

国内外生物信息学 WWW 服务器及数据库

朱佳鸣 郭 勇 唐 权 祖云霞

(东南大学医学院图书馆情报技术部 南京 210009)

摘要: 生物信息学是一门新兴的交叉学科, 近年来各类生物信息学数据库不断出现, 其规模呈爆炸趋势增长, 同时数据结构日趋复杂。目前生物信息学数据库服务已实现了高度的计算机和网络化。介绍国内外主要的生物信息学 WWW 服务器及其所含数据库的种类、作用。

关键词: 生物信息学, WWW 服务器, 数据库

中图分类号: Q93 **文献标识码:** A **文章编号:** 0253-2654 (2002) 01-0099-03

生物信息学 (bioinformatics) 是近年来发展并完善起来的热门交叉学科。数据库是生物信息学的基础, 几乎覆盖了生命科学各个领域, 包括核酸序列数据库, 蛋白质序列数据库, 蛋白质片段数据库, 蛋白质三维结构数据库, 文献数据库等各种数据库。

爆炸性增长和复杂程度的增加是生物信息学数据库的重要特征。由于其覆盖面广, 分布分散且格式不统一, 因此一些生物计算中心将多个数据库整合在一起提供综合服务, 以方便使用。

1 ExPASy Molecular Biology Server: <http://www.expasy.org/>

ExPASy 由瑞士生物信息学学会 (<http://www.isb-sib.ch/>) 建立, 主要提供蛋白质序列和结构分析数据库、研究工具和软件。

ExPASy 包含的主要数据库有: (1) SWISS-PROT + TrEMBL、SWISS-2DPAGE、PROSITE; 提供蛋白质家族, 区域数据。(2) SWISS-3DIMAGE: 提供已知三维结构的

收稿日期: 2001-01-12, 修回日期: 2001-09-21

生物高分子图象数据。(3) SWISS-MODEL Repository: 提供自动生成的蛋白质结构模型。(4) ENZYME: 提供酶的命名信息。(5) CD40Lbase: 提供 CD40 抗原缺陷与高 IgM 综合症的临床与分子生物学数据。(6) SeqAnalRef: 蛋白质序列分析文献书目数据库。

2 European Bioinformatics Institute (EBI): <http://www.ebi.ac.uk/>

欧洲生物信息学实验室 (<http://www.embl.org/>) 提供核酸、蛋白质序列, 高分子结构等生物信息数据库服务, 同时提供各种研究工具, 数据提交, 并提供数据库和软件 FTP 下载。

该服务器包含的主要数据库有: (1) EMBL Nucleotide Sequence Database: 收集欧洲研究 DNA、RNA 基因序列的个人、项目组提交的数据, 包括基因序列专利申请。(2) Macromolecular Structure Database (MSD): 生物高分子 NMR 光谱、X 射线晶体衍射和电镜结构数据库。(3) Ensembl: 人类基因组数据库。(4) InterPro Resource of Protein Domains and Sites: 蛋白质区域和功能的综合站点。(5) Complete Genomes: 提供 EMBL 已完成的基因组数据。(6) Proteome Analysis: 基因蛋白质表达分析数据库。(7) GeneQuiz: 可综合利用各种检索和分析方法, 对最新的蛋白质和 DNA 数据库进行研究, 得到基因、蛋白质序列及其生物学功能数据。

3 National Center for Biotechnology Information (NCBI): <http://www.ncbi.nlm.nih.gov/>

美国国立生物技术信息中心 WWW 服务器提供的主要数据库有: (1) GenBank: 具有基因数据的递交、修改和查询功能, 提供与之相关的生物学信息和参考文献。(2) Molecular database: 包括核苷酸序列数据库、蛋白质序列数据库、结构模型数据库和生物分类数据库。核苷酸序列数据库包括核苷酸的 ENTREZ 检索系统 (Entrez Nucleotides) 及表达序列标签数据库 (database of expressed sequence Tags, sbEST), 序列标签位点数据库 (database of sequence tagged sites, adSTS), 寡核苷酸多态性数据库 (database of single nucleotide polymorphisms, dbSNP) 等专业数据库; 蛋白质序列数据库包括蛋白质的 ENTREZ 检索系统 (Entrez Proteins), 蛋白质域数据库 (Conserved domain database, CDD), Blast 蛋白质同源性分析系统, web 蛋白质综述 (Protein Reviews On the Web) 等; 结构模型数据库包括蛋白质分子三维结构数据库 (molecular modelling database, MMDb), Cn3D3.0 蛋白质三维结构浏览器, 蛋白质 VAST 矢量对比搜索工具 (Vector Alignment Search Tool), VAST 结构同源性检索等; 生物分类数据库包括 NCBI 生物分类数据库及其浏览器等。(3) 文献数据库 (Literature Databases)。(4) 在线人类孟德尔遗传数据库 (Online Mendelian Inheritance in Man, OMIM) 列出与人类疾病相关的基因的序列的文本、图谱和相关文献等。(5) Genomes and Maps: 包含了人类基因组资源 (human genome resources), 疾病相关基因, 唯一基因序列族, 人类基因组测序索引, 癌相关基因组计划, 基因表达的序列分析, 人鼠同源性基因比较图谱, 同源进化群簇等专题数据库。

4 DNA Data Bank of Japan (DDBJ): <http://ddbj.nig.ac.jp/>

日本国立遗传研究所的日本 DNA 数据库, 具有数据递交功能和数据库检索和分析功能, 包括 getentry 的基因序列号途径; SRS, FASTA 和 BLAST 同源性检 (下转 103 页)

(上接 100 页) 索; TXSearch 生物分类数据库检索; CLUSTALW 蛋白质和 DNA 数据分析; GIB 基因组信息分析; GTOPI 基因组、蛋白的结构和功能分析; PDB 蛋白质结构浏览器; SS-Thread 蛋白质二级结构预测和 LIBRA 三维结构预测系统等。

5 北京大学生物信息服务器: <http://mdl.ipc.pku.edu.cn/mirror/mirror.html>

北京大学生物信息服务器 (PeKing University BIOinformatics Server, PKUBIOS) 提供了各种数据库的镜像, 免费软件及生物信息的在线检索。

镜像数据库包括: (1) PDB: The Brookhaven Protein Data Bank, 蛋白质、核酸、碳水化合物、病毒的 3D 结构数据库。(2) SCOP: The structural classification of proteins, 描述蛋白质结构及蛋白质进化关系数据库。其它镜像有: SWISS-PROT、PIR、PROSITE 等。

利用该服务器可进行 DNA、蛋白质序列, 高分子晶体、计算机辅助的药物设计、同源性和结构预测、蛋白质结构和功能、分子模型等研究。

6 其它生物信息 WWW 服务器有: 澳大利亚国立基因信息服务器 ANGIS (Australian National Genomic Information Service: <http://morgan.angis.su.oz.au/>), 瑞士肿瘤实验研究学会的 ISREC 生物信息服务器 (Institute Sulesse de Recherche Expérimental sur le Cancer: <http://www.isrec.isb-sib.ch/>), 英国医学研究学会下属的人类基因组计划数据中心管理的 BIOSCI 服务器 (bionet Electro-nic Newsgroup Network for Biology: <http://www.bio.net/>), 欧洲生物信息网瑞士站点 (Swiss node of EMBnet: <http://www.ch.embnet.org/>), 日本京都大学的日本基因组服务器 (Japanese GenomeNet: <http://www.genome.ad.jp/>) 等。

参考文献

- [1] Iliopoulos I, Tsoka S, Andrade M A. Genome Biology, 2000, 2 (1): interactions 0001. 1~3.
- [2] Stoesser G, Baker W, Break van den A. Nucleic Acids Research, 2001, 29 (1): 17~21.
- [3] Bairoch A, Gasteiger E, Hoogland C. <http://www.expasy.ch/doc/expasy.rtf>: Version of Feb. 2001, 14.
- [4] Benson D A, Karsch-Mizrachi I, Lipman D J. Nucleic Acids Res, 2000, 28 (1): 15~18.
- [5] DNA Data Bank of Japan: <http://ddbj.nig.ac.jp/>.
- [6] PeKing University BIOinformatics Server: <http://mdl.ipc.pku.edu.cn/mirror/mirror.html>.