

棉花基因组与生物育种研究进展

张艳, 王省芬, 马峙英*

河北农业大学农学院, 河北省作物种质资源重点实验室, 作物种质资源研究与利用教育部重点实验室, 保定 071001

* 联系人, E-mail: mzhy@hebau.edu.cn

2026-04-17 收稿, 2026-06-05 修回, 2026-06-18 接受, 2026-06-22 网络版发表

国家自然科学基金(32372194)、河北省重大科技支撑计划(252N7501D)和河北省高等学校科学研究项目(PTZX2026014)资助

摘要 棉花是全球重要的经济作物, 其基因组复杂、多倍体特性显著, 给遗传研究与分子育种带来挑战. 近年来, 随着高通量测序、长读长组装、泛基因组学及基因编辑等技术的发展, 棉花基因组与生物育种研究取得了突破性进展. 本文系统综述了二倍体棉花基因组、陆地棉基因组、海岛棉基因组以及棉花泛基因组、群体遗传学和生物育种技术领域的最新研究进展. 二倍体棉花端粒到端粒(telomere-to-telomere, T2T)无缺口基因组的完成为解析多倍体棉花演化奠定了基础; 陆地棉从早期品种TM-1参考基因组到现代品种中棉所24(ZM24)、农大棉8号(NDM8)、中棉113(ZM113)及农大棉13号(NDM13)的T2T基因组组装, 标志着棉花基因组研究进入精准设计时代; 海岛棉基因组已完成从草图到高质量参考基因组再到泛基因组的跨越. 泛基因组研究揭示了棉花广泛的遗传变异、结构变异及其与重要农艺性状的关联; 群体遗传学通过规模化重测序和全基因组关联研究(genome-wide association study, GWAS)鉴定出大量与纤维品质、产量、抗病性相关的位点; 而分子标记辅助选择、全基因组选择、CRISPR/Cas9基因编辑及遗传转化技术的优化, 为棉花精准育种提供了高效工具. 本文旨在为棉花功能基因挖掘和分子设计育种提供系统的理论参考.

关键词 棉花, 基因组学, 群体遗传学, 生物育种

棉花(*Gossypium* spp.)是全球最重要的天然纤维作物, 也是食用油和蛋白质的重要来源, 在世界农业经济中占有举足轻重的地位. 我国位居世界棉花生产国和消费国的前列, 棉花产业可持续发展直接关系到纺织工业的原料安全和数以亿计棉农的生计. 生产上主栽品种为异源四倍体陆地棉和海岛棉, 其由A和D两个二倍体祖先基因组融合而成, 基因组庞大(约2.5 Gb)、重复序列占比高、亚基因组间存在广泛的不对称演化与结构变异. 异源四倍体棉花基因组研究是解析多倍体演化机制、挖掘优异基因资源及推动分子育种的核心基础, 有助于理解多倍性及棉属内基因组大小变异及其对农艺性状影响的重要性, 为突破栽培品种面临的高产、优质、多抗协同改良的育种瓶颈提供科学指导.

进入21世纪, 高通量测序技术的革命性突破极大地推动了棉花基因组学的发展. 从早期基于二代测序的基因组草图, 到当前采用PacBio HiFi、ONT超长读长、Hi-C等技术的端粒到端粒(telomere-to-telomere, T2T)完整基因组组装, 棉花参考基因组的连续性和完整性发生质的飞跃. 与此同时, 泛基因组学突破了单一参考基因组的局限, 系统捕获了种内和种间丰富的遗传变异; 群体遗传学借助大规模重测序和全基因组关联分析, 通量定位了控制纤维品质、产量、抗病性等关键农艺性状的遗传位点; 功能基因组学整合转录组、三维基因组、表观组等多维数据, 揭示了纤维发育和逆境响应的分子调控网络; 而以CRISPR/Cas9为代表的基因编辑技术和高效遗传转化体系的建立, 则使棉

引用格式: 张艳, 王省芬, 马峙英. 棉花基因组与生物育种研究进展. 科学通报, 2026, 71: 4723–4734

Zhang Y, Wang S, Ma Z. Research progress on cotton genome and biological breeding (in Chinese). Chin Sci Bull, 2026, 71: 4723–4734, doi: [10.1360/CSB-2026-0512](https://doi.org/10.1360/CSB-2026-0512)

花育种正在经历从“依赖自然变异”向“定向创造变异”的新阶段的转变;国产育种芯片的问世,为我国棉花绘制出一张重要性状及其关联基因的“高精度地图”,在基因体检、品种指纹图谱鉴定、基因定位、人工智能育种等方面发挥重要作用.本文系统梳理了近十年来棉花基因组及生物育种领域的重要研究进展,旨在为棉花功能基因挖掘和分子设计育种提供系统的理论参考与前沿资讯.

1 二倍体棉花基因组组装与升级

雷蒙德氏棉(*Gossypium raimondii*, D₅)被认为是异源四倍体棉花的D亚基因组供体,其基因组草图是解析四倍体棉花基因组的重要资源^[1,2],近期完成的雷蒙德氏棉基因组质量获得大幅提升,实现T2T无缺口基因组组装,并进一步揭示了棉花着丝粒的演化机制^[3].在A基因组研究方面,草棉(*G. herbaceum*, A₁)基因组的成功组装以及亚洲棉(*G. arboreum*, A₂)基因组质量的提升,提出了两个A基因组独立演化且均起源于已灭绝的A₀基因组的假设^[4].Wang等人^[5]和Sheng等人^[6]构建了二倍体棉花物种的泛基因组、泛三维基因组,鉴定出大量结构变异,其中第5与第13染色体之间的易位是澳大利亚谱系(C、G、K基因组)所特有.Cai等人^[7]结合PacBio SMRT长读长、Illumina短读长、BioNano光学图谱和Hi-C技术获得了澳洲棉*G. australe*高质量参考基因组.通过泛基因组比较,发现二倍体与四倍体棉花的纤维品质的平行选择依赖不同的遗传位点,且调控网络高度分化;揭示了转座子驱动的结构变异是基因表达调控分化和驯化性状形成的重要驱动力^[8].Peng等人^[9]利用多个代表性二倍体棉花物种作为系统发育外群估算了四倍体亚基因组的分化时间,为棉花多倍体演化提供了时间框架.此外,研究者还发现,A基因组(~1.75 Gb)约为D基因组(~0.88 Gb)的两倍其主要是由于长末端重复(long terminal repeat, LTR)反转座子在A基因组中的近期扩增所致;尽管D基因组小,却贡献了四倍体棉花纤维品质和抗病性的关键遗传资源^[9].综上,从2012年雷蒙德氏棉基因组草图(contig N50约22 kb)到2024年T2T完整组装(contig N50达67 Mb),其组装连续性提升近3000倍,且全部13条染色体实现端粒到端粒无缺口组装.类似地,亚洲棉基因组从早期版本(contig N50约150 kb)升级至当前T2T级别(contig N50>60 Mb).这些数据为四倍体亚基因组溯源和结构变异精准鉴定提供了坚实基础.

2 陆地棉基因组组装研究进展

基于雷蒙德氏棉(D₅)、亚洲棉(A₂)等二倍体祖先基因组测序,为四倍体AD₁基因组的亚基因组溯源与分化分析提供了关键参照,近年来,四倍体棉花基因组组装经历了三个里程碑阶段:2015~2017年间,基因组草图绘制,首次揭示了At和Dt亚基因组不对称性;2019~2021年间,高质量参考基因组(PacBio+Hi-C)使Contig N50跃升至Mb级别,完成多个现代品种基因组组装;2024~2026年间,T2T完整基因组的组装和泛基因组构建,实现了对结构变异、着丝粒、端粒的全景解析(表1)^[10~25].具体而言,2015年,陆地棉遗传标准系TM-1成为首个被测定的异源四倍体棉花,研究人员通过二代BAC-to-BAC、三代PacBio结合Hi-C等多种测序技术,先后完成了TM-1基因组的二代组装及高质量三代组装^[12,13,20,26],现代陆地棉品种的基因组组装是棉花分子育种走向精准化的重要基础,与TM-1等不同,现代品种基因组直接承载了育种选择留下的遗传印记^[27].面对多性状协同提升的生物技术育种的重大目标,解析抗病优质高产等多性状优良的陆地棉代表性品种对当前棉花育种改良意义重大.Ma等人^[18]率先完成综合性状优良的现代陆地棉主推品种NDM8高质量基因组组装,其基因组大小约2.29 Gb,contig N50达到13.15 Mb,染色体锚定率99.57%,共注释蛋白编码基因80124个;与TM-1比较,发现两个品种间存在76568个结构变异,其中28626个可在棉花自然群体中检测到,系统揭示了陆地棉栽培品种间的基因组结构变异全景.Yang等人^[15]完成了中棉所24的基因组组装,发现A08染色体上约30%区域存在大尺度结构变异并导致重组抑制.上述研究开启了陆地棉从“单一参考基因组”向“品种间比较”的研究升级,推进了棉花现代育种基因组学的发展.随着长读长测序技术的发展,Hu等人^[20]最近完成了早熟棉品种ZM113的T2T基因组组装,揭示了棉花基因组中着丝粒的进化动态以及一个与早熟性状相关的11 Mb倒位变异,这些发现为理解棉花基因组进化与短季棉适应性提供了新见解.针对我国中高端原棉长期依赖进口的现状,解析优质棉花品种的基因组对棉花品质生物育种意义重大.NDM13属于国标I型优质棉品种,同时具备抗病虫、耐盐碱、高产的优良特性,于2019年获得国家科技进步奖二等奖;河北农业大学等近期完成了NDM13的T2T基因组组装,完整解锁了包括着丝粒、端粒和45S rDNA在内的所有染色

表 1 代表性棉花基因组研究进展
Table 1 Advances in representative cotton genome research

物种/品种	倍性	发表年份	基因组(Mb)	Contig N50(Mb)	注释基因数	参考文献
雷蒙德氏棉(D ₅)	二倍体	2012	748	0.14	37505	[1]
		2012	744	0.04	40976	[2]
		2024	776	/	53167	[3]
亚洲棉(A ₂)		2014	1561	0.07	41330	[10]
		2018	1710	1.10	40960	[11]
TM-1	早期陆地棉	2015	2160	0.03	70478	[12]
		2015	2090	0.08	76943	[13]
		2019	2282	1.89	70199	[14]
		2019	2286	4.76	73624	[15]
		2020	2302	5.02	74350	[4]
		2020	2302	0.78	75376	[16]
		2025	2299	108.20	79642	[17]
ZM24	近现代陆地棉	2019	2309	1.98	73707	[15]
NDM8		2021	2291	13.15	80124	[18]
Jin668		2025	2370	109.28	74457	[19]
ZM113		2025	2299	89.27	78471	[20]
NDM13		2026	2294	108.32	80450	[21]
3-79	海岛棉	2015	2573	0.02	80876	[22]
3-79		2019	2223	2.15	71297	[14]
Hai7124		2019	2191	0.08	75071	[23]
Pima90		2021	2210	9.20	79613	[26]
Pima-S6		2023	2301	0.05	70838	[24]
Junhai-1		2025	2267	68.79	75729	[25]

体复杂区域^[21]。这些高质量基因组资源显著提升了棉花基因组学研究的精度，基于T2T参考基因组的GWAS能够精准定位性状变异，为全基因组选择、基因编辑等生物育种技术提供了可靠靶点。

3 海岛棉基因组组装与发展

海岛棉(*Gossypium barbadense*)是优质长绒棉的主要栽培种，其基因组为AADD型异源四倍体(2n=4x=52)，具有与栽培陆地棉类似的基因组特征。2015年，首个海岛棉(3-79)基因组草图成功组装，获得总长约2.57 Gb的序列，注释得到80876个蛋白编码基因，发现At亚基因组(1.50 Gb)与Dt亚基因组(0.85 Gb)大小显著不对称，进一步揭示了Gypsy类转座子的扩张是核心影响因子^[22]。2019~2024年，随着PacBio单分子实时测序和Hi-C技术的成熟，国内外团队相继完成Pima 90^[18]、新海21^[14]、H7124^[23]、Pima-S6^[24]等多个海岛棉品种的高质量参考基因组组装，contig N50提升至Mb级别。

2025年Meng等人^[25]基于PacBio HiFi测序技术，对涵盖野生种、农家种和栽培种的12份海岛棉进行从头组装，结合已发表数据构建了包含17份材料的海岛棉基因组图谱，单个基因组大小介于2.21 Gb~2.25 Gb之间，contig N50达到55 Mb~78 Mb，BUSCO评估基因组完整性超99.5%，该研究突破了单一参考基因组的局限，系统揭示了海岛棉群体的遗传变异特征，为精准挖掘优异等位基因和解析性状驯化机制奠定了基因组学基础。

4 陆地棉和海岛棉比较基因组学研究

陆地棉与海岛棉虽同为异源四倍体(AADD)，但在基因组大小、结构变异、品质相关基因拷贝数及育种选择特征上存在显著分化。在基因组大小与结构方面，早期海岛棉3-79草图(2.57 Gb)因重复序列注释原因而导致偏大^[22]，后来的高质量组装显示陆、海基因组实际大小相近(2.21 Gb~2.30 Gb)，而海岛棉略小^[14]；海岛棉的重复序列比例(约62%)略低于陆地棉(约65%)，可

能与转座子活性差异有关。两物种均呈现At亚基因组大于Dt亚基因组的非对称特征,而海岛棉的Dt亚基因组中与纤维强度相关的基因家族(如*FLAs*、*Sus*、纤维素合酶)拷贝数显著扩增,为其优质纤维的形成提供了基因组基础。在结构变异与重组抑制方面, Wang等人^[14]在TM-1与3-79间鉴定出14条染色体上存在大片段倒位; Hu等人^[20]在陆地棉ZM113中发现A08染色体上约11 Mb的着丝粒周边倒位,该变异与早熟性状密切相关,而在海岛棉中未发现该倒位的存在。比较基因组和染色体片段代换系群体研究发现海岛棉在纤维长度、强度相关位点上具有显著优势,而陆地棉在衣分、产量和适应性位点上富集更多有利等位基因^[14,22]。综上所述,陆地棉与海岛棉在基因组水平上的系统比较不仅明确了两者在产量与品质性状上的遗传基础差异,也为通过种间渐渗和基因编辑实现优异性状聚合提供了精确靶点。

5 棉花泛基因组研究迅速发展

单一参考基因组难以全面涵盖棉花种内及种间丰富的遗传变异,泛基因组作为整合群体全部或大部分遗传组成的新型研究框架在棉花中快速发展。采用“map-to-pan”策略,华中农业大学等先后对包含7个代表性二倍体野生棉、3个A基因组以及陆地棉群体重测序结果初步构建了棉属泛基因组,随后该团队构建了包含22个二倍体棉花物种及其近缘野生种的高质量超级泛基因组图谱,揭示了不同棉种遗传多样性形成的基因组演化特征及棉纤维品质性状演化的遗传机制^[5,28,29]。Yang等人^[30]基于107份从野生到栽培连续谱系的高质量陆地棉基因组组装,构建了更大规模的异源四倍体棉花泛基因组,解析了6个大型基因组结构变异(structural variation, SV),并揭示了海岛棉向陆地棉持续渐渗对栽培种遗传多样性的作用。Meng等人^[25]通过从头组装12个涵盖野生到驯化谱系的海岛棉基因组,并整合17个已发表的四倍体棉花基因组构建了图形泛基因组,揭示了海岛棉与陆地棉的分化进化路径及后续基因交流情况。重要的是,我国棉花经历百年的育种历史,在这场跨越百年的“基因竞赛”中,棉花品种不断改良,纤维产量翻了十倍,但科学家一直苦于看不清这张完整基因“图纸”。最近, Zhang等人^[14]从棉花育种发展角度出发,选择跨域100年育种历史、涵盖全球主要产棉国生产上代表性的28个陆地棉品种,整合PacBio HiFi、ONT超长读长、Hi-C等尖端测序技术,首次构

建了陆地棉T2T泛基因组,绘制出前所未有的高精度遗传“全景地图”,对着丝粒、端粒这些“基因荒漠”区域的变异及SD、CNV等复杂基因组变异清晰呈现。基于28个基因组鉴定出51551个高度保守的核心基因,同时发现近一半基因属于“可变基因”(仅存在于部分品种中),暗示棉花基因组的显著动态性。研究进一步刻画了SV在基因组中的热点区域,并证明SV可通过影响基因表达水平或导致基因存在/缺失变异,直接关联纤维品质、抗病性、耐逆性等育种性状^[14]。通过比较现代品种与早期骨干亲本(Deltapine15、徐州209),鉴定出一批品种特异性SV,从结构变异角度揭示了人工选择对基因组的塑造效应。该研究将泛基因组从“基因区覆盖”提升至“全基因组(含重复区)覆盖”,并率先在棉花中系统揭示了以往被视为“基因荒漠”的着丝粒、端粒区域富集了大量功能性SV。上述研究在基础科学之外,具有重要的应用价值,与传统的依赖大量田间试验和经验判断的棉花育种方法不同,基于T2T泛基因组的SV图谱使育种家能够直接定位控制优良性状的结构变异,从而加速新品种选育^[21]。然而,当前SV-性状关联主要基于群体遗传推断,多数热点SV尚未通过独立遗传背景或多环境表型验证其育种效应的稳定性;此外,从SV图谱到可直接用于育种的分子标记(如竞争性等位基因特异性PCR(kompetitive allele specific PCR, KASP)、芯片探针)仍存在不同程度的转化断层,这是后续研究需要重点解决的应用瓶颈。

6 棉花群体遗传学研究进展

群体遗传学通过大规模重测序和全基因组关联分析,系统解析了棉花种质资源的遗传多样性、群体结构及重要农艺性状的遗传基础。不同研究在二倍体与四倍体棉花中揭示了共性规律:第一,群体分化与适应性演化的驱动变异类型多样。例如,在二倍体亚洲棉中, Du等人^[11]对230份亚洲棉和13份非洲棉的重测序揭示了亚洲棉与非洲棉的群体分化及其在中国的迁移历史,并鉴定与抗病性、纤维发育等性状相关位点;在四倍体陆地棉中, He等人^[31]进一步发现, A06与A08染色体之间的倒位变异是栽培陆地棉生态型适应性分化的关键结构变异。第二,纤维品质与产量性状的遗传控制呈现多基因、多变异类型(单核苷酸多态性(single nucleotide polymorphism, SNP); SV与复杂变异并存)特征。Fang等人^[32]在陆地棉中率先鉴定到多个与纤维发育和产量相关的SNP位点。Ma等人^[33]在此基础上进一步拓

展至棉花多性状,利用419份陆地棉核心种质在黄河流域、长江流域及西北内陆棉区等多年累计12个环境下获得20余万的囊括纤维产量、品质、熟性、黄萎病抗性等13个性状指标的表型数据,揭示了陆地棉13个重要性状的遗传基础与性状优异位点的分布规律,首次揭示D亚组较A亚组集聚着更多的纤维品质优异位点,纠正了以往A亚组对纤维产量与品质起主导作用的传统认知;鉴定出7383个与13个性状显著关联的基因组SNP位点,发现D11染色体不同区域是控制纤维品质与黄萎病抗性的关键区段;以及D03染色体在控制棉花开花与熟性中的关键作用;该团队在此基础上分别以现代陆地棉NDM8和NDM13号为参考基因组以及T2T泛基因组变异图谱,以及在全球1600余份种质资源样本中鉴定出的76万余个SV和来自我国长江流域、黄河流域和西北内陆三大棉区22个环境(地点、年份)的产量、品质、抗病性表型鉴定数据,捕获了大量影响关键育种性状的SV;上述研究结合T2T泛基因组构建的SV图谱,对已有群体重测序数据进行SV重挖掘,有效提升了SV的检测能力和覆盖范围,更全面地评估了不同变异类型对育种性状的相对贡献。该团队公布的表型数据与鉴定到的重要性状位点与基因,也被不同棉花研究者用于全基因组关联分析与开展重要功能基因分子调控网络研究^[30,34,35]。

在海岛棉中,Yu等人^[36]对240份材料进行了重测序,鉴定到与纤维强度和衣分相关的候选基因;Zhao等人^[37]利用336份海岛棉重测序数据,鉴定出5个参与纤维长度、强度、衣分和枯萎病抗性相关的候选基因;Wang等人^[38]分析陆地棉遗传渐渗对海岛棉群体结构的影响,共鉴定到与纤维品质、成熟期等性状相关的315个渐渗区间;Li等人^[39]基于326份海岛棉种质的重测序数据,鉴定出143个显著关联的性状位点。上述成果有效推动了棉花功能基因组学研究,在四倍体栽培棉种重要性状关键基因功能与调控机制解析、分子标记开发与人工创造优异变异等方面发挥了重要作用。上述研究从SNP和SV两个层面揭示了棉花重要性状的遗传结构远比单一变异类型复杂。除SNP和SV外,基因组上还存在着倒位、易位等复杂结构变异,当前多数GWAS以SNP和SV为标记主体,基因组复杂结构变异的遗传效应相关研究还远远不够。未来结合T2T泛基因组构建的变异图谱结合基于变异断点开发的标记对群体进行基因分型,有望进一步推动倒位、易位等复杂变异对育种性状的贡献研究。

7 棉花生物育种研究进展

7.1 棉花分子标记辅助育种的发展

分子标记辅助育种在缩短育种周期、提升选择精准度方面具有显著优势。分子标记辅助育种的核心目标是将基因型与表型的关联转化为可高效选择的检测工具。简单重复序列(simple sequence repeat, SSR)、插入缺失(insertion-deletion, InDel)、SNP及单倍型标记等技术已在作物遗传育种中逐步得到应用。在棉花中,基于GWAS和QTL定位结果,研究者已从全基因组水平和候选基因水平两个路径开发了多种分子标记。在全基因组水平, Ma等人^[18]通过对1081份陆地棉种质的重测序,系统解析了InDel变异的全基因组分布特征,并鉴定出与黄萎病抗性和产量性状显著关联的关键InDel位点,为高通量芯片设计提供了基础。在候选基因水平,研究呈现两条技术主线:一是功能基因直接开发标记,如Li等人^[40]针对纤维长度和强度相关基因开发的KASP标记, Liu等人^[41]基于陆地棉与海岛棉同源基因序列差异开发了30余对纤维长度与强力有效选择的基因功能标记,可实现对海岛棉与陆地棉直系同源基因的有效区分以及纤维品质分子辅助选择;二是结构变异驱动标记开发,如Zhang等人^[42]揭示了RPL6基因启动子区基因组结构变异对海岛棉和陆地棉黄萎病抗性差异的影响,并据此开发了抗黄萎病功能标记,可对具有海岛棉抗病优异单倍型进行有效选择; Jia等人^[43]利用海岛棉GbSER02与陆地棉GhSER02间的功能位点开发KASP记,可精准筛选携带优良等位基因的种质材料。Cao等人^[44]开发了棉花衣分基因GhLPA1的KASP标记,可用于选择棉花高衣分优势单倍型。由上述研究不难发现,当前棉花分子标记开发已从早期的通用SSR标记转向基于功能变异(尤其是SV和单倍型)的精准标记,这是重要进步。但当前多数标记的开发依赖于单一群体或有限环境下的关联信号,其不同遗传背景中的选择效果缺乏大规模验证;而且,标记从“成果发表”到“育种实际应用”之间存在相对较长的滞后,未来应建立多环境、多遗传背景的标记效果评估体系,并将已验证的功能标记系统整合到育种芯片中,实现从“标记开发”到“育种决策”的高效畅通。

7.2 棉花基因编辑与遗传转化技术优化与创新

CRISPR/Cas9基因编辑技术可实现从“被动利用现有变异”向“主动创造定向变异”的转变, Zhang等人^[45]

和Lei等人^[46]构建了棉花皱缩叶病毒介导的CRISPR/Cas9基因编辑系统,通过病毒混合接种实现了多基因同时编辑,并建立了首个CLCrV介导的腺嘌呤碱基编辑系统,实现了GhPEBP基因的A-to-G精准转换且脱靶率极低.近年来我国棉花领域基因编辑研究发展迅速,以金双侠教授团队为代表,开发了*Cas9*、*Cas12a*、*Cas12b*、*CBE*、*ABE*、*ABE8e*、*dCas9-TV*、*TDE*、*TME*等基因编辑工具,能够在靶标位点进行敲除、敲入、敲高、敲低、SNP点突变等精准修饰,并利用这些工具创制了棉籽无棉酚、高油酸、耐草剂、抗虫、彩色棉花等优异棉花新种质,为棉花功能基因组研究和分子设计育种提供了高效工具箱和技术体系^[19,28,47-55].同时,遗传转化技术是实现基因编辑策略的核心执行环节.在棉花转化受体改良与驯化方面,金双侠团队创新了棉花遗传转化受体Jin668并揭示了其具备高转化效率与高再生能力特性的基因调控机制,实现了通过下胚轴诱导愈伤途径的高效快速遗传转化,Jin668被国内外多家科研单位引进利用;在遗传转化技术方面,Ge等人^[56]建立了一种基于茎尖分生细胞介导的转化系统,以棉花种子顶端分生组织干细胞为外植体,结合农杆菌和超声波处理,实现了跨越种间/种内、无基因型限制的高效遗传转化;Ribeiro等人^[57]开发了农杆菌与基因枪相结合的策略,在棉花中获得了60%的转化效率和5.5%的再生率;Zhang等人^[45]利用棉花卷叶病毒载体将sgRNA导入棉花细胞,成功实现了*GhAGL16*、*GhPDS*和*GhCLA1*等基因的敲除.花粉管通道法也被用于棉花遗传转化;Li等人^[58]改良了真空渗透辅助的农杆菌介导花粉转化法,将*acsA*与*acsB*基因整合至棉花基因组,显著提升了纤维长度、强度及纤维素含量.上述生物技术在棉花中的不断发展与创新在棉花基因功能研究与种质新创制方面发挥了重要作用.

7.3 棉花快速育种技术的建立与发展

快速育种(speeding breeding, SB)技术是一种缩短育种周期、加速作物改良的创新解决方案.育种家们面临着开发产量更高、营养更好、抗病抗虫、适应气候变化的新作物的严峻挑战.迫切需要缩短育种周期,加快育种进程,以满足全世界对食品和健康日益增长的需求,而快速育种已被认为是养活未来100亿人的关键技术之一^[59].在SB技术方面,河北农业大学早在20年前率先开展棉花快速育种研究,揭示了棉花幼胚发育规律,原创棉花幼胚快速成苗技术,通过研制幼胚快速

发育培养基并不断创新,从最初的一年2~3代不断迭代实现当前的一年5代,并进一步整合遗传育种学、作物栽培学、生物信息学、分子生物学及自动化高通量检测等学科技术,通过精准调控棉花生长所需的光、温、水、肥等环境因子,结合化学调控等辅助技术手段,结合育种芯片全基因组选择、基因编辑等现代设计育种技术,创建了一套高效、标准化、可推广的棉花快速育种技术体系^[60-62],为棉花珍稀资源胚挽救、转基因材料世代快速推进及基因快速聚合与性状靶向提升提供了高效工具,对提高作物育种计划中的遗传增益具有重要意义.

7.4 棉花全基因组选择技术的发展

随着测序技术的飞速发展,海量基因组数据的积累为在全基因组水平挖掘育种优异等位变异/基因,构建新型的育种模式,实现作物基因组精准设计育种奠定了重要基础.2022年,浙江大学以TM-1为参考基因组,基于近1000份陆地棉材料的变异位点和等位基因频率,开发了一套棉花40K液相芯片—浙大棉芯1号^[60-62],该芯片可用于定位关键QTL位点,发掘棉花重要功能基因.河北农业大学联合华智生物和新疆国欣种业有限公司,基于1081份来自全球主要产棉国跨越100年育种历史的陆地棉资源群体深度重测序和13个重要性状14个环境下的海量表型数据以及抗病优质高产农大棉8号高质量基因组,开发了国内首款以现代陆地棉为参考面向棉花基因组选择的自主知识产权高密度育种芯片“棉芯一号”^[60-62].该芯片基于对现代陆地棉群体重要性状关联位点的系统评价,筛选出在多环境中表现稳定且遗传效应显著的11159个高质量位点进行探针设计,实现了育种标记从“数量覆盖”到“功能精准”的跨越.该款芯片是目前性状覆盖最全面,涵盖高产、优质、抗病、早熟、抗虫、耐除草剂、耐低磷低氮及耐盐等13个重要性状,实现了从单一性状向多性状协同选择的技术升级;芯片整合了3981个性状功能位点和7178个反映核心种质遗传多样性的背景位点,兼顾了目标性状选择与遗传背景监控的双重需求.上述两款芯片的研发为棉花全基因组选择育种与早期优良性状鉴定提供了高通量、高精度、低成本的分子工具,可以大幅提高性状选择效率、缩短育种周期.

8 结论与展望

近十年来,棉花基因组与生物育种研究在三个维

度上实现了系统性跨越(图1)。首先,基因组资源从“单一个体草图”升级为“群体T2T泛基因组”。二倍体棉花T2T基因组的完成为多倍体演化研究奠定了基础;四倍体栽培棉花基因组从TM-1参考基因组发展到多个现代品种(NDM8、中棉113、NDM13等)的T2T组装^[18,20],并进一步构建了覆盖百年育种历史的T2T泛基因组,首次实现了对着丝粒、端粒等复杂区域的结构变异系统解析^[21]。其次,性状关联研究从“SNP-GWAS”深入到“SV-性状”的解析,群体遗传学与泛基因组的结合揭示了结构变异在纤维品质、产量、抗病性等性状中的重要作用,并以代表性材料为例评估了人工选择对SV的累积效应。第三,育种工具从“单标记辅助”拓展为“基因组设计”,CRISPR/Cas9基因编辑、无基因型限制的遗传转化、一年5代快速育种技术以及整合功能位点的全基因组选择芯片的建立,使棉花育种走向变异创造、精准设计、高效选择的发展阶段。

当前关于棉花基因组与生物育种仍存在若干不足。第一,技术成本与研究通量的矛盾。现阶段T2T级组装仍依赖多平台数据(HiFi+ONT+Hi-C),费时费力且单个品种成本较高,难以覆盖大规模育种群体;未来应发展单平台长读长T2T组装算法及代表性材料选择策略以降低成本和门槛。第二,SV效应的验证不足。多数SV-性状关联来自群体统计计算,缺乏独立遗传背景或多环境验证,建议建立SV近等基因系库或结合CRISPR编辑进行功能验证,准确评估其育种价值。第三,CRISPR技术虽功能强大但仍面临特异性、精准性、递送效率和安全性的挑战。未来可从借助于机器学习进行脱靶预测、开发智能递送载体、研发超大装载方案以及开发多元编辑器、规避染色体断裂等方面进行技术优化和升级。第四,从图谱到工具的转化仍存在不同程度断层。尽管T2T泛基因组鉴定出大量潜在功能性SV,但转化为育种可直接使用的标记(KASP、芯片探针)的比例较低,未来需发展自动化SV-to-marker设计流程,并推动

近十年来棉花基因组学与育种技术研究进程

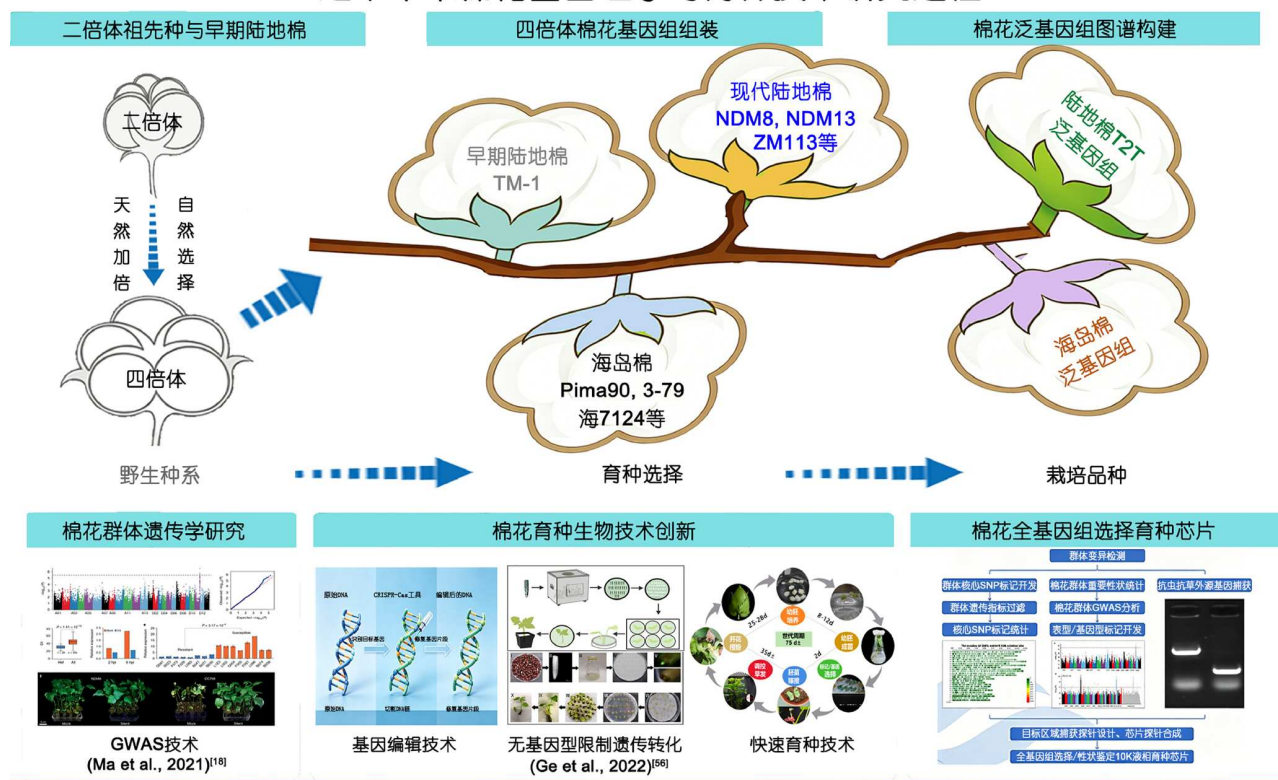


图1 近年来棉花基因组学与生物育种研究进展。图中“基因编辑技术”图片元素在作者的指导下借助人工智能工具豆包制作
Figure 1 Research progress in cotton genomics and biobreeding in recent years. The graphical elements of “gene editing technology” in the figure were created with the assistance of the artificial intelligence tool Doubao under the supervision of the author

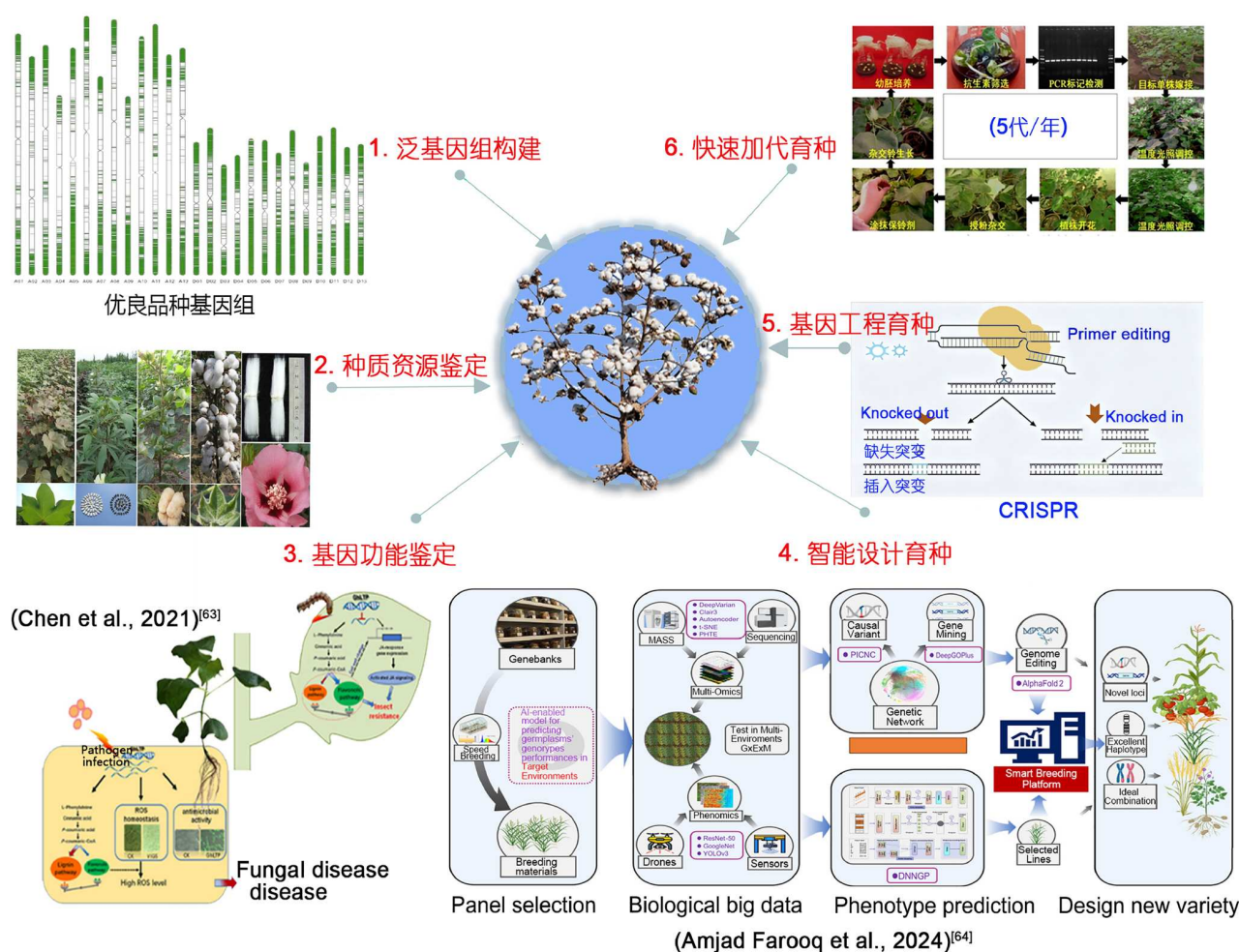


图 2 棉花未来育种展望示意图
Figure 2 Schematic illustration of future cotton breeding prospects

下一版本育种芯片的SV位点整合与芯片升级. 第五, 多性状协同改良的预测模型匮乏. 当前全基因组选择模型多以单一性状为目标, 难以处理性状间的相关性与权衡, 未来应结合机器学习与多性状基因组选择模型, 利用T2T泛基因组的SV信息, 实现高产、优质、抗病等多性状的协同优化.

综上所述, 棉花育种正从“经验驱动”向“数据驱动”转型. 近年来在基因组组装、变异图谱构建、功能基

因鉴定及群体遗传学研究方面取得的突破性进展, 为这一转型奠定了关键的底层数据基础. 未来, 随着单细胞组学、空间转录组技术的成熟以及合成生物学的发展, 棉花多组学资源将日趋完善. 下一步的关键在于降低成本、验证功能^[63]、打通转化、建立模型^[64], 这四个环节的系统实现, 将有效助推棉花从“基因组图谱”到“新品种培育”的育种闭环, 使品种改良完成从“经验选择”向“精准设计”的根本性转变(图2).

参考文献

- 1 Paterson A H, Wendel J F, Gundlach H, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*, 2012, 492: 423–427
- 2 Wang K, Wang Z, Li F, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*, 2012, 44: 1098–1103
- 3 Huang G, Bao Z, Feng L, et al. A telomere-to-telomere cotton genome assembly reveals centromere evolution and a Mutator transposon-linked

- module regulating embryo development. [Nat Genet](#), 2024, 56: 1953–1963
- 4 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
 - 5 Wang M, Li J, Qi Z, et al. Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*. [Nat Genet](#), 2022, 54: 1959–1971
 - 6 Sheng K, Sun Y, Liu M, et al. A reference-grade genome assembly for *Gossypium bickii* and insights into its genome evolution and formation of pigment glands and gossypol. [Plant Commun](#), 2023, 4: 100421
 - 7 Cai Y, Cai X, Wang Q, et al. Genome sequencing of the Australian wild diploid species *Gossypium australe* highlights disease resistance and delayed gland morphogenesis. [Plant Biotechnol J](#), 2019, 18: 814–828
 - 8 Li J, Liu Z, You C, et al. Convergence and divergence of diploid and tetraploid cotton genomes. [Nat Genet](#), 2024, 56: 2562–2573
 - 9 Peng R, Xu Y, Tian S, et al. Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons. [Proc Natl Acad Sci USA](#), 2022, 119: e2208496119
 - 10 Li F, Fan G, Wang K, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. [Nat Genet](#), 2014, 46: 567–572
 - 11 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
 - 12 Zhang T, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. [Nat Biotechnol](#), 2015, 33: 531–537
 - 13 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
 - 14 Wang M, Tu L, Yuan D, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. [Nat Genet](#), 2019, 51: 224–229
 - 15 Yang Z E, Ge X E, Yang Z R, et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. [Nat Commun](#), 2019, 10: 2089
 - 16 Chen Z J, Sreedasyam A, Ando A, et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. [Nat Genet](#), 2020, 52: 525–533
 - 17 Yan H, Han J, Jin S, et al. Post-polyploidization centromere evolution in cotton. [Nat Genet](#), 2025, 57: 1021–1030
 - 18 Ma Z, Zhang Y, Wu L, et al. High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. [Nat Genet](#), 2021, 53: 1385–1391
 - 19 Xu Z, Wang G, Zhu X, et al. Genome assembly of two allotetraploid cotton germplasms reveals mechanisms of somatic embryogenesis and enables precise genome editing. [Nat Genet](#), 2025, 57: 2028–2039
 - 20 Hu G, Wang Z, Tian Z, et al. A telomere-to-telomere genome assembly of cotton provides insights into centromere evolution and short-season adaptation. [Nat Genet](#), 2025, 57: 1031–1043
 - 21 Zhang Y, Sun Z, Tian S, et al. A pangenome reference and population studies link structural variants with breeding traits in *Gossypium hirsutum*. [Nat Genet](#), 2026, 58: 928–939
 - 22 Yuan D, Tang Z, Wang M, et al. The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. [Sci Rep](#), 2015, 5: 17662
 - 23 Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. [Nat Genet](#), 2019, 51: 739–748
 - 24 Chávez Montes R A, Ulloa M, Biniashvili T, et al. Assembly and annotation of the *Gossypium barbadense* L. ‘Pima-S6’ genome raise questions about the chromosome structure and gene content of *Gossypium barbadense* genomes. [BMC Genomics](#), 2023, 24: 11
 - 25 Meng Q, Xie P, Xu Z, et al. Pangenome analysis reveals yield- and fiber-related diversity and interspecific gene flow in *Gossypium barbadense* L. [Nat Commun](#), 2025, 16: 4995
 - 26 Wang M, Tu L, Lin M, et al. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. [Nat Genet](#), 2017, 49: 579–587
 - 27 Benevenuto J, Ferrão L F V, Amadeu R R, et al. How can a high-quality genome assembly help plant breeders? [GigaScience](#), 2019, 8: giz068
 - 28 Li J, Yuan D, Wang P, et al. Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. [Genome Biol](#), 2021, 22: 119
 - 29 Wang X, Lu H, Zhao Y, et al. A super pan-genome map provides genomic insights into evolution of diploid cotton species. [iMetaOmics](#), 2024, 1: e15
 - 30 Yang Z, Yang Z, Gao C, et al. Graph pan-genome illuminates evolutionary trajectories and agronomic trait architecture in allotetraploid cotton. [Nat Genet](#), 2026, 58: 218–229

- 31 He S, Sun G, Geng X, et al. The genomic basis of geographic differentiation and fiber improvement in cultivated cotton. *Nat Genet*, 2021, 53: 916–924
- 32 Fang L, Wang Q, Hu Y, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet*, 2017, 49: 1089–1098
- 33 Ma Z, He S, Wang X, et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat Genet*, 2018, 50: 803–813
- 34 Zhang X, Tian X, Gao X, et al. Integrated metabolomic and transcriptomic analyses identify MYB genes regulating key metabolites and agronomic traits in upland cotton *Gossypium hirsutum*. *Nat Genet*, 2025, 57: 2819–2830
- 35 Wang P, Li Z, Liu R, et al. Identification of genetic basis underlying mature cotton fiber length through metabolite-based genome-wide association analysis. *New Phytol*, 2025, 247: 2210–2230
- 36 Yu J, Hui Y, Chen J, et al. Whole-genome resequencing of 240 *Gossypium barbadense* accessions reveals genetic variation and genes associated with fiber strength and lint percentage. *Theor Appl Genet*, 2021, 134: 3249–3261
- 37 Zhao N, Wang W, Grover C E, et al. Genomic and GWAS analyses demonstrate phylogenomic relationships of *Gossypium barbadense* in China and selection for fibre length, lint percentage and *Fusarium wilt* resistance. *Plant Biotechnol J*, 2022, 20: 691–710
- 38 Wang P, Dong N, Wang M, et al. Introgression from *Gossypium hirsutum* is a driver for population divergence and genetic diversity in *Gossypium barbadense*. *Plant J*, 2022, 110: 764–780
- 39 Li H, Wang Z, Peng Z, et al. Integrative genomic structural variation analysis unveils genetic architecture underlying important traits in *Gossypium barbadense*. *Theor Appl Genet*, 2025, 138: 201
- 40 Li L, Sun Z, Zhang Y, et al. Development and utilization of functional competitive allele-specific PCR markers for key genes underpinning fiber length and strength in *Gossypium hirsutum* L. *Front Plant Sci*, 2022, 13: 853827
- 41 Liu Z, Sun Z, Ke H, et al. Transcriptome, ectopic expression and genetic population analysis identify candidate genes for fiber quality improvement in cotton. *Int J Mol Sci*, 2023, 24: 8293
- 42 Zhang D, Wang Y, Gu Q, et al. Cotton RLP6 interacts with NDR1/HIN6 to enhance *Verticillium wilt* resistance via altering ROS and SA. *Mol Plant Pathol*, 2025, 26: e70052
- 43 Jia H, Zhang D, Gu Q, et al. GbSER02 interacts with GhVOZ1 to promote fiber elongation by modulating gibberellin synthesis in cotton. *Adv Sci*, 2025, 12: e17578
- 44 Cao Y, Qiu Z, Wang Z, et al. Integrative BSR-seq and association analysis identify *GhLPA1* as a candidate gene associated with lint percentage in cotton. *J Integr Agr*, 2026, doi: 10.1016/j.jia.2026.03.004
- 45 Zhang J, Dai P, Weng Z, et al. Efficient CRISPR/Cas-based gene editing in cotton induced by cotton leaf crumple virus. *J Biotechnol*, 2026, 409: 33–43
- 46 Lei J, Li Y, Dai P, et al. Efficient virus-mediated genome editing in cotton using the CRISPR/Cas9 system. *Front Plant Sci*, 2022, 13: 1032799
- 47 Li J, Wang M, Li Y, et al. Multi-omics analyses reveal epigenomics basis for cotton somatic embryogenesis through successive regeneration acclimation process. *Plant Biotechnol J*, 2019, 17: 435–450
- 48 Qin L, Li J, Wang Q, et al. High-efficient and precise base editing of C·G to T·A in the allotetraploid cotton (*Gossypium hirsutum*) genome using a modified CRISPR/Cas9 system. *Plant Biotechnol J*, 2020, 18: 45–56
- 49 Li B, Liang S, Alariqi M, et al. The application of temperature sensitivity CRISPR/LbCpf1 (LbCas12a) mediated genome editing in allotetraploid cotton (*G. hirsutum*) and creation of nontransgenic, gossypol-free cotton. *Plant Biotechnol J*, 2020, 19: 221–223
- 50 Wang Q, Alariqi M, Wang F, et al. The application of a heat-inducible CRISPR/Cas12b (C2c1) genome editing system in tetraploid cotton (*G. hirsutum*) plants. *Plant Biotechnol J*, 2020, 18: 2436–2443
- 51 Wang G, Xu Z, Wang F, et al. Development of an efficient and precise adenine base editor (ABE) with expanded target range in allotetraploid cotton (*Gossypium hirsutum*). *BMC Biol*, 2022, 20: 45
- 52 Wang F, Liang S, Wang G, et al. CRISPR-Cas9-mediated construction of a cotton CDPK mutant library for identification of insect-resistance genes. *Plant Commun*, 2024, 5: 101047
- 53 Sun L, Alariqi M, Wang Y, et al. Construction of host plant insect-resistance mutant library by high-throughput CRISPR/Cas9 system and identification of A broad-spectrum insect resistance gene. *Adv Sci*, 2024, 11: e2306157
- 54 Hui F, Tang X, Li B, et al. Robust CRISPR/Mb2Cas12a genome editing tools in cotton plants. *iMeta*, 2024, 3: e209
- 55 Wang P, Zhang J, Sun L, et al. High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. *Plant Biotechnol J*, 2018, 16: 137–150
- 56 Ge X, Xu J, Yang Z, et al. Efficient genotype-independent cotton genetic transformation and genome editing. *J Integr Plant Biol*, 2023, 65: 907–917
- 57 Ribeiro T P, Lourenço-Tessutti I T, de Melo B P, et al. Improved cotton transformation protocol mediated by *Agrobacterium* and biolistic

- combined-methods. [Planta](#), 2021, 254: 20
- 58 Li X, Wang X D, Zhao X, et al. Improvement of cotton fiber quality by transforming the *acsA* and *acsB* genes into *Gossypium hirsutum* L. by means of vacuum infiltration. [Plant Cell Rep](#), 2004, 22: 691–697
- 59 Hickey L T, N. Hafeez A, Robinson H, et al. Breeding crops to feed 10 billion. [Nat Biotechnol](#), 2019, 37: 744–754
- 60 Wang X F, Wang Y X, Zhang G Y, et al. An integrated breeding technology for accelerating generation advancement and trait introgression in cotton. [Plant Breeding](#), 2011, 130: 569–573
- 61 Yan Z, Liqiang W, Zhikun L, et al. Targeted transfer of trait for *Verticillium wilt* resistance from *Gossypium barbadense* into *G. hirsutum* using SSR markers. [Plant Breeding](#), 2016, 135: 476–482
- 62 Wang G, Sun Z, Yang J, et al. The speed breeding technology of five generations per year in cotton. [Theor Appl Genet](#), 2025, 138: 79
- 63 Chen B, Zhang Y, Sun Z, et al. Tissue-specific expression of *GhnsLTPs* identified via GWAS sophisticatedly coordinates disease and insect resistance by regulating metabolic flux redirection in cotton. [Plant J](#), 2021, 107: 831–846
- 64 Farooq M A, Gao S, Hassan M A, et al. Artificial intelligence in plant breeding. [Trends Genet](#), 2024, 40: 891–908

Summary for “棉花基因组与生物育种研究进展”

Research progress on cotton genome and biological breeding

Yan Zhang, Shengfen Wang & Zhiying Ma^{*}

College of Agronomy, Hebei Agricultural University, Hebei Key Laboratory of Crop Germplasm Resources, North China Key Laboratory for Crop Germplasm Resources of Ministry of Education, Baoding 071001, China

^{*} Corresponding author, E-mail: mzhy@hebau.edu.cn

Cotton (*Gossypium* spp.) is a globally important cash crop for its natural fibers and high edible oil and protein content, yet its complex genome with polyploid nature has long posed challenges for genetic research and molecular breeding. In recent years, advances in high-throughput sequencing, long-read assembly, pan-genomics, and gene editing have driven breakthrough progress in cotton genomics and biotechnological breeding. This review systematically summarizes the latest research on diploid cotton genomes, upland cotton (*G. hirsutum*) genomes, sea-island cotton (*G. barbadense*) genomes, cotton pan-genomes, population genetics, and biotechnological breeding techniques. The completion of telomere-to-telomere (T2T) gap-free genomes for diploid cotton of *G. raimondii*, with contig N50 increasing nearly 3,000-fold from early drafts to 67 Mb, laid a foundation for deciphering polyploid evolution. In upland cotton, the transition from the early reference genome of the standard line TM-1 to T2T assemblies of modern cultivars such as ZM24, NDM8, Zhongmian 113, and NDM13 (contig N50 up to 108 Mb) marks the advent of precision design breeding. For sea-island cotton, genomes have progressed from draft assemblies to high-quality references and then to a 17-accession pan-genome, revealing interspecific gene flow and domestication trajectories. In upland cotton, pan-genomic studies have uncovered extensive structural variations (SVs); the 28-accession upland cotton T2T pan-genome identified 51,551 core genes and nearly half variable genes, with SV hotspots directly linked to fiber quality, disease resistance, and stress tolerance. Population genetics, through large-scale resequencing and GWAS, has identified thousands of loci associated with fiber quality, yield, and disease resistance, etc. This research exposed the discovery that the D subgenome harbors more fiber-quality superior alleles than the A subgenome, correcting traditional views. Moreover, the optimization of molecular marker-assisted selection, genomic selection, CRISPR/Cas9 gene editing (enabling targeted knockout, base editing, and creation of gossypol-free cotton), and genotype-independent transformation has provided efficient tools for precise breeding. Rapid breeding technologies, achieving five generations per year and high-density breeding chips (“Mianxin No. 1”) integrating 11,159 stable markers (including functional and background loci) across 13 traits, have further accelerated genetic gain. Despite these advances, challenges remain in reducing T2T assembly cost, validating SV effects via near-isogenic lines or CRISPR, and bridging the gap between pan-genome maps and breeder-friendly markers. Future integration of single-cell omics, spatial transcriptomics, synthetic biology, and machine learning will drive the transition from experience-driven to data-driven breeding, enabling collaborative improvement of high yield, superior quality, and multi-resistance in cotton. This review provides a systematic theoretical reference for functional gene discovery and molecular design breeding in cotton.

cotton, genomics, population genetics, biotechnological breeding

doi: [10.1360/CSB-2026-0512](https://doi.org/10.1360/CSB-2026-0512)