

水稻基因组工程

洪国藩 主编

上海科学技术出版社

洪国藩 主编

水稻基因组工程

上海科学技术出版社

内 容 提 要

21 世纪被认为是基因组时代。本书通过对水稻基因组遗传图、物理图构建原理及其应用的详尽描述,通过对基因组大规模 DNA 测序的系统分析,全面介绍了我国水稻基因组的研究成果,包括光敏不育基因、广亲和基因、杂种低温敏不育基因的定位和水稻基因组 BAC 物理图的构建等。

这是一部学术专著。全书以大量的原始论文,配上典型的图表,特别是以作者所在研究组的工作为基础,对所涉及内容作深入而精确的阐述。同时,提供了在遗传图和物理图构建中、在基因组大规模测序中所涉及的各种具体技术细节和实验操作步骤,使本书亦具有实用性。

本书可作为从事基因组及分子生物学学科的研究人员和高等院校师生的参考书。

图书在版编目(CIP)数据

水稻基因组工程 / 洪国藩主编. — 上海: 上海科学技术出版社, 1999. 10
ISBN 7-5323-5089-4

I. 水... II. 洪... III. 水稻 - 基因组 - 研究
IV. S511. 03

中国版本图书馆 CIP 数据核字 (1999) 第 53644 号

上海科学技术出版社出版发行

(上海瑞金二路 450 号 邮政编码 200020)

上海中华印刷有限公司印刷 新华书店上海发行所经销

1999 年 10 月第 1 版 1999 年 10 月第 1 次印刷

开本 787 × 1092 1/16 印张 13 插页 4 字数 160 000

印数 1—1 500 定价: 62.00 元

本书如有缺页、错装或坏损等严重质量问题,
请向本社出版科联系调换

序

序

与美国启动人类基因组计划时的情形相仿，水稻基因组计划在我国启动的时候，也经历了曲折。“花了很多钱，将基因组的DNA顺序全测出来了，也是一本不为人所看得懂的天书。”这是反对者的主要理由。纯粹的学术上争论是科学发展的一种动力，是件好事。当然，DNA顺序本身不是天书。关于这一点，现在已是国际科学界的共识了。1996年，在众多国家实验室的共同努力下，完成了酿酒酵母1206.8万个核苷酸的全序列测定。从中确定了约5885个蛋白质基因，140个rRNA，40个SnRNA和275个tRNA基因。人们从此第一次知道，要维持一个像酵母这样最简单的真核生物的生命，究竟需要多少个基因。现在，已有近20种生物的基因组DNA序列被测定出来了。结果显示，在每种生物中都发现了大量潜在的新基因。平均地说，新发现的基因约占了已知基因的40%！科学发展到今天，不去掌握维持重要生命（如人类和水稻）所需的全部基因和控制这些基因活动的DNA序列，怎能说是真正全面了解了这些生物的遗传信息，又怎能用这些知识有效地为人类自身服务呢？

我在这里特别提到了控制基因活动的DNA序列，这是因为，现在的确有一些人似乎有一种倾向，认为只要把注意力集中到基因就可以了，就能解开这种生物的全部遗传秘密。其实，这种看法是很不全面的。生命活动的种种现象，如生长、发育、繁殖、变异……当然是基因活动的综合结果。但是，如果没有特定的元件在时间、空间和程序上协调地控制各种相关基因的活动，或者这些控制元件在基因活动的某些环节中，发生了偏差或失去了作用，均会导致完全不同的后果，包括出现畸胎、疾病、生理障碍等。如何鉴定出调控基因的元件，在一定程度上比鉴定基因本身还难。一般来说，高级生命体（包括人类）的基因的总量不会超过该生命体总DNA的5%。那么，其余的95%DNA是否都是“废物”？当然不是。大量调控顺序的存在就是一例。因此，广义上讲，基因组研究应是基因组DNA序列及其功能的全面研究，是对基因及其余DNA顺序所含的全部生命信息的研究。

基因组的研究正在世界范围内全面展开。在国际上，现在已经开始形成了一个以基因组研究为基础的新兴的生命科学工业。据 *Science* 和

Nature近年的连续报道,这个工业将成为世界经济中发展最快的部门。

1992年8月,国家科委(现国家科技部)宣布,中国开始实施水稻基因组计划,目的是要在分子(核苷酸)水平上解开水稻的遗传之谜,以及将研究中所获得的结果应用于水稻品种的改良。从长远的观点看,我国的这一决策既为日益增长的粮食需求作了准备,又创造了一种条件,使我国的有关研究力量有可能汇聚到自1990年10月才正式开始出现的国际基因组研究的主流中去。这是一个及时的、搭知国际研究脉搏的决策。不久,水稻基因组物理作图及测序便得到了国家科技部、中国科学院和上海市人民政府的联合支持。数年后的今天,在广大科技工作者的艰苦努力下,不论在水稻遗传图的研究和应用上,还是在水稻基因组物理图的构建中,均取得了可喜的成果。同时,我国其他基因组的研究也进展瞩目。1996年,随着水稻基因组BAC物理图在我国问世,测定整个水稻基因组的DNA序列已成为可能。实际上,水稻第4号染色体大规模DNA测序工作已在我国进行。由于国家的正确决策和科技人员的努力,当21世纪世界进入“基因组时代”之际,我国已是这个时代的成员了。

水稻养活着几乎全球一半的人口,也是我国近一半人口的主食。十分幸运的是,对于提供如此庞大粮食的水稻,它的基因组却是禾本科作物中最小的(如只有小麦的约四十分之一)。这就使得水稻基因组很快便成为全球禾本科作物的首选研究目标。同样幸运的是,近来研究表明,水稻中基因的含量和基因的排列次序与其他禾本科作物是类同的。这就是说,如果通过对水稻基因组DNA的研究,搞清它所含有的基因及其排列次序,就可帮助找到小麦、玉米、高粱等其他禾本科作物的相关基因。从这个意义上讲,水稻基因组的研究已超越了其自身的意义。水稻成为全球禾本科作物的研究模型,已是国际科学家的共识了。

正如本书第1章(基因组学)中指出的那样,生命活动远不是各自无关的单个基因的孤独行为的结果。一个简单的生命现象背后,往往牵动着一大群基因有序而和谐的活动。比如,酵母总共有约6000个基因,但当它遇到不利的环境而开始形成孢子时,竟然牵动了近1000个基因的变化!水稻比酵母复杂得多了,其经济价值也大得多了。水稻被认为有约35000个基因。它的产量,它的米质,它的生命周期,它对病虫害、盐碱、温度等周围环境的反应,该有多少基因的参与!因此,如果不在科学上搞清楚这个秘密,也不搞清是什么DNA序列在控制着这个极其

复杂的基因网络,那么,要让水稻定向地为人类提供尽可能多的高产优质的粮食是难以想象的。而要搞清这些秘密,则必须对水稻基因组进行DNA测序,并研究所测序列的功能。这不仅是当今科学家的共识,而且已成为他们的一致行动了。从1998年开始,亚洲、欧洲、美洲的国家和地区均已起动了各自的水稻基因组DNA测序计划。实际上,水稻基因组测序已成为一个国际的行为。

水稻基因组的研究应尽可能地与水稻育种的研究相互渗透。我们愿将本书的出版看作为是一种好的兆头,希望在同样的数年后,国家的决策和广大科技工作者的努力,能使我国的这一研究结出更丰硕的成果。

本书介绍了有关基因组研究的原理和当前的进展,并着重阐述了我国在水稻基因组研究中取得的成果。本书的一个可贵之处是,书的作者们均为长期在研究第一线活跃的科技工作者,因而他们有着第一手的研究经验和科学储备,这最有利于对书中所涉及的问题作深入而精确的阐述。本书的另一可贵之处是,作者在各自的章节中努力使所涉及的内容实用化。这样,在读完有关的章节后,可望帮助有关读者在自身的工作中应用书中的原理和方法。比如,利用水稻遗传图寻找基因;用指纹法构建基因组的物理图;根据物理图展开大规模的基因组DNA测序,等等。为了配合上述目的,对书中所涉及的实验部分,包括较经典的和最新的技术,均作了详细的描述。不仅如此,书中大部分的实验方法均已经过作者们实验室的反复考验。这就使得书中所描述的技术和实验操作步骤具有了可靠性,使书的这部分内容具有实验室手册的性质。为了读者阅读方便,在书后还附了必要的名词解释。上述一系列努力之目的,是希望本书同时也能够对从事水稻基因组以外的科技工作者有所帮助。

生命科学正处在兴旺时期,更多的生命规律会被揭示出来,相应的生物新技术也会不断出现,人类改善自身生存质量的前景比以往任何时候都好。但是,危机同样存在。如何正确利用生命科学知识,正是人类面临的巨大挑战。如果允许利用生命科学知识去玩弄生命,从而导致现存生物格局的破坏,其后果将是毁灭性的。

基因组学是一个新兴的研究领域,发展很快,涉及许多交叉学科,书中不妥之处在所难免,欢迎指正。

洪国藩

1999年5月6日于上海

目 录

序

第1章 基因组学	1
1.1 遗传物质 DNA	2
1.2 结构基因组	6
1.2.1 基因组 DNA 测序的意义	7
1.2.2 基因密度和未知基因	14
1.2.3 典型基因组分析	16
1.3 功能基因组	23
1.3.1 SAGE 技术	24
1.3.2 DNA 芯片和微阵技术	25
1.4 比较基因组	26
1.4.1 微生物基因组的比较	27
1.4.2 单细胞和多细胞生物基因组比较	31
1.4.3 直向同源物集簇	32
参考文献	33

第2章 遗传图谱	37
2.1 遗传图谱的构建	38
2.1.1 基本概念	38
2.1.2 DNA 分子标记	39
2.1.3 构建遗传图谱的方法	43
2.1.4 水稻基因组遗传图谱	46
2.2 禾谷类基因组研究的模式植物	48
2.2.1 公共标记	48
2.2.2 水稻染色体连续区段的保守性	49
2.2.3 模式植物的确定	51
2.3 水稻遗传图谱的育种应用	51
2.3.1 亚洲栽培稻的遗传变异性	52
2.3.2 基因定位和标记辅助选择	57
2.3.3 水稻产量性状遗传机理和标记辅助高产育种	64
参考文献	77

第3章 遗传图谱与育种	79
3.1 应用分子标记研究水稻种质资源遗传多样性	80
3.1.1 遗传多样性的度量和分解	81
3.1.2 水稻种质遗传多样性的特点	82
3.2 光敏不育性的遗传和基因定位	88
3.2.1 光敏不育性的遗传	88
3.2.2 光敏不育基因的定位	89
3.3 籼粳杂种不育性和广亲和性的遗传基础	93
3.3.1 籼粳品种间的遗传分化与杂种不育的关系	93
3.3.2 广亲和性的遗传分析和基因定位	95
3.3.3 杂种低温敏不育基因的定位	101
3.4 杂种优势的遗传和分子基础	103
3.4.1 杂合性和杂种优势的预测	104
3.4.2 杂种优势遗传基础的 QTL 分析	107
3.4.3 上位性组分在杂种优势遗传基础中的重要性	108
3.4.4 基因表达与杂种优势的关系	110
3.4.5 分子育种	112
参考文献	113
 第4章 水稻基因组物理图的构建	 117
4.1 指纹锚标法原理	118
4.1.1 水稻基因组计划	118
4.1.2 水稻基因组物理图	119
4.1.3 BAC-指纹-锚标战略	121
4.2 指纹算法	128
4.2.1 Sanger 中心软件包	128
4.2.2 指纹图形的输入、编辑和数字化	129
4.2.3 指纹数据的分析和匹配	133
4.2.4 重叠群的形成	134
4.3 获取基因和大规模测序	135
4.3.1 分子标记	135
4.3.2 测序战略	138
4.4 构建物理图的实验程序	141
4.4.1 BAC DNA 的小量制备	142

4.4.2 指纹图谱的制作	143
4.4.3 重叠群在染色体上的定位	146
参考文献	149
第5章 基因组测序	151
5.1 基因组DNA测序战略	152
5.1.1 全基因组随机测序战略	152
5.1.2 以物理图为依据的测序战略	154
5.1.3 随机定向战略	157
5.1.4 随机定向测序	163
5.2 DNA操作	170
5.2.1 测序用DNA的制备	170
5.2.2 PCR、测序、电泳、标记和杂交	178
5.2.3 实验配方	187
参考文献	191
附录 专业名词简释	193



基因组学

从 20 世纪末开始, 科学家已有能力开始研究单个生物的全部遗传信息。从研究单个基因或一组基因到研究生物体的全部遗传信息是一个质的飞跃。因为, 从此人们可以根据完整的遗传背景解释生命, 掌握生命的规律, 使其更好地为人类服务。基因组学就是研究生命体全部遗传信息的一门科学。

1.1 遗传物质 DNA

在开始介绍基因组学之前, 先回顾一下 DNA 的研究进展。基因组 DNA 是一种奇妙的分子, 生命的密码蕴含在它的碱基的无穷排列方式之中。正因为如此, 对基因组的研究吸引了全世界科学家。到目前为止, 已有近 20 种生物的基因组 DNA 的顺序被测定出来了。其中包括 2 种真核生物 (啤酒酵母、线虫) 和目前再度出现并具抗药性而令人担忧的结核杆菌、梅毒及可能导致胃癌的幽门螺旋菌。对这样的研究, 人们有充分理由相信, 不久将可看到丰厚的回报。

40 多年来, 仅在核酸及其相关领域的诺贝尔奖获得者就有数十位^[1]。这些突破性的工作表明, 当今之所以有条件展开基因组研究, 是多年来科学发展和积累的结果, 绝不是偶然的。表 1-1 列出自 1957 年至 1997 年期间, 有关核酸领域中获诺贝尔奖的科学家名单及其主要贡献。

其实, 在 1957 年以前的诺贝尔奖项目中, 还有一些科学家。例如, 德国的 A. Kossel, 他的研究组发现了除鸟嘌呤以外常见的其他 4 种碱基, 即腺嘌呤、尿嘧啶、胞嘧啶和胸腺嘧啶, 于 1910 年获生理医学奖; 美

国的T.M.Morgan因发表基因与染色体连锁的学说,于1933年获生理医学奖;1946年美国的H.J.Muller,因发现X-射线能促使基因变异频度增高而获生理医学奖。上述一系列里程碑式的发现和发明,最终为基因组研究的大规模展开铺平了道路。

再往前看,在18世纪后期,发现的与核酸有关的化学成分是从鸟粪和人尿中得到的尿酸,这是世界上发现的第一个嘌呤化合物。到了19世纪,又从膀胱结石中找到了黄嘌呤,从鸟粪中找到鸟嘌呤,从牛脾中找到次黄嘌呤。现在看来,这些都是在核酸被发现以前的重要发现,虽然在当时没有人知道这些嘌呤化合物有什么生物学意义或功能,更不知道它们的结构。就在那个时候,细胞学家发现了细胞中的细胞核。于是人们开始研究细胞核的化学成分。1869年,瑞士的青年科学家Friedrich Miescher(1844~1895)在德国F.Hoppe-Seyler实验室工作时,从伤员绷带上的脓细胞中提取到细胞核,又从这种细胞核中提取到一类原先不曾认识的化合物。它呈酸性,不溶于稀酸,但能溶于稀碱,而且含有大

表 1-1 1957~1997 年核酸及其相关领域获诺贝尔奖情况

科学家	国籍	年份	奖项	发明、发现
A.R.Toood	英	1957	化学	确定核苷酸结构,合成二核苷酸等
G.W.Beadle E.L.Tatum	美 美	1958	生理、医学	发现基因的作用是对某些化学过程的调控
J.Kederberg	美	1958	生理、医学	发现细菌的基因重组,提出新的遗传观点
S.Ochoa	美	1959	生理、医学	酶促合成核糖核酸
A.Lornberg	美	1959	生理、医学	模型合成脱氧核糖核酸
J.D.Watson F.H.C.Crick	美 英	1962	生理、医学	提出 DNA 双螺旋结构学说
M.H.F.Wilkins	英	1962	生理、医学	DNA 的 X-射线衍射研究
F.Jacob A.M.Lwoff J.L.Monod	法 法 法	1965	生理、医学	调控基因的发现
R.W.Holley H.G.Khorana M.W.Nirenberg	美 美 美	1968	生理、医学	对遗传密码的解释及其在蛋白质合成中功能的揭示
M.Delbrück A.D.Hershey S.E.Luria	美 美 美	1969	生理、医学	基因结构和病毒复制机制的发现
E.W.Sutherland	美	1971	生理、医学	发现 3', 5'-环 AMP 和激素作用机制
H.M.Temin D.Baltimore	美 美	1975	生理、医学	发现反转录酶

(续表)

科学家	国籍	年份	奖项	发明、发现
R.Dulbecco	意	1975	生理、医学	肿瘤病毒与细胞遗传物质之间的相互作用
W.Arber	瑞士	1978	生理、医学	发现细菌限制性内切酶
H.O.Smith	美	1978	生理、医学	发现限制性内切酶作用方式的特点
D.Nathans	美	1978	生理、医学	用限制性内切酶制成肿瘤病毒的基因组图谱
P.Berg	美	1981	化学	建立 DNA 重组技术
W.Gilbert F.Sanger	美 英	1981	化学	DNA 一级结构测定方法
A.Klug	英	1982	化学	建立晶体电子显微技术测定核酸-蛋白复合体的构造
B.McClintock	美	1983	生理、医学	发现可移动的转座基因
根进利川	日	1987	生理、医学	发现产生抗体多样性的遗传原理
M.Bishop H.E.Varmus	美 美	1989	生理、医学	发现逆转录病毒癌基因的细胞起源
T.Cech S.Altma	美 美	1989	化学	RNA 具有催化功能的发现
L.Robert P.Sharp	英 美	1993	生理、医学	发现断裂基因
C.Mullis	美	1993	化学	发明聚合链式反应(PCR)体外扩增DNA技术
M.Smith	美	1993	化学	发明寡核苷酸介导的 DNA 定点诱变技术
E.B.Lewis C.Nusslein- Volhard E.F.Wieschaus	美 德 美	1995	生理、医学	控制果蝇体节发育的基因的鉴定
S.B.Prushier	美	1997	生理、医学	发现 prion (朊粒) 的致病机理

量磷元素。这引起了Miescher的重视。他把新发现的化合物命名为“核质”(nuclein),并将实验结果写成论文送给Hoppe-Seyler。当Hoppe-Seyler重复了这一结果后,将此结果发表了。他们所发现的所谓“核质”,现在看来是含有蛋白质的核酸制品。20年后,R.Altmann制备得到了不含蛋白质的物质,并首次将它命名为“核酸”。

自此之后,核酸吸引了化学家的注意。但是核酸的结构、生物学意义以及两者的关系,直到20世纪初仍无任何线索。甚至研究了核酸几十年的Kossel也认为核酸没多大重要性,因而改行研究细胞核中的碱性蛋白质去了。不久,第一次世界大战爆发,由于当时的世界核酸研究中心德国是参战的主要一方,使核酸研究被迫停顿。

第一次世界大战结束,德国战败,核酸的研究也就分散到其他国家,

特别是美国和西欧。20世纪发生的第二次世界大战,使核酸的研究又受到严重冲击而进展迟缓。然而就在战争结束前,即1944年, O.T.Avery, C.M.Macleod和M.McCarthy发表了他们谨慎研究达14年的结果:在研究肺炎球菌的转化现象时,发现从一种类型的肺炎球菌制备得到的DNA可以引入到另一类型的肺炎球菌体内,并使后者的遗传性状变成前者。于是,他们得出结论:遗传物质是DNA。这对当时流行的看法——蛋白质为遗传物质——是一个极大的冲击,从而DNA吸引了物理学家和化学家们的兴趣。

英国的化学家A.R.Toood,在1951年阐明了核酸链中的3',5'-磷酸二酯键;美国的E.Chargaff研究了核酸中A, T, C, G 4种碱基的相对含量;英国物理学家则用X-射线确定了DNA具有规则结构的事实和参数。建筑在如此坚实的基础上,剑桥大学的卡文迪许(Cavendish)实验室J.Watson和F.Crick于1953年,提出了DNA结构的双螺旋模型。这个学说不但阐明了DNA的结构,并且为一个DNA分子如何复制成两个相同结构的DNA分子,以及DNA怎样传递生物体的遗传性状,提供了合理的说明。这项作为现代分子生物学和分子遗传学奠定了基础,可被看作是当代生物学的最大突破,受到科学界的普遍重视。

直至如今,核酸及遗传领域的新发现、新发明一直不断,进展之快,涉及范围之广,对人类未来影响之深远,在其他研究领域是少见的。大体说来,有下面这些工作。

(1) 核酸的生物合成 包括DNA和RNA的酶促合成。这一点很重要,因为DNA序列的测定技术主要依赖酶促合成法。这项发现直接为DNA的一级结构测定奠定了基础。

(2) 细胞核外(即细胞质中)发现DNA 真核生物细胞的线粒体、叶绿体以及原核生物细胞的质粒都属于核外DNA。原核质粒的发现为基因重组技术提供了必要工具。

(3) DNA上三联密码的确定 这项发现直接证明了DNA作为遗传信息载体的理论推测。

(4) 重要核酸工具酶的发现 它们包括核酸酶、聚合酶、限制性内切酶、连接酶等。这些发现使人们直接有目的地干预遗传物质成为可能,并导致基因工程的产生。

(5) 基因的不连续性的发现 这一发现加深了人们对基因的认识,为从基因组角度探索遗传信息的组织结构创造了条件。

(6) 核酸之间以及核酸和蛋白质之间的相互作用的发现 核酸蛋白质复合体的发现打开了研究基因之间相互作用之门。

(7) 核酸的人工合成 化学合成核酸使人们可能有目的地设计特定的遗传信息, 也为PCR的发明创造了条件。

(8) 发现转座基因 这个发现一开始不为人们相信, 因当时的流行看法为, 基因是静止地呈线性排列在染色体上的。此发现30多年后, 随着更多事实的积累, 人们意识到基因组的这种动态特征乃是一种普遍现象。这不仅解释了生物多样性, 也为人工产生基因突变创造了条件。

(9) 免疫系统体细胞基因组的遗传重组多样性 这一发现展示给人们的是, 基因组在参与体细胞分化过程中的精妙行为。

(10) 癌基因细胞起源的发现 这为从基因组角度理解癌指示了方向。

(11) DNA的定点诱变技术和体外扩增(PCR)技术 这是基因工程领域的新工具, 使人们对遗传物质的干预能力进一步增强。

(12) 控制果蝇体节发育基因的鉴定 这个发现为阐明基因组信息如何指导多细胞生物体的形态建成开辟了道路。

以上是在核酸及遗传领域的发现中, 对已获诺贝尔奖的有关简介。近年来还有一系列重大进展, 其中值得一提的是, 如克隆羊技术、全基因组的微矩阵分析方法、人工染色体的构建成功和生物信息学的完善等, 这些都直接或间接地为基因组研究得以全面开展作出了贡献。

1.2 结构基因组

基因组学(genomics)最初是由Thomas Roderick于1986年提出的^[2], 其内容是指基因组作图(mapping)和测序(sequencing)。近年来, 这个概念有新的发展。现在人们倾向于将基因组学分为结构基因组学(structural genomics)和功能基因组学(functional genomics)两大部分。

功能基因组学是建筑在结构基因组学基础上的, 因此, 结构基因组学是第一阶段的工作。它的目标是构建出高分辨率的遗传图(genetic map)和物理图(physical map); 图内又可包含转录图(transcript map)。这里没有将测定的基因组DNA序列单独列出, 因为基因组DNA序列就是最高分辨率的物理图, 其分辨率为1个核苷酸。功能基因组学

是基因组分析的更进一步阶段。它的主要内容是：利用结构基因组学所提供的生物材料和信息，全基因组或全系统(global, genome wide, system wide)地理解某种生物的遗传体系。功能基因组学的研究必须结合统计学和计算机科学，应用所谓高产出(high throughput)和大规模的实验技术得以实现^[2]。

虽然，科学界现在越来越热烈地谈论功能基因组学，并开始有关这方面的研究工作，但实际上，目前还处在结构基因组学的阶段。

在结构基因组学中，基因组测序是最基本的研究领域。因为，只有完成了这一阶段的研究，才有可能在分子(核苷酸)的水平上破译生物的遗传之谜。在人类基因组计划的带动下，迄今已完成近20个不同生物的基因组DNA序列的测定，其中包括多个病原微生物和模式生物基因组的测序。流感嗜血杆菌(*Haemophilus influenzae*)^[3]是能独立复制的小基因组，而大肠杆菌(*Escherichia coli*)^[4]是多年来在生物学上极为频繁使用的、也是被研究得十分深入的实验室“工具”。啤酒酵母(*Saccharomyces cerevisiae*)^[5-12]是基因组序列被测定的第一个真核生物，秀丽线虫(*Caenorhabditis elegans*)^[13]则是第一个基因组序列被测定的多细胞动物。

自1990年开始大规模展开基因组DNA测序以来，进展非常迅速，参与这类研究的国家也不断增多。在本书1.2.2和1.2.3两节中，概括了这方面所取得的主要成果。通过基因组DNA测序，已经揭示和正在揭示出许多过去人们所未知的遗传信息。

在10年前，还有不少人怀疑基因组DNA测序的意义，现在情况已有了根本的扭转。基因组学的发展必将给生物学和人类社会带来巨大的冲击，引起深刻的变化，这一点已是确定无疑的了。

1.2.1 基因组DNA测序的意义

生命科学经过一二百年的发展，已由定性描述阶段开始进入到定量实验阶段，由宏观到微观分子生物学，所取得的成果是巨大的。现在已进入了基因组时代。通过基因组测序，可获得用其他手段所无法获得的蕴藏在DNA序列之中的生物体的生命信息，为加深人们对生命本质的认识，揭开生命之谜提供了一把钥匙。这一点，可以通过对大肠杆菌和酵母基因组DNA的分析显示出来。

大肠杆菌和酵母，可能是科学家们了解得比较清楚的两个生命系

统。因为许多分子生物学上的关键性突破,比如主要生物合成途径的阐明,基因调节机理的最初发现,遗传三联体密码的阐明,对基因的转录和蛋白质翻译的认识,DNA复制、突变和修复的机制,等等,几乎都是通过大肠杆菌的研究而得到的。所以长时间来,许多生物科学家认为,对大肠杆菌的了解已很透彻了。其实情况并不是如此。近来,完成了大肠杆菌基因组的DNA测序,并对所测序列进行了全面的计算机分析,结果令人吃惊。原来,大肠杆菌基因组中有40%左右的基因,其功能却一无所知(表1-2)^[4]。可以想象,如果搞清楚这40%左右基因的功能,则对大肠杆菌生命系统的了解肯定要全面得多。

长期以来,酵母是作为真核细胞的一种模式生物而被科学家们所研究。根据长期的遗传学分析得知,酵母基因组约含4000个基因,但在全基因组DNA序列测出后,发现酵母基因组中,有将近6000个编码蛋白质的基因。与大肠杆菌的情况相似,在酵母中也有50%左右的基因不为人们所知(表1-2)^[5-12]。

从此,基因组DNA测序成为世界各国科学家的一个活跃的研究领域。

为什么基因组DNA测序能发现这么多可能存在的基因呢?这是因为,传统的分子生物学和分子遗传学方法,往往是根据已知基因的产物(如蛋白质)去寻找相应的基因。而问题是,由于许多基因的产物十分微量,于是基因就不易被发现。不仅如此,基因在生物体的不同组织、不同器官、不同发育阶段,表达也各不相同,这又造成了发现新基因的困难。过去,大量基因是通过遗传突变才发现的。科学家们通过生物体的遗传突变,观察到生物体的表型变化,进而推测造成这种表型变化的基因,从而找出这些相应的基因。这是长期来一个十分有效的寻找基因的方法。问题是,虽然有时基因被改变了,但由于观察手段的限制,大量的表型变化并不被人们所观察到。鉴于上述种种原因,基因组中所含的许多基因,不能被鉴别出来,于是基因组测序及对序列功能的研究,便成为强大的揭示生命奥秘的武器。

生命现象是基因组中所有有关的基因,包括在复制、转录、转录后加工、翻译、翻译后加工等多个层次上,并在不同的时间和空间上进行运作、相互调节、相互作用的结果。这是一个高度错综复杂的网络系统,是一个天然的平衡整体。只有获得生物体基因组的全部DNA序列的信息后,才能在整个基因组水平上同时研究细胞中所有的基因及其产物,