

## 建议尽快组建国家级的生物医学信息中心

郝柏林

1999年6月10日

从细菌到人类，众多物种的基因和蛋白质数据正在以科学史上从未有过的高速度增长。目前已测定出二十多种细菌，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14个月翻一番，每个月至少测出150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，在短短几年内必将影响生物、医学、农业乃至军事的众多领域。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，即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。

发达如美国，目前也面临有关研究跟不上需求，人才严重缺乏的局面。1999年6月3日美国国家卫生署(NIH)的一个专家委员会建议，迅速在大学和研究机构中建立20个生物计算中心，给予每个中心每年可达8百万美元的支持，从事研究和培养人才。这一建议可能从2001年开始实施。

然而，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年开始的基因库GenBank，现在由1988年成立的国家生物

信息中心 (NCBI) 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 (EMBNNet), 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。目前绝大部分核酸和蛋白质数据由美国、欧洲和日本三家产生。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。其它国家如德国、法国、意大利、澳大利亚、丹麦、以色列等, 在分享网络资源同时, 还纷纷建立自己的生物信息中心, 为本国服务。同时应当指出, 随着生物信息容量、成本和重要性上升, 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, 长期无赏享用成果的状况很容易出问题。我们必须从战略高度, 早作对策。

中国在生物信息方面与国际差距虽大, 然而近来国内生物信息学又开始“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, 同一所校园从不同渠道由国外引入两套同名数据库的情况也已经出现。培养出的少数青年人才迅即流失。全国范围重复引进和建设, 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。现在必须从国家全局着眼, 打破部门壁垒, 采取紧急措施。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共同支持建立一个国家级的、免费为全国研究和教学单位服务的“中国生物医学信息中心”(以下简称“中心”)。
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. 在完成以上两项任务的前提下, “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，以不断提高自己的学术水平和国际地位。

5. 至于各单位已经通过各种渠道引入和建立的镜像点，根本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。(凡是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”大量提供高效率、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，以节省自己的人力物力。
6. 就目前条件看，“中心”应建立在北京，随即应在上海、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，就是集中支持国内最有条件的一个单位。就我的观察了解，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。它们在1996年欧洲EMBNET扩大到欧洲之外时已经正式成为中国结点(每个国家只有一个结点)，目前已经有五十多种生物数据库的经常更新的镜像点。三年来它们已经组织过多次国内和地区的培训班和会议，有着较广泛的国际联系。然而，软硬件资源的不足，限制着他们发挥更大作用。一旦建立，北京地区高等院校，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，1000万元人民币的投入就可以使这个生物信息中心为全国的服务立即改观。以后的经常费用和各地分中心的建设可从容规划。这些“中心”在初期会以生物为主，随后逐步发展医学信息服务。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。

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，解决目前几大网络相对顺畅通往国外，彼此连接速度甚慢，国家花在网络建设上的大量投资因为管理问题而不能尽量发挥作用的不合理局面，是科技部和邮电部应协商解决的大问题，科技和教育界已经反映过多次意见，一直得不到解决。这也是影响生物信息网发挥作用的重要因素，这里不再复述。