

人类基因组计划大事记

刘 威

(中国科学院生物物理研究所, 北京 100101)

今年2月中旬,《Nature》与《Science》分别发表了人类基因组工作框架图 (the draft genome), 这是人类基因组计划 (Human Genome Project, HGP) 实施以来所取得的最重大进展, 也是生命科学领域中的一个里程碑. 自人类基因组计划提出以来, 到现在一共 15 年. 这 15 年来科学家们所走过的是一条充满艰辛的路, 同时也是一条谱写辉煌的路. 回顾这段将在人类历史上永载史册的光辉历程, 相信对每一位科学殿堂里的探索者都会大有启迪. 为此, 参考有关资料, 我们将这段时期围绕基因组测序所发生的重大事件整理纪录如下.

1985 年: 5 月, R. Sinsheimer 在加州大学 Santa Cruz 分校主持会议, 讨论将人类基因组全部测序的可行性; 12 月, 西特斯公司 (Cetus Corp.) 的 K. Mullis 和他的同事们发明了 PCR 技术, 利用这项技术, 科学家们能够在短时间内精确复制成百万的 DNA 片段, 从而不再为遗传物质的用量而担忧.

1986 年: 3 月, 美国能源部 (Department of Energy, DOE) 在新墨西哥州召开会议, 讨论人类基因组测序计划; 6 月, 科学家们齐聚纽约州冷泉港实验室, 以“人类的分子生物学”为题讨论人类基因组计划可能带来的利益和前景; 同月, 加州理工学院的 L. Hood 和 L. Smith 共同发明了第一台 DNA 自动测序仪; 9 月, DOE 从 1987 年的财政预算中拨款 530 万美元, 资助 C. DeLisi 开始基因组研究.

1987 年: 2 月, W. Gilbert 从美国国家研究委员会 (National Research Council, NRC) 辞职, 组建 Genome 公司, 同时宣称该公司将参与人类基因组测序工作, 并将为测序结果申请专利; 4 月, 一个专家小组建议 DOE 在未来 7 年内拨款 10 亿美元, 用以进行人类基因组图谱测定和基因组测序, DOE 主持的人类基因组计划开始实施; 5 月, 华盛顿大学的 D. Burke, M. Olson 和 G. Carle 发明酵母人工染色体 (YAC) 克隆技术, 使插入片段的长度扩大了 10 倍; 10 月, H. Donis-Keller 发表了第一张带有 403 个遗传标记的遗传图谱, 由此在全世界范围内引发了测序竞争; 同月, 杜邦公司将荧光链终止双脱氧技术引入到 DNA 快速测序中; 这一年, Applied Biosystems 公司将第一台自动测序仪推上市场.

1988 年: 2 月, 在一份至关重要的报告中, NRC 批准了人类基因组计划, 并提出分阶段实施和每年 2 亿美元的拨款申请; 3 月, 在许多专家的提议下, J. Wyngaarden 在弗吉尼亚州 Reston 召开的会议上宣布, 美国国立卫生研究院 (National Institutes of Health, NIH) 应代替 DOE 成为人

类基因组计划中的主要负责机构. 在这之后, Wyngaarden 被任命为 NIH 主席; 6 月, 第一届基因组年会在冷泉港召开; 9 月, NIH 宣布成立人类基因组研究办公室, 并任命发现 DNA 双螺旋结构的 J. Watson 为主任; 10 月, NIH 与 DOE 达成谅解备忘录, 双方同意在人类基因组计划中团结协作.

1989 年: 1 月, 洛克菲勒大学的 N. Zinder 主持了第一次 HGP 顾问委员会会议; 9 月, Olson, Hood, Botstein 和 Cantor 共同描绘了基因图谱测定的新战略; 同月, DOE 与 NIH 共同组建了一个委员会, 以解决基因组计划可能带来的伦理、法律和社会问题; 10 月, NIH 的人类基因组研究办公室升级为人类基因组研究国家中心 (National Center of Human Genome Research, NCHGR), 并被授予拨款权.

1990 年: L. Smith, B. Karger 和 N. Dovichi 分别领导的 3 个小组各自独立地发展了毛细管电泳技术, 这项技术后来被广泛应用于大规模测序中; 4 月, NIH 和 DOE 一起发表了一个五年计划, 内容包括测定完整的遗传图谱和物理图谱 (每 100 kb 含一个标记), 并在 2005 年之前完成模式生物的基因组测序; 8 月, NIH 决定对四种模式生物开始大规模测序尝试. 这四种模式生物是: 支原体 (*Mycoplasma capricolum*)、大肠杆菌 (*Escherichia coli*)、线虫 (*Caenorhabditis elegans*) 和酿酒酵母 (*Saccharomyces cerevisiae*); 10 月, NIH 和 DOE 宣布这一年的 10 月 1 日为人类基因组计划的正式启动时间. 同月, 生物技术信息中心 (National Center of Biotechnology Information, NCBI) 的 D. Lipman 与 E. Myers 发表了用于进行同源序列比较的 BLAST 算法.

1991 年: 6 月, NIH 的生物学家 J. C. Venter 发表了一种发现表达基因的新战略, 即使用表达序列标签 (Expressed Sequence Tag, EST); 一个月之后, Venter 在国会听证会上披露, NIH 堆积了上千份表达基因的专利申请, 由此引发了一场关于基因能否获得专利的争论; 10 月, 日本开始对水稻基因组测序; 12 月, 用于基因查找的第一个程序 GRAIL 开发成功.

1992 年: 4 月, 就基因申请专利的问题, Watson 在与 B. Healy 争论之后辞去了 NCHGR 主任的职位, 后者随后被任命为 NIH 主席; 6 月, Venter 离开 NIH 组建了基因组研究所 (The Institute of Genome Research, TIGR), 这是一个位于马里兰州 Rockville 的非盈利机构; 7 月, 英国慈善机构 Wellcome Trust 投资 9 500 万美元, 加入人类基因组计划; 9 月, 加州理工学院的 M. Simon 和他的同事们发明细

菌人工染色体 (BAC) 克隆的新技术, 使大规模克隆成为可能, BAC 后来成为基因组研究中的“硬通货”; 10 月, 美国法国的两支研究队伍分别完成了人类 Y 染色体和第 21 条染色体的第一张物理图谱; 12 月, 经过长时间讨论, NIH 和 DOE 为鼓励资源共享放宽了对数据资源的限制, 规定研究人员必须在 6 个月之内将新的数据传送到公共数据库中; 同月, 美国法国的上述两个研究机构又分别完成了鼠和人的遗传图谱。

1993 年: 4 月, 密歇根大学的 F. Collins 被任命为 NCHGR 的领导人; 10 月, NIH 和 DOE 发布了新的五年计划, 按照这个计划, 到 1998 年底, 将有 80 Mb 的 DNA 被测序, 而人类基因组将在 2005 年被全部测序。与此同时, 位于英国剑桥的 Sanger 中心在 J. Sulston 的领导下加入人类基因组计划, 并成为一个重要的测序中心。

1994 年: 9 月, 爱荷华大学的 J. Murray 与法国 Génethon 的 Cohen 及他们的同事们完成了人类基因组中的第一个完整遗传连接图谱。

1995 年: 5~8 月, 测序染色新技术和热稳定聚合酶开发成功; 7 月, 第一个基因组——流感嗜血杆菌 (*Haemophilus influenzae*) 的全序列发表, 大小为 1.8 Mb; 9 月, 日本政府拨款 1 590 万美元, 用以资助 3 个研究机构从事基因组测序工作; 10 月, 斯坦福大学的 P. Brown 和他的同事们发表了第一篇关于完整 cDNA 探针芯片的论文。

1996 年: 2 月, 各国科学家聚集在百慕大群岛召开会议, 讨论并通过了数据资源共享的问题, 规定此后的测序结果必须在 24 小时内传送到公共数据库中, 这就是著名的百慕大原则; 4 月, NIH 资助 6 个研究小组尝试人类基因组的大规模测序; 同月, Affymetrix 将 DNA 芯片商品化; 10 月, 酿酒酵母 (*Saccharomyces cerevisiae*) 的基因组全部测序完成; 11 月, RIKEN 的 Y. Hayashizaki 小组完成了第一套小鼠全长 cDNA 的测序。

1997 年: 1 月, NCHGR 升级为美国国家基因组研究所 (National Human Genome Research Institute, NHGRI), DOE 也成立了相应的联合基因组研究所 (Joint Genome Institute); 9 月, 大肠杆菌 (*Escherichia coli*) 基因组全部测序完成, 全长 5 Mb; 同月, 毛细管测序仪出现。

1998 年: 1 月, NIH 公布了一项新计划, 其目的是在浩如烟海的 DNA 碱基序列中寻找单核苷酸多态性 (single nucleotide polymorphisms, SNPs); 5 月, PE Biosystems 公司开发成功 PE Prism 3700 毛细管测序仪; 同月, Venter 宣布

成立一个新公司, 取名 Celera, 并宣称将在 3 年内, 投资 3 亿美元, 采用“全基因组鸟枪法” (whole genome shotgun) 完成人类基因组的全部测序, 同时 Venter 宣布 Celera 公司的数据将不遵从百慕大原则。与此同时, Wellcome Trust 将对人类基因组计划的投资加倍, 达到了 3 亿 3 千万美元, 以此作为对 Celera 公司的回应。从此, 人类基因组测序在“公” (HGP) 与“私” (Celera) 之间展开了激烈竞争; 10 月, NIH 和 DOE 宣布将在 2001 年完成人类基因组工作框架图, 并将全部测序的最后期限从 2005 年提前到 2003 年; 12 月, 线虫 (*Caenorhabditis elegans*) 基因组测序完成。

1999 年: 3 月, 作为对 Celera 公司的应战, NIH 再次宣布将人类基因组工作框架图的完成时间提前到 2000 年春季, 大规模的测序工作主要集中在以下五个测序中心: 麻萨诸塞州的 Whitehead 生物医学研究所、位于圣路易斯的华盛顿大学、位于休斯敦的 Baylor 医学院、英国剑桥的 Sanger 中心和位于加利福尼亚州 DOE 下属的联合基因组研究所; 4 月, 10 家公司与 Wellcome Trust 签署协议, 合作开展 SNP 研究, 并计划将公共数据库中的 SNP 数据每季度更新一次; 9 月, NIH 宣布将在 3 年内投资 1 亿 3 千万美元, 完成小鼠基因组的测序; 12 月, 英国、日本和美国的科学家们共同完成了第一条人染色体——第 22 条染色体的全部测序工作。

2000 年: 3 月, Celera 公司完成果蝇 (*Drosophila melanogaster*) 基因组 (180 Mb) 的全部测序工作, 在当时所有已测序的基因组中是最大的, 同时这也证明了“全基因组鸟枪法”的可行性; 3 月, 由于对数据资源共享的政策持不同观点, HGP 与 Celera 公司的合作计划破产; 5 月, 德国和日本的科学家公布了人类第 21 条染色体的测序结果; 6 月 26 日, 这是一个值得纪念的日子, HGP 与 Celera 公司的领导人在白宫的庆祝仪式上共同宣布了人类基因组工作框架图的完成, 并约定将双方的研究成果同时发表; 12 月, 第一个植物基因组——拟南芥 (*Arabidopsis thaliana*) 基因组被全部测序, 大小为 125 Mb。

2001 年: 2 月, HGP 将自己测定的人类基因组工作框架图发表在《Nature》(2001, 409(6822)) 上; Celera 公司则将它们的结果发表在《Science》(2001, 291(5507)) 上。发表的“工作框架图”已能覆盖人类基因组的 97%, 其中至少 92% 的序列已组装得准确无误, 并显示其中只有 3 万到 4 万个编码蛋白质的基因, 这是人类基因组研究的一个重要里程碑。