





普通高等教育“十一五”国家级规划教材

# 基因组学

第4版

杨金水 编著

高等教育出版社·北京

## 内容提要

基因组学是当代生命科学发展最为迅速、关注度较高的学科之一。本书是国内首部全面介绍基因组学及其研究进展的图书,自2002年第1版、2007年第2版、2013年第3版出版发行以来,在专业领域产生较大影响。本次修订突出系统性和简明实用性,重点阐述了基因组学的基本概念、基本理论,介绍了研究基因组的基本思路与技术手段,充分吸收了近几年国内外学科重要进展。全书共分14章,分别是:基因组、遗传图绘制、物理图绘制、基因组测序与序列组装、基因组序列注释、基因组解剖、基因的转录调控、转录物组、蛋白质组、基因组表观遗传、基因组的复制、基因组进化的分子基础、基因组进化的模式、基因组与生物进化。

与第3版相比,第4版在如下章节进行了调整、补充与修改:1.3.4异常结构基因、2.3.2连锁分析、3.2.3指纹作图、3.5.3玉米基因组最短重叠路径整合图、4.1.3第三代DNA测序、4.4.4单细胞DNA测序、4.5.1人类基因组的测序策略、4.5.3人类基因组测序计划相关的重大事件、5.3.3基因组编辑、5.4.1突变库构建、5.5.5DNA元素百科全书、6.1.3泛基因组、7.4.1转录调控的顺式元件与反式因子、7.4.4转录场、8.2.4mRNA的定位与降解、8.3基因组小RNA、8.4长非编码RNA、9.2.3翻译的专一性调控、10.1.3表观遗传机制、10.3DNA甲基化与表观遗传、10.4组蛋白修饰与表观遗传、10.5表观遗传通路、11.2.3复制的延伸、11.3.3真核生物DNA复制叉上的事件、11.5.1基因组复制与细胞的分裂、12.1突变、12.2.3双链断裂重组模型、13.2.3水平基因转移、13.4.3远缘物种中基因与调控序列的保守性。各个章节根据学科进展和教学需要,对有关数据、图表、参考文献等也进行了更新和调整。同时,一体化设计了教学视频、教学课件、自测题等数字课程资源,参考文献、索引也在数字课程加以呈现。

本书可供综合性大学、高等师范院校、农林院校及医学院校本科生使用,也可供研究生及有关科研人员参考。

## 图书在版编目(CIP)数据

基因组学 / 杨金水编著. --4版. --北京:高等教育出版社, 2019.9

ISBN 978-7-04-052684-4

I. ①基… II. ①杨… III. ①基因组-高等学校-教材 IV. ①Q343.2

中国版本图书馆CIP数据核字(2019)第188439号

JYINZUXUE

策划编辑 吴雪梅 王莉 责任编辑 单冉东 封面设计 张楠 责任印制 刁毅

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 100120  
印刷 天津文林印务有限公司  
开本 850mm×1168mm 1/16  
印张 27  
字数 710千字  
购书热线 010-58581118  
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
网上订购 <http://www.hepmall.com.cn>  
<http://www.hepmall.com>  
<http://www.hepmall.cn>  
版 次 2002年6月第1版  
2019年9月第4版  
印 次 2019年9月第1次印刷  
定 价 52.00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 52684-00

数字课程 (基础版)

# 基因组学

(第4版)

杨金水 编著

## 登录方法:

1. 电脑访问 <http://abook.hep.com.cn/52684>, 或手机扫描下方二维码、下载并安装 Abook 应用。
2. 注册并登录, 进入“我的课程”。
3. 输入封底数字课程账号 (20 位密码, 刮开涂层可见), 或通过 Abook 应用扫描封底数字课程账号二维码, 完成课程绑定。
4. 点击“进入学习”, 开始本数字课程的学习。

课程绑定后一年为数字课程使用有效期。如有使用问题, 请点击页面右下角的“自动答疑”按钮。



## 基因组学 (第4版)

基因组学数字课程与纸质教材一体化设计, 紧密配合。数字课程配合教材各章节, 设置了教学视频、教学课件、自测题等内容。参考文献、索引也在数字课程加以呈现。可供各类高等院校不同专业的师生根据实际需求选择使用, 也可供相关科学工作者参考。

用户名:  密码:  验证码:  5369 忘记密码?

<http://abook.hep.com.cn/52684>



扫描二维码, 下载 Abook 应用

## 第4版前言

基因组学是现代生命科学最具活力的研究领域之一。作为遗传学学科分支的基因组学，虽然诞生时间不长，但借助许多全新的高科技技术的发展与创新，特别是生物信息学的进展，基因组学研究持续地呈现日新月异的变化。根据基因组学新近的研究进展和教学需求，在《基因组学》(第3版)的基础上，作者对原有教材内容进行了调整、补充与修改，编写了第4版教材。

新版仍然保持前版的编写风格，结构框架仍包括三大部分，即结构基因组(1~6章)、功能基因组(7~10章)和基因组进化(11~14章)。对各章有关知识点、图表等均有篇幅不等的更新与修订，较为重要的修订内容包括：在第4章，因当前第三代DNA测序已经商业化，故新增了第三代DNA实时测序的原理与方法；还增加了一节专门介绍单细胞DNA测序。在第5章有关基因组功能注释方法中，增加了基因组编辑的内容。在第6章有关基因组解剖内容中增加了泛基因组。在第7章基因的转录调控中，增加了一节转录场。在第8章转录物组中，对8.3节的内容做了调整，将miRNA、siRNA和piRNA整合为基因组小RNA，并分别叙述；另外增加了一节专门叙述长非编码RNA。在第10章有关基因组表观遗传的内容中，删改了原有的10.4节染色质重建与表观遗传，改为组蛋白修饰与表观遗传；此外，删除了先入模型与动态模型，增改为组蛋白修饰和染色质重建，并对DNA甲基化、组蛋白甲基化和乙酰化等内容做了修改与补充。在第11章基因组的复制中，增添了原核生物和真核生物DNA复制体的组成及其功能。在第12章基因组进化的分子机制中，对DNA的碱基突变修复机制和双链断裂修复机制进行了重要修改与补充。以上内容修改与补充均依相关文献，读者可在参考文献查找原文出处并作深入阅读。第4版教材也“一体化”设计了诸多数字课程资源，如教学视频、教学课件、自测题，以及有关拓展内容，方便读者更好地掌握有关知识并提升自主学习效果。为节约篇幅，将参考文献、索引等内容在数字课程中加以呈现。

本书的再版编写工作得到复旦大学生命科学学院各方面的支持与帮助。作者实验室的王莹博士在文献收集与引用核对、插图绘制与书稿的文字勘误等方面做了大量工作，此外本人家人在书稿的文字输入和资料整理等方面也给予了诸多帮助，在此一并表示衷心感谢！

囿于篇幅，仍然有许多基因组学研究领域的重要进展未能在新版中体现。限于作者的能力和水平，新版内容取材和编排上难免会有疏忽和遗漏，若有不当之处，诚望得到同行专家与读者的谅解与指正。

杨金水

2019年3月30日于复旦大学

## 第3版前言

由于 DNA 测序技术的革新与发展,目前国内外基因组测序计划已经覆盖到几乎所有具有重要经济价值和理论研究意义的物种。同时,由于 DNA 测序技术的改进提高了 DNA 测序的效率,极大地降低了测序成本,促使了转录组研究的蓬勃与深入发展,获取了许多此前无法探知的基因组表达谱信息。此外,生物信息技术的开发与应用,加快与深化了对结构基因组和功能基因组海量数据的分析、整理与归纳,使研究者得以从生命的整体视角,从生物体的不同结构层次和活性水平了解与认识基因组结构与功能、保守与进化的生物学意义。

《基因组学》第2版出版至今已经过去了5个年头。5年时间在科学研究的历史长河中只是一个瞬间,但对基因组学领域的研究而言,过去的5年取得了许多值得关注的进展。为了尽可能地反映基因组学领域的研究现状,为读者学习基因组学提供一些有益的参考,第3版沿袭原有的结构框架,在第2版的基础上,对原有的一些内容做了修改或重新撰写。在第4章中添加了第三代 DNA 测序,第5章中增加了基因本体以及采用 Gene Ontology 注释基因的方法。第8章的题目由原来的“RNA 的修饰与加工”改为“转录组”,第9章的题目由原来的“蛋白质的合成与加工”改为“蛋白质组”,相关内容进行更新。此外,近年来在基因组小 RNA 和非编码 RNA 研究领域取得了许多突破性进展,为在不同层次的调控水平解读基因功能提供了新的思路,为此在第8章中专门补充了许多相关的内容。

修订过程中也对各章节的参考文献进行了一次全面的梳理,尽可能地列出了相关领域的最新文献。限于篇幅,仍然有许多基因组学研究领域的重要进展未能在新版中体现,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。作者实验室的王莹同学和刘爽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,并提出了一些好的建议,在此表示衷心感谢。

杨金水

2012年8月于复旦大学

## 第2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2202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,及时反映基因组学领域近年来所取得的重要成果与重大进展,作者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,仅将原第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,读者如何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的工作,在此表示衷心感谢。

杨金水

2006年11月8日于复旦大学

# 第1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，对生物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行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，目的是为了读者提供更多的思维空间，或许能从中找出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

# 数字课程“教学视频”目录



- 第 1 章 基因组学——基因组计划缘起
  - 基因组学——基因组学重要进展
  - 基因组——DNA、RNA 与蛋白质
  - 基因组——基因组结构特征
- 第 2 章 遗传图绘制
- 第 3 章 物理图绘制
- 第 4 章 基因组测序技术
  - 全基因组测序
- 第 5 章 基因组注释——信息分析
  - 基因组注释——实验验证
  - 基因组注释——基因组编辑
- 第 6 章 基因组解剖——真核生物基因组
  - 基因组解剖——原核生物及细胞器基因组
- 第 7 章 基因的转录调控——Pol I 与 Pol III 类基因
  - 基因的转录调控——Pol II 类基因
- 第 8 章 转录物组——组成与加工
  - 转录物组——定位与降解
  - 转录物组——miRNA
  - 转录物组——siRNA
  - 转录物组——lncRNA
- 第 9 章 蛋白质组——翻译调控
  - 蛋白质组——折叠与降解
- 第 10 章 基因组表观遗传——表观遗传现象
  - 基因组表观遗传——DNA 甲基化
  - 基因组表观遗传——组蛋白甲基化
  - 基因组表观遗传——组蛋白乙酰化
- 第 11 章 基因组复制——原核生物
  - 基因组复制——真核生物
- 第 12 章 基因组进化的机制——突变与修复
  - 基因组进化的机制——重组与转座
- 第 13 章 基因组进化的模式——基因组水平的进化
  - 基因组进化的模式——基因水平的进化
- 第 14 章 基因组与生物进化——生命的起源与人类的进化

# 目 录

<b>第 1 章 基因组</b> .....	1
1.1 遗传的分子基础 .....	1
1.1.1 DNA 的化学与生物学 .....	2
1.1.2 RNA 的化学与生物学 .....	7
1.1.3 蛋白质的结构与生物学 .....	10
1.2 基因组序列复杂性 .....	13
1.2.1 C 值与 C 值悖理 .....	13
1.2.2 序列复杂性 .....	14
1.2.3 基因组的序列组成 .....	15
1.3 基因与基因家族 .....	16
1.3.1 编码 RNA 的基因 .....	16
1.3.2 编码蛋白质的基因 .....	17
1.3.3 基因家族 .....	18
1.3.4 异常结构基因 .....	19
1.3.5 假基因 .....	20
1.4 染色体 .....	21
1.4.1 真核生物染色体 .....	21
1.4.2 原核生物染色体 .....	22
1.5 基因组 .....	23
1.5.1 人类基因组 .....	23
1.5.2 其他生物基因组 .....	23
<b>第 2 章 遗传图绘制</b> .....	25
2.1 遗传图与物理图 .....	26
2.2 遗传作图标记 .....	27
2.2.1 基因标记 .....	27
2.2.2 DNA 标记 .....	27
2.3 遗传作图的方法 .....	30
2.3.1 孟德尔遗传学简介 .....	30
2.3.2 连锁分析 .....	31
2.3.3 不同模式生物的连锁分析 .....	36
2.4 遗传图绘制 .....	43
2.4.1 人类遗传图 .....	43
2.4.2 水稻遗传图 .....	43
<b>第 3 章 物理图绘制</b> .....	45
3.1 限制性作图 .....	46
3.1.1 限制性作图的基本方法 .....	46
3.1.2 限制性作图的局限 .....	47
3.2 基于克隆的基因组作图 .....	50
3.2.1 大分子 DNA 的克隆载体 .....	51
3.2.2 重叠群组建 .....	53
3.2.3 指纹作图 .....	54
3.3 原位染色体连锁图 .....	55
3.3.1 同位素或荧光标记探针的原位杂交 .....	56
3.3.2 染色体原位杂交 .....	56
3.4 辐射杂种作图 .....	57
3.4.1 序列标签位点 .....	58
3.4.2 辐射杂种作图的程序与方法 .....	58
3.5 基因组整合图 .....	62
3.5.1 人类基因组整合图 .....	62
3.5.2 水稻基因组整合图 .....	63
3.5.3 玉米基因组最短重叠路径整合图 .....	63
<b>第 4 章 基因组测序与序列组装</b> .....	66
4.1 DNA 测序的方法 .....	66
4.1.1 第一代 DNA 测序 .....	66
4.1.2 第二代 DNA 测序 .....	70
4.1.3 第三代 DNA 测序 .....	73
4.2 基因组测序 .....	78
4.2.1 基因组测序的策略 .....	78
4.2.2 基因组测序的覆盖面 .....	79

4.2.3 序列间隙与物理间隙·····	79	5.5 功能基因组学·····	122
4.2.4 插入片段的两端测序·····	81	5.5.1 组学简介·····	123
4.3 序列组装·····	81	5.5.2 转录物组·····	124
4.3.1 作图法测序与序列组装·····	82	5.5.3 蛋白质组·····	125
4.3.2 鸟枪法测序与序列组装·····	82	5.5.4 基因本体·····	128
4.3.3 不同测序路线与序列组装策略的 比较·····	85	5.5.5 DNA 元素百科全书·····	132
4.4 基因组测序的其他路线·····	88	<b>第6章 基因组解剖</b> ·····	135
4.4.1 重要区域的优先测序·····	88	6.1 原核生物基因组解剖·····	135
4.4.2 EST 测序·····	89	6.1.1 原核生物基因组的物理结构·····	135
4.4.3 共栖 / 共生生物基因组测序·····	89	6.1.2 原核生物基因组的遗传组成·····	138
4.4.4 单细胞 DNA 测序·····	89	6.1.3 泛基因组·····	141
4.5 人类基因组测序与组装·····	91	6.2 真核生物基因组解剖·····	142
4.5.1 人类基因组的测序策略·····	91	6.2.1 真核生物核基因组·····	142
4.5.2 人类基因组测序的伦理学问题·····	94	6.2.2 真核生物细胞器基因组·····	149
4.5.3 人类基因组测序计划相关的重大 事件·····	95	6.3 转座因子与分散重复序列·····	154
<b>第5章 基因组序列注释</b> ·····	98	6.3.1 DNA 转座子·····	154
5.1 搜寻基因·····	98	6.3.2 逆转录因子与分散重复序列家族·····	154
5.1.1 根据基因结构特征搜寻基因·····	99	6.3.3 真核生物分散重复序列的比较·····	158
5.1.2 同源基因查询·····	100	6.4 串联重复序列及其分布·····	159
5.1.3 实验确认基因·····	104	6.5 人类基因组的结构与组成·····	160
5.1.4 基因的命名与分类·····	105	6.5.1 人类基因组编码基因·····	160
5.2 基因注释·····	109	6.5.2 人类基因组非编码基因·····	163
5.2.1 计算机预测基因功能·····	110	6.6 拟南芥基因组的结构与组成·····	164
5.2.2 蛋白质结构域在功能预测中的 意义·····	110	6.6.1 蛋白质编码基因·····	164
5.2.3 根据协同进化注释基因功能·····	112	6.6.2 RNA 编码基因·····	165
5.3 基因功能检测·····	113	<b>第7章 基因的转录调控</b> ·····	167
5.3.1 基因失活是基因功能分析的主要 手段·····	113	7.1 原核生物基因的转录·····	167
5.3.2 基因的过量表达用于基因功能 检测·····	115	7.1.1 转录起始调控·····	167
5.3.3 基因组编辑·····	116	7.1.2 转录延伸与终止调控·····	169
5.4 高通量基因功能的研究方法·····	119	7.2 真核生物基因的转录·····	173
5.4.1 突变库构建·····	119	7.2.1 RNA 聚合酶与转录因子·····	173
5.4.2 RNA 干扰与基因功能检测·····	121	7.2.2 真核生物 Pol I 基因的转录起始与 终止·····	174
5.4.3 蛋白质互作·····	121	7.2.3 真核生物 Pol II 基因的转录起始与 终止·····	177
		7.2.4 真核生物 Pol III 基因的转录起始与 终止·····	181

7.2.5 细胞器基因的转录·····	184	9.3.2 蛋白质折叠·····	251
7.3 古菌基因的表达调控·····	185	9.3.3 化学修饰·····	253
7.4 转录调控·····	185	9.4 蛋白质降解·····	255
7.4.1 转录调控的顺式元件与反式因子·····	185	9.4.1 蛋白质降解标记——泛素化·····	255
7.4.2 转录因子与调控序列的互作·····	187	9.4.2 蛋白酶体·····	256
7.4.3 转录因子家族·····	191	9.4.3 蛋白质降解是调控细胞活性的 重要环节·····	257
7.4.4 转录场·····	194		
<b>第 8 章 转录物组</b> ·····	197	<b>第 10 章 基因组表观遗传</b> ·····	260
8.1 细胞中的 RNA 组分·····	197	10.1 什么是表观遗传·····	260
8.1.1 mRNA·····	198	10.1.1 表观遗传定义·····	260
8.1.2 非编码 RNA·····	198	10.1.2 表观遗传现象·····	261
8.1.3 前体 RNA 及其修饰·····	199	10.1.3 表观遗传机制·····	261
8.2 mRNA 的修饰与加工·····	200	10.2 位置效应与表观遗传·····	263
8.2.1 mRNA 的 5' 加帽·····	200	10.2.1 座位控制区·····	263
8.2.2 mRNA 的 3' 端多聚腺苷酸化·····	202	10.2.2 绝缘子·····	265
8.2.3 前体 mRNA 的剪接加工·····	205	10.2.3 副突变·····	267
8.2.4 mRNA 的定位与降解·····	214	10.2.4 单等位基因表达·····	269
8.3 基因组小 RNA·····	218	10.3 DNA 甲基化与表观遗传·····	269
8.3.1 miRNA·····	218	10.3.1 DNA 甲基化·····	269
8.3.2 siRNA·····	223	10.3.2 DNA 甲基化与基因调控·····	274
8.3.3 piRNA·····	226	10.3.3 DNA 甲基化与转座子沉默·····	274
8.4 长非编码 RNA·····	228	10.3.4 基因组印记·····	275
8.4.1 长非编码 RNA 的普遍性·····	229	10.4 组蛋白修饰与表观遗传·····	277
8.4.2 长非编码 RNA 的特征·····	229	10.4.1 核小体与基因表达·····	277
8.4.3 长非编码 RNA 的功能·····	230	10.4.2 组蛋白修饰·····	279
8.4.4 非编码 RNA 的生物学意义·····	231	10.4.3 染色质重建·····	282
<b>第 9 章 蛋白质组</b> ·····	236	10.5 表观遗传通路·····	283
9.1 蛋白质的合成·····	236	10.5.1 表观遗传诱导·····	284
9.1.1 tRNA 与氨酰化·····	236	10.5.2 表观遗传起始·····	286
9.1.2 密码子与反密码子的互作·····	238	10.5.3 表观遗传密码·····	288
9.1.3 蛋白质合成中核糖体的作用·····	240	10.5.4 表观遗传维持·····	293
9.2 蛋白质翻译调控·····	242	<b>第 11 章 基因组的复制</b> ·····	296
9.2.1 翻译的起始·····	242	11.1 DNA 复制的问题·····	296
9.2.2 翻译起始的整体调控·····	244	11.1.1 DNA 复制的拓扑学·····	297
9.2.3 翻译的专一性调控·····	244	11.1.2 DNA 的半保守复制·····	297
9.3 蛋白质翻译后加工·····	250	11.1.3 DNA 拓扑酶及其功能·····	297
9.3.1 蛋白质的剪接加工·····	250	11.1.4 DNA 复制的特点·····	298

11.2 原核生物基因组的复制 .....	299	13.1.2 基因组的起源 .....	362
11.2.1 复制起始点 .....	299	13.1.3 生命三域 .....	363
11.2.2 复制的起始 .....	299	13.2 新基因的产生 .....	364
11.2.3 复制的延伸 .....	300	13.2.1 基因与基因组加倍 .....	366
11.2.4 复制的终止 .....	305	13.2.2 外显子洗牌与蛋白质创新 .....	372
11.2.5 古菌基因组的复制 .....	307	13.2.3 水平基因转移 .....	375
11.3 真核生物核基因组的复制 .....	307	13.2.4 重复基因的命运 .....	377
11.3.1 酵母 DNA 复制起始点 .....	307	13.3 非编码序列的扩张 .....	379
11.3.2 高等真核生物 DNA 复制起始点 .....	308	13.3.1 真核生物基因组非编码序列的组成 .....	379
11.3.3 真核生物 DNA 复制叉上的事件 .....	310	13.3.2 转座子与基因组进化 .....	380
11.3.4 端粒复制 .....	313	13.3.3 内含子的起源 .....	383
11.4 细胞器基因组的复制 .....	317	13.4 比较基因组学 .....	385
11.4.1 线粒体基因组的复制 .....	318	13.4.1 基因组同线性 .....	385
11.4.2 叶绿体基因组的复制 .....	320	13.4.2 基因岛和基因协同进化 .....	387
11.5 基因组复制的调控 .....	320	13.4.3 远缘物种中基因与调控序列的保守性 .....	388
11.5.1 基因组复制与细胞的分裂 .....	320		
11.5.2 细胞 S 期的控制 .....	323		
<b>第 12 章 基因组进化的分子基础</b> .....	<b>328</b>	<b>第 14 章 基因组与生物进化</b> .....	<b>391</b>
12.1 突变 .....	328	14.1 分子系统发生学 .....	391
12.1.1 突变的机制 .....	328	14.1.1 表征学和分支系统学 .....	391
12.1.2 突变的效应 .....	332	14.1.2 分子系统发生学 .....	392
12.1.3 超突变与程序性突变 .....	335	14.1.3 DNA 系统发生树 .....	392
12.1.4 DNA 修复 .....	337	14.2 分子系统发生学与生物进化 .....	395
12.1.5 DNA 单链的非对称性进化 .....	345	14.2.1 生命的起源 .....	395
12.2 重组 .....	345	14.2.2 人类的起源 .....	396
12.2.1 同源重组 .....	346	14.2.3 现代人的起源 .....	402
12.2.2 位点专一性重组 .....	349	14.3 基因组与生物多样性 .....	412
12.2.3 双链断裂重组模型 .....	351	14.3.1 生物多样性的遗传基础 .....	412
12.2.4 染色体重排 .....	352	14.3.2 生物多样性的分子机制 .....	413
12.3 转座 .....	355	14.3.3 基因调控的进化与生物多样性 .....	413
12.3.1 DNA 转座 .....	356		
12.3.2 逆转录转座 .....	356		
<b>第 13 章 基因组进化的模式</b> .....	<b>360</b>	<b>参考文献</b>	<b>索引</b>
13.1 遗传系统的起源 .....	360		
13.1.1 RNA 世界 .....	360		

# 第 1 章

## 基因组

**基**因组 (genome) 一词系由德国汉堡大学威克勒教授于 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存在 DNA, 因此又将线粒体或叶绿体所携带的 DNA 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 (genomics) 一词系由罗德里克 (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。

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。

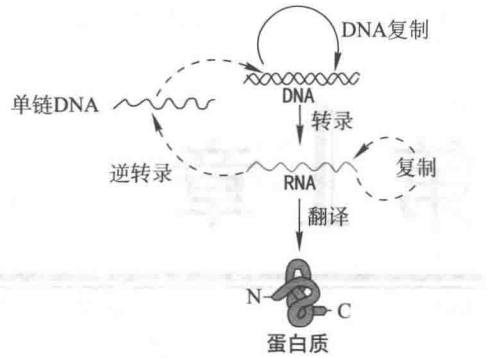


图 1.1 中心法则——遗传信息流  
遗传信息流的总的方向是从 DNA 到 RNA, 再从 RNA 到蛋白质。逆转录可将 RNA 的遗传信息转移到 DNA, 但逆转录的遗传信息的表达仍然要经由 DNA 到 RNA, 再从 RNA 到蛋白质的流程, 并不改变遗传信息流的方向

### 1.1.1 DNA 的化学与生物学

生命活动无论在何种水平, 每一种特定的生物学功能总有特定的结构基础与之对应。DNA 几乎是所有生命形式都必须具备的遗传物质。DNA 的生物学功能主要是储存遗传信息, DNA 的物理结构和 DNA 的化学性质与其功能是相适应的。

#### 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  $\beta$ -N-糖基键 ( $\beta$ -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。当未注明具体的脱氧核糖核苷酸, 而是泛指 4 种核苷酸时, 可以写成 dNTP, “N” 代表 A、C、G 和 T 中的任意一种。

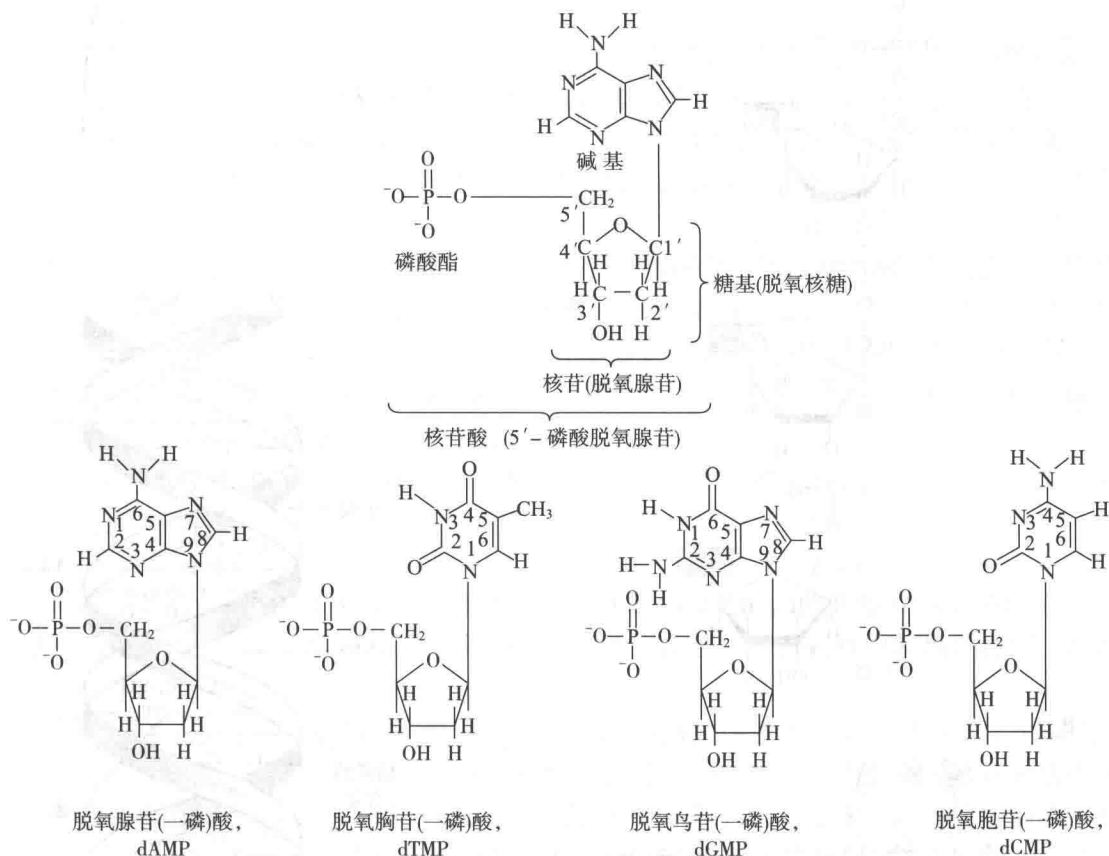


图 1.2 DNA 的组成单元为脱氧核苷酸

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

由于核苷酸单体中提供聚合反应所需能量的三磷酸基团直接与核糖的 5'-碳原子相连，发生 DNA 聚合反应时，下一个核苷酸单体与前一个核苷酸单体的聚合总是由下一个核苷酸单体的 5'-三磷酸与前一个核苷酸单体的 3'-OH 缩合生成磷酸酯键。因此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。

**核酸序列记号** 虽然在大多数场合下，DNA 序列中的碱基记号只涉及上述 4 种，即 A, C, G 和 T。但在有些场合常常会在核酸序列中出现一些表示含混的碱基位点，如转录因子结合位点 -TRCCATSRN-，这里的 R、T 和 S 分别代表某些可选碱基（Arguinano AAA 等，2017）。为了方便交流，NC-IUB 组织（Nomenclature Committee of the International Union of Biochemistry，国际生物化学联合会命名委员会）推荐了一套核酸序列记号规则，供研究者在必要时参考使用。该规则如下所述：R，代表 A+G（即该位点可以是 A，也可以是 G）；Y，C+T；M，A+C；K，G+T；S，C+G；W，A+T；H，A+C+T；B，C+G+T；V，A+C+T；N，

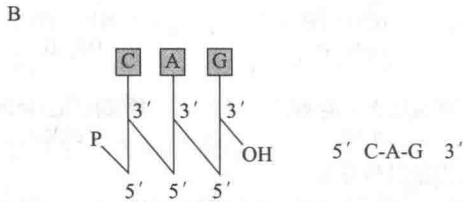
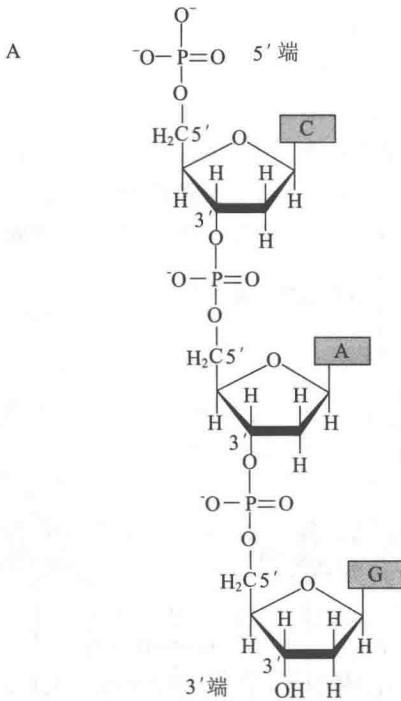


图 1.3 多聚核苷酸分子结构

A. 4 种脱氧核苷酸以磷酸二酯键连接, 由此聚合成链分子。  
 B. 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

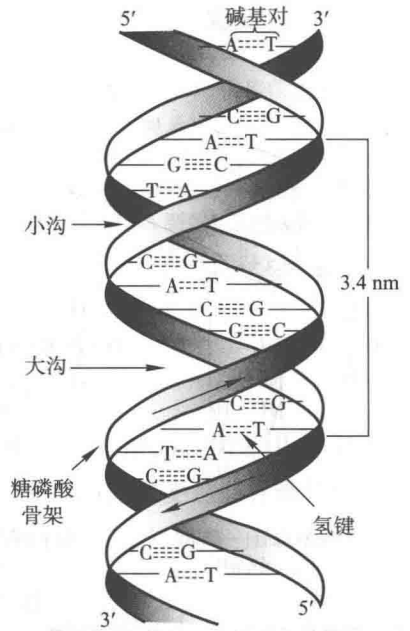


图 1.4 DNA 分子的双螺旋结构  
每旋转一圈约 10.5 个核苷酸, 两个大小凹槽沿 DNA 分子交替排列

A+C+G+T ( Nomenclature Committee of the International Union of Biochemistry, 1986 )。

### 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动, 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