

美国能源部资助的微生物基因组研究概况*

艾云灿

(中山大学生命科学院 科技部海洋生物功能基因组学研究开放实验室 广州 510275)

A SURVEY OF MICROBIAL GENOMICS PROJECTS SUPPORTED BY THE US DEPARTMENT OF ENERGY

Ai Yuncan

(*The Open Lab for Functional Genomics of Marine Organisms, Ministry of Science and Technology of China,
School of Life Sciences, Zhongshan University, Guangzhou 510275, China*)

关键词: 微生物, 基因组计划, 项目遴选, 项目管理

中图分类号: Q93 文献标识码: A 文章编号: 0001-6209 (2001) 06-0757-05

自 1995 年 12 月完成第一个微生物(流感嗜血杆菌)全基因组测序以来^[1], 微生物基因组学策略和技术提高很快, 促进了其他生命形式(包括人类)的基因组学发展, 开创了生物基因组计划新时代^[2]。短短几年内, 互联网数据库中各种生物基因组数据迅速增长。1998 年 2 月 20 日美国 Science 杂志评选的当年世界十大科技进展, 将微生物基因组图谱的构建列为其中第六项, 同年 3 月 27 日 Science 杂志编者指出: 生物基因组革命可以同工业革命和计算机革命所带来的变化相比, 是“第三次技术革命”^[3]。微生物全基因组测序, 不仅是人类开始得最早和首先完成的第一种生物的全基因组分析, 也是迄今为止完成测序基因组种类最多的领域。美国国家能源部(DOE)和国立卫生研究院(NIH)是美国的微生物基因组计划(MGP)的启动和发展者, 两个部门所关注的分别是资源微生物和病原微生物, 其研究发展策略对全球 MGP 研究发展趋势起着决定性的影响作用。本文谨重点介绍 DOE 的 MGP 项目发展简况和资助思想。简言之, DOE 认为 MGP 不仅是研究和应用极端环境微生物特定功能的科学基础, 而且将是 21 世纪微生物学的信息基础^[2]。

1 科学思维方式革命

微生物基因组是有关微生物的完整知识, 使微生物学科不断丰富发展, 并正在深刻影响着生命科学门类的所有领域, 影响到关于基因组信息与生物学特性及行为之间逻辑关系的新思维。长期以来一部很成功的“还原学派”(reductionist)科学历史, 就是如何探索采取更精细方式来拆分细胞和系统, 使之成为越来越基本的组件, 我们由此已经接近认识的极限。现在, 应该考虑怎样以一种更具有挑战性的方式, 把这些“部件”组装起来成为一个整体, 以便深化基础研究和更广泛的应用领域。一种生物的完整基因组就是该生物的最大“组件列表”, 完成基因组测序获得大量组件之后, 我们就有基础反过来思考怎样“组装”它们。目前 DOE 资助的 MGP 已经完成 30 多个微生物基因组测序, 更多重要微生物基因组测序也将在 10 年内迅速完成^[2]。MGP 所涉及的这些微生物在系统发生学上具有代表性, 科学家能够通过一一对应地比较它们的基础序列变异程度, 研究它们的进化关系。对这些信息的贮存、归档、共享、重验

* 国家自然科学基金(39870579)、教育部回国留学人员科研启动基金、高等学校骨干教师资助计划、广东省自然科学基金资助项目(001271)

作者简介: 艾云灿(1963-), 男, 湖北江陵人, 中山大学生命科学院副教授, 博士。研究方向是功能基因组学、分子生理与分子药理学、微生物分子生物学与生物技术。

收稿日期: 2000-10-16, 修回日期: 2000-12-01

证,都需要依赖一系列数据库和通讯技术以及国际合作。MGP 对于科学观和方法论的最大影响,就是它正在改变传统生物科学研究模式,即那种依赖于孤立地观察行为和遗传,来对生物的遗传控制进行还原性思考的模式^[2]。

2 资源微生物学价值观

微生物构成 60% 以上的地球生物量。它们在地球上生存和进化了 37 亿年,能够在任何极端环境(包括极端热、冷、辐射、压力、盐、酸、黑暗等)条件下存在。以往研究依赖于可培养性,而绝大多数微生物在实验室培养困难,所以仅仅只有 1% 的微生物种类被认识,绝大多数都未被重视研究。微生物对环境的适应范围和多样性变化,意味着它们有能力解决那些令我们百思不得其解的大量问题。在基因组水平上研究微生物多样性群体,可能为解决医学、农业、工程过程、能量生产和应用、环境和废物治理等问题提供新途径^[4,5]。科学家们正在转而探求微生物生命活动在全球气候变化过程中的作用,并期望由此更多地获得关于微生物影响气候变化,及在地球全球性总量代谢中的作用等方面的知识。

3 MGP 计划启动及发展方向拓展

MGP 由美国 DOE 在 1986 年开始启动,它与人类基因组计划(HGP)相伴而生,现在变成 DOE 与国家卫生研究院(NIH)的国家人类基因组研究所(NHGRI)合作计划的组成部分。自 1994 年起,DOE 项目转向对那些可能为 DOE 所用的非病原性微生物的完整基因组序列进行测定。这些微生物包括(i)环境或能源相关的;(ii)系统发生学相关的;(iii)潜在商业应用性的。许多微生物都被遴选出来作为候选菌。DOE 召集一系列小规模工作会议,邀请来自基因组学、微生物学和环境科学领域的专家们参加,请他们建议和提出优先候选对象用于微生物基因组测序。1999 年 11 月举办的主题论坛是“纤维素降解微生物”。纤维素是地球上数量最多的生物多聚物,是几乎所有植物性生命形式的重要组成部分,而绝大多数动物都没有必需的酶类来降解纤维素,它们通常依靠细菌的酶类来提供,例如牛瘤胃和白蚁后肠等。加强这个领域的研究,将促进 DOE 利用巨大生物量的研究开发步伐。近 6 年来许多微生物种类已被纳入 MGP,遴选标准也逐渐扩展为如下几点:(i)非人类病原菌;(ii)环境相关性,即能源生产、废物整治、生物兴趣、工业过程;(iii)基因组较小(特别是小于 10Mb);(iv)遗传背景好,便于探讨遗传工程对 DOE 的应用;(v)科学兴趣,如在进化树中系统发育学位置;可培养性,易于扩增和纯化基因组 DNA^[2]。

4 DOE 资助的 MGP 部分重点项目简介

4.1 由 TIGR 为主完成的微生物基因组测序

1995 年初始,就有四个项目经过严格遴选获准确定为 DOE MGP 首批资助计划。基因组研究所(The Institute of Genome Research, TIGR)于 1995 年底完成嗜血流感杆菌(*Haemophilus influenzae*)全基因组测序^[1]之后,又获得再次资助,同伊利诺斯大学共同获得三年期合作协议,完成詹氏甲烷球菌(*Methanococcus jannaschii*)^[6]及其他合适的微生物全基因组测序。*M. jannaschii* 是极端嗜热菌,只能在高温下生长,是从加利福尼亚一个海湾的 8000 英尺深的热泉底部分离到。它拥有小基因组(只有 1.66Mb),又是耐压菌(生活在高压下),产甲烷,最适生长温度为 90℃,严格厌氧和自养(存在于无氧环境中,需要二氧化碳或碳酸盐为唯一碳源,简单无机氮化物用于代谢合成)。令人惊异的是,当完成其全部基因测序时,发现 65% 的关键基因都与以前发现的其他任何基因没有同源性,引发了关于古细菌的测序及意义的研究^[6]。对这些微生物 DNA 测序,使得可能搜寻这 3 个嗜热菌的遗传相似性,并比较它们在系统发育树上不同分支地位,及 3 个细菌的产甲烷情况,还证实了“生命树”的确存在 3 个分支,这是伊利诺斯大学的 Carl Woese 20 多年前首先开创的假说^[7]。*M. jannaschii* 序列确证了它在“生命树”上的独立地位,也明确了“生命树”有 3 个而不是 5 个分支^[6,7]。TIGR 还完成了 *Mycoplasma genitalium* 全基因组测序,这种真细菌是自养细菌中基因组最小(580,000bp)的,因此成为阐释足够用于自养生存的最小基因组的一种模式

菌^[8]。由 TIGR 测序的 *Archaeoglobus fulgidis* 是从油井中分离到的,该菌在油井中添加硫,败坏原油品质,增加了商业油提炼困难^[9]。1999 年 TIGR 还测定了 *Deinococcus radiodurans* 全序列^[10],该菌具有不可思议的抗辐射能力,能够在 150 万里拉辐射下存活,即是 3000 倍于人类的致死辐射剂量。另外,*D. radiodurans* 还被遗传改造用于部分降解甲苯,用于转化二价汞(Hg^{2+})释放毒性较低形式。这些工作建立了两种可能性:遗传工程增强该菌的附加功能,且该宿主菌遭受辐射环境时还能够维持其附加功能。这种微生物惊人的 DNA 修复能力,通过遗传工程改造,在用于清除放射性废物方面很有潜力^[10]。

4.2 由 GTC 等为主完成的微生物基因组测序

DOE 还分别与基因组治疗公司(Genome Therapeutics Corporation, GTC)、犹他大学、重组生物催化剂有限公司(Recombinant BioCatalysis, Inc.)等签署合作协议,共同测定 *Methanobacterium thermoautotrophicum*^[11]、*Pyrococcus furiosus*^[12,13]、*Aquifex aeolicus* VF5^[14] 等全基因组序列。这些微生物都是极端高温古细菌,被认为是地球上最古老的生命。*M. thermoautotrophicum* 是从污水淤泥中分离到的,生长适宜温度为 65℃,基因组大小约为 1.7Mb。对这个厌氧自养古细菌的研究,获得了大部分关于生物转化 CO_2 为 CH_4 的生物化学知识,因此它成为研究甲烷形成的一个模式菌^[11]。*P. furiosus* 是 1980 年代中期发现的海洋异养极端嗜热菌,最适生长温度 100℃,具有 2Mb 基因组,是研究嗜热菌的模式种^[12,13]。*A. aeolicus* VF5 是海洋极端嗜热细菌,代表细菌链上分化程度最远的分支^[14]。GTC 还开展了 *Clostridium acetobutylicum* 的全基因组测序,该菌能制造和分泌溶剂丙酮。*C. acetobutylicum* 的故事已成为现代历史的一部分。在第一次世界大战中,丙酮因战事极度缺乏,正是利用这个微生物,Chaim Weizmann 在 1916 年供应丙酮(那时关键的军事原料)给英国,替代因战争而稀缺的原料。基于这样的荣誉和贡献,英国政府征询 Weizmann 的意见:什么爵位或奖赏才能与之相配。Weizmann 对个人荣誉没有兴趣,而是要求英国政府能否帮助犹太人在巴勒斯坦建立自己的家园。结果在 1917 年的 Balfour 宣言里正式托付给英国政府筹建,这就是后来于 1948 年建立的以色列国,Weizmann 成为首任总统。

5 MGP 计划成果开发利用的考虑

5.1 拓展研究视野的深度和广度

采用现代技术完成较小的微生物基因组测序速度很快,最近 DOE 资助的 *Enterococcus faecium* 全基因组测序(大小为 2.8Mb)仅在一天内完成^[15]。DOE MGP 所有序列数据库都尽快地通过公共数据库向公众开放,这些数据增加了探索微生物多样性的新空间,赋予科学家研究地球生命起源和各种环境下细胞结构与功能的遗传基础的新视野。一些私立机构也开始涉足微生物基因组测序,但遗憾的是这些序列并不总是对公众开放的。为此,美国微生物学会 2000 年度咨文报告中提出了公私投资和技术合流风险利益均沾权利义务均衡的新型合作的具体策略性建议^[2]。随着微生物基因组被大量测定,分析和标识它们的 DNA 序列,识别所有潜在基因和潜在功能,也是 DOE MGP 的关键部分。

5.2 揭示新的独特基因

研究者们接连不断地发现,已测序的微生物基因组中有大约 35% 的可预测的新基因,都是以前没发现的。这些新基因的存在,预示着今后在基础研究和生物资源开发利用方面存在令人激动的机会。目前令人诧异的是,尽管所获得的基因组数量在逐渐增加,但是 35% 这个数字比例却并未下降。这个现象可能正好强调了人类关于微生物基因组和它们适应“技巧”的认识实在是太少了。

5.3 关注基因水平转移与进化

大量的微生物基因组序列中都出现了遗传组件——不是单个基因而是许多个基因的积木块结构——它们似乎是在微生物进化中从其他位于 Woese’s “生命树”上十分遥远分支的微生物中所获得的。一个典型例子是从 Sicily 火山岛分离的极端嗜热菌 *Thermotoga maritima*^[16],它似乎从另一种古细菌中获得了四分之一的基因组,而这些古细菌显然已经经历了数十亿年的分道进化。由此突出了被称作“水平基因转移”(lateral gene transfer)的现象^[16]。这些现象对于关于微生物进化的知识提出了新挑战。虽然目前我

们还不知道微生物世界里怎样启动,或在多少程度上发生着基因的水平交换,但是它们的“足迹”(footprints)都清晰地存在于已经完成测序的微生物基因组序列中。微生物基因是其他进化程度更高的生物(包括人)的许许多多基因的始主,许许多多的微生物基因后来分离成为其他生物基因,成为更加复杂的多细胞生物的必需新功能基因。序列相似性分析,常常导致研究者能够将某些人类或动物基因追踪到它们的微生物起源,以致于不得不重新构建进化途径假说。当然,这类重新构建往往取决于人们所关注的是哪些具体基因。

5.4 研究基因组的组织结构与调控

完成基因组测序后,可以进一步研究基因组的组织结构。全基因组信息可以帮助我们深入了解有关微生物响应环境的调控机制。反过来,这些知识对于应用微生物整治和清除废物等潜在应用也十分重要。在宿主免疫障碍情形下,很多微生物通常能够变成人类病原微生物(例如 *Pneumocystis carinii* 和 AIDS)。我们已经对那些容易了解的或是引起疾病的微生物有相当认识,但是对那些通常不会引起疾病的微生物的认识则相对较少,对它们的病原性转变和对环境依赖的选择特性还知之不多。最主要的障碍可能在于研究环境中这些微生物间的相互作用时,通常要涉及到无数种微生物的复杂区系,而对这种区系中的种间通讯,我们还一无所知。不断发掘新方法,研究基因组的组织结构与调控,在探讨上述疑难问题方面将能够大有作为。

6 MGP 发展的新重点

6.1 DOE 需要

DOE MGP 的主要注意力,仍将放在与 DOE 之需要相关的微生物基因组测序上。DOE 项目领域包括能源利用和生产、生物整治,特别是自然和加速生物整治研究(NABIR)项目,最近还启动了研究气候变化的新领域。DOE 强调大气 CO₂ 和其他温室气体所导致的全球气候变化。在这类新启动的项目中,DOE MGP 突出了能消费 CO₂ 或生产替代燃料资源(如甲烷或氢)的微生物的序列测定。DOE 希望从这些努力中发现微生物和酶是怎样调节这些重要气体的机理。根据 DOE 新的遴选标准和近期发展需求,一批新目标菌的全基因组测序已经立项,受资助单位的范围也大大扩展了。目前,这些正在进行中的项目包括 *Carboxydotherrnus hydrogeniformans*、*Caulobacter crescentus*、*Chlorobium tepidum*、*Clostridium acetobutylicum*、*Dehalococcoides ethenogenes*、*Desulfovibrio vulgaris*、*Geobacter sulfurreducens*、*Methylococcus capsulatus*、*Nitrosomonas europaea*、*Nostoc punctiforme*、*Prochlorococcus marinus*、*Pyrococcus furiosus*、*Rhodospseudomonas palustris*、*Synechococcus* spp.、*Thiobacillus ferrooxidans* 等(<http://www.tigr.org/>)。

6.2 功能基因组学

MGP 还重点拓展研究新策略,以便经济有效地完成那些与已知完整基因组序列的及其近缘的微生物基因组序列的测定。对这类微生物的序列测定,没有必要从头开始。DOE MGP 正在开发用于鉴定、操纵和模拟完整反应途径或调控网络的新策略和新工具,率先研究那些成群地工作共同决定特定产物或特定行为的一群基因,这些目标不仅是“功能基因组学”的核心内容,而且其重要性将随着更多基因组序列的完成,以及对其比较研究所揭示的新线索不断涌现,而显得越来越重要。

6.3 生物信息学

生物信息学依赖计算机分析和巨大的微生物基因组序列信息。发展生物信息学,将会极大地促进发现新基因和新功能特征的研究步伐,因此成为 DOE MGP 的又一个重要领域。总之,处在世纪之交的微生物学,正处在令人神往的时代,数量日益突增的微生物全基因组序列信息正在迅速改变着微生物学的所有方面。信息爆炸,又使我们变得越来越对微生物世界充满好奇和惊喜,特别是关于海洋微生物学知识的探索和积累,使我们越来越意识到人类获得的关于微生物学的知识还是多么的微不足道,越来越感叹微生物的潜在益处对于我们这个地球世界是多么的重要——只要有足够的智慧去发掘它们!

参 考 文 献

- [1] Fleischmann R D, Adams M D, White O, *et al.* *Science*, 1995, **269**:496 ~ 512.
- [2] Relman D A, Evelyn M D, Strauss E. *Microbial Genomes: Blueprints for Life. A Report from the American Academy of Microbiology*. Washington D. C. : ASM Press, 2000.
- [3] Abelson P H. *Science*, 1998, **279**:2019 ~ 2109.
- [4] Staley J T, Castenholz R W, Colwell R R, *et al.* *The Microbial World: Foundation of the Biosphere. A Report from the American Academy of Microbiology*. Washington D. C. : ASM Press, 1997.
- [5] Pace N R. *Science*, 1997, **276**:734 ~ 740.
- [6] Bult C J, White O, Olsen G J, *et al.* *Science*, 1996, **273**:1058 ~ 1073.
- [7] Woese C R, Fox G E. *Proc Natl Acad Sci*, 1977, **74**:5088 ~ 5090.
- [8] Fraser C M, Gocayne J D, White O, *et al.* *Science*, 1995, **270**:397 ~ 403.
- [9] Klenk H P, Clayton R A, Tomb J F, *et al.* *Nature*, 1997, **390**:364 ~ 370.
- [10] White O, Eisen J A, Heidelberg J F, *et al.* *Science*, 1999, **286**:1571 ~ 1577.
- [11] Smith D R, Doucette-Stamm L A, Deloughery C, *et al.* *J Bacteriology*, 1997, **179**:7135 ~ 7155.
- [12] Bartlett M S, Thomm M, Geiduschek E P. *Nat Struct Biol*, 2000, **7**:782 ~ 785.
- [13] Maeder D L, Weiss R B, Dunn D M, *et al.* *Genetics*, 1999, **152**:1299 ~ 1305.
- [14] Deckert G, Warren P V, Gassterland T, *et al.* *Nature*, 1998, **392**:353 ~ 358.
- [15] Fox J L. *ASM News*, 2000, **66**:388.
- [16] Nelson K E, Clayton R A, Gill S R, *et al.* *Nature*, 1999, **399**:323 ~ 329.