



# Functional Plant Genomics

# 植物功能基因组学



王元英 时 焦 主译

J.-F. Morot-Gaudry

P. Lea

J.-F. Briat

éditions  
**Quæ**

中国农业科学技术出版社

Functional Plant  
Genomics

植物功能基因组学

王元英 时 焦 主译

J.-F. Morot-Gaudry

P. Lea

J.-F. Briat

éditions  
Quæ

中国农业科学技术出版社

## 图书在版编目 (CIP) 数据

植物功能基因组学/ (法) 高德里等著; 王元英, 时焦主译. —北京: 中国农业科学技术出版社, 2009. 11

ISBN 978 - 7 - 80233 - 860 - 9

I. 植… II. ①高…②王…③时… III. 植物 - 基因组 - 研究 IV. Q943

中国版本图书馆 CIP 数据核字 (2009) 第 199423 号

Published by arrangement with INRA, Paris.

English edition: Functional Plant Genomics

ISBN 978 - 1 - 57808 - 506 - 4

©2007, Copyright reserved

French edition: LA GÉNOMIQUE EN BIOLOGIE VÉGÉTALE

©INRA, Paris, 2004

ISBN 2 - 7380 - 1167 - 5

责任编辑 沈银书 贺可香

责任校对 贾晓红

出版者 中国农业科学技术出版社

北京市中关村南大街 12 号 邮编: 100081

电 话 (010)82109709(编辑室)(010)82109704(发行部)

(010)82109703(读者服务部)

传 真 (010)82109709

网 址 <http://www.castp.cn>

经销者 新华书店北京发行所

印刷者 北京科信印刷厂

开 本 787 mm × 1 092 mm 1/16

印 张 36.75

字 数 900 千字

版 次 2009 年 12 月第 1 版 2009 年 12 月第 1 次印刷

定 价 98.00 元

# 《植物功能基因组学》翻译人员

主 译 王元英 时 焦

副主译 孙玉合 刘贯山 王树声 徐秀红 徐宜民 刘艳华

译校人员 (以姓氏笔画为序)

王卫锋 王元英 王凤龙 王绍美 王树声

太帅帅 冯全福 任 民 刘好宝 刘贯山

刘艳华 许立峰 孙玉合 杜咏梅 杨爱国

李凤霞 时 焦 迟立鹏 张兴伟 张怀宝

张忠锋 陈德鑫 罗成刚 胡海洲 侯婉莹

贾兴华 徐秀红 徐宜民 高晓明 龚达平

梁晓芳 戴培刚

# 中文版前言

1990年人类首次发起国际拟南芥基因组计划，2000年12月公布了拟南芥全基因组序列，这是世界上完成的第一个高等植物基因组测序，它为确定基因对植物生长发育及环境应答的决定性影响构筑了新的研究平台，极大地加快了人类认识自然和改造自然的步伐。随着水稻等其他高等植物基因（组）测序的完成，植物科学的研究已经步入了后基因组时代。

植物后基因组时代研究的核心内容是功能基因组学，即利用基因组序列的信息和高通量的系统分析技术，在基因组水平上研究植物组织结构与功能在细胞、有机体和进化上的关系，它强调发展和应用整体的实验方法分析基因组序列信息、阐明基因功能，其特点是采用高通量的方法结合大规模的数据统计计算，从而能够快速获得有关基因功能的知识。

植物功能基因组学是一个新的研究领域，其研究方法正在不断完善。在植物功能基因组学研究 and 应用不断取得重要进展的背景下，法国农业科学院（INRA）的 J.-F. Morot-Gaudry、P. Lea 和 J.-F. Briat 教授组织有关的科学家用法语撰写了题为《LA GÉNOMIQUE EN BIOLOGIE VÉGÉTALE》的专著，于2004年首次由 INRA-Editions 出版；2005～2006年期间对该书进行了更新与补充，其英文版《Functional Plant Genomics》于2007年由美国 Science Publishers 出版。该书得到了许多学者的高度评价，其中美国佐治亚大学植物基因组图谱实验室主任，著名的科学家 Andrew H. Paterson 教授于2008年1月3日在 Nature 杂志 451 卷 19 页发表了对该书的评价。他在文章中指出，该书远远超出了它的题目；通过基因组测序、注释和功能解剖等手段，从基因组学的创始分析发展到功能信息的翻译，从模式植物研究发展到作物改良，从克隆与测序到将序列转换为信息的计算机技术；既直观地解释了植物学的基本方法，又详细地描述了有关知识与资源的适用范围；同时还对植物多样性的众多特殊现象给予了极大关注，总之，他认为该书拓宽了从事生命科学工作读者的视野。

全书共三十二章，涵盖了小麦、玉米、水稻、甘蔗、葡萄、苜蓿、番茄、拟南芥、蒺藜苜蓿、高粱、杨树等多种植物，内容极为丰富，包

括植物核基因组序列测定、核基因组的物理结构、突变体库、基因组学与植物育种、生物信息学导论、蛋白质组学,根据基因组序列分析蛋白质功能、大规模蛋白质序列比对、转座因子和植物多样性分析、代谢网络结构、质子核磁共振代谢指纹与代谢谱分析、代谢通量的定性和定量分析方法、植物 DNA 微阵列等研究领域。

本书的中文翻译出版得到 INRA-Editions 与原作者的授权,也得到中国农业科学技术出版社的大力支持。全书翻译与校对由中国农业科学院烟草研究所(中国烟草总公司青州烟草研究所)的科研人员完成,译校人员充满了对生物学新领域的极大兴趣和热情,期望为我国植物功能基因组学研究做一点微薄工作,然而,由于植物功能基因组学是一门全新的研究领域,我们对新的实验方法和技术的掌握尚存在不足,加上时间仓促,译文中难免存在不妥和错误之处,恳请广大读者包涵与批评指正。

王元英

中国农业科学院烟草研究所

中国烟草总公司青州烟草研究所

2009 年 11 月 6 日于青岛

# 序言 1

十字花科模式植物拟南芥 (*Arabidopsis thaliana*) 核基因组序列分析结果的发表, 是植物科学在新千年里通往创新性发展的里程碑。至少在理论上首次使植物基因组中操控基因的完整序列编目及其功能的系统分析成为可能。染色体序列中假定基因的识别, 可以根据注释方法利用转录研究加以验证, 通过鉴定每个注释基因的一组特异序列, 按照研究拟南芥基因转录设计的 AFFYMETRIX 和 CATMA 基因芯片所描述的, 利用对应序列合成探针, 并建立微阵列, 从而在基因组规模上进行基因鉴定及其功能分析。然而, 根据预测估算进行基因注释不是百分之百的可靠, 在外显子与内含子边缘经常出现预测误差, 因而需要采用补充方法进行注释验证。利用系统测序获得全长 cDNA 序列并对其分析已经成为一种有效的验证方法, 不同的研究机构, 如日本的 RIKEN 研究所, 美国的 Salk 研究所和法国国家基因测序中心 (Genoscope) 等, 在这方面都做出了贡献, 他们不仅改进注释, 而且进行功能研究资源创新, 这方面研究包括特殊转录物的过量表达, 利用标记酶 (例如 GFP, 荧光蛋白) 融合技术进行相应蛋白的细胞内定位, 或在异源系统中进行蛋白表达来研究蛋白的生物化学特性和 3D 结构。由于全长 cDNA 序列集总是不完整的, 美国 Salk 研究所的 J. Ecker 研究小组利用覆盖整个基因组的寡核苷酸阵列 (oligonucleotide-based array) 设计了一种替代方法进行转录研究。

了解基因组序列可为基因组中基因的功能分析铺平道路。根据随机插入已知序列 (T-DNA, 转座子) 能够抑制基因表达的原理, 在已经测序的基因组序列中, 进行插入位点的识别和序列编目, 这种方法已经用于拟南芥基因组 T-DNA 插入位点的序列编目, 以及在数据库中保存相应的信息, 几项程序在这方面做出了贡献, 其中包括 SIGNAL、Genoplante、GABI 以及其他一些程序。此类方法也不能使拟南芥基因组所有注释基因达到插入饱和。一种补充方法是利用基因的 TILLING, 亦即通过筛选大量  $M_2$  植株进行特殊基因的识别, 另外采用 RNA 干扰的方法在化学上可以足够有效地对拟南芥基因组中的基因进行系统钝化, 该方法在 AGRIKOLA -

EC 项目中进行了描述。只要一系列有待功能分析的基因得到了鉴定，这两种方法还可应用于任何植物基因组。然而，基因组学也能从其他系统方法中得到益处。可以通过利用已经注释基因组序列中的预测蛋白质序列，使纯化特殊细胞器和细胞重要组成成分（叶绿体、液胞、质膜），以及利用质谱技术鉴定纯化组分中不同蛋白质的过程变得简化，因此，可以批量鉴定细胞内蛋白质的位置。可以利用酵母双杂交方法（Y2H）和近期使用的 TAP-TAG 方法进行的“互作组”特性描述完成基因组序列编目。最近，还利用比较基因组方法鉴定基因启动子的调节动因，以及系统鉴定非编码 RNA 转录物，其中 miRNA 目前被认为具有关键的调节功能，是精细研究的主题。因此，基因组学集合了一系列基因功能分析的分子水平的方法，基因组学过去很可能处于分子遗传学的初期，但目前更多资源的可用性使其得到了更新。新的数据挖掘使未知基因的假设功能得到研究，有时证实这种假设的工作量并不大，这是由于大量的特殊基因的信息（突变体表型、基因表达谱、细胞内定位以及与其他基因的相似）已经散布在各种不同的数据库中。虽然植物的基因组学耗资很大，但在某种意义上它节约资金。测序已经商业化，许多研究基因功能所需要的方法可以合并高效率地用于大量的分析。这对于 T-DNA 插入位点的序列编目已经成为事实，并且所花费的时间也很短。

拟南芥等模式植物的基因组学对作物基因组研究产生了重要影响，与赤霉素信号传导相关的 Gai 基因的鉴定完美地阐述了进化过程中基因功能的保守性，已证明该基因是小麦矮秆基因（Rht）的功能直系同源基因（functional ortholog），其保守性与染色体上的基因顺序相关。很有必要选择少数物种作为模式植物进行共线性（synteny）开发，如对拟南芥和水稻进行精细研究，并将其结果应用于其他植物基因组研究中。关于基因注释，共线性预测不是百分之百可靠，但至少可以建立一种参考框架，用于研究某种作物基因组的某个感兴趣的未知位点，还可以快速构建高密度的遗传图谱，进行有价值农艺性状基因的定位克隆。基因组学是作物生物多样性研究的重要内容，也是控制农艺性状位点等位基因多样性序列的重要内容，其最终目标是实现分子育种。

Michel Caboche

王元英译



## 序言 2

拟南芥全基因组序列分析完成之后，植物种之间基因含量、基因功能、基因结构的许多保守性的假设，将使拟南芥基因组知识作为植物基因组测序的通用线路图被充分利用。的确，已经证明拟南芥可以作为植物许多性状和代谢途径基因功能极端相关的一种通用模型，但对基因含量和顺序的参考价值较小，尤其对系统发育远缘的物种。

幸运的是人们以拟南芥为蓝本，启动了多项研究计划，对水稻、番茄等具有农艺价值的物种进行基因组测序研究，并逐步扩展测序的物种数量。由于基因组的体积和结构等问题，起初人们认为玉米、小麦、甘蔗、葡萄等物种难以完成基因组测序，目前也已经相继开展了测序研究。从而通过表达序列标签（EST）测序发现基因，并且破译植物家族之间基因的复杂网络。当今，随着几种植物全基因组测序的完成，另一些植物的高密度物理图谱的构建，以及一种植物或一组相关物种的不同基因型的序列草图的绘出，人类获得了大量的数据。应该牢记“数据不是知识”，生物信息学作为稳定这种信息的关键学科而出现，系统发育学和进化方法成为了解基因和基因组之间关系的重要途径。

正如 DNA 测序的快速发展一样，功能基因组学也发生了类似的变化。反向遗传学正在通过 RNA 干扰方法、定向诱导基因组局部突变技术（Targeting Induced Local Lesions In Genomes, TILLING）和 T-DNA 系统创造群体等方法研究候选基因的功能。自然发生的遗传变异分析和人们区别遗传漂移变异和适应性变异的兴趣正在发展成为一个关键课题。而且，就标记或个体分析而言，高通量基因分型（high-throughput genotyping）技术将改变我们制作精细基因图谱的能力，对基因组结构的了解和决定感兴趣农艺性状位点的图谱将产生关键的影响。对于基础应用学家来说，利用基因分型和基因型评价得到其“组学”状态。

数量性状遗传图谱的关键环节正在由基因型分析转向表现型分析，了解环境怎样影响数量性状表达是基础与应用研究的关键问题，揭示这些秘密及开发综合模型，可以促进对至今尚未开发的遗传多样性资源特性的了解，加快其在作物改良中的应用。

基因组学的研究信息正在为植物生物学提供新的丰富多彩的机会，从进化和分子生理学基础研究到聚焦于作物改良的应用研究，该领域的许多论题涵盖在本书的各个章节之中。

A. Charcosset

法国农业科学院 (INRA) 植物遗传育种系

王元英译

# 前 言

由于概念和方法学的变化，在过去 50 年中生命科学研究已经取得显著的进步。最近，人类成功地完成了多个真核多细胞有机体基因组完整序列的测序，植物生物学研究也加入到这一宏伟的事业中。然而，全基因组测序应该被看作是起点，而不是终点。由功能基因组学提供的新视野使有机体生物学 (organism biology) 与分子生物学成为一体，为了定义综合生物学，应该吸收新的观点，即一个表现型如何在环境和基因型相互作用的基础上建立。

那是在法国农业科学院 (INRA) 植物生物系工作之前的几年我们就构思写本书，因为我们所处的特殊地位，能够观察到植物生物学科的主要发展变化，以致于我们决定将收集到的资料汇编成本书，书中多个题目与基因组学有关。这些题目对植物生物学及其相关学科产生了积极的影响，改变了相关的研究方法。感谢来自 INRA、CNRS、CIRAD、IRD 和大学的同事，正是因为他们的专业知识才使该书涵盖了一个广泛的基因组学的概念和方法，内容包括从国际基因组测序计划到最新的生物信息学方法、基因表达的分析方法，以及最终表达产物、特异组织和细胞分化。这些新知识能够综合理解基础生理功能和发育过程。基因组学与遗传和植物育种之间的紧密关系揭示了利用分子标记、分析遗传变异和研究数量性状的共线性 (synergy)。这些方法的综合利用是切实可行的，这种综合也可以影响生态生理学，并为植物群体研究模式的应用开拓了光明的前景。

因此，我们正在期盼新的植物生物学的诞生，同时我们也希望本书将对新的实践的传播做出贡献。

本书适合于高等院校生物学和农学类专业的学生、教师以及期望用相对精细和系统的方式快速了解植物基因组学的科学工作者。

Jean-François Briat

Jean-François Morot-Gaudry

A. Charcosset

王元英译

· VII ·

# 目 录

中文版前言 .....	王元英 ( I )
序言 1 .....	王元英译 ( III )
序言 2 .....	王元英译 ( V )
前言 .....	王元英译 ( VII )

## 结构基因组学和生物信息学

第一章 植物核基因组的物理结构 .....	李凤霞译, 刘贯山校 ( 1 )
一、经典细胞遗传学 .....	( 1 )
二、分离基因组 DNA 的物理方法 .....	( 3 )
三、采用变性复性动力学分析 DNA 分子 .....	( 4 )
四、基于克隆和杂交的方法分析植物 DNA 的物理结构 .....	( 6 )
五、基因组的物理作图 .....	( 9 )
六、染色质的结构 .....	( 11 )
七、染色质的异质性与动力学 .....	( 13 )
第二章 植物核基因组的测序 .....	太帅帅译, 时 焦校 ( 17 )
一、引言 .....	( 17 )
二、真核细胞核基因组的测序策略 .....	( 17 )
三、拟南芥基因组 .....	( 24 )
四、水稻基因组 .....	( 27 )
五、玉米基因组 .....	( 31 )
六、蒺藜苜蓿基因组 .....	( 32 )
七、百脉根基因组 .....	( 33 )
八、白杨基因组 .....	( 34 )
九、番茄基因组 .....	( 34 )
十、葡萄基因组 .....	( 34 )

十一、小麦基因组 .....	(35)
十二、其他正在测序的植物基因组 .....	(36)
十三、共线性 .....	(36)
十四、生物学及其技术的未来发展方向 .....	(37)
<b>第三章 生物信息学导论</b> .....	王元英译, 孙玉合校 (41)
<b>第四章 生物学数据库</b> .....	孙玉合译, 侯婉莹校 (43)
一、数据库的类型 .....	(43)
二、生物学的主要公共数据库 .....	(45)
三、专题数据库 .....	(48)
四、单一生物体或分类群的数据库 .....	(49)
五、一些植物的专业数据库 .....	(50)
六、生物学数据库的未来 .....	(50)
七、专业词汇 .....	(51)
<b>第五章 基因预测</b> .....	刘贯山译, 孙玉合校 (55)
一、从头计算的 DNA 序列模拟和预测 .....	(55)
二、外在方法 .....	(61)
三、基因预测的成功率与目前的局限性 .....	(62)
<b>第六章 基于基因组序列的蛋白质功能分析</b> .....	刘贯山, 陈德鑫译, 时 焦校 (69)
一、蛋白质“功能”概念的准确性 .....	(70)
二、序列内在信息的分析方法 .....	(71)
三、同源性搜索方法 .....	(74)
四、基因临近序列的分析方法 .....	(79)
五、注释过程中常见错误的来源 .....	(81)
六、结论 .....	(83)
<b>第七章 大规模蛋白质序列比对</b> .....	戴培刚译, 刘贯山校 (87)
一、蛋白质或核酸 .....	(88)
二、序列比对 .....	(88)
三、相似性标准 .....	(90)
四、蛋白质全序列比对及其算法 .....	(93)
五、集群构建 .....	(94)
六、共线性搜寻 .....	(96)
七、结论 .....	(96)

## 功能基因组学——从序列到功能

<b>第八章 植物转基因</b>	时 焦译, 徐宜民校	(99)
一、引言		(99)
二、外源基因在植物细胞中表达的检测		(100)
三、转基因的主要技术		(103)
四、DNA 在受体植物基因组中的整合		(106)
五、转基因——科学研究的手段		(107)
六、转基因——植物育种的工具		(108)
七、结论与展望		(110)
<b>第九章 突变体库——高通量的拟南芥反向遗传学系统</b>	刘贯山译, 时 焦校	(117)
一、插入突变		(118)
二、T-DNA		(120)
三、转座子		(122)
四、基因诱捕和激活诱变		(124)
五、插入群体在反向遗传学中的应用		(126)
六、PCR 筛选		(127)
七、插入的计算机搜索		(128)
<b>第十章 植物 DNA 微阵列</b>	王卫锋译, 刘贯山校	(135)
一、引言		(135)
二、DNA 阵列流程		(137)
三、微阵列实验设计		(142)
四、植物 DNA 微阵列实验		(147)
五、DNA 微阵列的局限性及发展方向		(150)
六、结论		(151)
<b>第十一章 独立成分分析测定基因表达的统计设计与分析</b>	王绍美, 许立峰译, 时 焦校	(158)
一、数据分析		(158)
二、实验设计		(175)
三、微阵列分析		(177)
四、软件和数据		(178)

---

<b>第十二章 蛋白质组学</b> .....	迟立鹏译, 刘艳华校 (182)
一、蛋白质组学的工具 .....	(182)
二、蛋白质组学在植物中的应用 .....	(185)
三、新的方法学和技术学手段 .....	(192)
四、结论 .....	(193)
五、附录——质谱技术在蛋白质鉴定中的应用 .....	(194)
六、补充内容 .....	(202)
<b>第十三章 植物细胞基因表达跟踪——显微及生物成像技术</b> .....	龚达平译, 王凤龙校 (206)
一、引言 .....	(206)
二、显微镜 .....	(206)
三、显微镜配件 .....	(214)
四、成像的光探测 .....	(215)
五、图像后处理 .....	(220)
六、荧光染料和探针的优化 .....	(222)
七、光在生物成像中的应用 .....	(225)
八、结论与展望 .....	(226)
<b>第十四章 植物细胞基因表达跟踪——功能基因组学的新探针</b> .....	王元英译, 杨爱国校 (232)
一、引言 .....	(232)
二、细胞和组织中基因表达及其产物的检测 .....	(232)
三、基因和蛋白的报告子——荧光蛋白 .....	(240)
<b>第十五章 代谢组学</b> .....	罗成刚, 胡海洲译, 刘艳华校 (262)
一、引言 .....	(262)
二、术语 .....	(263)
三、分析技术 .....	(264)
四、代谢组学面临的挑战和困难 .....	(267)
五、统计和生物信息学工具 .....	(268)
六、功能基因组学的重要工具 .....	(272)
七、对代谢的认识 .....	(273)
八、实体生物等效性的唯一决定性证据 .....	(275)
九、总结和展望 .....	(276)

<b>第十六章 质子核磁共振代谢指纹与代谢谱分析</b> .....	杨爱国译, 刘艳华校 (282)
一、核磁共振波谱技术 .....	(282)
二、样品制备 .....	(283)
三、质子核磁共振代谢指纹分析 .....	(284)
四、质子核磁共振定量代谢谱分析 .....	(286)
<b>第十七章 代谢通量的定性和定量分析方法</b> .....	贾兴华译, 冯全福校 (291)
一、数据分析 .....	(291)
二、终产物积累净通量与单向通量 .....	(291)
三、同位素标记 .....	(292)
四、原理 .....	(297)
五、优点、局限性及应用 .....	(299)
六、通过 <sup>13</sup> C 同位素异构体分析所获得的信息 .....	(300)
七、自旋标记 .....	(301)
八、结论 .....	(306)
<b>第十八章 代谢网络结构——对最简单代谢网络的认识</b> .....	刘艳华译, 徐秀红校 (310)
一、引言 .....	(310)
二、代谢网络 .....	(310)
三、代谢网络的数学表示——化学计量学矩阵 .....	(313)
四、初级通量模式 .....	(318)
五、代谢网络在初级模式分析中的应用 .....	(318)
六、初级通量模式分析存在的问题 .....	(319)
七、结论 .....	(320)

## 基因组学模式植物

<b>第十九章 模式植物——拟南芥</b> .....	王树声译, 时 焦校 (322)
一、引言 .....	(322)
二、植物学和生物学性状 .....	(322)
<b>第二十章 谷类植物基因组学的模式植物——水稻</b> .....	时 焦译, 徐宜民校 (333)
一、水稻基因组的一般特性 .....	(333)
二、水稻基因组测序 .....	(336)
三、水稻基因组的特殊特性 .....	(337)



四、水稻基因组与其他谷类植物基因组的比较 .....	(338)
五、水稻功能基因组学 .....	(339)
六、结语 .....	(342)
<b>第二十一章 豆科模式植物——蒺藜苜蓿</b> .....	梁晓芳译, 王凤龙校 (345)
一、豆科作物的重要性和特性 .....	(345)
二、模式植物研究对豆科植物生物学研究和育种的必要性 .....	(346)
三、蒺藜苜蓿作为豆科模式植物的主要特性 .....	(349)
四、国际基因组计划对豆科植物基因组学研究的促进 .....	(349)
五、蒺藜苜蓿功能基因组学 .....	(350)
六、蒺藜苜蓿的结构基因组学和基因组测序 .....	(356)
七、蒺藜苜蓿生物信息学和网络资源 .....	(359)
八、展望 .....	(361)
<b>第二十二章 茄科植物基因组学的模式植物——番茄</b> .....	刘好宝译, 李凤霞校 (372)
一、作为模式物种的几个特点 .....	(372)
二、遗传资源——近缘种、突变体和遗传材料 .....	(372)
三、基因和 QTL 图位克隆的遗传图谱 .....	(373)
四、优良基因克隆 .....	(375)
五、果实的主要生理特性 .....	(376)
六、QTL 定位 .....	(377)
七、染色体步移克隆的第一个 QTL .....	(377)
八、表达序列标签和微阵列的收集 .....	(377)
九、比较基因组和共线性——保守同源位点标记 .....	(378)
十、通过反向遗传学的基因功能分析 .....	(379)
十一、展望 .....	(382)
<b>第二十三章 基因组学在葡萄改良中的应用</b> .....	冯全福译, 徐秀红校 (388)
一、葡萄改良的进展与展望 .....	(388)
二、葡萄科与蔷薇分支的密切关系 .....	(389)
三、连锁不平衡的研究基础——种质资源 .....	(389)
四、遗传图谱技术在重要性状遗传研究中的应用 .....	(391)
五、物理图谱与遗传图谱的整合效应 .....	(393)
六、EST 测序的作用 .....	(394)
七、反向遗传学研究的高效遗传转化技术 .....	(395)
八、葡萄基因组测序前景 .....	(396)