



普通高等教育“十一五”国家级规划教材

基因组学

第2版

杨金水 编著



高等教育出版社
Higher Education Press

普通高等教育“十一五”国家级规划教材

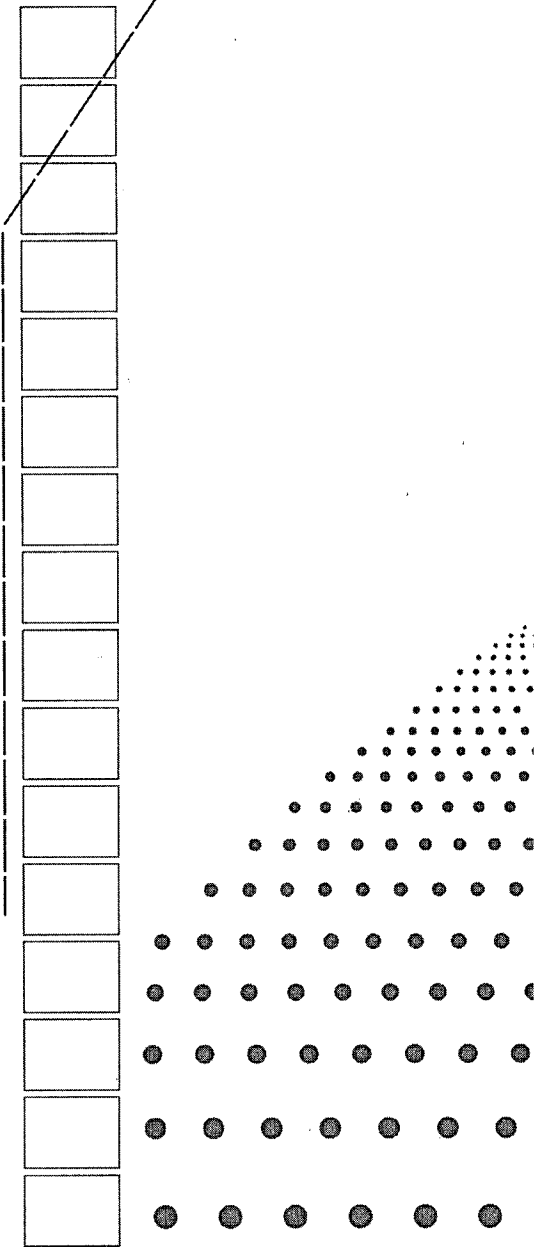
基因组学

第2版

杨金水 编著



高等教育出版社



内容提要

为了适应基因组学蓬勃发展的现状,及时反应基因组学领域近年来所取得的重要成果与重大进展,作者在《基因组学》(第一版)的基础上,对已有的内容进行了扩充与调整。再版《基因组学》仍然保持原书的结构框架,全书共分14章。第1章至第5章主要涉及基因组的结构及其测序的路线与方法。第6章至第10章着重介绍基因组的功能及其研究的策略。第11章至第14章讲述基因组的进化及其进化的机制。《基因组学》(第一版)第10章“染色质结构与基因表达”现改为“基因组表观遗传”,并相应地增补了与之有关的内容。

本书可作为高等院校生命科学类高年级本科生和研究生教材,也可供从事生命科学研究的人员作为基因组学基础知识和相关研究的参考书。

图书在版编目(CIP)数据

基因组学/杨金水编著. —2版. —北京:高等教育出版社,2007.7

ISBN 978-7-04-021462-8

I. 基… II. 杨… III. 基因组-高等学校-教材
IV. Q343.2

中国版本图书馆CIP数据核字(2007)第070806号

策划编辑	吕庆娟	责任编辑	田 军	封面设计	张 楠	责任绘图	朱 静
版式设计	陆瑞红	责任校对	张 颖	责任印制	尤 静		

出版发行 高等教育出版社
社 址 北京市西城区德外大街4号
邮政编码 100011
总 机 010-58581000

经 销 蓝色畅想图书发行有限公司
印 刷 北京铭成印刷有限公司

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landraco.com>
<http://www.landraco.com.cn>
畅想教育 <http://www.widedu.com>

开 本 850×1168 1/16
印 张 24.5
字 数 610 000

版 次 2002年6月第1版
2007年7月第2版
印 次 2007年7月第1次印刷
定 价 38.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 21462-00

再版前言

截至 2006 年 11 月 1 日,国际上已完成测序的基因组为 456 个,其中古细菌 29 个,真细菌 384 个,真核生物 44 个。正在测序的古细菌 56 个,真细菌基因组 995 个,真核生物基因组 632 个,环境微生物基因组 63 个,已完成和正在进行的基因组测序计划总共为 2 202 个。由于 DNA 测序方法的改进,基因组测序计划的数目正在迅速地增加。与此同时,基因组注释与功能基因组的研究也呈现百花齐放的局面。基因组学的研究已深入到生命科学的各个领域,正在深刻地影响着生命科学未来的发展方向。

为了适应基因组学蓬勃发展的现状,及时反应基因组学领域近年来所取得的重要成果与重大进展,作者在《基因组学》(第一版)的基础上,对已有的内容进行了扩充与调整。再版《基因组学》仍然保持原书的结构框架,仅将第一版第 10 章“染色质结构与基因表达”改为“基因组表观遗传”,并相应地增补了与之有关的内容。虽然各个章节都在原有内容的基础上作了更新,但在基因组注释和基因组进化方面增添了更多的新的内容,这也是近年来基因组学研究中最令人瞩目的领域。

再版《基因组学》仍将重点放在一些基本概念的阐述上。由于篇幅有限,许多研究进展只能简要提及,读者如有兴趣可以查阅相关的参考文献。

本书的再版编写工作得到了复旦大学生命科学院各方面的支持与帮助。作者实验室的洪芳同学、王玉锋同学和查笑君同学在本书的文字校对、文献查阅以及编写内容安排等方面做了大量工作,在此表示衷心感谢。

杨金水

2006 年 11 月 8 日于复旦大学

第一版前言

在人类基因组计划的影响下，分子生物学的主要目标已经从传统的单个基因的研究转向对生物整个基因组结构与功能的研究。生命科学正从全新的视角研究与探讨生长与发育、遗传与变异、结构与功能以及健康与疾病等生物学与医学基本问题的分子机理，并形成了一门新的学科分支——基因组学。基因组学研究的对象涉及原核生物和真核生物不同的种属，其所研究的内容触及生命学科各个领域，对生命科学的未来发展将产生重大影响。

为了向青年学生介绍基因组学的基本概貌，作者在已有教学的基础上，参考有关的书籍，收集与整理了近年来基因组学研究的最新资料，编著了《基因组学》一书。本书共分14章，第1章至第5章主要涉及基因组的结构，重点介绍基因组遗传图与物理图绘制的原理与方法，这是基因组测序与序列组装的基础，同时对基因组序列诠释的依据与注解的方法进行了分析与探讨。第6章至第10章着重介绍基因组的功能，包括基因水平以及基因组水平的表达与调控。第11章至第14章讲述基因组的进化，内容涉及基因组进化的分子机制以及进化的模式与生物多样性的关系。

基因组学是一门年青的学科，并处在迅速发展之中。书中所介绍的许多方法甚至某些实验结果在读者看到本书时或许已有改变或修正，因此希望读者随时关心基因组学相关领域的最新进展，不必囿于已有的结论。此外书中不少章节还介绍了基因组学研究中一些尚未定论的观点以及某些目前还缺少有效方法进行研究的难题，目的是为读者提供更多的思维空间，或许能从中找到自己现在或将来感兴趣的研究方向。

本书的编写工作得到赵寿元先生和乔守怡教授的全力支持与帮助。作者实验室的钱晓茵博士撰写了本书第13章“比较基因组学”一节，柯越海博士为作者提供了有关人类起源、进化与迁徙相关的基因组学的大量资料，左开井博士、王东和黄骥同学为本书绘制插图出力不少，在此一并表示衷心感谢。

杨金水

2002年2月19日于复旦大学

目 录

第1章 基因组1

1.1 遗传的分子基础2

1.1.1 DNA 的化学与生物学2

1.1.2 RNA 的化学与生物学6

1.1.3 蛋白质的结构与生物学9

1.2 基因组序列复杂性12

1.2.1 C 值与 C 值悖理12

1.2.2 序列复杂性13

1.2.3 基因组的序列组成14

1.3 基因与基因家族15

1.3.1 编码 RNA 基因15

1.3.2 编码蛋白质基因16

1.3.3 基因家族16

1.3.4 异常结构基因17

1.3.5 假基因17

1.4 染色体18

1.4.1 真核生物染色体18

1.4.2 原核生物染色体19

1.5 基因组19

1.5.1 人类基因组19

1.5.2 其他生物基因组20

复习思考题21

参考文献21

第2章 遗传图绘制23

2.1 遗传图与物理图24

2.2 遗传作图标记25

2.2.1 基因标记25

2.2.2 DNA 标记25

2.3 遗传作图的方法28

2.3.1 孟德尔遗传学简介28

2.3.2 连锁分析28

2.3.3 不同模式生物的连锁分析32

2.4 遗传图绘制38

2.4.1 人类遗传图38

2.4.2 水稻遗传图39

复习思考题39

参考文献39

第3章 物理图绘制41

3.1 限制性作图42

3.1.1 限制性作图的基本方法42

3.1.2 限制性作图的局限43

3.2 基于克隆的基因组作图46

3.2.1 大分子 DNA 的克隆载体46

3.2.2 重叠群组建49

3.2.3 指纹作图49

3.3 染色体细胞图50

3.3.1 同位素或荧光标记探针的 原位杂交50

3.3.2 原位杂交51

3.4 辐射杂种作图52

3.4.1 序列标签位点53

3.4.2 辐射杂种作图的程序与方法	53	5.1.3 实验确认基因	91
3.5 基因组整合图	56	5.1.4 基因的命名与分类	92
3.5.1 人类基因组整合图	56	5.2 基因功能预测	95
3.5.2 水稻基因组整合图	57	5.2.1 计算机预测基因功能	96
复习思考题	59	5.2.2 蛋白质结构域在功能预测中的意义	96
参考文献	59	5.2.3 根据协同进化注释基因功能	98
第4章 基因组测序与序列组装	61	5.3 基因功能检测	98
4.1 DNA 测序的方法	61	5.3.1 基因失活是基因功能分析的 主要手段	99
4.1.1 链终止法	62	5.3.2 基因的过表达用于功能检测	102
4.1.2 化学降解法	64	5.4 高通量基因功能的研究方法	103
4.1.3 自动化测序	65	5.4.1 突变体库构建	103
4.1.4 非常规测序	67	5.4.2 RNA 干扰与基因功能检测	104
4.2 基因组测序	69	5.4.3 蛋白质互作	105
4.2.1 基因组测序的策略	69	5.5 功能基因组学	106
4.2.2 基因组测序的覆盖面	69	5.5.1 组学简介	107
4.2.3 序列间隙与物理间隙	70	5.5.2 转录物组	107
4.2.4 插入片段的两端测序	70	5.5.3 蛋白质组	109
4.3 序列组装	71	复习思考题	110
4.3.1 作图法测序与序列组装	72	参考文献	110
4.3.2 鸟枪法测序与序列组装	72	第6章 基因组解剖	114
4.3.3 不同测序路线与序列组装 策略的比较	74	6.1 原核生物基因组解剖	114
4.4 基因组测序的其他路线	77	6.1.1 原核生物基因组的物理结构	114
4.4.1 重要区域的优先测序	78	6.1.2 原核生物基因组的遗传组成	117
4.4.2 EST 测序	78	6.2 真核生物基因组解剖	119
4.4.3 序列浏览	78	6.2.1 真核生物核基因组	119
4.5 人类基因组测序与组装	79	6.2.2 真核生物细胞器基因组	125
4.5.1 人类基因组的测序策略	79	6.3 转座因子与分散重复序列	130
4.5.2 人类基因组测序的伦理学问题	81	6.3.1 DNA 转座子	130
4.5.3 人类基因组测序计划相关的 重大事件	82	6.3.2 逆转录因子与分散重复序列家族	131
复习思考题	83	6.3.3 真核生物分散重复序列的比较	132
参考文献	83	6.4 串联重复序列及其分布	134
第5章 基因组序列注释	86	6.5 人类基因组的结构与组成	134
5.1 搜寻基因	86	6.5.1 人类基因组编码基因	134
5.1.1 根据基因结构特征搜寻基因	87	6.5.2 人类基因组非编码基因	137
5.1.2 同源基因查询	88	6.6 拟南芥基因组编码基因	138
		6.6.1 蛋白质编码基因	138
		6.6.2 RNA 编码基因	139

复习思考题	140	8.3.4 干扰 RNA 与转录后调控	193
参考文献	140	8.4 细胞中 mRNA 的定位与降解	194
第7章 基因的转录调控	143	8.4.1 细胞中 mRNA 的分布与定位	194
7.1 原核生物基因的转录	143	8.4.2 mRNA 的定位机制	195
7.1.1 转录起始调控	143	8.4.3 mRNA 的降解路线	196
7.1.2 转录延伸与终止调控	145	复习思考题	199
7.2 真核生物基因的转录	149	参考文献	199
7.2.1 RNA 聚合酶与转录因子	149	第9章 蛋白质的合成与加工	203
7.2.2 真核生物 Pol I 基因的转录 起始与终止	150	9.1 蛋白质的合成	203
7.2.3 真核生物 Pol II 基因的转录 起始与终止	152	9.1.1 tRNA 与氨酰化	203
7.2.4 真核生物 Pol III 基因的转录 起始与终止	155	9.1.2 密码子与反密码子的互作	205
7.2.5 细胞器基因的转录	159	9.1.3 蛋白质合成中核糖体的作用	207
7.3 古细菌基因的表达调控	160	9.2 蛋白质翻译调控	208
7.4 基因的转录调控	160	9.2.1 翻译的起始	208
7.4.1 转录调控的顺式元件	160	9.2.2 翻译的整体调控	210
7.4.2 转录调控的反式因子	162	9.2.3 翻译的专一性调控	211
7.4.3 转录因子与调控序列的互作	163	9.3 蛋白质翻译后加工	214
7.4.4 转录因子家族	165	9.3.1 蛋白质的剪切加工	215
复习思考题	168	9.3.2 蛋白质折叠	216
参考文献	168	9.3.3 化学修饰	218
第8章 RNA 的修饰与加工	171	9.4 蛋白质降解	219
8.1 细胞中的 RNA 组分	171	9.4.1 蛋白质降解标记——泛素化	220
8.1.1 编码 RNA	172	9.4.2 蛋白酶体	221
8.1.2 非编码 RNA	172	9.4.3 蛋白质降解是调控细胞活性的 重要环节	221
8.1.3 前体 RNA	174	复习思考题	223
8.2 mRNA 的修饰与加工	174	参考文献	223
8.2.1 mRNA 的 5'加帽	175	第10章 基因组表观遗传	226
8.2.2 mRNA 的 3'端多聚腺苷酸化	176	10.1 什么是表观遗传	226
8.2.3 前体 mRNA 的剪接加工	179	10.1.1 表观遗传定义	226
8.3 非编码 RNA 的加工与修饰	187	10.1.2 表观遗传现象	227
8.3.1 rRNA 前体和 tRNA 前体的 剪切加工	187	10.1.3 表观遗传机制	227
8.3.2 rRNA 和 tRNA 的化学修饰与加工	189	10.2 位置效应与表观遗传	229
8.3.3 mRNA 的编辑	191	10.2.1 座位控制区	230
		10.2.2 绝缘子	231
		10.2.3 副突变	233
		10.3 DNA 甲基化与表观遗传	235

10.3.1 DNA 甲基化	235
10.3.2 甲基化与基因调控	236
10.3.3 DNA 甲基化与转座子沉默	238
10.3.4 基因组印记	240
10.3.5 甲基化组	241
10.4 染色质重建与表观遗传	242
10.4.1 核小体与基因表达	243
10.4.2 先入模型	246
10.4.3 动态模型	248
10.4.4 非编码 RNA 与染色质结构的 动态变化	249
复习思考题	250
参考文献	250

第 11 章 基因组的复制

11.1 DNA 复制的问题	254
11.1.1 DNA 复制的拓扑学	255
11.1.2 DNA 的半保守复制	255
11.1.3 DNA 拓扑酶及其功能	256
11.1.4 DNA 复制的特点	256
11.2 原核生物基因组的复制	257
11.2.1 原核生物 DNA 复制起始点	257
11.2.2 复制的起始	258
11.2.3 复制的延伸	258
11.2.4 复制的终止	262
11.2.5 古细菌基因组的复制	263
11.3 真核生物核基因组的复制	263
11.3.1 酵母 DNA 复制起始点	263
11.3.2 高等真核生物 DNA 复制起始点	264
11.3.3 真核生物 DNA 复制叉上的事件	266
11.3.4 端粒复制	268
11.4 细胞器基因组的复制	273
11.4.1 线粒体基因组的复制	273
11.4.2 叶绿体基因组的复制	275
11.5 基因组复制的调控	276
11.5.1 基因组复制与细胞的分裂	276
11.5.2 细胞 S 期的控制	279
复习思考题	280
参考文献	280

第 12 章 基因组进化的分子基础

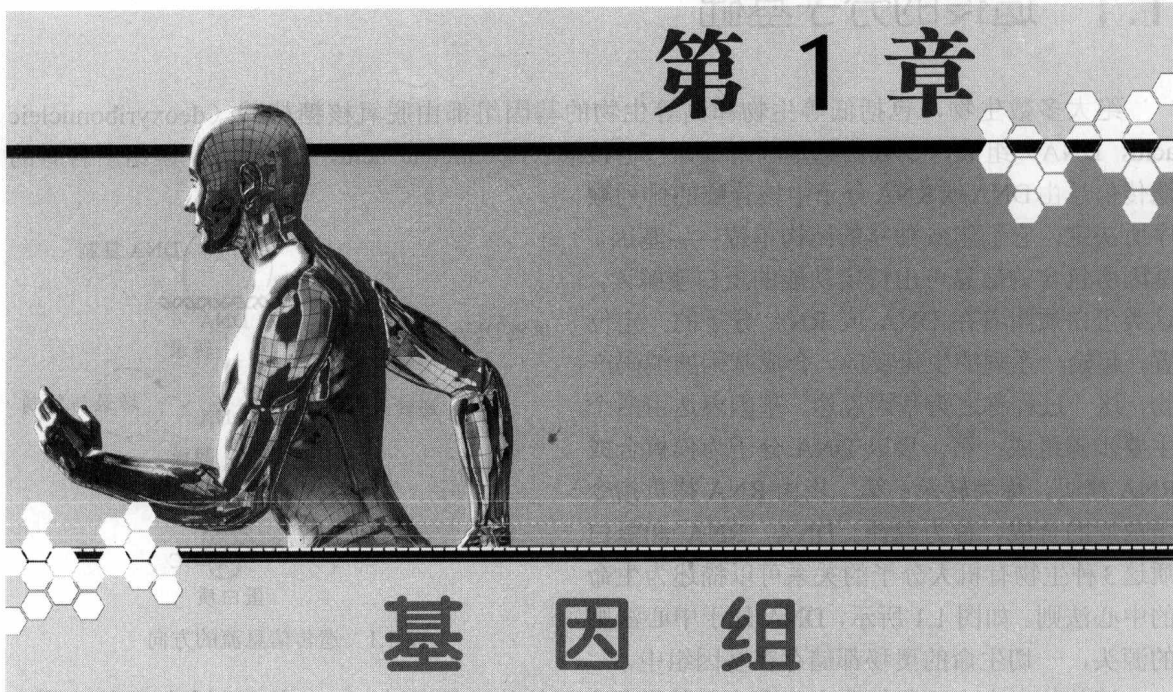
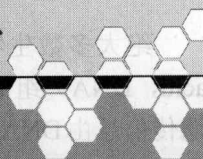
12.1 突变	283
12.1.1 突变的机制	283
12.1.2 突变的效应	286
12.1.3 超突变和程序性突变	288
12.1.4 DNA 修复	290
12.1.5 DNA 单链的非对称性进化	294
12.2 重组	295
12.2.1 同源重组	296
12.2.2 位点专一性重组	299
12.2.3 双链断裂重组模型	301
12.2.4 染色体重排	302
12.3 转座	304
12.3.1 DNA 转座	304
12.3.2 逆转录转座	304
复习思考题	307
参考文献	307

第 13 章 基因组进化的模式

13.1 遗传系统的起源	309
13.1.1 RNA 世界	309
13.1.2 基因组的起源	311
13.1.3 生命三界	312
13.2 新基因的产生	313
13.2.1 基因与基因组加倍	314
13.2.2 外显子洗牌与蛋白质创新	319
13.2.3 DNA 水平转移	321
13.2.4 重复基因的命运	323
13.3 非编码序列的扩张	325
13.3.1 真核生物基因组非编码 序列的组成	325
13.3.2 转座子与基因组进化	326
13.3.3 内含子的起源	328
13.4 比较基因组学	330
13.4.1 基因组同线性	330
13.4.2 基因岛和基因协同进化	332
13.4.3 远缘物种中基因与调控序列的 保守性	332

复习思考题	334	14.2.2 人类的起源	344
参考文献	334	14.2.3 现代人的起源	348
第 14 章 基因组与生物进化	339	14.3 基因组与生物多样性	357
14.1 分子系统发生学	339	14.3.1 生物多样性的遗传基础	358
14.1.1 表征学和分支系统学	340	14.3.2 生物多样性的分子机制	358
14.1.2 分子系统发生学	340	14.3.3 基因调控的进化与生物多样性	359
14.1.3 DNA 系统发生树	340	复习思考题	362
14.2 分子系统发生学与生物进化	343	参考文献	362
14.2.1 生命的起源	343	名词解释	366

第 1 章



基 因 组

基因组 (genome) 一词系由德国汉堡大学 H. 威克勒教授于 1920 年首创，用以表示真核生物从其亲代所继承的单套染色体，或称染色体组。更准确地说，基因组是指生物的整套染色体所含有的全部 DNA 序列。由于在真核细胞的线粒体和植物的叶绿体中也发现存在遗传物质，因此又将线粒体或叶绿体所携带的遗传物质称为线粒体基因组或叶绿体基因组。原核生物基因组则包括细胞内的染色体和质粒 DNA。此外非独立生命形态的病毒颗粒也携带遗传物质，称为病毒基因组。所有生命都具有指令其生长与发育，维持其结构与功能所必需的遗传信息，本书中将生物所具有的携带遗传信息的遗传物质总和称为基因组。

基因组学 (genomic) 一词系由 T. 罗德里克 (T. Roderick) 于 1986 年首创，用于概括涉及基因组作图、测序和整个基因组功能分析的遗传学学科分支，并已用来命名一个学术刊物 *Genomics*。基因组学是伴随人类基因组计划的实施而形成的一个全新的生命科学领域。

基因组学与传统遗传学其他学科的差别在于，基因组学是在全基因组范围研究基因的结构、组成、功能及其进化，因而涉及大范围高通量收集和分析有关基因组 DNA 的序列组成，染色体分子水平的结构特征，全基因组的基因数目、功能和分类，基因组水平的基因表达与调控以及不同物种之间基因组的进化关系。基因组学的研究方法、技术和路线有许多不同于传统遗传学的特点，各相关领域的研究仍处于迅速发展和不断完善的过程中。

1.1 遗传的分子基础

绝大多数生物，包括低等生物和高等生物的基因组都由脱氧核糖核酸（deoxyribonucleic acid, DNA）组成，少数病毒基因组则为核糖核酸（ribonucleic acid, RNA）。基因组所含有的遗传信息由DNA或RNA分子中核苷酸的排列顺序所决定，它们组成独立的结构单位——基因。基因所包含的信息可由特定功能的蛋白质解读，这类蛋白质附着在DNA或RNA分子的一定位置，起始一系列的生化反应，合成基因的编码产物，这一过程称之为基因表达。基因表达由两个主要步骤组成：第一步以DNA分子为模板合成RNA拷贝，称为转录；第二步由RNA拷贝指令蛋白质的合成，称为翻译。DNA、RNA和蛋白质这3种生物有机大分子的关系可以描述为生命的中心法则。如图1.1所示，DNA处于中心法则的源头，一切生命的奥秘都储存在基因组中。

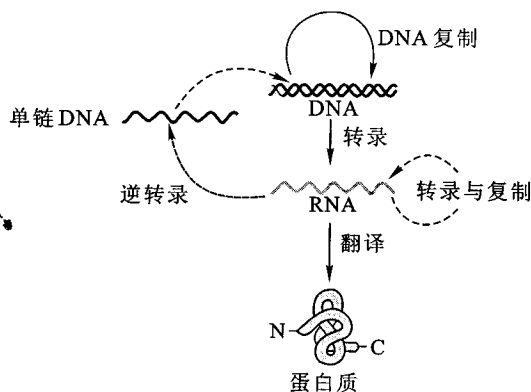


图 1.1 遗传信息流的方向

遗传信息在上下代细胞之间和个体的世代之间的传递是通过DNA的复制和细胞的分裂完成的。复制使亲代DNA加倍，通过细胞分裂将两份相同的DNA拷贝分配到两个子代细胞中。DNA在复制时偶尔会发生突变与重组，使其所携带的遗传信息改变，它们是生命进化与生物多样性的源泉。生命的所有现象都与DNA、RNA和蛋白质的结构与功能有关。

1.1.1 DNA的化学与生物学

核苷酸与多聚核苷酸

DNA是一种长链多聚分子，由4种核苷酸组成，这4种核苷酸可以任何次序排列连接成可达数百上千万个核苷酸的长链分子。每个核苷酸分子都含有3个组分（图1.2）：

2'-脱氧核糖（2'-deoxyribose） 这是一种五碳糖，即由5个碳原子组成的核糖。2'-脱氧核糖系指核糖的2'-碳原子上连接的羟基（—OH）基团由氢原子取代。

含氮碱基（nitrogenous base） 共有4种，即2个嘧啶分子，分别为胞嘧啶（cytosine）和胸腺嘧啶（thymine）；2个嘌呤分子，即腺嘌呤（adenine）和鸟嘌呤（guanine）。这些碱基通过 β -N-糖基键（ β -N-glycosidic bond）分别与嘧啶环的1位氮原子和嘌呤环的9位氮原子共价连接。

磷酸基团 磷酸基团与核糖分子的5'-碳原子相连。由糖分子与碱基组成的分子称为核苷（nucleoside）；加上磷酸后成为核苷酸。核苷酸含有的磷酸基团可分为3类，即单磷酸，双磷酸，三磷酸。

虽然细胞中均含有单磷酸，双磷酸和三磷酸基团的核苷酸，但只有核苷三磷酸才是合成DNA的底物。4种核苷三磷酸的全称为：2'-脱氧腺嘌呤-5'-三磷酸，2'-脱氧胞嘧啶-5'-三磷酸，2'-脱氧鸟嘌呤-5'-三磷酸，2'-脱氧胸腺嘧啶-5'-三磷酸。这4种核苷酸的简称依次分别为dATP，

dCTP, dGTP 和 dTTP, 当注明为 DNA 的序列时, 可简写成 A, C, G 和 T。

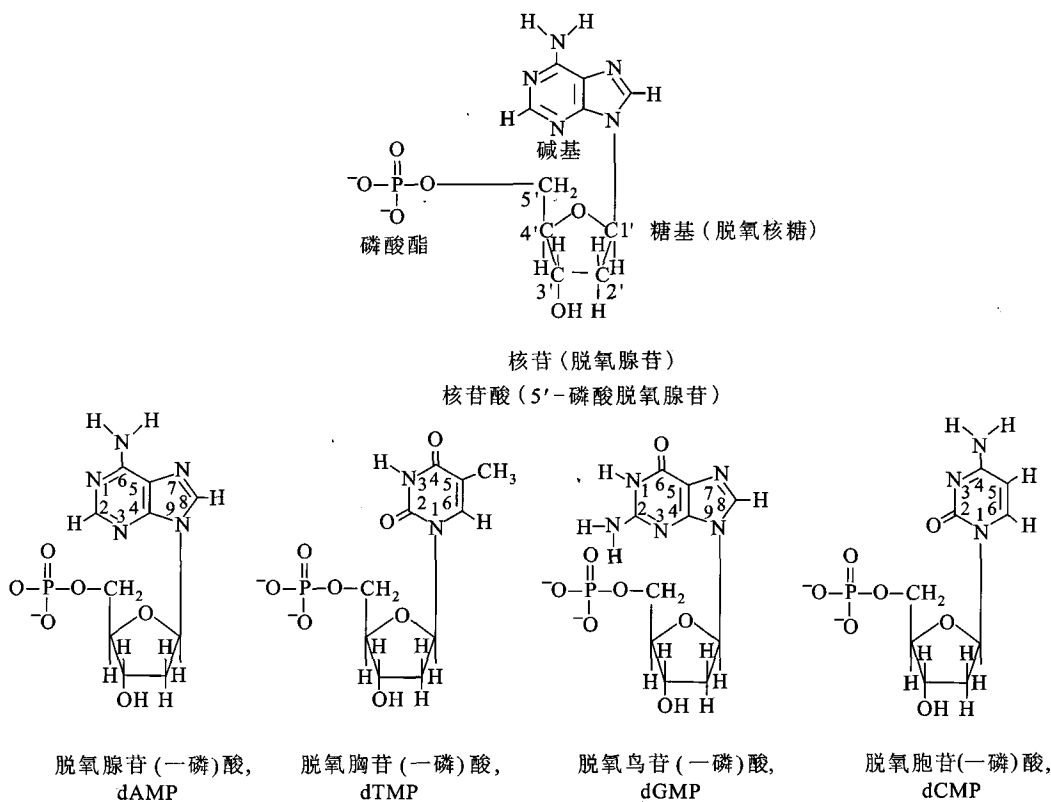


图 1.2 DNA 的组成单元为脱氧核苷酸

每个脱氧核苷酸均由一个 2'-脱氧核糖, 一个磷酸基团和一个碱基组成。碱基有 4 种, 即腺嘌呤、胞嘧啶、鸟嘌呤和胸腺嘧啶。磷酸基团与脱氧核糖的 5'-碳原子相连, 碱基则连接在核糖的 1'-碳原子上

多聚核苷酸链的化学反应只能是 5'→3'方向, 所有天然的 DNA 聚合酶都只执行 5'→3'的合成, 同样的极性也表现在以 DNA 为模板合成 RNA 拷贝的反应中(图 1.3)。核苷酸聚合反应的方向性使双链 DNA 的复制复杂化。因为 DNA 的两条互补单链化学极性正好相反, 在复制时必须采取不同的策略, 即 3'链采取连续复制, 而 5'链采取间断复制。

DNA 的双螺旋结构

2 条反向平行的 DNA 单链彼此相互缠绕组成双螺旋分子, 有 2 种化学作用稳定双螺旋结构:

碱基配对 (base-pairing) 位于 2 条 DNA 单链中的碱基可相互配对。配对只发生在腺嘌呤 (A) 与胸腺嘧啶 (T) 或鸟嘌呤 (G) 与胞嘧啶 (C) 之间, 因为只有 A 与 T 和 G 与 C 配对才使 DNA 双螺旋成为最稳定的状态 (图 1.4), A 与 T 或 G 与 C 称为互补碱基对。

碱基堆积 (base-stacking) 碱基堆积系指与 DNA 双螺旋主轴垂直的相邻碱基对杂环之间的互作, 可增加双螺旋的稳定性。

碱基配对具有重要的生物学意义。在双链 DNA 复制时, 每条单链均可作为模板以碱基互补的方式合成新链, 它们的序列组成与亲链完全相同, 使遗传信息以极其简单而准确的方式忠实地传递给下一代。

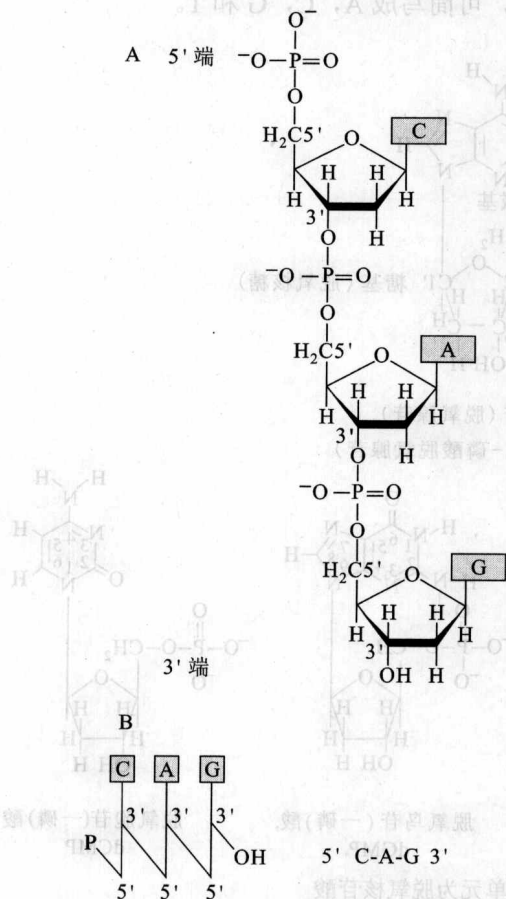


图 1.3 4 种脱氧核苷酸以磷酸二酯键连接，由此聚合成链分子
DNA 分子的合成按 5'→3' 方向进行，加入的每个核苷酸总是以 5'-磷酸与 DNA 单链的最后一个核苷酸 3'-OH 缩合形成磷酸二酯键

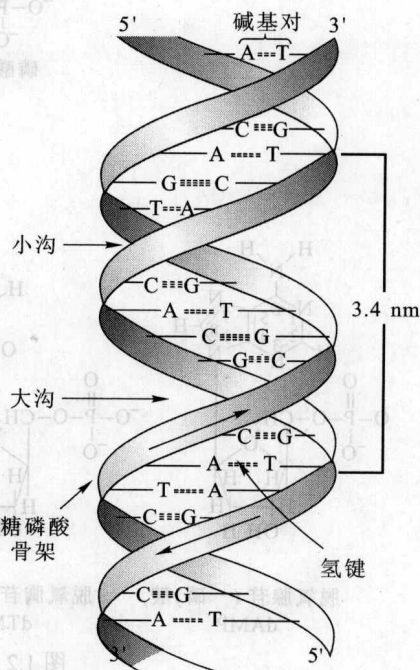


图 1.4 DNA 分子的双螺旋结构

每旋转一圈约 10.5 个核苷酸，两个大小凹槽沿 DNA 分子交替排列

碱基配对还有另一特征，即 A/T 碱基对与 G/C 碱基对之间氢键的差异。A/T 碱基对含两对氢键，G/C 碱基对为 3 对氢键，因此 G/C 碱基对比 A/T 碱基对更加稳定。或者说，A/T 碱基对比 G/C 碱基对更易解链。正是这一特点，基因组 DNA 中的某些序列为适应特别的功能便含有更多的 A/T 或 G/C。如 DNA 转录与复制的起始区要求迅速解链，该区段含有较多的 A/T 碱基对。

DNA 单链彼此缠绕时，沿着双螺旋的走向交替分布两个凹槽，一个较宽且较深的凹槽称为大沟 (major groove)，另一个较窄且较浅的为小沟 (minor groove)。DNA 双螺旋中两个交替分布的大小沟具有特征性的结构信息，在基因表达中起重要作用。DNA 结合蛋白的特定功能域可伸入大小沟中，通过氨基酸侧链与双螺旋大小沟中碱基杂环上的基团互作，由此识别 DNA 序列所包含的信息。

除少数单链 DNA 病毒外，活体中的 DNA 分子均以双链形式存在。单链 DNA 病毒的复制也必须先转变为双链 DNA，然后以半保守方式复制，再合成单链 DNA 病毒。

DNA 双螺旋构象

Watson 和 Crick 描述的 DNA 双螺旋是天然 DNA 的构象之一，即 B 型 DNA。其他的 DNA

双螺旋构象还包括 B'-, C-, C'-, C''-, D-, E- 和 T-DNA, 所有这些构象都是右旋, 此外还有左旋构象的 Z-DNA。

双螺旋 DNA 分子的不同构象影响到蛋白质接触双螺旋内部的程度, 沟槽内表面化学基团的组成与位置提供了 DNA 结合蛋白可识别的空间信息。DNA 结合蛋白自身所具有的结构能使其阅读 B 型 DNA 大小沟槽中特定的碱基顺序, 是基因调控的核心内容。一些特别的序列组成也能影响 DNA 构型, 如果周期性即每隔 10 个碱基对重复出现连续的 A/T 碱基对, 所在的 DNA 区段将发生明显的弯曲。锥虫 (trypanosome) 动基体 (kinetoplast) 的线性 DNA 片段电泳时迁移率降低, 因为 DNA 分子中出现周期性重复的 (A/T)₅₋₆ 与 (G/C)₄₋₆ 序列, 使其构型偏离一般的线形 DNA, 从而影响电泳行为。现已证明弯曲 DNA 在基因的表达调控中起重要作用。

DNA 拓扑学

DNA 有两种构型, 即线性 DNA 与环状 DNA。细菌质粒、大肠杆菌染色体、线粒体、叶绿体以及哺乳类病毒基因组均由共价连接的双链 DNA 组成。 λ 噬菌体染色体在生活史中有线性 DNA 与环状 DNA 两种状态。

活体中的 DNA 总是与蛋白质结合在一起。结合的蛋白质可使 DNA 的双螺旋轻微解旋, 因此单位长度的与蛋白质结合的 DNA 其螺旋圈数要少于游离的 B 型 DNA。当除去结合的蛋白质后, DNA 分子又恢复正常的 B 型螺旋圈数。对线性 DNA 而言, 因存在游离的末端, 增加螺旋圈数所产生的张力可由游离的末端释放。双链闭环 DNA 的情况却与此不同, 因为闭环 DNA 的螺旋圈数在拓扑学上是固定的。如果没有其他的力学补偿, 闭环 DNA 的双螺旋圈数是不能改变的。因此当结合蛋白质离开天然环状 DNA 后, 会同时发生两种事件: ①DNA 的双螺旋圈数恢复到正常 B 型 DNA 的数值; ②环状 DNA 发生扭曲, 扭曲方向与右旋方向相反, 但圈数相同, 这种 DNA 分子称为负超螺旋。如果超螺旋 DNA 的一条单链中出现单个缺口, 超螺旋构型即将消失, 拓扑学的约束力也不再存在 (图 1.5)。基因的转录, DNA 的复制、修复与重组均涉及 DNA 分子的拓扑学变化。

DNA 甲基化

无论高等生物还是低等生物, 其活体中的 DNA 都有不同程度的化学修饰, 主要为甲基化。细菌中甲基化大多发生在腺嘌呤的 6 位氮原子与胞嘧啶 5 位碳原子。高等生物 DNA 的甲基化主要涉及胞嘧啶 5 位碳原子。

细菌中有专门的甲基化酶, 它们可识别特异的 4~8 个碱基对序列并对其中特定碱基甲基化。哺乳动物中甲基化主要涉及双碱基 CpG, 使其转变为甲基化 ^mCpG。高等植物中甲基化包括回文序列 CpG 和 CpNpG, N 为任意碱基。

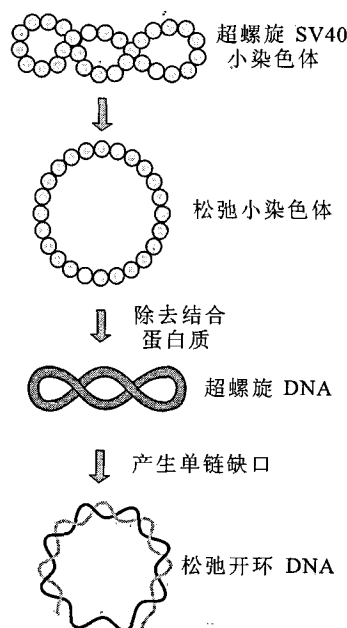


图 1.5 图示 SV40 小染色体的拓扑学结构
当除去与超螺旋小染色体结合的蛋白质后, 为了补偿因回复 B 型 DNA 产生的张力, 双链闭环 DNA 出现负超螺旋。当在超螺旋双链 DNA 的单链中产生缺口时, 单链 DNA 释放张力转变为开环结构

基因由 DNA 组成

20 世纪 20 年代人们即已完成了细胞核的细胞化学研究,并证实染色体由等量的 DNA 和蛋白质组成。由于 DNA 的组成过于简单,当时人们猜测只有蛋白质才有足够的多样性成为遗传物质,因此认为基因应由蛋白质组成。直到 20 世纪中期,2 项关键的实验报道才改变了人们关于基因分子基础的看法。

第一项实验由哥伦比亚大学 Avery、Macleod 和 McCarty 于 1944 年完成。此前,一位英国人已经证实,肺炎链球菌 (*Streptococcus pneumoniae*) 死去的细胞中有一种未知的成分可使非致病肺炎球菌活细胞转变为致病品系。Avery 及其合作者完成了一系列细心的实验,结果表明,与预期的想法相反,转化的物质不是蛋白质而是 DNA。这一重要发现当时并未引起人们的重视,许多人也不接受 DNA 是遗传物质的结论,部分原因是大多数微生物学家还不清楚转化是一种遗传现象还是一种生理现象。另一种批评意见认为, Avery 采用的脱氧核糖核酸的纯度值得怀疑,因为不能排除在纯化脱氧核糖核酸过程中存在痕量蛋白质污染的可能。

第二项实验是由纽约冷泉港美国国家实验室的 Alfred Hershey 和 Martha Chase 于 1952 年完成的,其结论无懈可击。他们的实验直接针对噬菌体生活史中感染的第一步,并清楚地证明只有 DNA 才具有遗传活性。在 Hershey-Chase 实验完成前一年, Watson 和 Ole Maaloe 就已尝试用放射性标记追踪噬菌体感染中 DNA 的命运,可惜他们以及其他一些人的实验都未能给出结论性的证据,主要因为无法区分沾污在细菌细胞表面的噬菌体与感染中产生的新的噬菌体。Hershey 和 Chase 采取同样的放射性标记方法,但在程序上作了重要修改。他们用放射性标记的噬菌体 T2 感染大肠杆菌,保温数分钟以使噬菌体吸附在细胞表面并将其内部的 DNA 注入到细胞中。然后将混合物在振荡器上振动,使空心的噬菌体从细胞表面脱落,再经离心收集含有噬菌体基因的细菌。结果证实有 70% 的噬菌体 DNA 和 20% 的噬菌体蛋白质保留在细菌组分中。在完成生活史循环后,新产生的噬菌体中仍含有 50% 原来的 DNA,而蛋白质降为 1%。这些结果表明,噬菌体接种到细菌中的基因与注入的 DNA 呈平行关系, DNA 是噬菌体的遗传物质并可遗传到下一代噬菌体中。

1.1.2 RNA 的化学与生物学

细胞中 RNA 的总量是 DNA 的 5~10 倍。RNA 的主要功能是传递遗传信息,参与基因的表达与调控。某些病毒,如逆转录病毒和许多动物、植物及昆虫的单链与双链病毒基因组由 RNA 组成。

细胞中的 RNA 种类很多,其中含量较高的包括核糖体 RNA (ribosomal RNA, rRNA), 转运 RNA (transfer RNA, tRNA) 和信使 RNA (messenger RNA, mRNA)。此外,大多数细胞中均含有其他一些胞质内小 RNA (small cytoplasmic RNA, scRNA), 真核生物还含有核内小 RNA (small nuclear RNA, snRNA)。细胞中约 80% 的 RNA 由 rRNA 和 tRNA 组成, mRNA 约占 RNA 总量的 5%。

近年来在真核生物中还发现一类在基因表达调控中起重要作用的干扰 RNA, 即 miRNA (微小 RNA) 和 siRNA (small interfering RNA, 小干扰 RNA) (He L 等, 2004)。这类非编码 RNA 分子的生物学功能多姿多彩, 几乎涉及基因组表达调控的所有方面, 是后基因组 (post-genome) 时代重要的研究内容之一。

RNA 的化学组成

RNA 与 DNA 在结构上有很大的相似性, 也由 4 种核苷酸组成, 即腺嘌呤 (A), 鸟嘌呤

呤 (G), 尿嘧啶 (U) 和胞嘧啶 (C) (图 1.6)。RNA 多聚分子的合成方式与 DNA 相似, 相邻核苷酸之间亦由 3', 5'-磷酸二酯键连接, 具有相同的化学极性。不同之处在于, RNA 核苷酸中连接在核糖 2'-碳原子上的基团为羟基而非氢原子。这一看似微小的差别极其重要, 因为 2'-OH 比 2'-H 的化学性质更为活泼, 造成了 RNA 与 DNA 整个理化性质与生物学功能的不同 (图 1.7):

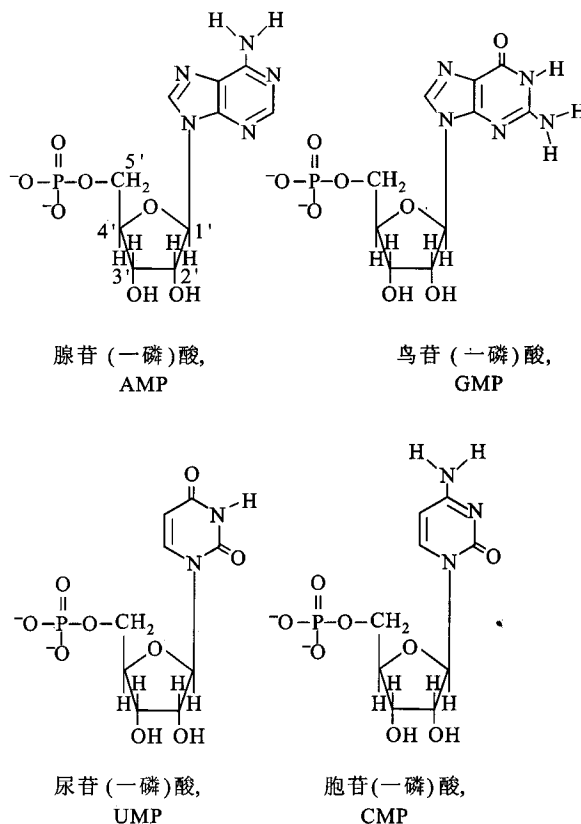


图 1.6 组成 RNA 分子的 4 种核苷酸

与脱氧核糖核苷酸的主要差别是连接核糖 2'-碳原子的基团为羟基 (—OH)

① 2'-OH 非常靠近连接两个核苷酸的磷酸二酯键位置, 使 RNA 的磷酸二酯键对碱性环境变得十分敏感, 即使在正常细胞内略微偏碱的 pH 条件下也很容易降解。

② 2'-OH 的化学作用使 RNA 构型的选择范围受到很大限制, 通常 RNA 双螺旋区段均在数十碱基对以下。

③ 长链 RNA 增加了 2'-OH 与磷酸二酯键互作引起分子断裂的可能, 限制了 RNA 长度的增加, 这是 RNA 远比 DNA 相对分子质量小得多的主要原因。由于不能形成很长的多聚分子, RNA 只能储存有限的遗传信息。

④ 2'-OH 可参与同磷酸或碱基的互作从而稳定 RNA 分子的折叠构型, 所以 RNA 分子比 DNA 分子更易形成稳定而紧凑的三级结构, 并使 RNA 分子获得了某些重要的催化功能。在一些蛋白质为主体的复合物如介导核糖体在内质网上翻译的信号识别颗粒 (signal recognition particle, SRP) 中, 7SL RNA 是重要的骨架分子。