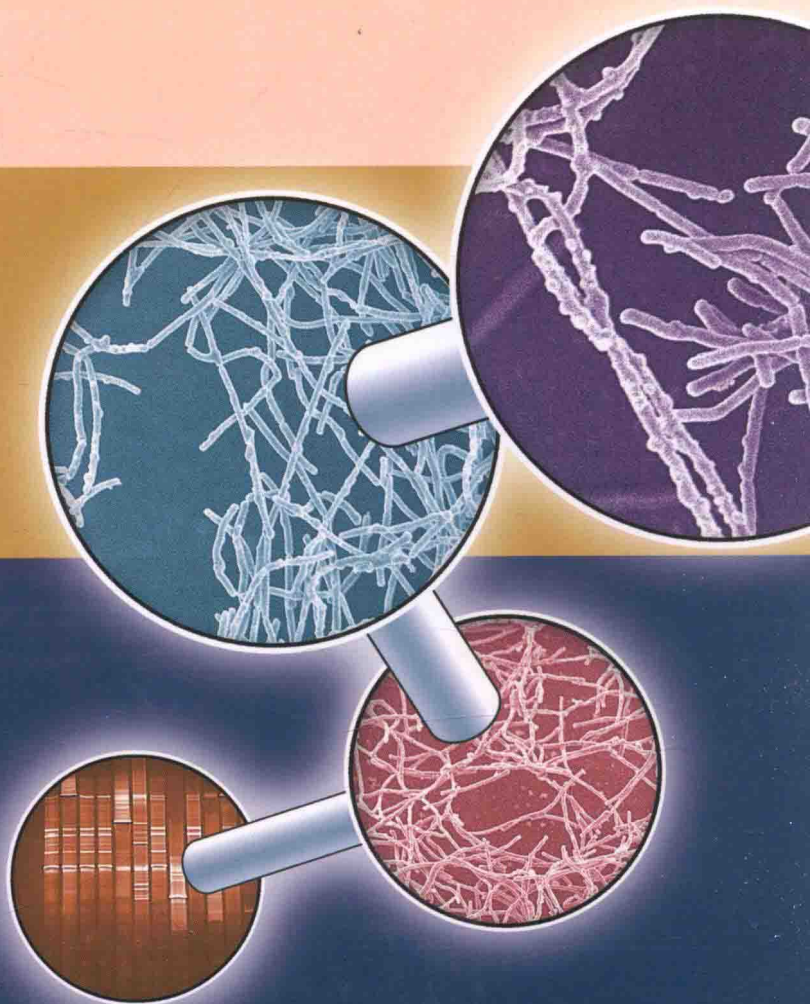


高等院校“十二五”规划教材 / 食品科学与工程系列

高等微生物学

● 主编 崔艳华



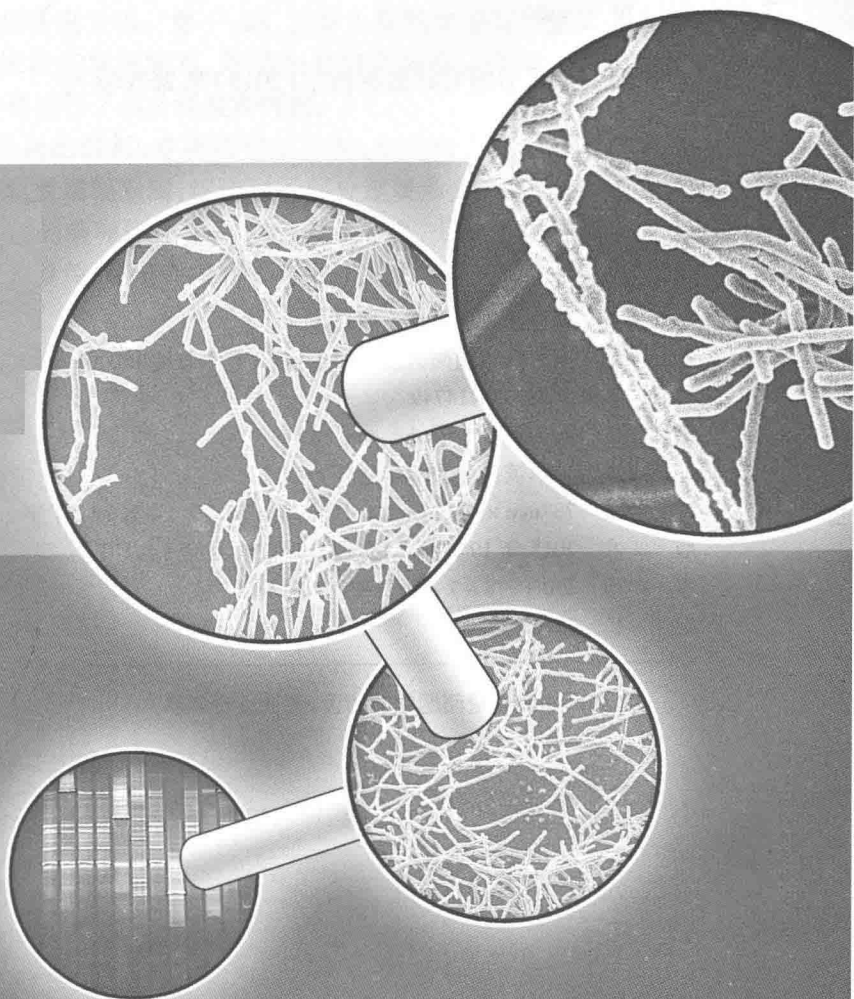
哈尔滨工业大学出版社

高等院校“十二五”规划教材 / 食品科学与工程系列

高等微生物学

● 主 编 崔艳华

● 副主编 曲晓军 代翠红



哈尔滨工业大学出版社

内 容 简 介

本书从微生物基因组学、极端环境微生物、微生物与环境、微生物代谢产物及其分子调控、微生物进化与多样性、微生物菌种选育等方面,就当前微生物学研究领域的最新研究进展进行了系统性的阐述。

本书可作为食品科学与工程、食品质量与安全、生物学等专业的硕士研究生教材,也可作为相关研究人员的参考用书。

图书在版编目(CIP)数据

高等微生物学/崔艳华主编. —哈尔滨:
哈尔滨工业大学出版社,2015.10
(食品科学与工程系列)

ISBN 978-7-5603-5707-2

I. ①高… II. ①崔… III. ①微生物学-研究生-教材 IV. ①Q93

中国版本图书馆 CIP 数据核字(2015)第 267068 号

策划编辑 杜 燕

责任编辑 郭 然

出版发行 哈尔滨工业大学出版社

社 址 哈尔滨市南岗区复华四道街 10 号 邮编 150006

传 真 0451-86414749

网 址 <http://hitpress.hit.edu.cn>

印 刷 哈尔滨工业大学印刷厂

开 本 787mm×1092mm 1/16 印张 16.75 字数 381 千字

版 次 2015 年 10 月第 1 版 2015 年 10 月第 1 次印刷

书 号 ISBN 978-7-5603-5707-2

定 价 39.80 元

(如因印装质量问题影响阅读,我社负责调换)

前 言

微生物学是生命科学领域内重要的学科之一,是现代高新生物技术理论与技术的基础。随着分子生物学研究的飞速发展,以微生物为材料的研究成果层出不穷,使整个微生物学研究发生了深刻的变化。本书从微生物基因组学、极端环境微生物、微生物与环境、微生物代谢产物及其分子调控、微生物进化与多样性、微生物菌种选育等方面,就当前微生物学研究领域的最新研究进展进行了系统性的阐述。本书在内容安排上本着由浅入深、循序渐进、理论联系实践的原则,注重系统性、广泛性、前沿性和实用性。本书可作为食品科学与工程、食品质量与安全、生物学等专业的硕士研究生教材,也可作为相关研究人员的参考用书。

本书由崔艳华(哈尔滨工业大学)担任主编,曲晓军(黑龙江省科学院微生物研究所)、代翠红(哈尔滨工业大学)担任副主编。第一、四、五章由崔艳华编写,第二、三章由代翠红编写,第六章由曲晓军、崔艳华共同编写,附录由曲晓军编写。全书由崔艳华统稿,并对各位参编人员编写的内容进行了部分修改和调整。

本书在编写过程中参考了大量的教材、专著和相关资料,在此向各位专家表示衷心的感谢。由于参考文献数量较多,受篇幅限制,仅选列了部分重要参考文献,如有疏漏之处,敬请各位专家批评指正。本书的编写倾注了每位编者的心血,但限于编者能力和水平有限,书中疏漏和不足之处在所难免,敬请专家和读者批评指正。

编 者

2015 年 9 月于哈尔滨

目 录

第一章 微生物基因组学	1
第一节 基因组学研究概述	2
第二节 基因组学的基础知识	8
第三节 微生物基因组	35
第四节 微生物结构基因组学	52
第五节 微生物功能基因组学	59
参考文献	67
第二章 极端环境微生物	70
第一节 极端环境微生物概述	70
第二节 极端嗜热菌	72
第三节 极端嗜冷菌	75
第四节 极端嗜酸菌	80
第五节 极端嗜碱菌	83
第六节 极端嗜盐菌	87
第七节 嗜压菌	92
第八节 极端微生物与古细菌的关系	96
参考文献	100
第三章 微生物与环境	102
第一节 微生物与环境污染	102
第二节 微生物与环境监测	107
第三节 污染环境的微生物治理与修复	109
第四节 污水的微生物处理	112
第五节 废气与固体废物的微生物处理	117
参考文献	119
第四章 微生物代谢产物及其分子调控	121
第一节 微生物代谢概述	121
第二节 微生物的代谢产物	132
第三节 微生物代谢的自动调节	139

第四节 微生物的总体调控	154
第五节 微生物代谢的人工控制	170
参考文献	178
第五章 微生物进化与多样性	180
第一节 微生物进化	180
第二节 微生物多样性	187
第三节 微生物分类鉴定技术	198
参考文献	214
第六章 微生物菌种选育	216
第一节 微生物传统选育方法概述	216
第二节 原生质体融合育种	226
第三节 基因工程育种	232
第四节 分子定向进化育种	238
参考文献	248
附录	251

第一章 微生物基因组学

本章引言 提到基因组,人们立刻会想到国际人类基因组计划(Human genome project, HGP)。人类基因组计划是由多国科学家共同完成的庞大的科研计划,耗资约 30 亿美元,历时 10 年,测序工作量高达 30 亿碱基对,成绩斐然,令世人瞩目。人类基因组计划的完成对人类有着非常深远的影响。与人类基因组相比,人们对于微生物基因组却了解甚少。但这并不意味着微生物基因组研究无足轻重,恰恰相反,微生物基因组研究具有举足轻重的地位。

由于微生物个体微小,基因组相对较小,易操作,因此对于它的研究往往是先行一步。早在 1985 年,人类基因组计划启动之前已经完成了数个病毒的全基因组测序。在微生物基因组学研究过程中所取得的理论和技术的进展,为人类基因组计划提供了极为有利的借鉴。在最初的人类基因组研究中,研究者选择了包括大肠杆菌和酵母在内的 5 个模式生物先行研究。同时,人类基因组计划的巨大资金投入也极大地促进了微生物基因组研究的飞速发展。

自 1995 年 12 月完成了第一个微生物(流感嗜血杆菌)全基因组测序以来,微生物基因组研究发展迅猛,并促进了其他生命形式(包括人类)的基因组学的发展,开创了生物基因组研究时代。互联网数据库中各种生物基因组数据日益增长。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 6 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”。

微生物全基因组测序,不仅是人类开始的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。截至 2015 年 9 月,已经测得了 5 899 种微生物基因组,并且这个数字正以惊人的速度不断增长。

美国国家能源部(Department of energy, DOE)和国立卫生研究院(National institutes of health, NIH)是美国微生物基因组计划(Microbial genome project, MGP)的启动和发展者,两个部门所关注的领域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及作用。已经完成基因组测序的微生物包括多种重要病原微生物以及资源微生物,其中包含人们熟知的引起胃炎、胃溃疡的幽门螺杆菌,引发结核病的结核杆菌等病原菌。这些病原微生物全基因组序列的完成,有利于人们更加深入地认识这些病原菌的致病分子机制,从而为疾病的诊断、预防和治疗开辟了新的途径。

资源微生物基因组的测序完成,将获得更多的关于微生物有用代谢产物调控的信息,为利用微生物生产更多有用代谢产物提供可能,进而为解决当前能源紧张的困境提

供新的途径。同时,人们对人类健康有益的一系列的乳酸菌进行了全基因组测序。人们希望从全基因组的信息中,挖掘乳酸菌的代谢调控机制,发现关键的功能基因,更好地为人类的健康服务。微生物基因组学在人类的健康、能源、环境等方面起着重要作用。微生物基因组研究正以惊人的速度向前发展,势必对人类产生极其深远的影响。本章主要分为两部分:第一部分是基因组学概论,介绍基因组学的基本知识;第二部分是阐述微生物基因组的研究进展,主要包括微生物结构基因组学、微生物的功能基因组学以及当今微生物基因组研究的最新进展。

本章要点 近年来,微生物学研究已经进入了全基因组研究时代。本章系统阐述微生物基因组学研究发展的历史、基因组学研究的基础知识和研究方法、微生物结构基因组学和功能基因组学研究概况,并结合当前科学家最新的研究成果,重点阐述当前食品微生物基因组学的最新研究进展。主要内容包括:基因组学的基本概念,基因组学研究的关键支撑技术,基因组学研究策略,微生物的结构基因组学和功能基因组学,微生物基因组的研究概况、特点以及代表性微生物基因组。

第一节 基因组学研究概述

一、基因组学的基本概念

(一) 基因组

关于遗传物质,科学家们早有所臆测。1860—1870年,奥地利科学家遗传学奠基人孟德尔(Gregor Johann Mendel, 1822—1884)根据豌豆杂交实验提出遗传因子概念,并总结出孟德尔遗传定律。孟德尔提出,生物的每个性状是通过“遗传因子”(Hereditary factor)进行传递的,遗传因子是一些独立的遗传单位。孟德尔把可观察的表观性状和控制它的内在的遗传因子区分开来。由此,遗传因子作为基因的名词雏形诞生了。

1909年,丹麦遗传学家约翰逊(Wilhelm Johannsen, 1859—1927)在《精密遗传学原理》一书中根据希腊语“给予生命”之义,创造“基因”(Gene)一词来代替孟德尔假定的“遗传因子”。从此,基因便成为遗传因子的代名词一直沿用至今。在孟德尔的成果获得承认后,生物界普遍观点是遗传因子(即基因)决定了生物的性状。但是,基因究竟在细胞内的什么位置?不得而知。摩尔根(Thoman Hunt Morgan, 1866—1945)以果蝇为实验对象,做了大量的研究,回答了这一问题——基因在染色体上。摩尔根于1926年出版的《基因论》中首次完成了当时最新的基因概念的描述:基因是在染色体上呈线性排列的遗传单位,它不仅是决定性状的功能单位,也是一个突变单位和交换单位。至此,人们对基因概念的理解更加具体和丰富。

德国汉堡大学 Hans Winkler 教授于1920年首次提出基因组(Genome)这一概念。基因组是基因和染色体(Chromosome)两个词组合而成的,代表生物的全部基因和染色体组

成,即基因组包含了生物的全部遗传信息。

(二) 基因组学

尽管基因组的概念提出很早,但是当时对基因的认知很肤浅,只知道基因是决定生物性状的遗传单位,其化学本质完全是一个未知数,因此谈不上基因组学的研究。直至通过肺炎双球菌转化实验、噬菌体侵染实验、烟草花叶病毒重组实验等3个经典实验证明DNA是遗传信息的载体,人们对基因才有了进一步的认识。美国分子生物学家詹姆斯·沃森(James Dewey Watson, 1928—)和英国物理学家弗朗西斯·克里克(Francis Harry Compton Crick, 1916—2004)于1953年根据英国分子生物学家莫里斯·威尔金斯(Maurice Wilkins, 1916—2004)和英国物理化学家与晶体学家罗莎琳·富兰克林(Rosalind Franklin, 1920—1958)所进行的DNA晶体X射线衍射分析,提出了著名的DNA双螺旋结构模型。进一步研究证明,基因是DNA分子的一个区段。每个基因由成百上千个脱氧核苷酸组成,一个DNA分子可以包含几个乃至几千个基因。基因的化学本质和分子结构的确定具有划时代的意义,它为基因的复制、转录、表达和调控等方面的研究奠定了基础,开创了分子遗传学的新纪元。至此,人们才真正了解了基因的本质。随后在1973年发明的DNA测序技术、1985年发明的PCR技术,为基因组学的出现提供了必要的理论知识和技术手段。

在理论知识不断完善和技术手段不断涌现之后,1986年美国科学家Thomas Roderick首先提出了基因组学(Genomics)的概念。次年,Thomas创办了以Genomics命名的杂志。现今Genomics已成为基因组学研究领域的顶级期刊。

基因组学是对生物的所有基因进行基因组作图、核苷酸序列分析、基因定位和基因功能分析的科学。基因组学研究包括两部分内容:一部分是全基因组测序;另一部分是基因功能的鉴定。前者属于结构基因组学(Structural genomics)研究的范畴,后者以基因功能鉴定为目标的研究,为功能基因组学(Functional genomics)研究的内容。

(三) 结构基因组学

结构基因组学为基因组分析的早期阶段,是基因组学的一个重要组成部分和研究领域,它是以全基因组测序为目标,通过基因作图、核苷酸序列分析以确定基因组成、基因定位的科学,以建立具有高分辨率的生物体基因组的遗传图谱、物理图谱、序列图谱以及转录图谱为主要内容。结构基因组学的研究为后期功能基因组学的研究奠定了基础。

(四) 功能基因组学

功能基因组学是利用结构基因组学所提供的信息及产物,系统地研究基因的功能及调控机理的一门科学,是以高通量、大规模实验方法、统计与计算机分析为主要特征。功能基因组学的一个近期的目标是采用高通量、大规模、自动化的方法,加速遗传分析进程;长远目标是避开传统遗传分析的局限,采用系统化的途径及数据采集方法阐明复杂的生物学现象。

功能基因组学为基因组分析的后期阶段,因此也被称为后基因组学。多种模式生物,比如大肠杆菌、酿酒酵母、秀丽线虫已经进入了功能基因组学,也就是后基因组学的研究阶段。自 2000 年,人类基因组工作草图完成以来,人类基因组研究也已进入了所谓的后基因组时代。

(五) 比较基因组学

比较基因组学(Comparative genomics)是在基因组图谱及测序的基础上,即在结构基因组的基础上,对来源不同生物的基因和基因组结构进行比较,以了解基因的功能、表达机理及物种进化的科学。

例如,利用比较基因组学,揭示人类的疾病基因在动物中的同源基因,为人类疾病研究提供动物模型。同时,可以利用比较基因组学研究物种的进化。水稻的起源研究是一个很好的例子。水稻分为粳稻和籼稻两个亚种。人们一直对是粳稻先出现还是籼稻先出现这个问题争论不休。但一直苦于没有更加切实的实验证据验证结论的正确性。2002 年,人们分别获得了两个亚种的全基因组序列。通过比较基因组的研究发现,研究人员断言,这两个亚种起源相同,因为它们之间的基因关系比它们与在印度或中国发现的任何野生稻品种之间的基因关系都更为密切,并认为粳稻和籼稻的分化是在大约 3 900 年前。

比较基因组学在德氏乳杆菌保加利亚亚种(*L. delbrueckii* subsp. *bulgaricus*)和嗜热链球菌(*S. thermophilus*)的共生关系上也得以应用。长久以来,人们就认识到上述两个菌株在牛奶发酵过程中彼此促进生长并产生酸化,但是这个过程涉及的机制并不十分清楚。最为普遍接受的观点是:*L. delbrueckii* subsp. *bulgaricus* 具有一个胞外细胞壁结合蛋白酶,可以降解牛奶蛋白。通常 *S. thermophilus* 不具有此类蛋白酶,进而 *L. delbrueckii* subsp. *bulgaricus* 分解牛奶蛋白为 *S. thermophilus* 能提供肽和氨基酸。同时 *S. thermophilus* 产生的甲酸和 CO_2 ,将促进 *L. delbrueckii* 的生长。

研究者通过 *L. delbrueckii* subsp. *bulgaricus* 与 *S. thermophilus* 的全基因组比较分析,发现还存在着其他因素在二者协作上扮演着重要角色。*L. delbrueckii* subsp. *bulgaricus* 基因组序列编码了一整套合成叶酸的基因。叶酸是许多代谢反应的辅助因子,也是人类所需的重要维生素。然而,*L. delbrueckii* subsp. *bulgaricus* 没有产生叶酸的组成成分 PABA(对氨基苯甲酸)的途径,而 *S. thermophilus* 拥有合成 PABA 和叶酸所必需的酶。因此,二者共培养时可以产生较高水平的 PABA 和叶酸。

同时研究者发现,*S. thermophilus* 基因组具有较低的 GC 含量(39.1%),而 *L. delbrueckii* subsp. *bulgaricus* 基因组的 GC 含量则为 49.7%。基因组分析表明,*S. thermophilus* *pepD* 基因中含有一个发生基因水平转移的 GC 含量异常区域。该区域内的 3 600 bp 大小片段,与 *L. delbrueckii* subsp. *bulgaricus* 的 DNA 区域同源性高达 95%,编码了独特的含硫氨基酸代谢相关基因簇 *cbs* - *cblB*(*cglB*) - *cysE*(又称 *cysM2* - *metC* - *cysE2*),该基因簇是含硫氨基酸(如甲硫氨酸、半胱氨酸)合成所必需的。以上发现表明,*S. thermophilus*

与 *L. delbrueckii* subsp. *bulgaricus* 长期存在一个生境中,基因组之间发生了基因水平转移,便于相互协作和共生。

二、基因组学研究的历史与现状

(一) 人类基因组计划回顾

基因组学和人类基因组计划是相伴产生的,因此基因组学的历史要从人类基因组计划谈起。人类基因组计划与曼哈顿原子弹计划、阿波罗登月计划并称为人类自然科学史上的三大计划,其规模和意义超过了曼哈顿原子弹计划和阿波罗登月计划,被公认为是生命科学发展史上的里程碑。

之所以提出人类基因组计划,源于人们对健康的考虑。其实,早在 20 世纪 60 年代,美国总统肯尼迪(John Fitzgerald Kennedy, 1917—1963)就提出了两个科学计划,一个是阿波罗登月计划,另一个就是与人类健康相关的攻克肿瘤计划。阿波罗登月计划最终得以实现,但是由于人类遗传信息的复杂性,攻克肿瘤这个难题一直悬而未决。人们一直在寻找攻克这一难题的方法和途径。毫无疑问,人类全基因组序列的获得,将为人们提供人类全部的遗传信息,进而在整体上破解人类遗传信息的奥秘,从 DNA 的层面上找出致癌的原因。由此,人类基因组计划应运而生。

该计划旨在获悉人类基因组的全部 DNA 碱基序列,进而为阐明人类所有基因在染色体上的位置以及其结构与功能,破译人类遗传信息,解码人类生命的奥秘奠定基础。1984—1990 年,美国能源部组织召开了多次会议,对人类基因组计划实施的可行性进行了讨论,并且成立了国家人类基因组研究中心。DNA 双螺旋的发现者之一詹姆斯·沃森担任了该中心的第一任主任。

自 20 世纪 70 年代起,英国生物化学家弗雷德里克·桑格(Frederick Sanger, 1918—2013)等研究者已开始进行病毒的全基因组测序,到 1985 年已经完成了噬菌体 Φ X174 (5 386 bp)、Epstein-Barr 病毒(172 281 bp)、T7 噬菌体(39 937 bp)、花叶病毒基因组(8 024 bp)、大肠杆菌丝状噬菌体 I_{ke}(6 883 bp)和人线粒体基因组(16 000 bp)全基因组的测序。尽管它们的基因组只有几千个到十几万个核苷酸,不及人类基因组的万分之一,却已经激发了科学家们破译人类基因组全序列的巨大信心。随后在 1985 年发明的 PCR 技术,更为基因组学研究注入了新的活力。在基本理论知识和必要的技术手段具备之后,人类基因组计划在 1990 年 10 月正式启动。

经过近 10 年的努力,在 2000 年 6 月 26 日,由美国国家人类基因组研究中心和塞莱拉(Celera)遗传公司同时宣布人类基因组工作草图完成,测出人类基因组约含 30 亿碱基对,预测编码基因 5 万~8 万个。时任美国总统克林顿在白宫发表讲话,向世人宣布人类基因组工作草图完成。Celera 遗传公司的人类基因组计划的领导人 Craig Venter 和国家人类基因组研究中心的负责人 Francis Collins 也出席了会议。

人类基因组计划是一个跨国的国际科研计划,由美国、日本、德国、法国、英国、中国等六国科学家协作完成。1999 年,中国加入人类基因组计划,承担 1% 的测序任务,成为

参与这一国际科研计划的唯一发展中国家。

2001 年 2 月,在国际顶级学术刊物 *Nature* 和 *Science* 杂志上分别发表了国际公共领域和 Celera 遗传公司及其合作者完成的人类基因组工作草图以及初步分析结果,报告人类基因组共有 30 亿个碱基对,预测编码基因 31 000 个,比最初预测的 10 万个编码基因数大大减少。

人类基因组计划重要时间表见表 1.1。

表 1.1 人类基因组计划重要时间表

时间	事件
1984 年	在犹他州(Utah)的 Alta 城,R. White 和 M. Mendelsonhn 受美国能源部的委托主持召开了一个小型专业会议讨论,测定人类整个基因组 DNA 序列的意义和前景
1985 年 5 月	在加利福尼亚州 Santa Cruz 由美国能源部的 R. L. Sinsheimer 主持的会议上提出了测定人类基因组全序列的提议,形成了美国能源部的“人类基因组计划”草案
1986 年 3 月	在新墨西哥州的 Santa Fe 讨论了这一计划的可行性,随后美国能源部宣布实施这一计划
1986 年	诺贝尔奖得主杜尔雷纳托·杜尔贝科(Renato Dulbecco)在 <i>Science</i> 周刊撰文回顾肿瘤研究的进展,指出要么依旧采用“零敲碎打”的策略,要么从整体上研究和分析人类基因组
1988 年	美国成立了“国家人类基因组研究中心”,由 James Dewey Watson 出任第一任主任
1990 年 10 月	美国能源部与国立卫生研究院共同启动了被誉为生命科学的“阿波罗登月计划”国际人类基因组计划,总体计划在 15 年内投入至少 30 亿美元进行人类全基因组的分析
1994 年	中国人类基因组计划在吴旻、强伯勤、陈竺、杨焕明的倡导下启动,最初由国家自然科学基金会和“863”高科技计划的支持下,先后启动了“中华民族基因组中若干位点基因结构的研究”和“重大疾病相关基因的定位、克隆、结构和功能研究”
1998 年	在上海成立南方人类基因组中心
1998 年 5 月	一批科学家在美国罗克威尔组建 Celera 遗传公司,目标投入 3 亿美元,到 2001 年绘制完整的人体基因组图谱,与国际人类基因组计划展开竞争
1999 年 3 月 15 日	英国威尔科姆基金会宣布,由于科学家加快工作步伐,人类基因组工作草图将提前至 2000 年输出
1999 年	在北京成立北方人类基因组中心

续表 1.1

时间	事件
1999 年 7 月	中国积极加入这一研究计划,负责测定人类三号染色体短臂上约 30 Mb 区域的测序工作,该区域约占人类整个基因组的 1%。中国是继美国、英国、日本、德国、法国之后第 6 个国际人类基因组计划参与国,是参与这项计划的唯一的发展中国家
1999 年 12 月 1 日	国际人类基因组计划联合研究小组宣布,他们完整地破译出人体第 22 对染色体的遗传密码,这是人类首次成功地完成人体染色体基因完整序列的测定
2000 年 3 月 14 日	美国总统克林顿和英国首相布莱尔发表联合声明,呼吁将人类基因组研究成果公开,以便世界各国的科学家都能自由地使用这些成果
2000 年 4 月 6 日	美国 Celera 遗传公司宣布破译出人类的完整遗传密码。但是不少欧美科学家对此表示质疑,认为该公司的研究“没有提供有关基因序列的长度和完整性的可靠参数”,因而是“有漏洞的”
2000 年 4 月底	中国科学家按照国际人类基因组计划的部署,完成了 1% 的人类基因组的工作框架图
2000 年 5 月 8 日	由德国和日本等国科学家组成的国际科研小组宣布,他们已基本完成了人体第 21 对染色体的测序工作
2000 年 6 月 26 日	科学家公布人类基因组工作草图

(二) 模式生物基因组测序

人类基因组计划从 1990 年启动到 2000 年人类基因组工作草图完成,历时 10 年得以完成。在人类基因组计划实施的初期,也就是在该计划启动后的前 5 年的时间里,科学家们选取了大肠杆菌、酿酒酵母、秀丽线虫、果蝇和小鼠 5 种模式生物作为人类基因组测序的先行生物,为后期人类基因组测序积累经验。在 1996 年率先完成了酿酒酵母的全基因组测序,随后在 1997 年大肠杆菌的全基因组序列测序完成。接着,在 1998 年、2000 年和 2005 年分别完成了秀丽线虫、果蝇、小鼠的全基因组测序。至此,最初选定的 5 种模式生物的基因组测序工作全部完成。5 种模式生物基因组测序信息见表 1.2。

表 1.2 5 种模式生物基因组测序信息

名称	基因组大小 /Mb	预测基因数量 /个	平均基因长度 /kb	完成时间/年
大肠杆菌(<i>Escherichia coli</i>)	4.6	4 288	1	1997
酿酒酵母(<i>Saccharomyces cerevisiae</i>)	12.16	6 352	1.9	1996
秀丽线虫(<i>Caenorhabditis elegans</i>)	97	18 424	5.3	1998
果蝇(<i>Drosophila melanogaster</i>)	130	13 601	10	2000
小鼠(<i>Mus musculus</i>)	3 000	40 000 ~ 100 000	30 ~ 70	2005

同时,人类基因组计划的实施也带动了其他生物的基因组测序工作的展开。例如,在 2000 年,完成了模式植物拟南芥的全基因组测序。这是第一个完成的植物全基因组

测序工作。在 2002 年,完成了水稻即籼稻和粳稻基因组全基因测序。值得一提的是,我国科学家在生物基因组测序领域走在了国际前列,已经完成了水稻、家鸡、家蚕以及大量微生物基因组测序工作。但在基因功能研究上,我国与发达国家相比还有很大差距。基因测序工作有专门的仪器和程序化的流程,其科学原创性较小,与其说是科学研究,倒不如说是一个项目工程。而基因功能的研究是一个原创性的新发现,当然有知识产权保护,从而可以控制下游研究(如基因药物研发、基因诊断、基因治疗等),具有重要的科学意义和巨大的商业开发前景。目前,新基因的发现与日俱增,功能基因的研究处于蓬勃发展的时代,竞争日趋激烈,因此应加强这方面的研究。

(三)微生物基因组的研究概况

由于微生物基因组相对较小,通常在几百万个碱基对左右,易于操作,故对微生物基因组的研究先行一步,起到“先行官”的作用。但作为微生物基因组计划的正式开始,是在 1994 年由美国能源部立项启动的。这项计划一经启动,其发展就甚为神速。1995 年 7 月, *Science* 首次刊登了 TIGR (The institute for genomic research) 采用全新测序方法完成的流感嗜血杆菌 (*Haemophilus influenzae*) 的全基因组测序的论文,这是人类完成的第一个单细胞微生物的全基因组序列的测定,标志着基因组时代的真正开始。由于微生物种类的多样性,微生物基因组计划将扩展至数以万计的微生物基因组。可以预计,今后微生物基因组计划的总投入和测序工作都将可能超越人类基因组计划。它对本学科、其他学科以及人类本身产生的深远影响不言而喻。

研究者对与人类健康、能源开发、环境保护、食品等方面密切相关的微生物展开了一系列的基因组研究。与人类健康相关的多种病原微生物如生殖器支原体、幽门螺杆菌、包柔螺旋体、结核杆菌、梅毒螺旋体、金黄色葡萄球菌、幽门螺杆菌等已经测序完成,为揭示其致病分子机制,设计诊断、预防和治疗新方法、新途径开辟了无限可能。大量的资源微生物,如产谷氨酸的谷氨酸棒杆菌、产抗生素的天蓝色链霉菌、产丙酮丁醇的丙酮丁醇梭菌、产乙属还原固定能力的微生物,如荧光假单胞菌、恶臭假单胞菌、金属还原地杆菌等被测序,揭示了生物降解相关基因,为利用基因工程方法改造微生物,更好地为人类服务成为可能。大量的食品微生物,如酿酒酵母、乳酸乳球菌、嗜热链球菌、植物乳杆菌等均已测序完成,为更好地开发食品微生物的优良特性提供了参考与借鉴。

第二节 基因组学的基础知识

一、基因组学研究的支撑技术

基因组学和人类基因组计划的提出并不是空穴来风,它们是建立在人们对基本理论知识足够的认识,同时掌握必要技术手段的基础之上的。如果人们没有对 DNA 化学本质、DNA 双螺旋结构的认识,没有 DNA 测序技术、PCR 技术,就不可能进行基因组学研究,无法开展人类基因组计划。

在基因组学研究的前期阶段,即结构基因组学研究中主要有 4 个关键技术,分别为 PCR 技术、DNA 测序技术、DNA 自动化测序技术以及生物信息学分析技术。

(一) PCR 技术

PCR(Polymerase chain reaction)是聚合酶链反应的简称,PCR 技术是基因组测序的基础,为基因组测序提供足够的测序样品。

PCR 技术是在核酸研究的基础上发展起来的。早在 1869 年瑞士的年轻医生 Miescher 就开始对核酸进行研究,至今已经有 100 多年的历史。自 1930 年正式提出 DNA 和 RNA 两种核酸的概念后,人们对核酸的化学组成、结构及其功能进行了更为深入的研究,特别是 1953 年詹姆斯·沃森(James Dewey Watson)和弗朗西斯·克里克(Francis Harry Compton Crick)根据 DNA 晶体 X 射线衍射图形、各种化学分析数据提出的 DNA 双螺旋结构以及克里克半保留复制模型,是生物学史上的一个飞跃,这使得对基因在体外克隆、表达、调控等方面的研究取得了长足的进展,并由此拉开了基因工程的序幕。

20 世纪 70 年代以来,人们采用两种思路尝试建立基因的无性繁殖体系。一种思路是重组 DNA 技术,它是从基因文库中分离出单个的目的基因,将其拼接到载体中构建成重组 DNA。因载体是具有独立复制能力的质粒或噬菌体,当重组 DNA 引入细菌细胞后,经多次复制便可得到足够数量的 DNA 克隆片段。DNA 的连接酶和限制性内切酶的发现,为重组 DNA 技术和基因克隆铺平了道路,现已经成为最为常用的基因克隆技术。

另一种思路是生物化学家科兰纳(Har Gobind Khorana, 1922—2011)等在 1971 年提出的,其观点是:DNA 在体外经变性后,与适当引物结合,再用 DNA 聚合酶延伸引物并且不断重复该过程便可克隆 DNA。这种核酸体外扩增的设想由于当时不能合成寡聚核苷酸的引物,同时很难进行 DNA 测序而渐渐被人们所忽略。利用美国生物化学家 Arthur Kornberg(1918—2007)在 1958 年发现并分离的 DNA 聚合酶,美国 PE-Cetus 公司人类遗传研究室的年轻科学家凯利·穆利斯(Kary Banks Mullis, 1944—)在 1983 年实现了具有划时代意义的聚合酶链反应,使得 Khorana 的这一设想终于付诸实现。1987 年,第一台 PCR 仪问世,使该技术自动化成为现实,进而得以广泛推广和应用。由于 PCR 技术具有特异性强、简便快速、灵敏度高等优点,赢得了研究者的广泛认可和使用。现在,PCR 技术已经成为生命科学实验室获得某一目的 DNA 片段的常规技术。PCR 技术的发明者凯利·穆利斯也因此获得了 1993 年的诺贝尔化学奖。

PCR 技术的原理与细胞内发生的 DNA 复制过程十分类似。首先是在临近沸点的温度下,双链 DNA 分子解链,分离成两条单链的 DNA 分子,然后 DNA 聚合酶以单链 DNA 分子为模板,并且利用反应混合物中的 4 种脱氧核苷三磷酸(dNTPs)合成新生的 DNA 互补链。此外, DNA 聚合酶同样需要一小段双链 DNA 来启动“引导”新链的合成。因此,新合成的 DNA 链的起点,实际上是由加入在反应混合物中的一对寡聚核苷酸引物在模板 DNA 链两端的退火位点决定的。

在为每一条链均提供一段寡核苷酸引物的情况下,两条单链 DNA 都可作为合成新生互补链的模板。由于在 PCR 反应中所选用的一对引物,是按照与扩增区段两端序列彼此互补的原则设计的。因此,每一条新生链的合成都是从引物的退火结合位点开始,并沿着相反链延伸。这样,在每一条新合成的 DNA 链上都具有新的引物结合位点。然后,反应混合物经再次加热使新、旧两条链分开,并加入下一轮的反应循环,即引物杂交、DNA 合成和链的分离。PCR 反应的最后结果是,经过 n 次循环后,反应混合物中所含有

的双链 DNA 分子数,即两条引物结合位点之间的 DNA 区段的拷贝数,理论上是 2^n 。

PCR 反应只需要模拟 DNA 的体内复制条件,在试管中提供 DNA 合成所需要的物质以及适合的温度和作用时间,就可以完成 DNA 在体外的复制。PCR 反应需要模板 DNA、一对上下游引物、DNA 聚合酶、底物(即 4 种脱氧核糖核苷酸)以及合适的缓冲体系。设置邻近沸点的温度作为变性温度。退火温度主要取决于引物的解链温度(T_m 值)。延伸温度一般为 72 ℃,该温度主要取决于 PCR 所用的 Taq 酶的来源。

简述 PCR 反应的过程如下:首先,由一个双链 DNA 充当模板。在邻近沸点的温度即变性温度下,一般采用 94 ℃,双链 DNA 会分离成两条单链的 DNA 分子,充当扩增的模板;然后当温度下降达到退火温度以下时,上、下游引物分别与两条 DNA 模板相结合。温度略有升高,达到延伸温度(72 ℃),在 DNA 聚合酶作用下,引物沿着模板链方向,利用反应混合物中的 4 种脱氧核苷三磷酸作为底物,合成新生的 DNA 互补链。如此进行反复循环,从而得到大量的 PCR 产物。PCR 产物的量取决于 PCR 的循环数。

(二) DNA 测序技术

DNA 测序的方法很多,主要有链终止法、化学降解法、焦磷酸法等,前两种快速、有效的 DNA 测序方法于 20 世纪 70 年代中期建立,二者几乎同时出现,但存在明显的差异。链终止法是利用 DNA 聚合酶合成与模板互补的标记拷贝;而化学降解法是利用一套碱基专一的化学试剂作用于标记好的 DNA 链。两种方法开始时都很受欢迎,但是链终止法后来成为主要的测序方法,特别是在基因组测序当中。其中的原因主要有两方面,一方面是因为化学降解法中的化学试剂具有毒性,另一方面是链终止法更容易自动化。基因组计划需要大量的测序反应,手工测序将花费很多年的时间。因此在可接受的时间内完成测序计划,就必须有自动测序方法。

1. 链终止法

链终止法(The chain termination method) 又称 Sanger 双脱氧末端终止法,是由英国生物化学家弗雷德里克·桑格(Frederick Sanger, 1918—2013)建立的(图 1.1)。在利用 PCR 进行 DNA 复制过程中,脱氧核糖核苷酸(dNTPs)通过与模板碱基的特异性配对,被掺入到新合成的 DNA 链中。链的延伸是通过引物生长端的 3' - 羟基与被掺入的脱氧核糖核苷酸的 5' - 磷酸基反应形成 3',5' - 磷酸二酯键。新链由 5' 向 3' 方向延伸。链终止法测序中,在 PCR 反应体系中加入 4 种单脱氧核糖核苷酸(即 dATP, dCTP, dGTP, dTTP)的同时,还分别加入 2,3 - 双脱氧核糖核苷酸(即 ddATP, ddCTP, ddGTP, ddTTP)。在加入 ddATP 的反应中,ddATP 会随机地代替 dATP 参加反应,一旦 ddATP 加入了新合成的 DNA 链,由于其 3 位的羟基变成了氢,不能继续延伸,所以反应中所产生的 DNA 链都是到 A 就终止了;同理,当反应中分别加入 ddCTP, ddGTP 或 ddTTP 时,链的延伸分别终止于 C, G, T, 最终分别获得以 C, G, T 为末端的一系列长短不一的引物延伸链。

由此,通过合成与单链 DNA 互补的多核苷酸链,因为 2,3 - 双脱氧核糖核苷酸的加入,合成的互补链可在不同位置随机终止反应,产生只差一个核苷酸的 DNA 分子,进行变性聚丙烯酰胺凝胶电泳,读取 DNA 的核苷酸顺序。变性聚丙烯酰胺凝胶电泳能够把长度只差一个核苷酸的单链 DNA 分子区分开。这就意味着在长 10 ~ 1 500 个核苷酸的范围内,所有分子经电泳后可成为一系列可分辨的条带。链终止法测序产物的平均链长取

决于 ddNTP 与 dNTP 的比例,比例高时得到较短的产物。

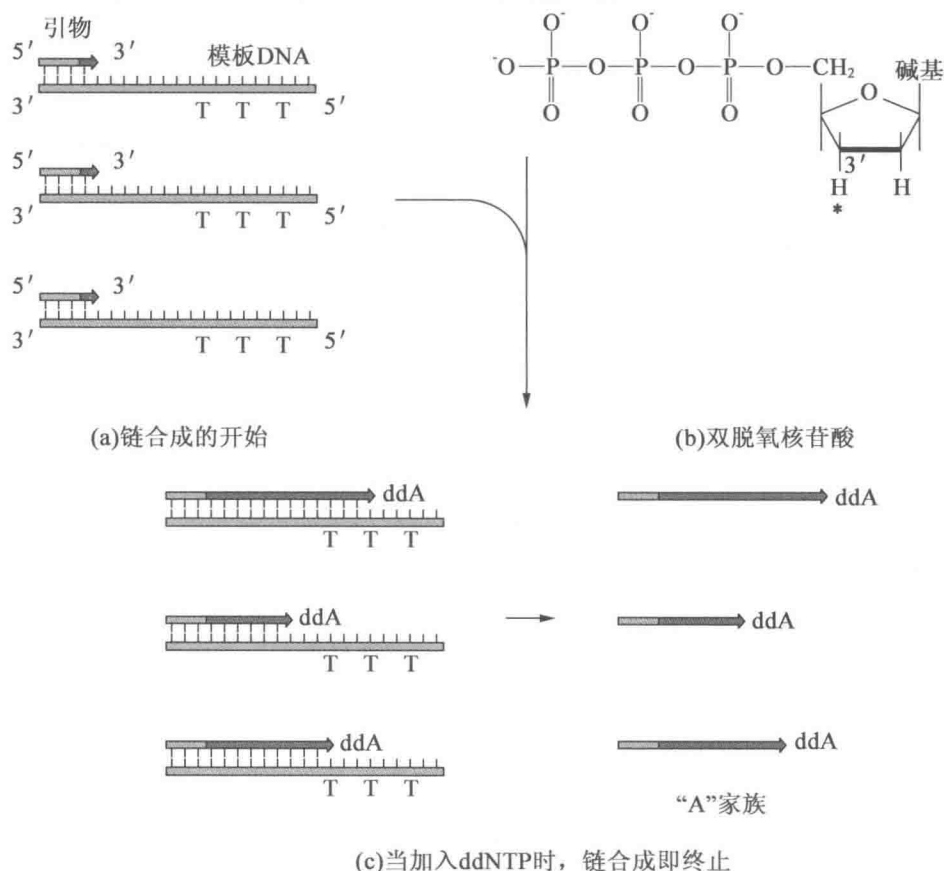


图 1.1 链终止法测序

注 (a)链终止法测序包括合成与单链模板互补的新链 DNA; (b)链合成不能持续地进行,因为体系中含有少量的 4 种双脱氧核苷酸,由于它们的 3' 碳原子上连接的是氢原子,而不是羟基,故能阻断进一步延伸; (c) ddATP 的掺入引起合成链在与模板中 T 相对的位置上终止,这就产生了“A”家族终止分子,掺入其他类型的双脱氧核苷酸就产生了“C”“G”以及“T”家族。* 代表 dNTP 中的—OH 被—H 所代替

2. 化学降解法

化学降解法又称为 Maxam - Gilbert 化学法,是 Allan Maxam 和 Walter Gilbert 于 1977 年发展起来的化学测序方法。化学降解法与链终止法的区别在于得到分子的 A, C, G, T 系列的方法不同。在化学降解法中,不是通过酶促合成,而是利用在特定核苷酸位置特异切割的化学试剂处理而得到这样的一系列分子。起始样品可以是双链 DNA,但是开始测序前必须转变为单链,并且每个分子都必须在一端标记。

其基本原理是在选定的核苷酸碱基中引入化学基团,再用化合物处理,使 DNA 分子在被修饰的位置降解。首先将双链 DNA 样品变为单链,每个单链的同一方向末端都用放射性同位素标记,以便显示 DNA 条带。接着分别用不同的化学方法修饰和裂解特定碱基,从而获得一系列长度不一的、只差一个核苷酸,且一端被放射性标记的 DNA 片段群体。这些以特定碱基为末端的片段群,通过变性聚丙烯酰胺凝胶电泳分离,再经过放