

经验交流

英特网上生物信息资源的利用

李 勇 倪福太¹⁾ 贺福初

(北京放射医学研究所, 北京 100850)

摘要 充分利用英特网上完善的生物信息资源是十分有益且必要的. 介绍了利用网上生物资源的几种主要方式和一些重要的站点和生物信息搜索引擎.

关键词 英特网, 生物信息

学科分类号 Q7

在生命科学高速发展的今天, 生物信息成指数增长, 这些信息多以数据库的形式存储于计算机中, 通过英特网可以免费查阅它们, 这就是网上生物信息资源^[1]. 这种资源形式的优点是: a. 便于查寻, 只要点击鼠标即可; b. 全面, 包括了生物学的各个方面且不断充实更新; c. 便于自学, 无论多么复杂和庞大的数据库、软件或机构, 在其网站的主页都有内容介绍或 FAQ (Frequent Ask Question), 上面列举了访问者最常遇到的问题及答案. 在当今知识爆炸的时代, 充分利用这些资源, 及时得到所需信息是十分有益且必要的. 以下是利用网上生物资源的几种主要方式:

1.1 查阅文献

MEDLINE 是美国国立医学图书馆创立的规模最大、权威性最高的文献数据库. 它收录了 1966 年至今 70 多个国家出版的生物医学期刊 4 000 多种, 内容涉及医学、动植物、微生物、食品等多个领域. 其高级检索页面 (<http://www.healthgate.com/healthgate/medline/search-advanced.shtml>) 可同时以主题词和作者进行检索, 还可选择所查阅文献的年代. MEDLINE 也有缺陷, 一是它以医学为主, 植物学等方面收录不够全面, 另一个是只能查到摘要. 这些缺陷可由其他在线杂志搜索引擎弥补. 搜索引擎 yahoo (<http://www.yahoo.com/science/biology/journals>) 以学科分类收录了上千种各类生物学杂志. 哈佛大学的在线图书馆 (<http://www.mcb.harvard.edu/library/ejbio.htm>) 以字母顺序列举了除医学之外的几千种生物学杂志, 其中一些杂志可下载全文, 包括图表.

1.2 检索数据库

最主要的生物学数据库是核酸、蛋白质序列数

据库及其三维结构数据库. DNA 序列数据库主要有: 美国国立卫生研究院全国生物技术研究中心 (NCBI) 的 GenBank (<http://www.ncbi.nlm.nih.gov>), 欧洲生物信息学研究所的 EMBL 数据库, 日本的 DNA 数据库 (DDBJ) (<http://www.nig.ac.jp>); 蛋白质序列数据库主要是瑞士日内瓦大学的 SWISS-PROT; 蛋白质结构数据库主要是美国 Brookhaven 国家实验室的 PDB (Protein Data Bank). NCBI 开发的 ENTREZ 系统综合了上述各大数据库的信息和 MEDLINE 的文献信息, 其特点是对于每一个序列给出了类似的序列或相关的结构、功能和文献信息^[2].

此外, 还有一些专项数据库, 如载体序列数据库、人类突变基因数据库、酵母数据库等.

1.3 使用网上软件

虽然有 GCG、PC GENE 等功能强大的商品化软件, 但价格昂贵、操作复杂、需要经过专业培训才能掌握, 而且每隔一段时间要升级的缺点限制了它们的应用. 而使用网上软件则避免了上述缺点. 网上软件根据功能可分为三类:

1.3.1 同源性检索软件: 利用这类软件检索数据库得到与靶序列具有同源性的序列, 在一定程度上起到核酸杂交实验的作用, 因而被称为电子杂交. NCBI 的 BLAST 软件 (Basic Local Allynnet Search Tool, 局部比对基本检索工具) 是较常用的一种同源性检索工具, 由一组功能不同的程序组成, 其中 BLASTN 用于在核酸序列库检索核酸序列; BLASTP 用于在蛋白质序列库中检索氨基酸序列;

¹⁾ 四平师范学院生物系, 四平 136000.

收稿日期: 1998-10-07, 修回日期: 1999-01-15

TBLASIN 则可以在核酸序列中检索氨基酸序列. 另一种常用的检索程序是 FASTA. 可以对同一序列使用这两种方法进行检索以增加找到同源序列的可能性^[3].

1.3.2 未知序列分析软件: 随着人类基因组计划的开展, 争夺有限基因资源的竞争日益激烈. 利用网上软件分析未知序列, 迅速得到尽可能多的信息显得尤为重要. 仅仅在几年前, 网上只零散地分布着一些功能简单的小程序用于新基因的鉴别, 现在这些软件有了长足的发展, 其中一些优秀的软件根据功能归类被集中在一个网页上 (<http://bioinformatics.weizmann.ac.il/gdp/gdp.html>), 极大地方便了使用.

1.3.3 常用工具软件, 如: DNA 序列酶切位点分析的软件 (<http://www.medkem.gu.se/cutter>), PCR 引物设计软件 (<http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>) 等.

1.4 学术交流

网上有上千个与生物学有关的消息组, 每个消息组以某一研究领域为主题. 活跃的专题讨论组就象一个没有围墙的专家系统, 能够对一些专业问题迅速作出反应, 有些用户还会公布实验中创新的技术方法和心得体会, 以方便他人的工作; 组织完善的消息组不但是交流的场所, 还可以提供大量信息. 如: 细胞因子消息组 (<http://bioinformatics.weizmann.ac.il/cytokine>) 的主页上, 汇集了有关杂志、产品、会议通知、网址等信息, 可以说是细胞因子的百科全书.

除了专一主题的消息组外, 还有更大规模的生物学家电子论坛, 如: BIONET 和 BIOMOO. BIONET (<http://www.bio.net>) 由美国科学基金会创立于 1991 年, 目的是使全世界的生物学家利用网络进行交流, 它根据不同专题分为几十个消息组. 用户以电子邮件的形式参与消息组的活动; BIOMOO (<http://bioinformatics.weizmann.ac.il/BioMOO>) 是生物学家网上聚会的场所, 即所谓的虚拟会场. 目前有上千名生物学家经常参加 BIOMOO 的活动.

2 生物信息搜索引擎

英特网上生物学站点之繁多如天上星辰, 信息

量之巨大如浩淼海洋, 如何以最短的时间找到所需的信息? 生物信息搜索引擎是良好的开端. 哈佛大学的生物信息搜索引擎 Biology Links (<http://golgi.harvard.edu>) 是 Internet 生物学的总汇, 将生物资源分成: 生物学相关机构、在线杂志、模式生物数据库、分子生物学数据库、软件等几类, 收录全面. Pedro's 工具 (http://www.public.iastate.edu/~pedro/research_tools.html) 是较为直接的搜索引擎, 包括生物分析、文献检索和帮助工具三部分, 其中每部分以字母顺序列举了上百个站点的超链接, 通过超链接可直接进入所需网页. 此外, 德国基因组和生物信息学研究所 (<http://bioinformatics.weizmann.ac.il/index.html>) 的搜索引擎以其 FAQ 见长, 回答了用户查询生物信息时最常遇到的问题, 方便了初学者的使用.

对于种类繁多的分子生物学数据库, SRS (Sequence Retrieval System) (<http://www.ebi.ac.uk/srs>) 是功能强大的检索系统, 它能检索 45 个数据库, 而且这些数据库通过超链接整合在一起, 通过这些超链接能方便地不同数据库间跳跃浏览.

参 考 文 献

- 1 Dracopoli C N, Haines L J, Korf B R, *et al.* Current Protocols in Human Genetics. USA: John Wiley & Sons Inc, 1998. Supplement, 11. 6
- 2 夏云, 雷二庆, 王槐春. 英特网实用技术与生物医学应用. 北京: 军事医学科学出版社 (Xia Y, Lei E Q, Wang H C. Internet Technology and Application in Biology, Medicine. Beijing: Military Medical Science Press), 1997. 317~ 318
- 3 奥斯伯 F, 部伦特 R, 金斯顿 R E, 等. 精编分子生物学实验指南. 北京: 科学出版社 (Ausubel F M, Brent R, Kingston R E, *et al.* Short Protocols in Molecular Biology, Beijing: Science Press), 1998. 240~ 251

Make Good Use of Bioinformatical Resource on Line. LI Yong, NI Fu-Tai¹⁾, HE Fu-Chu (*Institute of Radiation Medicine, Beijing 100850, China;*
¹⁾ *Siping Teacher's College, Siping 136000, China*).

Abstract Some important bioinformatical web sites and searching engines were introduced, in order to make good use of biological resource on line.

Key words internet, bioinformatics