




普通高等教育农业部“十二五”规划教材

# 基因分析和 操作技术原理

JIYIN FENXI HE  
CAOZUO JISHU YUANLI

吕慧能·主编

 中国农业出版社

普通高等教育农业部“十二五”规划教材

# 基因分析 操作技术原理

吕慧能 主编



中国农业出版社

## 图书在版编目 (CIP) 数据

基因分析和操作技术原理/吕慧能主编. —北京:  
中国农业出版社, 2015. 9  
普通高等教育农业部“十二五”规划教材  
ISBN 978-7-109-20936-7

I. ①基… II. ①吕… III. ①基因工程—高等学校—  
教材 IV. ①Q78

中国版本图书馆 CIP 数据核字 (2015) 第 224335 号

中国农业出版社出版

(北京市朝阳区麦子店街 18 号楼)

(邮政编码 100125)

策划编辑 刘 梁

文字编辑 陈睿曈

---

北京中兴印刷有限公司印刷 新华书店北京发行所发行  
2016 年 6 月第 1 版 2016 年 6 月北京第 1 次印刷

---

开本: 787mm×1092mm 1/16 印张: 27.75

字数: 675 千字

定价: 39.50 元

(凡本版图书出现印刷、装订错误, 请向出版社发行部调换)

## 内 容 简 介

本教材以基因分析和操作为目的，在介绍了必要的基因和基因组基础知识与工具后，从方法与原理的角度，比较全面并且有一定深度地介绍基因与基因组研究的技术，基础的如核酸的分离纯化和特异的检测分析技术，较前沿的如新一代测序技术和蛋白质组学技术，综合性的如动、植物基因和基因组操作，同时配备了相应的教学实验。本教材适用于生物类和大农学类各专业。对于本科生，可以在老师指导下选取各章较为基础的部分进行学习；对于研究生，可以选取完整的章节进行学习。

## 编写人员名单

主 编 吕慧能 (中国, 南京农业大学)

副主编 骆 严 (中国, 浙江大学)

编 者 (按姓名拼音排序)

陈翠英 (比利时, 根特大学)

郝运伟 (中国, 军事医学科学院)

骆 严 (中国, 浙江大学)

吕慧能 (中国, 南京农业大学)

马玉良 (美国, San Diego)

钱 进 (美国, Esense Biolab)

蔚文峰 (中国, 军事医学科学院)

吴松锋 (中国, 军事医学科学院)

应万涛 (中国, 军事医学科学院)

张养军 (中国, 军事医学科学院)

朱嘉明 (中国, 暨南大学)



毫无疑问，基因和基因组研究是最近半个多世纪生命科学中研究最活跃、进展最快的一个领域，已成为现代生物学的基础之一。2014年《Nature Methods》10周年纪念刊中选出的10年十大技术中的8项：新一代测序技术、基因组工程技术、单分子技术、以质谱为基础的蛋白质组学技术、结构生物学技术、细胞重编程技术、光遗传学技术和合成生物学技术都是以基因分析和操作为基础，或与之交叉紧密结合的技术。基因和基因组研究在相当长的一段时间里会一直作为生命科学的中心课题。多个学科科学家参与了基因和基因组的研究，逐步形成了现在轮廓较为完整的基因技术体系。

从基因组学的诞生至今，研究历史并不长，然而，无论是在基础理论研究领域，还是在生产实际应用方面，都已经取得了惊人的成绩，给生命科学研究带来了深刻的变化。包括植物转基因在内的基因操作技术是解决人口增长和环境恶化导致的饥饿与粮食安全问题的有力手段。作为生命科学的基础研究和应用研究人员，了解和学习基因分析和操作技术是必需的。

基因和基因组领域的研究技术极其繁多，介绍研究技术的实验手册也相当丰富。其中最为著名的当数《Molecular Cloning》和《Current Protocols in Molecular Biology》，分别被誉为分子生物学实验技术的“圣经”和“红宝书”。这两本书合起来基本上包罗了本领域最主要的实验技术，内容浩繁，是极有价值的参考书。此外，各种资料还散见于网上文档、BBS等。各种技术手段不但信息量大，而且在不断更新，旧的技术逐渐被淘汰。对于初学者而言，要在短时间内全面掌握如此巨量的实验技术显然是不可能的。为了能较好地引导学生入门，同时对初进实验室的操作人员也有所帮助，我们在吕慧能根据多年授课经验编写的讲义基础上进行扩展和充实，组织编写了本教材。除了必要的基础知识外，希望把那些真正好用的，接近或处在前沿的，既是基础性的又有未来发展前景的技术原理和方法介绍给大家，使初学者能在较短时间里了解和掌握这个领域的技术全貌，并且达到一定深度。

吕慧能统筹全书编写并整理书稿，各章节的编写人员如下：

## 第一部分 理论和技术原理

第一章：吕慧能、骆严、马玉良、钱进。

第二章：吕慧能、骆严、马玉良、朱嘉明、钱进、陈翠英。

第三章：吕慧能、马玉良、朱嘉明、钱进、骆严。

第四章：吕慧能、钱进、马玉良、骆严。

第五章：吕慧能、马玉良、钱进、骆严。

第六章：吕慧能、钱进、骆严、马玉良。

第七章：吕慧能、骆严、马玉良、钱进。

第八章：吕慧能、马玉良、骆严、钱进。

第九章：马玉良、吕慧能、骆严、钱进。

第十章：应万涛、张养军、吴松锋、郝运伟。

第十一章：钱进、马玉良、骆严、陈翠英、吕慧能。

第十二章：吕慧能、陈翠英、朱嘉明。

## 第二部分 实验操作

### 第一单元 基本技术

实验一至实验十三、实验十五：吕慧能、朱嘉明、马玉良、骆严、钱进。

实验十四：蔚文峰、应万涛。

### 第二单元 综合实验

综合实验一：吕慧能、马玉良、骆严。

综合实验二：吕慧能、陈翠英。

综合实验三：钱进、骆严、马玉良。

附录：吕慧能、陈翠英、骆严、马玉良。

基因和基因组研究技术所涉及的学科范围很广，要在较短时间里，一定深度地了解和掌握这些技术困难不少。本教材的编写是一个尝试。由于时间仓促，编者知识水平有所不逮，虽然编者尽了力，但仍会有不少地方存在疏漏，甚至错误。在此抛砖引玉，希望得到更多老师和同学的指正，使教材在使用和未来更新过程中不断完善。

本教材之所以能成书，首先要感谢那些为教材编写提供原始研究结果的研究人员，是他们开创性的探索成果成就了基因分析和操作技术体系。本教材参考了很多原始研究文献和其他可搜索到的资料。教材里的插图除了注明引用的以外，还有一部分是参考了原始文献或较为原始的综述和参考书而绘制的。由于篇幅的限制未能将参考文献一一详细标注，在此对这些资料的作者深表歉意。

感谢林海帆教授审阅并对 RNA 基因部分的编写提出了修改意见。

感谢骆严教授指导下的学生对本教材的认真校对和编辑。纪冰琰、丘楚平、钱琨浔、成梦、刘炳旭、熊嘉莉、曹致宁、张学云依序分别校对了第一至第八章。张梦雅和孙艳娣同时校对了第十一章，还分别校对了除实验十四、实验十五和综合实验三外的所有实验操作初稿。

还要感谢贺福初院士对教材编写的帮助和支持，感谢王琰女士在教材编写中帮助联系、协调北京的编者，以及陆红缨女士在教材编写过程中的资料文献服务工作。许多编者的同学与编者就本书内容进行了讨论和交流，使编者受益匪浅，在此表示衷心的感谢。

编 者

2015 年 7 月





## 前言

## 第一部分 理论和技术原理

第一章 基因和基因技术基础 .....	2	第五章 质粒载体 .....	153
第一节 基因和基因技术的发展 .....	2	第一节 质粒生物学和质粒载体 .....	153
第二节 基因和基因组基础 .....	10	第二节 质粒载体的功能元件 .....	158
第三节 生物信息学工具 .....	37	第三节 表达载体 .....	165
第四节 基因操作中的模式生物 .....	47	复习题 .....	174
复习题 .....	50	第六章 噬菌体衍生的载体 .....	175
第二章 基本技术和工具 .....	51	第一节 $\lambda$ 噬菌体 .....	175
第一节 DNA/RNA 的提取 .....	52	第二节 $\lambda$ 噬菌体载体 .....	178
第二节 电泳技术 .....	61	第三节 $\lambda$ 噬菌体衍生的质粒载体 ——柯斯质粒 .....	181
第三节 特异的检测技术 .....	67	第四节 单链 DNA 载体 .....	183
第四节 工具酶 .....	82	第五节 P1 噬菌体衍生的载体 .....	186
第五节 遗传转化 .....	88	复习题 .....	189
复习题 .....	89	第七章 基因克隆 .....	190
第三章 离体基因扩增 .....	91	第一节 基因组 DNA 文库 .....	191
第一节 聚合酶链式反应 .....	91	第二节 cDNA 文库 .....	195
第二节 PCR 反应的组成和设计 .....	93	第三节 克隆筛选的策略 .....	201
第三节 PCR 反应循环及参数 .....	98	复习题 .....	215
第四节 PCR 中的几个问题 .....	99	第八章 DNA 测序 .....	216
第五节 各种 PCR 扩增及其 应用 .....	102	第一节 双脱氧终止法 DNA 测序 .....	218
第六节 其他核酸扩增技术 .....	117	第二节 DNA 测序的自动化 .....	221
复习题 .....	123	第三节 基因组测序 .....	223
第四章 重组技术 .....	124	第四节 新一代测序技术 .....	228
第一节 特异的 DNA 内切酶及其 特点 .....	125	第五节 测序在研究中的应用 .....	233
第二节 连接和重组 .....	134	复习题 .....	235
第三节 同源重组技术 .....	145	第九章 离体突变和基因创建 .....	236
复习题 .....	151	第一节 DNA 和基因的化学合成 .....	236

第二节	离体突变 .....	237	第二节	毕赤酵母和昆虫细胞 表达系统 .....	297
第三节	分子进化 .....	252	第三节	动物细胞培养和转染 .....	300
复习题 .....	259	第四节	转基因动物 .....	313	
<b>第十章 蛋白质组学技术原理 .....</b>	<b>260</b>	第五节	动物基因改造的应用 .....	316	
第一节	蛋白质样本制备 .....	260	复习题 .....	319	
第二节	复杂样本的蛋白质分离 .....	263	<b>第十二章 植物转基因 .....</b>	<b>321</b>	
第三节	蛋白质酶解及肽段分离 .....	266	第一节	转基因载体和外源基因 导入技术 .....	322
第四节	肽混合物质谱分析 .....	269	第二节	植物转基因的选择和 筛选标记 .....	332
第五节	定量蛋白质组学 研究技术 .....	274	第三节	转基因结构和表达 .....	335
第六节	蛋白质组学数据分析技术 .....	280	第四节	基因开发和转基因植物 .....	342
复习题 .....	290	复习题 .....	348		
<b>第十一章 酵母和动物基因操作 .....</b>	<b>291</b>				
第一节	酿酒酵母基因操作 .....	292			

## 第二部分 实验操作

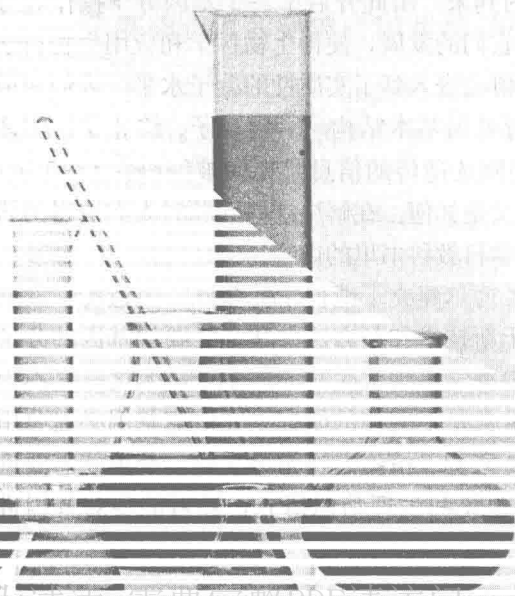
<b>第一单元 基本技术 .....</b>	<b>352</b>	实验十	DNA 片段的分离纯化 .....	376	
实验一	生物信息学实验: 基因检索 和表达克隆设计 .....	352	实验十一	DNA 末端的连接 .....	377
实验二	实验室基本操作 .....	355	实验十二	感受态细胞的制备 .....	378
实验三	大肠杆菌的培养和分离 .....	360	实验十三	大肠杆菌的遗传转化 .....	379
实验四	蛋白质 SDS-PAGE 电泳 .....	362	实验十四	蛋白质二维电泳分离 及胶内蛋白质质谱 分析 .....	380
实验五	质粒 DNA 的分离 .....	366	实验十五	SSR 分子标记 .....	385
实验六	基因组 DNA 提取 .....	367	<b>第二单元 综合实验 .....</b>	<b>389</b>	
实验七	总 RNA 的提取 .....	371	综合实验一	EGFP 蛋白质的 克隆表达 .....	389
实验八	聚合酶链式反应 (PCR) 扩增和电泳观察 .....	372	综合实验二	植物转基因 .....	394
实验九	质粒 DNA 酶切及凝胶 电泳 .....	374	综合实验三	动物细胞基因操作 .....	398
<b>附录 .....</b>	<b>403</b>				
一、常用生物信息学网址 .....	404	四、基因的命名 .....	408		
二、常用核酸、蛋白质数据 .....	406	五、糖类的结构 .....	409		
三、度量衡 .....	407	六、MS 培养基 .....	410		

参考文献 .....	411
索引 .....	413

# 第三章 好因基味因基 第一版

## 第一部分

# 理论和技术原理



# 第一章 基因和基因技术基础

## 第一节 基因和基因技术的发展

### 一、基因、基因分析和操作技术

绝大部分生物基因的载体是 DNA，一些病毒是 RNA。基因是一段有功能或表型效应的 DNA 或 RNA 序列，通常基因编码 RNA 或多肽分子。基因组指一个生命体的所有遗传物质。对于绝大部分生物来说，它是细胞内正常遗传的 DNA 分子总和。对于一些病毒来说，基因组是正常遗传的 RNA 分子总和。一个基因组的基因数可以少至几百个（类菌体约 500 个），也可以多至数十万个。据估计，人类本身的基因数为 2 万~2.5 万，这也是哺乳动物的基因数大概范围。高等植物基因组估计的基因数有较大的变化范围。水稻的基因数估计 3 万左右，小麦估计有 94 000~96 000 条基因。

人们对基因的正确认识最早要追溯到 1866 年孟德尔（Gregor Johann Mendel）提出的抽象、颗粒式的遗传因子。1909 年，丹麦科学家 W. Johansen 提出基因（gene）这个名词。经过半个多世纪的探索，在 20 世纪中期，人们终于将抽象的遗传因子落实到了细胞核内的脱氧核糖核酸（DNA）这一具体的物质上。在这半个多世纪的历程中，首先将基因和具体细胞结构联系起来的是摩尔根（Thomas Hunt Morgan）及其研究团队的遗传的染色体理论（1902 年，遗传的染色体理论；1911 年，基因在染色体上线性排列）。他们将基因落实到染色体这个具体的细胞结构上。1941 年，有了“一个基因一个酶”的假说；通过遗传转化试验（Oswald Theodore Avery 等，1944）和噬菌体感染试验（Alfred Hershey 和 Martha Chase，1952），人们确信 DNA 是我们要寻找的基因的物质载体。差不多用了半个世纪，人们认识到如果要研究基因，就要对 DNA 进行研究。也就是说，从此基因的研究确定了具体的研究物质对象。20 世纪 50 年代，双螺旋模型的建立标志着一个新的时代的到来，由此开启了一个基因分子操作和分析技术带动的基因和基因组研究迅速发展的年代。它们的发展，使得生物科学和应用生物科学研究迎来了革命性的变化。人们渴望的解码生命的研究进入到了实质性的分子水平。人们知道了一系列关于基因结构的基本概念，知道需要进行基因基本结构——启动子、终止子、阅读框、内含子、外显子和密码子等——的研究；需要阐述遗传的信息是如何被编码在 DNA 上，它们又是如何表达的；转录、翻译的具体分子过程又是如何。在解决这些问题时人们取得的研究结果迅速成为遗传学、生物化学、分子生物学等学科教科书里的内容。

从遗传学和现代生物学发展历史可以看出：孟德尔建立了遗传学的基础，但遗传学发展和突破的关键却常常在于技术和方法的突破，在于新技术和新方法的建立。要说基因的分子分析和操作技术已成为遗传学、生物化学、分子生物学等现代生物学的主干技术之一毫不为过。在现代生物学基础研究中，没有基因的研究常常是不完整的研究。要研究基因必然要用到基因分析和操作技术。不管研究的是人类、家禽家畜、植物，还是微生物，都可能需要进行基因的检测、分离（克隆），进行基因的结构、生化和生理功能研究，还可能要对基因进

行改造以更好地为我们所用,开发和生产基因产品,设计、改造生命体等,所有这些都需要一些通用的基因分析和操作技术,这些技术已渐成规模和体系。虽然它们是多学科交叉发展的结果,但也已形成了一个相对独立和完整的知识体系,有自己的发展目标。我们需要对它们进行系统的归纳和整理。

这些技术覆盖面之广使得几乎没有人能在一两个学期内把它们都做一遍。但在生命科学和应用生命科学研究中了解和认识这些技术又是必需的。了解了它们的方法、技术原理和技术设计的思路,才能更好地开发和创新基因和基因组分析与操作技术,在实践中应用并改进它们,更好地进行基因和基因组的基础研究和应用开发研究。

遗传学是研究生物的遗传和变异的规律的学科。分子生物学是从分子水平研究、探讨生命活动,尤其是生命基本活动的分子机制和分子过程。生物化学研究生命体内的化学反应,研究生命新陈代谢的具体过程。它们都和基因的研究有关。这些课程都从各自的角度介绍了基因研究的成果。它们主要或首先要介绍的是这些学科研究的结论和结果。和它们不同,基因分析和操作技术原理要介绍的是基因研究的技术和方法,介绍使生命科学研究获得突破性进展的技术和方法。也就是说,基因分析和操作技术原理要介绍的是基因研究的过程。

基因和基因组分析、操作技术包括了多个学科的多种技术,它们今后的发展也将是如此。它们的发展体现了现代科学多学科交叉发展的特征,一定程度上代表了生物学的发展水平。留意一下的话,这个领域是获诺贝尔奖很多的一个领域。有的是化学奖,有的是生理学或医学奖。但它们包含共同的概念:基因、基因组的分析和操作。

在现代基因研究中,理论上的三大发现消除了人们头脑中的疑惑,并指出了研究的基本方向:证明基因是DNA使基因的分析操作有了具体的目标,基因工程有了理论先导;DNA的双螺旋结构和半保留复制机制解决了基因遗传和突变的机制和基础问题,进一步具体地确定了理论和技术研究的研究方向;遗传信息传递的中心法则使人们有了对基因及其活动研究的具体层次和入手点。

现在,这些理论上的突破使对基因的研究触手可及。可是,如果回到20世纪的六七十年代,你会发现研究的技术是研究取得进展的关键。在取得理论上的基本突破后,研究的进展更多地取决于技术的进步。回顾一下基因和基因技术发展的历史,或许会给我们新的启发。

## 二、基础积累和突破:20世纪70年代以前

1868年,瑞士科学家Friedrich Miescher发现分离的细胞含有特异的磷化合物,取名为“nuclein”;1871年,又从精子头部分离到一种酸性化合物(今天的核酸)和一种碱性化合物的复合物“protamine”。1889年,Richard Altman发表一种制备无蛋白质的核酸(nucleic acid)的方法,认识到核酸总是伴随着蛋白质存在。19世纪的这些生物化学的研究和遗传学、基因的概念是分隔、不连接的。

首先将基因和细胞内细胞结构联系起来的是Thomas Hunt Morgan和他的研究团队:他把基因落实到具体的细胞结构染色体上,并证明基因在染色体上是线性排列的。

1941年,遗传学家George Beadle和生物化学家Edward Tatum通过对营养缺陷突变体的研究,提出了“一个基因一个酶”的理论。尝试着将生物化学的研究和遗传学、基因的研究联系起来。而真正在这方面取得突破的是1944年的Oswald T. Avery的肺炎双球菌的转化试验,证据确凿地说明DNA是遗传物质。但是,由于当时仍有少数科学家固执地认为蛋

白质才是遗传密码的承受者，而 DNA 只是遗传物质赖以存在的“框架”，Avery 的结论未能马上得到普遍的肯定。最终使所有人消除怀疑，肯定“DNA 是遗传物质”的是 1952 年 Alfred Hershey 和 Martha Chase 的 T2 噬菌体感染实验。他们用<sup>32</sup>P 标记噬菌体 DNA、<sup>35</sup>S 标记噬菌体的蛋白质外壳，在产生的后代的噬菌体颗粒中，发现子代噬菌体含有的是<sup>32</sup>P 标记的 DNA，而不是<sup>35</sup>S 标记的蛋白质。再次证明了遗传给后代的是 DNA，而不是蛋白质。

诺贝尔奖小知识：1969 年诺贝尔生理学或医学奖授予了发现病毒复制和遗传机制的 3 位科学家 Max Delbrück、Alfred D. Hershey 和 Salvador E. Luria。此时，Avery 已不在人世了，因而错过了诺贝尔奖。

1951 年，Erwin Chargaff 发现 DNA 中 4 种碱基 A、C、G、T 组成的规律：A : T 和 C : G 总是 1 : 1。在这些背景下，加上 Rosalind Elsie Franklin 实验室拍摄的清晰的 DNA 的 X 射线衍射照片所表明的结构，1953 年，James Watson 和 Francis Crick 提出 DNA 双螺旋模型，它成了现代生物学和基因研究重要的里程碑。它回答了 DNA 作为遗传物质如何进行自我复制和信息传递功能的问题。生物科学的一个新纪元诞生了。

诺贝尔奖小知识：1962 年诺贝尔生理学或医学奖授予了提出 DNA 双螺旋模型的科学家 Francis Crick、James Watson 和 Maurice Wilkins。Crick 和 Watson 在剑桥大学卡文迪许实验室进行 DNA 结构的研究，同为研究生物大分子结构的物理学家 Wilkins 和 Rosalind Elsie Franklin 在伦敦的伦敦国王学院 (King's College London)。在研究的关键阶段，Watson 遇到 Wilkins，Wilkins 给 Watson 看了由 Rosalind Elsie Franklin 的研究生 Raymond Gosling 拍摄的 DNA 的 X 射线衍射照片（史称 51 号照片，Photo 51）。此 X 射线衍射照片清楚地表明，DNA 至少有 2 条链，直径约为 2.2 nm，0.34 nm 一个基团，3.4 nm 一个重复单位。此照片最终使 Crick 和 Watson 提出了著名的 DNA 双螺旋模型。虽然 Franklin 在 DNA 结构的研究中贡献卓著，但在授予诺贝尔奖的 1962 年，她已经不幸患乳腺癌去世了。根据诺贝尔奖规则，1962 年的奖授予了还在人世的 Francis Crick、James Watson 和 Maurice Wilkins。

1956 年，Arthur Kornberg（1959 年被授予诺贝尔生理学或医学奖）在 *E. coli* 中发现并分离了 DNA 聚合酶 I。这是人类分离到的第一个 DNA 聚合酶。1958 年，Matthew Meselson 和 Franklin Stahl 通过试验证明了 DNA 的半保留复制的机制，从而结束了关于半保留复制（Watson 和 Crick 提出双螺旋模型时提出的）还是全保留复制（两条链作为模板合成新的两条链组成的子代双螺旋）的争论。同年，Crick 提出中心法则，它最终使人们清楚地知道遗传信息的传递方式。

1959—1960 年，Severo Ochoa（和 Kornberg 分享 1959 年诺贝尔生理学或医学奖）发现 RNA 聚合酶和信使 RNA，证明 mRNA 决定了蛋白质分子中的氨基酸序列。1961 年，Marshall Warren Nirenberg 等破译了第一个遗传密码：他们在大肠杆菌提取液中，借助于添加合成的多聚尿嘧啶核苷酸而合成出聚苯丙氨酸多肽，这说明苯丙氨酸的密码子是 UUU。此后多年，经过多位科学家（Robert W. Holley，H. Gobind Khorana，Marshall W. Nirenberg，S. Ochoa，F. Crick 等）的努力，到 1966 年，遗传密码全部被破译。Robert W. Holley、H. Gobind Khorana 和 Marshall W. Nirenberg 由于他们在破译密码子中的贡献获得了 1968 年诺贝尔生理学或医学奖。

1960 年，François Jacob 和 Jacques Monod 提出了调节基因表达的操纵子模型。他们和发现噬菌体感染及生长机制的科学家 André Michel Lwoff 一起分享了 1965 年的诺贝尔生理

学或医学奖。

进入 20 世纪 70 年代后, 基因技术的发展基础基本建立, 进入了具体技术开发和应用的时期。一系列的工具酶被发现并迅速地应用于基因和基因技术的研究中。1970 年 Hamilton O. Smith 和 K. W. Welcox 从流感嗜血杆菌 (*Haemophilus influenzae*) 中分离到一种限制酶, 能够特异性地切割 DNA, 这个酶后来命名为 Hind II。这是分离到的第一个 II 类限制性内切酶。此后, 发现了包括来源于大肠杆菌的 EcoR I 和 EcoR II 以及来源于流感嗜血杆菌的 Hind III 等。1970 年, H. M. Temin 和 D. Baltimore 从 RNA 肿瘤病毒中发现反转录酶。

诺贝尔奖小知识: 1975 年度诺贝尔生理学或医学奖授予了发现致癌病毒和反转录酶的 3 位科学家 David Baltimore、Renato Dulbecco 和 Howard Martin Temin。1978 年诺贝尔生理学或医学奖授予了发现 (II 类) 限制性内切酶并应用它们于分子遗传学研究中的三位科学家 Werner Arber、Daniell Nathans 和 Hamilton O. Smith。

### 三、里程碑

除了上面所述的 DNA 双螺旋模型等重大突破外, 在基因和基因技术发展史上还有一些值得我们记忆的重要事件。

#### (一) 重组 DNA 技术的诞生

1972 年, 美国斯坦福大学的 Berg 获得了 SV40 和  $\lambda$  噬菌体 DNA 重组的 DNA 分子。1973 年, 美国斯坦福大学的 Stanley N. Cohen (另外有一位叫 Stanley Cohen 的生物化学家和同事 Rita Levi-Montalcini 因发现生长因子分享了 1986 年的诺贝尔生理学或医学奖) 与加州大学 Herbert W. Boyer 等发展了 DNA 重组技术, 获得第一个有新功能的重组 DNA 分子和第一个基因工程的细菌基因克隆。当时 Stanley N. Cohen 有很好的质粒载体和大肠杆菌转化技术, 但缺乏对质粒进行特异性切割的限制性内切酶。在一次学术会议上, 遇到了做报告介绍 EcoR I 限制性内切酶的 Herbert W. Boyer。于是, 他们很快地走到了一起并迅速地取得了成功。他们用 EcoR I 酶切大质粒 R6-5 后转化大肠杆菌, 挑选一个具有卡那霉素抗性的克隆, 此中的质粒记为 pSCI02, 再用 EcoR I 酶切后发现 3 个片段。将大肠杆菌的抗四环素 ( $Tc^R$ ) 质粒 pSCI01 和 pSCI02 用 EcoR I 酶切, 混合后进行连接酶连接处理或不经连接酶连接处理, 两种混合物转化大肠杆菌, 在含四环素和卡那霉素的平板中, 选出了抗四环素和抗卡那霉素的重组菌落 (连接处理后的克隆数是未连接处理的 8 倍)。挑选其中一个克隆进行分析, 发现此中的质粒 (记为 pSCI05) 用 EcoR I 酶切后有一条和 pSCI01 相同的片段, 此外还有一条和 pSCI02 用 EcoR I 酶切切出的 3 个片段中的一个片段相同的片段。类似地, 他们通过离体重组还获得四环素和链霉素双抗的重组质粒 pSCI09。他们的论文一发表, 人们便迅速地认识到一项新的产业——基因工程从此诞生了, 这一年被定为基因工程诞生的元年。1974 年, Cohen 与 Boyer 等合作, 将非洲爪蟾编码核糖体 RNA 结构基因同 pSCI01 质粒构成重组 DNA 分子, 导入大肠杆菌。发现在大肠杆菌中有非洲蟾蜍的核糖体 RNA 结构基因的表达。实现异源真核生物基因在 *E. coli* 中的表达。

1976 年, 时年 29 岁的风险投资家 Robert Swanson 敏锐地捕捉到离体 DNA 重组技术 (基因工程关键技术) 的商业机会。他邀请 Herbert Boyer 一起喝咖啡, 热情的鼓动使 Boyer

也为之振奋，同意合伙成立 Genentech (Genetic Engineering Technology) 公司，共同探索重组 DNA 的商业化之路。如今，该公司是最为成功的现代生物工程公司之一。

## (二) DNA 测序技术诞生

1975 年，Frederick Sanger 建立了加减法 DNA 测序技术。它主要是利用 DNA 聚合酶的 3'→5' 的外切酶活性和 5'→3' 聚合酶活性，在不同条件下得到特定碱基结尾的片段，从而测得 DNA 的序列。1977 年，测得第一个基因组：5.38 kb 的大肠杆菌噬菌体  $\Phi$ X174。但很快地，加减法便被更可靠和更易进行大规模测序的 Walter Gilbert 和 Allan Maxam (1977) 发明的特异断裂法和 Sanger (1977) 发明的双脱氧链终止法所替代。Gilbert 和 Sanger 因而与得到第一个离体重组 DNA 分子的 Berg 分享了 1980 年诺贝尔化学奖。Sanger 成了少数两次获诺贝尔奖的科学家之一。

1982 年，Sanger 和同事完成  $\lambda$  噬菌体基因组 48 502 bp 全序列的测定。全基因组测序成为了当时的测序目标。1990 年，以美国为主，全球合作的人类基因组计划正式启动。1996 年前后，作为大规模测序技术建立的试验，先完成了酵母基因组 ( $1.25 \times 10^7$  bp) 等模式生物全基因组序列测定。2000 年，人类基因组工作框架图完成；2003 年，人类基因组测序工作完成。重要粮食作物——水稻——的基因组草图和完整全基因组序列分别于 2001 年和 2005 年完成测定。

到 20 世纪末，在人们大规模应用 Sanger 双脱氧链终止法的同时，认识到为了提高测序的通量，降低测序成本到 1 000 美元/人基因组 (2014 年有公司宣布在测序成本上实现了这个目标) 需要开发新的测序技术。于是，新一代的测序技术便产生了。这些新一代测序技术能以极高的速度产生大量的测序数据，称为深度测序，又分别称为第二代测序技术和第三代测序技术。它们的目标之一是使测序技术能够进入临床，为建立精准医疗方案提供基本的个人基因组数据等。新一代的测序技术的开发将给我们的研究和应用带来新的机会，我们对生物界基因和基因系统的了解将到达一个前所未有的准确和全面的程度。以它为代表和起始，一个在各层次上全面地刻画生命系统的时代已经来临。基因组水平、转录组水平、蛋白质组水平、代谢组水平等的全景研究成了生命科学研究最为活跃的领域。虽然很难预测它们研究的进展如何取得重大突破，但无疑生命研究的框架正在重塑。

基因组测序的一个重要的特点是：它无意之中加强了科学研究的国际合作。其他的学科 (如空间科学) 有的更多的是国际竞争，但基因组测序则更多的是国际合作。全球实验室以一种从未有过的合作精神交流和共享各自的研究结果。

## (三) 特异的基因检测技术的建立

最早代表分子生物学或基因研究水平的一项检测技术是 1975 年由 Edwin Mellor Southern 建立的 DNA 印迹检测技术，这项技术以发明者的姓氏命名，它便是几乎为所有基因研究者所熟知的 Southern 印迹技术。当时，人们对核酸的变性-复性动力学进行了较好的研究，限制性内切酶得到了发展和应用，各种载体系统有了一定的开发，因而方便了特异探针的制备。在这些基础上，建立这个特异的印迹检测技术是顺理成章的事。它的意义和应用是持久和广泛的。此后的一系列其他的检测技术和它的原理都有相通之处。1977 年，建立了 RNA 印迹检测技术被称为 Northern 印迹检测技术。尔后，蛋白质、脂类的特异检测技术也沿用这种命名被称为 Western 印迹、Eastern 印迹等。作为 Southern 印迹检测技术的发展，1982 年，建立了由 378



个探针组成的 DNA 微阵列（类似于基因芯片），用于肿瘤细胞表达基因的筛选。此后，数万个探针被同时用来进行基因的全景表达研究，检测所有已知的蛋白质基因和 RNA 基因的 RNA 表达情况。相似地，还发明了蛋白质芯片对表达的蛋白质进行高通量的检测和分析。

1983 年，美国 PE-Cetus 公司人类遗传研究室的 Kary B. Mullis 等发明了具有划时代意义的 DNA 特异扩增技术——PCR 技术。1988 年，PE-Cetus 公司推出了第一台 PCR 热循环仪，从而使 PCR 技术的自动化成为现实。PCR 技术为基因的特异检测和其他分析操作提供了非常方便、有效的的基本技术。由它可以开发出方便有效的检测技术，也可以建立不是以文库为基础的基因克隆技术（此前的克隆都是以文库为基础的）。Mullis 因此与建立定点诱变技术的 Michael Smith 分享了 1993 年诺贝尔化学奖。

这些技术上的突破和进展，在应用上的反映是 20 世纪 70 年代后期和 80 年代初期人胰岛素、人  $\alpha$  干扰素、人白细胞介素、人促红细胞生成素、人粒细胞集落刺激因子和延迟成熟的转基因番茄等一系列基因新产品的开发和上市。

#### （四）具有催化功能的 RNA 和 RNA 干扰现象的发现

随着基因研究的纵深发展，分子水平的生命活动的立体画面逐渐呈现出来。Sidney Altman 和 Thomas R. Cech 在研究 tRNA 和 rRNA 剪接时，分别发现了一些内含子具有自我催化剪接功能，催化剪接的活性中心是 RNA。这些具有催化功能的 RNA 在更多的场合被发现，被称为核酶（ribozyme）。它们的发现使人们对生命的进化有了更深入的认识：生命很可能起源于 RNA 世界。早期生命编码基因和催化生化反应很可能同时由一种大分子——RNA——所承担。随着生命的进化，编码遗传信息和催化生化反应才分别由更为合适和高效的 DNA、蛋白质所承担。由于他们的发现具有重大意义，因此获得了 1989 年诺贝尔化学奖。

Andrew Z. Fire 和 Craig C. Mello 在研究用外源反义 RNA 调控线虫的基因表达时，发现双链 RNA 能高效地介导同源 RNA 的降解。他们将这种现象称为 RNA 干扰。他们的发现提示了一种新的基因表达调控的途径，也为基因表达调控和功能研究提供了新的技术途径。因此，他们获得了 2006 年诺贝尔生理学或医学奖。

#### （五）转基因动植物的开发上市

1983 年，几个独立的团队集中宣布了转基因植物成功的报道。从植物基因工程的角度，第一个获得转基因植株的是比利时的 Jeff Schell 和 Marc van Montagu 团队（农杆菌-原生质体共培养法转基因）和美国 Monsanto 公司的 Robert Fraley、Stephen Rogers 和 Robert Horsch 等（农杆菌-原生质体共培养法和叶圆盘法转基因）。1989 年，转基因抗虫棉获批准进行田间试验。1992 年，我国批准了世界第一例商品化转基因植物——抗病烟草——的种植。1994 年，第一个在市场销售的转基因植物产品——Calgene 公司的 Flavr Savr 番茄——在美国上市。它是聚半乳糖醛酸酶（polygalacturonase, PG）反义 RNA 转基因的延熟番茄。其果实能延长货架期。它意味着第一批基因工程植物产品的上市。

在 DNA 技术发展的同时，细胞操作技术也取得了突破，使得它们的结合有了更广泛的基础。1981 年，R. D. Palmiter 和 R. L. Brinster 等获得第一只转基因动物（鼠）；A. C. Spradling 和 G. M. Rubin 得到转基因果蝇。1989 年，Mario R. Capecchi、Martin Evans 和 Oliver Smithies