

生物信息

微生物资源的信息化

张何 莫湘涛 夏立秋*

(湖南师范大学生命科学学院微生物学系 长沙 410081)

摘要: 随着计算机技术的飞跃发展和互联网的渗透普及, 微生物研究者已经能够通过网络获取微生物许多领域的信息, 甚至包括某一微生物的完整基因组, 这些在以前是不敢想象的, 信息的快速收集已经极大程度扩展了微生物研究人员的研究范围以及研究能力; 同时, 高度发达的的网际互联为微生物学的信息交流、资源共享和国际合作带来了前所未有的机遇。

关键词: 国际互联网, 微生物, 信息交流

中图分类号: Q93 文献标识码: A 文章编号: 0253-2654 (2002) 03-0098-05

INFORMATIONIZATION OF MICROBIAL RESOURCES

ZHANG He MO Xiang-Tao XIA Li-Qiu

(Department of Microbiology, College of Life Science, Hunan Normal University, Changsha 410081)

Abstract: Along with the development of computer techniques and the dissemination of Internet, many investigators of microorganisms already can acquire a lot of knowledge of many fields on microbe via Internet, extremely including the whole genome of a certain microbe. This was considered unimaginable in the past. Rapid collection of information also to a great extent expands the researching ranges and researching ability of microbial researcher, and at one time, the highly developed Internet provides a unprecedented opportunity for intercommunication of information, share of resources and international cooperations of microbiology.

Keywords: Internet, Microorganism, Intercommunication of information

在电脑技术突飞猛进的今天, 国际互联网已经渗入到政治、经济、教育等许多领域, 并且发挥出越来越重要的作用, 成为信息交流、资源共享和国际合作的重要形式, 生命科学领域更是如此。微生物学是一门综合性很强的学科, 作为现代分子生物学发展的三大支柱之一, 其在整个生物学发展中作用是不可磨灭的。随着现代生物核酸测序技术的进步, 自全长 1.8Mb 的嗜血流感杆菌 (*Haemophilus influenzae*, Rd) 基因组序列于 1995 年发表以来, 已有 54 个模式生物的完整基因组已被测序完成, 其中包括 9 个古细菌、31 个原核真细菌、14 个真核生物的完整基因组或它们的完整染色体, 还有另外的 13 个古细菌、177 个真细菌及 127 个真核生物基因组正在测序当中 (数据引自 <http://wit.integratedgenomics.com/GOLD/>), 而且这些数据都归入了相关数据库并能为全球科研工作者共享, 这就为人类更深刻的研究微生物提供了巨大的信息资源^[1]。

1 形式众多的生物学数据库

数据库最初是由美国国家科学基金会 (NSF)、国家航空航天局 (NASA) 和高级研

* 通讯联系人, E-mail: Xialiqu@263.net

收稿日期: 2001-04-12, 修回日期: 2001-12-30

究规划局 (ARPA) 共同规划, 共同收集、存储和组织信息, 并且通过搜寻、检索使这些资料成为随时可以得到的信息资源。目前, 世界上最著名的核酸一级数据库有 3 个, 即由美国国立生物技术信息中心 (NCBI) 建立和维护的 GeneBank、由欧洲生物信息学研究所 (EBI) 维护的 EMBL 核酸序列数据库和日本的 DDBJ 数据库^[2]。另外伴随着人类基因组计划 (HGP) 的实施, 产生了为此计划保存和处理基因组图谱数据的基因组数据库 (GDB)。蛋白质数据库主要包括 PSD (PIR 国际蛋白质序列数据库)、SWISS-PROT (蛋白质序列注释数据库)、PROSITE (蛋白质位点和序列模式数据库)、PDB (蛋白质空间结构数据库)、SCOP (蛋白质结构分类数据库)、和 COG (蛋白质直系同源簇数据库) 等。相关功能数据库主要包含 KEGG (京都基因和基因组百科全书)、DIP (相互作用的蛋白质数据库)、ASDB (可变剪接数据库)、TRRD (转录调控区数据库)、TRANSFAC (转录因子及其结合位点数据库)^[3]。还有一些数据库是将多个数据库整合在一起以便于提供服务, 例如 EBI 的 SRS (sequence retrieval system) 包含了核酸序列库、蛋白质序列库、三维结构库、基因组等 30 多个数据库, 使功能更为强大^[4]。近年来, 由于 DNA 测序技术的成熟, 人类获得的基因组信息在成指数增长, 例如 GeneBank 中储存的碱基数目大约每 14 个月就增加一倍, 到 1999 年 12 月其数目已达 30 亿, 它们来自于 47000 种生物。数据库中 2000 年 4 月 DNA 碱基数目是 60 亿, 但到了 2001 年初, 这一数目已达 110 亿。各种生物的 EST 序列也已达 600 多万条, 其中人类的 EST 序列已超过 300 万条, 估计覆盖人类基因 90% 以上; UniGene 的数目约达 7 万个; 自 1999 年初单核苷酸多态性 (SNPs, Single Nucleotide Polymorphisms) 数据库出现以来, 到 2000 年 3 月 20 日 SNP 的总数仅仅 26569, 但到现在已超过 350 万^[5], 可见生命信息的积累速度是十分惊人的。同时这些数据库中还包含着数十种模式微生物的基因组, 当然也包含许多人类还不能免疫方法控制的病原菌的基因组, 随着传统疫苗的逐渐淘汰, 亚单位疫苗的兴起, 数据库对于新疫苗的产生将起到至关重要的作用。因为某一特定病原菌整个基因组方面的信息将会构成寻找亚单位疫苗三种实验策略的基础^[6]。因此这些经过整理的核酸和蛋白质数据的共享将为微生物研究者提供极大的便利, 仅仅通过网络就可以获得所需基因、蛋白质方面的数据。

2 虚拟图书馆

虚拟图书馆是在网络中组织好的一套多级链接。这些项目包括资料、软件、图象、数据库等。虚拟图书馆的目的在于能使使用者方便而迅速地查阅许多专业方向的重要信息, 比如您可以在许多虚拟图书馆中了解世界上各著名机构组织的研究侧重点和工作进展; 您可以同各专业方向的研究人员进行学术交流; 您可以通过电子出版物掌握最新的学术动态和重要成果。所以对于一个微生物学研究者, 虚拟图书馆将能为您的科研提供良好的信息服务, 现在的互联网上存在着许多虚拟图书馆, 例如清华大学虚拟图书馆 (<http://www.lib.tsinghua.edu.cn/chinese/virtual/>)、The virtual library: biosciences (<http://www.vlib.org/Biosciences.html>) 等等, 有关微生物方面的一些虚拟图书馆见表 1。

3 电子刊物

由于分子生物学的飞速发展, 科学研究所产生的文献的积累速度也越来越快。例

表1 微生物相关虚拟图书馆

菌种及数据库名称	国家	虚拟图书馆网址
<i>Bacillus subtilis</i>	瑞典	http://anka.livstek.lth.se:2080/Bsubtil.htm
<i>Escherichia coli</i>	美国	http://web.bham.ac.uk/bcm4gt6/res.html
Yeasts	美国	http://genome-www.stanford.edu/Saccharomyces/VL-yeast.html
<i>Filamentous fungi</i>	美国	http://biodiversity.uno.edu/~fungi/fteach.html#BFF
<i>Haemophilus influenzae</i>	美国	http://www.tigr.org/tdb/mdb/hidb/hidb.html
Actinomyces	美国	http://molbio.cbs.umn.edu/asirc/
Protozoa	美国	http://www.aan18.dial.pipex.com/parasite.htm
<i>Mycoplasma genitalium</i>	美国	http://www.tigr.org/tdb/mdb/mgdb/mgdb.html
<i>Saccharomyces erevisiae</i>	美国	http://genome-www.stanford.edu/Saccharomyces/
www virtual library: Microbiology & Virology	美国	http://microbiol.org/vlmicro/vleduc.htm

如分子生物学和遗传学的文献积累从60年代中期的接近10万篇迅速增长至60年代末期的20多万篇,即在3~4年间,翻了一番。此后,至80年代中期,上升至约30万篇,即平均每年增长6~7千篇。至90年代中,文献数已上升至40多万篇;即在10年中,平均每年增长1万篇。到2000年,则增长至约50万篇,即在约5年间,又增长了10万篇(根据<http://www.ncbi.nlm.nih.gov>有关PubMed数据整理)。如此众多的文献资源,如果没有先进的检索系统,没有实时服务系统,要在其中搜索相关内容的文献将意味着巨大的工作量。令人兴奋的是现代计算机技术的发展与互联网的普及已经使快速文献查找成为现实。除了文献外还有大量与微生物信息资源紧密相关的电子图书、电子期刊等。同时还产生了一系列在线期刊服务网站,通过互联网为用户提供文献检索或论文索取服务,使现在的科研工作者能快速、高效的获取所需信息。

现在,世界上许多生物类杂志都有了电子版本,包括一些有权威性的学术刊物,比如 Science (<http://science-mag.aaa.org/science>), Nature (<http://www.nature.com>), Nucleic Acids Research (<http://www.oup.co.uk/nar/>)等。同时,国内的许多杂志也都推出了电子版本,如万方数据资源系统收集了大量期刊的电子版本(<http://www.periodicals.com.cn/qikan/>),其中包括了微生物学报(<http://www.periodicals.com.cn/gyjs.asp?ID=228051>)、微生物学通报(<http://www.periodicals.com.cn/gyjs.asp?ID=228052>)等许多微生物相关杂志的电子版本。另外用户还可以在 Electronic Journals VL (<http://www.edoc.com/ejournal>)和 NewJour (<http://gort.ucsd.edu/newjour/>)上找到许多有用的网址和简讯。

4 微生物基因组测序计划在线

由于科学研究的需要,越来越多微生物的基因组测序工作已经开始实施,尤其是英国的 Sanger 中心。近年来,该中心在 Beowulf Genomics 等的资助下,对许多微生物基因组进行测序,为了便于将最新的测序结果公布出来,建立了一系列相关微生物基因组测序进展网站,以方便广大微生物研究人员参考。表2仅列出一部分测序的菌种名称及相关网址,更多的链接请见 Sanger 中心的主页(<http://www.sanger.ac.uk/>)。

5 微生物菌种分类数据库

虽然微生物学的发展只有短短的100多年,但是发现的微生物却是层出不穷。而且,随着信息技术的不断发展,国内和国外已经形成了一系列微生物分类数据库,用

表2 相关微生物基因组测序在线信息

测序菌种名称	相关网址
<i>Aspergillus fumigatus</i>	http://www.im.ac.cn/database/catalogsc.html
<i>S. cerevisiae</i>	http://www.sanger.ac.uk/Projects/S_cerevisiae/
<i>Bordetella bronchiseptica</i>	http://www.sanger.ac.uk/Projects/B_bronchiseptica/
<i>Bordetella parapertussis</i>	http://www.sanger.ac.uk/Projects/B_parapertussis/
<i>Burkholderia pseudomallei</i>	http://www.sanger.ac.uk/Projects/B_pseudomallei/
<i>Mycobacterium tuberculosis</i>	http://www.sanger.ac.uk/Projects/M_tuberculosis/
<i>Neisseria meningitidis</i>	http://www.sanger.ac.uk/Projects/N_meningitidis/
<i>Dictyostelium discoideum</i>	http://www.sanger.ac.uk/Projects/D_discoideum/
<i>Plasmodium falciparum</i> chromosome 5	http://www.sanger.ac.uk/Projects/P_falciparum/

户可以根据菌种的部分特征搜索所要查找的微生物,并能通过网络获得相关微生物的许多在线信息,比如生化鉴定指标、分类依据以及培养条件等等,这些都为微生物研究者提供了巨大的信息资源,大大促进了微生物科学事业的蓬勃发展。有关微生物分类数据库见表3。

表3 微生物相关的分类数据库

数据库名称	数据库网址
中国微生物菌种目录	http://www.im.ac.cn/database/catalogsc.html
细菌名称数据库	http://www1.im.ac.cn/bacteria/bacteria.htm
革兰氏阴性杆菌编码鉴定数据库	http://www.im.ac.cn/database/gnb/gnb.shtml
微生物物种编目数据库	http://www1.im.ac.cn/species/speciesnew.htm
中国经济真菌数据库	http://www.im.ac.cn/ecofungi/fungimenu.shtml
国际计算机用微生物性状编码系统	http://www.im.ac.cn/RKC.html
国际微生物菌种数据网络 MSDN	http://www.im.ac.cn/msdn.shtml
法国细菌名称数据库	http://srs.im.ac.cn/srs6bin/cgi-bin/wgetz? -page + LibInfo + -lib + LBSN-Bacterial-Names
西藏大型经济真菌	http://www1.im.ac.cn/xzzj/fungi-xz.htm

6 展望

随着计算机技术和互联网技术的飞速发展,资源的信息化以及信息的网络化使微生物科研工作者足不出户就能获取大量的有用信息。特别是近几年生物信息学的高速进步与发展,由此产生了一系列的服务高度计算机化和网络化的数据库,并且许多数据库也已经实现了一体化,使数据库的功能更为强大^[7]。同时还建立了服务器客户模式,用户可以直接从数据库下载信息,也可以直接上传数据到主服务器以便他人也能共享自己的数据,因此资源的信息化程度以及对资源信息的利用程度已经成为衡量生物学发展水平的一个重要标准。当然人类利用的信息资源越多,所产生的成果就越多,现代不断信息化的微生物资源将能更快更好的为科技进步和社会经济发展做出贡献。

参考文献

- [1] Krause A, Vingron M. *Bioinformatics*, 1998, 14: 430-438.
- [2] Baxevanis A D. *Methods Biochem Anal*, 1998, 39: 1-15.
- [3] Borsani G, Ballbio A, Barfi S. *Hum Mol Genet*, 1998, 7: 1641-1648.

(下转 107 页)

(上接 101 页)

[4] Stoesser G, Moseley M A, Sleep J, *et al.* *Nucleic Acids Res*, 1998, 26: 8 ~ 15.

[5] Baker P G, Brass A. *Curr Opin Biotechnol*, 1998, 9: 54 ~ 58.

[6] Guido G. *Trends in Biotechnology*, 2001, 5: 181 ~ 188.

[7] 欧阳曙光, 贺福初. *科学通报*, 1999, 44: 1457 ~ 1468.