



现代生物农业·植物科学

COMPARATIVE SEQUENCE ANALYSIS OF  
HOMOLOGOUS REGIONS OF TILLERING  
CONTROL GENE IN THE *ORYZA* GENUS

# 稻属植物分蘖控制基因 同源区比较研究

张胜利/著



科学出版社

# 稻属植物分蘖控制基因 同源区比较研究

张胜利 著

科学出版社

北 京

## 内 容 简 介

进行亲缘关系较近的不同植物中特定长度同源区的 DNA 序列比较分析是深入探索同源区区段进化的重要途径。本书介绍的主要内容有植物同源区研究背景知识、稻属分蘖控制基因 *MOC1* 同源区的新基因形成、多倍体中重复基因或基因组中串联复制基因的去功能化、保守非编码序列鉴定、同源区基因保守性与微重排、重复序列的类型及其对基因组结构的影响等方面。本书中介绍的内容对小麦等其他禾本科粮食植物开展重要性状基因同源区研究提供了一定参考。

本书可供作物遗传育种、比较基因组学、分子生物学、生物科学等相关领域或专业的科研院所的科研人员、高校教师、本科生及研究生参考。

### 图书在版编目 (CIP) 数据

稻属植物分蘖控制基因同源区比较研究 / 张胜利著. —北京: 科学出版社, 2015.9

ISBN 978-7-03-045464-5

I. ①稻… II. ①张… III. ①稻属—植物基因工程—研究  
IV. ①Q949.71 ②Q943.2

中国版本图书馆 CIP 数据核字(2015)第 192669 号

责任编辑: 李 迪 / 责任校对: 鲁 素

责任印制: 徐晓晨 / 封面设计: 北京铭轩堂广告设计有限公司

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>

北京京华虎彩印刷有限公司印刷

科学出版社发行 各地新华书店经销

\*

2015 年 9 月第 一 版 开本: 720 × 1000 B5

2015 年 9 月第一次印刷 印张: 10 3/8

字数: 198 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换)

# 前言

水稻作为世界上最重要的粮食植物之一，如何进一步提高水稻的产量和品质进而满足全世界不断增长的粮食需求显得越来越重要。众所周知，20 世纪中叶从作物种质资源中开发和利用的一些优异基因使作物的产量和品质取得了突破性进展。例如，由于水稻、小麦资源中少数几个矮秆基因的开发利用而引发了第一次“绿色革命”；野生稻资源中“野败”的发现与应用对我国的杂交水稻取得巨大的成功起着关键的作用。然而自 20 世纪末以来，作物育种的水平一直没有大的突破。研究发现，这主要是由于育种亲本遗传基础狭窄。因此，如何高效持续地发掘利用种质资源，尤其是野生资源中的优良基因，是目前水稻遗传改良迫切需要解决的一个重要的科学问题。

广泛地分布在热带、亚热带地区的野生稻具有丰富的遗传多样性，以及许多优异的栽培稻中已灭绝了的基因。因此，野生稻已成为栽培稻的产量、品质、抗性及其他农艺性状改良的重要基因库。为了充分有效地发掘和利用野生稻资源，开展稻属植物的进化关系研究显得日益迫切和重要。随着以新一代高通量测序技术为标志的生命科学新技术的不断涌现、新设备的不断开发与利用，稻属植物分类及进化关系研究已经由原来的形态学、细胞学、分子标记等水平发展至 DNA 序列比较水平。

分蘖是水稻、小麦等禾本科作物所具有的一个重要农艺性状，它直接决定水稻、小麦等的穗数及其产量，与创建理想株型密切相关。分蘖数目长期以来一直都是该类作物的一个十分重要的育种目标，是培育“超级水稻”等超高产作物新品种的一个重要方面，因而广受世界各国关注。我国学者李家洋研究组于 2003 年率先搞清了水稻分蘖控制基因 *MOC1* 对分蘖调控的分子机制，对将这一重要农艺性状控制基因应用于水稻等禾谷类作物超级品种的培育具有重大的理论意义与实践价值。研究表明，*MOC1* 属于很保守的称为 GRAS 的转录因子家族一员，在众多植物（包括一些双子叶植物）中都有其同源基因。作者采用分子生物学、比较基因组学及生物信息学相结合的方法对整个稻属各基因组类型植物中 *MOC1* 同源区进行了比较分析，在栽培稻起源演化、同源区基因保守性与微重排、多倍体形成后重复基因去功能化分子机制、保守非编码序列鉴定、新基因形成的 *de novo* 机制、转座子与反转录转座子对基因表达及基因组进化影响的分子机制等方面都进行了较为深入、详细的探讨，以期为有效地发掘和利用野生资源中相关优异基

因提供一定的理论参考和技术支持。

本书中介绍的内容对小麦等其他禾本科粮食植物开展重要性状基因同源区研究也提供了一定参考。对于基因组相对比较大的植物来说(小麦 16~17Gb),同时进行多个物种整个基因组测序及其比较分析对普通研究组来说是不现实的。这种工作花费巨大,且需要有良好的软硬件、优秀的研究团队、技术体系比较成熟的良好平台才能胜任。如果撇开整个基因组而从重要农艺性状基因所在同源区入手,进行亲缘关系较近的不同植物中特定长度基因组同源区的比较序列分析,则不但能大大降低科研费用,还可以进行基因组同源区精细序列比较分析,这将对多个物种整个基因组比较分析得出的宏观结论的有益完善和补充,且能为一般研究组所胜任,更易于在广大普通课题组中推广,使各个普通课题组有机会对自己所关注的重要农艺性状进行研究,得出更加深入的结论,并能为品种相应性状遗传改良提供更多有价值的基因组序列信息。

本书是对作者近十年来在稻属植物基因组同源区研究方面工作的阶段性总结,全书分为两篇:上篇介绍了植物基因组同源区研究背景知识,主要介绍植物比较基因组学研究进展、生物信息学软件在基因组测序和注释中的应用、基因组测序与植物分子育种、生物倍性进化与新基因形成等内容,其中对新一代高通量DNA测序新方法及其应用进行重点阐述;下篇是本书的重点,介绍了稻属分蘖控制基因 *MOCI* 同源区比较研究情况,本部分涉及的内容主要有稻属概况及其基因组学基础、野生稻中优异基因与利用、稻属分蘖控制基因 *MOCI* 同源区的新基因形成、多倍体中重复基因或基因组中串联复制基因的去功能化、保守非编码序列鉴定、同源区基因保守性与微重排、重复序列的类型及其对基因组结构的影响、稻属各基因组类型系统发育关系等方面。

本书的大部分试验过程是在恩师中国科学院遗传与发育生物学研究所陈明生研究员的指导下完成的,还得到了植物基因组学国家重点实验室贾士琴老师及本课题组吴玉峰、刘铁燕、鲁非、陈金锋、隋毅等同学的大力支持与帮助!试验期间还得到恩师西北农林科技大学张改生老师的大力支持,本书在写作过程中还得到了科学出版社的大力支持和热情帮助,同时,本书还得到了国家自然科学基金(NSFC)-河南人才培养联合基金项目(U1204315)及国家自然科学基金面上项目(31571667)的大力资助,在此一并表示衷心的感谢!

由于作者水平所限,书中难免存在不妥之处,敬请读者谅解并提出宝贵意见。

张胜利

2015年8月

# 目 录

## 前言

## 上篇 植物基因组同源区研究背景知识

|       |                                  |    |
|-------|----------------------------------|----|
| 1     | 植物比较基因组学研究进展                     | 3  |
| 1.1   | 禾本科植物比较基因组学研究进展                  | 4  |
| 1.2   | 十字花科植物比较基因组学研究进展                 | 11 |
| 2     | 生物信息学软件在基因组测序和注释中的应用             | 14 |
| 2.1   | 生物信息学定义                          | 14 |
| 2.2   | 基因组学研究中常用的生物信息学软件                | 14 |
| 2.2.1 | 序列组装与质量检测                        | 15 |
| 2.2.2 | 序列注释                             | 16 |
| 3     | 基因组测序与植物分子育种                     | 20 |
| 3.1   | 基因组定义                            | 21 |
| 3.2   | 基因组测序策略                          | 21 |
| 3.2.1 | BAC by BAC 法                     | 21 |
| 3.2.2 | 全基因组霰弹法                          | 21 |
| 3.3   | 关于基因组测序完成标准                      | 22 |
| 3.4   | DNA 测序新方法                        | 22 |
| 3.4.1 | 454 测序系统 (GS 系统)                 | 22 |
| 3.4.2 | Genome Analyzer 测序系统             | 28 |
| 3.4.3 | SOLiD 测序系统                       | 33 |
| 3.4.4 | DNA 测序技术回顾、总结与展望                 | 39 |
| 3.5   | 基因组信息在植物分子育种中的应用                 | 42 |
| 4     | 生物倍性进化与新基因形成                     | 43 |
| 4.1   | 生物多倍体化与重新二倍体化                    | 43 |
| 4.2   | 新基因形成                            | 45 |
| 4.2.1 | 外显子重排 (exon shuffling) 介导的新基因形成  | 45 |
| 4.2.2 | 基因复制 (gene duplication) 介导的新基因形成 | 46 |
| 4.2.3 | 反转录转座 (retrotransposon) 介导的新基因形成 | 46 |
| 4.2.4 | 可移动元件 (mobile element) 介导的新基因形成  | 46 |

|                                                     |    |
|-----------------------------------------------------|----|
| 4.2.5 横向基因转移 (lateral gene transfer) 介导的新基因形成 ..... | 47 |
| 4.2.6 基因断裂/融合 (gene fusion/fission) 介导的新基因形成 .....  | 47 |
| 4.2.7 从头形成( <i>de novo</i> origination) .....       | 47 |

## 下篇 稻属植物分蘖控制基因 *MOC1* 同源区比较研究

|                                            |    |
|--------------------------------------------|----|
| 5 稻属植物分蘖控制基因 <i>MOC1</i> 同源区研究选题依据 .....   | 53 |
| 5.1 水稻遗传改良上面临的主要问题 .....                   | 53 |
| 5.2 解决问题的突破口 .....                         | 53 |
| 5.3 稻属植物分蘖控制基因 <i>MOC1</i> 同源区研究的重要性 ..... | 54 |
| 6 稻属概况及其基因组学基础 .....                       | 55 |
| 6.1 稻属分类概况及栽培稻起源 .....                     | 56 |
| 6.1.1 稻属分类研究进展 .....                       | 56 |
| 6.1.2 栽培稻起源 .....                          | 58 |
| 6.2 稻属植物基因组学基础 .....                       | 59 |
| 6.2.1 稻属植物染色体组型的确定 .....                   | 59 |
| 6.2.2 稻属不同染色体组型间的亲缘关系 .....                | 60 |
| 7 野生稻中的优异基因与利用 .....                       | 62 |
| 7.1 野生稻中的优异基因 .....                        | 62 |
| 7.2 野生稻中优异基因的利用 .....                      | 63 |
| 7.2.1 抗病性 .....                            | 63 |
| 7.2.2 抗虫性 .....                            | 63 |
| 7.2.3 抗非生物胁迫性 .....                        | 64 |
| 7.2.4 品质 .....                             | 65 |
| 7.2.5 产量性状 .....                           | 65 |
| 7.2.6 细胞质雄性不育基因 .....                      | 65 |
| 8 稻属 <i>MOC1</i> 同源区研究手段 .....             | 66 |
| 8.1 研究材料及工具 .....                          | 66 |
| 8.1.1 植物材料 .....                           | 66 |
| 8.1.2 载体及菌株 .....                          | 66 |
| 8.1.3 分子生物学试剂 .....                        | 66 |
| 8.1.4 仪器设备 .....                           | 66 |
| 8.2 研究方法 .....                             | 67 |
| 8.2.1 植物材料的培养 .....                        | 67 |
| 8.2.2 基因组 DNA 的提取 .....                    | 67 |
| 8.2.3 质粒 DNA 的提取 .....                     | 68 |

|        |                                         |     |
|--------|-----------------------------------------|-----|
| 8.2.4  | BAC DNA 的提取                             | 70  |
| 8.2.5  | Southern 探针的制备                          | 70  |
| 8.2.6  | BAC 文库 Southern 杂交                      | 71  |
| 8.2.7  | BAC 文库阳性克隆鉴定                            | 72  |
| 8.2.8  | BAC DNA 酶切与转膜                           | 78  |
| 8.2.9  | BAC 克隆 Southern 杂交复选                    | 78  |
| 8.2.10 | BAC 测序                                  | 79  |
| 8.2.11 | BAC 序列组装与质量检测                           | 79  |
| 8.2.12 | BAC 序列注释                                | 79  |
| 8.2.13 | 植物总 RNA 的提取与纯化                          | 79  |
| 8.2.14 | RT-PCR                                  | 80  |
| 8.2.15 | 籼稻 93-11 的 <i>MOC1</i> 同源区序列“缺口”(gap)填补 | 81  |
| 8.2.16 | 引物合成与序列测定                               | 82  |
| 8.2.17 | Fgenesh 对水稻基因的注释                        | 82  |
| 8.2.18 | 公共数据来源                                  | 82  |
| 8.2.19 | 计算环境                                    | 82  |
| 8.2.20 | 序列分析用到的生物信息学软件                          | 82  |
| 9      | 籼稻 93-11 的 <i>MOC1</i> 同源区序列“缺口”填补      | 84  |
| 9.1    | “缺口”的大小、数目及补“缺口”扩增结果                    | 84  |
| 9.2    | “缺口”估计错误的发现                             | 85  |
| 9.3    | 拼接错误的发现                                 | 86  |
| 9.4    | 基因组中的难测区段                               | 89  |
| 9.5    | 注释分析                                    | 90  |
| 9.6    | 本章小结                                    | 91  |
| 10     | Fgenesh 对水稻基因的注释                        | 92  |
| 10.1   | 注释结果                                    | 92  |
| 10.2   | 注释准确性评价                                 | 93  |
| 10.3   | 高支持度外显子和基因的长度统计                         | 95  |
| 10.4   | 关于序列注释                                  | 96  |
| 10.4.1 | 基因注释                                    | 97  |
| 10.4.2 | 基因预测软件的评价                               | 98  |
| 10.5   | 本章小结                                    | 99  |
| 11     | 野生稻中 <i>MOC1</i> 同源 BAC 的筛选与鉴定          | 100 |
| 11.1   | 野生稻基因组 BAC 文库的初选                        | 100 |
| 11.1.1 | 关于“锚定探针”                                | 100 |
| 11.1.2 | 野生稻基因组 BAC 文库初选结果                       | 100 |

|        |                                         |     |
|--------|-----------------------------------------|-----|
| 11.2   | 对初选出的阳性 BAC 的复选                         | 103 |
| 11.3   | 本章小结                                    | 106 |
| 12     | 野生稻中 <i>MOC1</i> 同源 BAC 的序列测定及注释        | 107 |
| 12.1   | 野生稻中 <i>MOC1</i> 同源 BAC 的序列测定           | 107 |
| 12.2   | 野生稻中 <i>MOC1</i> 同源 BAC 的序列注释           | 108 |
| 12.2.1 | 基因注释                                    | 108 |
| 12.2.2 | 重复序列注释                                  | 110 |
| 12.3   | 本章小结                                    | 110 |
| 13     | 稻属 <i>MOC1</i> 同源区基因保守性与重排              | 111 |
| 13.1   | 基因结构保守性                                 | 111 |
| 13.2   | 基因排列顺序保守性与微重排                           | 111 |
| 13.3   | 易位-倒位引起的多基因重排                           | 113 |
| 13.4   | 未知机制导致的基因组片段移动                          | 113 |
| 13.5   | 关于基因组共线性和微重排的探讨                         | 114 |
| 13.6   | 本章小结                                    | 115 |
| 14     | 基因结构验证                                  | 116 |
| 14.1   | 目标基因的确定                                 | 116 |
| 14.2   | 验证结果                                    | 117 |
| 14.3   | 关于异源多倍体中同源基因对的表达变化的探讨                   | 122 |
| 14.4   | 本章小结                                    | 123 |
| 15     | 稻属 <i>MOC1</i> 同源区中新基因的形成及保守非编码序列的鉴定    | 124 |
| 15.1   | 稻属 <i>MOC1</i> 同源区中新基因的形成               | 124 |
| 15.1.1 | 新基因的头形成                                 | 124 |
| 15.1.2 | 通过基因复制模式产生新基因                           | 126 |
| 15.2   | 保守非编码序列的鉴定                              | 126 |
| 15.3   | 本章小结                                    | 127 |
| 16     | 稻属 <i>MOC1</i> 同源区中的重复序列                | 128 |
| 16.1   | 重复序列注释                                  | 128 |
| 16.2   | 重复序列的类型及其对基因组结构的影响                      | 129 |
| 16.3   | 本章小结                                    | 134 |
| 17     | 稻属植物不同基因组类型 <i>MOC1</i> 同源区间的比较研究       | 135 |
| 17.1   | 二倍体基因组间 <i>MOC1</i> 直向同源区比较             | 135 |
| 17.1.1 | 二倍体 AA 基因组间 <i>MOC1</i> 直向同源区比较与亚洲栽培稻起源 | 135 |
| 17.1.2 | 二倍体 AA 基因组与非 AA 间 <i>MOC1</i> 直向同源区比较   | 136 |
| 17.2   | 二倍体基因组与四倍体基因组间 <i>MOC1</i> 同源区的比较       | 137 |

---

|                                       |            |
|---------------------------------------|------------|
| 17.3 四倍体基因组间 <i>MOC1</i> 同源区的比较 ..... | 139        |
| 17.4 本章小结 .....                       | 139        |
| <b>18 稻属各基因组类型系统发育关系 .....</b>        | <b>140</b> |
| 18.1 稻属 <i>MOC1</i> 基因系统发育分析 .....    | 140        |
| 18.2 关于植物系统发育问题探讨 .....               | 141        |
| 18.3 本章小结 .....                       | 142        |
| 参考文献 .....                            | 143        |
| 附表                                    |            |
| 附表 1 .....                            | 151        |
| 附表 2 .....                            | 154        |
| 缩略词 .....                             | 157        |

# 上 篇

## 植物基因组同源区研究背景知识

以已经完成全基因组测序的模式物种为参考，进行植物重要性状基因的同源区比较研究，有助于比较彻底地搞清重要性状基因所在基因组区段的基因、转座子或顺式调控元件等调控因子的组成情况，尤其是搞清野生材料（或物种）的基因组同源区中重要性状基因的基因组织情况、基因结构与调控区情况等，为优良性状基因的利用提供坚实理论基础，为实现分子模块育种提供理论依据。什么是直向同源区？当前哪些植物中进行了该研究？进展如何？基因组同源区研究需要用到哪些软件、技术？生物倍性进化对基因组同源区中基因功能、序列组成及新基因形成等有何影响？这些问题都将在本部分进行阐述。

本部分涉及的内容主要有与人们粮食、蔬菜供给相关的重要科（禾本科、十字花科）植物比较基因组学研究进展、生物信息学软件在基因组测序和注释中的应用、基因组测序与植物分子育种、生物倍性进化与新基因形成等方面。



# 1 植物比较基因组学研究进展

简单地说，比较基因组学（comparative genomics）就是进行不同生物基因组比较研究的科学。具体来讲，它是基于基因组遗传图谱、物理图谱和序列基础上，对不同生物的基因和基因组结构进行比较分析，进而了解不同生物中同源基因的功能相似性、同源或非同源基因表达时空差异、基因组共线性、基因组中保守功能元件及物种进化的学科。进行模式生物基因组如拟南芥、水稻等和对人类生活有重要影响的植物如小麦、玉米、高粱等基因组之间的比较研究，有助于人们发现基因组间的共线性关系，进而大大加快从复杂、目前还没有完成基因组测序的生物中克隆保守的直向同源基因（orthologous gene）；有助于揭示基因功能进化及同生物适应特定环境有关的基因表达差异机制；有助于阐明基因组的内在结构、物种进化关系和新基因的形成等。通过比较基因组学研究，目前已得出了一些基本规律：内含子和外显子的结构组织比较保守，剪切位点在多种生物中一致；DNA 冗余，即重复序列较多；绝大多数的核心生物功能由相当数量的直向同源基因所编码的蛋白质承担；同一连锁群的同源基因在不同的基因组中有相同的连锁关系等。

新兴起的学科尤其是边缘学科一般都有较多的新名词出现，为了叙述和理解上的方便有必要在此作一阐述。

相似性与同源性：相似性（similarity），表示两者的相似程度，与进化起源是否同一、亲缘关系的远近没有必然关系。同源性（homology），是指两者在进化上有相同的起源。从进化生物学上讲，相似性和同源性是两个不同的概念，相互之间没有直接的等同关系。在进化起源上相同的序列，特别是具有重要生物学功能的保守区段或基因，一般表现为相似。但相似性只是表观的描述性词汇，没有实质性生物学意义。相似的不一定同源，而同源的序列经过长期进化改变，也不一定有较高的相似性。

直向同源与平行同源（图 1.1）：直向同源或直系同源（ortholog），是指两者在进化上起源于同一个祖先基因并垂直传递，分歧于物种形成之后。直向同源基因的特点是：分布于不同物种的基因组中；功能高度保守，甚至在近缘物种中可以互换；基因结构相似；组织特异性和亚细胞分布相似。直向同源基因的鉴定和比较分析是比较基因组学研究的重要内容。平行同源或旁系同源（paralog），是指在同一基因组（或同系的物种基因组）中，由于原始基因的加倍而横向扩增的同源基因。直向同源和平行同源的共性是同源。它们的区别在于：在进化起源上，直

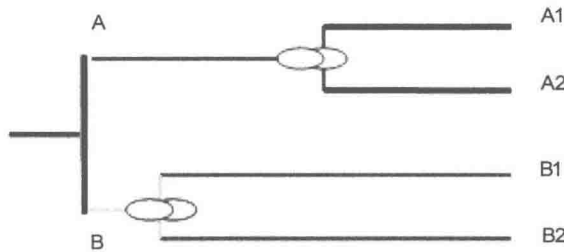


图 1.1 直向同源基因与平行同源基因 (彩图可扫描封底二维码获取)  
Fig. 1.1 Orthologous genes and paralogous genes

图中 A、B 代表不同的两个物种；部分重叠的两个椭圆表示基因或基因组加倍事件；A1、A2、B1、B2 表示两个物种中不同的基因；A1 与 B1、B2，A2 与 B1、B2，B1 与 A1、A2，B2 与 A1、A2 之间为直向同源基因；A1 与 A2、B1 与 B2 之间为平行同源基因（本书彩图请扫书后二维码查看，余同）

向同源强调在不同基因组中的垂直传递，是随着新物种的形成而产生，它的变化能在一定程度上代表物种的进化；而平行同源则是在同一基因组中的横向加倍，形成了一个个成员数目不一的基因家族，是基因或基因组加倍后发生分歧的。在功能上，直向同源一般功能高度相似，而平行同源对功能没有严格要求，相反，很多平行同源基因功能产生了较大分化。

本部分仅就与人们生产、生活密切相关并且研究又比较深入的两个科（即禾本科与十字花科）的植物比较基因组学研究概况进行概述。

### 1.1 禾本科植物比较基因组学研究进展

禾本科属于单子叶植物纲，禾本目。禾本科是一个大科，有 620 多个属，1 万多个种。从种的数量看，在显花植物中居于第四位。禾本科植物遍及全世界，是陆地植被的主要成分，尤其是各种类型草原的主要组成者。禾本科是单子叶植物中经济价值最高的一个科。人类赖以生存的主要粮食作物如水稻、小麦、玉米、粟、高粱等，以及经济作物如甘蔗、竹类和很多牧草等都属于禾本科。禾本科植物不但是人类粮食、动物饲料的主要来源，还为造纸、制糖、制药、酿造、编织等方面提供丰富的原材料，而且在绿化环境、保护堤岸、水土保持等方面也具有重要意义。

正是由于其特殊重要的地位，人们对该科植物的研究开展得较早而且较深入。早期的禾本科植物比较基因组学研究主要是在细胞遗传学水平上进行的。DNA 分子水平上的比较基因组学研究始于 RFLP 技术出现以后。限制性片段长度多态性 (restriction fragment length polymorphism, RFLP) 是根据不同个体基因组的限制性内切酶酶切位点的变化（如碱基突变、碱基的插入、缺失、甲基化等）而导致酶切片段产生多态性，转膜后通过特定探针杂交进行检测，从而可比较不同个体 DNA 水平上的差异（即多态性），多个探针的比较可以确立生物的进化和分类关系。所

用探针可以为同种或不同种位于染色体不同位点的基因组 DNA 的克隆或 cDNA 克隆, 当某个性状(基因)与某个(些)分子标记共分离时, 表明该性状(基因)与这个(些)分子标记连锁, 它们之间交换值的大小, 即表示目标基因与分子标记之间的距离, 从而将基因定位于分子图谱上。RFLP 技术自从 20 世纪 80 年代诞生以来就被广泛用于基因组遗传图谱构建、基因定位及生物进化和分类的研究, 而且至今仍被许多遗传学家、育种学家所广泛应用, 随后还陆续开发出了 SSR(1984 年)、STS(1989 年)、RAPD(1990 年)、AFLP(1993 年)、SRAP(2001 年)等分子标记。这为进行不同生物间的比较基因组学研究提供了强有力的手段。

较早的利用 RFLP 技术进行比较基因组学研究的是 Bonierbale 等(1988)和 Chao 等(1988)。前者利用 RFLP 技术并结合跨物种探针构建了番茄和马铃薯之间的高密度比较遗传图谱, 后者构建了六倍体小麦中 A、B、D 基因组间的比较遗传图谱。之后, Ahn 和 Tanksley(1993)基于一套不同物种中的 cDNA 克隆在水稻和玉米中所鉴定的直向同源基因位点构建了这两个物种的高密度遗传图谱(图 1.2)。他们发现这两个物种间保守的连锁区占了这两个物种基因组的 2/3 以上, 甚至某些整条染色体或者染色体臂从基因排列顺序和种类上看都是完全保守的, 如玉米的 2 号、10 号染色体和水稻的 4 号染色体。不管是从外部形态、开花授粉习性上, 还是从生长环境上, 两个有着显著差异的物种, 其基因组间却有如此高的保守性, 因此, Ahn 和 Tanksley(1993)的工作激起了众多从事植物科学研究的科学家的兴趣, 使禾本科植物的比较基因组学研究成了长期以来的研究热点。例如, Kurata 等(1994)对水稻和小麦之间共线性的研究; Devos 等(1994)对玉米和小麦之间共线性的研究; Grivet 等(1994)对甘蔗、玉米、高粱的比较作图的研究; van Deynze 等(1995)将研究范围从水稻、玉米、小麦之间的比较扩大到二倍体的燕麦; Dufour 等(1996)对玉米、甘蔗和高粱间的共线性研究; Devos 等(1998)对水稻和谷子间共线性的比较研究等。Moore 等(1995)对前人及他们自己关于禾本科植物基因组间保守性研究结果进行了总结, 发现可以将水稻的染色体划分成 19 个连锁片段(linkage segment), 其他禾谷类作物如水稻、谷子、甘蔗、高粱、玉米、小麦等的染色体都可以用与之同源的水稻上的不同连锁片段或其组合来代表(图 1.3), 并推断出了禾谷类作物的“祖先染色体”。Moore 等(1995)利用上述作物间的共线性信息还构建了这 6 种作物间的基因组比较图谱。3 年后, Gale 和 Devos(1998)将燕麦加入了 Moore 等提出的共线性图谱中, 对禾谷类作物的共线性图谱进行了完善和补充(图 1.4)。禾谷类作物共线性图谱的构建对人们开展禾谷类作物的研究有着十分重要的意义。首先, 它可以帮助人们在基因组信息较少的作物中利用共线性关系来发展 DNA 或蛋白质标记; 利用基因的直向同源性来定位并克隆另一种作物中的直向同源基因; 也可以为人们弄清一些禾谷类作物在物种形成后通过重复序列的非随机累积而导致该作物特定基因组区段膨胀的原因提供一些线索。

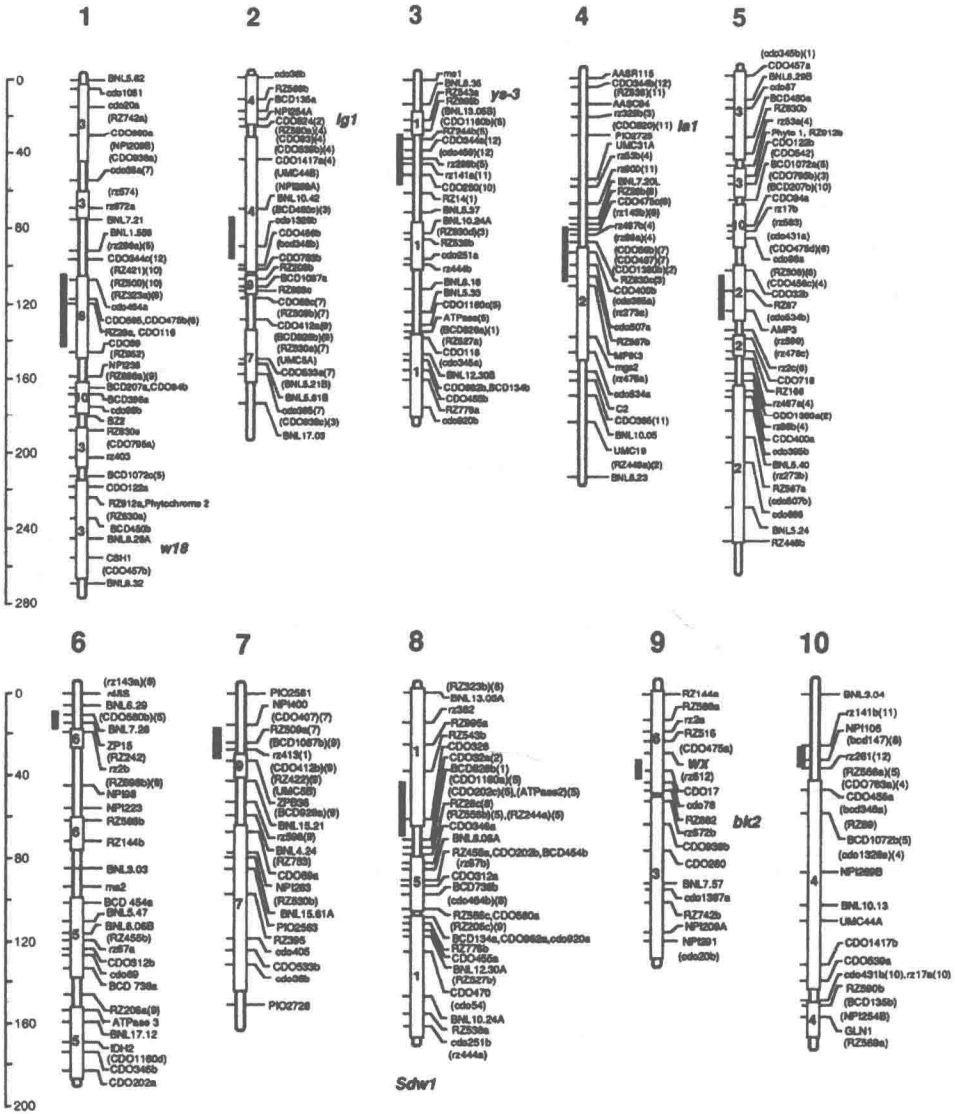


图 1.2 玉米和水稻的 RFLP 比较遗传连锁图 (Ahn and Tanksley, 1993)

Fig. 1.2 RFLP comparative linkage map of maize and rice

图中最左侧为标尺 (cM)；编号 1~10 及其下面长条形图表示玉米的 10 条染色体；染色体右侧的位点名称表示不同的 cDNA 克隆：RZ 表示水稻叶片 cDNA，CDO 表示燕麦叶片 cDNA，BCD 表示大麦叶片 cDNA，BNL 表示玉米基因组克隆；内含有不同数字编号的长方形框表示玉米的某染色体或其一部分与水稻的某染色体一部分的保守区域；位点名称后括号中的数字表示其水稻中同源位点的染色体位置；各个染色体左侧的实心长条表示着丝粒的大概位置；染色体右侧不同位置的 *ys3* (黄条斑)、*Sdw1* (半矮秆)、*bk2* (脆秆)、*w18* (幼苗白斑)、*lg1* (无叶舌)、*wx* (蜡质)、*la1* (惰性) 等表示形态标记基因的大概位置

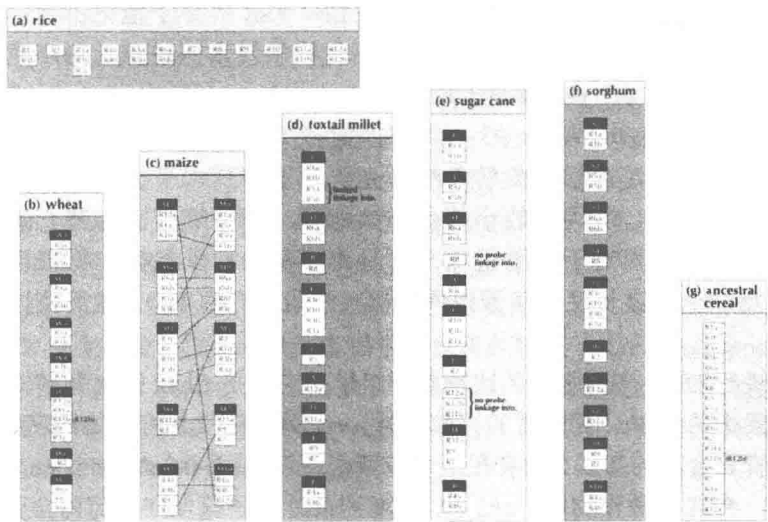


图 1.3 基于水稻连锁片段的禾谷类作物基因组比较 (Moore et al., 1995) (彩图可扫描封底二维码获取)

Fig. 1.3 The comparison of grass genome based on rice linkage segments

(a) 水稻染色体被划分成的连锁片段 (图中的白色方框, 下同); 基于与水稻不同连锁片段同源性的禾谷类作物染色体的重建[小麦 (b)、玉米 (c)、谷子 (d)、甘蔗 (e)、高粱 (f)]; 重建的玉米染色体中用连线连接的两个连锁片段表示该两个片段为复制片段; (g) 基于水稻不同连锁片段而重建的禾谷类作物祖先染色体

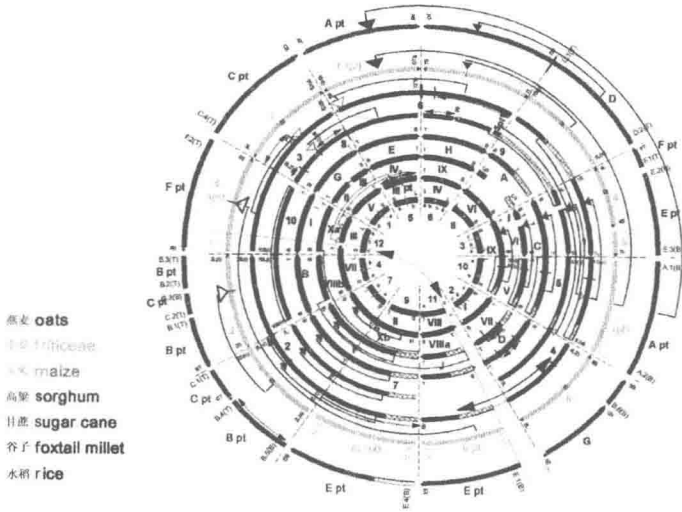


图 1.4 禾本科作物基因组比较图谱 (Gale and Devos, 1998) (彩图可扫描封底二维码获取)

Fig. 1.4 The comparative map of grasses

图中共列举了水稻、谷子、甘蔗、高粱、玉米、小麦、燕麦 7 种禾本科作物的比较基因组图谱, 各物种基因组用与其名称颜色一致的实心粗线来表示; 带有连线的实心箭头表示染色体间的倒位和易位; “△”及“□”分别表示各染色体端粒和着丝粒的位置; 各染色体中阴影的区域表示目前能用于比较的信息较少