



植物基因组学 与蛋白质组学

Plant Genomics and Proteomics

[英] C. A. 卡利斯 著

Christopher A. Cullis

冯晓燕 等译



化学工业出版社

现代生物技术与医药科技出版中心

植物基因组学 与蛋白质组学

Plant Genomics and Proteomics

[英] C. A. 卡利斯 著

Christopher A. Cullis

冯晓燕 等译



化学工业出版社

现代生物技术与医药科技出版中心

· 北 京 ·

图书在版编目 (CIP) 数据

植物基因组学与蛋白质组学/ [英] 卡利斯 (Cullis C. A.) 著; 冯晓燕等译. —北京: 化学工业出版社, 2006. 5

书名原文: Plant Genomics and Proteomics

ISBN 7-5025-8862-0

I. 植… II. ①卡…②冯… III. ①植物-基因组-研究②植物-蛋白质-研究 IV. ①Q943②Q946. 1

中国版本图书馆 CIP 数据核字 (2006) 第 055037 号

Plant Genomics and Proteomics/by Christopher A. Cullis

ISBN 0-471-37314-1

Copyright©2004 by John Wiley & Sons, Inc. All rights reserved.

Authorized translation from the English language edition published by John Wiley & Sons, Inc.

本书中文简体字版由 John Wiley & Sons, Inc. 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。

北京市版权局著作权合同登记号: 01-2005-0873

植物基因组学与蛋白质组学

[英] C. A. 卡利斯 著

冯晓燕 等译

责任编辑: 李 丽 郎红旗

责任校对: 于志岩 王素芹

封面设计: 关 飞

*

化学工业出版社 出版发行
现代生物技术与医药科技出版中心
(北京市朝阳区惠新里 3 号 邮政编码 100029)

购书咨询: (010)64982530

(010)64918013

购书传真: (010)64982630

<http://www.cip.com.cn>

*

新华书店北京发行所经销
大厂聚鑫印刷有限责任公司印刷
三河市延风装订厂装订

开本 720mm×1000mm 1/16 印张 13½ 字数 164 千字

2006 年 8 月第 1 版 2006 年 8 月北京第 1 次印刷

ISBN 7-5025-8862-0

定 价: 28.00 元

版权所有 违者必究

该书如有缺页、倒页、脱页者, 本社发行部负责退换

译者名单

主 译：冯晓燕

翻译人员：李若楠 顾海涛 李 斌 尉连伶

译者的话

随着人类基因组计划、植物基因组计划和微生物基因组计划的出现和完成，生命科学的研究已进入后基因组时代，必须从整个基因组及其全套蛋白质产物的结构、功能、机理以及相互作用的深度去了解生命活动的全貌。由于植物生长发育规律和所处环境的特殊性及其复杂性加大了科学研究的难度，使植物生命活动的研究总是落后于人类和其他动物生命活动的研究。

《植物基因组学与蛋白质组学》概述了近年来植物基因组研究中涉及的分子生物学技术及其应用情况，包括基因组测序、数据库的应用等，是一本综合了基因组研究前沿的专著。它既有教科书的系统性，又充分反映了植物分子生物学领域的最新进展。它系统介绍了植物基因组的结构、获得功能基因组数据、基因组测序的策略、基因表达的调控、功能基因组学及植物与外界环境的相互作用，还讲述了生物信息学的运用，并对生物技术和植物基因组学的未来作了展望，是一本综合了基因组研究前沿的著作。

我们在翻译过程中始终坚持忠实原文的指导思想，遵循全国名词委审定公布的名词及名词原命名原则处理相关名词，对于个别尚未审定统一的词汇，我们按照“约定俗成”的原则进行翻译。这本书翻译工作的顺利完成离不开北京大学生命科学院几位老师和同学的努力。李若楠、顾海涛、李斌和尉连伶等做了大量细致的工作，并且在稿件加工、文字润色等方面做出了很大努力和贡献。化学工业出版社的李丽编辑在翻译文字、审校和定稿工作过程中，为确保本书的翻译质量和出版工作的顺利进行付出了

大量精力。在此，我谨向所有参加本书译校、促成本书出版的工作人员表示感谢。

这本书的翻译专业性强，参加翻译的人员都是工作在第一线
的研究人员，但由于知识有限，时间仓促，在翻译过程中难免存
在错漏，在文字修饰方面也会有不当之处，谨请读者批评指正。

冯晓燕

2006年6月于北京大学生命科学院

中文版序

随着人类基因组测序的完成和全球基因组计划的迅猛发展，近 200 种生物体的基因测序工作接近或已经完成，其中包括一些高等植物，如拟南芥、水稻等。但是，生命的遗传奥秘并没有伴随生物基因组的碱基排列顺序的测定而揭开。基因的功能最终是由它们编码的蛋白质或非编码 RNA (non-coding RNA, ncRNA) 来表现的，即一个基因所含的遗传信息通过一系列复杂的反应，最终导致了相应的蛋白质及 ncRNA 形成，蛋白质及相关 RNA 再参与到生命的各种活动中去。所以，要想真正揭开生命的奥秘，仅仅了解基因组的碱基排列顺序是远远不够的，还必须认识基因的产物——蛋白质及 RNA。长期以来，研究人员在蛋白质的结构和功能研究方面进行了大量的工作。近年随着基因组计划的完成，在全基因组水平研究蛋白质表达、功能以及蛋白质-蛋白质的相互作用已愈来愈广泛，已经成为植物科学发展的必然趋势。

《植物基因组学与蛋白质组学》(plant genomics and proteomics) 一书的英文版于 2004 年由 John Wiley & Sons Inc. 出版，书中概述近年来植物基因组的研究进展，系统介绍了植物基因组的结构和研究策略，综述了蛋白质组学的研究方法及其在植物研究中的应用，是一本体现基因组研究前沿的专著。概括了在植物基因组和蛋白质组研究中使用的研究技术和进展。强调目前生命科学的研究必须从整个基因组及其全套蛋白质产物的结构、功能、机理以及相互作用的深度去了解生命活动的全貌，并系统地整合有关生命科学的全部知识，这样才能够完全揭开生命

之谜。

本书是由工作在北大-耶鲁植物分子遗传学及农业生物技术联合研究中心的青年科技工作者们进行翻译的。本中心近年来在植物蛋白质组和拟南芥转录调控因子蛋白质组学研究方面取得了重要成果，克隆了 1303 个转录调控因子基因，并制作相关的蛋白质芯片，积累了一定的实际工作经验。但是，由于时间和水平所限，难免有不尽如人意的地方，还请读者提出宝贵意见。衷心地希望《植物基因组学与蛋白质组学》能够成为广大学生和研究人员进行植物蛋白质组研究的有力助手！



博士

北大-耶鲁植物分子遗传学及农业技术
联合研究中心主任
耶鲁大学分子与细胞发育生物学系教授

致 谢

本书的出版与两个人的努力分不开。首先，我要感谢我的妻子 Margaret，她阅读了本书草稿，并努力润色文字使其表达清晰明了，她对本书所作的贡献是无价的。书中出现的模糊或错误之处并非她的责任，反而是由于她努力说服我却并没能使我改变主意造成的。其次，感谢我的儿子 Oliver，他和我一起尝试了写作工作，并为早期的草图作了简洁的注释。

目 录

绪论	1
第 1 章 植物基因组的结构	7
1.1 导言	7
1.2 DNA 的变化——数量	7
1.3 染色体变异	10
1.4 DNA 变异的起源	13
1.5 多基因组的结果	23
1.6 小结	26
1.7 参考文献	27
第 2 章 基本工具箱——获得功能基因组数据	29
2.1 导言	29
2.2 克隆系统	30
2.3 测序策略和自动化技术	43
2.4 蛋白质分析	44
2.5 数据库与信息学	46
2.6 小结	49
2.7 参考文献	49
第 3 章 测序的策略	51
3.1 导言	51
3.2 完全基因组序列的获得	53
3.3 物理和遗传学图谱	54
3.4 制作物理图谱	55

3.5	基因组分段	57
3.6	基因组分段的方法	58
3.7	基因组测序策略	63
3.8	基因组分级分离途径	66
3.9	测序和数据加工	67
3.10	最有效的方法	68
3.11	“完全”基因组序列的可能目标	68
3.12	参考文献	71
第4章	寻找基因	73
4.1	导言	73
4.2	从基因组序列鉴定基因	74
4.3	全长 cDNA 的产生	78
4.4	全长 cDNA 测序	80
4.5	基因组注释	81
4.6	共线性	83
4.7	利用诱变确定基因	83
4.8	小结	91
4.9	参考文献	91
第5章	基因表达调控	93
5.1	导言	93
5.2	基因表达调控需要特异启动子序列	95
5.3	增强子元件在基因表达上的作用	96
5.4	5'端序列在基因表达中的地位	97
5.5	3'端序列在基因表达中的地位	98
5.6	内含子在基因表达中的地位	98
5.7	真核生物启动子的保守序列	99
5.8	反式作用因子调控基因表达	100
5.9	mRNA 的稳定性	101

5.10	植物中化学调控基因的表达	102
5.11	染色质结构的影响	103
5.12	翻译调控	105
5.13	小结	105
5.14	参考文献	105
第 6 章	功能基因组学	107
6.1	导言	107
6.2	表达谱	109
6.3	蛋白质组	120
6.4	小结	127
6.5	参考文献	127
第 7 章	植物体与外界环境的相互作用	129
7.1	导言	129
7.2	生物因子的相互作用	130
7.3	非生物作用	139
7.4	参考文献	142
第 8 章	复杂性状的确定和操作	145
8.1	导言	145
8.2	分子图谱	147
8.3	分子标记系统	147
8.4	确定 QTL	150
8.5	分子标记辅助选择	158
8.6	参考文献	162
第 9 章	生物信息学	165
9.1	导言	165
9.2	数据库	166
9.3	信息处理工具	169

9.4	蛋白质鉴定和比较	178
9.5	比较基因组学	180
9.6	数据的意义	181
9.7	小结	184
9.8	参考文献	184
第 10 章	生物伦理学和植物基因组学的未来	187
10.1	导言	187
10.2	安全问题	189
10.3	贸易问题	190
10.4	公众的态度	190
10.5	公共策略问题	194
10.6	小结	196
10.7	参考文献	196
后记		198
索引		199

绪 论

如果写一本关于植物界基因组学的书，可能的理论基础是什么呢？显然，植物和动物在许多基本特征上有巨大差异。一般地，植物不能随意移动、可能有较长的生命周期、并且几乎都是自己制造营养，因而它们的生长仅需要矿物质、阳光、水和空气。植物基因组必须编码支持全部必需代谢过程的酶，包括光合作用、呼吸作用、中间代谢、矿物质吸收，以及脂肪酸、脂类、氨基酸、核酸及辅因子的合成。而动物则是通过摄食来获得这些物质的。在技术水平上研究基因组学，是从整体上评价基因组信息以及它是如何决定一个组织的形态和功能的，从而揭示一个适用于任何系统的普遍线索。但是植物所表现的特异偏好和呈现的特殊问题，使之不能单单从一种模式植物中反映出来，所以经常需要用特定的植物研究某一领域。植物基因组学是建立在几个世纪以来许多植物观察和实验研究基础上的。基于这种历史的原因，许多实验数据和观察记录跨越了不同的植物材料，而不是集中研究一种方便的单一模式植物发育的各个方面。光合作用的模型是藻类植物，它们为研究光合作用提供了有用的相关基因的提示。但反过来，藻类植物对于理解在植物的生长过程中干旱或盐分对根的胁迫是如何影响光合作用的就没有帮助。利用基因组学方法研究植物生物学，会加强对基因结构、功能和可变性的理解。这些新知识的应用必将引发增加作物产量的新技术的产生，满足我们日益增长的食品需求。

本书认为，植物与其他生物体相对应的不同之一在于植物界核 DNA 含量（基因组大小）变化范围很大，即使亲缘关系很近

的物种也如此。因此，由于额外 DNA（即使没有确实的 DNA 序列）可能会引起表型特征——例如，核 DNA 含量在一年生或多年生植物的生活周期中起作用——描述典型植物基因组的真实特征很难。更为复杂的是，在进化过程中由多倍体化循环引起的多倍体基因组结构重建经常发生。基因组重建往往引起在原始多倍体事件中获得的一些额外 DNA 的丢失。因此，研究多个植物基因组的详细特点，而不是集中研究一个或几个模式植物，对于加深人们理解功能和进化对植物基因组大小的约束方面更为重要。学术界普遍认为，尽管在每个细胞里的 DNA 含量有着巨大的变化，大多数植物以同样数目的基因和相似的遗传蓝图控制着生长和发育。

正如第 1 章所述，通过研究大多数的可调控系统，得到了许多关于生理过程的丰富的数据，如细胞壁的合成、光合作用及植物的抗病性，这可以帮助理解特定的生理现象。然而，许多植物模型只有相当少的基因组学信息资源，比如缺少足够的序列数据、丰富的突变体及相关的信息，在其他方面也没有非常深入的研究背景。因此，源自这些系统中的每一信息都不得不在一个熟知的模式植物中去验证，以期理解许多交叉的发育过程在分子水平的综合和协调能力。这些独立事件在这些经过透彻研究的系统中再进行研究几乎是不可能的。百日草属植物（*Zinnia*）提供了一个极好的研究导管元件分化的模型，因为该属植物中分离的叶肉细胞可以在体外被同步诱导形成这些元件。因此，这种同步化使分子与生化事件年表及其建立成为可能，这些事件与细胞特异分化和木质部分化相关基因鉴定有关。然而，百日草属植物（*Zinnia*）不具有用于实验的基础结构，研究人员不便于对它的其他重要生物过程进行广泛的基因组学研究。因此，已获得的详细信息需要整合入另一种更为详细描述的模式植物，虽然这种信息在没有特定的实验系统的条件下很难识别。因此，积累基因组信息对于植物界是必需的，随着信息的整合合成，也许最终研究

将会集中在很少的几个模式物种里。为了能够揭开植物是如何生长、发育、繁殖以及与环境相互作用，相应的研究方法包括在特定分子水平上对来自许多物种的无数植物途径进行研究的复杂方法。

拟南芥和水稻全基因组序列的公布，使植物和动物在基因序列水平的比较更加简易化。早期比较显示，大部分过程有很大分歧，但一些过程在跨界物种中似乎是保守的，如跨膜运输与DNA重组和修复过程。在已研究的植物基因组中发现了许多可能涉及植物中的多种功能而在动物和微生物中没有的新基因。

植物基因组序列及所有其他的基因组学工具唾手可得，如标签突变体的收集、基因芯片和蛋白质组学技术，这基本上改变了植物科学研究。现在研究人员可以以新的活力和热情来解决难于处理的老问题。例子之一是植物基因组计划（Floral Genome Project，参见 <http://128.118.180.140/fgp/home.html>）解决了达尔文的“恼人之谜”，即开花植物的起源问题。该问题已长达一个多世纪没有答案。植物基因组学不仅回答了达尔文提出的问题，而且花的起源和多样化还是植物界的基本问题。因为花的结构在植物繁殖和农业上的重要地位，对它的研究具有进化和经济学上的意义。

基于实验研究，对植物的两个部分，即气生部分（茎、叶和花）和地下部分（根）的处理是不同的（对植物的地上部分明显可以进行更加准确的视觉上的描述和生化鉴定），部分原因是研究根比较困难。不仅由于正常情况下它们处于肥沃的环境，被有益和有害的微生物困扰，而且即使仅将它们与土壤的物理基质分离也很困难。随着基因组工具不断发展，研究相应的微生物和植物根的特征，甚至研究根和土壤微环境的相互关系变得越来越容易。相对于根和病原物之间的破坏性关系，了解植物根和微生物如根瘤菌和菌根菌的有益关系特别有趣。

植物和病原物的接触界面也很重要，这涉及到植物的气生部

分。越来越多地了解病原物的基因组、病原菌和感染宿主的反应，将二者结合可以开辟一个控制作物病害的新途径。对植物和病原物相互影响的进一步理解可能形成对许多破坏性的植物病症的持久抗性，使得世界范围的食物供应的安全性上升。因此，这些由基因组学研究提供信息的新技术，对于增加产量和降低由于现代农业大量使用除草剂和杀虫剂带来的环境危险非常重要。

光是植物的基本能量来源，也是许多发育过程的调节因子。叶绿素合成基因、许多核及叶绿体编码基因的诱导受光质量和强度的影响。在这方面，核基因组和叶绿体基因组的紧密联合是另一个植物体特有的过程。许多光应答的生化过程已被很好地证明，但是对细胞器基因组中转移至细胞核中的基因的识别能力不仅可以有助于揭示这些应答的协同调控，而且也可以有助于揭示进化史、胁迫和限制因素。另外，对于藻类和其他微生物基因组的研究会给这些研究带来方便。

在高等植物中细胞壁的合成和后来的修饰显然是重要过程。拟南芥基因组的最初注释发现了多于 420 个预计在细胞壁聚合物的合成和修饰中起作用的基因。这些基因中许多属于结构相关酶家族的事实也说明植物基因组的基因冗余现象。然而，正如在本书中讨论的那样，不管这个冗余真实与否，基因家族中的一个成员是否能够有效地替代任何一个其他成员，或者是否这只是表面的冗余，许多基因具有不同底物的特异性或在不同发育阶段发挥作用仍然有待进一步确定。

植物合成许多次级代谢物。综合所有物种，大约有超过十万种次级代谢物。大部分这些代谢物其确切特征和功能都有待于研究确认。综合基因序列、表达图谱和代谢图谱等信息，可以帮助我们确定相关基因、表达模式和这些代谢物之间的相互关系。对一个基因家族中特定成员在特定组织中表达和参与的特定反应的理解，可以揭示控制合成许多化合物的功能基因的冗余水平。

许多在动物中已知的调节或控制发育的过程，如染色质结构