第 4-3 章　生物信息学数据库和在线分析工具

第一节　重要门户网站和分子数据库

一、重要门户网站

·美国国家生物技术信息中心，National Center for Biotechnology Information（NCBI）：www.ncbi.nlm.nih.gov/

·欧洲生物信息学研究所/欧洲分子生物学实验室，The European Bioinformatics Institute（EBI）/European Molecular Biology Laboratory（EMBL）：www.ebi.ac.uk

·蛋白质序列分析专家系统（Expasy），瑞士生物信息学研究所，Swiss Institute of Bioinformatics：www.expasy.org

·国际生物信息学动态及会议发布网站：www.bioinformatics.org

·SeqAnswer 国际生物信息技术问答论坛：http://seqanswers.com

·Biostar 生物信息技术问答论坛（BioStar：An Online Question & Answer Resource for the Bioinformatics Community）：www.biostars.org

二、主要分子数据库

数据库类型及名称	网址
1. DNA 和 RNA 序列数据库	
GenBank	http://www.ncbi.nlm.nih.gov/
EMBL	http://www.ebi.ac.uk
DDBJ	http://www.ddbj.nig.ac.jp/
2. 蛋白质序列数据库	
SWISS-PROT/TrEMBL/UniProtKB	http://www.uniprot.org/
PIR（Protein Information Resource）	http://pir.georgetown.edu/
3. 蛋白质功能域数据库	
InterPro	http://www.ebi.ac.uk/interpro/
PFAM	http://pfam.xfam.org/
PROSITE	http://prosite.expasy.org/
SMART	http://smart.embl-heidelberg.de/
CDD（Conserved Domain Databases）	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
iProClass	http://pir.georgetown.edu/pirwww/dbinfo/iproclass.shtml

续表

数据库类型及名称	网址
ProDom	http://prodom.prabi.fr/prodom/current/html/form.php
PRINTS Database	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/index.php
4. 蛋白质结构数据库	
PDB	http://www.rcsb.org/pdb
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop
CATH	http://www.biochem.ucl.ac.uk/bsm/cath
NRL-3D	http://pir.georgetown.edu/pirwww/search/textnrl3d.html
HSSP	http://www.sander.embl-heidelberg.de/hssp
PDBE	http://www.ebi.ac.uk/pdbe/
PDBJ	http://pdbj.org/
5. 蛋白质组学数据库	
PRIDE	http://www.ebi.ac.uk/pride/archive/
CORUM	http://mips.gsf.de/genre/proj/corum/index.html
ConsensusPathDB	http://cpdb.molgen.mpg.de
SWISS-2DPAGE	http://world-2dpage.expasy.org/swiss-2dpage/
6. 代谢途径和细胞调控数据库	
KEGG	http://www.genome.jp/kegg/
BRENDA	http://www.brenda-enzymes.org
IMP	http://imp.princeton.edu/
PlantCyc（Plant Metabolic Pathway Database）	http://www.plantcyc.org/
MANET	http://www.manet.uiuc.edu/
NCBI BioSystems	http://www.ncbi.nlm.nih.gov/biosystems
Pathguide	http://www.pathguide.org
MapMan	http://mapman.gabipd.org/web/guest
Human Metabolome Database	http://www.hmdb.ca/
Microme	http://www.microme.eu/
SMPDB	http://smpdb.ca/
7. 非编码 RNA 数据数据库	
miRBase	http://www.mirbase.org
piRNAbank	http://pirnabank.ibab.ac.in/
GtRNAdb	http://gtrnadb.ucsc.edu/
SILVA	http://www.arb-silva.de/
lncRNAdb	http://www.lncrnadb.org/
lncRNADisease	http://cmbi.bjmu.edu.cn/lncrnadisease

续表

数据库类型及名称	网址
LncRNAWiki	http：//lncrna.big.ac.cn/index.php/Main_Page
Noncode	http：//noncode.org/
siRNA database	http：//web.mit.edu/sirna/；http：//sirna.sbc.su.se/
CircNet	http：//circnet.mbc.nctu.edu.tw/
circbase	http：//www.circbase.org/
Rfam	http：//rfam.xfam.org/
8. 基因功能分类数据库	
GO	http：//geneontology.org/
KEGG	http：//www.genome.jp/kegg/
9. 人类基因组及遗传、疾病相关数据库	
UCSC（University of California Santa Cruz）genome browser	http：//genome.ucsc.edu/
The 1000 Genomes Project	http：//www.1000genomes.org/
Personal Genome Project	http：//www.personalgenomes.org/
Online Mendelian Inheritance in Man（OMIM）	http：//omim.org/
TCGA（The Cancer Genome Atlas）	http：//cancergenome.nih.gov/
Human Epigenome Project（HEP）	http：//www.epigenome.org/
EpiFactors	http：//epigenie.com/epigenetic-tools-and-databases/
MethDB（the database for DNA methylation and environmental epigenetic effects）	http：//www.methdb.de/
The database of human DNA Methylation and Cancer（MethyCancer）	http：//methycancer.psych.ac.cn/
Oncomine（癌症芯片及整合数据挖掘平台）	https：//www.oncomine.org/
TIGR DATAbase	http：//www.tigr.org/tdb/hcd/overview.html
HUPO	https：//www.hupo.org/
10. 模式植物基因组数据库	
Phytozome	https：//phytozome.jgi.doe.gov/pz/portal.html
PlantGDB	http：//www.plantgdb.org/
EnsemblPlants	http：//plants.ensembl.org/index.html
TAIR	http：//arabidopsis.org/
BRAD	http：//brassicadb.org/
Sol genomics	https：//www.solgenomics.net/

续表

数据库类型及名称	网址
RGAP	http://rice.plantbiology.msu.edu/
RAP-DB	http://rapdb.dna.affrc.go.jp/
Gramene	http://www.gramene.org/
MaizeGDB	http://www.maizegdb.org/
PineRefSeq	http://pinegenome.org/pinerefseq/
11. 模式动物基因组数据库	
UCSC genome browser	http://genome.ucsc.edu/
EnsemblMetazoa	http://metazoa.ensembl.org/index.html
Flybase	http://flybase.org/
Mouse Genome	http://www.informatics.jax.org/ http://rgd.mcw.edu/
Zebrafish genome	http://zfin.org/
Xenbase	http://www.xenbase.org/entry/
Wormbase	http://www.wormbase.org/
Gorilla	http://asia.ensembl.org/Gorilla_gorilla/Info/Index/
Nematode	http://nematode.net/
12. 主要微生物基因组数据库	
EnsemblFungi	http://fungi.ensembl.org/index.html
EnsemblBacteria	http://bacteria.ensembl.org/index.html
Yeast genome（SGD）	http://www.yeastgenome.org/
iMicrobe	http://data.imicrobe.us/

第二节　生物信息学主要在线分析工具

工具类型及名称	网址
1. 序列搜索和检索	
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi
HMMER	http://www.ebi.ac.uk/Tools/hmmer/
ENTREZ	http://www.ncbi.nlm.nih.gov/gquery/
Batch Entrez	http://www.ncbi.nlm.nih.gov/sites/batchentrez
BLAT	http://genome.ucsc.edu/cgi-bin/hgBlat? hgsidcommand=start;
SRS	http://www. ncbi. nlm. nih. gov/Class/NAWBIS/Modules/Protein/protein10.html

续表

工具类型及名称	网址
2. 多序列联配/比对	
CLUSTALW	http://www.genome.jp/tools/clustalw/
MAFFT	http://mafft.cbrc.jp/alignment/server/
MUSCLE	http://www.ebi.ac.uk/Tools/msa/muscle/
T−Coffee	http://tcoffee.org/Projects/tcoffee/
M−COFFEE	http://www.tcoffee.org/Projects/mcoffee/
3. 基因注释/预测	
FgeneSH	http://www.softberry.com/berry.phtml
GENSCAN	http://genes.mit.edu/GENSCAN.html
Gene Finder	http://rulai.cshl.org/tools/genefinder/
GenomeScan	http://genes.mit.edu/genomescan.html
AUGUSTUS	http://bioinf.uni−greifswald.de/webaugustus/
4. 序列开放阅读框(ORF)查找和翻译工具	
ORFINDER	http://www.ncbi.nlm.nih.gov/gorf/gorf.html
Transeq	http://www.ebi.ac.uk/Tools/st/emboss_transeq/
5. 蛋白质分析工具	
蛋白质二级结构预测	
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/；http://www.cmpharm.ucsf.edu/~nomi/nnpredict；http://www.embl−heidelberg.de/predictprotein/SOPMA； http://www.embl−heidelberg.de/sspred/ssprd_info.html
ProteinPredict	https://www.predictprotein.org/
GOR	https://npsa−prabi.ibcp.fr/
NetSurfP	http://www.cbs.dtu.dk/services/NetSurfP/
Jpred	http://www.compbio.dundee.ac.uk/jpred/
APSSP	http://imtech.res.in/raghava/apssp/
蛋白质三级结构预测	
SWISS−MODEL	http://expasy.hcuge.ch/swissmod/SWISS−MODEL.html
CPHmodels	http://www.cbs.dtu.dk/services/CPHmodels/
ESyPred3D	http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/
GENO3D	https://geno3d−prabi.ibcp.fr/

<div align="right">续表</div>

工具类型及名称	网址
Phyre	http://www.sbg.bio.ic.ac.uk/
Pymol	https://www.pymol.org/

6. 基因组分析方面

Multiple tools for genome analysis	http://molbiol-tools.ca/Genomics.htm

7. 系统进化树构建

PhyML	http://www.atgc-montpellier.fr/phyml/
RAxML	http://embnet.vital-it.ch/raxml-bb/
Phylogenyfr	http://www.phylogeny.fr/
Phylogeny	http://www.ebi.ac.uk/Tools/phylogeny/clustalw2_phylogeny/
iTOL – interactive Tree Of Life	http://itol.embl.de/

第三节　生物信息学主要开放分析软件

软件类型及名称	网址
1. 序列搜索和检索	
BLAST	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST
HMMER	http://hmmer.org/
2. 多序列联配/比对	
CLUSTALW	http://www.clustal.org/
MAFFT	http://mafft.cbrc.jp/alignment/software/
MUSCLE	http://www.drive5.com/muscle
3. 基因注释/预测	
SOAPdenovo	http://soap.genomics.org.cn/soapdenovo.html
ALLPATHS-LG	http://www. broadinstitute. org/science/programs/genome – biology/crd
AUGUSTUS	http://bioinf.uni-greifswald.de/augustus/
GeneMark	http://exon.gatech.edu/GeneMark/
PASA	http://pasapipeline.github.io/
EVidenceModeler	https://evidencemodeler.github.io/
Maker	http://www.yandell-lab.org/software/maker.html
4. 蛋白质分析工具	
InterProScan（Protein domain identification）	http://www.ebi.ac.uk/interpro/download.html
PSIPRED（Secondary structure prediction）	http://bioinfadmin.cs.ucl.ac.uk/downloads/psipred/

续表

软件类型及名称	网址
ProteoWizard	http://proteowizard.sourceforge.net/index.shtml
5. 非编码 RNA 分析工具	
MIREAP	https://sourceforge.net/projects/mireap/
The ViennaRNA Package	http://www.tbi.univie.ac.at/RNA/
The UEA small RNA Workbench	http://srna-workbench.cmp.uea.ac.uk/downloadspage/
CPC（Coding Potential Calculator）	http://cpc.cbi.pku.edu.cn/download/
find_circ（circular RNA finding）	https://github.com/marvin-jens/find_circ
CIRCexplorer（circular RNA finding）	http://yanglab.github.io/CIRCexplorer/
circ RNA_pocket（circRNA viewing）	http://ibi.zju.edu.cn/bioinplant/tools/computational_tools.htm
6. 基因组可视化	
Circos	http://circos.ca/software/
Gbrowser	http://hgdownload.soe.ucsc.edu/downloads.html
JBrowser	http://jbrowse.org/
IGV	https://www.broadinstitute.org/igv/
SeeDNA	www.itp.ac.cn/~hao/SeeDNA.tar.gz
7. 系统进化树构建	
MEGA	http://www.megasoftware.net/
RAxML	http://sco.h-its.org/exelixis/web/software/raxml/
Phylip	http://evolution.genetics.washington.edu/phylip.html
PAML	http://abacus.gene.ucl.ac.uk/software/paml.html
MrBayes	http://mrbayes.sourceforge.net/
TreeView	http://jtreeview.sourceforge.net/
FastTree	http://meta.microbesonline.org/fasttree/
CVTree	http://tlife.fudan.edu.cn/cvtree3/