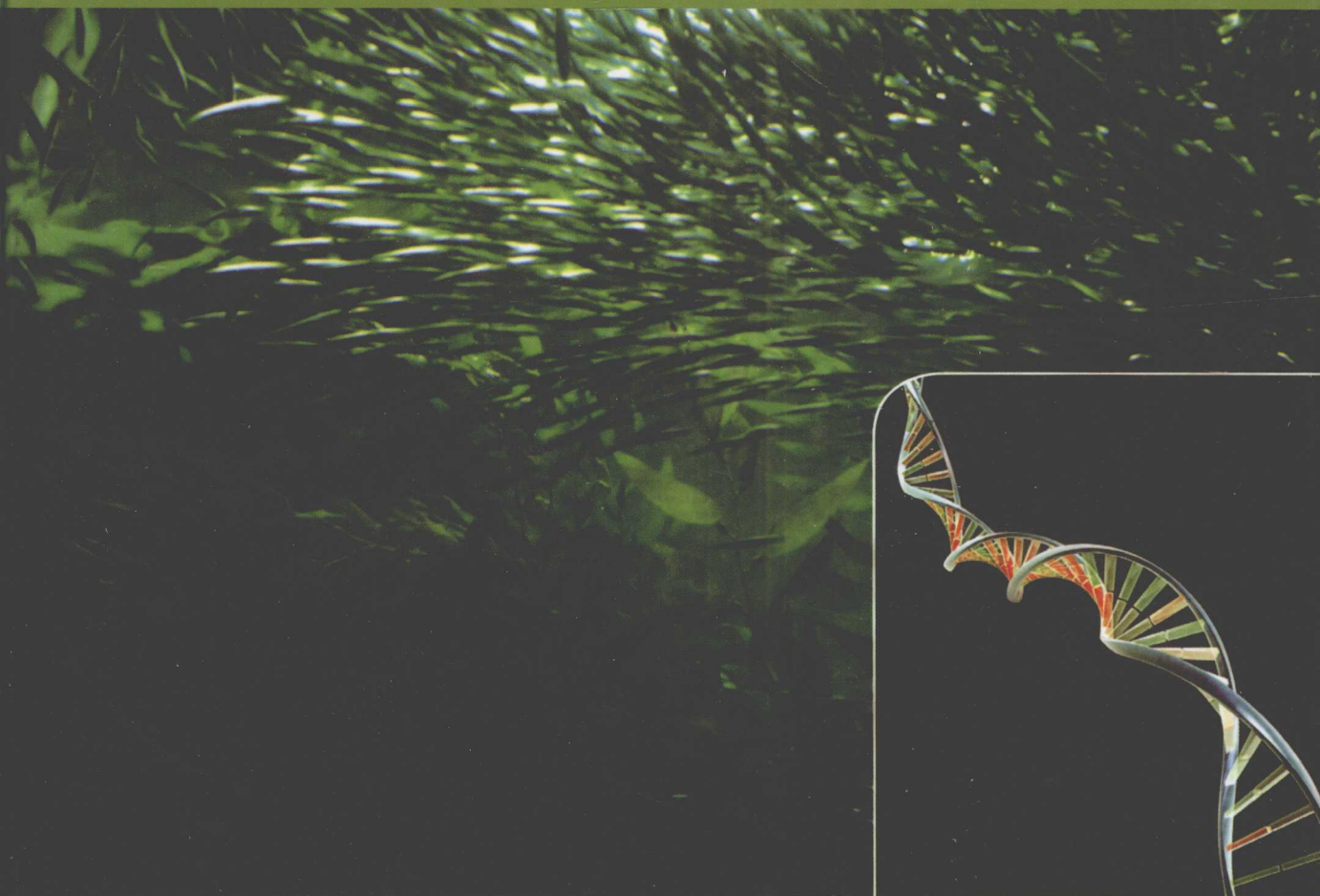


水产基因组学技术

AQUACULTURE GENOME TECHNOLOGIES



[美] 刘占江

主编

Zhanjiang (John) Liu

鲍宝龙 王志勇 张士瑾 等译



化学工业出版社

水产基因组学技术

AQUACULTURE GENOME TECHNOLOGIES

[美] 刘占江 主编

Zhanjiang (John) Liu

鲍宝龙 王志勇 张士瑾 等译



化学工业出版社

· 北 京 ·

本书云集了当前国际上从事水产动物基因组学研究的一线科学家,基于理论联系实际的原则,根据水产动物基因组的特点,详细介绍了基因组学各种技术及其原理,涵盖了水产动物基因组标记技术、基因组作图技术和全基因组测序技术等各个方面。

本书极具权威性和适应性,可作为水产养殖、动物育种、海洋生物技术、生物学等专业的教学用书,更是从事分子育种、海洋生物技术等研究人员必备的工具书。

图书在版编目(CIP)数据

水产基因组学技术/[美]刘占江主编;鲍宝龙等译. —北京:化学工业出版社,2011.1

书名原文: Aquaculture Genome Technologies

ISBN 978-7-122-09774-3

I. 水… II. ①刘… ②鲍… III. 水产动物-基因组-研究 IV. S96

中国版本图书馆CIP数据核字(2010)第207392号

Zhanjiang (John) Liu is a Distinguished Alumni Professor, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, and Director, Aquatic Genomics Unit at Auburn University.

Aquaculture Genome Technologies/by Zhanjiang (John) Liu.

ISBN 978-0-8138-0203-9

Copyright© 2007 by Blackwell Publishing. All rights reserved.

Authorized translation from the English language edition published by Blackwell Publishing Ltd, Oxford.

Responsibility of the accuracy of the translation rests solely with Chemical Industry Press and is not the responsibility of Blackwell Publishing Ltd.

本书中文简体字版由 Blackwell Publishing Ltd 授权化学工业出版社独家出版发行。

未经许可,不得以任何方式复制或抄袭本书的任何部分,违者必究。

北京市版权局著作权合同登记号: 01-2008-5017

责任编辑: 刘亚军

文字编辑: 高霞

责任校对: 宋玮

装帧设计: 刘丽华

出版发行: 化学工业出版社(北京市东城区青年湖南街13号 邮政编码100011)

印刷: 北京永鑫印刷有限责任公司

装订: 三河市万龙印装有限公司

787mm×1092mm 1/16 印张23½ 字数596千字 2011年4月北京第1版第1次印刷

购书咨询: 010-64518888(传真: 010-64519686) 售后服务: 010-64518899

网址: <http://www.cip.com.cn>

凡购买本书,如有缺损质量问题,本社销售中心负责调换。

定 价: 98.00 元

版权所有 违者必究

撰稿人名单

Jonas S. Almeida

Department of Biostatistics and
Applied Mathematics
University of Texas
MD Anderson Cancer Center, Unit 447
1515 Holcombe Boulevard
Houston, TX 77030 USA

Louis Bernatchez

Réseau Aquaculture Québec (RAQ)
Pavillon C-H Marchand
Université Laval
Québec, QC
Canada G1K 7P4

Robert W. Chapman

Marine Resources Research Institute
South Carolina Department of
Natural Resources
Charleston, SC 29412 USA

Alan Christoffels

Computational Biology Group
Temasek Life Sciences Laboratory
1 Research Link
National University of Singapore
Singapore

Karl J. Clark

Department of Animal Sciences
University of Minnesota
St. Paul, MN 55108 USA

Avner Cnaani

Hubbard Center for Genome Studies
University of New Hampshire
Suite 400, Gregg Hall
35 Colovos Road
Durham, NH 03824 USA

Glenn A. Cooper

Centre for Biomedical Research
University of Victoria
Victoria, British Columbia
Canada V8W 3N5

Jason P. Curole

Department of Biological Sciences
University of Southern California
3616 Trousdale Parkway, AHF 107
Los Angeles, CA 90089-0371 USA

Roy G. Danzmann

Department of Integrative Biology
University of Guelph
Guelph, Ontario
Canada N1G 2W1

William S. Davidson

Department of Molecular Biology
and Biochemistry
Simon Fraser University
8888 University Drive
Burnaby, British Columbia
Canada V5A 1S6

Chunguang Du

Department of Biology and
Molecular Biology
Montclair State University
Montclair, NJ 07043 USA

Pierre Duchesne

Réseau Aquaculture Québec (RAQ)
Pavillon C-H Marchand
Université Laval
Québec, QC
Canada G1K 7P4

Karim Gharbi

Division of Environmental and
Evolutionary Biology Institute of
Biomedical and Life Sciences
University of Glasgow
Glasgow, Scotland UK G12 8QQ

Paul J. Grobler

Faculty of Natural and
Agricultural Sciences
University of the Free State
P.O. Box 339, Bloemfontein 9300
South Africa

Ximing Guo

Haskin Shellfish Research Laboratory
Institute of Marine and Coastal Studies
Rutgers University
6959 Miller Avenue
Port Norris, NJ 08349 USA

Perry B. Hackett

Department of Genetics, Cell Biology
and Development
Arnold and Mabel Beckman Center for
Transposon Research
6-106 Jackson Hall
University of Minnesota
Minneapolis, MN 55455 USA

Eric M. Hallerman

Department of Fisheries and
Wildlife Sciences
Virginia Polytechnic Institute and State
University Blacksburg, VA 20461-0321
USA

Limei He

Department of Soil and Crop Sciences
Texas A&M University
College Station, TX 77843 USA

Dennis Hedgecock

Department of Biological Sciences
University of Southern California
3616 Trousdale Pkwy, AHF 107
Los Angeles, CA 90089-0371 USA

Jess W. Jones

U.S. Fish and Wildlife Service
Blacksburg, VA 24061-0321 USA

Thomas D. Kocher

Hubbard Center for Genome Studies
University of New Hampshire
Suite 400, Gregg Hall
35 Colovos Road
Durham, NH 03824 USA

Ben F. Koop

Centre for Biomedical Research
University of Victoria
Victoria, British Columbia
Canada V8W 3N5

Abraham Korol

Institute for Evolution
Haifa University
Haifa, Israel 31905

Huseyin Kucuktas

Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Bo-Young Lee

Hubbard Center for Genome Studies
University of New Hampshire
Suite 400, Gregg Hall
35 Colovos Road
Durham, NH 03824 USA

Yaning Li

Department of Plant Pathology
Agricultural University of Hebei
Biological Control Center of Plant
Disease and Plant Pests of Hebei
Province
Baoding, China 071001

Lei Liu

W.M. Keck Center for Comparative
and Functional Genomics
University of Illinois at
Urbana-Champaign
330 Edward R. Madigan Laboratory
1201 W. Gregory Dr.
Urbana, IL 61801 USA

Zhanjiang Liu

Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Eric Peatman

Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Caird E. Rexroad III

USDA/ARS National Center for Cool
and Cold Water Aquaculture
11861 Leetown Road
Kearneysville, WV 25430 USA

Matthew L. Rise

The Ocean Sciences Centre
Memorial University of Newfoundland,
1 Marine Lab Road
St. John's, NL
Canada A1C 5S7

Max F. Rothschild

Department of Animal Science and the
Center for Integrated Animal Genomics
2255 Kildee Hall
Iowa State University
Ames, IA 50011 USA

Anatoly Ruvinsky

The Institute for Genetics and
Bioinformatics
University of New England
Armidale, Australia NSW 2351

Chantel Scheuring

Department of Soil and
Crop Sciences
Texas A&M University
College Station, TX 77843 USA

Andrey Shirak

Agricultural Research Organization
Institute of Animal Science
Bet Dagan, Israel 50250

Kristian R. von Schalburg

Centre for Biomedical Research
University of Victoria
Victoria, British Columbia
Canada V8W 3N5

Shaolin Wang

Department of Fisheries and

Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Yongping Wang

Haskin Shellfish Research Laboratory
Institute of Marine and Coastal Studies
Rutgers University
6959 Miller Avenue
Port Norris, NJ 08349 USA

Gregory W. Warr

Department of Biochemistry and
Molecular Biology
Marine Biomedicine and Environmen-
tal Sciences Center
Hollings Marine Laboratory
Medical University of South Carolina
Charleston, SC 29412 USA

Peng Xu

Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Zhe Xu

Haskin Shellfish Research Laboratory
Institute of Marine and
Coastal Studies
Rutgers University
6959 Miller Avenue
Port Norris, NJ 08349 USA

Hong-Bin Zhang

Department of Soil and Crop Sciences
Texas A&M University
College Station, TX 77843 USA

《水产基因组技术》

翻译人员

主要译校人员及联系方式

鲍宝龙 上海海洋大学, email: blbao@shou.edu.cn

王志勇 集美大学, email: zywang@jmu.edu.cn

张士瑾 中国海洋大学, email: sczhang@ouc.edu.cn

其他译校人员及联系方式 (按汉语拼音顺序排列)

龚小玲, 上海海洋大学, email: xlgong@shou.edu.cn

韩芳, 集美大学, email: hanfangyc@jmu.edu.cn

李家乐, 上海海洋大学, email: jlli@shou.edu.cn

刘贤德, 集美大学, email: xdliu@jmu.edu.cn

丘高峰, 上海海洋大学, email: gfqiu@shou.edu.cn

王成辉, 上海海洋大学, email: wangch@shou.edu.cn

魏华, 上海海洋大学, email: hwei@shou.edu.cn

武祥伟, 中国海洋大学, email: wxw2050@sina.com

姚翠鸾, 集美大学, email: clyao@jmu.edu.cn

邹曙明, 上海海洋大学, email: smzou@shou.edu.cn

中译本序

分子生物学从上个世纪后半页萌生以来仅有非常短暂的历史。但是，人类在过去半个世纪对生命本质的理解，远远超过以往历史长河积累的总和。分子生物学成果与技术的开发和利用，使基于分子设计的动植物遗传改良变为现实，一批抗病、抗虫、快速生长、高产优质的转基因动植物新品系相继出现，翻开了人类育种史上全新的一页。

人类对未知的探索和新知识的创造，继续以前所未有的速度前进。新近出现的现代基因组研究，几乎突然之间渗透到生命科学的所有领域。一个个基因组学的先进研究手段和一批批基因组研究的成果不断涌现，簇拥着基因组革命时代的到来。利用基因组学技术全面揭示基因组的整体结构、基因分布和演化、基因功能以及功能调控网络等已逐渐成为现实。与之相应的，一批以模式和经济水产动物为先导的水生生物全基因组研究已经拉开序幕。

美国奥本大学刘占江教授编著的《水产基因组学技术》一书，全面介绍了基因组学的理论、技术及其在水产生物中的应用，内容全面、出版及时。刘占江20世纪80年代中期留学美国，获明尼苏达大学博士学位，长期从事水生生物基因组学研究。他学术造诣深厚，是国际上水生生物基因组学研究的领军人物之一。本人曾与他有过密切合作，他作为一个华人的豪情和自信、过人的精力、勤勉治学的态度和对事业孜孜不倦的追求，给我留下十分深刻的印象。这样一本好书的出版，正是刘教授长期努力钻研结晶的印证。

而今，这本书已由上海海洋大学鲍宝龙教授、集美大学王志勇教授、中国海洋大学张士璠教授等译为中文。我相信，中译本的出版将为推动我国水产基因组学的研究发挥积极作用。作为世界上最大的水产国家的科学技术人员，我们有责任不失时机地开展水产基因组学的研究，推动水产科学和技术与时俱进，为经济社会发展、为全世界的水产事业做出应有的贡献。

谨此，我向全国农业、水产及生物科学技术工作者及大专院校的本科生、研究生推荐这本值得一读的好书。

朱作言

中国科学院院士、发展中国家科学院院士

2010年11月10日

译者的话

在过去的十余年，基因组学已经给生命科学带来革命性的影响，相关技术和方法已经深入医学、农业科学、进化生物学等各个领域。目前，一些重要水产养殖物种的全基因组测序工作已经开始，有的已经接近完成。如何利用大量基因组数据信息，为选育具有优良水产养殖性状的品种服务，对于水产养殖基础生物学家、遗传学家、育种工作者、渔业管理人员等来说，都是一个全新的挑战。虽然已有众多介绍基因组学的书籍，但水产养殖对象广泛，又有各自的特点，因此，特别需要有一部针对水产动物基因组的理论和技术方面的书籍。

由 Wiley-Blackwell 出版社出版、刘占江主编的《Aquaculture Genome Technologies》，作者汇集了全世界水产基因组学领域的著名科学家。刘占江是国际上从事水产动物基因组学和生物技术研究权威的专家，各章作者也是在基因组学、水产养殖和渔业科学领域富有经验的专家。该书根据各水产动物基因组的特点，详细介绍了基因组学各种技术及其原理，涵盖了水产动物基因组标记技术、基因组作图技术和全基因组测序技术等各个方面，对于国内目前如火如荼的水产动物基因组学的研究，必将会起到重要的指导和推动作用。

为了使读者更好地理解该书的内容，我们邀请了上海海洋大学、集美大学和中国海洋大学从事水产动物基因组学和水产生物技术与教学的老师组成了本书的翻译团队。本书翻译的分工是：鲍宝龙译前言和第 1、14、15 章，校第 2、4、5、7、22、23、26、28、29 章；谢彩霞校前言、第 1、14、15 章；王成辉译第 2、3、7 章，校第 3 章；陈琴译第 2 章；王志勇译第 4、5 章，校第 6、8、9、10、11、12、13、18、24、27 章；张静译第 6 章；翁朝红译第 9 章；刘贤德译第 8、9、10、11、12、18 章；韩芳译 13 章；郭华荣译第 16 章；张士璀校第 16、17、19、20、25 章；李红岩译 17 章；刘振辉译 19 章；梁宇君译第 20 章；丘高峰译第 21、22 章；张恩帆校第 21 章；沈春燕校第 22 章；魏华译第 23 章；姚翠鸾译第 24、27 章；汲广东译第 20、25 章；龚小玲译第 26 章；邹曙明译第 28 章；李家乐译第 29 章。全书最后由鲍宝龙、王志勇和张士璀三位教授修改定稿。感谢中国科学院院士朱作言先生为本书中文版作序，感谢原著主编刘占江教授在本书翻译过程中给予的指导和鼓励！化学工业出版社在该书的引进、翻译到出版的整个过程都提供了相当大的支持。在此一并表示感谢！本翻译书稿的出版得到了上海市水生生物学重点学科的资助！

水产动物基因组学技术是一门新生的交叉技术，由于时间仓促，谬误在所难免，切望读者不吝赐教。

译者

2011 年 1 月

英文原著序

15年前,受人类基因组计划启动的激发,畜牧基因组学诞生了。人类基因组学技术和大量基因组比较分析数据可应用到畜牧品种,大多数的比较分析数据来自哺乳动物。得益于这些全基因组测序计划的努力成果,已经完成或接近完成的有鸡、奶牛、猪、马的测序,以及其他如绵羊和火鸡的图谱资源。经济性状和生理学差异特征已经被定位到图谱上,相关基因正逐步被发现。在技术发展与基因组信息在开发食品上的利用方面,水产养殖物种的生物多样性给基因组学研究带来了独特的难题。现在,本书展示基因组技术在水产养殖物种中的应用和在一些主要物种中的快速进展,是对动物基因组学文献的有益补充。

水产养殖对象范围很广,既有脊椎动物又有无脊椎动物,所涉及的问题较多,如基因组大小、基因组冗余和各种繁殖策略等。除了硬骨鱼类能从斑马鱼和河鲢基因组学研究中得到帮助外,其他水产动物基因组研究很难从人类到各种哺乳类乃至鸡基因组数据中得到有益的比较基因组资料。然而,DNA标记、遗传连锁图谱和物理图谱、转录组分析技术等基因组技术是有普遍应用价值的,现在正在热议的每Gb几千美金的测序费用的新测序技术的发展也是值得期待的。在水产养殖品种的基因组技术开发、图谱构建和最终的全基因组测序等方面的努力,不仅对提高水产食品来源的健康指数和产量有很多好处,最为重要的是有助于阐明水产养殖物种的基因组和生理学多样性,解决其遗传学研究上面临的许多令人头痛的问题。

在许多水产养殖对象的基因组图谱构建和功能研究方面,已经开发出了许多令人惊讶的技术。随着高水平测序技术潜力的开发,水产基因组学的前途是光明的。作为一个喜欢钓鱼和海鲜的从事哺乳动物研究的遗传学家,我为本书所介绍的水产基因组技术的进展和前景感到高兴。

James E. Womack

美国科学院院士,德州农工大学杰出教授

前言

人类基因组计划的完成让全世界为之振奋，也引发了一场基因组学革命。伴随这场革命的是生命科学领域研究方式的完全改变；这场由基因组学引起的革命，也不例外地影响到水产基因组学和水产遗传学总体上的研究。最近 10 年，在全世界还没有一个大规模的水产基因组测序计划！1997 年秋天，在曼彻斯特的达尔茅斯成立的第一个水产基因组工作小组，可能是水产基因组学官方主导的一个开端。今天，有几个重要水产养殖物种的全基因组测序工作即将开始，这对水产养殖遗传学家、育种工作者、渔业管理人员等提出了新的挑战，那就是如何最有效地利用目前已有的大量基因组数据信息，如何在水产养殖和渔业上管理和运用迅速发展的基因组技术。

本书目的是为将来的水产养殖和渔业科学家介绍基因组技术，也为农业科学专业的学生提供合适的参考书。写这本书，有这样几个紧迫的原因。首先，你很容易找到有关基因组学的书籍，但很难发现非常详尽介绍基因组学技术概念和基本原理的书籍。本人的学术背景是农业，但我近年来主要从事基础基因组学的研究，本人的经验，加上为研究生讲授农业基因组学课程的经历，使我能有效地抓住基因组学的关键问题，即对基因组技术的理解是必需的，假如能对基因组技术的基本原理有非常清楚的解释，将会更加有帮助，因为许多学生没有系统地学习过分子生物学、遗传学、生物化学、生物信息学等。其次，大多数基因组学书籍利用经典的模式动物来介绍纯粹的基因组学方法，没有考虑到基因组技术在某些实际情况下的潜在的应用前景，在如何把基础的基因组学运用到人类健康研究之外的领域方面，亦存在严重的不足；本书详尽介绍了基因组技术以及这些技术在水产养殖和渔业上的应用。最后，水产养殖和渔捞对象有其独特的生物学特性，目前的基因组技术需要适当的改进和调整以后才能更好地运用。尽管本书没有任何章节是有关起源于水产养殖或渔捞对象的全新的基因组技术，但几乎每一章节都是介绍基因组技术在水产养殖和渔业或农业上的具体运用。

本书共分 29 章，分别由来自全世界的著名科学家撰写，他们在基因组学、水产养殖和渔业领域均富有经验，能从独特的角度探讨基因组技术。这对于专业研究人员、研究科学家、农业专业的研究生和本科生、水产养殖和渔业专业的学生，将有最好的帮助。本书主要关注水产养殖和渔业，也适合动物科学、家禽科学、农学、园艺学、昆虫学和植物病理学的学生。我完全同意本书的一个作者，来自弗吉尼亚理工大学的 Eric Hallerman 博士写作本书时的感受，正如他在给我的电子邮件中写到的，“这章节最终成了非常必要的一个章节，我从写作中得到了比预期还要多的收获，我最终学到很多，这也是我同意参加的部分原因（教授学生是参加本书的主要动机）”。是的，更加有效地教授学生同样也是我在准备此书漫长过程中的主要的动机和热情来源。

本书分为 5 篇。第一篇介绍基因组标记，包括各种分子标记技术的概念、原理和应用。第二篇介绍基因组作图，包括遗传连锁图谱、QTL 图谱、物理图谱、辐射杂交图谱和比较基因组图谱等各种基因组作图技术，也包含了利用分子标记选育的原理和应用。第三篇的主题是基因组表达和功能的分析，包括 EST 分析、微阵列分析、环境基因组学和功能基因组学。第四篇应该是水产基因组测序，但由于目前没有任何水产基因组被测序，现在的题目

是基因组测序前的准备，这一部分讨论已有的测序技术以及今后可能出现的新的测序技术，关于水产基因组测序的策略也在这部分进行了讨论。第五篇是水产养殖物种基因组学面临的问题，举例说明了对鱼类基因组加倍、平行基因功能研究的复杂性、高繁殖能力、牡蛎基因分离异常、牡蛎等双壳类极高多态性等水产动物基因组独特的生物学和特点。本书为研究人员今后碰到此类问题提供了解决问题的捷径。

我感谢参与本书撰写的所有作者，他们都是水产基因组学领域的真正专家，正是因为他们愿意与大家一起分享他们的知识和经验，才使本书在客观上成为可能。我荣幸地邀请到了国际上在畜牧基因组学领域最杰出的科学家之一、美国科学院院士、来自美国德州农工大学的 James Womack 博士给本书撰写序言。我很高兴我的学生 Eric Peatman、Peng Xu、Shao-lin Wang 和 Jason Abernathy，以及我的同事 Huseyin Kucuktas 博士帮助审阅部分章节。很高兴与 Blackwell 出版社的编辑助理 Erica Judisch，Blackwell 出版社植物科学、农业和水产养殖学的执行主编 Justin Jeffryes 经历长达一年的愉快合作。最后，我必须感谢我生命中两位最为重要的女性，我的母亲王友珍和我的夫人高东亚，母亲激励我成功，夫人相信我一定能成功。

刘占江

Zhanjiang (John) Liu

目 录

第 1 章 基因组和基因组学的概念	1
-------------------	---

第一篇 基因组标记

第 2 章 限制性片段长度多态性 (RFLP)	8
第 3 章 随机扩增多态性 DNA (RAPD)	13
第 4 章 扩增片段长度多态性 (AFLP)	18
第 5 章 微卫星标记和标记效用的评价	26
第 6 章 单核苷酸多态性 (SNP)	33
第 7 章 等位酶和线粒体 DNA 标记	41
第 8 章 水产养殖中基于个体的基因型分析方法	48
第 9 章 应用 DNA 标记进行群体遗传学分析	61

第二篇 基因组作图

第 10 章 水产生物的连锁作图	76
第 11 章 水产生物经济数量性状位点的检测与分析	92
第 12 章 水产生物的标记辅助选择	108
第 13 章 大片段细菌克隆文库的构建和应用	116
第 14 章 细菌人工染色体文库 (BAC) 为基础的水产基因组物理图的构建	134
第 15 章 利用 BAC 末端测序分析基因组物理特征	142
第 16 章 基因组概貌: 基因组重复结构鉴定	151
第 17 章 应用荧光原位杂交分析基因组	159
第 18 章 水产生物的放射性杂交作图	173
第 19 章 比较基因组学与定位克隆	178

第三篇 基因组表达与功能分析

第 20 章 基于表达序列标签的转录组分析	186
第 21 章 基因芯片技术的基础知识: 基本原理及在水产养殖业中的应用	195
第 22 章 鲑鱼 DNA 芯片技术及其他用于功能基因组学研究的方法	203
第 23 章 水域环境基因组学相关的大型数据集分析的計算性挑战	226
第 24 章 功能基因组学	233

第四篇 全基因组测序的前期准备

第 25 章 DNA 测序技术	254
第 26 章 全基因组测序	261

第 27 章 生物信息学	270
--------------------	-----

第五篇 水产养殖动物基因组测序的特殊性

第 28 章 硬骨鱼类的重复基因组	282
第 29 章 双壳类基因组学：所面临的复杂性、挑战及展望	291

参考文献

第 1 章

基因组和基因组学的概念

Zhanjiang Liu 著

当你想了解基因组学的基本概念时，你可以发现有许多关于基因组学的定义，如：“关于基因及其功能的研究”、“基因组的研究”、“一个物种所有基因的分子特征”、“一个细胞或一种生物体所有遗传信息的综合研究”、“同时关于大量基因的结构和功能的研究”等。为了对基因组学有一个较好的理解，让我们先来了解基因组的概念，以及其与基因组表达和功能的关系。

一、基因组和基因组学的概念

“基因组”这个词指一个生物体所有的遗传物质，严格地说，生物体的遗传物质包括动植物的核基因组和线粒体基因组，以及植物的叶绿体基因组。因为线粒体基因组和叶绿体基因组比较小而且其基因的数目有限，所以基因组研究主要关注核基因组。本章将主要介绍核基因组。

让我们根据遗传中心法则（图 1.1）为基因组学下一个狭义的定义。在多数情况下，双脱氧核苷酸（DNA）转录为核糖核酸（RNA），再翻译为蛋白质。尽管遗传信息是储存在 DNA 里，但它都将转录成中间分子 RNA，最后绝大多数都会被翻译成有生物功能的蛋白质。因此，一个生物体的全部 DNA 被称为基因组，全部 RNA 被称为转录组，全部蛋白质被称为蛋白质组。研究基因组的学科称为基因组学，研究转录组的学科为转录组学，研究蛋白质组的学科为蛋白质组学。尽管有这些分支，基因组学经常是涵盖狭义意义上的基因组学和转录组学，有时也包括蛋白质组学。



图 1.1 基因组和基因组学的概念与遗传中心法则的关系。生物体全部 DNA（基因组）转录为 RNA（生物体全部的 RNA 被称为转录组），RNA 再翻译成蛋白质（蛋白质组）。

基因组学、转录组学和蛋白质组学是分别研究基因组、转录组和蛋白质组的学科

基因组学包括结构基因组学和功能基因组学。结构基因组学主要研究基因组的结构、组织和进化，功能基因组学研究基因组的表达和功能。因为基因组功能反映了转录和编码蛋白，所以基因组学也包括对转录组和蛋白质组的研究。

需要指出的是，绝大多数类型的细胞中的基因组是比较稳定的，但一些免疫相关的细胞由于存在基因重排现象，基因组不是很稳定，转录组变动也很大。转录类型和相对表达水平受组织特异性、发育阶段、生理状况和环境的影响。例如，一种生物有 25000 个基因，但不是所有基因在每一种细胞都表达，那些编码细胞基本结构和功能的基因可能会在所有细胞、组织和器

官中表达,而每一种类型的细胞会表达该类型细胞的特异性基因。许多基因会在整个发育阶段表达,而有些基因只在特定的发育阶段表达。生理状态也会明显地影响基因表达,譬如,促性腺激素基因仅在垂体和性腺表达,在鱼类繁殖周期的产卵季节,表达水平会大大增强。多种环境因素如温度、pH值、水质、各种刺激压力、溶解氧等都能诱导或抑制许多基因的表达。

除了以上因素能造成转录组波动之外,同一组基因由于可变转录的缘故,也会造成转录组的差异,目前普遍认为转录组的复杂性要高于基因组。绝大多数的可变转录是可变剪接造成的,可变剪接指一个基因转录为异源的核 mRNA (hnRNA),通过不同的剪接,产生一种以上的 mRNA,造成一个转录本的内含子在另一个转录本可能成为外显子的现象。产生可变转录组的第二个机制是不同启动子造成的。一个基因,有不只一个的启动子在起作用,导致几个不同但相近的转录本。此外,不同的 PolyA 加尾信号位点也会造成不同的转录本。因此,储存在基因组的信息,在转录组水平被进一步放大和多样化了。尽管一个转录本只能编码一个蛋白质,但蛋白质可通过糖基化、乙酰化、磷酸化及其他翻译后的修饰来产生不同的活性,从而产生更多的生物学功能。由相关的基因编码的蛋白,其生物功能可能截然相反,例如,IL-1 II 受体是 IL-1 的假靶点,其通过结合 IL-1 来阻碍 IL-1 的功能。因此,遗传中心法则在遗传信息的基本流向、转录和翻译的主要功能方面无疑是正确的,而由同一组遗传信息的放大和多样性造成的复杂性,导致具有不同生物学功能分子的产生。各种基因功能的综合,自然就会表现出产生无数不同的生物学现象。

作为学科的一个分支,基因组学有其特定的研究范围,有自身独特的研究方法和技术。与传统的分子遗传学不同,基因组学不是仅关注一个或几个基因,而是同时把所有基因当作一个全天候的动态系统,来确定它们之间是如何相互作用的,以及如何影响生物学途径、网络、生理和整个生物个体。本书所关注的基因组技术,已经发展到能同时研究数万个基因的水平,和其他宏观技术一样,也需要借助地标(分子标记)给巨大的基因组定位。遗传连锁图和物理图可以帮助理解基因组的结构和组织,以及与基因组表达和功能相关的基因组环境和进化的情况。目前已开发了一些基因组学独特的方法来处理大量的基因,用于基因发掘、克隆和特征分析,或基因表达分析。通过对来自高度均匀化 DNA 或 cDNA 文库的大规模的序列表达标签(expressed sequence tags, EST)的分析,可以快速发掘基因和克隆几万个基因。这些技术需要一些其他基因组学技术的支撑,如强大的自动测序技术,使在工业流水线规模上挖掘和鉴定基因成为可能。基因表达定位到整个基因组的水平,有时就称为基因组表达,在系统生物学的概念上,指对复杂基因功能的相对复杂的基因表达调控。通过对几万个同时表达的基因的连续性监测,可以检测它们之间的相互作用和网络关系。信号传导不再是“无法察觉”的分子事件,而是可以通过在特定发育阶段、生理状态或环境条件下的各基因表达情况来观察到。研究基因及其功能,不再像传统分子生物学技术那样同时只研究一个或几个基因,而是要求在整个基因组环境、网络和相互作用的范围内开展。这些研究需要非常强有力的基因表达分析技术(如微阵列技术)的支持,应用这些先进技术,会产生大大超出生物学家所能理解的海量数据,需要整合整个生物学科以及其他学科的科技人员一起来分析处理这些数据。计算机和数学是最早介入基因组研究的两个学科,当处理这些来自基因组、基因组表达和基因组功能的大量数据时,会遇到许多困惑,比如观察到的现象是真实的,还是仅仅是系统生物学上正常波动的结果,因此也需要统计学家加入进来。不同国家的研究团队使用的语言不同(如英语、法语、汉语等),理解各团队科学家所用的语言并在不同学科的科学家中发挥作用,成为从事生物信息学研究的被称为“生物信息学家”的一大群科学家的工作目标。很显然,没有生物信息学,基因组学不能称为一门科学。因而,基因组

学的定义变得越来越复杂,读者也可以根据自己的理解来给基因组学下个定义。

基因组学的蓬勃发展使一批“组学”学科涌现出来 (http://genomicglossaries.com/content/genomics_glossary.asp)。基因组学的分支学科也已大量出现,包括:农业基因组学、应用基因组学、行为基因组学、生化基因组学、化学基因组学、临床基因组学、组合基因组学、计算基因组学、推断基因组学、生态毒理基因组学、环境基因组学、进化基因组学、正向基因组学、功能基因组学、免疫基因组学、工业基因组学、整合基因组学、反向基因组学、横向基因组学、纳米基因组学、网络基因组学、癌基因组学、药物基因组学、系统发育基因组学、生理基因组学、种群基因组学、预测基因组学、逆向基因组学、结构基因组学、毒物基因组学、翻译基因组学、比较基因组学等。

二、细胞、细胞核、染色体、基因组和基因组 DNA

基因组可以以各种方式存在,一个基因组可以是 RNA 或 DNA,以单链或双链的方式存在。例如,人类免疫缺陷病毒 (human immunodeficiency virus, HIV) 是一种逆转录病毒,其基因组是单链 RNA,这类异常的基因组主要存在于病毒和噬菌体。像细菌之类的原核生物,没有细胞核,基因组是由环状或线性的双链 DNA 组成。例如,大肠杆菌的基因组是由单个环状 DNA 链组成,而博氏疏螺旋体是由近 100 万个碱基组成的线性染色体组成的。真核生物基因组以染色体的形式存在,包含两个或更多的线性双链 DNA。

真核细胞染色体位于细胞核中,相同物种有固定数目的成对染色体 ($2n$),它们有固定的形状、大小和中心着丝粒位置,这些染色体的形态,就是通常所称的“核型”。二倍体生物所有体细胞都有相同数目的成对染色体,而在减数分裂后,精子和卵子只随机拥有其中的一组染色体,在卵子 (n) 和精子 (n) 结合后,胚胎又恢复到拥有两组染色体的二倍体。

染色体在细胞核中呈线状结构,含有基因和其他 DNA 序列,不同的生物拥有不同数目的染色体,人有 46 条染色体,其中 44 条为常染色体,2 条为性染色体。父母各贡献一半的染色体,因此,孩子的染色体一半来自母亲,另一半来自父亲。有性繁殖非常重要,配子(例如精子和卵子)是单倍体,受精后胚胎又恢复到二倍体。每一物种的染色体数目通常是稳定的,而种之间可能染色体数目差异很大,如果蝇只有 4 条染色体,而一种蕨类植物,心叶瓶尔小草 (*Ophioglossum reticulatum*),却有 1260 条 (630 对) 染色体,是目前已知染色体数目最多的物种。染色体数目最少的生物是一种蚂蚁 (*Myrmecia pilosula*),雌性只有 1 对染色体,而雄性只有 1 条染色体,这种生物的繁殖是通过一种单双套倍性系统来完成的,即受精卵 (二倍体) 发育成雌性个体,而未受精卵 (单倍体) 发育为雄性个体。

每个染色体是基因组的一部分,所有染色体组成整个基因组,尽管所有的染色体均有稳定的结构,但也可看作基因组的一部分。因此,基因组的总长度是所有染色体长度的总和。每个细胞在正常情况下的 DNA 片段数等于染色体数目。需要强调的是,利用常规的分子生物学手段得到染色体上的全部 DNA 是基本不可能的,在基因组 DNA 抽提过程中,染色体 DNA 很容易断裂,即使技术最熟练的人员也很难避免。多数情况下,一次 DNA 抽提需要几百万个细胞。用分子生物学分析基因组 DNA 时需要多个拷贝的基因组,因为每个细胞基因组 DNA 是随机断裂的,所以这些多拷贝的基因组会有多份相互重叠的 DNA 片段。

三、基因组大小

各种生物的基因组大小差异较大,最大差别可达 100000 倍,细菌基因组一般有几百万个碱基对 (Mbp),而目前已知一种非洲肺鱼 (*Protopterus aethiopicus*) 基因组最大,为