Cotton Genomics and Genetics 2025, Vol.16, No.6, 259-268
http://cropscipublisher.com/index.php/cgg

Review and Progress **Open Access**

Structural Variations Drive Phenotypic Divergence in Upland and Pima Cotton

Pingping Yang, Xian Zhang, Shujuan Wang
Hainan Provincial Key Laboratory of Crop Molecular Breeding, Sanya, 572025, Hainan, China
Corresponding email: shujuan.wang@hitar.org

Cotton Genomics and Genetics, 2025, Vol.16, No.6 doi: 10.5376/cgg.2025.16.0026
Received: 01 Sep., 2025
Accepted: 11 Oct., 2025
Published: 02 Nov., 2025

Preferred citation for this article:
Yang P.P., Zhang X., and Wang S.J., 2025, Structural variations drive phenotypic divergence in upland and pima cotton, Cotton Genomics and Genetics, 16(6): 259-268 (doi: 10.5376/cgg.2025.16.0026)

Abstract Structural variations (SVs) are major genomic alterations that contribute to species diversity and phenotypic differentiation. In this review, we explored how SVs drive the phenotypic divergence between Upland cotton (*Gossypium hirsutum*) and Pima cotton (*Gossypium barbadense*), two economically significant species with distinct fiber characteristics. We first summarized advances in sequencing technologies that facilitate the detection and characterization of SVs and analyzed their types, frequency, and lineage-specific patterns across cotton genomes. We then discussed the functional impact of SVs on gene expression, dosage, and regulation, emphasizing their role in modifying fiber traits, stress tolerance, yield, and plant architecture. Mechanistic insights revealed that transposable elements, homologous recombination, and epigenetic modifications are key forces shaping SV formation and genome plasticity. A case study on a major inversion on chromosome A07 further demonstrated how SVs influence fiber quality and provide new opportunities for marker-assisted selection. Finally, we highlighted the integration of SV data into breeding and genome-editing programs to enhance cotton improvement. This review underscores the central role of structural variations in cotton evolution and breeding innovation, offering a genomic foundation for future research on trait diversification and molecular breeding strategies.

Keywords Structural variations; *Gossypium hirsutum*; *Gossypium barbadense*; Fiber quality; Genomic evolution

1 Introduction

In global cotton production, upland cotton (*Gossypium hirsutum*) and Pima cotton (*Gossypium barbadense*) almost occupy the entire market. The former is highly productive and hardy, and is the main force of the world's cotton. The latter has earned the reputation of "high-end cotton" by virtue of its long fibers and high strength. However, its output is not large and its market share is relatively limited (Lu et al., 2022). Interestingly, the story of these two types of cotton dates back over A million years-when an ancient polyploidization event led to the "fusion" of the ancestors of the A genome and the D genome, and then each embarked on an independent path of domestication, eventually forming the two types with distinct traits today (Yang et al., 2019).

When it comes to the diversity of cotton, the concept of "structural variation" (SVs) cannot be avoided. Insertions, deletions, inversions and translocations, these seemingly trivial changes, are precisely the key forces driving the genetic diversity of plants. They not only influenced the evolution of cotton, but also determined the quality of fibers, yield and even disease resistance (Jin et al., 2023). With the continuous improvement of sequencing technology, a large number of structural variations between upland cotton and island cotton, as well as within the same species, have been gradually revealed. These variations will rewrite the expression mode of genes and reshape the regulatory network, thereby causing differences among phenotypes (Ma et al., 2021). It is particularly worth noting that many of these variations are hidden in regions that are difficult for single nucleotide polymorphisms (SNPS) to reach-that is to say, they explain the traits that SNP analyses often "overlook".

This study focuses on a core issue: How exactly does structural variation drive the differentiation of land-based cotton and Pima cotton in terms of morphology and agronomic characteristics? We have integrated the achievements of high-quality genome assembly, pan-genome and genome-wide association studies to sort out the formation, function and potential value of SV in breeding from different perspectives. The focus is on how they affect fiber quality, yield performance and stress resistance, with the aim of providing new clues for future cotton improvement.

2 Genomic Landscape of Structural Variations in Cotton

2.1 Advances in sequencing technologies enabling SV detection

To truly understand the appearance of the cotton genome is not something that can be achieved at the very beginning. Although early sequencing technology solved the problem of "being able to see", it was far from being able to "see clearly". Especially those regions rich in repetitive sequences and close to the filaments have almost become blind spots where scientists have no idea where to start. It was not until recent years that the situation gradually began to improve. Technologies such as single-molecule real-time sequencing (SMRT), BioNano optical spectroscopy and Hi-C have been integrated and applied successively. Finally, the reference genomes of *Gossypium hirsutum* and *Gossypium barbadense* were fully pieced together (Chang et al., 2024). Those vague fragments of the past, the parts that were once unpieced together or simply ignored, have finally revealed their details this time. Subsequently, the addition of Oxford nanopore and the PacBio platform further enhanced the resolution, enabling the identification of even complex structures such as large areas of inversion and centromere repositioning. It is worth noting that the significance of these advancements goes far beyond the "more comprehensive" data. For the first time, they have enabled us to truly understand the structural differences between different lineages, and have also made comparative genomic research on cotton more solid and convincing.

2.2 Types and frequency of SVs in cotton genomes

The genome of cotton is no quiet world, where various variations occur one after another: insertions, deletions, inversions, translocations, copy number changes-they can be found on almost every chromosome. Studies have found that some variations are so extensive that they span large areas between or within the arms and can affect hundreds of genes at one time (Meng et al., 2023). Among these chromosomes, A08 is particularly special, with approximately one-third of the regions undergoing structural variations, often accompanied by recombination or gene loss. Meanwhile, genome-wide analysis also revealed tens of thousands of lineage-specific SNPS and SVS, which gradually widened the functional gap between upland cotton and island cotton. It should be noted that the distribution of variations is not uniform: some areas are like ignited "hotspots", with dense variations; Some areas, however, remain almost motionless, perhaps being "quiet zones" left over from long-term domestication or selection under pressure. This interwoven genomic pattern of cold and hot makes the evolutionary process of cotton not monotonous; instead, it is more like a dynamic map that is constantly being rewritten.

2.3 Comparative genomic analyses reveal lineage-specific SVs

If the differences between land cotton and island cotton are understood as a simple sequence rearrangement, it is obviously an underestimation of their complexity. After polyploidy, long-term domestication and natural selection have continuously left traces at the genomic level-large fragment inversions, translocations, and even centromeric repositioning have occurred successively during this period (Fan et al., 2020). These changes, layer upon layer, gradually reshaped the way genes were organized and also widened the gap in traits. Some variations are related to the expansion or expression bias of gene families, and ultimately manifest as differences in agronomic traits such as fiber quality and stress resistance. From a deeper perspective, structural variations seem to be involved in the evolution of centromeres and also affect the overall stability of the genome. Different families of repetitive sequences have formed their own independent centromeric structures in each subgenome, causing the two species to also present a pattern of "each going its own way" at the chromosomal level. It can be said that in the differentiation process of upland cotton and Pima cotton, structural variation was not an accidental side branch but a driving force that accompanied them all the way.

3 Functional Impact of SVs on Gene Expression and Regulation

3.1 SVs alter gene dosage and copy number

Sometimes, changes in gene expression are not caused by mutations themselves, but by alterations in the "quantity". Structural variations (SVs) such as insertions, deletions or duplications can directly change the gene dose or copy number, thereby pulling up or down the expression level. This phenomenon is not uncommon in cotton (Cheng and Zhang, 2025). Studies have found that hundreds and thousands of genes are affected by structural variations. Some genes disappear due to regional loss, while new genes emerge due to repetitive or presence/deletion variations (PAVs). When the dose of genes increases, their expression may also increase accordingly, especially for those genes involved in basal metabolism. However, there are exceptions-some dose-insensitive genes are often concentrated in stress responses or signaling pathways instead. These seemingly subtle dose differences are precisely one of the important driving forces behind the phenotypic differentiation between upland cotton and Pima cotton.

3.2 Disruption of regulatory elements by SVs

Not all variations occur within the genes themselves; more often than not, they appear in the "controlled regions" of the genes. Regulatory elements-such as promoters, enhancers or topologically associated domain (TAD) boundaries-can easily disrupt the balance of gene regulation once they are affected by structural variations (SVS). In the study of cotton, thousands of structural variations have been discovered in these regions, many of which are related to significant expression differences between upland cotton (*G. hirsutum*) and sea island cotton (*G. barbadense*) (You et al., 2022). For instance, a small 16 bp fragment was added to the promoter of the RLP6 gene in sea island cotton (which was not present in upland cotton). As a result, the expression of RLP6 was enhanced, and the resistance to yellow wilt was also improved accordingly (Figure 1) (Zhang et al., 2025). What is more complex is that SV can also reshape TAD boundaries, change the three-dimensional structure of the genome, and recombine originally adjacent or cooperative gene regulatory networks (Long et al., 2021). In addition, the replacement of transcription factor binding sites (TFBS) also occurred, further promoting the formation of new expression patterns, especially in fiber-related genes, which was most evident.

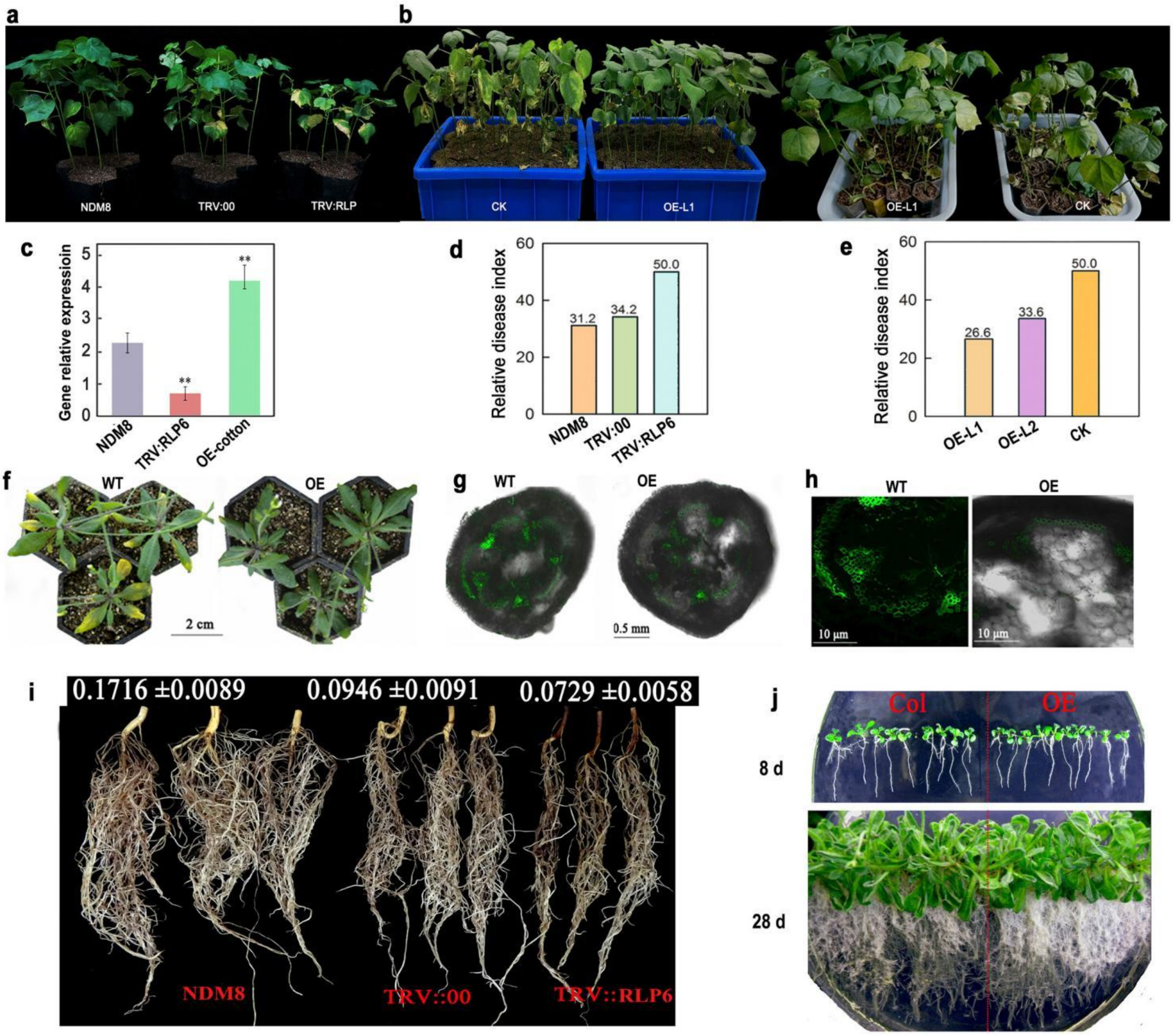

Figure 1 Functional analysis of*RLP6* gene via gene silencing and overexpression assays(Adopted from Zhang et al., 2025)

3.3 SV-mediated creation of novel gene fusions or isoforms

The arrangement of genes is not always regular. Occasionally, structural variations (SVS) can disrupt the original sequence, resplicing the boundaries between exons and introns, and even generating new splicing sites in unexpected places. The originally intact genes are thus "rewritten", with new fusion genes or alternative splicing isomers emerging. In cotton, this situation is actually not rare and is often related to the movement of the rotating seat component. Their insertion disrupts the splicing rhythm, enabling the same gene to produce multiple mRNA isomers and thereby generate different versions of proteins. The results of the comparative transcriptome also confirmed this point-many direct homologous genes showed significant differences in the number of isomers and splicing patterns (Shahzad et al., 2024). These differences are not merely minor variations but bring about linea-specific expression patterns and even new functions. Interestingly, the isomer changes triggered by SV are not confined to the developmental stage; they are also involved in regulating fiber growth and sometimes affect cotton's responsiveness under stress.

4 SVs and Phenotypic Traits in Upland and Pima Cotton

4.1 Fiber length, strength, and fineness

The difference in fiber quality between *Gossypium hirsutum* and *Gossypium barbadense* has long been well known, but it is not easy to explain the genetic basis of this difference. Researchers are increasingly inclined to believe that structural variations (SVs) play a key role in this. According to the results of genome-wide association studies (GWAS) and pan-genome studies, many variations significantly associated with traits such as fiber length, strength, and fineness are concentrated in the D subgenome (Chen et al., 2022). These variations are usually distributed near the genes that regulate fiber development. Their effects may not be direct, but they often accumulate continuously during the development process. Interestingly, although some mutations can bring about superior traits, they may not be retained. Take the *GhROPGEF5* gene as an example. The deletion of one of its bases can significantly increase fiber length and strength. However, this allele is not common in upland cotton because it may lead to a decrease in yield (Wang et al., 2024). In contrast, researchers are more inclined to adopt a "leveraging" approach-introducing the high-quality gene fragments of Pima cotton into upland cotton to improve the fiber properties. Through QTL mapping and transcriptome analysis, multiple candidate genes and their regulatory networks have been identified (Li et al., 2024). It is worth noting that many differences in fibrous traits that SNP analysis fails to capture are often caused by these structural variations deeply hidden in the genome.

4.2 Stress tolerance and environmental adaptation

The performance of upland cotton and Pima cotton in responding to environmental pressure is also not the same, and this difference is closely related to structural variation. In certain gene families, such as the NDA family, there is a significant association between internal structural variations or presence/absence variations (PAVs) and enhanced salt-alkali tolerance (Fan et al., 2022). In comparative genomic studies, the structural variation density of the D subgenome is higher, which may explain why terrestrial cotton (*G. hirsutum*) can adapt to a wider range of environments. Unlike Pima cotton, which is mainly concentrated in tropical and subtropical regions, upland cotton can even maintain a stable yield in high-latitude areas. Some studies have pointed out that this adaptability also benefits from the combined effect of gene infiltration and structural variation, which provide upland cotton with stronger disease resistance, such as resistance to Fusarium wilt (Wang et al., 2022). In addition, in many areas where cis-type elements related to light, hormones and environmental signal responses are located, structural variations often occur, enabling plants to adjust their physiological responses more flexibly to adapt to complex environments.

4.3 Yield components and plant architecture

Yield is not the result determined by a single gene, but rather a combination of traits influenced by structural variations. A large number of structural variations found in the A subgenome are closely related to traits such as cotton boll weight, cotton fluff rate and seed index. GWAS analysis has identified nearly a hundred SVS significantly associated with yield traits, many of which directly affect genes regulating plant type or reproductive development (Jiang et al., 2024). There is a typical example: by introducing the sub-okra leaf shape gene of sea island cotton into upland cotton, the canopy structure of short-season cotton varieties was optimized, the photosynthetic efficiency was improved, and ultimately the yield increased. What is more concerning is that when favorable SV haplotypes are aggregated, the originally unbalanced relationship between high yield and superior fibers may also be disrupted (Zhao et al., 2024). This means that by rationally utilizing structural variations, it is expected to achieve a variety improvement direction that takes into account both yield and quality.

5 Mechanistic Insights: How SVs Drive Divergence

5.1 Role of transposable elements in SV formation

In the cotton genome, there are always some components that are "restless". The turntable component (TE) is one of the most active types. They are like vagrants in the genome, from time to time jumping out of their original positions and inserting into new sites, disrupting the originally orderly sequence. Such "turmoil" may seem unregulated, but it often leads to far-reaching consequences-genomic expansion, changes in the number of genes, and even the reorganization of regulatory networks. Studies have shown that in diploid and polyploid cotton, the activation of long-end repeat (LTR) retrotransposons-particularly in the Gypsy and Copia families-is closely related to differences in genome size and has also given rise to some species-specific genes (He et al., 2024). Sometimes, a single TE insertion is sufficient to rewrite the fate of a gene: it may disrupt the original structure, causing an existence-deletion variation (PAV). It is also possible that due to carrying new regulatory fragments, the expression mode of genes is completely renewed (Tian et al., 2025). In addition, the activities of TE will also promote gene replication, fusion, and even facilitate new regulatory pathways. These changes have invisibly widened the gap in traits between upland cotton and Pima cotton. However, the TE activities of each lineage are not the same. Some are almost silent, while others will suddenly "explode" under multiple amplification or environmental stress. A single amplification is enough to accelerate the process of differentiation.

5.2 Homologous recombination and genome plasticity

The source of structural variation is not entirely due to the "disruption" of the turntable components. In cotton, reorganization is also a force that cannot be ignored. After polyploidy, mismatches or fragment exchanges often occur between homologous chromosomes. This recombination does not always occur at allelic sites; sometimes it even "accidentally collides" between non-allelic homologous sequences, thereby triggering complex structural changes such as inversions, deletions or duplications. The seemingly chaotic movement of genes has instead endowed the cotton genome with greater flexibility. As this plasticity increases, gene families begin to expand or contract, with some functions gradually lost while new ones are born. It is worth noting that non-homologous end joining and DNA repair mechanisms tend to be more active in stressful environments, which invisibly increase the frequency of structural variations (Krasileva, 2019). However, these genome-level adjustments are not isolated events. They are intertwined with the domestication process-selective pressure, on the one hand, eliminates unfavorable rearrangements, while on the other hand, retains variations that can enhance adaptability. Over time, the morphological and physiological differences between upland cotton and Pima cotton gradually took shape in such a dynamic balance.

5.3 Epigenetic consequences of SVs

When studying the cotton genome, people often focus first on sequence changes, while the epigenetic effects hidden behind them are frequently overlooked. But once you look deeper, you will find that their influence is not light. Especially when it comes to structural variations involving transposition elements (tes), such changes are often not merely about rewriting DNA sequences; they can also disrupt the state of local chromatin, causing alterations in methylation levels or the pattern of histone modifications. Take transposon-related variations (TRV) as an example. Differences in methylation often affect the transcriptional activity of adjacent genes, thereby influencing the development of cotton fibers (Liu et al., 2025). However, the problem is not limited to this. Structural variations sometimes also disrupt the boundaries of topologically associated domains (TAds), causing the three-dimensional structure of chromatin to refold and the connections between regulatory elements and target genes to be recombined. Such disturbances sometimes bring about stable expression patterns, while at other times they make regulation more flexible, enabling different lineages of cotton to respond more quickly to environmental changes and gradually form their own unique phenotypes.

6 Case Study: SVs in a Key Locus Influencing Fiber Quality

6.1 Identification of a major inversion on chromosome A07

Research on the quality of cotton fibers has always focused on chromosome A07. Researchers have discovered a significant structural variation on this chromosome-a major inversion, which is closely related to multiple fibrous traits. Using high-throughput resequencing technology, Inv(A07) p1.09-p2.23 inlocation was detected in dark brown fiber cotton. This structural change not only causes microdeletions near the breakpoints but also leads to the destruction or abnormal expression of some genes. It is worth noting that this inversion only occurs in dark brown fiber cotton and is significantly correlated with properties such as fiber color, length and strength (Figure 2) (Wen et al., 2020). The nucleotide diversity in the affected areas has declined, recombination has been suppressed, and linkage disequilibrium has intensified, suggesting that it may have special significance in evolution and breeding. In addition, there is a 1.4Mb inversion on chromosome A07 that is connected to the GhTT2_A07 gene. This variation activates the biosynthesis of flavonoids, thereby affecting fiber color and potential quality traits. Multiple QTL mapping and association analyses also repeatedly identified stable QTL clusters in the same region, suggesting that this might be a concentrated area of beneficial alleles.

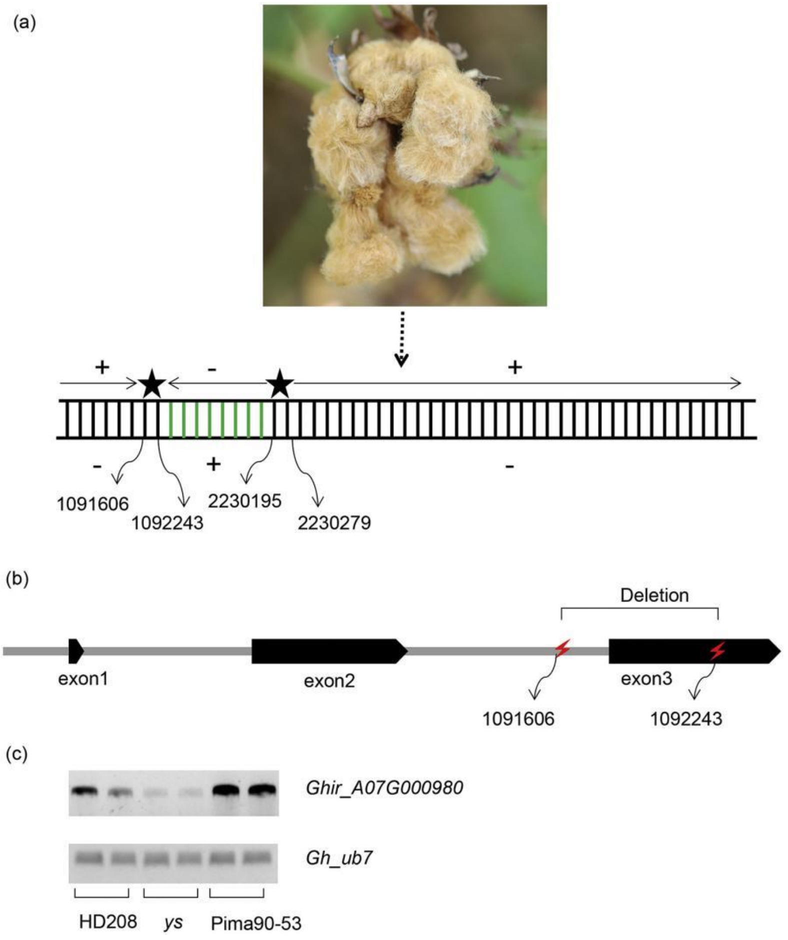

Figure 2 Inversion structure of Inv(A07)p1.09p2.23 andgene disruption of *Ghir_A07G000980*. (a) Inversion structure in the dark-brown fibre mutant. “+” and “-” indicate the sense and antisense strands in the reference genome, respectively. The stars indicate the truncated fragments. The green colour indicates the inverted fragment. The numbers means the physical position in genome. The arrows indicate the direction of alignment with the reference genome. (b) Structure of *Ghir_A07G000980*disrupted by breakpoints and micro-deletion. (c) Transcript of *Ghir_A07G000980* derived from RNA of HD208, *ys*, and Pima90-53. *Gh_ub7* was used as the control (Adopted from Wen et al., 2020)

6.2 Functional analysis and expression profiling

The inversion in the A07 region is not merely a "static decoration" in the genome; it does indeed alter the expression pattern of genes. Functional analysis revealed that some key genes were abnormally expressed in the context of inversion, such as Ghir_A07G000980. Both its functional disorder and expression imbalance were closely related to the differences in fiber development. To verify this, researchers introduced fragments of island cotton (*G. barbadense*) into the upland cotton (*G. hirsutum*) strain and conducted transcriptome analysis on them. The results showed that there were a number of differentially expressed genes in the A07 region, mainly involved in processes such as cell wall tissue formation, polysaccharide metabolism and hormone signal transduction (Lu et al., 2017; Song et al., 2020). These seemingly basic metabolic and regulatory activities actually directly affect the strength and elongation of fibers. It is worth noting that in this QTL cluster, GhRBB1_A07 has been confirmed as an important candidate gene for high-quality fibers. There is a significant association between its sequence variation and the change in expression level and the improvement of traits. In other words, the inversion of A07 is not merely a structural change, but also a molecular "clue" behind the difference in fiber quality.

6.3 Breeding implications and marker development

The significance of the inverted position of A07 is not limited to revealing the mechanism; it also brings practical value. Researchers screened out a batch of SSR and SNP loci near the QTL peak of chromosome A07 and developed a marker system that can be used for molecular marker-assisted selection (MAS) based on this (Fang et al., 2020). This enables breeders to screen high-quality fiber materials more efficiently and precisely. What is more concerning is that qFS-c7-1 QTL shows stable performance and strong transferability in different populations and generations, and has become a core reference marker for multiple breeding projects. Meanwhile, functional markers like GhRBB1_A07 have also been gradually established, which can help researchers identify and aggregate favorable fragments to achieve a balanced improvement in quality and yield (Li et al., 2019). It can be said that these breeding strategies based on structural variations are bringing the improvement of upland cotton and Pima cotton into a new stage that is more precise and efficient.

7 Applications in Breeding and Genetic Improvement

7.1 SV-informed breeding strategies

Cotton breeding does not solely rely on traditional genotype selection; the intervention of structural variations (SV) makes this process more targeted. Structure-level changes such as insertions, deletions, inversions, and presence/deletion variations (PAVs) often conceal clues that affect key traits. In recent years, with the establishment of high-quality reference genomes and the application of resequencing chips, genome-wide association studies (GWAS) have revealed hundreds of structural variations significantly associated with fiber quality, yield, and disease resistance (Liu et al., 2023). Interestingly, the SVS corresponding to different traits are not evenly distributed-for instance, those related to fibers are mostly in the D subgenome, while the signals for yield traits are concentrated in the A subgenome. Now, researchers are attempting to apply these SV markers or haplotype blocks (especially those derived from gene infiltration) to molecular breeding to guide the screening and combination of favorable alleles. In this way, not only can the genetic diversity of the population be maintained, but also the environmental adaptability of the breeding materials can be enhanced. Compared with the traditional selection method that only relies on SNPS, integrating SV data can help discover those complex but easily overlooked superior traits (Li et al., 2021).

7.2 Genome editing approaches targeting SV regions

In the past, structural variations were almost all "natural products", and people were more likely to observe and record them rather than intervene. It was not until the emergence of CRISPR/Cas technology that this situation was completely broken. Nowadays, researchers can not only manipulate genes but also precisely edit those regions related to structural variations (SVS). The CRISPR/Cas9 and Cas12a systems have been successfully applied in cotton, enabling the creation of deletions, insertions, and simultaneous modification of multiple sites (Sheri et al., 2025). This enables people to directly verify for the first time exactly what role a certain structural variation has, and even artificially create new alleles. It is worth noting that with the continuous optimization of the guide RNA design and verification system, the complex genomic structure of polyploid cotton is no longer an obstacle, and the editing efficiency has been significantly improved (Hui et al., 2024). It is precisely for this reason that the results of gene modification in the laboratory are gradually being brought into breeding practice to improve the quality, yield and stress resistance of fibers(Wang and Zhang, 2024). CRISPR technology is not intended to replace traditional breeding, but it has indeed turned precise improvement from theory into reality.

7.3 Pan-genomics and SV diversity

If genome sequencing has revealed the full picture of cotton genetics, then pan-genomics tells us that the genome of each variety is not the same. By integrating hundreds or even thousands of cotton germplasm data, researchers have drawn a more complete genetic picture: tens of thousands of structural variations have been identified, a considerable number of which are related to key traits such as fiber quality and yield. These works not only reveal the non-reference genes that were lost during domestication and selection, but also bring rare and linea-specific variations back into view. For breeders, this means being able to mine materials from a broader genetic pool. Meanwhile, the construction of graph-based pan-genomes and graph lineage genomes has also made the association between structural variations and phenotypes clearer. Such tools are helping to design a new generation of breeding strategies-those that can maximize the utilization of beneficial SV alleles are gradually becoming possible.

8 Concluding Remarks and Future Perspectives

After years of research, people have gradually come to realize that the genetic diversity of cotton is far more complex than imagined. The latest genomic sequencing and pan-genomic studies have shown that there are a large number of structural variations (SVs) in the genomes of upland cotton and Pema cotton, ranging from insertions and deletions to inversions and presence/deletion variations (PAVs), which are almost ubiquitous. These variations not only enrich the genomic structure but also directly promote phenotypic differentiation. The division of labor among different subgenomes in trait regulation is particularly evident: subgenome D is more associated with fiber quality, while subgenome A leans towards yield-related traits. Many SVS can also regulate gene expression and network structure. The trait differences that are often "invisible" in SNP analysis are often precisely what they are at play. Gene infiltration and long-term domestication have further shaped the distribution pattern of these variations and also left valuable allele resources for breeding.

However, there are also quite a few problems. The cotton genome is large and complex, especially for polyploid species, which often makes comprehensive detection and verification of structural variations difficult. Many SVS are located near repetitive sequences or centromeres-assembly and annotation work in these areas remains challenging to this day. What is more troublesome is that although the identification of SV is increasing, the functional analysis cannot keep up, and it is often still undetermined which variations truly affect the traits. Furthermore, the true application of SV data to the breeding process is still in its infancy. The current typing technology is costly and has limited throughput, and there is still a long way to go before it can be widely promoted in practice.

Future work should focus more on "filling in the gaps". Constructing a more complete and gap-free reference genome is crucial, especially enhancing the ability to identify SVS in complex genomic regions. Meanwhile, functional genomics methods, such as CRISPR editing or transcriptome analysis, can be employed to verify the specific impact of candidate SVS on gene expression and phenotype. Expanding the coverage of pan-genome resources and combining SV information with other genomic, epigenomic and phenotypic data is also expected to discover new superior alleles. Finally, if efficient and user-friendly SV typing and marker-assisted selection tools can be developed, these research achievements will be able to enter the actual breeding process more quickly, truly achieving the transformation from genomic discovery to field application.

Acknowledgments

We are grateful to Mr. Ma for critically reading the manuscript and providing valuable feedback that improved the clarity of the text.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Chang X., He X., Li J., Liu Z., Pi R., Luo X., Wang R., Hu X., Lu S., Zhang X., and Wang M., 2024, High-quality *Gossypium hirsutum* and *Gossypium barbadense* genome assemblies reveal the landscape and evolution of centromeres, Plant Communications, 5(2): 100722. https://doi.org/10.1016/j.xplc.2023.100722

Chen H., Han Z., Ma Q., Dong C., Ning X., Li J., Lin H., Xu S., Li Y., Hu Y., Si Z., and Song Q., 2022, Identification of elite fiber quality loci in upland cotton based on the genotyping-by-target-sequencing technology, Frontiers in Plant Science, 13: 1027806.
https://doi.org/10.3389/fpls.2022.1027806

Cheng J.H., and Zhang J., 2025, High-yield cotton cultivation practices in arid regions, Molecular Soil Biology, 16(1): 27-36.
https://doi.org/10.5376/msb.2025.16.0003

Fan K., Mao Z., Zheng J., Chen Y., Li Z., Lin W., Zhang Y., Huang J., and Lin W., 2020, Molecular evolution and expansion of the KUP family in the allopolyploid cotton species *Gossypium hirsutum* and *Gossypium barbadense*, Frontiers in Plant Science, 11: 545042.
https://doi.org/10.3389/fpls.2020.545042

Fan Y., Zhang Y., Rui C., Zhang H., Xu N., Wang J., Han M., Lu X., Chen X., Wang D., Wang S., Guo L., Zhao L., Huang H., Wang J., Sun L., Chen C., and Ye W., 2022, Molecular structures and functional exploration of NDA family genes responding to alkaline stress in *Gossypium hirsutum* L., Biological Research, 55(1): 4.
https://doi.org/10.1186/s40659-022-00372-8

Fang D., Zeng L., Thyssen G., Delhom C., Bechere E., Jones D., and Li P., 2020, Stability and transferability assessment of the cotton fiber strength QTL qFS-c7-1 on chromosome A07, Crop Journal, 9(2): 380-386.
https://doi.org/10.1016/j.cj.2020.06.016

He X., Qi Z., Liu Z., Chang X., Zhang X., Li J., and Wang M., 2024, Pangenome analysis reveals transposon-driven genome evolution in cotton, BMC Biology, 22(1): 92.
https://doi.org/10.1186/s12915-024-01893-2

Hui F., Tang X., Li B., Alariqi M., Xu Z., Meng Q., Hu Y., Wang G., Zhang Y., Zhang X., and Jin S., 2024, Robust CRISPR/Mb2Cas12a genome editing tools in cotton plants, iMeta, 3(3): e209.
https://doi.org/10.1002/imt2.209

Jiang H., Ma X., Shi J., Gao M., Zhang X., Zhang C., Chai Q., Wang Y., Wang X., Wang J., Chen Y., Zhang D., Li F., Zhao W., and Zhao J., 2024, Sub-okra leaf shape conferred via chromosomal introgression from *Gossypium barbadense* L. improves photosynthetic productivity in short-season cotton (*Gossypium hirsutum* L.), Frontiers in Plant Science, 15: 1393396.
https://doi.org/10.3389/fpls.2024.1393396

Jin S., Han Z., Hu Y., Si Z., Dai F., He L., Cheng Y., Li Y., Zhao T., Fang L., and Zhang T., 2023, Structural variation-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons, Molecular Plant, 16(4): 678-693.
https://doi.org/10.1016/j.molp.2023.02.004

Krasileva K., 2019, The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes, Current Opinion in Plant Biology, 48: 18-25.
https://doi.org/10.1016/j.pbi.2019.01.004

Li J., Yuan D., Wang P., Wang Q., Sun M., Liu Z., Si H., Xu Z., Y. H., Zhang B., Pei L., Tu L., Zhu L., Chen L., Lindsey K., Zhang X., Jin S., and Wang M., 2021, Cotton pan-genome retrieves the lost sequences and genes during domestication and selection, Genome Biology, 22(1): 119.
https://doi.org/10.1186/s13059-021-02351-w

Li P., Chen Y., Yang R., Sun Z., Ge Q., Xiao X., Yang S., Li Y., Liu Q., Zhang A., Xing B., Wu B., Du X., Liu X., Tang B., Gong J., Lu Q., Shi Y., Yuan Y., Peng R., and Shang H., 2024, Co-expression network analysis and introgressive gene identification for fiber length and strength reveal transcriptional differences in 15 cotton chromosome substitution segment lines and their upland and sea island parents, Plants, 13(16): 2308.
https://doi.org/10.3390/plants13162308

Li S., Liu A., Kong L., Gong J., Li J., Gong W., Lu Q., Li P., Ge Q., Shang H., Xiao X., Liu R., Zhang Q., Shi Y., and Yuan Y., 2019, QTL mapping and genetic effect of chromosome segment substitution lines with excellent fiber quality from *Gossypium hirsutum* × *Gossypium barbadense*, Molecular Genetics and Genomics, 294(5): 1123-1136.
https://doi.org/10.1007/s00438-019-01566-8

Liu S., Cheng H., Zhang Y., He M., Zuo D., Wang Q., Lv L., Lin Z., Liu J., and Song G., 2025, Cotton transposon-related variome reveals roles of transposon-related variations in modern cotton cultivation, Journal of Advanced Research, 71: 17-28.
https://doi.org/10.1016/j.jare.2024.05.019

Liu S., Zuo D., Cheng H., He M., Wang Q., Lv L., Zhang Y., Ashraf J., Liu J., and Song G., 2023, Cotton pedigree genome reveals restriction of cultivar-driven strategy in cotton breeding, Genome Biology, 24(1): 282.
https://doi.org/10.1186/s13059-023-03124-3

Long Y., Liu Z., Wang P., Yang H., Wang Y., Zhang S., Zhang X., and Wang M., 2021, Disruption of topologically associating domains by structural variations in tetraploid cottons, Genomics, 113(5): 3405-3414.
https://doi.org/10.1016/j.ygeno.2021.07.023

Lu Q., Shi Y., Xiao X., Li P., Gong J., Gong W., Liu A., Shang H., Li J., Ge Q., Song W., Li S., Zhang Z., Rashid M., Peng R., Yuan Y., and Huang J., 2017, Transcriptome analysis suggests that chromosome introgression fragments from sea island cotton (*Gossypium barbadense*) increase fiber strength in upland cotton (*Gossypium hirsutum*), G3: Genes, Genomes, Genetics, 7(10): 3469-3479.
https://doi.org/10.1534/g3.117.300108

Lu X., Chen X., Wang D., Yin Z., Wang J., Fu X., Wang S., Guo L., Zhao L., Cui R., Dai M., Rui C., Fan Y., Zhang Y., Sun L., Malik W., Han M., Chen C., and Ye W., 2022, A high-quality assembled genome and its comparative analysis decode the adaptive molecular mechanism of the number one Chinese cotton variety CRI-12, GigaScience, 11: giac019.
https://doi.org/10.1093/gigascience/giac019

Ma Z., Zhang Y., Wu L., Zhang G., Sun Z., Li Z., Jiang Y., Ke H., Chen B., Liu Z., Gu Q., Wang Z., Wang G., Yang J., Wu J., Yan Y., Meng C., Li L., Li X., Mo S., Wu N., L. H., Chen L., Zhang M., Si A., Yang Z., Wang N., Wu L., Zhang D., Cui Y., Cui J., Lv X., Li Y., Shi R., Duan Y., Tian S., and Wang X., 2021, High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement, Nature Genetics, 53(9): 1385-1391.
https://doi.org/10.1038/s41588-021-00910-2

Meng Q., Gu J., Xu Z., Zhang J., Tang J., Wang A., Wang P., Liu Z., Rong Y., Xie P., Hui L., Udall J., Grover C., Wendel J., Jin S., Zhang X., and Yuan D., 2023, Comparative analysis of genome sequences of the two cultivated tetraploid cottons, *Gossypium hirsutum* (L.) and G. barbadense (L.), Industrial Crops and Products, 196: 116471.
https://doi.org/10.1016/j.indcrop.2023.116471

Shahzad K., Zhang M., Mubeen I., Zhang X., Guo L., Qi T., Feng J., Tang H., Qiao X., Wu J., and Xing C., 2024, Integrative analyses of long and short-read RNA sequencing reveal the spliced isoform regulatory network of seedling growth dynamics in upland cotton, Functional & Integrative Genomics, 24(5): 156.
https://doi.org/10.1007/s10142-024-01420-0

Sheri V., Mohan H., Jogam P., Alok A., Rohela G., and Zhang B., 2025, CRISPR/Cas genome editing for cotton precision breeding: mechanisms, advances, and prospects, Journal of Cotton Research, 8(1): 4.
https://doi.org/10.1186/s42397-024-00206-w

Song Z., Chen Y., Zhang C., Zhang J., Huo X., Gao Y., Pan A., Du Z., Zhou J., Zhao Y., Liu Z., Wang F., and Zhang J., 2020, RNA-seq reveals hormone-regulated synthesis of non-cellulose polysaccharides associated with fiber strength in a single-chromosomal-fragment-substituted upland cotton line, The Crop Journal, 2020, 8(2): 273-286.
https://doi.org/10.1016/j.cj.2019.11.003

Tian X., Wang R., Liu Z., Lu S., Chen X., Zhang Z., Liu F., Li H., Zhang X., and Wang M., 2025, Widespread impact of transposable elements on the evolution of post-transcriptional regulation in the cotton genus *Gossypium*, Genome Biology, 26(1): 60.
https://doi.org/10.1186/s13059-025-03534-5

Wang J.M., and Zhang J., 2024, Assessing the impact of various cotton diseases on fiber quality and production, Field Crop, 7(4): 212-221.
https://doi.org/10.5376/fc.2024.07.0021

Wang P., Dong N., Wang M., Sun G., Jia Y., Geng X., Liu M., Wang W., Pan Z., Yang Q., Li H., Wei C., Wang L., Zheng H., He S., Zhang X., Wang Q., and Du X., 2022, Introgression from *Gossypium hirsutum* is a driver for population divergence and genetic diversity in *Gossypium barbadense*, Plant Journal, 110(3): 764-780.
https://doi.org/10.1111/tpj.15702

Wang W., Liu D., Zhang T., Guo K., Liu X., Liu D., Chen L., Yang J., Teng Z., Zou Y., J. C., Wang Y., Yang X., Guo X., Sun X., Zhang J., Xiao Y., Paterson A., and Zhang Z., 2024, Natural variation in *GhROPGEF5* contributes to longer and stronger cotton fibers, New Phytologist, 245(3): 1090-1105.
https://doi.org/10.1111/nph.20286

Wen T., Yao T., You C., and Lin Z., 2020, A case study of a micro-inversion event in dark brown fibre cotton (*Gossypium hirsutum*), Crop Journal, 8(4): 577-585.
https://doi.org/10.1016/j.cj.2020.02.002

Yang Z., Ge X., Yang Z., Qin W., Sun G., Wang Z., Li Z., Liu J., Wu J., Wang Y., Lu L., Wang P., Mo H., Zhang X., and Li F., 2019, Extensive intraspecific gene order and gene structural variations in upland cotton cultivars, Nature Communications, 10(1): 2989.
https://doi.org/10.1038/s41467-019-10820-x

You J., Lin M., Liu Z., Pei L., Long Y., Tu L., Zhang X., and Wang M., 2022, Comparative genomic analyses reveal cis-regulatory divergence after polyploidization in cotton, Crop Journal, 10(6): 1545-1556.
https://doi.org/10.1016/j.cj.2022.03.002

Zhang D., Wang Y., Gu Q., Liu L., Wang Z., Zhang J., Meng C., Yang J., Zhang Z., Z. L., Wang X., and Zhang Y., 2025, Cotton RLP6 interacts with NDR1/HIN6 to enhance Verticillium wilt resistance via altering ROS and SA, Molecular Plant Pathology, 26(1): e70052.
https://doi.org/10.1111/mpp.70052

Zhao Y., Chen B., Li H., Wang J., Jia Y., Pan Z., Hu D., Peng Z., Li Y., Gao X., Zhang P., Wang L., Peng J., He S., and Du X., 2024, Pyramiding effects of favorable haplotypes of loci on major fiber yield and quality traits in upland cotton (*Gossypium hirsutum* L.), Industrial Crops and Products, 217: 118805.
https://doi.org/10.1016/j.indcrop.2024.118805