Important Databases Related to Genomes

9

Abstract

Genome databases are the locations, which permits storing, sharing, retrieving and comparison of the information related to the genomes of various individuals and organisms. Traditionally the databases were confined to the updated information of certain vital model organisms. Rapid development of technology and high speed internet facilities have created an explosion of databases resulting in the development of specific databases of almost all model organisms and a group of organisms with a common specificity. In the present chapter, details of the databases related to the genomes of viruses, archaea, bacteria, cell organelle, invertebrates, vertebrates, plants, and human beings are provided in a table format which provides an instant information, about different databases and their URLs.

9.1   Virus Databases

Database	Description	URL
CoVDB	Coronavirus genes and genomes	http://covdb.microbiology.hku.hk/
DPV web	Central source of information about viruses, viroids and satellites of plants, fungi, and protozoa	http://www.dpvweb.net/
euHCVdb	European hepatitis C virus database	http://euhcvdb.ibcp.fr/
HCV database	Hepatitis C virus database	http://hcv.lanl.gov/
HERVd	Human endogenous retrovirus database	http://herv.img.cas.cz/
HFV database	Hemorrhagic fever virus sequence database	http://hfv.lanl.gov/
HIV drug resistance DB	Compilation of mutations in HIV genes that confer resistance to anti-HIV drugs	http://www.hiv.lanl.gov/content/sequence/RESDB/
Influenza research DB	Influenza virus	http://www.fludb.org

(continued)

NCBI viral genomes	Viral genomes resource	http://www.ncbi.nlm.nih.gov/genomes/VIRUSES/viruses.html
Papillomavirus episteme	A database of Papillomaviridae family of viruses	http://PaVE.niaid.nih.gov
phiSITE	Gene regulation in bacteriophages	http://www.phisite.org
Poxvirus.org	Poxvirus genomic sequences and gene annotation	http://www.poxvirus.org/
VBRC	Viral bioinformatics resource center	http://www.biovirus.org/
ViTa	microRNAs targets of the influenza virus	http://vita.mbc.nctu.edu.tw/
ViralZone	Molecular and epidemiological information on viral genera and families	http://viralzone.expasy.org/
ViralGenomes	Viral sequence data resource by NCBI	https://www.ncbi.nlm.nih.gov/genome/viruses/

9.2   Archaeal Databases

Database	Description	URL
Ensembl	A browser for archaeal genomes	https://bacteria.ensembl.org/index.html
EZ cloud	EZ cloud for archaea	http://www.ezbiocloud.net/ezgenome
IMG	Integrated microbial genomes and microbiomes	https://img.jgi.doe.gov/
MicrobesOnline resource	A collection of 3707 microbial genomes	http://www.microbesonline.org/
BacDive	The bacterial diversity metadatabase contains detailed information on more than 20,000 strains of bacteria	https://bacdive.dsmz.de/
Wikipedia	List of sequenced archaeal genomes	https://en.wikipedia.org/wiki/List_of_sequenced_archaeal_genomes
UCSC archaeal genome browser	Integrated genome browser for archaeal genomes	http://archaea.ucsc.edu/

9.3   Bacterial Databases

Database	Description	URL
Ensembl bacteria	A browser for bacterial and archaeal genomes	https://bacteria.ensembl.org/index.html
MBGD	Database for comparative analysis of completely sequenced microbial genomes	http://mbgd.genome.ad.jp/
Human microbiome project	Characterisation of microbes in healthy humans	https://hmpdacc.org/

GOLD	Genomes online database	https://gold.jgi.doe.gov/
EZ cloud	EZ cloud for bacteria	http://www.ezbiocloud.net/ezgenome
IMG	Integrated microbial genomes and microbiomes	https://img.jgi.doe.gov/
MicrobesOnline resource	A collection of 3707 mmicrobial genomes	http://www.microbesonline.org/
BacDive	The bacterial diversity metadatabase contains detailed information on more than 20,000 strains of bacteria	https://bacdive.dsmz.de/
BIGSdb	Bacterial isolate genome sequence database	https://bigsdb.readthedocs.io/en/latest/
metaMicrobesOnline	Website for browsing and comparing the genomes of 3527 microbial genomes	http://meta.microbesonline.org/
MetaRef	A pan-genomic database for comparative and community microbial genomics	https://omictools.com/metaref-tool
MetaBioME database	MetaBioME is a web resource to find novel homologs for known commercially useful enzymes (CUEs) in metagenomic datasets and completed bacterial genomes	https://metasystems.riken.jp/metabiome/
PATRIC	Pathosystems resource integration center	https://www.patricbrc.org/
TBDB	Tuberculosis database integrates genomic sequences and data for *Mycobacterium* species relevant to drug discovery, vaccines, and biomarkers	http://www.tbdb.org/
EuPathDB	Eukaryotic pathogen database resources is a collection of individual databases, each focusing on specific pathogens, accessible through a common portal	https://eupathdb.org/
coliBase	A database for *E. coli*, *Salmonella* and *Shigella*	http://xbase.warwick.ac.uk/colibase/
DEG	Database of essential genes from bacteria and yeast	http://www.essentialgene.org
Essential genes in *E. coli*	First results of an *E. coli* gene deletion project	http://www.genome.wisc.edu/resources/essential.htm
GenoBase	Escherichia coli K-12 in-frame, single-gene knockout mutants	http://ecoli.naist.jp/
HGT-DB	Horizantal gene transfer database is a genomics database that includes G+C content, codon and amino-acid usage, as well as information on which genes deviate in these parameters for prokaryotic complete genomes	http://genomes.urv.es/HGT-DB/
Pseudomonas genome database	Consists of information about Pseudomonas genome	http://www.pseudomonas.com/
PAIDB	Pathogeneticity islands database	http://www.paidb.re.kr/about_paidb.php

9.4   Cell Organelle Databases

Database	Description	URL
Chloroplast genome database	Chloroplast genome	http://chloroplast.cbio.psu.edu/
OrganelleDB	A web-accessible database cataloging proteins localized to a known organelle	http://organelledb.lsi.umich.edu/
Organelle genomes	Organelle genome resource at NCBI	www.ncbi.nlm.nih.gov/genomes/ORGANELLES/organelles.html
Plant organelles database	Images of plant organelles and protocols for plant organelle research	http://podb.nibb.ac.jp/Organellome/
PLProt	*Arabidopsis thaliana* chloroplast protein database	http://www.plprot.ethz.ch/
Plastid-LCG base	Plastid lineage-based conserved gene-pair database	http://lcgbase.big.ac.cn/plastid-LCGbase/
HMPD	Human mitochondrial protein database	http://bioinfo.nist.gov/hmpd/
HmtDB	Human mitochondrial DataBase	http://www.hmtdb.uniba.it/
MamMiBase	Database of complete mitochondrial genome sequences of over 100 mammalian species	http://www.mammibase.lncc.br/
MitoCarta	Mouse and human mitochondrial proteins	http://www.broadinstitute.org/pubs/MitoCarta
MitoDrome	Annotating D. melanogaster nuclear genes encoding mitochondrial proteins and support studies on human diseases related to mitochondrial dysfunctions	http://mitodrome.ba.itb.cnr.it/
MitoPorteome	Mitochondrial protein sequences encoded by mitochondrial and nuclear genes	http://www.mitoproteome.org/
OGRe	Database containing information on completely sequenced animal mitochondrial genomes of 473 species	http://ogre.mcmaster.ca/
Arabidopsis mitochondrial protein database	Database comprising of experimentally identified proteins in *Arabidopsis* mitochondrial genome	www.plantenergy.uwa.edu.au/applications/ampdb/index.html

9.5   In-Vertebrate Genome Databases

Database	Description	URL
AphidBase	Genome database of pea aphid (*Acyrthosiphon pisum*)	http://www.aphidbase.com/
AppaDB	Database of nematode *Pristionchus pacificus*	http://appadb.eb.tuebingen.mpg.de
BeetleBase	Database of the beetle *Tribolium castaneum*	http://www.beetlebase.org
CeNDR	*C. elegans* natural diversity resource	http://www.elegansvariation.org

Comparasite	Database for comparative studies of transcriptomes of parasites	http://comparasite.hgc.jp/
CuticleDB	Structural proteins of arthropod cuticle	http://bioinformatics.biol.uoa.gr/cuticleDB/
FlyBrain	Database of drosophila nervous system	http://flybrain.neurobio.arizona.edu/
Helminth.net	Parasitic nematode sequencing project	http://www.helminth.net
InsectBase	Insect genomes and transcriptomes	http://www.insect-genome.com/
Hymenoptera genome	Genome sequences and annotation for honey bee and the parasitoid wasp *Nasonia vitripennis*	http://HymenopteraGenome.org
InsectBase	Insect genomes and transcriptomes	http://www.insect-genome.com/
SilksatDB	Comprehensive relational database of silkworm microsatellite	http://www.cdfd.org.in/SILKSAT/index.php
SpodoBase	Genomics of the butterfly *Spodoptera frugiperda*	http://bioweb.ensam.inra.fr/spodobase/
VectorBase	Invertebrate vectors of human pathogens	http://www.vectorbase.org
RNAiDB	RNAi phenotypic analysis of *C. elegans* genes	http://www.rnai.org/
NEMBASE	Nematode sequence and functional data database	http://www.nematodes.org/
PLANMINE	Platform for archiving and mining planarian biodiversity	http://planmine.mpi-cbg.de/
WORMBASE	Database of *Caenorhabditis elegans*	http://www.wormbase.org
WorfDB	Worm ORF database which contains predicted proteins from *C. elegans*	http://worfdb.dfci.harvard.edu/

9.6   Vertebrate Genome Databases

Database	Description	URL
FANTOM	Functional annotation of mouse full-length cDNA clones	http://fantom.gsc.riken.jp/
GWIPS-viz	Genomes of human beings and vertebrates	http://gwips.ucc.ie
Xenbase	Human and Vertebrate Genomes	http://www.xenbase.org
Mammalian gene collection	Identify and sequence cDNA clones of FL or ORF for human, mouse and rat genes	http://mgc.nci.nih.gov/
PlasmID	A repository for collection and distribution of plasmid clones for human and vertebrate genomes	http://plasmid.hms.harvard.edu/
Mouse genome database	Mouse genome database for integrated data on the genetics, genomics and biology of the laboratory mouse	http://www.informatics.jax.org

9.7   Human Genome Databases

Database	Description	URL
3DSNP	Database for comprehensively annotating the regulatory function of human noncoding SNPs by exploring their 3D interactions with genes and genetically associated SNPs mediated by chromatin loops	http://biotech.bmi.ac.cn/3dsnp/
ENCODE at UCSC	Encyclopedia of DNA elements an international a consortium of investigators funded to analyze the human genome with the goal of producing a comprehensive catalog of functional elements	http://genome.ucsc.edu/ ENCODE
GeneAnnot	Revises and improves the annotation of affymetrix probe sets	http://genecards.weizmann.ac.il/ geneannot/
Locus reference genome sequences	Each sequence is a stable genomic DNA sequence for a region of the human genome	http://www.lrg-sequence.org
QTL match maker	Database which integrates QTLs of human mouse and rat genomes to annotate their functional genomes	http://pmrc.med.mssm.edu:9090/ QTL/jsp/qtlhome.jsp
TIARA	Total integrated archive of short-read and array (TIARA) accumulates raw-level personal genomic data from whole genome next-generation sequencing (NGS) and comparative genomic hybridization (CGH) arrays	http://tiara.gmi.ac.kr
TRBASE	Database that relates tandem repeats to the gene locations and disease genes in the human genome	http://trbase.ex.ac.uk/
UniSTS	A comprehensive database of sequence tagged sites (STSs), each of which is defined by a distinct pair of primer sequences designed for use in a PCR reaction	http://www.ncbi.nlm.nih.gov/ probe
X:MaP	Annotation and visualization of genome structure for Affymetrix exon array analysis	http://annmap.picr.man.ac.uk/
EVOLA	A database of human genes and their vertebrate orthologs	http://jbirc.jbic.or.jp/hinv/evola/
H-InvDB	HumanInvitational database is an integrated database of human genes and transcripts	http://www.h-invitational.jp/
HGPD	Human gene and Protein database is a unique database that stores information from human gateway entry clones and in vitro expression data related to human proteins	http://hupex.hgpd.jp/entrance/ index.html

HRPD	Human protein reference database that stores the data related to domain architecture, post-translational modifications, interaction networks and disease	http://www.hprd.org/
LIFEdb	A database on the localisation, interaction, functional assays and expression of human proteins	http://www.dkfz.de/gpcf/lifedb.php
NetAffx	A database of public affymetrix probesets and annotations	http://www.affymetrix.com
ORFDB/ ORFeome	Invitrogen initiated program intended to provide the open reading frames (ORFs) of all human proteins	http://orf.invitrogen.com/
dbGaP	Database of genotypes and phenotypes	http://www.ncbi.nlm.nih.gov/gap
MSY breakpoint mapper	STSs on human Y chromosome	http://breakpointmapper.wi.mit.edu/
Gene cards	An automated, integrated database of human genes, genomic maps, proteins and diseases	http://www.genecards.org/

9.8   Plant Genome Databases

Database	Description	URL
AgBase	A curated, open-source, web-accessible resource for functional analysis of agricultural plant and animal gene products	http://www.agbase.msstate.edu/
autoSNPdb	Plant SNP discovery database	http://autosnpdb.appliedbioinformatics.com.au/
BarleyBase	Database for barley microarrays holds the data for over 1000 hybridizations from the affymetrix Barley1 GeneChip and genome arrays	http://www.barleybase.org/
Cereal small RNA database	An integrated resource for small RNAs such as miRNAs, siRNAs, ta-siRNAs expressed in rice and maize	http://sundarlab.ucdavis.edu/smrnas/
CR-EST	Crop EST database	https://apex.ipk-gatersleben.de/apex/f?p=116:1
EURISCO	European catalogue for plant genetic resources	http://eurisco.ecpgr.org/
GABiPD	Genome analysis of the plant biological system	http://www.gabipd.org/
GoMapMan	Unified plant-specific gene ontology	http://www.gomapman.org
GrainGenes	Molecular and phenotypic information on grain yielding plants such as wheat, barley, rye, triticale, and oats	http://wheat.pw.usda.gov/

GreeNC	Green non-coding database (GreeNC) is a repository of long non-coding RNAs annotated in 37 plant species and six algae	http://greenc.sciencedesigners.com/wiki/Main_Page
MIPS plants database	Plant database resource for integrative and comparative plant genome research	http://mips.gsf.de/proj/plant/jsf/
NIAS	Databases for plant genetic resources and plant disease information	http://www.gene.affrc.go.jp/databases_en.php
P-MITE	Database of plant miniature inverted-repeat transposable elements (MITEs)	http://pmite.hzau.edu.cn/django/mite/
PGDD	A plant genome duplication database	http://chibba.agtec.uga.edu/duplication/
Phytozome	A database for green plant genomics	http://www.phytozome.net/
PIECE	Database of plant intron exon comparison and evolution	https://wheat.pw.usda.gov/piece/
Plant DNA C-values database	A database that provides a one-stop, user-friendly database where the plant genome sizes can be readily accessed and compared	http://www.kew.org/genomesize/homepage.html
Plant ontology database	This database is a collection of plant systematics, botany and genomics involved in developing simple and controlled vocabularies that accurately reflect the biology of plant structures	http://www.plantontology.org/
Plant SnoRNA database	A database of snoRNA genes in plant species	http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home
Plant stress-responsive gene catalog	Stress-responsive gene in various plant species	http://dayhoff.generationcp.org/
Plant GDB	A database of plant genomic sequences and ESTs	http://www.plantgdb.org/
Plant PAN	Plant promoter analysis navigator database	http://PlantPAN2.itps.ncku.edu.tw
PlantProm	A database of plant promoter sequences	http://www.epd.isb-sib.ch/
PlantsP/PlantsT	PlantsP and PlantsT are plant-specific curated databases that combine sequence derived information with experimental functional genomics data	http://plantsp.sdsc.edu/
Plant tribes	Families of protein-coding genes from five sequenced plant species	http://fgp.huck.psu.edu/tribedb/index.pl
PlantTFDB	Plant transcription factor database provides a comprehensive, high-quality resource of plant transcription factors, regulatory elements and interactions between them	http://planttfdb.cbi.pku.edu.cn
PLAZA	A plant-oriented online resource for comparative, evolutionary and functional genomics	https://bioinformatics.psb.ugent.be/plaza

PMDBase	A database of plant microsatellites and marker development	http://www.sesame-bioinfo.org/PMDBase
TIGR database	Plant transcript assemblies database	http://plantta.tigr.org/
TROPGene	A database comprising of genetic and genomic information about tropical crops such as banana, coconut, cotton, cocoa, etc.	http://tropgenedb.cirad.fr/
ForestTree DataBase	A resource that centralizes large-scale EST sequencing results from several tree species	http://foresttree.org/ftdb
LegumeIP	Model legumes integrative database platform	http://plantgrn.noble.org/LegumeIP/

9.9   Genomes of Economically Important and Model Plants

Database	URL
Medicago trunculata database	http://www.medicago.org/
Brassica genome	http://www.brassicagenome.net/
Coffee genome hub	http://coffee-genome.org/
CottonGen	http://www.cottongen.org
Maize genetics and genomics database (MaizeGDB)	http://www.maizegdb.org/
The Arabidopsis information resource (TAIR)	https://www.arabidopsis.org/
Oryza sativa (ssp. japonica) genome	http://www.plantgdb.org/OsGDB/
The rice annotation project database	https://rapdb.dna.affrc.go.jp/
PoMaMo—potato maps and more	http://www.gabipd.org/projects/Pomamo/
Shanghai rapeseed database	http://rapeseed.plantsignal.cn/
Genome database for Rosaceae	http://www.rosaceae.org
Solanaceae genomics network	http://solgenomics.net/
SoyBase	http://soybase.org/
SoyKB—soybean knowledge base	http://soykb.org
TED—tomato functional genomics database	http://ted.bti.cornell.edu/
TIGR maize database	http://maize.tigr.org/
TomatEST DB	http://biosrv.cab.unina.it/tomatestdb/
TomatEST DB	http://biosrv.cab.unina.it/tomatestdb/
PANZEA	http://www.panzea.org/
Diatom EST database	http://www.biologie.ens.fr/diatomics/EST/