

*Discovering*  
**探索**

GENOMICS,  
**基因组学**  
PROTEOMICS,  
**蛋白质组学和**  
& BIOINFORMATICS  
**生物信息学**

(第二版)

[美] A. 马尔科姆·坎贝尔 著  
劳里 J. 海尔 著  
孙之荣 主译

A T G C T C A T C G G C A T T G



科学出版社  
www.sciencep.com

# 探索基因组学、蛋白质组学 和生物信息学

(第二版)

[美]A. 马尔科姆·坎贝尔 劳里J. 海尔 著

孙之荣 主译

科学出版社

北京

图字:01-2006-6137号

## 内 容 简 介

本书围绕对基因组学、蛋白质组学和生物信息学的概括描述,向读者介绍了近几年来生物学以及医学生物学研究方法上的发展,以及这些方法在研究思维上的深远影响。作者以通俗易懂的语言将这三个方面组成一个研究、探索问题的互动平台,使读者对生物信息学有一个系统的认识和了解,利于更深入的研究。内容主要包括基因组序列、基因组变异、基因组表达、DNA芯片、蛋白质组学、全基因组学以及基因组学在医学病例中的应用等,同时穿插问题探讨、数学备忘录等使内容更加丰富。

本书可作生物学专业本科生、研究生的生物信息学教材或教学辅导书,亦可供分子生物学、生物化学、细胞生物学以及医学、药学等领域的科研人员参考。

Simplified Chinese edition copyright 2007 by PEARSON EDUCATION NORTH ASIA LIMITED and SCIENCE PRESS.

Original English language title: Discovering Genomics, Proteomics, and Bioinformatics, by A. Malcolm Campbell and Laurie J. Heyer, Copyright 2007

All Rights Reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Benjamin Cummings.

This edition is authorized for sale only in the People's Republic of China (excluding the Special Administrative Regions of Hong Kong and Macao).

本书封面贴有 Pearson Education 出版集团激光防伪标签,无标签者不得销售。

### 图书在版编目(CIP)数据

探索基因组学、蛋白质组学和生物信息学/(美)坎贝尔(Campbell, A. M.), (美)海尔(Heyer, L. J.)著;孙之荣主译.—2版.—北京:科学出版社,2007

ISBN 978-7-03-018903-5

I. 探… II. ①坎…②海…③孙… III. ①基因组-研究②蛋白质-研究③生物信息论-研究 IV. ①Q343.1 Q51

中国版本图书馆CIP数据核字(2007)第062580号

责任编辑:夏 梁 虎在堂 彭克里 孙晓洁/责任校对:邹慧卿

责任印制:钱玉芬/封面设计:王 浩

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2004年7月第 一 版 开本:787×1092 1/16

2007年5月第 二 版 印张:35

2007年5月第二次印刷 字数:794 000

印数:3 001—6 000

定价:75.00元

(如有印装质量问题,我社负责调换〈环伟〉)

## 译者序

21 世纪的生物学研究对每一个生物学工作者来说既是机遇又是挑战。随着人类基因组测序工作的初步完成,基因组生物学的研究由测序基因组向功能基因组学转移。作为功能基因组学的重要研究领域之一的蛋白质组学,与基因组学一起,成为当今生命科学的热点和前沿。基因组学、蛋白质组学这一新兴的研究领域孕育了生物信息学这一新兴交叉学科的产生和发展。基因组学、蛋白质组学和生物信息学密不可分。这本书的作者将基因组学、蛋白质组学和生物信息学放在一起加以介绍,别具特色。

《探索基因组学、蛋白质组学和生物信息学》围绕基因组学、蛋白质组学和生物信息学,详细地展示了近几年来生物学及医学生物学研究方法上的革命,以及它们在研究思维方式上的深远影响。

本书涵盖了现代生物学的大部分研究领域。作者将本书设计成一个探索基因组学和蛋白质组学问题的互动平台,能够帮助读者学会假设、推理、分析、建立模型等研究现代生物学的特定模式的能力。所以这本书无论是对教学人员,还是对科研工作者都是一本适用的参考书。

本书的翻译工作是由清华大学生物科学与技术系生物信息研究所的师生集体完成的。参加本书翻译工作的有孙之荣教授、柳树群博士,以及郑家顺、季星来、夏雪峰、资治科、李薇、陈虎、张扬和沈金诚等。全书由孙之荣教授负责组织和审校工作。希望本书的翻译和出版将对我国生物信息学以及基因组生物学、蛋白质组生物学的发展具有良好的促进作用。

孙之荣  
清华大学生物科学与技术系

# 序

基因组学这个名词来自于基因组，基因组指的是生物体内所有 DNA 的总和，基因组学是研究基因组的学科。当基因组和基因组学的概念流行起来后，大量以 omes 和 omics 结尾的名词开始在出版物中频频出现。《探索基因组学、蛋白质组学和生物信息学》不仅仅是一本 omes 的书，因为该领域已经不局限于基因组学的狭隘定义，正如本书所体现的，基因组学包括细胞内各种分子（比如 DNA、蛋白质、脂质和糖等）的相互作用。本着探索的精神，我们将在本书中探讨一系列引发生物学研究方式变革的工具和问题。

《探索基因组学、蛋白质组学和生物信息学》采取了许多老师成功运用过的两条教学原则：教授感兴趣的问题并在求知欲的基础上引出新的知识。这其实是人们获取新信息的途径，也是引导学生学习的最好方法，这样他们才会有求知的欲望并牢记所学的知识。《探索基因组学、蛋白质组学和生物信息学》从文献中提取出研究例子，在解答这些研究案例所遇到的各种问题的同时，探索科学的内容。本书涵盖了现代生物学的大部分研究领域如基因组学（全基因组和基因变异）、蛋白质组学和生物信息学。

作者将本书设计成一个探索基因组学和蛋白质组学问题的互动平台。插图包含了许多数据，读者可以自行挖掘其中更多的内在信息；“在线数据库”鼓励读者像真正做研究一样使用数据库进行实时探索；“问题探讨”使读者专注于关键信息，并促使读者应用文章和图片中的工具和信息去思考。一般的教科书提供的是应付测试所需记忆的事实和细节，而基因组学要求读者学会分析、假设、推理、建立模型；本书正是为了让读者具备这种能力而设计。

21 世纪的生物学研究对每一个生物学工作者来说既是机遇又是挑战。本书的特色在于：读者必须反复使用计算机去获取最新信息。为了让读者深入理解基因组学、蛋白质组学和生物信息学，本书列举了大量的真实而有说服力的研究事例，激发读者的学习兴趣。

## 写作风格

文体通俗易懂，虽然引入了一些新名词，但尽量避免了不必要的术语。

## 问题探讨

通过“问题探讨”来引导读者思考，问题探讨不是在每章末尾才提出，而是放在每个研究案例之中。问题探讨让读者的注意力集中在实验设计、数据解释、数据对观点的支持等关键概念上，分析从文献中提取出来的数据，让读者得出自己的结论，而作者只是引导读者去寻找相应数据。为了能够回答其中的一些问题，读者必须自己使用在线数据库，这些数据库是经常更新的。

所有的问题探讨都可以从与本书配套的网站得到，便于读者与在线资源互动，读

者也可以通过电子邮件把答案送给自己的导师批阅。

### 问题探讨

46. 在 NCBI 上的 MeSH(医学部分标题)检索“CAMP 因子”。你会看到一条叫做“CAMP 蛋白,链球菌[物种名]”。在最右边,点击“Link”的链接并选择 NLM(国家医学图书馆)MeSH 浏览器。CAMP 因子能够对我们的血细胞有什么作用?

### 工具

《探索基因组学、蛋白质组学和生物信息学》的配套网站(www. GeneticsPlace. com)是一个帮助学习基因组学的好工具。

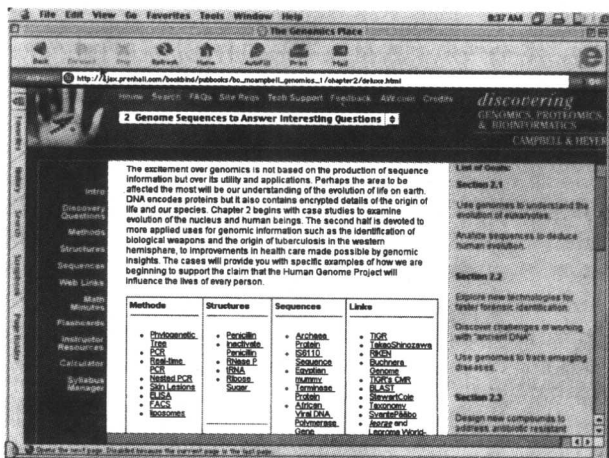
**方法:**这些网页解释了分子和基因组学方法是如何运用并产生了什么样的数据,如果读者以前没有学过相关知识,他们将提供相关的背景知识作为补充。

**数据:**在多个问题探讨中,读者要运用在线的生物信息学工具分析蛋白质或者 DNA 序列。为了节省读者的时间并避免排版错误,所有序列都在网页上提供,读者可以复制粘帖到分析工具中去。一些非序列数据也提供了。

**结构:**三维结构对每个蛋白质来说都是非常重要的。其中带有有声教程的网页有助于读者更好地理解蛋白质的结构特征。Jmol 是一个基于 Java 的三维结构软件,并将在随后几年中替代 Chime。我们很高兴成为第一本采用新标准的 Jmol 的教材。

**链接:**每章都有两种链接。第一种链接直接连到在线数据库和生物信息学分析工具的相关网址,例如,NCBI 和 PDB;第二种链接链接到了相关实验室的主页上,当读者对某个案例的研究内容格外感兴趣时可以使用该种链接。

**数学备忘录:**一些数据集和交互电子资源的链接可以供你研究时使用。一些数学备忘录和探索问题的回答要求你利用这些文件。



**相关方法**

Quick PDB

**结构**

环加氧酶

**序列**

未鉴定蛋白

**相关链接**

保守域

PDB

PREDATOR

**工具菜单**

工具菜单贯穿每一章,提醒读者使用于本书配套的网站资源。在该网站上读者可以阅读更多更深入的方法描述,获取序列信息,观看三维结构图,链接到相关网站。这些工具帮助你更好地参与到理解基因组学核心内容的互动过程中来。

**数学备忘录**

许多细胞/分子领域的生物学家在他们的工作中用到的数学不多,但是基因组学、蛋白质组学和生物信息学正在改变这个事实,这些领域都非常依赖数学。为了让读者更好地理解数据分析的过程和数学在生物学研究中扮演的角色,我们设立了数学备忘录,以便读者了解生物和数学之间的联系。数学备忘录(如下例)用研究案例作为统计分析、概率和计算方法的简明教程。

**数学备忘录 6.2 Sup35 是蛋白质相互作用网络的中心蛋白质吗?**

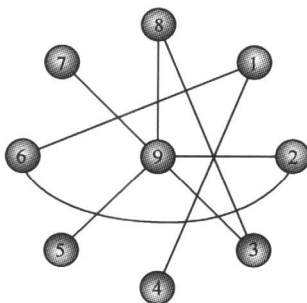
图 6.16 中,Sup35 看起来位于相互作用蛋白质的网络中心。但是,在这样一个复杂的网络中,观察者看到的很可能是假象。因此,Sup35 与网络中的其他蛋白质相比是否能与绝大多数蛋白质发生相互作用,还需要进一步的量化处理。

数学家将图 6.16 的相互作用蛋白网络叫做图。在图论里,节点间的连线被称为弧或边。定向图的边上具有箭头,以指明两个节点间的信息流动方向。在图 6.16 中,节点的连接和关系是相当重要的,但是,因为边上没有箭头,所以并没有给出方向信息。与节点接触的边的数目叫做节点的度。用图论的术语讲,目前的问题在于节点 Sup35 是否具有比图中“普通”节点大得多的度。如果这样,研究者便能确定 Sup35 的确在整个蛋白质相互作用网络中发挥中心作用。

**相关链接**

DIP 数据库

解决图 6.16 问题的方法是建立一个图概率模型,称之为随机图。在这个模型里,使用一定的概率画出每对节点之间的边,每个边独立于其他边。高级概率理论以及图论提供了精确测定超过一定数量的大图中最大度的概率方法。但是,我们有一个简单的假设,即度数近似于正态分布(见数学备忘录 8.1)因此,我们能估计出 Sup35 节点的度是否显著大于度数的平均期望值和标准误差。由于图 6.16 作为例子讲解过于复杂,我们来看一下下面这个小网络图:



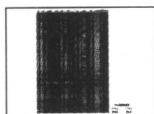
在该图中,节点4、5、7的度为1;节点2、3、6的度为2;节点1和8的度是3;节点9的度是5。平均度则是这9个度数的平均值(2.22),9个度数的标准偏差是1.23,用它来衡量这9个节点的度是否正常。特别地,如果节点9的度大于平均度的程度为标准偏差的两倍,那么便可认为节点9具有非正常的大度。由于 $5 > 4.68 = 2.22 + 2 \times 1.23$ ,因此我们说节点9具有非正常的大度。同样的方法可以应用于图6.16,以量化方法测定 Sup35 在蛋白质网络图中是否发挥中心作用。

## 插图

丰富详细的插图再现并扩展了正文提供的原始信息,如下例所示。通过分析插图,读者可以增进理解并获取更多信息。

## 同伴网站上的图

书中所有的图片都能在同伴网站(<http://aw-bc/genomics>)上面找到。有些图片因为是彩图或者格式较大,在书中显示为小图,也可以在网站上找到这些图。



压力促使 900 个基因的表达聚类为 ESR  
原图见 [www.GeneticsPlace.com](http://www.GeneticsPlace.com)

## 从遗传学到基因组学

人们通常认为基因组是高容量的遗传学,但是一个基因组比起它的各部分之和更为复杂,我们必须使用新方法和新思路进行研究。最后一篇(第四篇)包含了3章的内容,帮助读者从遗传学到基因组学过渡,指出了以前“一个基因、一个蛋白、一个表型”这种看法的错误。研究案例包含了真实的数据,以便于读者理解和发现“细胞网络”的相互连接。这种探索方法将培养读者科学的分析数据、建立模型的能力。

## 致教师

电子版的《教师手册》(<http://aw-bc/genomics>)提供彩图。例如,如果学生画一幅图画或者做一副线路图,电子版将提供一个可能答案。这些图片加上网站上的图片,可用于讲课时做幻灯投影。只有采用这本书做课程使用的老师才可以拿到答案,学生不给。要密码可以联系 Benjamin Cummings 销售代表。

## 采用本书者的奖励

出版一本教科书很大的一个缺点是无法最快的更新。因此我们对采用本书作为教学用书的老师给了一个奖励。包括许多教学案例。另外,定期会更新一些新资源,如实验室课程、电子文档和以 DNA 芯片和合成生物学为重点的数据分析课程。向 Benjamin Cummings 销售代表索取教师用户名和密码。



# 前 言

最近一些年生物世界已经经历了一个令人吃惊的转型。在人类基因组测序和核酸、蛋白质及它们相关特性的数据库出现的帮助下，生物学最终加入到了化学和物理的行列成为一门真正的定量科学。

然而，许多挑战仍然摆在面前。我们才刚刚开始接触到基因组学和蛋白质组学扩展生物学知识和改进人类健康的能力。为了实现这个革命的所有利益，所有的科学家必须掌握必要的知识，找到揭开生命最大奥秘的线索，创造性地挖掘这些生物学中巨大的宝藏。

随着国家人类基因组研究中心对于基因组时代蓝图的提出，为了能够用一种整合的方式更好的理解生物通路、网络和分子系统，这就需要来自各个层面的信息。在普通层面上，目前对生物学家最主要的一个问题是如何找到一种能够同时控制一个细胞中所有基因的表达方法。更有挑战性的目标是发明一项技术能够实时的测量细胞中所有蛋白质的丰度、定位、修饰和活性。在分析层面上，目前计算方法对于现代生物学研究非常重要。随着基因组和蛋白质组数据的量和复杂程度的增加，并且随着提出的问题高度复杂，生物信息学创新的重要性注定是要增长起来的。

考虑到基因组学、蛋白质学和生物信息学的飞速发展，看上去能出版一本教科书非常难。然而，不仅仅是列出一些容易过时的内容，《探索基因组学、蛋白质组学和生物信息学》这本书已经给学生提供了扎实的知识框架来探索这些学科，使他们在科学中前进。这本书清楚地提出了基因组学和蛋白质组学的基本原理，同时也强调了生物信息学在分析这些学科产生的数据集的整合作用。

通过一种交互的、基于问题的指导性方法将生物学带到生活中来，新版在合成生物学、比较基因组学和原核基因组学上作了更多的强调。升级和扩充的部分也包含了更多的“数学备忘录”内容，学生可以练习使用统计和概率计算来增加对数据集解释的定量准确性。

课本的多学科领域集中反映了 21 世纪生物学研究逐步增加的复杂性。当我们在人类基因组计划的基础上做研究时，我们在科学未开拓领域探索的能力随着生物学知识和计算机科学、物理、数学、临床研究、生物伦理学和一些其他学科的融合而改变。这种交互在单个层面、合作层面甚至在学科层面（在传统研究的交互区域出现新学科的时候）都是需要的。因此，现在的生物学课程需要张开一张大网，来发现学生们表现出的不同兴趣、技巧和观点。

这本书另外一个有远见的部分就是“问题探讨”，通过一些标准的科学提问，来鼓励学生更深地钻研随着基因组学和蛋白质组学带来的伦理、法律和社会问题。每位生物学家在着手一个项目或者提倡一种新技术之前，一定要考虑到在伦理、法律和社会应用方面的责任，这一点越来越重要。

很明显，一片美好的前景出现在下一代生物学家面前。现在很难去预测这些未来的探索会在什么方面或者会有什么样的发现。但是有一件事情是确定的：扎实地理解基因组学、蛋白质学和生物信息学对于在这个令人惊奇的新世界中成功非常关键。

## 致 谢

Davidson 学院的同事给了大力支持。Clark Ross, Bobby Vagt, Verna Case 的支持和领导是这本书成书的基础。许多系的同事给予了作者多种形式的支持, 他们是 Karen Bernd, Karen Hales, Dave Wessner, Barbara Lom, Pam Hay, Julio Ramirez, David Brown, Jeanne O'neill, Suzanne Churchill, Chris Paradise, Mur Muchane, Peggy Maiorano 和 Betty Hartsell。感谢 Waksman 基金、国家科学基金、Duke 基金和南部大学联盟提供资金支持。最重要的, 我要感谢我的 14 位勇敢的学生, 本书的第一稿就是他们写的, 他们是: Sean Burke, Amber Hartman, Ben Havard, Julie Hwang, Dennis Jones, Peter Lowry, Jennifer Madden, Emily Oldham, La Powell, Lisa Robinson, Elizabeth Sellars, Liz Shafer, J. D. Willson 和 Marisa Wilson。

本书第一稿成稿之时, 我正在西雅图度假, Mary Claire King、Maynard Olson 和 Leroy Hood 在这段时间给予了极大支持。华盛顿基因组中心的多位成员给了我技术上和精神上的支持。系统生物学研究所 (ISB) 的人位建立了一个极好的学习基因组学和蛋白质组学的环境。另外, Eric Davidson 在 ISB 年会上的演讲给了我很大启发。同时也感谢全世界这领域的研究者, 因为能够使我免费共享他们的工作成果并且受到其热忱的帮助, 能够成为这个知识充分共享的社会的一个成员令我感到非常欣慰。就个人而言, 我在西雅图的朋友和他们的家庭使我感到宾至如归, 特别是 Hope-Young 一家、Hill 一家、Ginger Armbrust 和 Susan Francis, Walker 一家, 以及在 Meadbrook Co-op 幼儿园的 20 个家庭。

每个项目都需要一大批人的参与与创新, 这本书从最初模糊的想法到最后成稿能如此迅速, 得益于我的朋友 Michele Sordi 的指导, 没有她就不会有如此硕果。本书的出版也归功于 Peggy Williams 的勤奋和公正。她的劳动对于这样一本新领域中的新书的问世至关重要。另外, 在 John Inglis 的领导下, 冷泉港实验室出版社的编辑们工作非常敬业。在 Benjamin Cummings, 许多人对这本书作出了重要的贡献, 包括 Steve McEntee 和 Mary Ann Tenorio, 他们的艺术才华使得书中的图片显得十分真实, 更方便用于教学。Benjamin Cummings 的 Jamie Sue Brooks、Larry Lazopoulos 和 Left Coast Group 成员的设计方案令人赞叹。感谢 Andrew Ogus 给本书设计外观, 感谢 Yvo Rezebios 为本书设计封面, Michael McArdle 作为出版助理, 很好地使每个人的工作团结合作起来。

最后, 最重要的支持来自于离家最近的人们。这本书是一位非常了解生物学的教学家和一位尽力了解数学的生物学家彻夜合作的结晶, 为此, 特别感谢 Laurie。感谢 Genome Consortium for Active Teaching (GCAT) 的成员们提供了我所需的大学支持体系。Susan、Paulina 和 Celeste 长途跋涉了两次, 给予我继续完成本书的毅力。我的大家庭也给了我极大支持, 没有他们, 我将还呆在工厂里制作地板瓷砖。

感谢数学系的同事们,尤其是系主任 Stephen Davis 和好朋友 Donna Molinek 给予了我无穷的支持和鼓励。我的学生们是源源不断的力量和灵感的源泉,我尤其受惠于计算生物学的学生们: Frank Chemotti, Amber Hartman, Soren Johnson, Jennifer Kawwass, Peter Leese, Rachel Patton McCord, Celilia Mendiondo, Emily Oldham, Ma Powell, Talbot Presley, Stanley Prybe, Lang Robertson, Lisa Robinson 和 Megan Shafer。这个多学科交叉的团队制作了一级 Kyte Doolittle 和层次聚类的网页。基因组学和计算生物学的学生都测试了许多数学备忘录,尤其感谢对撰写数学备忘录 6.1 有帮助的那些问题的提出者。

我至今仍然能够从那些引导我深入该领域的老师和朋友们那里获得重要的灵感,他们是: Mike Waterman, Simon Tavare, Gary Stormo 和 John Williamson。他们与我的朋友们和家人一起,让我能对这本书有所贡献。Malcolm, 一个愿意与新同事分享他的基因组学观点的人,让这一切成为现实。最后,我感谢我的丈夫 Bill,他是如此的耐心和如此的爱我,他给了我最大的支持。

——Laurie J. Heyer

# 目 录

译者序  
序  
前言  
致谢

## 第一部分 基因组序列

第 1 章 我的孩子究竟怎么了? .....	3
1.1 首先要关心的问题 .....	3
1.1.1 第一阶段:临床表现 .....	3
1.1.2 第二阶段:家庭病史研究 .....	5
1.1.3 第三阶段:染色体组型及连锁分析 .....	5
1.1.4 第四阶段:DNA 序列分析 .....	8
数学备忘录 1.1 什么是 E 值? .....	9
本节小结 .....	10
1.2 了解这一疾病的下一步工作 .....	11
1.2.1 我们需要一个动物模式系统 .....	11
1.2.2 同时检索出的另一种蛋白究竟是什么? .....	12
1.2.3 Utrophin 蛋白是否也参与了肌营养不良症的形成? .....	12
1.2.4 Dystrophin 蛋白究竟具有什么样的功能? .....	13
数学备忘录 1.2 这幅图究竟有什么特别? .....	16
1.2.5 为什么 DMD 的病人在出生 3 年后就会发生肌肉坏死? .....	17
1.2.6 是否可能患有 DMD 症却不出现肌营养不良的症状呢? .....	18
1.2.7 如果他们的 <i>dystrophin</i> 基因是正常的,那他们为什么会出现肌营养不良的症状呢? .....	18
数学备忘录 1.3 很不可能是什么意思? .....	19
1.2.8 肌肉营养障碍的研究领域在何处? .....	25
1.2.9 6 个造成肌肉营养障碍的分子起因的议题 .....	25
1.2.10 会议的开始 .....	25
1.2.11 结构上的弱点 .....	26
1.2.12 功能缺失型突变 .....	28
1.2.13 新范例:肌肉营养不良的非结构性的致病原因 .....	30
数学备忘录 1.4 cGMP 的产生是否增多了? .....	33
1.2.14 最后一场演讲 .....	34
本节小结 你最终的想法 .....	35

本章总结 .....	36
<b>第 2 章 基因组序列的获取和分析</b> .....	<b>38</b>
2.1 基因组是如何被测序的? .....	38
2.1.1 什么是基因组学? .....	39
2.1.2 基因组是如何被测序的? .....	39
数学备忘录 2.1 你可以从一个点图中了解什么? .....	46
数学备忘录 2.2 你怎样寻找模板? .....	49
2.1.3 我们可以从 DNA 序列中预测蛋白质功能吗? .....	54
数学备忘录 2.3 什么是“positives”,它们和 E-value 有何关系? .....	54
2.1.4 这些蛋白质什么形状? .....	57
2.1.5 结构提示了功能吗? .....	57
2.1.6 为什么数据库中有那么多不完整序列? .....	58
2.1.7 在线注释基因组 .....	63
2.1.8 一个基因可以产生多少蛋白? .....	64
2.1.9 基因组能否不改变 DNA 序列来改变基因表达 .....	65
2.1.10 DNA 的第五种碱基是什么? 甲基胞嘧啶 .....	67
2.1.11 印迹,甲基化和癌症 .....	68
本节小结 .....	69
2.2 我们从单细胞基因组中学到了什么? .....	70
2.2.1 为什么我有这么多的粉刺? .....	70
2.2.2 哪一个基因导致粉刺? .....	72
2.2.3 我们体内所有细菌都是有害的吗? .....	74
2.2.4 微生物基因组是否可以依赖于人类基因? .....	76
2.2.5 最小基因数可能是多少? .....	77
2.2.6 所有病毒基因组都小于细菌基因组吗? .....	79
2.2.7 Mimivirus 是活的吗? .....	80
2.2.8 基因组是否反映生物的生态位(ecological niche)? .....	81
数学备忘录 2.4 你能估计点阵图中倒置次数吗? .....	84
2.2.9 为什么 MED4 的基因组这么小? .....	87
2.2.10 多大的基因组变化导致新的物种? .....	88
2.2.11 什么生物导致疟疾(Malaria)? .....	89
2.2.12 疟原虫有什么样的基因组? .....	91
2.2.13 预测的蛋白质组同样不寻常吗? .....	92
2.2.14 是否存在真核生物的模式基因组? .....	95
2.2.15 研究者对未来基因组研究作何预测? .....	98
2.2.16 酵母基因组的尾声 .....	98
本节小结 单细胞基因组分析 .....	101
2.3 我们从多细胞生物基因组中学到了什么? .....	101
2.3.1 动物基因组更难完成吗? .....	101

什么是多线染色体(polytene chromosomes)?	103
2.3.2 果蝇与其他真核生物有何不同?	105
2.3.3 果蝇仍然是一个好的模式生物吗?	105
2.3.4 果蝇基因组尾声	107
2.3.5 我们需要两个植物基因组序列吗?	108
2.3.6 植物看起来比动物简单,它们的基因组也如此吗?	109
2.3.7 我们能从草图中得出什么结论?	111
2.3.8 我们学到了什么?	113
2.3.9 水稻尾声	113
2.3.10 我们可以从河豚 puffer fish 基因组中学到什么?	116
2.3.11 这个基因组有什么惊人之处吗?	117
2.3.12 可以从 Tetraodon 学到什么更重要的?	117
2.3.13 什么使得人类与众不同?	119
数学备忘录 2.5 你怎样用一条直线来拟合数据?	121
2.3.14 我们能否描述一个典型的人类基因?	123
2.3.15 人类基因组尾声	127
2.3.16 人类基因组的下一个目标是什么?	129
本节小结 多细胞基因组分析	131
本章总结	131
<b>第3章 进化论和医学研究中的比较基因组学</b>	<b>134</b>
3.1 比较基因组学	134
3.1.1 致命的大肠杆菌怎么能够同时存在于我们的肠道中呢?	135
数学备忘录 3.1 你怎么区分碱基组成的不同?	136
3.1.2 二百个基因组:对于原核生物,我们能够从比较基因组学中得到什么?	138
3.1.3 所有的原核生物都只含有一个圆形的染色体吗?	138
3.1.4 基因组是否仍在变化?	140
3.1.5 到底有多少基因组?	141
3.1.6 通过比较不同物种的整个基因组我们能够得到什么?	145
3.1.7 从染色体的角度我们可以得到什么?	148
本节小结	151
3.2 基因组的进化	151
3.2.1 什么物种位于生命树的根部呢?	151
3.2.2 我们细胞核中基因的来源是什么?	156
数学备忘录 3.2 引用计数的差别是否显著?	159
3.2.3 基因组进化过程中有中间步骤存在的证据吗?	161
3.2.4 你要吃那个东西吗?	162
3.2.5 圣经完美平衡中的缺失环节?	165
3.2.6 离开了共生,细胞核能够进化吗?	167
3.2.7 我们人类的起源是什么?	170

3.2.8 我们都是非洲人的后裔吗? .....	171
数学备忘录 3.3 怎么知道进化树是正确的? .....	173
3.2.9 我们停止进化了吗? .....	175
本节小结 .....	176
3.3 基因组鉴定 .....	176
3.3.1 我们怎样检查生物武器? .....	176
3.3.2 DNA 能存活多久? .....	181
3.3.3 肺结核是怎样进入北美的? .....	184
3.3.4 怎样鉴定新疾病? .....	189
3.3.5 有什么其他疾病即将爆发? .....	194
本节小结 .....	196
3.4 生物学基因组研究 .....	197
3.4.1 我们能利用基因组序列制造新疫苗吗? .....	197
3.4.2 我们能够开发出新型抗生素吗? .....	200
3.4.3 我们能够发明新型药物吗? .....	202
3.4.4 有任何方法可选来抑制 RNA 么? .....	205
3.4.5 还有更稳定的 RNA 基因组靶么? .....	206
本节小结 .....	209
本章总结 .....	209
<b>第 4 章 基因组变异</b> .....	<b>213</b>
4.1 环境因素案例分析 .....	213
基因组多样性能影响全球变暖吗? .....	214
数学备忘录 4.1 如何衡量遗传变异? .....	217
数学备忘录 4.2 你怎样模拟种群的差异(多样性)? .....	219
本节小结 .....	223
4.2 人类基因组变异 .....	223
4.2.1 人类基因组有多大差异? .....	223
突变体和等位基因之间有什么不同? .....	225
数学备忘录 4.3 所有 SNP 都真的是 SNP 吗? .....	226
4.2.2 我们为何关注 SNP? .....	228
4.2.3 SNP 引起疾病的案例 .....	231
4.2.4 在非疾病 QTL 中存在由 SNP 导致的变化吗? .....	234
知识框 4.1 专利法和基因组学 .....	235
4.2.5 为什么 SNP 如此狂热? 药物基因组学! .....	236
本节小结 .....	239
4.3 最终的基因组表型——死亡? .....	239
4.3.1 我们为什么衰老? .....	239
4.3.2 延长寿命有隐含的成本吗? .....	241
4.3.3 细菌也有基因组的利弊平衡吗? .....	242

本节小结 .....	245
4.4 基因组变异的伦理后果 .....	245
4.4.1 基因改造生物有害吗? .....	245
4.4.2 遗传检测有益吗? .....	248
数学备忘录 4.4 阳性测试结果表示什么? .....	251
4.4.3 基因组多样性库和小群体 .....	253
4.4.4 谁将从基因组医学中获益? .....	254
4.4.5 复杂基因组的简单应用存在吗? .....	255
知识框 4.3 我需要做遗传检测吗? .....	255
4.4.6 人类可以被克隆吗? .....	257
本节小结 .....	259
本章总结 .....	259

## 第二部分 基因组表达

<b>第 5 章 为什么我们不能只通过服食一颗药丸来实现减肥计划</b> .....	265
对知识的渴望 .....	265
10月21日,星期六,早上7:30 .....	265
周六图书馆早上8:30开门 .....	266
为体重的动态平衡构建模型 .....	267
初步的结论 .....	268
克隆肥胖基因 .....	268
肥胖蛋白 leptin 的功能检测 .....	272
是该去和祖母解释的时间了 .....	272
祖母给我留了家庭作业! .....	274
本章总结 .....	284
<b>第 6 章 DNA 芯片的基础研究</b> .....	286
6.1 DNA 芯片简介 .....	286
6.1.1 我的“家酿”怎么了? .....	287
6.1.2 探针在哪里? .....	289
6.1.3 芯片数据看起来不错,但是真实可靠吗? .....	291
6.1.4 你如何分析这些数据? .....	292
数学备忘录 6.1 怎样对数据进行处理以避免分数比值? .....	293
数学备忘录 6.2 怎样评估表达模式的相似性? .....	294
数学备忘录 6.3 怎样进行基因聚类? .....	296
6.1.5 芯片能够发现调控序列吗? .....	302
6.1.6 我们能用这些数据推导可供验证的预测吗? .....	303
6.1.7 芯片看起来很好,但是可能不真实,是这样吗? .....	305
6.1.8 从经受外界环境变化的酵母身上,我们能够了解到什么? .....	307
6.1.9 菌类感受到压力了吗? .....	308



6.1.10	发现了什么……	309
6.1.11	为什么有些基因有很多拷贝而其他基因则没有？	310
6.1.12	启动子怎么调控基因表达？	311
6.1.13	启动子可以逆向工作吗？	313
	本节小结	313
6.2	DNA 芯片的其他应用	314
6.2.1	为什么许多无关的基因会拥有相同的表达谱？	314
	数学备忘录 6.4 比较基因表达矩阵列有什么用？	315
6.2.2	细胞能够对自己的基因加以验证吗？	318
6.2.3	预测的基因是真实基因还是假基因？	319
6.2.4	DNA 芯片能够改善基因的注释么？	320
6.2.5	DNA 芯片能够验证全基因组的注释么？	320
	本节小结	322
	本章总结	322
<b>第 7 章</b>	<b>DNA 芯片的应用研究</b>	<b>324</b>
7.1	癌症和基因组芯片	324
7.1.1	有更好地诊断癌症的方法吗？	325
	数学备忘录 7.1 什么是签名基因？如何使用签名基因？	327
7.1.2	能否用生物芯片对乳腺癌也进行分类呢？	330
7.1.3	癌细胞中的基因组发生了什么样的变化呢？	333
7.1.4	我们能否对数据进行验证？	336
	本节小结	337
7.2	利用 DNA 芯片改进卫生保健	337
7.2.1	为什么肺结核疫苗效果减弱？	337
7.2.2	我们能够对每种癌症选择最有效的治疗么？	341
7.2.3	我们能否预知化疗的有效性？	342
7.2.4	脂肪的积累会发生什么？	343
7.2.5	Leptin 对于 <i>wt</i> 小鼠的脂肪组织有什么作用？	347
	本节小结	349
	本章总结	349
<b>第 8 章</b>	<b>蛋白质组学</b>	<b>351</b>
8.1	前言	352
8.1.1	所有这些蛋白质能做什么？	352
8.1.2	这些蛋白质到底定位在细胞的什么地方？	357
8.1.3	不同的培养条件下需要哪些不同的蛋白质？	358
	数学备忘录 8.1 如何得知样品中有足够数量的细胞？	361
	本节小结	363
8.2	蛋白质三维结构	364
8.2.1	蛋白质的结构能否揭示它的功能？	364