

建议尽快组建国家级生物医学信息中心

郝柏林*

(理论物理研究所 北京 100080)

关键词 生物医学信息中心

从细菌到人类,众多物种的基因和蛋白质数据正以科学史上从未有过的高速度增长。目前已测定出 20 多种细菌以及比细菌更高等的一些物种如酵母菌和线虫的完全基因组,人类基因组,即一个典型的“人”的全部基因,也将提前在 2003 年完全测定。现在,基因数据的总量每 12 个月翻一番,每个月至少测出 160 种蛋白质的结构,增长速度本身也在加速。

这种情况不仅反映了科学知识的深化和研究方式的转变,而且必将在短短几年内影响到生物、医学、农业乃至军事等众多领域。生物学不再是恩格斯所说“数学等于零”的学科,也不再是仅仅基于观察和实验的科学。理论和计算将发挥日益巨大的作用,数学、物理、计算机科学将越来越多地把生物问题作为当然的研究课题。事实上,如果没有跨学科的发展,仅仅靠生物工作者不可能充分利用如此迅猛增长的海量数据。人本身当然是研究的核心。没有两个人的基因组完全相同。人类基因组计划的完成,只是更为细致的人群乃至个体的正常和病理的基因及其表达的研究的出发点。预计 10 年内,如何利用生物信息库和生物计算手段,将成为广大临床医师和农林畜牧工作者基本训练的一部分。生物信息对未来军事和国防的影响也不容忽视。

欧美发达国家在生物信息方面早有积累。手工搜集的蛋白质数据库早在 60 年代就在美国开始建立。美国洛斯阿拉莫斯国家实验室 1979 年开始建立的基因库 GenBank,现在由国家生物信息中心(NCBI,1988 年成立)管理维护。欧洲分子生物学实验室 1982 年开始服务的 EMBL 数据库和随后建立的欧洲生物网(EMBLnet),1994 年改由当年建在英国剑桥的欧洲生物信息研究所(EBI)管理。日本 1984 年着手建立国家级的核酸数据库 DDBJ,1987 年正式服务。以上三家共同组成了 DDBJ/EMBL/GenBank 国际核酸序列数据库,每天交换数据,同步更新,目前绝大部分核酸和蛋白质数据由这三家产生。其它国家如德国、法国、意大利、丹麦、以色列等,在分享网络资源的同时,也纷纷建立自己的生物信息中心,为本国服务。即使如此,美国目前仍面临有关研究跟不上需求、缺乏人才的局面。1999 年 6 月 3 日,美国国家卫生署(NIH)的一个专家委员会建议,要在 2001 年在大学和研究机构中建立 5—20 个生物计算中心,每年给予每个中心可达 800 万美元的支持,以从事研究和培养人才。

应当看到,随着生物信息容量、成本和重要性的上升,免费使用数据的情况已经开始改变。近两年,瑞士蛋白质数据库、德国转录因子数据库等都已对商业性用户收取费用,但对学术性

* 中国科学院院士
收稿日期:1999 年 12 月 1 日

用户仍继续免费。将来某些数据库征收费用,限制读取,甚至对我关闭的情形也可能出现。中国作为日益强大的世界大国,在对数据贡献甚微的条件下,长期无偿享用成果的状况很容易出问题。我们必须从战略高度,及早采取对策。

近年来,国内生物信息学开始“炒”热。规模甚小、实力不足的生物信息研究所或“中心”相继上马,全国范围重复引进和建设,甚至有同一所学校从不同渠道由国外引入两套同名数据库的情况,而大家都没有力量达到国家规模和接近国际水平。这种在其它领域曾多次出现、令人痛心疾首的局面已经开始呈现。必须从国家全局着眼,打破部门壁垒,采取紧急措施。为此,我建议:

(1)由科技部、教育部、中国科学院和国家自然科学基金委员会共同支持建立一个国家级的“中国生物医学信息中心”(以下简称“中心”),免费为全国研究和教学单位服务。

(2)“中心”的首要任务,是引进国际上一切有科学价值的生物医学数据库和免费共享软件,同时把我国在生物信息方面有特色的贡献提供给国际科学界。

(3)“中心”的第二项任务,是不断研制适合我国用户的“界面”和“接口”。常年分期分批培训各层次、各领域的教学科研人员和科技管理干部,使数据和软件充分发挥作用。

(4)在完成上述任务的前提下,“中心”应开展数据库管理、算法和计算生物学方面的研究,组织生物信息学和计算生物学方面的国内外学术活动,不断提高学术水平和国际地位。

(5)对于各单位已引入和建立的镜像,不必用行政命令去集中和归并,只要“中心”设置通向它们的链接,要它们为全国服务即可。(凡用国家经费引入的数据库,就必须为全国服务,否则国家应削减其经费。)将来“中心”提供高效率、高水平服务之时,不少单位会自动把“小本经营”的数据库转给“中心”管理,以节省人力物力。

(6)就目前条件看,“中心”应建立在北京,随即在上海、天津、南京、西安等地依托学校或研究所,设立以镜像点为主的分中心。

迅速见效而非“新起炉灶”的办法,就是集中支持目前国内最有条件的一个单位。就我的了解,北京大学生命科学院的生物信息中心(网址 <http://www.cbi.pku.edu.cn>),应是首选单位,它在1996年欧洲EMBNET扩大到欧洲之外时已正式成为中国结点(每个国家只有一个结点),目前已有60多种生物数据库的经常更新的镜像点。三年来,它已组织过多次国内和地区的培训班及会议,有着较广泛的国际联系。然而,软硬件资源的不足,限制着它发挥更大的作用。一旦成为“中心”,北京地区的高等院校、医学和军事医学以及中国科学院的有关专家可以实质性(非挂名)兼职方式,参与建设。

作为第一次专门资助,1 000万元人民币的投入就可以使这个生物信息中心的服务大为改观,其后的运行费用和各地分中心的建设则可从容规划。这些“中心”在初期会以生物信息为主,以后逐步发展到医学信息服务。北京大学和北京医科大学的合并,对于“中心”在医学信息方面的建设会起到积极作用。

应彻底改革国内教育科研单位计算机网络的收费方式和价格。国内目前几大网络通往国外相对顺畅,而彼此连接却速度甚慢,形成网络建设因管理问题而不能充分发挥作用的不合理局面,这是科技部和信息产业部应协商解决的大问题,科技和教育界已多次反映过意见,一直未见解决,这也是影响生物信息网发挥作用的重要因素。