www.le.ac.uk/genetics/collagen/	Human type I and type III collagen gene mutations
iARC TP53 Database	http://www.iarc.fr/p53/	Human TP53 somatic and germline mutations
KinMutBase	http://www.uta.fi/imt/bioinfo/KinMutBase/	Disease-causing protein kinase mutations
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 Animals	http://www.angis.org.au/omia	Catalog of animal genetic and genomic disorders
Online Mendelian Inheritance in Man	http://www.ncbi.nlm.nih.gov/Omim/	Catalog of human genetic and genomic disorders
PAHdb	http://www.mcgill.ca/pahdb/	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 (RB1) gene
SV40 Large T-Antigen Mutant Database	http://bigdaddy.bio.pitt.edu/SV40/	Mutations in SV40 large tumor antigen gene
Pathology		
BayGenomics	http://baygenomics.ucsf.edu	Identification of genes relevant to cardiovascular and pulmonary disease
FIMM	http://sdmc.krdl.org.sg:8080/fimm/	Functional molecular immunology data
GOLD.db	http://gold.tugraz.at	Genes, proteins, and pathways implicated in lipid-associated disorders
INFEVERS	http://fmf.igh.cnrs.fr/infevers	Familial Mediterranean Fever and hereditary inflammatory disorder mutation data

Table 1. *Continued*

MetaFMF	http://fmf.igh.cnrs.fr/metaFMF/index_us.html	Familial Mediterranean Fever phenotype-genotype correlation
Mouse Tumor Biology Database (MTB) genetic	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
PGDB	http://www.ucsf.edu/PGDB	Genes and genomic loci related to the prostate and prostatic diseases
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 protein family of positive regulators in bacteria
ASPD	http://wwwmgs.bionet.nsc.ru/mgs/gnw/aspd/	Artificial Selected Proteins/ Peptides Database
CSDBase	http://www.chemie.uni-marburg.de/~csdbase/	Cold shock domain-containing proteins
DAtA	http://luggagefast.Stanford.EDU/group/arabprotein/	Annotated coding sequences from *Arabidopsis*
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
GenProtEC	http://genprotec.mbl.edu	*E. coli* K-12 genome, gene products and homologs
GPCRDB	http://www.gpcr.org/7tm/	G protein-coupled receptors
Histone Database	http://research.nhgri.nih.gov/histones/	Histone and histone fold sequences and structures
HIV Molecular Immunology Database	http://hiv-web.lanl.gov/immunology/	HIV epitopes
HIV RT and Protease Sequence Database	http://hivdb.stanford.edu	HIV reverse transcriptase and protease sequences
Homeobox Page	http://www.biosci.ki.se/groups/tbu/homeo.html	Information relevant to homeobox proteins, classification and evolution
Homeodomain Resource genomic	http://research.nhgri.nih.gov/homeodomain	Homeodomain sequences, structures and related genetic and genomic information
HORDE	http://bioinfo.weizmann.ac.il/HORDE/	Olfactory receptor genes and proteins
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/	Polymorphic sequences of human MHC and related genes
IMGT/MHC Database	http://www.ebi.ac.uk/imgt/mhc/	Major histocompatibility complex sequences
InBase	http://www.neb.com/neb/inteins.html	All known inteins (protein splicing elements): properties, sequences, bibliography
InterPro	http://www.ebi.ac.uk/interpro	Protein families and domains
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
Lipase Engineering Database	http://www.led.uni-stuttgart.de/	Integrated information on sequence, structure and function of lipases and esterases
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
MitoDrome	http://bighost.area.ba.cnr.it/BIG/MitoDrome	Drosophila nuclear genes encoding proteins targeted to the mitochondrion
MHCPEP	http://wehih.wehi.edu.au/mhcpep/	MHC-binding peptides

Table 1. *Continued*

MPIMP	http://millar3.biochem.uwa.edu.au/~lister/index.html	Mitochondrial protein import machinery of plants
Nuclear Protein Database (NPD)	http://npd.hgu.mrc.ac.uk	Proteins localized in the nucleus
Nuclear Receptor Resource	http://nrr.georgetown.edu/nrr/nrr.html	Nuclear receptor superfamily
NRMD	http://www.receptors.org/NR/	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
ooTFD	http://www.ifti.org/	Transcription factors and gene expression
PANTHER	http://panther.celera.com	Gene products organized by biological function
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
PIR-NREF	http://pir.georgetown.edu/pirwww/pirnref.shtml	Non-redundant reference database with comprehensive protein sequences
PKR	http://pkr.sdsc.edu	Protein kinase sequences, enzymology, genetics and molecular and structural properties
PLANT-PIs	http://bighost.area.ba.cnr.it/PLANT-PIs	Plant protease inhibitors
PlantsP/PlantsT	http://plantsp.sdsc.edu	Functional geomics databases focusing on protein involved in plant phosphorylation and membrane transport, respectively
PPMdb data	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
Protein Information Resource (PIR)	http://pir.georgetown.edu	Comprehensive, annotated, non-redundant protein sequence databases
ProtoNet	http://www.protonet.cs.huji.ac.il/	Hierarchical clustering of SWISS-PROT
Ribonuclease P Database	http://www.mbio.ncsu.edu/RNaseP/home.html	RNase P sequences, alignments and structures
RTKdb	http://pbil.univ-lyon1.fr/RTKdb/	Receptor tyrosine kinase sequences
S/MARt dB	http://transfac.gbf.de/SMARtDB/	Nuclear scaffold/matrix attached regions
SDAP	http://fermi.utmb.edu/SDAP	Sequences, structures and IgE epitopes of allergenic proteins
SENTRA	http://wit.mcs.anl.gov/WIT2/Sentra/HTML/sentra.html	Sensory signal transduction proteins
SEVENS	http://sevens.cbrc.jp	7-transmembrane helix receptors
SRPDB	http://bio.lundberg.gu.se/dbs/SRPDB/SRPDB.html	Structural and functional information on signal recognition particles
SWISS-PROT/TrEMBL	http://www.expasy.ch/sprot	Curated protein sequences
TIGRFAMs	http://www.tigr.org/TIGRFAMs	Functional identification of proteins
TRANSFAC	http://transfac.gbf.de/TRANSFAC/index.html	Transcription factors and binding sites
trEST, trGEN, Hits	http://hits.isb-sib.ch	Hypothetical protein sequences
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		
ASC—Active Sequence Collection	http://crisceb.unina2.it/ASC/	Biologically-active short amino acid sequences
Blocks	http://blocks.fhcrc.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 domains	http://www.ebi.ac.uk/interpro/	Integrated documentation resource for protein families, domains, and sites
*i*ProClass	http://pir.georgetown.edu/iproclass/	Annotated protein database with family, function and structure information
NESbase 1.0	http://www.cbs.dtu.dk/databases/NESbase	Nuclear export signals
NLSdb	http://cubic.bioc.columbia.edu/db/NLSdb/	Nuclear localization signals
O-GLYCBASE	http://www.cbs.dtu.dk/databases/OGLYCBASE/	*O*- and *C*-linked glycosylation sites in proteins
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

Table 1. *Continued*

ProClass patterns	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.org/sbase	Protein domain sequences and tools
SMART	http://smart.embl-heidelberg.de	Simple Modular Architecture Research Tool
SUPFAM	http://pauling.mbu.iisc.ernet.in/~supfam	Grouping of sequence families into superfamilies
SYSTERS, GeneNest, SpliceNest	http://cmb.molgen.mpg.de	Integrated database of protein families, EST clusters and their genomic positions
TMPDB	http://bioinfo.si.hirosaki-u.ac.jp/~TMPDB/	Experimentally-characterized transmembrane topologies
Proteome Resources		
AAindex	http://www.genome.ad.jp/aaindex/	Physicochemical and biological properties of amino acids
GELBANK	http://gelbank.anl.gov	2D-gel electrophoresis patterns from completed genomes
PEP: Predictions for Entire Proteomes	http://cubic.bioc.columbia.edu/pep/	Summarized analyses of protein sequences
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.org/ch2d/	Annotated two-dimensional polyacrylamide gel electrophoresis database
Retrieval Systems and Database Structure		
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 Ribosomal RNA Mutation Database	http://www.fandm.edu/Departments/ Biology/Databases/RNA.html	16S and 23S ribosomal RNA mutations
5S Ribosomal RNA Database	http://biobases.ibch.poznan.pl/5SData/	5S rRNA sequences
ACTIVITY	http://util.bionet.nsc.ru/databases/activity.html	Functional DNA/RNA site activity
ARED	http://rc.kfshrc.edu.sa/ared	AU-rich element-containing mRNAs
Database for mobile group II introns	http://www.fp.ucalgary.ca/group2introns/	Database for mobile group II introns
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 *C. elegans*; alignments of *C. briggsae* and *C. elegans* genomic sequences
IRESdb	http://ifr31w3.toulouse.inserm.fr/IRESdatabase/	Internal ribosome entry sites
NCIR	http://prion.bchs.uh.edu/bp_type/	Non-standard base-base interactions in known RNA structures
Noncoding regulatory RNAs database	http://biobases.ibch.poznan.pl/ncRNA/	Noncoding RNAs with regulatory functions
PLANTncRNAs	http://www.prl.msu.edu/PLANTncRNAs/	Plant non-protein coding RNAs with relevant gene expression information
Plant snoRNA DB	http://www.scri.sari.ac.uk/plant_snoRNA/	snoRNA genes in plant species
PLMItRNA	http://bighost.area.ba.cnr.it/PLMItRNA/	Mitochondrial tRNA genes and molecules in photosynthetic eukaryotes
PseudoBase	http://wwwbio.leidenuniv.nl/~Batenburg/PKB.html	Structural, functional and sequence data related to RNA pseudoknots
Rfam	http://www.sanger.ac.uk/Software/Rfam/	Non-coding RNA families
Ribosomal Database Project (RDP-II)	http://rdp.cme.msu.edu	rRNA sequence data, analysis tools, alignments and phylogenies
RISCC	http://ulises.umh.es/RISSC	Ribosomal 16S–23S RNA gene spacer regions
RNA Modification Database	http://medlib.med.utah.edu/RNAmods/	Naturally modified nucleosides in RNA
SELEXdb	http://wwwmgs.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
Subviral RNA Database	http://penelope.med.usherb.ca/subviral/	Database of viroids and viroid-like RNAs
tmRDB	http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html	tmRNA (10Sa RNA) sequences and alignments
tRNA Sequences	http://www.uni-bayreuth.de/ departments/biochemie/trna/	tRNA and tRNA gene sequences
tmRNA Website	http://www.indiana.edu/~tmrna	tmRNA sequences, foldings, and alignments
UTRdb/UTRsite	http://bighost.area.ba.cnr.it/srs6/	5′- and 3′-UTRs of eukaryotic mRNAs and relevant functional patterns
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
BioMagResBank acids	http://www.bmrb.wisc.edu/	NMR spectroscopic data from proteins peptides, and nucleic acids
CADB	http://144.16.71.148	Conformation angles of protein structures, with associated crystallographic data
CATH	http://www.biochem.ucl.ac.uk/bsm/cath_new	Protein domain structures
CE	http://cl.sdsc.edu/ce.html	CE: a resource to compute and review 3D protein structure alignments
CKAAPs DB	http://ckaap.sdsc.edu	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/%7Efaculty/mini/ dsdbase/dsdbase.html	Native and modeled disulfide bonds in proteins
DSMM	http://projects.eml.org/mcm/database/dsmm	Database of Simulated Molecular Motions
E-MSD	http://www.ebi.ac.uk/msd	Collected data on macromolecular structures
FAMSBASE	http://famsbase.bio.nagoya-u.ac.jp/famsbase/	Protein three-dimensional structural models
Gene3D	http://www.biochem.ucl.ac.uk/ bsm/cath_new/Gene3D/	Precalculated structural assignments for genes within whole genomes
GTOP	http://spock.genes.nig.ac.jp/~genome/gtop.html	Protein fold predictions from genome sequences
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; structurarlly-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 linked	http://www.ncbi.nlm.nih.gov/Structure/	All experimentally-determined three-dimensional structures, linked to NCBI Entrez
MolMovDB	http://MolMovDB.org	Database of macromolecular movements
ModBase	http://guitar.rockefeller.edu/modbase	Annotated comparative protein structure models
NDB	http://ndbserver.rutgers.edu/	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://ncbs.res.in/%7Efaculty/mini/campass/pass.html	Structural motifs of protein superfamilies
PDB	http://www.pdb.org/	Structure data determined by X-ray crystallography and NMR
PDB-REPRDB	http://www.cbrc.jp/pdbreprdb/	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
PSSH	http://srs3d.ebi.ac.uk/	Alignments between protein sequences and tertiary structures
RESID	http://www-nbrf.georgetown.edu/pirwww/dbinfo/resid.html	Protein structure modifications
RNABase	http://www.rnabase.org	RNA-containing structures from PDB and NDB
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
Structure-Superposition Database	http://ssd.rbvi.ucsf.edu	Pairwise superposition of TIM-barrel structures
SUPERFAMILY	http://supfam.org	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 alignments	http://www-igbmc.u-strasbg.fr/BioInfo/BAliBASE2/index.html	Benchmark database for comparison of multiple sequence alignments
Cytokine Gene Polymorphism in Human Disease	http://bris.ac.uk/pathandmicro/services/GAI/cytokine4.htm	Cytokine gene polymorphism literature database
DBcat	http://www.infobiogen.fr/services/dbcat/	Catalog of databases
Global Image Database	http://www.gwer.ch/qv/gid/gid.htm	Annotated biological images
GlycoSuiteDB	http://www.glycosuite.com	N- and O-linked glycan structures and biological source information
Imprinted Genes and Parent-of-Origin Effects	http://www.otago.ac.nz/IGC	Imprinted genes and parent-of-origin effects in animals
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/	Names of all organisms that are represented in the genetic databases with at least one nucleotide or protein sequence
probeBase	http://www.probeBase.net	rRNA-targeted oligonucleotide probe sequences, DNA microarray layouts and associated information
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/	MEDLINE and Pre-MEDLINE citations
RefSeq	http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html	Reference sequence standards for genomes, genes, transcripts and proteins
RIDOM	http://www.ridom.de/	rRNA (16S and ITS) sequence-based identification of medical microorganisms
SWEET-DB	http://www.dkfz-heidelberg.de/spec2/sweetdb/	Annotated carbohydrate structure and substance information
The Pharmacogenomics and Pharmacogenetics Knowledge Base	http://www.pharmgkb.org	Variation in drug response based on human variation
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://viroligo.okstate.edu	Virus-specific oligonucleotides for PCR and hybridization