



棉花基因编辑工具开发及分子育种进展与展望

杨光琴¹, 王冠英¹, 余露¹, 张献龙¹, 聂新辉^{2*}, 李建^{2*}, 金双侠^{1*}

1. 湖北洪山实验室, 作物遗传改良全国重点实验室, 华中农业大学, 武汉 430070

2. 石河子大学南疆研究院, 图木舒克 843900

* 联系人, E-mail: jsx@mail.hzau.edu.cn; 17699436688@163.com; xjnxh2004130@126.com

2023-11-06 收稿, 2023-12-29 修回, 2024-01-03 接受, 2024-01-08 网络版发表

国家自然科学基金杰出青年科学基金(32325039)、国家自然科学基金面上项目(32272128)和湖北省洪山实验室重点项目(2021hszd013)资助

摘要 基因编辑技术能够在DNA及RNA水平上实现高效、精准修饰调控,是生命科学领域的革命性技术. 基于规律成簇间隔短回文重复序列及其相关蛋白(clustered regularly interspaced short palindromic repeats/CRISPR-associated protein, CRISPR/Cas)的基因编辑系统,凭借其简单、快速、高效的特点成为植物领域应用最广泛的基因编辑系统. 棉花是一种重要的经济作物,主要栽培种陆地棉(*G. hirsutum* L.)是一种异源四倍体植物物种,其基因组复杂,含有At和Dt两个亚基因组,基因拷贝数多,功能冗余现象普遍,因此基于常规的T-DNA插入及化学、物理诱变等技术创造突变体的方法在棉花中效率很低. 近年来,我国棉花科研工作者先后建立了10多套高效的基因编辑工具(包括基因敲除系统CRISPR/Cas9/12a/12b、碱基编辑器CBE/ABE/ABE8e/CABE、RNA编辑器Cas13a/b/c/d、转录激活系统dCas9-TV等),利用这些基因编辑工具能够对目标基因进行敲除、敲入、敲高、敲低及点突变,同时还建立了基于全基因组重测序评估基因组脱靶效应的技术体系. 基因编辑系统能否有效地传递到植物是基因编辑的重要步骤,受体材料的基因型在主要由农杆菌介导的基因转化中,对基因编辑系统的有效传递起着决定性的作用. Jin668是一种遗传转化效率高的棉花材料,是目前世界上广泛应用的受体材料. 本文详细介绍了获得高转化效率的Jin668的连续再生驯化(successive regeneration acclimation, SRA)策略,并对其再生机理进行了分析,未来利用Jin668中的遗传密码,有望打破基因型的限制从而进一步促进基因编辑工具的应用. 利用上述高效的遗传转化体系和基因编辑工具,科研工作者创造了包含棉花纤维品质、抗逆(病、虫、高温、盐碱)、棉籽品质、单倍体诱导等性状的5000多个基因的突变体库,有效地推动了棉花的基础研究及分子育种的发展. 本文还讨论了基因编辑系统本身的局限性,以及基因编辑系统在传递过程中存在的问题及解决方案. 最后,对棉花基因编辑研究的未来发展方向进行了展望. 育种的4.0时代,即分子设计育种的阶段,是其发展趋势. 对于棉花,“精确基因组+高效基因编辑策略”有望推动未来棉花高效分子设计育种体系的建立.

关键词 棉花, 基因编辑, CRISPR/Cas9/12/13, 碱基编辑器, 转录激活, 分子育种

基因编辑(genome editing)指的是利用基因编辑技术对生物基因组的特定目标基因进行比较精准的定点插入、突变或敲除的过程,从而达到定点修饰或改变目标基因的目的. 近年,新一代基因编辑技术随着一系列人工核酸内切酶的出现逐渐发展起来,如: 人工锌指

核酸酶(zinc finger endonuclease, ZFNs)、类转录激活因子效应物核酸酶(transcription activator-like effector nuclease, TALENs)、规律成簇间隔短回文重复序列及其相关蛋白(clustered regularly interspaced short palindromic repeats/CRISPR-associated, CRISPR/Cas)等.

引用格式: 杨光琴, 王冠英, 余露, 等. 棉花基因编辑工具开发及分子育种进展与展望. 科学通报, 2025, 70: 2495–2508

Yang G Q, Wang G Y, Yu L, et al. Progress and prospect of genome editing tools development and molecular breeding in cotton (in Chinese). Chin Sci Bull, 2025, 70: 2495–2508, doi: 10.1360/TB-2023-1138

ZFNs和TALENs分别是第一和第二代人工核酸内切酶，都是由结合蛋白与核酸内切酶*Fok I*融合而来，可以在基因组的特定位置制造DNA双链断裂^[1]。早期被广泛应用于人类、动物和植物细胞的基因编辑，但由于ZFNs存在较高的脱靶效应、ZFNs和TALENs系统的载体构建耗时费力等原因(表1)，阻碍了其广泛利用^[2]。CRISPR/Cas作为第三代人工核酸内切酶，来自对古细菌、细菌天然免疫系统的改造，该系统是细菌、古细菌抵抗噬菌体DNA增殖和质粒转移的天然防御机制。CRISPR/Cas9系统是第三代人工核酸内切酶的典型代表，该系统由反式激活crRNA(TracrRNA)、crRNA和Cas9蛋白组成，crRNA-tracrRNA结构指引Cas9核酸内切酶识别并结合在目的基因片段上，对这个在PAM(proto spacer adjacent motif, 原间隔序列临近基序；特征为NGG，这里N可为任意碱基)位点前携带20 nt的与crRNA序列互补的外源DNA进行切割，使得入侵噬菌体DNA或外源DNA双链发生断裂^[3,4](图1(a))。与第一、二代人工核酸内切酶相比，CRISPR/Cas9系统能够实现目标基因的定点敲除、定点敲入，也可同时对多位点进行突变，具有成本低、制作简单、突变效率高等特点(表1)^[5]，因此成为当前应用最广泛的基因编辑系统^[6-8]。

以上基因编辑工具通过切割目标DNA的两条链，造成DNA的双链断裂(double strands break, DSB)，可诱导DNA损伤后的修复机制，即同源重组(homologous recombination, HR)和非同源末端连接(non-homologous end joining, NHEJ)修复，而细胞中NHEJ修复占据绝对主导地位，在NHEJ修复过程中往往会发生高频率的修复错误(如碱基的缺失或插入)，使得编辑个体的目标基因的编码框发生移码突变^[9]。

Cas12核酸酶是植物中应用第二广泛的Cas蛋白，特别是Cas12a(Cpf1)已在许多植物物种中得到应

用^[10-12]。与Cas9识别NGG基序的PAM位点不同，Cas12a识别富含A/T的PAM，PAM序列位于靶点上游。Cas12a的编辑过程不需要tracrRNA的参与，precrRNA可以自我修饰为成熟的crRNA。Cpf1与成熟的crRNA结合形成二元复合物，然后该复合物与目标DNA结合形成三元复合物^[13]。因此，CRISPR/Cas12a系统的guide-RNA序列要比CRISPR/Cas9系统的sgRNA序列短一半左右，在多基因编辑应用中可能具有一定的优势(在T-DNA载体系统中可以串联更多的guide-RNA序列)。另一种Cas12蛋白是Cas12b(C2c1)。Cas12b系统需要crRNA和反激活crRNA(tracrRNA)的共同参与^[14-16]。这两种Cas12蛋白在目标位点均制造具有黏性末端的DNA双链断裂(图1(e)、(f))，这个突出的黏性末端可以作为同源重组修复的模板，实现基因的敲入^[17]。

碱基编辑(base editing)技术是基于CRISPR系统衍生的一种新兴的基因靶向修饰技术。它通常是DNA切割活性丧失但保留DNA结合活性的Cas9突变体(nCas9, nickase Cas9; 或dCas9, dead Cas9)和人工进化的胞嘧啶脱氨酶或腺嘌呤脱氨酶，在guide RNA的引导下让nCas9或dCas9结合到目标序列特定位置，脱氨酶利用短的肽段和nCas9或dCas9交联在一起，可以对目标位点特定碱基进行精确的脱氨等化学反应，再在细胞碱基错配修复机制的作用下，使C•G转化为T•A或A•T转化为G•C^[18-20](图1(b)、(c))。哺乳动物细胞中开发了一系列胞苷碱基编辑器(cytosine base editors, CBEs)BE1、BE2和BE3，最常用的是第三代碱基编辑器BE3^[19]。腺嘌呤碱基编辑器(adenine base editors, ABEs)也已经开发出来，这些新系统已在植物中得到应用^[21-24]。

定向调控内源基因以实现解析目标基因表达网络一直是植物研究领域的热点。CRISPR激活(CRISPR activation, CRISPRa)基于dCas蛋白与转录激活因子结构

表1 ZFN、TALEN和CRISPR/Cas三种基因编辑系统的比较

Table 1 Comparison of three gene editing systems: ZFN, TALEN and CRISPR/Cas

比较项目	ZFN	TALEN	CRISPR/Cas
DNA识别模式	蛋白-DNA	蛋白-DNA	RNA-DNA
DNA结合元件	Zinc finger motif	TALE repeat	gRNA
切割元件	<i>Fok I</i> 二聚体	<i>Fok I</i> 二聚体	Cas-RNA复合物
载体构建	较简单	难	简单
特异性	低	高	高
脱靶效应	高	较低	低

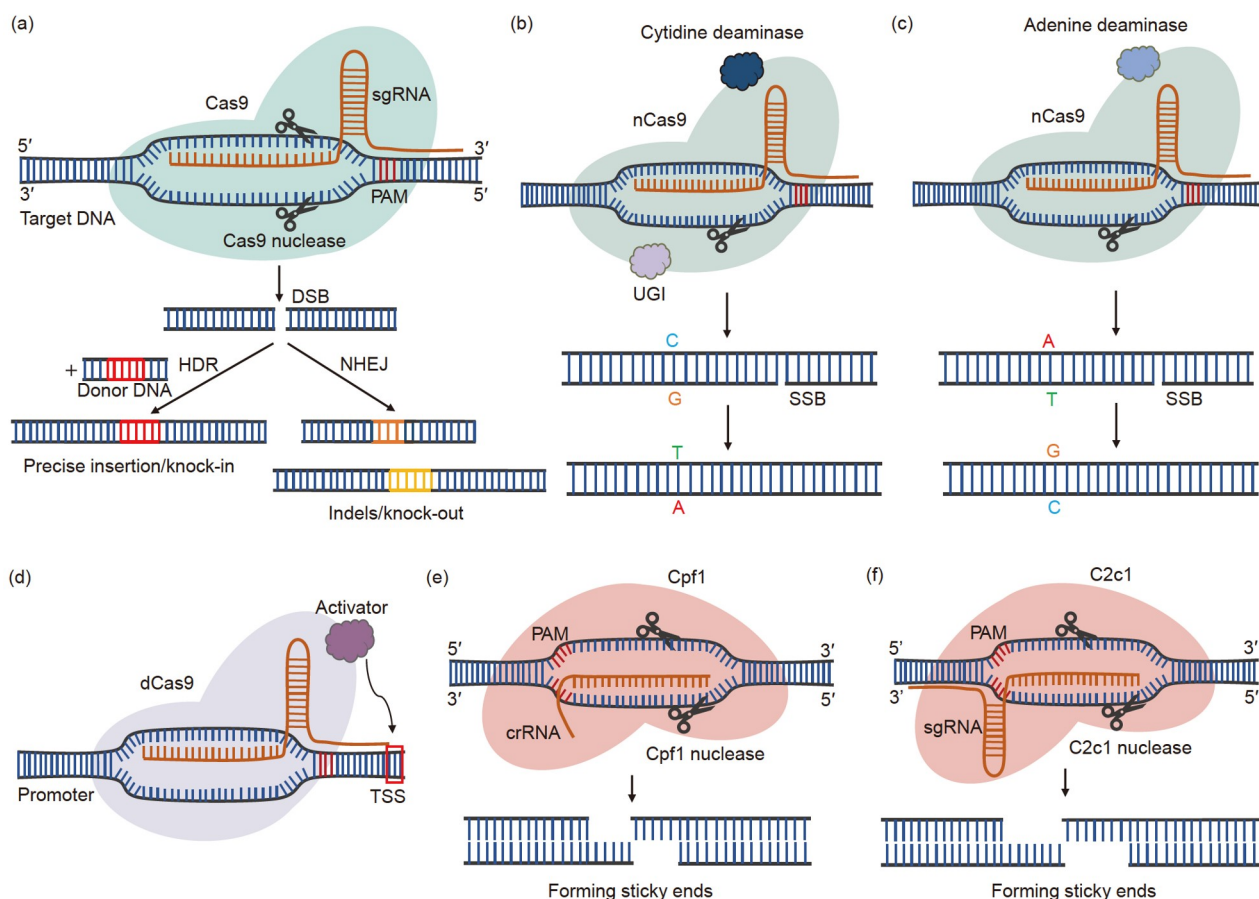


图 1 基于CRISPR/Cas的不同基因编辑系统的工作原理。(a) CRISPR/Cas9. 碱基编辑器: (b) 胞嘧啶碱基编辑器CBE; (c) 腺嘌呤碱基编辑器ABE. (d) 基于dCas9的转录激活系统: dCas9-TV. (e) CRISPR/Cpf1(Cas12a). (f) CRISPR/C2c1(Cas12b). DSB, 双链断裂; SSB, 单链断裂; nCas9, Cas9切口酶; dCas9, 核酸酶缺陷型Cas9; UGI, 尿嘧啶糖基化酶抑制剂; cytidine deaminase, 胞嘧啶脱氨酶; adenine deaminase, 腺嘌呤脱氨酶; TSS, 转录起始位点

Figure 1 Principles of different CRISPR/Cas based gene editing systems. (a) CRISPR/Cas9. Base editors: (b) cytosine base editor (CBE); (c) adenine base editor (ABE). (d) Transcriptional activation system based on dCas9. (e) CRISPR/Cpf1(Cas12a). (f) CRISPR/C2c1(Cas12b). DSB, double strands break; SSB, single strands break; nCas9, Cas9 nickase; dCas9, dead Cas9; UGI, uracil glycosylase inhibitor; TSS, transcription start sites

域的耦合,在guide RNA的指导下在目标基因的启动子区域进行结合,进而激活下游的基因转录(图1(d)),为植物中传统的基因过表达方法提供了一种有效的替代方案^[25]。从理论上讲,由于其RNA引导的性质,CRISPRa可以实现基因组中任何靶基因的特异性激活。与传统以过表达技术来提高基因表达水平的方法相比,CRISPRa系统具有以下明显优势: (1) 原位激活目标基因的表达,可以研究转录因子与目标序列的调控关系; (2) 可以实现多基因同时激活,只需要一段guide-RNA序列(100 nt左右)即可对任意基因进行激活,理论上一个T-DNA元件上可以串联20多条guide RNA序列,因此可以实现对不同遗传位点的多个基因或者同一染色体的基因簇或整个代谢通路的基因同步激活; (3) 从目前的

研究数据来看,CRISPR激活目标基因的效率要高于过表达系统的效率^[26,27]。目前CRISPR激活系统已经在植物中被开发出来,该系统将dCas9核酸酶融合到转录激活因子(如VP64)中,以调节内源基因的转录^[25,27,28]。

棉花是世界上最重要的经济作物之一,既是纺织品和可再生纤维的原料,也是食用油和蛋白质的来源^[29]。目前农业生产上主要种植的棉花是陆地棉,它是一种异源四倍体植物,含有At和Dt两个亚基因组,其基因组庞大而复杂,由于大部分基因存在多个同源拷贝容易形成基因的功能冗余,使得棉花基因功能研究存在极大挑战^[30,31]。随着棉花高质量基因组序列的公布^[31~34],高效遗传转化和基因编辑技术体系在棉花中相继应用起来,这些技术为四倍体棉花的基因功能研

究和优质新种质的创制提供了有力的工具。

1 基因编辑工具递送系统的研究进展

绝大多数转基因及基因编辑作物是通过农杆菌介导的遗传转化及植株再生(器官发生、体细胞胚胎发生等)方式获得的,这个过程非常耗时且高度依赖于植物的基因型和外植体类型。因此,转化和再生成为植物基因编辑系统能否成功应用的关键步骤。这个技术环节通常被称为基因编辑工具的递送系统,也是在许多植物中能否有效进行转基因和基因编辑工作遇到的重要瓶颈。

1.1 棉花高效转化受体材料Jin668的发现与再生机制解析

目前国际上只有珂字棉系列(Corker201/310/312)、泗棉3号和中12等少数棉花基因型能够通过体细胞胚胎发生途径获得再生植株,但这些材料的农艺性状普遍不佳、再生转化效率偏低及转化周期长等问题严重制约了棉花生物技术的发展。华中农业大学棉花遗传改良团队从20世纪90年开始,一直致力于棉花遗传转化体系的优化和改良。先后对数百个陆地棉材料进行了再生能力的比较,筛选到一个优良的受体材料YZ-1^[35]。YZ-1作为受体材料的转化效率和再生周期较以往的棉花受体材料都有较大提升,其转化效率提高到90%以上,是国际上棉花遗传转化模式材料-珂字棉系列转化效率的5倍左右,再生周期由18个月缩短到6~8个月左右,YZ-1目前已经成为国际上最重要的棉花转化受体之一^[35]。但YZ-1存在综合农艺性状差且不耐高温等不足,因此亟待培育新的、优良的棉花转化受体材料来代替YZ-1。此后,该团队通过以一个2000年以后选育的、农艺性状较好的棉花主栽品种Y668(豫668)为研究对象,经过连续再生驯化(successive regeneration acclimation, SRA)提高棉花转化效率的策略(通过获取经过体细胞胚胎发生过程获得的再生植株的种子,作为下一次组织培养的受体材料,连续重复这一过程多次),获得了遗传转化效率高达90%的材料,并命名为Jin668,其转化效率远高于母本材料Y668(30%)^[36,37]。目前全球超过60多家棉花科研单位利用Jin668成功进行了5000多个基因的遗传转化和基因编辑研究,是目前国际上应用最广泛的棉花转化受体材料,这项成果得到国际棉花基因组协会(International Cotton Genome Initiative, ICGI)/美国棉花公司(Cotton Incorporated)的

高度认可,2021年,该协会/公司授予这项成果的主要发明人金双侠教授棉花生物技术奖——Cotton Biotechnology Award。

此后,Jin668的高转化效率和高再生能力背后的基因调控机制成为该团队的重要研究方向,希望通过挖掘出一些重要的再生相关基因,解决其他不能再生的棉花品种再生的问题。不同的细胞类型,可重编程的能力不同,因此,在单细胞水平上准确分析体细胞重编程的过程,对于提高对植物再生调控机制的解析,挖掘决定植物再生能力的关键“密码”,为未来打破棉花再生的基因型限制,具有重要的理论和实践价值。近期,华中农业大学棉花遗传改良团队用Jin668和不可再生材料的陆地棉遗传标准系TM-1为实验对象,结合单细胞测序技术在单细胞层面解析棉花的体细胞胚胎发生机制。通过对不同再生能力材料的比较,挖掘到决定体细胞胚胎发生能力的关键基因并构建了棉花体细胞胚胎发生的基因调控网络。Jin668和TM-1的细胞类型具有高度的一致性,与生长素、细胞分裂素和伤口诱导相关的基因在初生维管组织细胞类型(包括初生木质部、薄壁组织和形成层细胞)中表达。因此推测,分化程度较低的细胞类型对激素的诱导更加敏感,是愈伤组织形成的主要细胞类型。同时发现,在初生维管组织细胞中鉴定到许多和再生相关的差异基因,在这两个品种中具有不同的表达模式。很多体细胞胚胎发生相关的基因在Jin668中上调表达,而在TM-1中下调表达。其中,维持生长素极性分布的基因以及生长素内流相关基因AUX/LAX在Jin668的形成层和薄壁组织细胞中特异表达,而一些生长素外流相关基因在这些细胞类型中下调表达,这表明生长素的分布和极性转运对于细胞的脱分化和再分化具有重要作用。并通过CRISPR/Cas9基因编辑和过表达实验证明了LAX2、LAX1和LOX3在愈伤组织增殖和植物再生中发挥重要作用。此外,他们还鉴定了参与再生的新基因,包括CSEF、PIS2、AFB2、ATHB2、PLC2和PLT3。目前,该团队正在对这些与体细胞胚胎发生相关的关键基因(“再生因子”)进行深入的功能解析,更全面地研究体细胞胚胎产生背后的机制,以期将这些“再生因子”导入到不能再生或再生能力较弱的基因型中,实现高效的体细胞再生,从而彻底打破棉花体细胞植株再生的基因型依赖,实现更多棉花品种的植株再生,为棉花功能基因组研究和分子育种实践奠定重要的技术平台^[38]。近期该团队还完成了Jin668的端粒到端粒(telomere-to-telomere, T2T)水平的

完整基因组(gap-free)组装,有望挖掘出更多与植物体细胞胚胎发生相关的遗传密码(数据未发表)。

1.2 再生及递送系统的改良促进棉花遗传转化

为了打破基因型的限制,中国农业科学院棉花研究所李付广团队^[39]在棉花中建立了高效的遗传转化系统——顶端分生组织细胞介导的转化系统(shoot apical meristem cell-mediated transformation system, SAMT)。该系统结合农杆菌和超声波处理,将外源载体传递到棉花种子顶端分生组织的干细胞中,诱导不定芽产生,目前已经在陆地棉、海岛棉和亚洲棉等基因型顽固的品种中完成了遗传转化。另外,利用该系统获得了CRISPR/Cas9介导的棉花基因编辑植株。肖荣等人^[40]通过对茎尖转化法获得的阳性株进行了系统的验证,提出了一套假阳性株鉴定方法,推测由于茎尖侵染的农杆菌与受体植株共生关系的存在,存在一定比例的假阳性植株。Zhao等人^[41]开发了基于磁性纳米颗粒基因载体的花粉磁珠转化方法。该方法以磁性纳米颗粒 Fe_3O_4 作为载体,利用磁场介导外源片段传递至棉花花粉内,进而利用人工授粉获得转化的种子。在辣椒、南瓜、西葫芦和百合中也进行了验证。然而,Vejlupkova等人^[42]的研究称,利用纳米磁珠无法在单子叶植物中实现花粉转化,在百合中也重复不出来。进一步,在玉米中,Wang等人^[43]的研究称,纳米磁珠可以通过花粉萌发孔将外源基因导入花粉中,并通过人工授粉和自然结实过程,将外源基因转入多种玉米自交系中,部分解决了玉米遗传转化过程中“依赖组培体系,受基因型限制”的瓶颈问题,再次证明了花粉纳米磁转在植物转化中的可行性。

2 棉花基因编辑体系的建立与优化

2.1 基于Cas9蛋白的基因编辑体系

2.1.1 棉花CRISPR/Cas9体系的建立及优化

从2013年开始,华中农业大学棉花遗传改良团队获得了CRISPR/Cas9骨架载体pRGEB32^[44]。考虑到物种特异性内源U6启动子,他们使用棉花内源启动子pGhU6.9对载体进行改造,构建了棉花CRISPR/Cas9载体——pRGEB32-GhU6.9-NPT II。DsRed2蛋白来源于珊瑚(*Discosoma* sp.),已作为报告蛋白应用于植物分子生物学研究^[45]。前期,该团队创建了DsRed2过表达的转基因棉花株系,其种子颜色为红色^[46]。AtCLA1负责植物

的叶绿体发育,其突变体植株具有白化表型。因此,选择了外源基因DsRed2和棉花内源基因GhCLA1作为CRISPR/Cas9的编辑靶点。基于tRNA的自剪切作用设计了tRNA-sgRNA 转录元件^[44],利用CRISPR/Cas9系统在单个载体上生成两个sgRNA,对四倍体棉花(Jin668)进行基因编辑。经CRISPR/Cas9编辑的DsRed2过表达转基因材料的T₀植株获得了有效的基因突变,整个植株失去了红色的性状恢复为野生型。此外,T₀植株突变的表型和基因型能够稳定遗传到T₁后代中。对于内源基因GhCLA1,75%的再生植株表现为白化表型,Sanger测序表明,棉花中CRISPR/Cas9系统的基因编辑效率为66.7%~100%^[47]。

考虑到棉花稳定转化方法耗时长,技术难度大,限制了基因编辑技术的应用。河南大学宋纯鹏团队使用瞬时表达系统,在棉花(TM-1)中瞬时表达CRISPR/Cas9,开发了一种快速验证不同sgRNAs效率的方法。通过该方法验证了针对单个基因(*GhPDS*、*GhCLA1*和*GhEFL*)的sgRNA,在棉花子叶中产生的编辑类型大部分是碱基缺失(约64%);验证了同时表达多个sgRNA可以实现多基因靶向和同源基因的突变。此外,还通过稳定转化(YZ-1)获得了含有CRISPR/Cas9诱导的GhCLA1基因突变的棉花植株(80.6%的转基因株系),具有典型的白化表型^[48]。接着,该团队又提出了新的方法:结合农杆菌侵染、GUS染色和活性测定,利用瞬时表达系统来评估和验证CRISPR/Cas9在棉花中的效率。通过使用该系统,可以在稳定转化之前选择出最有效的CRISPR/Cas9载体。同时还优化了现有的棉花CRISPR/Cas9系统,通过引入内源性GhU6.3启动子来提高sgRNA的表达水平,从而提高了基因编辑效率^[49]。

CRISPR/Cas9技术依赖于在gRNA指导下的Cas9核酸酶在PAM位点的上游3个碱基的位点切割目标DNA(On-target),但也会切割非靶标的位点(与sgRNA靶标位点序列相似,且具有PAM位点),这就造成了所谓的脱靶(Off-target),从而引起不可控的突变^[50,51]。相比二倍体植物而言,异源四倍体棉花中,多拷贝数的同源基因或基因家族理论上会增加脱靶的可能性。石河子大学孙杰团队^[52]开发了一种简单易操作的CRISPR/Cas9系统,并通过靶向GhALARP(富含丙氨酸蛋白的编码基因),证明了其在棉花中具有高效的编辑效率。并在网站预测的15个潜在脱靶位点中均未检测到脱靶突变事件。基于前期已成功构建的棉花CRISPR/Cas9系统,华中农业大学棉花遗传改良团队利用高通量测序的方法,

首次详细评估了该系统在棉花基因编辑过程中的脱靶效应。通过对野生型对照、阴性再生植株、Cas9编辑的阳性植株进行35×测序深度的全基因组测序,他们认为遗传材料内在的遗传变异和组织培养过程中的体细胞变异位点数量远远大于CRISPR脱靶位点的数量(4000多个预测位点中,只检测到4个脱靶位点),表明CRISPR/Cas9系统在多倍体植物基因组编辑具有高度特异性,脱靶率极低^[53]。

在DNA双链断裂的修复过程中提供足够的gRNA和供体模板可以提高基因编辑(敲入和敲除)的效率。双生病毒具有快速侵染、复制、转录和表达而且不整合到植物基因组中的特点^[54],适合作为载体进行修饰,通过提供足够的转录物,从而提高细胞内基因编辑效率^[55]。华中农大棉花遗传改良团队选择使用已成功在植物细胞中高效表达的菜豆矮缩病毒(BeYDV)的温和株^[56]的滚环复制元件对已有的CRISPR/Cas9系统进行了改良。将棉花内源U6.7启动子及靶标序列的sgRNA片段插入到复制环内,结果表明,该双生病毒介导的基因编辑载体pBeYDV-Cas9-GhCLA1与原始载体pRGEB32-GhU6.7和pRGEB32-35s相比,它们对靶标位点的切割位点是一致的。此外,pBeYDV-Cas9-GhCLA1载体同时对棉花At和Dt亚基因组靶标基因的纯合突变编辑效率达到了73.3%。pBeYDV-Cas9-GhCLA1载体主要产生较大的DNA片段缺失,更容易引起靶基因的移码突变。作者认为可能是在基因编辑过程中,BeYDV双生病毒滚环复制元件可以提供更多的sgRNA片段,在有足够活性的Cas9蛋白的作用下,对目标基因靶位点的切割频率大大增加,使编辑效率得到了提升^[57]。利用该系统成功将*DsRed2*基因680 bp的全长片段插入到滚环复制元件中,实现了外源基因*DsRed2*的敲入,敲入效率在6%左右(未发表数据)。

2.1.2 棉花CBE单碱基编辑器的建立

棉花基因组中的许多等位基因具有高度同源性,只存在少数SNP位点的遗传变异,传统的CRISPR/Cas9系统无法实现对等位基因进行特定的点突变(碱基编辑)。因此,华中农业大学棉花遗传改良团队开发了棉花中首个CBE单碱基编辑系统。他们基于前期棉花基因编辑载体CRISPR/Cas9^[47],利用nCas9、胞嘧啶脱氨酶(APOBEC1)和尿嘧啶糖基化酶抑制剂(uracil glycosylase inhibitor, UGI)碱基编辑器单元,构建了适用于棉花遗传转化的碱基编辑系统(GhBE3)。并选择内源基因*GhCLA*和*GhPEBP*对系统进行验证。*GhPEBP*参与棉花

株型调控过程,具有显著的表型和重要的农艺价值,但该基因调控腋生开花和簇生棉铃的性状是由单个SNP决定的,因此无法通过传统的CRISPR/Cas9敲除来验证^[58]。通过该系统获得的材料中靶标基因*GhCLA*和*GhPEBP*的3个靶标均被编辑,具有很高的C-T碱基编辑效率(高达57.78%,与水稻和拟南芥的效率相当),以靶标位点上多个C位点同时编辑为主。将两个*GhCLA*基因编辑的单株和一个野生植株(Jin668)进行深度全基因组测序以检测脱靶效应,在全基因组水平预测的1500个潜在脱靶位点上没有发现脱靶突变,这表明单碱基编辑系统对棉花的基因组编辑具有较强的特异性^[59]。

2.1.3 棉花ABE单碱基编辑器的建立

近期,华中农业大学棉花遗传改良团队使用4种不同的腺苷脱氨酶(ABE6.3、7.8、7.9和7.10)与nCas9或dCas9融合,首次构建了棉花中的ABE系列载体(GhABE6.3n、GhABE6.3d、GhABE7.8n、GhABE7.8d、GhABE7.9n、GhABE7.9d、GhABE7.10n和GhABE7.10d)。此外,为了扩大ABEs在棉花中的靶点范围,该团队基于dCpf1还开发了一个新的ABE系统(GhABE7.10-dCpf1),首次成功应用于植物。通过San-ger和Barcode高通量测序显示,GhABE7.10n编辑效率最高,单株编辑效率高达64%。通过对GhABE7.10n编辑材料全基因组测序分析以及脱靶位点预测与验证表明,GhABE7.10n不存在DNA水平的脱靶。全转录组测序分析显示,GhABE7.10n在棉花中RNA水平脱靶也很低。识别PAM为TTTV的单碱基编辑载体GhABE7.10-dCpf1可以在棉花中工作,但编辑效率不高。通过对T₁代材料编辑效率检测显示,以上GhABEs产生的编辑位点都可以稳定地遗传给后代^[60]。

2.1.4 棉花转录激活系统的建立

基因表达的精确调控是功能基因组学研究、分子育种和种质创新的一种前沿方法。华中农业大学棉花遗传改良团队和中国棉花研究所范术丽团队合作将转录激活因子VP64、TAL和EDLL以不同的组合与dCas9融合,在烟草中评估其激活*LUC*报告基因的效率,发现dCas9-TV(dCas9-6×TAL-2×VP64)是最佳的组合。为了评估dCas9-TV系统在棉花中的效果,又选择了*GhTULP34*^[61]和*GhTLP19*基因作为靶标基因构建激活载体进一步验证。结果显示*GhTULP34*和*GhTLP19*的转录水平最高上调幅度分别可达35.5和30.5倍。气孔开度等数据证明,*GhTULP34*的定向上调提高了棉花的抗旱

能力。此后,他们通过该系统定向激活了棉花内源的*GhEPSPS*基因,成功创制了*GhEPSPS*表达量上调2.9~33.5倍不等的具有良好草甘膦抗性的棉花材料,并且这种抗性可以从 T_0 亲本稳定遗传到 T_1 后代。另外,利用该系统成功激活转录因子*GhPAPID*,其表达量最高上调了41.7倍,在植物组织培养过程中该基因的激活材料的愈伤组织出现了明显的红色或紫色表型,因此认为该基因可以作为棉花遗传转化过程中的报告基因来快速筛选转基因事件。此外,该基因的转基因植株的叶片、茎、枝条和花瓣呈现出明显的红/紫色,检测发现转基因株系中的原花青素和花青素含量显著提高,这一结果与这些植物的转录水平增加是一致的。这一研究为培育天然彩色棉花提供了一个有潜力的新策略,未来可以利用该系统结合组织特异性启动子来同时定向上调类黄酮途径中的多个基因(如*GhTT2*和*GhCHS*)实现棉花纤维颜色的改变^[62]。

2.2 基于Cas12蛋白的基因编辑体系

2.2.1 棉花CRISPR/Cpf1(Cas12a)体系的建立

目前在基因编辑中常用的Cpf1蛋白有3种(AsCpf1、LbCpf1和FnCpf1)。相关研究认为,LbCpf1是继Cas9系统之后的一种强大而精确的基因组编辑工具^[13]。为了验证LbCpf1系统在棉花中的效率,华中农业大学棉花遗传改良团队使用来源于毛螺菌科细菌的LbCpf1系统,设计了靶向棉花*GhCLA*基因的crRNA。Sanger测序和Hi-TOM结果显示,87%的 T_0 植株(80/92株)在目标位点检测到突变,大部分植株中编辑效率为20%~60%,与在水稻和玉米中的编辑效率相当。与1~3 bp小片段替换和插入相比,CRISPR/Cpf1系统更倾向于在棉花基因组编辑中产生较长片段(5~12 bp)的缺失。Sanger测序表明,在6个预测的脱靶位点均未检测到脱靶效应,表明CRISPR/Cpf1系统具有高度特异性,可以作为棉花基因组编辑的良好选择^[63]。为了进一步研究CRISPR/LbCpf1在棉花中的温度敏感性,他们构建了靶向棉花腺体基因*GhPGF*的LbCpf1载体,并与前期靶向*GhCLA*的LbCpf1载体获得的棉花植株同时验证Cpf1蛋白的温度敏感性。结果发现,在棉花中CRISPR/LbCpf1系统的最佳诱导温度为34℃。脱靶分析表明,LbCpf1系统在34℃下对棉花基因组没有脱靶作用。在此基础上,通过筛选获得了不含棉酚的无转基因成分(transgene free)、无腺体棉花材料,为棉花生物育种提供了重要的种质资源^[64]。

2.2.2 棉花CRISPR/C2c1(Cas12b)体系的建立

CRISPR/Cas12b是一个热诱导系统,该系统实现有效的基因编辑需要40~55℃的温度条件^[14,15]。棉花是一种喜温作物,可以在40~45℃的温度下生存一段时间^[65]。华中农业大学棉花遗传改良团队首次在棉花中构建了CRISPR/Cas12b(AacCas12b)基因编辑系统。通过设置不同的培养温度和时间处理转化的外植体,探究AacCas12b蛋白在棉花中的最适温度和处理时间。对获得上百株再生植株的编辑效率进行统计分析发现,在45℃处理4天时编辑效率最高,对愈伤组织的分化和存活几乎没有不利影响,因此认为45℃、4天的孵育时间是AacCas12b产生的棉花细胞培养和基因组编辑的最佳条件。CRISPR/AacCas12b系统编辑的类型主要是9~14 bp的长片段删除,这与CRISPR/Cas9在棉花中主要以1~5 bp较小片段的缺失、插入存在明显差异。同样地,该系统在预测脱靶位点没有检测到脱靶效应,其 T_0 代的编辑能够稳定地传递给子代。这些特征可以用于辅助Cas12b基因编辑系统在植物中更好地应用^[66]。目前这项技术获得了中国国家知识产权局的专利授权,是我国获批的第一个CRISPR/Cas12b基因编辑工具在高等植物中开发利用的授权专利,也是继中国农大开发的Cas12i和Cas12j后,又一把我国自主研发的“基因剪刀”,有效地弥补了我国在基因编辑核心底盘工具开发领域的不足。

综上,棉花基因编辑系统的建立及工具开发的大致发展历程如图2所示。

3 基因编辑技术在棉花生物育种中的应用

在抗虫方面,为了解析在棉花-刺吸式害虫互作过程中lncRNA的功能,华中农业大学棉花遗传改良团队^[67]通过qRT-PCR分析了两个lncRNAs(lncD09和lncA07)在刺吸式害虫侵染过程中的时空表达模式,用CRISPR/Cas9基因编辑技术创建了两个lncRNAs的功能丧失突变体,基因编辑后的植株均表现出显著的感虫特征,为研究lncRNA在棉花与刺吸式害虫互作过程中的作用提供了新的见解。近期,该团队开发了高通量基因编辑技术体系,利用sgRNA混合建库的方法,一次文库构建实现对600多个棉花抗虫相关基因的饱和突变,获得了包含2000多个独立的基因编辑株系的突变体库,并从中鉴定了一个广谱的棉花抗虫基因*GhMLP423*^[68]。

在抗病方面,Liu等人^[69]在高抗枯萎病的棉花品种

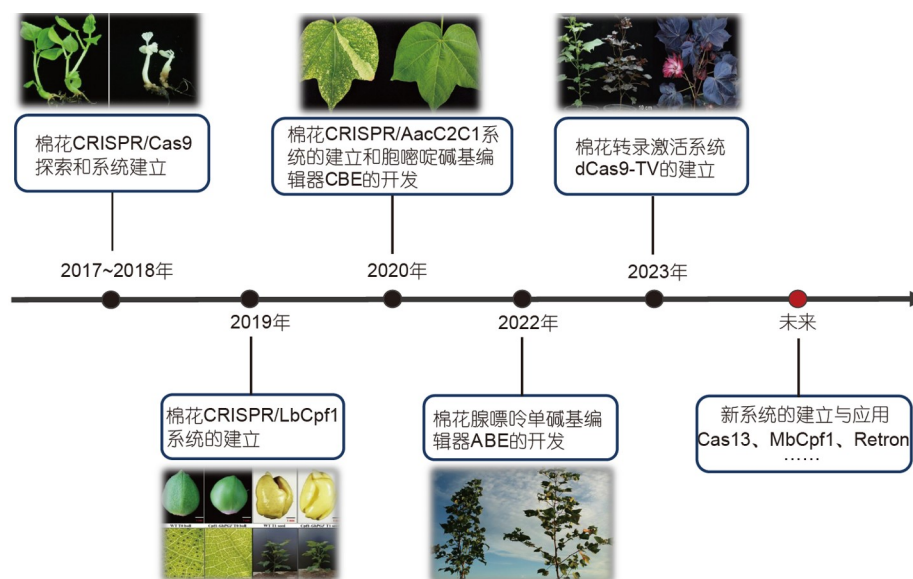


图2 棉花基因编辑研究的发展历程. 图片来源于文献^[47,59,60,62,64]

Figure 2 Development of cotton gene editing research. Pictures are from published references^[47,59,60,62,64]

中通过基因编辑突变 $GhGLR4.8$ 使得棉花品种变得极为感病. 根据 $GhGLR4.8$ 在抗病和感病品种中等位基因的SNP变异开发了基于PCR扩增的分子标记, 并建立了可区别 $GhGLR4.8$ 抗病基因型和感病基因型的检测体系. 该研究完善了植物与病原菌分子互作理论体系, 为棉花抗枯萎病分子育种提供了技术支撑, 并对棉花抗黄萎病研究提供了借鉴. Zhang等人^[70]利用CRISPR/Cas9对棉花中负调控黄萎病的基因 $Gh14-3-3d$ 进行编辑, 获得了两株不含T-DNA插入的突变体(transgene-free genome editing plants), 为棉花抗病育种提供了重要种质. 另外, Iqbal等人^[71]基于多重CRISPR/Cas9系统创制了抗棉花卷叶病(circumscribe cotton leaf curl disease, CLCuD)的材料.

在棉花抗旱方面, Li等人^[72]利用CRISPR/Cas9技术创建 $GhALKBH10B$ (去甲基化酶基因)的突变体, 苗期干旱实验分析表明突变 $GhALKBH10B$ 基因后植株的 m^6A 修饰显著上升, 干旱处理进一步增强植株 m^6A 修饰水平, 且 $alkbh10b$ 突变体植株抗旱性增强. 棉花侧根的生长发育影响棉花地上部分的生长及纤维产量, 而精氨酸酶的过表达会使得植株内NO含量明显下降, 抑制侧根的生长发育. Wang等人^[73]利用CRISPR/Cas9技术同时敲除陆地棉A、D亚基因组精氨酸酶编码基因 $Gh_A05G2143$ 和 $Gh_D05G2397$. 获得的双基因纯合株系在高氮或低氮条件下, 侧根生长发育都得到显著提升, 为培育耐干旱、耐贫瘠的棉花奠定了基础.

在棉籽、纤维品质改良方面, 华中农业大学和山东省农业科学院经济作物研究所合作以棉花 $GhFAD2-1A/D$ 基因为靶标, 利用pRGE32-GhU6.9-NPTII载体构建了 $GhFAD2-1A/D$ 基因的双靶点编辑载体, 通过农杆菌转化棉花品系Jin668, 获得种子中油酸含量显著升高和亚油酸含量显著降低的突变体. 该研究首次利用CRISPR/Cas9基因编辑技术创制出非转基因高油酸棉花新种质, 为改良棉籽油品质和高油酸棉花分子育种奠定了基础^[74]. Gao等人^[75]在有腺体棉花品种中CRISPR敲除 $CGP1$ 后棉花表现出与显性无腺体突变体 $Gl2e$ 相似的无腺体表型. 进一步, 在 $cgp1$ 突变体中, 棉酚合成路径相关基因表达量下调, 棉酚等萜类化合物的含量显著降低, 为研究腺体发育过程、腺体与棉酚及其他内含物关系提供新的视角和遗传资源. Lin等人^[76]使用CRISPR/Cas9介导的基因编辑技术对棉花中未知功能的 $GhDIR5$ 进行基因敲除, 突变体植株的所有组织中左旋棉酚含量显著下降, 而右旋棉酚和其他抗毒素的合成并未受到影响. $dir5$ 突变体棉籽对雄性小鼠的精细胞、胚胎成纤维细胞及人海拉细胞毒性显著降低. 该研究为开发脱毒安全的棉花品种奠定了理论与实践基础. Tian等人^[77]为了证实赤霉素诱导的GRF4对于激活独脚金内酯生物合成的重要性, 创制了 $GRF4$ 的CRISPR/Cas9敲除株系. 结果表明敲除 $GRF4$ 抑制了独脚金内酯的生物合成, 并干扰了与纤维细胞伸长和细胞壁厚度有关的独脚金内酯信号通路, 表明GRF4是连

接赤霉素和独脚金内酯之间的桥梁。

在棉花单倍体育种方面, 近期河南大学宋纯鹏团队和华中农业大学棉花遗传改良团队合作利用基因编辑技术首次创制出棉花单倍体育种材料。棉花传统育种主要通过自交或回交来固定重组基因型, 不仅耗时费力而且成本高昂, 极大地阻碍了棉花优良品种的选育。双单倍体技术可在两代内获得优良性状的纯合株系, 目前已经在多种作物中开始应用, 可以有效地缩短育种进程。自20世纪70年代以来, 棉花单倍体育种也受到了广泛关注。然而, 由于单倍体诱导系的制备困难重重, 因此进展缓慢。Long等人^[78]克隆了棉花中花粉精细胞特异表达的*GhDMP*基因, 并利用CRISPR/Cas9技术敲除*GhDMP*获得了高效的棉花单倍体诱导系。以ghdmp为父本的杂交F₁代中, 母本单倍体诱导率可以达到1.06%。该研究提供了一个不依赖于基因型的单倍体诱导系统, 实现了棉花单倍体育种的重大突破。

在细胞命运转变方面, Deng等人^[79]研究发现, CRISPR/Cas9介导的*GhTCE1*功能丧失导致活性氧(reactive oxygen species, ROS)过度积累, 阻止愈伤组织细胞伸长, 并增加不定器官发生, 为研究创伤诱导的愈伤组织早期细胞命运转变提供了重要见解, 对揭示分化状态的植物体细胞转变为全能细胞并最终形成胚胎的命运重塑机制有着重要意义。

在花药发育方面, Zhang等人^[80]通过CRISPR/Cas9基因编辑鉴定出棉花花药中的*GPAT12/25*的功能冗余, 且这两个基因的敲除导致完全棉花雄性不育, 花药异常, 小孢子不能存活, 为揭示棉花花药发育的调控机理提供了依据。Wu等人^[81]首次对棉花细胞核不育系1355A进行遗传定位, 并确认了陆地棉不育系1355A的败育恢复基因*GhNSP*。通过CRISPR/Cas9系统介导的突变分析证实了其功能, 并通过生化和免疫组化分析进行败育分子机制研究。此外, *GhNSP*基因在耐高温棉花花药中受高温诱导提前表达, 而在敏高温材料中不受诱导, 暗示其可能参与耐高温棉花种质高温下花粉外壁和花粉刺突的正常形成, 为耐高温棉花种质创新提供了基因资源。Zhang等人^[82]在常温(NT)条件下对棉花高温(HT)敏感系H05的花粉和花药与其他组织类型进行了转录组分析, 在NT和HT条件下对H05花粉进行了差异表达基因分析, 并创制了*Ghir_D12G012350*的基因编辑材料, 该基因是通过RT-PCR和qRT-PCR鉴定的16个候选*ASGs*(花粉特异基因)之一。与WT (Jin668)相比, 在NT条件下该突变体的花药开裂较差。结果表明,

*Ghir_D12G012350*基因突变影响了棉花花药发育。该研究为快速鉴定对HT有反应的*ASGs*和鉴定耐HT育种的关键基因提供了有用的方法。Li等人^[83]构建了首个常温 and 高温条件下四分体期花药的单细胞转录组图谱和染色质可及性图谱。他们发现负责花粉壁形成的绒毡层细胞对高温最为敏感, 并挖掘出影响高温下棉花花药育性的关键基因*GhCYP703A2*和*GhQRT3*。通过CRISPR/Cas9创制了这两个基因的敲除突变体, 在常温下均出现雄性不育表型。结果表明*GhCYP703A2*和*GhQRT3*是高温下棉花花药育性的关键正向调控因子。该研究不仅为植物花药发育的研究提供了大量线索, 也为未来作物耐热改良提供了思路。

在棉花植株生长发育方面, Wang等人^[84]利用过表达和CRISPR/Cas9基因编辑技术证实了*GhAP1-D3*基因在促进棉花提早成熟中具有重要作用。同时也发现*GhAP1-D3*能够显著地提升棉花早熟性, 而不降低产量和纤维品质等经济性状, 为深入解析陆地棉早熟性状的遗传基础提供理论依据。

4 总结与展望

对于植物的基础研究而言, 基因编辑技术的作用不只在基因敲除和碱基编辑, 还可以通过靶向目标位置, 实现基因的敲入、基因的转录调控(基因敲低和转录激活)、表观遗传修饰(DNA和RNA甲基化等)^[85,86]。因此, 除了以上系统, 扩展基因编辑系统的多功能性势在必行, 目前华中农业大学棉花遗传改良团队构建了用于DNA/RNA甲基化的表观修饰系统, 取得了初步的成功。基于双生病毒滚环复制元件的棉花基因敲入系统的工作也取得了初步成效。另外, 基于CRISPR/Cas9和Cas12a多基因编辑系统也取得了一定的进展, 已经成功实现一个载体同时对9个不同的基因进行精准编辑, 创制出的集食、纤、饲于一体棉花为多功能作物, 并能较好适宜新疆生产条件的早熟、矮秆、耐密植的要求。

虽然CRISPR/Cas9系统的应用很多, 但该系统本身也存在一些局限性, 比如, Cas9体积相对较大、有限的PAM位点选择等。需要开发新的工具来克服这些局限性, 实现更高效的基因编辑, 因而CRISPR/Cpf1和CRISPR/C2c1等这些蛋白更小、靶向富含T的PAM的新系统正在发展起来。为了使靶点选择范围更大, 新的CRISPR/Cas变体(spCas9-NG、xCas9等)也在不断涌现。另外, 靶向DNA的Cas14、靶向RNA的Cas13、基于非

编码RNA和逆转录酶的PE系统、Retron系统等新系统也相继出现^[87-90],有望将这些系统应用于棉花中。

除了基因编辑系统本身的因素以外,如何将基因编辑载体递送到植物细胞中依然是植物基因编辑工作的关键步骤。大多数植物需要复杂的组织培养系统来将CRISPR载体传递到它们的基因组中,然而以组织培养为基础的植物转化既费力又低效。近年来,植物转化领域的新成果,在一定程度上打破了转化过程中基因型的限制,比如使用形态发生调节因子促进了植物转化和再生;通过农杆菌转化技术的简化与创新克服了组培的复杂性;利用纳米颗粒和病毒载体递送质粒载体等。

高质量的基因组信息是设计精准的sgRNA靶标位点和评估脱靶效应的重要遗传基础,前期华中农业大学棉花遗传改良团队开展了基于全基因组重测序检测CRISPR/Cas9编辑的棉花材料的脱靶情况的研究^[52]。研究结果发现,受体材料基因组的准确性对于sgRNA的设计和脱靶效应的评估至关重要,当时棉花只有TM-1这个基因型有高质量的参考基因组,而用于遗传转化和基因编辑的受体材料——Jin668没有可用的基因组数据,这两个棉花基因型之间存在上百万的SNP和十多万Indel差异,这些遗传差异会对sgRNA靶标选择和精准设计造成极大影响:靶标位置的遗传差异(尤其是PAM位点的SNP)足以导致sgRNA的结合失败,造成编辑效率下降的同时会在非预期的位点产生脱靶,同时影响基因编辑的效率和精准性。因此我们认为获得转化受体材料的高质量基因组是进行有效基因编辑的重要前提和基础。华中农业大学棉花遗传改良团队近期完成了Jin668的T2T级别的基因组组装(未发表数据),鉴定了更多的与TM-1基因组的差异位点,尤其是在着丝粒区域高重复序列区段的遗传信息首次得到鉴定,以T2T级别的Jin668基因组为模板进行基因编辑能显著减少因为遗传材料背景差异造成的脱靶,大幅度提升基因编辑的效率和精准性。

现代生物育种发展历程大致经历了4个时代:(1)育种1.0时代,即经验育种时代,主要依赖耕作者的经验和主观判断,对作物表型变异的肉眼观察,筛选具有符合人类所需要的产量、品质与农艺性状的育种材料。

(2)育种2.0时代,即杂交育种时代,由职业育种家根据生物遗传规律以及田间统计学、数量遗传学等理论知识的运用,通过预先设计杂交育种试验选育现代栽培品种。(3)育种3.0时代,主要采用转基因技术及分子标记辅助选择的手段,实现单一目标性状的导入与修饰,同时基于连锁群体的功能基因克隆、自然群体的全基因组关联分析、复杂农艺性状的基因功能解析、各类高通量测序与芯片平台的发展等研究领域发展迅猛,使全基因组选择辅助育种技术逐步成为育种3.0时代的核心内容。(4)育种4.0阶段,即分子设计育种阶段,利用前沿生物技术与信息技术推动育种向着智能化的方向发展,即以基因组测序技术与人工智能图像识别技术为基础,通过基因型与表型数据的高通量、自动化获取与解析,实现组学大数据的快速积累;以生物信息学与机器学习技术为依托,通过常规育种数据、各类组学数据的整合,实现作物性状调控基因的快速挖掘与表型的精准预测;以基因编辑与合成生物学技术为主要手段,通过人工合成基因回路以及人工改造基因元件,赋予作物具备新的生物学性状,实现智能、高效、定向培育新品种。

国际上主要农作物已经逐渐步入育种4.0时代,而我国仍处于2.0向3.0时代过渡的阶段,仍有大量的关键核心技术亟待突破。我国棉花育种与主要粮食作物育种水平相比还有较大差距,未来如果要建立高效的棉花分子设计育种技术体系实现弯道超车,需要通过“精准基因组+高效基因编辑”的双引擎来驱动,高质量基因组是分子设计育种的“蓝图”,而基因编辑是绘制这幅蓝图的“画笔”。目前,我国棉花科研工作者在棉花高质量基因组测序(草图、染色体水平参考基因及T2T级别基因组完整图、泛基因组等)的研究处于国际领先地位。以华中农业大学棉花遗传改良团队为代表的科研团队建立了高效的棉花遗传转化体系并开发10余套高效基因编辑工具引领了国际棉花转基因和基因编辑技术的发展,在这些技术的支撑下,我们有理由相信在不远的将来,棉花分子设计育种的体系将会得到进一步完善,相关成果将会集中呈现,实现我国棉花生物育种跨越式发展,有效保障我国棉花供给安全。

参考文献

- 1 Urnov F D, Rebar E J, Holmes M C, et al. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, 2010, 11: 636–646
- 2 Gaj T, Gersbach C A, Barbas Iii C F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 2013, 31: 397–405

- 3 Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet*, 2011, 45: 273–297
- 4 Puchta H. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Curr Opin Plant Biol*, 2017, 36: 1–8
- 5 Mussolino C, Cathomen T. RNA guides genome engineering. *Nat Biotechnol*, 2013, 31: 208–209
- 6 Cong L, Ran F A, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, 339: 819–823
- 7 Mali P, Yang L, Esvelt K M, et al. RNA-guided human genome engineering via Cas9. *Science*, 2013, 339: 823–826
- 8 Shan Q, Wang Y, Li J, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol*, 2013, 31: 686–688
- 9 Symington L S, Gautier J. Double-strand break end resection and repair pathway choice. *Annu Rev Genet*, 2011, 45: 247–271
- 10 Lee K, Zhang Y, Kleinstiver B P, et al. Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol J*, 2019, 17: 362–372
- 11 Tang X, Lowder L G, Zhang T, et al. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat Plants*, 2017, 3: 17018
- 12 Tang X, Liu G, Zhou J, et al. A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol*, 2018, 19: 84
- 13 Xu R, Qin R, Li H, et al. Enhanced genome editing in rice using single transcript unit CRISPR-*Lb* Cpf1 systems. *Plant Biotechnol J*, 2019, 17: 553–555
- 14 Liu L, Chen P, Wang M, et al. C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism. *Mol Cell*, 2017, 65: 310–322
- 15 Yang H, Gao P, Rajashankar K R, et al. PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. *Cell*, 2016, 167: 1814–1828.e12
- 16 Wu D, Guan X, Zhu Y, et al. Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. *Cell Res*, 2017, 27: 705–708
- 17 Li S, Li J, He Y, et al. Precise gene replacement in rice by RNA transcript-templated homologous recombination. *Nat Biotechnol*, 2019, 37: 445–450
- 18 Kim J S. Precision genome engineering through adenine and cytosine base editing. *Nat Plants*, 2018, 4: 148–151
- 19 Komor A C, Kim Y B, Packer M S, et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 2016, 533: 420–424
- 20 Gaudelli N M, Komor A C, Rees H A, et al. Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature*, 2017, 551: 464–471
- 21 Hua K, Tao X, Yuan F, et al. Precise A·T to G·C base editing in the rice genome. *Mol Plant*, 2018, 11: 627–630
- 22 Kang B C, Yun J Y, Kim S T, et al. Precision genome engineering through adenine base editing in plants. *Nat Plants*, 2018, 4: 427–431
- 23 Li C, Zong Y, Wang Y, et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol*, 2018, 19: 59
- 24 Yan F, Kuang Y, Ren B, et al. Highly efficient A·T to G·C base editing by Cas9n-guided tRNA adenosine deaminase in rice. *Mol Plant*, 2018, 11: 631–634
- 25 Pan C, Sretenovic S, Qi Y. CRISPR/dCas-mediated transcriptional and epigenetic regulation in plants. *Curr Opin Plant Biol*, 2021, 60: 101980
- 26 Kampmann M. CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. *ACS Chem Biol*, 2018, 13: 406–416
- 27 Pan C, Wu X, Markel K, et al. CRISPR-Act3.0 for highly efficient multiplexed gene activation in plants. *Nat Plants*, 2021, 7: 942–953
- 28 Ming M, Ren Q, Pan C, et al. CRISPR-Cas12b enables efficient plant genome engineering. *Nat Plants*, 2020, 6: 202–208
- 29 Chen Z J, Scheffler B E, Dennis E, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol*, 2007, 145: 1303–1310
- 30 Wang M, Tu L, Lin M, et al. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. *Nat Genet*, 2017, 49: 579–587
- 31 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
- 32 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
- 33 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
- 34 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
- 35 Jin S, Zhang X, Nie Y, et al. Identification of a novel elite genotype for *in vitro* culture and genetic transformation of cotton. *Biol Plant*, 2006, 50: 519–524
- 36 Jin S X, Tu L L, Zhang X L. A method for improving the efficiency of cotton regeneration and transformation and its application (in Chinese). PRC Patent, CN106801065A, 2017-06-06 [金双侠, 涂礼莉, 张献龙. 一种提高棉花再生及转化效率的方法及应用. 中国: CN106801065A, 2017-06-06]

06]

- 37 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
- 38 Zhu X, Xu Z, Wang G, et al. Single-cell resolution analysis reveals the preparation for reprogramming the fate of stem cell niche in cotton lateral meristem. *Genome Biol*, 2023, 24: 194
- 39 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
- 40 Xiao R, Wei Y X, Wang Y, et al. Exploring of authenticity identification method of transgenic cotton plant by shoot tip (in Chinese). *Curr Biotechnol*, 2022, 12: 83–89 [肖荣, 魏云晓, 王远, 等. 茎尖法转基因棉花植株真实性鉴定方法探究. *生物技术进展*, 2022, 12: 83–89]
- 41 Zhao X, Meng Z, Wang Y, et al. Pollen magnetofection for genetic modification with magnetic nanoparticles as gene carriers. *Nat Plants*, 2017, 3: 956–964
- 42 Vejlupekova Z, Warman C, Sharma R, et al. No evidence for transient transformation via pollen magnetofection in several monocot species. *Nat Plants*, 2020, 6: 1323–1324
- 43 Wang Z, Zhang Z, Zheng D, et al. Efficient and genotype independent maize transformation using pollen transfected by DNA-coated magnetic nanoparticles. *J Integr Plant Biol*, 2022, 64: 1145–1156
- 44 Xie K, Minkenberg B, Yang Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci USA*, 2015, 112: 3570–3575
- 45 Nishizawa K, Kita Y, Kitayama M, et al. A red fluorescent protein, DsRed2, as a visual reporter for transient expression and stable transformation in soybean. *Plant Cell Rep*, 2006, 25: 1355–1361
- 46 Sun L, Alariqi M, Zhu Y, et al. Red fluorescent protein (DsRed2), an ideal reporter for cotton genetic transformation and molecular breeding. *Crop J*, 2018, 6: 366–376
- 47 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
- 48 Gao W, Long L, Tian X, et al. Genome editing in cotton with the CRISPR/Cas9 system. *Front Plant Sci*, 2017, 8: 1364
- 49 Long L, Guo D D, Gao W, et al. Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods*, 2018, 14: 85
- 50 Fu Y, Foden J A, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, 2013, 31: 822–826
- 51 Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337: 816–821
- 52 Zhu S, Yu X, Li Y, et al. Highly efficient targeted gene editing in upland cotton using the CRISPR/Cas9 system. *Int J Mol Sci*, 2018, 19: 3000
- 53 Li J, Manghwar H, Sun L, et al. Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. *Plant Biotechnol J*, 2019, 17: 858–868
- 54 Baltes N J, Gil-Humanes J, Cermak T, et al. DNA replicons for plant genome engineering. *Plant Cell*, 2014, 26: 151–163
- 55 Vu T V, Sivankalyani V, Kim E, et al. Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. *Plant Biotechnol J*, 2020, 18: 2133–2143
- 56 Regnard G L, Halley-Stott R P, Tanzer F L, et al. High level protein expression in plants through the use of a novel autonomously replicating geminivirus shuttle vector. *Plant Biotechnol J*, 2010, 8: 38–46
- 57 Li B, Fu C, Zhou J, et al. Highly efficient genome editing using geminivirus-based CRISPR/Cas9 system in cotton plant. *Cells*, 2022, 11: 2902
- 58 Si Z, Liu H, Zhu J, et al. Mutation of SELF-PRUNING homologs in cotton promotes short-branching plant architecture. *J Exp Bot*, 2018, 69: 2543–2553
- 59 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
- 60 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
- 61 Li Z, Wang X, Cao X, et al. GhTULP34, a member of tubby-like proteins, interacts with GhSKP1A to negatively regulate plant osmotic stress. *Genomics*, 2021, 113: 462–474
- 62 Yu L, Li Z, Ding X, et al. Developing an efficient CRISPR-dCas9-TV-derived transcriptional activation system to create three novel cotton germplasm materials. *Plant Commun*, 2023, 4: 100600
- 63 Li B, Rui H, Li Y, et al. Robust CRISPR/Cpf1 (Cas12a)-mediated genome editing in allotetraploid cotton (*Gossypium hirsutum*). *Plant Biotechnol J*, 2019, 17: 1862–1864
- 64 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*, 2021, 19: 221–223
- 65 Ekinci R, Basba S, Karademir E, et al. The effects of high temperature stress on some agronomic characters in cotton. *Pak J Bot*, 2017, 49: 503–508
 - 66 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
 - 67 Zhang J, Li J, Saeed S, et al. Identification and functional analysis of lncRNA by CRISPR/Cas9 during the cotton response to sap-sucking insect infestation. *Front Plant Sci*, 2022, 13: 784511
 - 68 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*, 2023, 11 [e2306157]
 - 69 Liu S, Zhang X, Xiao S, et al. A single-nucleotide mutation in a GLUTAMATE RECEPTOR-LIKE gene confers resistance to fusarium wilt in *Gossypium hirsutum*. *Adv Sci*, 2021, 8: 2002723
 - 70 Zhang Z, Ge X, Luo X, et al. Simultaneous editing of two copies of *Gh14-3-3d* confers enhanced transgene-clean plant defense against *Verticillium dahliae* in allotetraploid upland cotton. *Front Plant Sci*, 2018, 9: 842
 - 71 Iqbal Z, Sattar M N, Shafiq M. CRISPR/Cas9: A tool to circumscribe cotton leaf curl disease. *Front Plant Sci*, 2016, 7: 475
 - 72 Li B, Zhang M, Sun W, et al. *N*⁶-methyladenosine RNA modification regulates cotton drought response in a Ca²⁺ and ABA-dependent manner. *Plant Biotechnol J*, 2023, 21: 1270–1285
 - 73 Wang Y, Meng Z, Liang C, et al. Increased lateral root formation by CRISPR/Cas9-mediated editing of arginase genes in cotton. *Sci China Life Sci*, 2017, 60: 524–527
 - 74 Chen Y, Fu M, Li H, et al. High-oleic acid content, nontransgenic allotetraploid cotton (*Gossypium hirsutum* L.) generated by knockout of *GhFAD2* genes with CRISPR/Cas9 system. *Plant Biotechnol J*, 2021, 19: 424–426
 - 75 Gao W, Xu F, Long L, et al. The gland localized *CGPI* controls gland pigmentation and gossypol accumulation in cotton. *Plant Biotechnol J*, 2020, 18: 1573–1584
 - 76 Lin J L, Fang X, Li J X, et al. Dirigent gene editing of gossypol enantiomers for toxicity-depleted cotton seeds. *Nat Plants*, 2023, 9: 605–615
 - 77 Tian Z, Zhang Y, Zhu L, et al. Strigolactones act downstream of gibberellins to regulate fiber cell elongation and cell wall thickness in cotton (*Gossypium hirsutum*). *Plant Cell*, 2022, 34: 4816–4839
 - 78 Long L, Feng Y M, Shang S Z, et al. *In vivo* maternal haploid induction system in cotton. *Plant Physiol*, 2023, 194: 1286–1289
 - 79 Deng J, Sun W, Zhang B, et al. GhTCE1-GhTCEE1 dimers regulate transcriptional reprogramming during wound-induced callus formation in cotton. *Plant Cell*, 2022, 34: 4554–4568
 - 80 Zhang M, Wei H, Hao P, et al. *GhGPAT12/25* are essential for the formation of anther cuticle and pollen exine in cotton (*Gossypium hirsutum* L.). *Front Plant Sci*, 2021, 12: 667739
 - 81 Wu Y, Li X, Li Y, et al. Degradation of de-esterified pectin/homogalacturonan by the polygalacturonase GhNSP is necessary for pollen exine formation and male fertility in cotton. *Plant Biotechnol J*, 2022, 20: 1054–1068
 - 82 Zhang R, Zhou L, Li Y, et al. Rapid identification of pollen- and anther-specific genes in response to high-temperature stress based on transcriptome profiling analysis in cotton. *Int J Mol Sci*, 2022, 23: 3378
 - 83 Li Y, Ma H, Wu Y, et al. Single-cell transcriptome atlas and regulatory dynamics in developing cotton anthers. *Adv Sci*, 2023, 11: e2304017
 - 84 Wang C, Liu J, Xie X, et al. GhAPI-D3 positively regulates flowering time and early maturity with no yield and fiber quality penalties in upland cotton. *J Integr Plant Biol*, 2023, 65: 985–1002
 - 85 Manghwar H, Lindsey K, Zhang X, et al. CRISPR/Cas system: Recent advances and future prospects for genome editing. *Trends Plant Sci*, 2019, 24: 1102–1125
 - 86 Hassan M M, Zhang Y, Yuan G, et al. Construct design for CRISPR/Cas-based genome editing in plants. *Trends Plant Sci*, 2021, 26: 1133–1152
 - 87 Khan M Z, Haider S, Mansoor S, et al. Targeting plant ssDNA viruses with engineered miniature CRISPR-Cas14a. *Trends Biotechnol*, 2019, 37: 800–804
 - 88 Wolter F, Puchta H. The CRISPR/Cas revolution reaches the RNA world: Cas13, a new Swiss Army knife for plant biologists. *Plant J*, 2018, 94: 767–775
 - 89 Vu T V, Nguyen N T, Kim J, et al. Prime editing: Mechanism insight and recent applications in plants. *Plant Biotechnol J*, 2024, 22: 19–36
 - 90 Kong X, Wang Z, Zhang R, et al. Precise genome editing without exogenous donor DNA via retron editing system in human cells. *Protein Cell*, 2021, 12: 899–902

Summary for “棉花基因编辑工具开发及分子育种进展与展望”

Progress and prospect of genome editing tools development and molecular breeding in cotton

Guangqin Yang¹, Guanying Wang¹, Lu Yu¹, Xianlong Zhang¹, Xinhui Nie^{2*}, Jian Li^{2*} & Shuangxia Jin^{1*}¹ Hubei Hongshan Laboratory, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China² The Southern Xinjiang Research Institute of Shihezi University, TuMu ShuKe 843900, China

* Corresponding authors, E-mail: jsx@mail.hzau.edu.cn; 17699436688@163.com; xjnxh2004130@126.com

Genome editing technology can achieve efficient and accurate editing at the biological gene and transcription level, which is a revolutionary technology in the field of life science. The genome editing system based on Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein (CRISPR/Cas), has become the most widely used genome editing system in the field of plants with its simple, fast and efficient characteristics. Cotton is an important economic crop and the main cultivated species of upland cotton (*G. hirsutum* L.) is an allotetraploid plant species with complex genome. The upland cotton genome contains At- and Dt- two subgenomes with multiple gene copy, therefore, the gene functional redundancy is common. Therefore, the conventional T-DNA insertion and chemistry, physical mutagenesis and other technologies to create mutants are very low efficiency in cotton. In recent years, Chinese cotton researchers have established efficient genome editing tools including CRISPR/Cas9/12a/12b, CBE/ABE/ABE8e/CABE base editor, Cas13a/b/c and dCas9-TV transcriptional activation, which can achieve knock out, knock in, knock down, knock up and point mutation of the target genes. In addition, they have established a system to evaluate off-target effects of genome editing based on whole genome re-sequencing strategy. Whether the genome editing system can be effectively delivered to plants is an important step in gene editing, and the genotype of receptor material plays a decisive role in the effective delivery of genome editing system in the genetic transformation mainly mediated by *Agrobacterium*. Jin668, a cotton material with high genetic transformation efficiency, is a widely used receptor material in the world at present, and is an important receptor material for exploring the function of cotton endogenous genes and conducting gene editing research. Here, the Successive Regeneration Acclimation (SRA) strategy used by cotton researchers to obtain Jin668 with high transformation efficiency and the analysis of its regeneration mechanism is introduced in detail. It is expected to break the genotypic restriction of other cotton varieties during transformation and regeneration. In this review, the establishment process of cotton genome editing tools is introduced in particular. Using the above efficient genetic transformation system and genome editing tools, researchers have created a mutant library of more than 5,000 genes, including cotton fiber quality, resistance to disease, insect, high temperature, salt and alkali, cotton seed quality, haploid induction and other traits, effectively promoting the development of basic research and molecular breeding of cotton. This paper also discusses the limitations of the genome editing system itself and the problems in the delivery process of the genome editing system and the solutions. At the end of the paper, the future development direction of cotton gene editing research is also prospected. The era of breeding 4.0, namely the stage of molecular design breeding, is the trend of crop breeding in the future. For cotton, the strategy of “precise genome+efficient gene editing” is expected to drive the establishment of an efficient molecular design breeding system for cotton in the future.

cotton, genome editing, CRISPR/Cas9/12/13, base editor, transcription activator, molecular breedingdoi: [10.1360/TB-2023-1138](https://doi.org/10.1360/TB-2023-1138)