

作物科学研究



棉花纤维发育生物学

喻树迅 朱玉贤 陈晓亚 主编



科学出版社

棉花纤维发育生物学

喻树迅 朱玉贤 陈晓亚 主编

科学出版社

北京

内 容 简 介

本书内容包括棉花的起源与进化,棉花纤维基因资源的创制、挖掘与利用及遗传连锁图谱的构建;棉花基因组学、棉花纤维转录组学和蛋白质组学研究进展;棉花纤维发育相关基因、启动子的克隆与功能验证及转基因优质材料创制;棉花纤维品质和产量相关性状遗传分析方法研究及 QTL 定位、棉花抗逆高产分子改良的遗传基础及优质高产多抗棉花新品种的分子改良。

本书适合从事棉花生物学研究、植物细胞伸长研究及相关研究的科研人员参考。

图书在版编目 (CIP) 数据

棉花纤维发育生物学/喻树迅,朱玉贤,陈晓亚主编. —北京:科学出版社, 2016.8

ISBN 978-7-03-045099-9

I. ①棉… II. ①喻… ②朱… ③陈… III. ①棉纤维-发育生物学
IV. ①S562.01

中国版本图书馆 CIP 数据核字(2015)第 132957 号

责任编辑:韩学哲 贺窑青 / 责任校对:郑金红 张小霞

责任印制:肖 兴 / 封面设计:铭轩堂

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>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2016 年 8 月第 一 版 开本:787×1092 1/16

2016 年 8 月第一次印刷 印张:33 1/4

字数:770 000

定价:280.00 元

(如有印装质量问题,我社负责调换)

《棉花纤维发育生物学》编辑委员会

主 编 喻树迅 朱玉贤 陈晓亚

副主编 李付广 张献龙 刘进元 马峙英 祝水金 裴 炎

编 委 (按姓氏拼音排序)

杜雄明 范术丽 贾银华 林忠旭 刘传亮

陆才瑞 罗 明 庞朝友 彭 军 秦咏梅

单淳敏 商海红 上官小霞 沈法富 宋国立

宋美珍 孙君灵 涂礼莉 王寒涛 王凌健

王省芬 魏恒玲 徐海明 叶武威 于霁雯

袁有禄 张 兵 朱 军 邹长松

第一章 林忠旭 张献龙 杜雄明 贾银华 孙君灵 宋国立

邹长松 陆才瑞

第二章 朱玉贤

第三章 朱玉贤 马峙英 王省芬 秦咏梅

第四章 刘进元 张 兵

第五章 上官小霞 单淳敏 王凌健 陈晓亚 罗 明 裴 炎

刘传亮 于霁雯 李付广 涂礼莉 张献龙

第六章 喻树迅 祝水金 朱 军 徐海明 沈法富 马峙英

王省芬 袁有禄 叶武威 范术丽 宋美珍 于霁雯

魏恒玲 庞朝友 王寒涛 商海红

前 言

1997年6月4日，国家科技领导小组第三次会议制定了《国家重点基础研究发展规划》，科学技术部随即组织实施了国家重点基础研究发展规划（973计划）。973计划的主要任务是以认识自然现象、揭示客观规律为主要目的，并遵循基础研究的特点和规律，开展重点探索性基础研究，以解决在国民经济和社会发展，以及科学自身发展中所提出的重大科学问题为目的的定向性基础研究。973计划无论从理论水平还是从资助力度上都高于国家高技术研究发展计划（863计划）和自然科学基金（一般性的理论科学研究）。实施973计划的战略目标是加强原始性创新，在更深的层面和更广泛的领域解决国家经济与社会发展中的重大科学问题，以提高我国自主创新和解决重大问题的能力，为国家未来发展提供科学支撑。在项目管理上，973计划采用首席科学家负责制。

棉花是我国重要的经济作物，常年种植面积在8000万亩左右，在我国国民经济中具有举足轻重的地位。然而，我国每年进口优质原棉100多万吨，且优质原棉主要（95%）依赖进口，严重影响棉农收入和国民经济的健康可持续发展。提高棉花品质及产量是我国棉花安全的重大需求。随着项目研究的深入，需要在棉花纤维品质功能基因组学研究和优质高产的分子改良方面进行持续深入研究，需要开展纤维数量与品质间的调控模式研究，全面解析棉花纤维发育相关的基因互作网络，探讨棉花纤维改良的分子机制，全面提升国产原棉的国际竞争力，保证我国棉花产业的可持续发展。针对我国棉花产业发展中的重大需求和关键科学问题，科学技术部自2004年以来，通过连续两轮973计划对棉花纤维有关的基础研究进行立项资助，第一轮973计划“棉花纤维品质功能基因组学研究及分子改良”（2004CB117300），第二轮973计划“棉花纤维品质功能基因组研究及优质高产新品种的分子改良”（2010CB126000），均由中国农业科学院棉花研究所主持，喻树迅院士为首席科学家，主要承担单位有北京大学、清华大学、中国科学院上海生命科学研究院、浙江大学、华中农业大学、西南大学、河北农业大学、中国科学院微生物研究所、上海交通大学等。

棉花是天然纤维植物，我国是世界上的主要产棉国家，棉花生产在我国国民经济中起着举足轻重的作用，棉花纤维发育的机制研究是棉花纤维品质研究的重大基础课题。连续两轮973计划的实施，在科学技术部的大力支持下，棉花纤维品质相关基础研究实现了国家重大需求导向与科学目标的高度统一。以喻树迅院士为首席科学家的棉花纤维品质项目团队，经过连续10年两轮973计划的协同攻关，不仅在管理机制上实现了从基础研究到材料创新应用的有机衔接，而且在棉花纤维发育、品质形成与分子改良等方面取得了理论、方法和技术的创新及突破，大大提升了我国棉花科技的国际影响力。项目团队先后发现了乙烯可促进棉花纤维细胞伸长、超长链脂肪酸在转录组水平可调节乙烯生物合成与释放和调控纤维发育，以及生长素可促进纤维细胞起始并提高皮棉产量的分子机制，相继探明了乙烯促进棉花纤维伸长的信号通路（COBP），而且在国际上率先完

成 D 基因组草图绘制和 A 基因组测序,建立了棉花纤维蛋白质组学研究的新方法,并构建了与纤维发育密切相关的 235 个差异显示蛋白表达谱,建立了棉花规模化转基因技术体系、高效外源基因功能验证平台和棉花分子聚合育种体系,成功将海岛棉优质基因转移到陆地棉中,获得一批高强优质纤维材料,并成功培育出我国第一个国审优质棉新品种‘中棉 70’。10 年来,项目团队在 *Nature Genetics*、*Nature Biotech*、*Nature Commun*、*Plant Cell* 等国际知名学术期刊发表了一批有影响的论文,共荣获国家级科技奖励成果 8 项、省部级奖励 14 项,其中国家自然科学奖二等奖 1 项,国家技术发明奖二等奖 1 项,国家科学技术进步奖二等奖 6 项,省级科学技术进步奖一等奖 4 项,2 位科学家分别当选中国科学院院士和中国工程院院士,1 位科学家获得国家杰出青年科学基金项目资助,2 位科学家入选国家“万人计划”,打造出一支高水平的创新研究队伍。

根据科学技术部基础研究司“关于 2014 年 973 计划项目结题验收工作安排的通知”(国科基函[2014]24 号)要求,973 计划于 2014 年 9 月 20~21 日在北京召开课题结题验收。为了总结 973 计划实施 10 年来的研究进展,展示两轮 973 计划的巨大成就,我们编写这部专著,作为我国棉花纤维基础性研究的里程碑。

在科学技术部的资助下,973 计划顾问组、973 计划项目责任专家和项目各参加单位主持人及参加人在计划实施的 10 年中付出了巨大的艰辛,也取得了可喜的进展。如果说艰辛的 10 年,换回了我们在棉花纤维发育基础研究上前进的一小步,我们应该感谢前人的研究。在棉花纤维基础研究的长河中我们才刚刚起步,正如牛顿所说,“我不知道在别人看来,我是什么样的人;但在我自己看来,我不过就像是一个在海滨玩耍的小孩,为不时发现比寻常更为光滑的一块卵石或比寻常更为美丽的一片贝壳而沾沾自喜,而对于展现在我面前的浩瀚的真理的海洋,却全然没有发现。”

著者

2015 年 8 月

目 录

前言

第一章 棉花纤维基因资源与创新.....	1
第一节 棉花起源与进化.....	1
一、棉属的系统分类.....	1
二、二倍体棉种的起源与进化.....	5
三、四倍体棉种的起源与进化.....	6
第二节 棉花纤维基因资源挖掘与创制.....	14
一、棉花纤维基因资源创制.....	14
二、棉花纤维基因资源挖掘与利用.....	33
第三节 棉花遗传连锁图的构建.....	50
一、遗传群体的构建.....	50
二、分子标记的开发.....	56
三、高密度连锁图的构建.....	75
参考文献.....	87
第二章 棉花基因组学研究.....	90
第一节 植物基因组学研究进展.....	90
一、高通量 DNA 测序技术.....	90
二、模式植物基因组学研究进展.....	96
三、重要农作物基因组学研究进展.....	96
第二节 棉花 D 基因组（雷蒙德氏棉）研究.....	113
一、测序、组装与注释.....	113
二、古六倍体化事件和全基因组复制.....	116
三、转座子扩增与简单序列重复.....	118
四、纤维发育、棉酚生物合成相关基因分析.....	118
第三节 棉花 A 基因组（亚洲棉）研究.....	121
一、测序、组装与注释.....	121
二、A 基因组与 D 基因组比较分析.....	123
三、重要农艺性状相关功能基因分析.....	126
四、A 基因组资源应用前景.....	128
参考文献.....	130
第三章 棉花纤维转录组学研究.....	134
第一节 转录组学研究技术.....	135

一、功能基因组学核心技术.....	135
二、基因芯片数据分析技术及相关资源.....	137
三、基因表达系列分析技术.....	138
四、转录组分析和 RNA-Seq.....	138
五、棉花 cDNA 减法技术.....	140
第二节 乙烯与纤维发育.....	141
一、棉花中参与乙烯合成基因的研究.....	141
二、乙烯在棉花纤维伸长过程中起重要作用.....	142
三、油菜素内酯对纤维细胞发育的影响及与乙烯的相互作用.....	144
第三节 超长链脂肪酸与纤维发育.....	147
一、超长链脂肪酸的合成途径及其调控机制研究.....	147
二、植物体内超长链脂肪酸功能研究.....	150
三、棉花纤维中超长链脂肪酸功能研究.....	151
第四节 初生壁合成与纤维发育.....	158
一、植物细胞壁的结构与核苷糖代谢.....	158
二、棉花纤维中初生壁合成的研究.....	162
第五节 次生壁加厚期转录组研究.....	172
一、海陆 F ₂ 群体次生壁加厚期转录组研究.....	174
二、海陆 BC ₁ 群体次生壁加厚期转录组研究.....	180
三、次生壁加厚期转录组基因挖掘与分析.....	193
参考文献.....	212
第四章 棉花纤维蛋白质组学研究.....	216
第一节 棉花纤维蛋白质组学研究新方法.....	217
一、棉花纤维总蛋白的酚提取法.....	217
二、棉花纤维总蛋白提取方法评价.....	219
三、棉花纤维蛋白质组分析技术路线.....	219
第二节 长纤维伸长期差异显示蛋白的鉴定与图谱构建.....	221
一、长纤维伸长期蛋白质总量的动态变化.....	222
二、长纤维伸长期蛋白质组的动态变化.....	223
三、长纤维伸长期差异表达蛋白的质谱鉴定.....	223
四、长纤维伸长过程中的活跃生化途径.....	226
第三节 短纤维起始期差异显示蛋白的鉴定与图谱构建.....	227
一、短纤维起始期蛋白质组的动态变化.....	228
二、短纤维起始期差异表达蛋白的质谱鉴定.....	230
三、短纤维起始期重要蛋白的动态变化.....	231
四、短纤维起始期的差异表达蛋白调控网络.....	235
第四节 棉花纤维细胞差异显示磷酸化蛋白质的质谱鉴定.....	236
一、棉花纤维伸长过程中差异显示蛋白的磷酸化修饰程度.....	236

二、二维电泳凝胶中单个磷酸化蛋白质的质谱鉴定方法	237
三、磷酸化肽段的预测及其磷酸化位点的质谱鉴定	238
四、伸长期差异表达磷酸化蛋白质的功能分析	242
第五节 蛋白质组学分析所揭示的棉花纤维伸长过程中 活跃的蛋白质网络	245
一、碳水化合物代谢相关的蛋白质网络	246
二、蛋白质代谢相关的蛋白质网络	248
三、细胞骨架和囊泡运输相关的蛋白质网络图	248
四、渗透压调控相关的蛋白质网络	250
参考文献	251
第五章 棉花纤维基因克隆与功能验证	255
第一节 棉花纤维发育重要基因的克隆与功能验证	255
一、转录因子对棉花纤维发育的调控	255
二、植物激素对棉花纤维发育的调控	262
三、功能基因对棉花纤维发育的调控	265
四、相关基因的克隆与功能鉴定	276
第二节 棉花纤维表达启动子的克隆和功能验证	295
一、纤维表达启动子的研究意义和进展	295
二、棉花源纤维优势表达启动子	299
三、非棉花源种皮/纤维表达启动子	306
四、棉花纤维特异/优势表达启动子研究的问题和展望	310
第三节 转基因技术体系的改良及应用	311
一、转基因技术体系的改良	311
二、时空调控生长素合成基因增加棉花纤维数量	316
三、转基因优质纤维材料的创制	331
参考文献	345
第六章 优质高产棉花新品种的分子改良	355
第一节 棉花纤维品质和产量性状的遗传分析方法	355
一、基于混合线性模型的复合区间作图法	355
二、基于 MCIM 的多变量 QTL 分析方法	361
三、基于 MCIM 的种子性状 QTL 分析方法	368
四、基于 MCIM 的关联分析方法	372
五、基于 MCIM 的分析软件开发	379
第二节 棉花纤维优质高产分子改良的遗传基础	384
一、棉花纤维品质和产量相关性状的 QTL 定位	384
二、棉花种子品质相关性状的 QTL 定位	414
三、棉花纤维品质分子标记聚合选择	422
第三节 棉花抗逆高产分子改良的遗传基础	427
一、棉花抗衰老功能基因筛选	428

二、棉花抗旱耐盐性状的遗传分析与基因克隆.....	453
三、棉花抗黄萎病相关基因克隆与分析.....	466
第四节 优质高产多抗棉花新品种的分子改良.....	481
一、陆海渐渗优质材料创制.....	481
二、棉花抗旱耐盐新材料创制.....	489
三、棉花光敏雄性不育系材料创制.....	495
四、棉花抗旱衰老材料创制.....	500
五、棉花优质高产分子聚合新品种选育.....	507
参考文献.....	510
展望.....	518

第一章 棉花纤维基因资源与创新

棉花纤维是棉花经济价值最重要的部分，掌握棉花纤维资源数量的多少直接影响着棉花优质育种的广度和深度。棉花在长时期的进化历程中逐渐形成了以纤维为主要利用价值的现代栽培种，挖掘棉花中现有的优异纤维资源，或利用生物或物理化学诱变技术创造纤维资源是丰富和利用棉花纤维资源多样性的重要手段，对这些纤维基因的遗传定位在一定程度上加速了纤维优异基因聚合，为培育优质高产的棉花新品种提供了资源和技术支撑。

本章引证形态学、分子细胞生物学等多方面的证据介绍了棉属的分类、起源与进化，探讨了四倍体棉种的 A、D 亚基因组的起源。详细介绍了通过杂交回交结合系统选择、远缘杂交、物理化学诱变、基因工程等方法创造的纤维和产量新种质，及其对这些种质材料的分子鉴定。介绍了各类遗传群体的构建、分子标记的开发，以及棉花种间、种内遗传图谱的构建。

第一节 棉花起源与进化

一、棉属的系统分类

(一) 棉属分类

棉属 (*Gossypium*) 属于锦葵目 (Malvales)，锦葵科 (Malvaceae)，棉族 (Gossypiceae)。棉属，共分为 4 个亚属，包括 50 多个种，其中二倍体棉种 ($2n=2x=26$) 有 40 多个，四倍体棉种 ($2n=4x=52$) 有 5 个 (Fryxell, 1992)。棉属植物广泛分布于全世界，种间形态特征和生长习性存在广泛的多样性，而早期的棉属植物分类研究，基本上依据棉属植物的形态特征和地理分布，不同的分类学家所偏重的形态特征不完全相同。

自从细胞学研究手段被广泛应用之后，棉属的分类得到关键性的突破。Nikolajeva (1923) 提出，亚洲棉和草棉的体细胞染色体数目为 $2n=26$ ，陆地棉和海岛棉的染色体数目为 $2n=52$ 。在棉属分类学中作出划时代贡献的当推美国的 Beasley (1940; 1941)，他在总结前人已积累的棉属细胞遗传学知识的基础上结合他本人进行的大量的种间杂交研究成果，以杂种后代减数分裂时染色体的配对情况为依据，参照棉种的地理分布，提出了划分染色体组的细胞学分类原则，即凡种间染色体同源性较高的棉种归为同一染色体组，反之则列为不同的染色体组。二倍体棉种分为 A、B、C、D、E、F、G 和 K 等 8 个基因组 (Beasley, 1940; 1941; Endrizzi et al., 1985; Wendel et al., 2010)，其中 A、B、E 和 F 基因组代表亚洲-非洲棉；C、G 和 K 基因组代表澳洲棉；D 基因组代表美洲棉；四倍体棉种被认为是 A 基因组和 D 基因组种间杂交产生的双二倍体种，将其定为 AD 基因组 (图 1-1)。

关于棉属起源的时间，已有大量文献对此进行了推测。棉属约在 1250 万年前从最近的亲属分开，与跨海洋传播有关 (Johnson and Thein, 1970; Edwards et al., 1974)。由于缺乏清晰的化石记录，重建棉属的系统发育的方法主要是通过分子数据，如叶绿体基因组 (cpDNA) 限制位

点 (Wendel and Albert, 1992)、5S 核糖体基因及其间隔区的 DNA 序列 (Cronn et al., 1996)、叶绿体基因 *ndhF*、5.8S 基因及其侧翼内转录间隔区 (ITS) (Seelanan et al., 1997)。利用这些基因序列获得的棉属系统发育关系大部分都与基因组名称和地理分布一致 (图 1-1、表 1-1)。

(二) 二倍体棉种的分类

澳洲棉种 (Australian species): 澳洲棉种 (亚属 *Sturtia*) 包括 C、G、K 基因组, 分别包含 2 种、3 种和 12 种。DNA 序列数据表明这 3 个基因组属于自然家系, 与先前它



图 1-1 棉属的系统发生树

表 1-1 棉属各基因组编号和地理分布

编号	种名	中译名	基因组类型	地理分布
1	<i>G. hirsutum</i>	陆地棉	(AD) ₁	中美洲
2	<i>G. barbadense</i>	海岛棉	(AD) ₂	南美洲
3	<i>G. tomentosum</i>	毛棉	(AD) ₃	夏威夷
4	<i>G. mustelinum</i>	黄褐棉	(AD) ₄	巴西
5	<i>G. darwinii</i>	达尔文氏棉	(AD) ₅	加拉帕戈斯群岛
6	<i>G. herbaceum</i>	草棉	A ₁	中东和非洲
7	<i>G. arboreum</i>	亚洲棉	A ₂	远东和印度
8	<i>G. anomalum</i>	异常棉	B ₁	非洲
9	<i>G. triphyllum</i>	三叶棉	B ₂	非洲南部
10	<i>G. capitiviridis</i>	绿顶棉	B ₃	佛得角群岛
11	<i>G. senarensis</i>	桑纳氏棉	B	非洲

续表

编号	种名	中译名	基因组类型	地理分布
12	<i>G. trifurcatum</i>	三叉棉	B	非洲
13	<i>G. sturtianum</i>	斯特提棉	C ₁	澳大利亚中部
14	<i>G. nandewarensis</i>	南岱华棉	C _{1-n}	澳大利亚东南部
15	<i>G. robinsonii</i>	鲁滨逊氏棉	C ₂	澳大利亚西部
16	<i>G. thurberi</i>	瑟伯氏棉	D ₁	亚利桑那州
17	<i>G. armourianum</i>	辣根棉	D ₂₋₁	墨西哥
18	<i>G. harknessii</i>	哈克尼西棉	D ₂₋₂	墨西哥
19	<i>G. davidsonii</i>	戴维逊氏棉	D _{3-d}	墨西哥
20	<i>G. klotzschianum</i>	克劳茨基棉	D _{3-k}	加拉帕戈斯群岛
21	<i>G. aridum</i>	旱地棉	D ₄	墨科利马地区
22	<i>G. raimondii</i>	雷蒙德氏棉	D ₅	秘鲁
23	<i>G. gossypioides</i>	拟似棉	D ₆	墨互哈卡州
24	<i>G. lobatum</i>	裂片棉	D ₇	墨米却肯州
25	<i>G. trilobum</i>	三裂棉	D ₈	墨西哥西部
26	<i>G. laxum</i>	松散棉	D ₉	墨西哥雷罗
27	<i>G. turneri</i>	特纳氏棉	D ₁₀	摩索诺拉州
28	<i>G. schwendimanii</i>	施温迪茫棉	D ₁₁	墨西哥
29	<i>G. stocksii</i>	司笃克氏棉	E ₁	阿拉伯地区
30	<i>G. somalense</i>	索马里棉	E ₂	北非
31	<i>G. areysianum</i>	亚雷西亚棉	E ₃	南也门
32	<i>G. incanum</i>	灰白棉	E ₄	阿拉伯南部
33	<i>G. benadirensis</i>	伯纳迪氏棉	E	北非
34	<i>G. bricchettii</i>	伯里切氏棉	E	北非
35	<i>G. vollesenii</i>	佛伦生氏棉	E	索马里
36	<i>G. longicalyx</i>	长萼棉	F ₁	非洲东部
37	<i>G. bickii</i>	比克氏棉	G ₁	澳大利亚中部
38	<i>G. australe</i>	澳洲棉	G ₂	澳大利亚中北部
39	<i>G. nelsonii</i>	奈尔逊氏棉	G ₃	澳大利亚中部
40	<i>G. exiguum</i>	小小棉	K ₁	澳大利亚
41	<i>G. rotundifolium</i>	圆叶棉	K ₂	澳大利亚
42	<i>G. populifolium</i>	杨叶棉	K ₃	澳大利亚
43	<i>G. pilosum</i>	稀毛棉	K ₄	澳大利亚
44	<i>G. marchantii</i>	马全特氏棉	K ₅	澳大利亚
45	<i>G. londonderriensis</i>	伦敦德里棉	K ₆	澳大利亚
46	<i>G. enthyle</i>	林地棉	K ₇	澳大利亚
47	<i>G. costulatum</i>	皱壳棉	K ₈	澳大利亚
48	<i>G. cunninghamii</i>	肯宁汉氏棉	K ₉	澳大利亚最北部
49	<i>G. pulchellum</i>	小丽棉	K ₁₀	澳大利亚
50	<i>G. nobile</i>	显贵棉	K ₁₁	澳大利亚
51	<i>G. anapoides</i>	孪生叶面棉	K ₁₂	澳大利亚

们分别被分成 *Sturtia* 组 (C 基因组)、*Hibiscoidea* 组 (G 基因组) 和 *Grandicalyx* 组 (K 基因组) 一致。这 3 个基因组的分类, 仅对 C 基因组和 G 基因组研究得比较多, 其分类学地位已经明确 (Wendel and Albert, 1992; Seelanan et al., 1997)。K 基因组的棉种有不寻常的地理、形态和生态特征, 具有适应干旱的综合性状, 特别是具有多年生季节周期生长模式, 它们的分类地位还没有确定 (Wendel et al., 2010)。

非洲-亚洲棉种 (African-Asian species): 非洲-亚洲棉种包括 14 种。分布于棉属基因组的 A、B、E、F 基因组, 具有丰富的物种多样性。A 基因组包括 *Gossypium* 亚组的 2 个栽培种亚洲棉和草棉, B 基因组包括 *Anomala* 亚组的 3 个非洲种异常棉、三叶棉、绿顶棉。三叉棉也可能属于 B 组。唯一一个 F 基因组种长萼棉, 在细胞遗传学上有独特的特性 (Phillips, 1966), 在形态上与其他种相互分开, 这可能是因为它比其他二倍体棉种更适应湿地环境。

美洲二倍体棉种 (American diploid species): 美洲二倍体棉种为新世界二倍体 D 基因组种, 共 13 种, 包括 *Houzingenia* 组 (含亚组 *Houzingenia*、*Integrifolia*、*Caducibracteolata*) 和 *Erioxylum* 组 (含亚组 *Erioxylum*、*Selera*、*Austroamericana*)。其中, 亚组 *Houzingenia* 包括瑟伯氏棉 (D_1) 和三裂棉 (D_8); 亚组 *Integrifolia* 包括克劳茨基棉 (D_{3-k}) 和戴维逊氏棉 (D_{3-d}); 亚组 *Caducibracteolata* 包括辣根棉 (D_{2-1})、哈克尼西棉 (D_{2-2}) 和特纳氏棉 (D_{10}); 亚组 *Erioxylum* 包括旱地棉 (D_4)、裂片棉 (D_7)、松散棉 (D_9) 和施温迪茫棉 (D_{11}); 亚组 *Selera* 包括拟似棉 (D_6); 亚组 *Austroamericana* 包括雷蒙德氏棉 (D_5) (Wendel and Cronn, 2003)。

与澳洲棉和非洲-亚洲棉相比, 美洲棉的分类和系统发生学方面得到了广泛的研究。尽管 6 个亚组间的系统进化关系证据不是十分充足 (Wendel et al., 2010), 但 13 个 D 基因组种与 6 个亚组相对应被精细地分成家系, 这与 Fryxell (1992) 长期按照形态已经得到的系统发生关系并不十分一致。已有多项证据表明拟似棉是美洲棉, 并作为亚组 *Selera* 的唯一代表种, 与地理上隔离的亚组 *Austroamericana* 的唯一一种雷蒙德氏棉的亲缘关系最近 (Cronn et al., 2003)。

(三) 基于叶绿体基因组的棉属系统发育分析

植物分子系统发育学研究主要基于植物的 3 个基因组, 即核基因组、线粒体基因组和叶绿体基因组。植物核基因组较为复杂, 单拷贝或低拷贝的基因筛选较为困难, 不易进行分析, 而且测序进度比较缓慢, 序列完全拼接困难, 目前只有较少物种的核基因组得到发布, 如水稻、高粱、黄瓜和拟南芥等, 缺乏数据基础, 使得基于此的大规模的植物系统发育分析相对滞后。植物线粒体基因组为 300~600kb, 但在不同的植物类群中变异很大, 而且进化速率较慢, 是叶绿体基因组进化速率的 1/4, 且线粒体基因组还存在分子内重组现象。植物叶绿体基因组作系统发育分析具有很多优点: 第一, 叶绿体基因组是单拷贝的, 且大多是母系遗传, 不存在自由组合和连锁交换现象, 独立进化; 第二, 叶绿体基因组结构相对简单, 相对保守, 不同物种叶绿体基因组之间具有一定的共线性; 第三, 叶绿体基因组长度仅为 120~160kb, 容易测序和分析; 第四, 叶绿体基因组进化速率适中, 包含大量遗传信息, 而且其中的编码区和非编码区由于进化速率的差异, 可以分别适用于不同分类阶元的系统发育研究 (Clegg et al., 1994)。

利用变异度高和进化速率快的序列区域, 建立系统发育树 (图 1-2), 3 种方法构建的 ML、MP 和 NJ 树拓扑结构基本一致, 支持率也较高, 较为可信。整个棉属分类与生

物地理学十分吻合，一支为澳洲棉种，包括 C 组和 G 组棉；另一支为美洲棉种，包括 D 组和 (AD) 组四倍体棉种；而亚洲-非洲棉种位于两支之间，其中 B 组进入澳洲棉种分支，而 A 组、F 组和 E 组进入美洲棉种分支。棉种聚合情况与传统分类基本一致，但 G 组的 G_1 却与 C 组较近，符合“双系祖先”观点 (Wendel and Cronn, 2003)，即比克氏棉以斯特提棉的祖先种为母本，以澳洲棉和奈尔逊氏棉的祖先种为父本，经种间杂交而成，进化后代中斯特提棉的核基因组被删除。

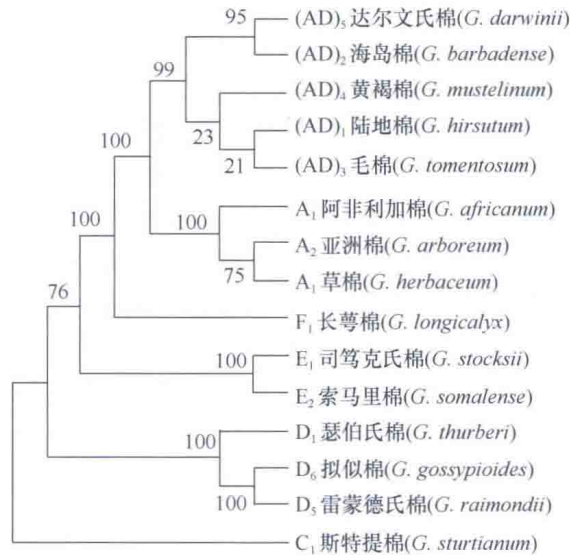


图 1-2 基于 10 个高变异叶绿体序列的 15 个棉种系统发育树

基于叶绿体基因组的系统发育分析，将棉种较为准确的聚合分类，与传统分类基本一致，各个染色体组自我聚合成一支。基因组内棉种的亲缘关系最近，二倍体棉种，A 组 (A_1 和 A_2)、B 组 (B_1 和 B_3)、C 组 (C_1 和 C_{1-n})、D 组 (D_1 、 D_5 和 D_6)、E 组 (E_1 和 E_2)、F 组 (F_1)、G 组 (G_2 和 G_3) 分别聚成一支；四倍体棉种大致分为 3 支 (Flagel and Wendel, 2010)，一支为 $(AD)_1$ 和 $(AD)_3$ ，一支为 $(AD)_2$ 和 $(AD)_5$ ，剩余的 $(AD)_4$ 自成一支。亲缘关系很近的四倍体棉种，叶绿体基因组相似度很高，所以 $(AD)_2$ 和 $(AD)_5$ 聚在一支可与其他三个棉种区别开来。

二、二倍体棉种的起源与进化

分属于 8 个基因组的二倍体棉种各染色体组间的演化关系是个很复杂的理论问题，有许多学者从不同方面进行过探讨。Phillips (1966) 根据二倍体棉种种间杂种 F_1 减数分裂过程中染色体的配对情况，提出分布于阿拉伯地区的 E 基因组可能是最原始的祖先种，因为 E 基因组和其他基因组杂交形成的杂种，其基因同源配对最低，出现最高频率的单价体；A、B 基因组则是由同一祖先分化而来的。因为 A、B 基因组的种间杂种平均单价体频率为 3；D 基因组与 C 基因组的分化在 A、B 基因组之前，因为 A、B 与 C、D 两个基因组的种间杂种单价体频率为 10~18。由于 D 基因组形成的杂种单价体频率又高于 C 基因组形成的杂种的单价体频率，所以 D 基因组应在 C 基因组之前先分化出来。按照

Phillips 的观点, A、B 基因组是由同一祖先进化而来的, 在细胞学上等同于两个基因组; 它们与 D、E 基因组杂交形成的杂交种单价体频率应该等同或相似, 但他的实验结果显示 A、B 基因组与 D、E 基因组形成的杂种与单价体频率却相差较大。说明仅靠杂种单价体频率并不能阐明所有二倍体种间的演化规律。

有人通过种子蛋白质电泳分析发现雷蒙德氏棉、裂片棉、松散棉、旱地棉、拟似棉 5 个 D 基因组种有一特异的条带, 而此带仅在 B 基因组内出现, E 基因组未出现, 随之把上述 5 种划分为 D_{β} 组, 其余为 D_{ϵ} 组。B 基因组与其他各组的平均相关系数最高 ($r=0.59$), 并考虑到 B 基因组的大面积的地理分布, 认为 B 基因组可能是最近的棉属祖先代表种。据此推断, D_{β} 组是由 B 基因组通过巴西至秘鲁, 最后到达墨西哥中部衍生而形成的; 而 D_{ϵ} 的形成则通过另一条途径, 即 B 基因组通过加勒比海至墨西哥衍生成的。E 基因组由 B 基因组向北衍生形成, A 基因组和 C 基因组则是由 B 基因组向南和向东南衍生而形成。这样便形成了现今二倍体棉种的地理分布格局。聂汝芝和李懋学 (1993) 通过 22 个二倍体棉种的核型分析也认为, 现有的二倍体棉种是 B 基因组最原始的祖先种所衍生的, 而 A、B、D、E 基因组则是由 B 基因组向不同方向扩展而分化形成的。根据表型性关系的原理, 估价了系统发育的性状, 发现 B 基因组是所有基因组中最具同质性的, 其组内种间平均相关系数最大, 而且与 A、E、D 基因组呈显著相关, 与其他基因组的平均相关系数最大, 因而基因组可看成是棉属其他二倍体的某种原型。但 Johnson 和 Thein (1970) 的蛋白质电泳发现, 南非作为 A 基因组的起源中心更合理, 因为它与 B 基因组的相关系数很低 ($r=0.15$), 且 A、B、 D_{β} 组在 3.5cm 处共有一条特异的深带, 此带除比克氏棉外, 其他基因组的棉种缺失此带。这样 A 基因组的起源问题就陷入悖论状态。

根据单 DNA 含量分析, 发现染色体最大核 DNA 含量最高的是分布于澳洲棉种的 C 基因组, DNA 含量最低的是分布于美洲棉种的 D 基因组, 据此提出 D 基因组是棉属中的原始类群, 其他基因组是由 D 基因组中的重复序列扩增而形成的; 然而仅根据 DNA 含量最少就认为是原始类群的观点似乎过于简单, 因为根据已进行过详细分析和比较研究的一些高等植物的属, 在属内二倍体种之间的进化趋势均伴随着 DNA 含量减少。此外, 把 D 基因组作为最原始的种群, 也与现今的观点相悖, 通常认为, 起源中心一般都是分化中心, 而现今美洲棉种仅分布一个 D 基因组, 而非洲棉种则有 A、B、D、E 和 F 共 5 个基因组。因此多数人认为非洲棉种是二倍体棉种的起源中心 (宋国立, 2007; Wendel et al., 2010)。

三、四倍体棉种的起源与进化

(一) 应用荧光原位杂交技术研究四倍体棉花的起源

目前荧光原位杂交 (FISH) 技术广泛应用于棉花基因组中单拷贝 DNA 序列、DNA 重复序列、多拷贝基因家族的染色体定位, 且可定位遗传图到特定的染色体或染色体臂上, 从而构建染色体的物理图谱, 确定遗传连锁图和物理图之间的关系, 对染色体和基因组的结构进行分析, 研究异源多倍体物种的进化机制等。

1. 棉花染色体核型分析

rDNA 的 FISH 信号是使用最多的染色体识别标记。45S rDNA 和 5S rDNA 位点产生杂

交信号, 杂交信号的有无、数目、位置和大小均可作为染色体识别的标记。用基因组 DNA 和 rDNA 作为探针, 定位除 K 基因组外的 A~G 和 AD 等 21 个棉种基因组的 45S rDNA, 发现随体或 45S rDNA 位点在棉种的分布有“主区域”的特点, 即集中于某些或某一序号染色体上, 如 D 基因组分布于以 5 号、9 号和 13 号染色体为中心的 3 个区域。9 个二倍体棉种定位了 5S rDNA, 其中 7 个棉种都与 45S rDNA 在同一染色体上, 推测棉属中两种 rDNA 可能有高度的共线性 (王坤波等, 1995; 王坤波, 2009)。

棉花上除了最常用的 rDNA 作为染色体标记外, 近年来随着棉花基因组学的快速发展, 还出现了以染色体着丝粒克隆和染色体特异的 BAC 克隆、端粒、基因组 DNA 等作为染色体标记, 进一步提高了棉花染色体核型分析的精细度。采用陆地棉 A 亚基因组的染色体特异的 22 个 BAC 克隆和 45S rDNA、5S rDNA 作为混合探针, 在 A 基因组亚洲棉有丝分裂中期染色体上进行 BAC-FISH, 将 45S rDNA 定位到了染色体 A5 和 A7 上 (Wang et al., 2008)。根据四倍体棉的同源性, 成功地将亚洲棉的 26 条染色体进行定位, 亚洲棉染色体被重新命名为 $A_1 \sim A_{13}$, 并进行了亚洲棉染色体的标准核型分析。陈丹 (2009) 将从 Pima90 的 BAC 文库中筛选出来的 D 亚基因组着丝粒 BAC 克隆 150-D-24 作为探针, 在 D 基因组瑟伯氏棉的粗线期染色体进行核型分析, 分析其核型公式为 $2n=2x=26=18m(SAT)+8sm$, 平均臂比为 1.48。凌键 (2008) 根据拟南芥端粒序列扩增到拟南芥类型的端粒序列, 对 7 个二倍体棉种进行核型分析。结果发现, G 基因组的比克氏棉、F 基因组的长萼棉和 A 基因组的草棉具有进化的核型。核型最原始的是 B 基因组的异常棉; 比较原始的核型有异常棉和 D 基因组雷蒙德氏棉, 它们的核型非常相似; E 基因组的索马里棉和 C 基因组的奈尔逊氏棉也是比较原始的核型。

2. 棉属四倍体及其供体种基因组 rDNA 的定位

核糖体 DNA (rDNA) 是研究植物基因组进化的非常有效的手段, 而 rDNA-FISH 则是 rDNA 定位的直观可视的方法。属内或种间 rDNA 位点的分析已经给分类学研究 (Bogunic et al., 2011)、种间进化、rDNA 本身的进化与功能研究提供了有力的证据。通过基于易位杂合子材料筛选的染色体特异 BAC 克隆, 系统地鉴定所有四倍体棉种、大部分 D 基因组二倍体野生棉种和 A 基因组二倍体棉种的 rDNA 的位点, 为 rDNA 在棉属分类、四倍体棉的供体种研究、二倍体和四倍体间的演化关系及 rDNA 位点的进化提供了有力的证据。

四倍体棉及其大多数二倍体供体棉种的 5S rDNA 位点的数目、在染色体上的位置及染色体定位都是高度保守的。这与豆科的落花生中 5S rDNA 位点结果是一致的。这些位点的保守分布可能是它们的近着丝粒区域分布的结果。因为小染色体的近中间区域很少与染色体结构重排有关, 且棉花染色体末端的重组率明显较高 (Wang et al., 2007)。

杂交信号的大小和强度差异可以反映串联重复单元拷贝数的差别。尽管棉花 rDNA 的拷贝数还没有确定, 但是 D 基因组棉种间或 A 基因组棉种间的 5S rDNA 位点基本一致的信号强度说明 D 基因组或 A 基因组 5S rDNA 有相近的拷贝数。5 个四倍体棉中, A 亚组或 D 亚基因组染色体上的 5S rDNA 位点大小和信号强度近似, 同样说明它们有相近的拷贝数。但是同一个四倍体棉中的 A 亚基因组和 D 亚基因组上的 5S rDNA 位点大小和信号强度不一致, 说明 A 亚基因组和 D 亚基因组间 5S rDNA 的拷贝数差异比较大。D 亚基因组染色体上的 5S rDNA 位点大小和信号强度明显比 D 基因组棉种的小, 说明其在