中国科学: 生命科学 2022 年 第 52 卷 第 9 期: 1407 ~ 1415
SCIENTIA SINICA Vitae lifecn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
进 展 中国知名大学及研究院所专栏 武汉大学生命科学学院专辑

武汉大学棉花功能基因组学研究的回溯与展望

王坤[1,2,3,4,5*], 朱玉贤[1,2,3,4*]

1. 武汉大学生命科学学院, 武汉 430072;
2. 武汉大学杂交水稻国家重点实验室, 武汉 430072;
3. 湖北洪山实验室, 武汉 430070;
4. 武汉大学高等研究院, 武汉 430072;
5. 武汉大学RNA研究所, 武汉 430072
*联系人, E-mail: zhuyx@whu.edu.cn; wangk05@whu.edu.cn

收稿日期: 2022-03-16; 接受日期: 2022-06-02; 网络版发表日期: 2022-08-29
国家自然科学基金(批准号: 31690090, 32070207)和湖北洪山实验室重点项目(批准号: 2021hszd014)资助

棉花是我国乃至全世界最重要的经济作物之一. 作为一种优良的天然纤维, 棉花纤维被广泛应用于纺织和其他工业, 占全世界纤维年均需求量的35%[1]. 全球有80多个国家种植棉花, 150多个国家从事棉花进出口贸易, 年均产值高达近5000亿美元. 我国的棉花年产量约600万吨, 占世界棉花产量的20%.

棉花实际上是棉属(*Gossypium*)植物的统称, 该属大部分植物的成熟种子被毛, 且富含纤维素, 棉花由此得名. 它们属于锦葵目(Malvales), 锦葵科(Malvaceae), 棉族(Gossypieae). 该属遗传资源丰富, 截至目前共发现有52个种, 包括45个二倍体($2n$=26)和7个四倍体($2n$=52)[2]. 棉属种间存在了巨大的基因组大小变异, 所有二倍体棉花物种都有相同的13条染色体, 然而其基因组大小却从D5基因组的~738 Mb到澳大利亚K基因组中的~2858 Mb不等分布[3]. 因此, 棉属是研究基因组大小进化的绝佳系统. 此外, 棉花纤维细胞是由棉花胚珠外被最外层的一部分表皮细胞突起伸长发育而来, 根据纤维细胞发育的不同时期, 可分为纤维起始、纤维伸长、次生壁发育和纤维成熟期等四个阶段, 因此其也是开展植物细胞分化、细胞伸长和细胞壁发育等调控机制研究的理想模式植物, 备受国际研究者青睐.

武汉大学棉花功能基因组研究团队一直围绕棉花生物学的基础科学问题开展研究工作, 在棉花基因组学、转录组学和棉花纤维发育调控机制等三个重要方向取得了较好的研究成果, 成为我国棉花生物学研究队伍中非常重要的一支. 以下从三个方面对该团队多年来取得的研究成绩进行梳理与总结, 并对武汉大学及全国、全世界棉花生物学研究的未来进行分析与展望.

1 棉花基因组学研究

20世纪50年代至21世纪的第一个10年, 国际棉花生物学研究先后经历了以植物形态与解剖学和植物生理学为主要研究方法的第一阶段, 以生化与分子生物学和遗传学为主要研究方法的第二阶段. 特别是从20世纪90年代开始, 伴随着经典的第一代测序技术的成熟和分子标记技术的发展, 分子遗传学技术被广泛地应用于棉花生物学研究. 基于分子遗传标记的图位克

引用格式: 王坤, 朱玉贤. 武汉大学棉花功能基因组学研究的回溯与展望. 中国科学: 生命科学, 2022, 52: 1407–1415
Wang K, Zhu Y X. Retrospectives and perspectives on cotton functional genomics studies at Wuhan University (in Chinese). Sci Sin Vitae, 2022, 52: 1407–1415, doi: 10.1360/SSV-2022-0037

© 2022 《中国科学》杂志社 www.scichina.com
https://engine.scichina.com/doi/10.1360/SSV-2022-0037

隆和分子辅助育种技术被广泛地应用于棉花农艺性状的遗传位点鉴定和基因克隆工作. 然而, 由于缺少完整的基因组信息, 研究者需要构建人工染色体文库(如细菌人工染色体文库BAC和酵母人工染色体文库YAC等), 并通过分子杂交技术筛选含有相关遗传标记或基因的BAC或YAC克隆, 最后通过一代测序获得候选基因序列. 上述方法费时费力, 严重地阻碍了棉花生物学研究的快速发展. 2000年左右, 随着高通量测序(第二代测序, NGS)技术的出现, 小型基因组植物拟南芥和水稻的全基因组序列得到测定组装, 这些高质量的物理图谱快速地促进了两种植物功能基因组学研究的发展.

21世纪初, 朱玉贤联合国内其他研究优势课题组, 包括喻树迅、王坤波、李付广等开展棉花基因组计划(Cotton Genome Project, CGP), 2012年8月完成并发表了全球第一个棉花基因组序列组装与分析工作[4]. 在这项工作中, 研究人员利用第二代测序技术对棉属二倍体雷蒙德氏棉(*G. raimodii*) D5(~800 Mb)基因组进行了累计78.7 Gb的深度测序(103.6倍的基因组覆盖度), 借助遗传连锁图将大于73%的序列锚定到13条假染色体上. 通过基因组保守区块分析, 发现雷蒙德氏棉基因组中的共2355个保守区块经历了大量的染色体重排; 另外还发现, 雷蒙德氏基因组在大约1.15~1.46亿年前和~1300万年前分别发生了一次双子叶植物共同祖先所经历的基因组六倍体化(paleohexaploidization)事件和一次棉属特异性的全基因组复制事件. 有意思的是, 该研究还通过对棉花与可可(*Theobroma cacao*)直系同源基因家族系统进化分析, 发现具有棉酚生物合成功能的CDN1基因家族的起源, 完美解释了棉属植物可以特异性生成棉酚等萜类化合物的分子机制.

2014年4月, 朱玉贤联合团队[5]又在国际上率先完成了二倍体亚洲棉(*G. arboreum*, A_2)全基因组图谱绘制工作. 通过对高纯系的亚洲棉品种-石系亚1号进行了全基因组测序, 获得193.6 Gb高质量序列(114.3倍基因组的覆盖度), 结合3.3万个BAC克隆的末端序列, 绘制出1694 Mb的亚洲棉基因组. 利用限制性酶切位点相关DNA测序构建的高分辨率的遗传图谱, 包含近2.4万个单核苷酸多态性(single nucleotide polymorphism, SNP)标记, 将1532 Mb(90.4%)的基因组组装序列锚定并挂载到13条染色体. 通过基因组注释分析, 发现亚洲棉(A2)含有与雷蒙德氏棉(D5)相近的蛋白质编码基因数目(约4.1万), 却富含高达68.5%的重复序列, 且其中95.12%是由长末端重复序列(Long Terminal Repeat, LTR)组成. 研究还通过基因组进化和比较基因组分析, 发现亚洲棉(A2)和雷蒙德氏棉(D5)在距今约500万年前由共同祖先分化而来, 在染色体水平上两个基因组呈现了高度的共线性. 相较于D5基因组, 亚洲棉的A2基因组在演化过程中发生了更大规模的转座子(transposable element, TE)插入, 这解释了A2基因组近乎2倍大于D5基因组的原因. 这一发现具有重要理论指导意义, 它第一次通过两个植物种的全基因组图谱序列比较揭示了植物基因组大小进化的一种重要方式. 此外, 该研究还第一次提出植物激素乙烯促进纤维伸长的阈值模型, 认为高浓度乙烯抑制纤维细胞的伸长, 而在低浓度范围则促进纤维伸长. 通过乙烯合成限速酶ACO基因的启动子序列比较分析揭示了雷蒙德氏棉、亚洲棉和陆地棉三者纤维发育差异的可能机制.

2015年4月, 在完成两个二倍体棉花, 亚洲棉(A_2基因组)和雷蒙德氏棉(D_5基因组)基因组图谱的基础之上, 联合团队又更进一步, 于业内领先完成了异源四倍体AD1基因组类型的陆地棉(*G. hirsutum*)遗传标准系TM-1基因组的测序、组装及注释分析工作[6]. 由于陆地棉可能是由二倍体A与D基因组棉花通过异源杂交加倍而成, 它不仅基因组庞大, 且包含大量的重复序列和高度相似的亚基因组同源序列. 这些因素都对高质量组装陆地棉基因组构成严峻挑战. 面对这一个艰巨任务, 团队综合运用了全基因组散弹枪法(测序基因组片段长度范围250~40 kb)获得445.7 Gb高质量序列数据(181倍的基因组覆盖度)、超过10万个细菌人工染色体BAC-BAC(bacterial artificial chromosome)末端测序、高密度SNP遗传图谱构建等策略, 成功地组装出2173 Mb的AD_1基因组. 借助39663个SNP标记的遗传连锁图, 将1923 Mb(88.5%)的序列锚定到26条染色体上. 该项研究首次从全基因组图谱水平揭示了异源四倍体陆地棉是由A基因组和D基因组两套染色体融合而成, 发现陆地棉基因组在异源多倍化之后出现大片段DNA的同源交换现象, 并初步通过比较基因组学方法揭示了植物祖先基因组到棉花基因组的演化路径, 进一步通过生信分析和分子实验证实了乙烯促进纤维伸长的阈值模型, 并提出乙烯调控纤维伸长的双向调节假说. 由于陆地棉是全球广泛种植(近95%)的栽培棉种, 解析它的基因组具有重大的社会与经济价值.

https://engine.scichina.com/doi/10.1360/SSV-2022-0037

该基因组图谱的绘制，极大地促进了棉花重要农艺性状分子机制的解析工作，对棉花新品种选育也具有重要指导意义，也为阐明棉属起源与演化，揭示植物多倍体物种形成机制打下良好基础.

随着第三代DNA测序技术的成熟，更长读长的测序为提升基因组质量提供了契机. 由于二代测序的技术限制，已组装完成的棉花基因组图谱仍需要填补与修正大量的空白和序列组装错误区. 2018年5月，朱玉贤的武汉大学团队联合中国农业科学院杜雄明利用PacBio单分子测序技术对亚洲棉A_2基因组进行了142.5 Gb(77.6倍的基因组覆盖度)的三代测序，并辅以Hi-C物理图谱(~125 Mb互作DNA对)新技术, 将基因组重叠群(Contig)数量从之前的4万多降低到近8000个, N50从先前的72 kb提升到1.1 Mb[7]. 此外, 基于新升级的亚洲棉基因组对230个亚洲棉自然群体和13个草棉(*G. herbaceum*)群体进行基因组重测序，获得包含17883108个SNPs和2470515个插入或者缺失(INDEL)的基因组变异图谱；利用SNP构建系统发育树和分析群体结构，发现草棉和亚洲棉具有明显的基因组序列特征差异；推断我国的亚洲棉是从南方辐射到黄河和长江流域; 最后结合全基因组关联(genome wide association study, GWAS)和数量性状位点(quantitative trait locus, QTL)定位技术鉴定了与种子油含量、抗病能力和棉纤维发育等重要性状相关的位点和基因.

2020年4月，朱玉贤的武汉大学团队[8]又进一步利用PacBio单分子测序技术和Hi-C物理图谱技术，结合最新的组装工具软件，首次高精度地解析了草棉(*G. herbaceum*)A_1基因组，并同时对亚洲棉A_2基因组和陆地棉AD_1基因组又进行了进一步的质量升级. 这些新基因组的质量得到显著性提升, 重叠群Contig的N50均达到Mb以上, 分别为1.9, 1.8和5.0 Mb. 更新后的A_2和AD_1基因组填补了先前近66%的缺口，成为真正意义上的染色体级别的高质量基因组. 重要的是，研究从谱系基因组学角度结束了长久以来困扰领域内的两个争议，即A_1和A_2基因组的演化关系以及四倍体A亚基因组供体. 研究通过揭示A_1和A_2基因组的序列和结构差异，结合群体遗传学数据，明确A_1起源于非洲的阿非利加棉(Aafricanum)，且A_1与A_2是独立起源演化，独立驯化; 通过构建系统发育树进行分析, 发现陆地棉A亚基因组真正的供体是非洲棉(A_1)和亚洲棉(A_2)的共同祖先A_0基因组棉，而并非之前一般认可的A_1或A_2起源学说. 该研究创新性地提出基于高斯概率密度方程的Gaussian probability density function (GPDF)算法, 精确计算棉花基因组中所有LTR类TE的插入数量和插入时间，发现陆地棉亚基因组、草棉和亚洲棉基因组都存在明显的多次大规模TE插入事件. 其中D和A亚基因组中2次TE插入事件的时间分别对应棉属产生的时间和四倍体形成的时间，A_1和A_2基因组中分别3次和2次TE插入事件对应棉属产生的时间以及草棉和亚洲棉的分化时间；通过GPDF精确计算发现陆地棉形成时间远早于草棉和亚洲棉分化时间，从而进一步论证了“四倍体A亚基因组供体来源于A_0基因组”这一重大发现[9].

在十多年的棉花基因组研究历程中，朱玉贤与合作者领衔中国棉花基因组研究大团队逐渐建立起了棉花基因组研究的理论框架，在雷蒙德氏棉基因组D_5、亚洲棉基因组A_2、草棉基因组A_1、陆地棉基因组AD_1等方面做出重要的技术和分析方法创新，取得了系列重要发现. 从2012年全球第一张棉花基因组草图公布之日起，在上述棉花基因组工作的引领带动下，截至目前全世界包括中国和美国等多家研究单位先后报道了不同基因组类型，不同种或品系的近50个棉种基因组[1](图1). 这些丰富的基因组数据信息为棉花群体遗传学和功能基因组学研究提供了宝贵的资源，极大地加快了优质高产棉纤维品种的分子育种改良工作.

2　棉花转录组学研究

为探究棉花纤维发育的分子机制，朱玉贤团队于20世纪90年代开始在棉花中开展分子生物学技术和方法的体系搭建工作. 通过改进常规RNA提取方法成功地利用陆地棉胚珠和纤维组织构建了高质量cDNA文库，大规模分离及鉴定棉纤维细胞快速伸长期特异表达基因[10]. 在此基础上，朱玉贤团队与当时的基因固相微阵列芯片企业合作，将此前构建的高质量cDNA文库转化为包含1.2个表达序列标签的EST，成功研制了国际上第一款棉花基因表达检测芯片[11]. 利用该基因芯片分析胚珠发育过程, 朱玉贤团队取得了有关“植物激素乙烯在纤维细胞伸长过程中发挥重要作用”的重要发现[11].

随着高通量测序技术的发展，通过二代和三代测序技术不仅可以更加灵敏高效地侦测基因的表达变

https://engine.scichina.com/doi/10.1360/SSV-2022-0037

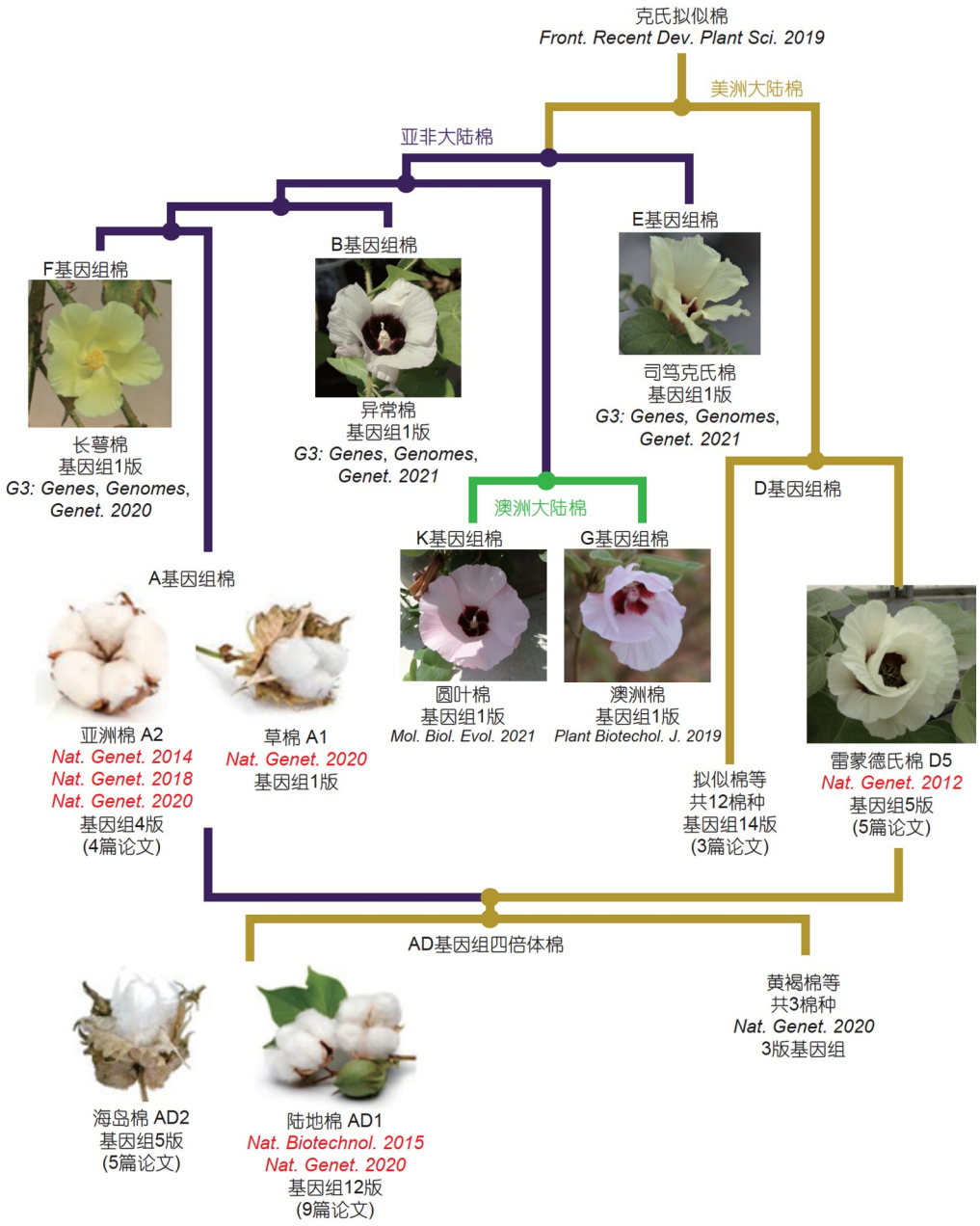


图 1 全球棉花基因组研究成果汇总图. 红色标记的研究成果由朱玉贤团队主持或主要参与(统计截至2022年2月)

Figure 1 Summary of global cotton genome studies. The research results marked in red are chaired or mainly participated by Zhu's team (Till Feb. 2022)

化，还可以探知基因的RNA剪接等转录变化. 朱玉贤团队敏锐地捕捉到了这一契机. 2014年，发表了棉花领域第一篇有关RNA剪接分析的研究报道[12]，将棉花分子生物学研究推向了更为精细的层次. 通过对雷蒙德氏棉进行深度的RNA测序，发现了10197个基因的16437个可变剪接事件. 通过比较分析发现TE插入导致的内含子保留(intron retension, IR)在植物中广泛存在. 研究还对TE的插入特点以及其可能与IR的关系进行了探讨，对动植物转录组在IR上的巨大差异进行了有意义的探索.

基于对亚洲棉基因组的高精度升级，朱玉贤的武汉大学团队[13]又系统性开展了对亚洲棉基因组转录

https://engine.scichina.com/doi/10.1360/SSV-2022-0037

全景的解析工作, 通过整合运用了4种高通量技术, 包括可获取全长转录本序列信息的三代Pacbio iso-seq测序、高深度定量表达和剪接信息的链特异性ssRNA-seq(strand-specific RNA-sequencing)以及可准确界定基因转录起始位点(transcription start site, TSS)和终止位点(transcription termination site, TTS)的5′RNA末端测序CAGE-seq(cap analysis of gene expression and deep sequencing)和3′RNA末端测序PolyA-seq(polyadenylation sequencing)等二代测序方法, 系统分析了亚洲棉16个不同组织器官的RNA转录情况. 通过编写IGIA软件系统整合包括Pacbio-seq 92.8 Gb, ssRNA-seq 455.1 Gb, CAGE-seq 158.2 Gb和PolyA-seq 339.2 Gb等高通量、高深度的测序数据, 对亚洲棉基因组的基因结构和转录进行了全方位的系统鉴定[13]. 该研究对亚洲棉基因组进行了系统精准的注释, 通过启动子单碱基分辨率等关键参数对比发现, 该研究对亚洲棉的基因组注释质量超过同期的拟南芥(*Arabidopsis thaliana*)和水稻(*Oryza sativa*)注释. 此外, 通过与已经发表的有关棉花重要农艺性状的遗传位点(GWAS和SNP等)的比对分析发现, 大量候选基因和遗传位点位于基因的非翻译区、RNA加工元件等区域, 进一步说明在棉花中开展精细转录分析的必要性.

随着高通量测序技术的发展, 已经完成多个棉花遗传材料的基因组测序与组装, 但缺乏对棉花基因组中非编码区域的深度调查. 这些区域包括小RNA(microRNA)和长链非编码RNA(long non-coding RNA, lncRNA)等非编码RNA. 武汉大学棉花团队在对转座子TE相关RNA转录的研究中发现, 棉花纤维发育相关的基因受到了附近TE的调控, 该结果表明棉花基因组中有大量非编码RNA, 其中部分是由TE演变而来[3]. 在随后的研究中, 通过对16个不同组织和5个非生物胁迫样本的多策略、高通量测序工作, 高质量鉴定了亚洲棉中9240个lncRNA, 其中有4405和4805个lncRNA转录本分别被CAGE-seq和PolyA-seq的RNA末端测序信号所支持[14]. 研究发现, lncRNA的转录和编码基因相似, 在不同组织和非生物胁迫条件下也频繁地呈现出转录起始位点TSS和转录终止位点TTS切换(TSS/TTS switch)的现象. 研究还发现, 大量的lncRNA以顺式作用的方式参与了邻近编码基因的转录调控, 有很多lncRNA可能参与邻近编码基因的TSS切换调节. 通过与之前GWAS数据的关联, 研究鉴定到了一个可能参与棉花纤维(短绒毛)发育的关键lncRNA; 此外, 研究团队利用在棉花中首次建立的染色质与RNA互作测序技术(chromatinIsolation by RNA purification sequencing, ChIRP-seq), 和病毒诱导的基因沉默VIGS实验验证了lnc-Ga13g0352的转录调控作用, 结果表明, 此lncRNA可以在基因组范围内通过激活或抑制不同靶标基因的转录而同时发挥完全相反的转录调控功能. 在对棉花miRNA的研究中, 他们发现了棉花所属锦葵科的特异性miRNA, 其可以通过调控锌离子转运蛋白编码基因的表达量, 从而控制胚珠发育过程的锌离子浓度[15]. 与此同时, 研究团队还关注棉花基因组中蛋白编码基因的非编码区, 如5′和3′UTR的功能. 他们通过整合转录组、翻译组和蛋白质组质谱数据, 在亚洲棉的基因组中鉴定出大量的小可读框(small open reading frames, sORFs), 即编码小于或等于100个氨基酸的短肽, 它们可能通过编码功能短肽或调控mRNA翻译而参与棉花的正常生长与发育过程[16]. 上述研究工作为开展棉花非编码基因的分子功能鉴定提供了极有价值的参考以及数据支持.

随着棉花生物学的快速发展, 不同类型和维度的测序数据爆发式增长, 不同数据间亟待有效整合利用. 研究团队也致力于大数据的互联互通, 构建了综合性的数据共享平台. 成功整合了棉花所在的锦葵科9个属共13个物种的374套多层次组学数据, 构建了首个锦葵科植物功能基因组学网站MaGenDB[17]. 该网站独立设计并开发出多种可视化组件与基因组在线比较工具, 内嵌基因组“浏览器”与“搜索引擎”, 实现数据的云端集成、快速检索和智能分析, 为开展对锦葵科特定生物学过程或基因家族的序列、结构、转录调控等的高效比较分析提供平台. 该云端集成检索极大地促进棉花研究者对现有棉花组学数据资源的利用, 有助于深入理解相关遗传变异、表型和基因之间的关系, 助力棉花的遗传育种和改良.

因此, 这一系列研究取得的重要基础发现对未来棉花功能基因组学以及基因功能的的进一步发现具有重要意义(图2). 中国农业科学院棉花研究所李付广研究员[18]在*Journal of Cotton Research*上发表专项述评, 认为“该系列工作揭示了植物中RNA转录调控的新机制, 也为棉花功能基因组学研究建立了一个极有价值的资源平台”, 认为该研究向彻底解构棉花转录组蓝图的方向迈进了一大步.

https://engine.scichina.com/doi/10.1360/SSV-2022-0037

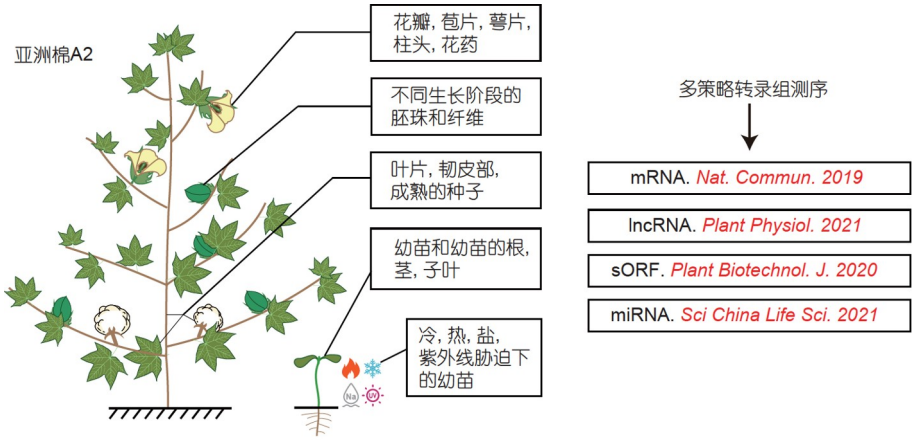


图 2 武汉大学棉花团队通过多组学策略研究亚洲棉A_2基因组转录调控

Figure 2 The studies of transcriptional regulation of Asian cotton A_2 genome via multi-omics strategy performed by the cotton team of Wuhan University

3 棉花纤维发育分子调控机制研究

纤维细胞的发育机制是棉花生物学研究的核心问题, 关乎棉花的产量和品质等重要农艺性状. 朱玉贤团队长期以来运用基因组学和转录组学等新兴技术开展基础研究, 并在此基础上开展棉纤维发育机制的研究, 为最终完成棉花纤维发育机制的完全拼图做出了重要贡献. 他们基于自己开发的世界第一款棉花基因芯片筛选到纤维伸长调控过程的重要基因*ACO*, 深入分析发现, 以*ACO*基因表达为限速步骤的乙烯合成途径在棉花纤维伸长的植物激素信号调控网络中起主导作用[11]. 通过比较蛋白质组学技术, 发现乙烯可以通过对促进植物细胞初生壁中果胶的生物合成, 主要是UGP, UGD, UER等酶活性调控, 来影响纤维等具有线性伸长机制的细胞伸长[19]. 进一步研究发现, 以*KCS*基因为关键调控步骤的超长链脂肪酸(very long chain fatty acid, VLCFA)合成途径可以调控乙烯的生物合成, 并控制棉花纤维的长度和强度[20,21]. 他们还发现*GhAPX1*参与调控纤维发育过程中的活性氧平衡, 而超氧离子可能通过与乙烯的交互作用调控棉纤维等植物细胞的伸长过程[22,23]. 进一步研究发现, VLCFA的合成和细胞壁纤维素合成共同受到植物激素赤霉素GA的正向调控[24]. 通过上述研究构建并确立“VLCFA-乙烯”是调控棉纤维伸长的核心组件, 勾勒出一个由乙烯为主导、多种分子参与的棉纤维伸长调控网络(图3), 为全面阐明棉纤维细胞伸长的分子机制提供了重要的理论基础. 该系列研究工作得到了国内外同行的广泛认可, 2014年, *McGraw-Hill Yearbook of Science & Technology*(《McGraw-Hill科学技术年鉴》)邀请朱玉贤撰写了题为“Regulation of cotton lint growth”的综述, 充分表明研究团队在世界棉花基础研究领域的重要地位.

在最新的研究中, 朱玉贤的武汉大学团队通过蓝绿非变性聚丙烯酰胺凝胶电泳(blue-native polyacrylamide gel electrophoresis, BN-PAGE), 证实在棉纤维次生壁形成期存在三十六聚体纤维素合酶复合体(cellulose synthase complex, CSC). 该三十六聚体的形成需要三个不同的纤维素合酶(cellulose synthase, CES)亚基GhCesA4, 7和8共同参与. 研究显示, 三个亚基的任意一个突变体株系的纤维细胞中仅可以检测到同源二聚体. 该研究第一次通过同步辐射光源表征技术(小角X射线散射和广角X射线散射线站)发现野生型陆地棉纤维次生壁中的纤维素微纤丝由72根葡聚糖链组成, 显著不同于其他组织或物种的18或24链. 该研究创制了世界上第一个没有次生壁沉积却大致能正常伸长的棉纤维细胞, 阐述了棉纤维及其他组织或物种中具有不同CES亚基数量的CSC与具有不同葡聚糖链数目的纤维素微纤丝合成的关系[25].

2011年, 朱玉贤团队[26]受邀在世界知名的三大植物科学综述刊物之一——*Current Opinion Plant Biology*撰写了题为“How cotton fibers elongate: a tale of linear cell-growth mode”的综述. 该文在世界上首次提出棉花纤维伸长可能是一种新的“线性细胞生长模型”(linear cell-growth mode), 综合当时的有关实验结

https://engine.scichina.com/doi/10.1360/SSV-2022-0037

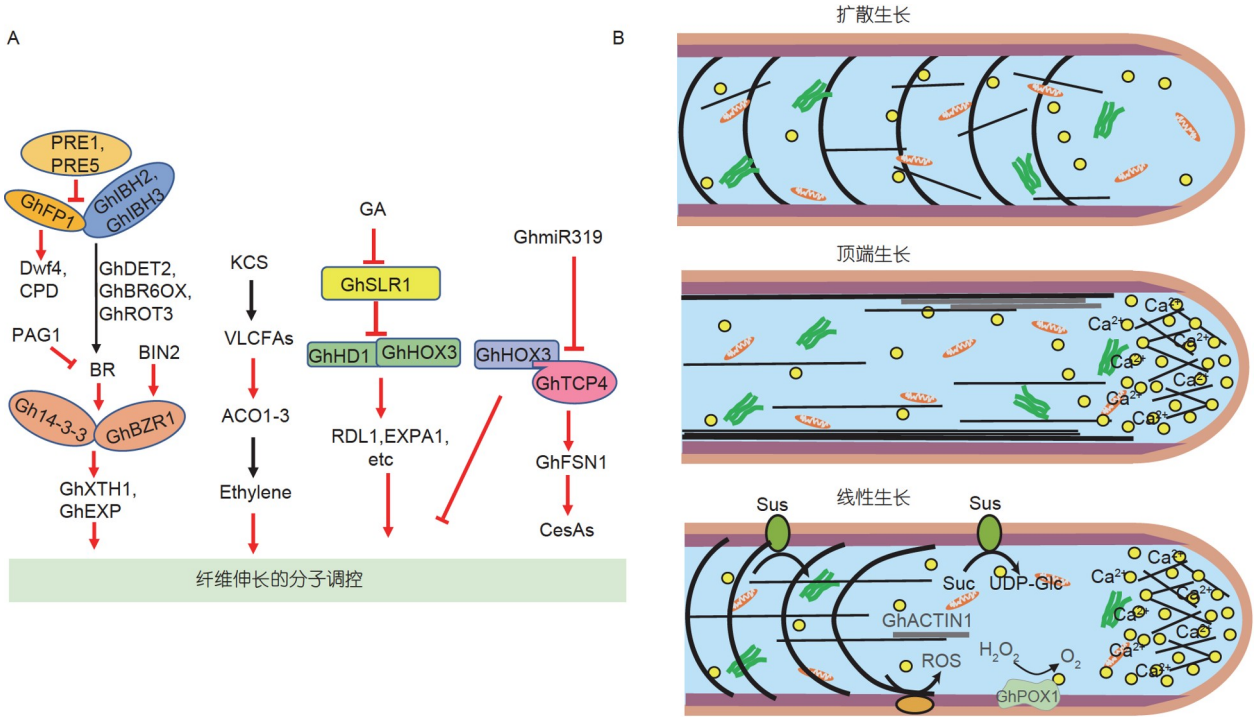


图 3　棉花纤维伸长调控的分子机制和模型. A: 棉纤维伸长调控的分子网络; B: 植物细胞伸长的三种模式[27], 其中棉花纤维的“线性伸长”模式由朱玉贤团队率先提出

Figure 3　Molecular mechanism and growth models of cotton fiber elongation regulation. A: Molecular network regulating cotton fiber elongation; B: three modes of plant cell elongation (adapted from ref. [27]), among which the “linear elongation” model was first proposed by Zhu’s team

果认为棉花纤维细胞伸长是一种结合了顶端生长(tip-growth)和扩散生长(diffuse-growth)两种模式特点的特殊伸长模式(图3). 这一大胆的学术猜想得到了后来的实验证据支持. 2019年, 研究者通过对棉花纤维的细胞骨架的荧光实时观测, 发现纤维细胞的伸长确实符合朱玉贤团队的猜想, 不同于经典的顶端或扩散生长方式, 是一种混合两种模式的特殊类型[27].

4　总结与展望

综上所述, 多年的研究历程中, 朱玉贤的棉花研究团队在棉花基因组学、转录组学和棉纤维发育的分子机制等三个主要方向取得了一系列重要发现, 这些研究成果在全世界棉花和植物学研究同行中产生了巨大反响, 为我国棉花生物学基础研究树立了典范. 目前, 武汉大学棉花研究团队已经成为国际棉花研究领域的一支重要力量. 2021年, 朱玉贤受邀在植物学经典年鉴*Annual Review of Plant Biology*发表题为“Recent advances and future perspectives in cotton research”的研究综述, 系统总结全球棉花生物学发展现状, 并对棉花生物学的研究未来进行展望.

当前, 我国棉花主产区出产的棉纤维整体品质仍然偏低. 2016~2018年, 我国高品质原棉年均短缺量近200万吨, 平均短缺率高达58.5%. 近五年国家每年需进口优质皮棉200多万吨, 也同时造成国产原棉的积压. 因此提升棉纤维品质仍然是我国棉花育种领域和棉花产业中的重要问题. 在未来的棉花生物学研究中, 武汉大学棉花研究团队将继续围绕国家战略需求, 并结合自身优势与特点, 有针对性的开展工作. 将研究力量集中投入到下面几个方向: (ⅰ) 棉属的谱系基因组学和泛基因组学. 构建棉属其他野生资源种的图形级染色体基因组, 并通过比较基因组和谱系基因组学分析, 厘清棉属的演化轨迹, 阐明植物基因组进化中的基础科学问题, 再通过泛基因组手段挖掘野生棉资源中纤维品质、抗性和适应性等优势性状的形成机制, 结合基因组编辑技术和分子育种技术, 结合传统育种技术快速改造并获得高产优质栽培棉花新品种. (ⅱ) 棉纤维细胞的发育调控机制研究. 运用单细胞生

https://engine.scichina.com/doi/10.1360/SSV-2022-0037

物学技术, 如单细胞转录组、蛋白组和代谢组等新技术在单细胞水平解析胚珠表皮细胞和纤维细胞的分化、伸长和细胞壁形成机制. (iii) 纤维素合酶机器的合成生物学. 探究应用交叉学科手段解析纤维素合酶复合体机器的运行机制, 为未来创造完全人工可控的纤维素体外发生器提供研究基础.

致谢 感谢武汉大学生命科学学院周一凡博士和胡晓博士在制图上的协助.

参考文献

1 Huang G, Huang J Q, Chen X Y, et al. Recent advances and future perspectives in cotton research. Annu Rev Plant Biol, 2021, 72: 437–462
2 Wang K, Wendel J F, Hua J. Designations for individual genomes and chromosomes in *Gossypium*. J Cotton Res, 2018, 1: 3–7
3 Wang K, Huang G, Zhu Y. Transposable elements play an important role during cotton genome evolution and fiber cell development. Sci China Life Sci, 2016, 59: 112–121
4 Wang K, Wang Z, Li F, et al. The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet, 2012, 44: 1098–1103
5 Li F, Fan G, Wang K, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet, 2014, 46: 567–572
6 Li F, Fan G, Lu C, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol, 2015, 33: 524–530
7 Du X, Huang G, He S, et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat Genet, 2018, 50: 796–802
8 Huang G, Wu Z, Percy R G, et al. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. Nat Genet, 2020, 52: 516–524
9 He P, Zhang Y, Xiao G. Origin of a subgenome and genome evolution of allotetraploid cotton species. Mol Plant, 2020, 13: 1238–1240
10 Ji S J, Lu Y C, Feng J X, et al. Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucl Acids Res, 2003, 31: 2534–2543
11 Shi Y H, Zhu S W, Mao X Z, et al. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell, 2006, 18: 651–664
12 Li Q, Xiao G, Zhu Y X. Single-nucleotide resolution mapping of the *Gossypium raimondii* transcriptome reveals a new mechanism for alternative splicing of introns. Mol Plant, 2014, 7: 829–840
13 Wang K, Wang D, Zheng X, et al. Multi-strategic RNA-seq analysis reveals a high-resolution transcriptional landscape in cotton. Nat Commun, 2019, 10: 4714
14 Zheng X, Chen Y, Zhou Y, et al. Full-length annotation with multistrategy RNA-seq uncovers transcriptional regulation of lncRNAs in cotton. Plant Physiol, 2021, 185: 179–195
15 Wen X, Huang G, Li C, et al. A Malvaceae-specific miRNA targeting the newly duplicated GaZIP1L to regulate Zn^{2+} ion transporter capacity in cotton ovules. Sci China Life Sci, 2021, 64: 339–351
16 Chen Y, Li D, Fan W, et al. PsORF: a database of small ORFs in plants. Plant Biotechnol J, 2020, 18: 2158–2160
17 Wang D, Fan W, Guo X, et al. MaGenDB: a functional genomics hub for Malvaceae plants. Nucl Acids Res, 2019, 48: D1076–D1084
18 Li F. Towards complete deconstruction of cotton transcriptional landscape. J Cotton Res, 2020, 3: 8
19 Pang C Y, Wang H, Pang Y, et al. Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and arabidopsis root hair elongation. Mol Cell Proteom, 2010, 9: 2019–2033
20 Qin Y M, Hu C Y, Pang Y, et al. Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. Plant Cell, 2007, 19: 3692–3704
21 Qin Y M, Pujol F M, Shi Y H, et al. Cloning and functional characterization of two cDNAs encoding NADPH-dependent 3-ketoacyl-CoA reductased from developing cotton fibers. Cell Res, 2005, 15: 465–473
22 Li H B, Qin Y M, Pang Y, et al. A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development. New Phytol, 2007, 175: 462–471
23 Mei W, Qin Y, Song W, et al. Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. J Genet Genom, 2009, 36: 141–150

https://engine.scichina.com/doi/10.1360/SSV-2022-0037

24 Xiao G H, Wang K, Huang G, et al. Genome-scale analysis of the cotton *KCS* gene family revealed a binary mode of action for gibberellin A regulated fiber growth. J Integr Plant Biol, 2016, 58: 577–589
25 Wen X, Zhai Y, Zhang L, et al. Molecular studies of cellulose synthase supercomplex from cotton fiber reveal its unique biochemical properties. Sci China Life Sci, 2022, doi: 10.1007/s11427-022-2083-9
26 Qin Y M, Zhu Y X. How cotton fibers elongate: a tale of linear cell-growth mode. Curr Opin Plant Biol, 2011, 14: 106–111
27 Yu Y, Wu S, Nowak J, et al. Live-cell imaging of the cytoskeleton in elongating cotton fibres. Nat Plants, 2019, 5: 498–504

Retrospectives and perspectives on cotton functional genomics studies at Wuhan University

WANG Kun[1,2,3,4,5] & ZHU YuXian[1,3,4]

1 College of Life Sciences, Wuhan University, Wuhan 430072, China;
2 State Key Laboratory of Hybrid Rice, Wuhan University, Wuhan 430072, China;
3 Hubei Hongshan Laboratory, Wuhan 430070, China;
4 Institute for Advanced Studies, Wuhan University, Wuhan 430072, China;
5 RNA institute, Wuhan University, Wuhan 430072, China

doi: 10.1360/SSV-2022-0037

https://engine.scichina.com/doi/10.1360/SSV-2022-0037