

Forum

From Sequencing to Genome Editing for Cotton Improvement

Renhai Peng,^{1,2,*}
 Don C. Jones ,³ Fang Liu,^{2,*}
 and Baohong Zhang ^{4,5,*}



Traditional breeding techniques are proven, but additional knowledge learned from genome sequencing provides vast new data that might help identify gene targets for improving cotton sustainability. CRISPR/Cas9 provides a powerful tool for precision cotton breeding. Here, we discuss the opportunities and challenges of genome sequencing and editing for cotton improvement.

Cotton is the leading fiber crop and an important source of protein, oil, and pharmaceuticals. Annual economic value of cotton is about US\$500 billion and provides income for 100 million families in about 150 countries [1]. However, as global warming and environmental changes continue, traditional cotton breeding techniques must be supplemented with new molecular tools so that cotton remains competitive against synthetic fiber. Although transgenic technology has enabled significant progress on breeding new cultivars with useful agronomic traits, the need remains for cotton to be more tolerant to numerous environmental extremes. Rapid development of cotton genome sequencing and gene editing may help solve such problems and improve cotton's sustainability footprint. In this forum article, we review the major progress of whole genome sequencing and gene editing, two rapidly developing fields in biotechnology. Then, we point out major problems associated with these two techniques and cotton breeding.

Integrated genome sequencing and gene editing will open a new era of precision in plant breeding and lead to new advancements in related research.

Sequencing Provides Genetic Resources/Gene Targets for Cotton Improvement

Sequencing cotton genomes and transcriptomes (Box 1) provides genetic information not previously available, including genomic structure, variation, and diversity. Sequencing also identified numerous genes and biological processes associated with agriculturally important traits such as fiber and stress response. Resequencing 737 cultivars identified genes that are associated with cotton fiber development and lint yield [2,3] and, interestingly, lint yield has stronger selection signatures than other traits during the long history of cotton evolution and domestication [2]. Cotton whole chromosome duplication and polyploidization doubled the number of fiber-associated genes. The expression of these fiber-related genes was significantly enhanced in cultivated cotton versus their wild type [1,4]. Genome sequencing also provides a landscape of disease-resistant genes in cotton that were diversified during cotton selection and domestication [1]. Additionally, sequencing analysis showed the epigenomic landmark and its importance in cotton, including DNA methylation and small regulatory RNAs, particularly for cotton fiber development and response to environmental stresses [1,5]. Understanding epigenetic inheritance may be a novel way to improve cotton's fiber yield, quality, and environmental sustainability.

Genome Editing Provides a Powerful Tool for Cotton Improvement

Although transgenic-based cotton improvement has been substantial, the community needs additional molecular tools. Recently developed CRISPR/Cas9 genome editing

technology has the potential to harness more genetic modifications of cotton traits and open an era for precision cotton breeding. CRISPR/Cas9 is a tool adopted from a prokaryotic adoptive immune defense system. Compared with transgenic technology, CRISPR/Cas9 technology not only can be used to knockout an individual gene for loss-of-function but also knock in and modify an individual gene at the epigenetic and transcriptional levels [6]. CRISPR/Cas9 genome editing has attracted attention from both academic and industrial communities and is becoming established in many laboratories for editing a specific gene in cotton for improving agriculturally important traits.

Studying cotton fiber initiation and development is a top priority for cotton improvement. Li and colleagues knocked out a *MYB-25like* transcription factor gene in cotton with high efficiency and no off-target impact [7]. The knockout line of *MYB-25like* displays a fibreless mutation without changing other phenotypic traits. By modulating the brassinosteroid signaling pathway, *14-3-3* plays a vital role in cotton fiber differentiation and early development. Overexpression of the *14-3-3l* gene produced longer fibers, while inhibiting its expression negatively affected cotton fiber initiation and elongation [8]. CRISPR/Cas9 knockout of key enzyme genes associating with gossypol biosynthesis may yield transgene-free gossypol-free cotton that could strengthen the value of cotton production (C. Li and B.H. Zhang, unpublished).

Another major application of genome editing in cotton improvement is to improve resistance to both biotic and abiotic stresses. CRISPR/Cas9 knockout of the *14-3-3d* gene significantly enhanced resistance to *Verticillium dahlia*, a serious pathogen in cotton [9]. CRISPR/Cas9 knockout of the *arg* gene significantly enhanced the total number of lateral roots and the total

Box 1. Cotton Genome Sequencing

In the past decade, genome and multiple omics sequencing (including transcriptome, proteomics, and noncoding sequencing) has become a powerful tool to study genome structure and evolution as well as genetic contribution to agriculturally important traits. Since the cotton D-genome diploid (*Gossypium raimondii*) was sequenced in 2012 [4,13], cotton genome sequencing has attracted attention from both academia and industry. Currently, among 52 cotton species, more than one dozen cultivated and uncultivated species have been fully sequenced and hundreds of cotton accessions have been resequenced. Table I summarizes the major genomes characterized of two major cultivated allotetraploid species, *Gossypium hirsutum* and *Gossypium barbadense*, and their two ancestral progenitors. There are four cultivated cotton species, two of them (*G. hirsutum* and *G. barbadense*) are allotetraploid species, the other two (*Gossypium herbaceum* and *Gossypium arboreum*) are diploid species. Due to their importance during the long history of evolution and cotton domestication, all four cultivated species have been fully sequenced. Comparing their genome structure and gene expression pattern as well as the copy number variation of many genes (orthologs and paralogs) during cotton evolution and domestication is essential for biotechnological improvement of cultivated cotton. Further studying the genome structure and evolution of cultivated cotton species and their wild species will enhance our understanding of the major agriculturally important traits, such as yield, quality, and response to climate change.

Table I. Summary of the Whole Genome Sequences of Cultivated Cotton Species and Their Two Potential Wild Species Ancestors

Genomic features	<i>G. hirsutum</i>	<i>G. barbadense</i>	<i>G. arboreum</i>	<i>G. raimondii</i>
Genome	AADD	AADD	AA	DD
Year sequenced	2015	2015	2014	2012
Genome size (Gb)	2.305	2.196–2.570	1.719	0.740
Number of genes ^a	76 943	80 876	41 330	40 976
Repeat sequences (%)	67.2		68.5	60
Refs	[1–3]	[14]	[15]	[4,13]

^aThe gene number refers the higher number reported.

root surface area in both normal and nitrogen deficiency conditions [10], which may help in water and nutrient uptake and strengthen resistance to other abiotic stresses, thereby improving cotton's environmental adaptability.

Finding Key Genes for Agriculturally Important Traits is Critical for Cotton Improvement

Identifying key genes that improve agriculturally important traits, such as fiber strength and length as well as tolerance to biotic and abiotic stresses, is still a bottleneck for cotton improvement. Although substantial whole genome and multi-omics sequencing have been conducted and genes associated with certain traits have been discovered, their exact function remains unclear. Additionally, there are many genes associated with a specific trait, but the key genes or master genes have not been elucidated. In the future, genome sequencing, other omics analysis, and gene editing

should be integrated to help identify the important genes for crop improvement, including those associated with cotton fiber initiation and development and tolerance to environmental stresses. Epigenetic regulation may also play an important role in these processes, so identifying genes controlling epigenetic regulation and its molecular mechanism is a critical objective for cotton improvement.

Integrated genotyping and phenotyping efforts are necessary to identify key genes controlling agriculturally important traits in cotton. Although many whole genome sequencing and resequencing projects have been performed, direct links between the genotype and phenotype still contains gaps. In the future, phenotyping based on genotyping may be used to identify key genes controlling cotton fiber development and response to extreme environmental stresses. The phenotyping should not be only limited to an individual plant, it should

be on a population scale with a variety of phenotypic data, including population survey on development-, stress-, and metabolism-related traits. Additionally, during the long history of cotton evolution, wild species have evolved several mechanisms to handle the external environment, so continuously sequencing these wild cotton species may identify useful genes for modern cotton improvement.

A Highly Efficient Transgenic and Genome Editing System Is Urgently Needed

Regenerating plants is still a major bottleneck for use of CRISPR/Cas9 genome editing in cotton. Currently agrobacteria-mediated gene transformation is the major approach for obtaining transgenic and genome editing plants. However, this method requires cotton tissue culture and plant regeneration. Currently there are only a few widely used genotypes, such as Coker 312 and Coker 201 in the US; YZ-1, Simian-3, and CRI24 in China; and Siokral 1-3 in Australia. This limits the application of genome editing in cotton. To enhance the plant's regeneration capacity, several labs have attempted to obtain germplasm with a high frequency of cotton somatic embryogenesis and plant regeneration. Thus, developing a highly efficient plant regeneration system in cotton will boost the application of transgenic and genome editing. To achieve this, more work should be conducted, including exploring various combinations of plant hormones and chemicals. Exploring different explants may also provide a new way to obtain regeneration plants in a genotype-independent manner. During the 1990s, rice and maize tissue culture experienced difficulty in obtaining regeneration, but the issue was resolved by using immature embryos as explants. For cotton, cotyledon and hypocotyl are common explants in tissue culture and transgenic research. Different explants, including the immature embryos and meristems, may be explored to develop a plant regeneration system independent of genotype.

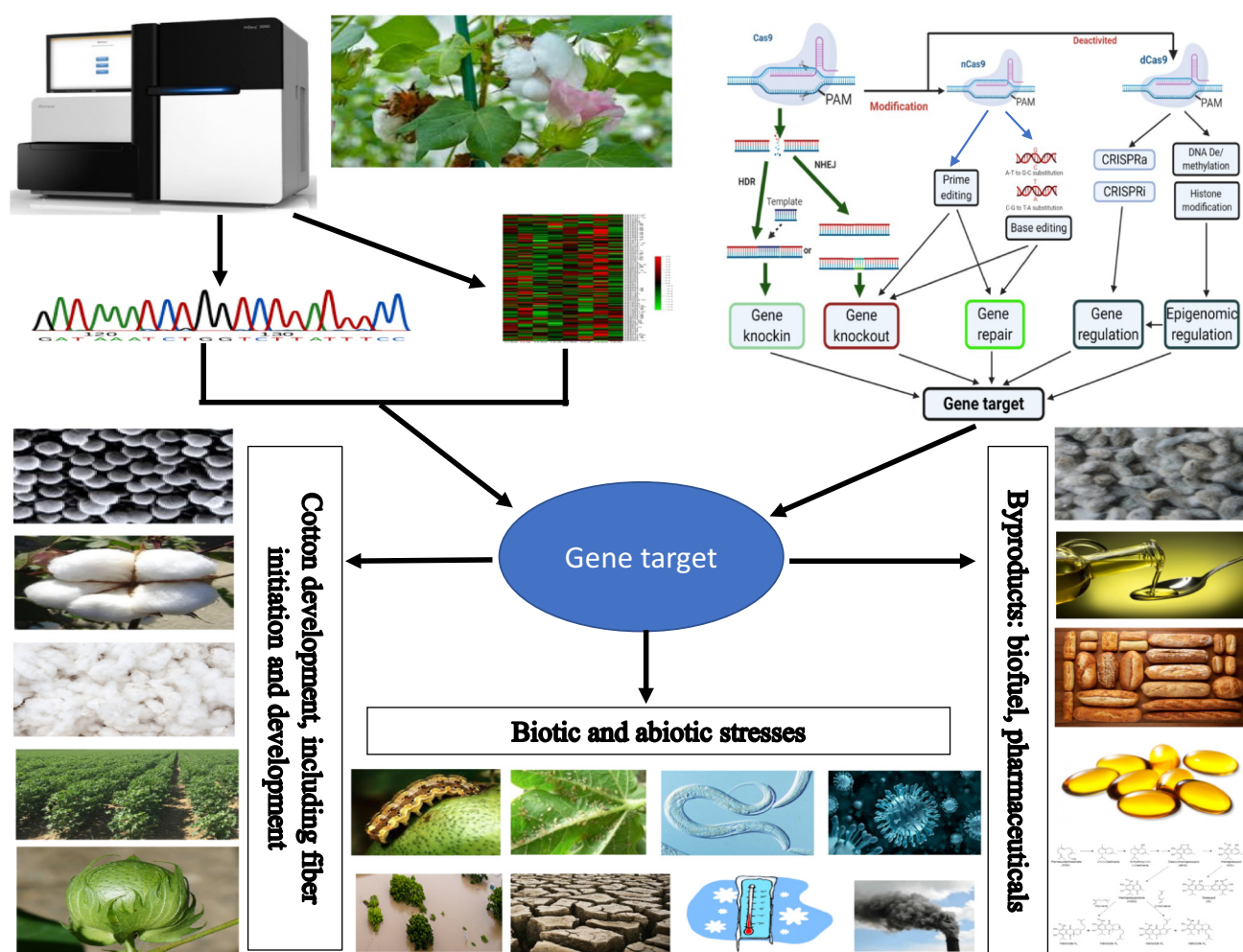
Exploring nontraditional transformation methods may be another way to escape the genotype requirements. For example, it may be possible to transfer CRISPR/Cas9 RNA or protein sequences, such as Cas9-sgRNA ribonucleoprotein (RNP), into cotton cells to test if they edit. RNP is a preassembled CRISPR/Cas9 system already used to edit genes in several plant species, which is a DNA-free genome editing approach [11]. This approach may

not only increase efficiency but also reduce off target effects as well as produce transgene-free plants. Additionally, Ma and colleagues recently developed a virus-based CRISPR/Cas9 genome editing method in which they engineered the virus to express CRISPR/Cas9 reagents and then used the engineered virus to infect plants. After the infection, the virus will express the CRISPR/Cas9 components and allow genome editing to occur [12]. This method

does not require plant tissue culture and can be done *in planta* without genotype limitation. This method may be a highly efficient genome editing method in cotton.

Concluding Remarks

Rapid development of deep sequencing and genome editing technology provides powerful opportunities for cotton improvement. Whole genome and multi-omics sequencing provide gene targets for genome



Trends in Biotechnology

Figure 1. Integrated Genome Sequencing and Gene Editing Will Harness the Precision Breeding and Biotechnological Improvement in Cotton. Sequencing-based omics, including whole genome sequencing, transcriptome and proteome sequencing, as well as small RNA sequencing will help identify key genes and their associated gene network for crop improvement. However, the function of an individual gene needs to be confirmed and CRISPR/Cas9 genome editing is particularly useful technology. There are huge potential applications of integrated genome editing and gene editing in precision cotton breeding, not only in improvement of cotton fiber yield and quality but also other major agriculturally important traits, including morphological improvement and resistance to various environmental abiotic (e.g., drought, salinity, and extreme temperatures) and biotic (e.g., pests, diseases, and nematodes) stresses. This new technology can also be employed to enhance cotton values by increasing pharmaceutical (e.g., vitamin E and gossypol) content as well as protein and oil content, two major nutrients for both human and livestock. Cotton is also a potential biofuel crop because it contains high pure cellulose and oil; the application of genome sequencing and gene editing will enhance the application of precision breeding on improving biofuel-related traits.

editing. Integrated application of both technologies will open a new era for precision cotton breeding for improving a variety of traits related to agriculture, biofuel, and pharmaceuticals (Figure 1) that should enhance the value of cotton, increase the income of cotton farmers, and strengthen cotton's global competitiveness against man-made fibers. However, the integrated application of genome sequencing and gene editing is also facing critical challenges, including both technical and regulation as well as intellectual property (IP)-related issues. Finding key targeting genes can be achieved by combining genotype and phenotype sequencing following gene function study by genome editing. Transgene-free-based genome editing is not regulated by US-based agencies in the same manner as transgenes were. However, IP policy may affect the quick application of both techniques, particularly genome editing on precision breeding of crops, including cotton. Thus, seeking new strategies on IP may enhance the application of genome editing.

Acknowledgments

Both genome sequencing and gene editing are hot research topics in cotton. We greatly appreciate the

significant contributions of many research laboratories to these two fields. However, due to the space limitation, we only can cite 15 papers in this short article and we apologize to all authors whose work could not be cited, but that do make significant contribution to these two fields. We greatly appreciate Dr Kunbo Wang for his great suggestion and comments on this manuscript. This work is partially supported by the National Natural Science Foundation of China (No. 31471548) and the National Key Research and Development Program of China (No. 2019YFD0100302) to R.P. B.Z. is supported in part by Cotton Incorporated and the National Science Foundation (Award No. 1658709).

¹Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

²State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China

³Cotton Incorporated, Cary, NC 27513, USA

⁴Department of Biology, East Carolina University, Greenville, NC 27858, USA

⁵<https://biology.ecu.edu/faculty/baohong-zhang/>

*Correspondence:

aydxprh@163.com (R. Peng), liufcri@163.com (F. Liu), and zhangb@ecu.edu (B. Zhang).

<https://doi.org/10.1016/j.tibtech.2020.09.001>

© 2020 Elsevier Ltd. All rights reserved.

References

- Chen, Z.J. *et al.* (2020) Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* 52, 525–533
- Fang, L. *et al.* (2017) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* 49, 1089–1098
- Ma, Z. *et al.* (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50, 803–813
- Paterson, A.H. *et al.* (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427
- Song, Q. *et al.* (2017) Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18, 99
- Ahmad, N. *et al.* (2020) A critical look on CRISPR-based genome editing in plants. *J. Cell. Physiol.* 235, 666–682
- Li, C. *et al.* (2017) A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in cotton (*Gossypium hirsutum* L.). *Sci. Rep.* 7, 43902
- Zhou, Y. *et al.* (2015) Cotton (*Gossypium hirsutum*) 14-3-3 proteins participate in regulation of fibre initiation and elongation by modulating brassinosteroid signalling. *Plant Biotechnol. J.* 13, 269–280
- Zhang, Z. *et al.* (2018) 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.* 9, 842
- Wang, Y. *et al.* (2017) Increased lateral root formation by CRISPR/Cas9-mediated editing of arginase genes in cotton. *Sci. China Life Sci.* 60, 524–527
- Woo, J.W. *et al.* (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164
- Ma, X. *et al.* (2020) Highly efficient DNA-free plant genome editing using virally delivered CRISPR–Cas9. *Nat. Plants* 6, 773–779
- Wang, K. *et al.* (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44, 1098–1103
- Liu, X. *et al.* (2015) *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Sci. Rep.* 5, 14139
- Li, F. *et al.* (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat. Genet.* 46, 567–572