Cotton Genomics and Genetics 2025, Vol.16, No.4, 163-172
http://cropscipublisher.com/index.php/cgg

Feature Review **Open Access**

Evolutionary Analysis of Fiber-Related Gene Families in Diploid and Polyploid Cotton

Delong Wang, Jin Zhou, Jin Zhang
Hainan Provincial Key Laboratory of Crop Molecular Breeding, Sanya, 572025, Hainan, China
Corresponding email: jin.zhang@hitar.org

Cotton Genomics and Genetics, 2025, Vol.16, No.4 doi: 10.5376/cgg.2025.16.0016
Received: 01 May, 2025
Accepted: 10 Jun., 2025
Published: 01 Jul., 2025

Preferred citation for this article:
Wang D.L., Zhou J., and Zhang J., 2025, Evolutionary analysis of fiber-related gene families in diploid and polyploid cotton, Cotton Genomics and Genetics, 16(4): 163-172 (doi: 10.5376/cgg.2025.16.0016)

Abstract Cotton is a globally important fiber crop. Both its diploid (varieties with normal chromosomes, such as Asiatic and Raymondia cotton) and polyploid (varieties with doubled chromosomes, such as Upland and Sea Island cotton) varieties provide valuable genetic resources for improving cotton fiber quality. This study, using comparative genomic approaches, conducted a detailed evolutionary analysis of gene families involved in fiber growth across different cotton varieties, primarily including cellulose synthase (CesA) genes responsible for cellulose synthesis, expansins, and the MYB transcription factor family. We also investigated how these gene families evolved through whole-genome duplications, small-scale duplications, and external selective pressures. Furthermore, we analyzed the expression of these genes at different stages of fiber development, their epigenetic regulation (influencing development by modulating gene activity), and their co-expression networks (how genes coordinate their function). Furthermore, we conducted an in-depth case study of the CesA gene in Upland cotton, including its classification, expression patterns during development, and functional validation. These studies can help us more clearly understand the evolution and functional differences of cotton fiber genes, and have practical significance for future molecular breeding, gene editing and protection of cotton genetic resources.

Keywords Cotton genomics; Fiber-related genes; Gene duplication; Polyploidy; Molecular breeding

1 Introduction

Everyone knows cotton is important, especially in agriculture. It serves as a fiber raw material and plays a fundamental role in the economies of many countries. However, its continued research stems from more than just its value. Its fiber structure is unique, making it a readily available model for those studying plant cell growth and cell walls (Yu, 2024). Furthermore, genetic improvements related to cotton yield, quality, and stress resistance ultimately revolve around these structures and mechanisms (Hu et al., 2019).

In the past, no one wanted to study the cotton genome. Honestly, it's incredibly complex and daunting. Strangely enough, this complexity has actually benefited research, as it allows for much deeper investigation. In recent years, gene sequencing technology (the process of determining the specific information about a gene) has advanced rapidly, achieving dramatically higher accuracy. Now, we can even directly visualize what genes are doing (Chen et al., 2020; Pan et al., 2020), something we never dared to imagine before. This is especially true when studying fiber-related genes. Problems that were once the most daunting are now being considered as breakthroughs. Of course, this isn't entirely due to improved technology; research tools are now plentiful. For example, common techniques like assembling genomes, analyzing the evolutionary relationships of gene families, and comparing gene expression patterns are now available. Simply put, cotton researchers are truly living in a golden age.

However, it would be wrong to say that these complexities are inherent to cotton. In fact, its genome has undergone a complex evolution. Initially a simple diploid, it subsequently doubled in size, becoming polyploid, and subgenomes were incorporated. This gradual accumulation of diversity has led to a natural diversification of fiber expression. Previously, polyploidy was daunting, often characterized by gene redundancy and structural overlap, making it difficult to understand (Zhang and Yang, 2024). This situation is no longer the case.

Research isn't about who lists the most genes; it's about whether these genes behave consistently across different cotton varieties. Some are normal in diploids, while others appear as extra copies in polyploid offspring, or even disappear altogether. Others change their "identity," taking on new roles, and even their regulatory mechanisms may be completely disrupted. These subtle changes at the genetic level may seem trivial, but they may be the underlying reason for the current improvement in fiber quality (You et al., 2022).

The goal of this study, therefore, is not to discover entirely new genes from scratch, but rather to integrate existing research and understand the key changes that have occurred in these fiber-related gene families during their evolution. How did they expand? Is there a clear division of labor between the two subgenomes? How did the regulatory logic change? In particular, in the context of polyploidization coupled with artificial selection, how were some key traits progressively driven? We hope to provide more reliable references for subsequent cotton genetic improvement through the integration of genomic, transcriptomic and functional studies.

2 Genomic Background of Cotton

2.1 Diploid cotton genomes (e.g., *G. arboreum*, *G. raimondii*)

G. arboreum and *G. raimondii* are considered the ancestors of modern cultivated cotton. With one A genome and one D genome, they diverged over millions of years, each pursuing its own path. Logically, given such a long time, their differences should be significant, but what happens? While their genome sizes differ, the positions of many genes still match remarkably. While there are a significant number of single-nucleotide polymorphisms (SNPs), 24 million, the overall structure remains relatively stable, with a relatively stable gene arrangement. Of course, there are some variations-some DNA is lost, and transposons have jumped around-but the overall framework remains.

2.2 Polyploid cotton genomes (e.g., *G. hirsutum*, *G. barbadense*)

In fact, the upland cotton and sea island cotton we commonly cultivate were originally hybridized from two completely different wild cotton species. In the earliest times, there were no tetraploid cottons (those with doubled chromosome sets). Instead, there were two different wild cotton species, each with its own genome (called the A genome and the D genome). Somehow, these two species came together, and their chromosome counts unexpectedly doubled, resulting in the allotetraploid we know today (Figure 1) (Li et al., 2024). This situation is usually quite troublesome, as the combination of two genes creates a mess of duplicated genes, disrupting the chromosomes and even disrupting gene expression (the process by which genes function). But cotton species are accustomed to this chaotic environment. Furthermore, they sometimes incorporate gene fragments from their wild relatives, patching them together like a patchwork, creating a surprisingly effective genetic combination. Is this a coincidence? Ultimately, this combination has helped cotton grow better and become much more adaptable to various environments.

2.3 Comparative genome analysis

Many details can't be discerned from a single genome. Comparing diploid and polyploid genomes side by side reveals which regions are "stabilized" and which are "turbulent." In some regions, the gene order is remarkably stable, much as it was millions of years ago; yet in others, the sequence is chaotic-inversions, translocations, and regulatory element shifts are common, especially in regions involved in fiber development. The same gene may perform completely different functions in the At and Dt subgenomes, one a primary operator and the other a minor one. Digging deeper, even more subtle differences emerge in the transcriptome and epigenetics, affecting not only expression levels but also chromatin folding (Wang et al., 2018; Han et al., 2022). These seemingly minor changes, however, are precisely what transformed cotton from a wild species into the high-yielding crop it is today, thanks to these "fine-tuning" changes.

3 Key Fiber-Related Gene Families

3.1 Cellulose synthase (CesA) gene family

Cotton fibers contain a large amount of cellulose, a component that becomes particularly important in later stages, when the fibers thicken or stiffen. Cellulose doesn't just form out of thin air; its synthesis relies on a specialized set of genes. Three genes, *GhCesA4*, *GhCesA7*, and *GhCesA8*, have been identified as crucial in this process.

However, these genes don't actually take the lead; they often simply "execute tasks." The timing and quantity of cellulose synthesis are orchestrated by MYB family genes like *GhMYB7* and *GhMYBL1* (Huang et al., 2021). However, MYB genes themselves are not always stable. Sometimes, when their rhythm is disrupted, the underlying *CesA* gene can malfunction: either causing unstoppable fiber growth or causing extremely thin cell walls, ultimately affecting the overall fiber quality (Wang et al., 2023b). So, while *CesA* genes play a significant role in cellulose synthesis, they are essentially under orders and cannot act on their own. Ultimately, fiber quality depends on the coordination of this regulatory system.

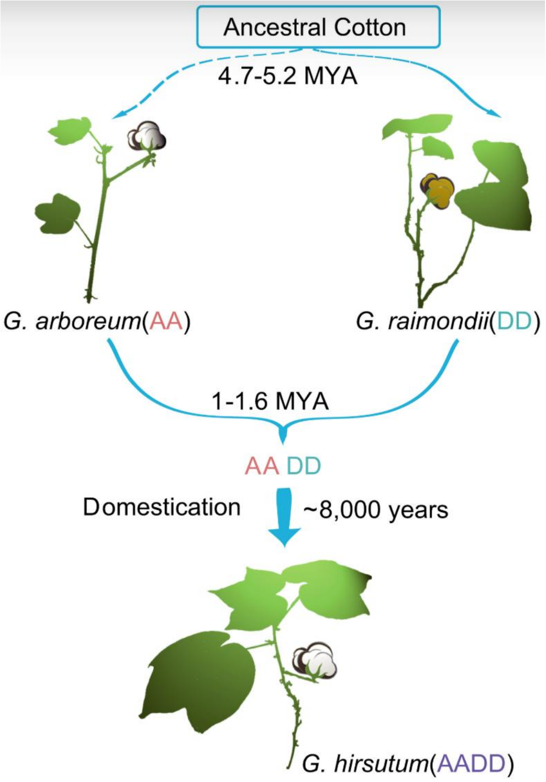


Figure 1 Schematic model of the evolutionary history of diploid and allotetraploid cotton (Adapted from Han et al., 2022)

3.2 Expansin gene family

In addition to cellulose, another class of "supporting forces" plays a crucial role in the elongation phase: expansins. These proteins' job is straightforward: loosen the cell wall to allow the cell to expand outward. Ninety-three related genes have been identified in cotton, but not all of them are always on call. Some, such as *GhEXPA4a* and *GhEXPA13a*, are particularly explosive during the rapid elongation phase (Lv et al., 2020); others, on the other hand, either come on early or quietly "come online" once the cell wall begins to thicken. They aren't always online, but as soon as their expression levels rise, the "jointing" stretching state of the cell often occurs, indicating that they play a role in regulating elongation.

3.3 MYB transcription factors

How do cotton fibers grow? Simply put, it's a group of genes that rotate in and perform their respective roles. The MYB gene family is particularly active within this group, controlling key processes. For example, some members are busy promoting cellulose synthesis, while others participate in fatty acid production, both of which are crucial for fiber structure (Suo et al., 2024). However, not every member works equally hard. GhMYB7 and GhMYB25 (Wu et al., 2018) are very active, often working on the front lines to spur the CesA genes to work. Meanwhile, GhMYB4 is more passive, slowing fiber growth and being considered a "speed reducer" (Duan et al., 2024). While one might imagine these opposing genes fighting, they don't. Instead, they appear at different times and don't interfere with each other. While these genes appear to have distinct personalities, they actually work in perfect harmony. It's this orderly rotation that allows for smooth fiber development. What may seem contradictory is actually the result of coordinated cooperation.

4 Evolutionary Mechanisms Driving Gene Family Diversification

4.1 Whole-genome duplication and gene retention

The emergence of upland cotton (*Gossypium hirsutum*) is actually the result of several major genome duplications (Wei et al., 2022). While this seemingly spectacular duplication of an entire gene set may seem, not every gene survives. Many replicated genes disappear after a while, leaving key genes involved in fiber development more likely to survive. These surviving genes are often grouped in pairs, located in similar locations within the At and Dt subgenomes. However, these repeated genome duplications have left some residual effects (Tao et al., 2021). Occasionally, genes that are usually quiet will suddenly "wake up" and stir in seemingly erratic patterns, but perhaps cotton's new characteristics emerge from this unstable state.

4.2 Tandem and segmental duplications

The expansion of the cotton genome doesn't necessarily require the large-scale duplication of entire chromosomes; it's more often the result of the gradual addition of small fragments. Gene families like the WAK and RING genes (which play important roles in plant growth regulation and stress tolerance (Shuya et al., 2023)) have gradually expanded through this type of localized duplication. While these small duplications may seem inactive during normal times, they can be highly effective in challenging environments like drought or salinity (Qiao et al., 2019), helping cotton cope with these challenges. Another type of "jumping genes" called transposable elements, such as Copia and Gypsy, are usually quite quiet, but once activated, they can roam the genome, sometimes carrying other genes along with them, like moving a whole load of luggage. While this seemingly chaotic movement may seem chaotic, this "order within chaos" may be precisely how the cotton genome evolves. What may seem unstable may actually be an opportunity for growth.

4.3 Selection pressure and functional divergence

Whether a gene persists after duplication depends on environmental support. Sometimes genes that maintain their original functions are more likely to survive (Malik et al., 2020), but others may branch out and develop new functions (Wu et al., 2024). In polyploid cotton (those with chromosomes duplicated in pairs), the situation is even more complex. Many genes have "backups," allowing mutations to be easily masked if one fails. However, this "masking" can surprisingly allow some harmful mutations to slip through. The At and Dt subgenomes have different abilities to handle these mutations, and over time, the differences between them have widened (Conover and Wendel, 2021). Recent studies have also revealed that some unique genes are not found in all cotton varieties, but are present only in certain varieties (Li et al., 2021). Although these genes are few in number and generally inconspicuous, they can be crucial, such as in the breeding of high-quality cotton.

5 Gene Expression and Regulatory Networks in Fiber Development

5.1 Stage-specific expression of fiber genes

Not all genes in cotton fiber production begin to function simultaneously. While many genes do indeed begin to activate immediately after pollination, many don't show significant changes until 16~17 days after flowering (Grover et al., 2024). Surprisingly, researchers have found that despite numerous experiments, only a few genes are consistently expressed across different cotton varieties (You et al., 2023). One gene, GhRALF1, which regulates cotton growth rhythms, is quite unusual (Wang et al., 2023), being particularly active during the day and inactive at night. This diurnal shift may affect fiber growth rates, or it may simply reflect the cotton plant's natural rhythm; the answer remains elusive.

5.2 Epigenetic regulation and non-coding RNAs

In addition to the genes themselves, several small molecules also secretly participate in cotton fiber development. For example, siRNAs, microRNAs that don't produce protein (but can control the activation of other genes), despite their small size, play a significant role in regulating key genes like MYB-MIXTA. Another type of long noncoding RNA, called lncRNAs, such as the recently mentioned MST23 (Wang et al., 2024), is usually less noticeable, but by activating the GhKCR2 gene, which controls fatty acid synthesis, it can indirectly influence the rate of fiber growth. Furthermore, epigenetic mechanisms such as DNA methylation and histone modifications (which regulate gene activity without altering the DNA sequence) also play a subtle role behind the scenes.

Although difficult to directly observe, these mechanisms can significantly regulate gene expression, sometimes even influencing the entire growth process. These seemingly insignificant factors actually have a significant impact on the rhythm of cotton fiber development.

5.3 Gene co-expression networks

Cotton fiber formation isn't a single gene's sole responsibility; it relies on the coordinated efforts of many genes. These genes typically follow a specific order of appearance, enabling them to coordinate steps like regulating cell turgor or cellulose deposition. However, in polyploid cotton (those with doubled chromosome composition), the situation is much more complex (Xiong et al., 2024). For example, the A subgenome primarily plays a coordinating regulatory role, while the D subgenome tends to express related genes at higher levels. However, this system is not entirely stable. Sometimes, foreign gene fragments inexplicably infiltrate (Chen et al., 2024), disrupting the original order of expression and causing the previously coordinated gene network to suddenly become somewhat disordered. While these "queue-jumping" changes are uncommon, they can indeed affect the normal rhythm of fiber development.

6 Case Study: Cellulose Synthase Genes in *G. hirsutum*

6.1 Gene identification and classification

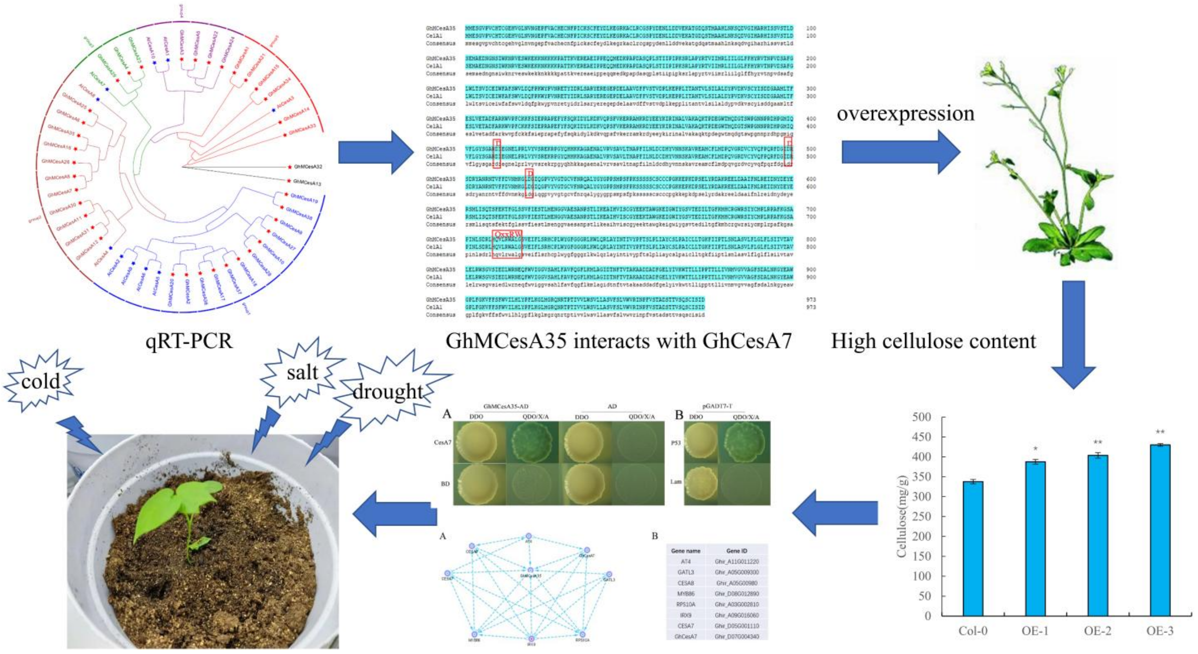
In the genome of upland cotton, the *CesA* gene can be regarded as one of the genes that are "given special attention." As soon as the early genome sequencing results came out, researchers locked in about 30 to 38 genes related to it. In fact, the arrangement of these genes can also be found in similar templates in other cotton plants, and the positional relationship is quite regular. Simply put, people divide them into two categories: one category is biased towards the construction of the primary cell wall, such as *GhCesA3*, *GhCesA5* and *GhCesA6*, and the other category is more active in the secondary wall stage, such as *GhCesA1*, *GhCesA2*, *GhCesA7* and *GhCesA8*. But classification is never clear-cut, and some *CesA* will also "cross the border" and play a role, and it is not uncommon for them to run both ends sooner or later.

6.2 Expression analysis during fiber development

These *CesA* genes aren't simply a one-on-one, one-off operation. Take *GhCesA1* and *GhCesA2*, for example. They typically become busy around 15 days after flowering, coinciding with the rapid elongation of fibers and the rapid increase in cellulose production. However, later steps require the help of other genes, such as *GhCesA4*, *GhCesA7*, and *GhCesA8*. These genes typically enter the scene as the secondary wall begins to thicken, serving as a kind of "reinforcement team," primarily responsible for maintaining fiber stiffness and strength. They don't work independently; more often, they work together as a large team, forming a so-called "supercomplex" to collectively complete cellulose synthesis (Figure 2). Those like *GhCesA2*, which activate early, may be crucial for the final fiber quality (Zhang et al., 2021). However, these rhythm arrangements are not static, and there may be some differences in different varieties, but the overall routine of "early start, mid-term outbreak, and late reinforcement" is quite methodical. In the final analysis, it seems like a formation designed in advance.

6.3 Functional characterization and validation

Expression is expression, but simply shouting slogans isn't enough. Whether these genes actually function requires real-world testing. Researchers simply eliminated one of *GhCesA4*, *GhCesA7*, or *GhCesA8*, and the results were unsurprising: the fiber secondary wall immediately became thinner, cellulose yield decreased, and quality naturally declined (Wen et al., 2022). Furthermore, despite similar names, these genes don't really work together. To verify this, researchers conducted overexpression and complementation experiments in *Arabidopsis thaliana*. The conclusion was clear: replacing them with other "close relatives" was ineffective; a corresponding homologous gene was required (Zhao et al., 2022). Furthermore, the promoter region of *GhCesA4* is quite interesting. It contains several regulatory elements that both respond to hormone signals and control tissue expression. This is like pre-installing a "time reminder" and "location navigation" for the genes, preventing the entire expression process from going astray or being accidentally triggered. Ultimately, this layered defense is designed to ensure that fibers grow in an orderly manner.



Gene name	Gene ID
AT4	Ghir_A11G011220
GATL3	Ghir_A05G009300
CESA8	Ghir_A05G00980
MYB86	Ghir_D08G012890
RPS10A	Ghir_A03G002810
IRX9	Ghir_A09G016060
CESA7	Ghir_D05G001110
GhCesA7	Ghir_D07G004340

Figure 2 The role of the interaction between GhMCesA35 and GhCesA7 in plant response to stress and cellulose synthesis (Adapted from Zhao et al., 2022)

7 Implications for Cotton Breeding and Biotechnology

7.1 Molecular breeding applications

In the past, cotton seed selection relied on experience, but today's breeding is no longer based on intuition. Technologies such as DNA markers, QTL mapping, and genome-wide association analysis have become the "essentials" of molecular breeding (Zhang et al., 2020; Kushanov et al., 2021). Marker-assisted selection (MAS) is particularly widely used because it allows breeders to directly identify cotton plants with high potential at the DNA level, eliminating the need to wait until cotton fluffs form to assess their performance (Shahzad et al., 2022). This naturally saves considerable time. The current trend is to combine high-throughput genotyping technologies with trait analysis, leveraging identified candidate genes and QTL loci to directly identify varieties with outstanding yield, fiber quality, or resistance (Yang et al., 2022; Joshi et al., 2023).

7.2 Genetic engineering and CRISPR-based approaches

Not every trait can be slowly selected through traditional breeding, especially in a complex crop like cotton. Sometimes, when faced with insufficient insect resistance, suboptimal fiber quality, or poor drought and salt tolerance, traditional methods prove insufficient. In these cases, genetic engineering becomes a shortcut. Gene-editing tools like CRISPR have seen extensive use in cotton in recent years (Wang et al., 2023a; Ahmed et al., 2024). Both CRISPR/Cas9 and Cas12 can not only target specific DNA segments but also create mutant libraries, effectively allowing researchers to test all suspected genes that may influence traits (Thangaraj et al., 2024; Wang et al., 2024). Of course, some goals are to enhance expression, while others aim to "delete" negative regulatory factors. Want to prevent a gene from doing all the harm it can? Simply delete it. A further step is to develop “edited” varieties without foreign DNA, thus circumventing the regulatory barriers that govern genetically modified crops (Khan et al., 2023; Kumar et al., 2024). In other words, this would make “modification” more covert and more efficient.

7.3 Conservation and utilization of gene diversity

Take a casual stroll through a cotton field these days and you'll find that most cotton plants look similar, as if they were all grown from the same "best" variety. But this is precisely where the problem lies: in the pursuit of high yield or quality, we repeatedly select for certain superior varieties, gradually losing less obvious but valuable genes. In fact, many old local varieties or wild cotton relatives (referring to wild cotton related to cultivated cotton; Baran et al., 2023; Abbas et al., 2024) may not look impressive or have high yields, but they may harbor genes that confer disease resistance, drought tolerance, or improved fiber properties. However, these advantages are usually invisible to the naked eye and require specialized techniques such as principal component analysis (PCA) or SSR molecular markers to identify them (Mangi et al., 2024). Failure to perform these preliminary tests before breeding can lead to the potential for incorrect parent selection (Salama et al., 2024). If we encounter pests or diseases or sudden climate change one day, it may be too late to go back and find these neglected resources. Therefore, even if these genetic materials are not used immediately, they should be preserved in advance, as they may come in handy at a critical moment.

8 Concluding Remarks

The development of cotton fiber into its present form isn't solely the work of any single mechanism. Many people's first reaction might be to the benefits of polyploidy, particularly the combination of the A and D genomes, which has indeed significantly improved cotton yield and quality. But this is only one part of the puzzle; the entire process is far more complex than one might imagine. Whole-genome duplication, the reconfiguration of regulatory networks, and the proliferation of transposable elements within the genome all came together at different points in time. Sometimes, "copy and paste" alone isn't enough to explain why, for example, a gene family becomes so large. Sometimes, it's precisely the disruption of transposable elements, their intervening movements disrupting previously stable structures. Furthermore, developmental stages aren't static. Gene expression can be upregulated one day and downregulated the next; chromatin can coil where it should and loosen where it should, creating a dynamic rhythm like a musical score. This is why, even among upland cotton plants, others' plants grow slender and shiny, while those in your own field are thick, short, and uneven. This is due to the hidden dynamics of genetic mechanisms.

The domestication of cotton is never a black-and-white story. Look at those two subgenomes: how could the current imbalance, one stronger than the other, have occurred without the alternating forces of natural selection and artificial breeding? Interestingly, this imbalance is linked to the expression rhythm of cis-regulatory elements-who could have imagined that such seemingly insignificant details could ultimately influence tangible traits like fiber length and strength? However, while we've learned a lot, much remains unknown. For example, those "orphan genes"-with no known relatives or provenance-are responsible for their hidden agendas. And the mitochondrial genome, while often discussed, remains a mystery.

What's next? While this sounds easy, the actual implementation presents several persistent challenges. For example, while the cotton genome has been sequenced a long time ago, its detailed assembly and minimal gaps are still far from satisfactory. Furthermore, most studies focus on a few mainstream varieties; sample sizes for less popular varieties, local species, and wild relatives are far from sufficient. Furthermore, while various omics data are readily available, they are often used independently. Expression data is one thing, metabolism is another, and few studies have managed to piece them together into a coherent picture. Other crops-rice, wheat, and others-are already experimenting with single-cell and spatial transcriptomics, and even AI is being brought into the picture. For cotton, the technology is plentiful and the foundation is strong; the key lies in how to use it effectively. Some key genes are well-known, and the tools are readily available. However, if the data can't be connected and fed back into breeding, these resources are, in essence, just a pile of goodies. Breeding today can't just be done on a whim, especially in an increasingly erratic climate. Truly ensuring cotton's resilience in all weather depends on integrating genomic data, functional research, and actual field performance into a practical approach, rather than simply adhering to separate "truths."

Acknowledgments

We thank Editor Huang for his professional guidance on formatting specifications, which ensured that the paper met the journal's academic presentation standards.

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

Abbas M., Youyi Z., Lu C., Khan M., Rahman M., Umer M., Liang C., Meng Z., Wang P., Askari M., Wei Y., and Zhang R., 2024, Insights into genetic diversity and functional significance of the bHLH genes in cotton fiber development, Industrial Crops and Products, 216: 118763.
https://doi.org/10.1016/j.indcrop.2024.118763

Ahmed A., Khan A., Negm M., Iqbal R., Azhar M., Khan S., and Rana I., 2024, Enhancing cotton resilience to challenging climates through genetic modifications, Journal of Cotton Research, 7(1): 10.
https://doi.org/10.1186/s42397-024-00171-4

Baran N., Shimira F., Nadeem M., Altaf M., Andirman M., Baloch F., and Temiz M., 2023, Exploring the genetic diversity and population structure of upland cotton germplasm by iPBS-retrotransposons markers, Molecular Biology Reports, 50(6): 4799-4811.
https://doi.org/10.1007/s11033-023-08399-0

Chen X., Hu X., Li G., Grover C., You J., Wang R., Liu Z., Qi Z., Luo X., Peng Y., Zhu M., Zhang Y., Lu S., Zhang Y., Lin Z., Wendel J., Zhang X., and Wang M., 2024, Genetic regulatory perturbation of gene expression impacted by genomic introgression in fiber development of allotetraploid cotton, Advanced Science, 11(40): 2401549.
https://doi.org/10.1002/advs.202401549

Chen Z., Sreedasyam A., Ando A., Song Q., De Santiago L., Hulse-Kemp A., Ding M., Ye W., Kirkbride R., Jenkins J., Plott C., Lovell J., Lin Y., Vaughn R., Liu B., Simpson S., Scheffler B., Wen L., Saski C., Grover C., Hu G., Conover J., Carlson J., Shu S., Boston L., Williams M., Peterson D., Mcgee K., Jones D., Wendel J., Stelly D., Grimwood J., and Schmutz J., 2020, Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement, Nature Genetics, 52(5): 525-533.
https://doi.org/10.1038/s41588-020-0614-5

Conover J., and Wendel J., 2021, Deleterious mutations accumulate faster in allopolyploid than diploid cotton (*Gossypium*) and unequally between subgenomes, Molecular Biology and Evolution, 39(2): msac024.
https://doi.org/10.1093/molbev/msac024

Duan Y., Shang X., Wu R., Yu Y., He Q., Tian R., Li W., Zhu G., and Guo W., 2024, The transcription factor GhMYB4 represses lipid transfer and sucrose transporter genes and inhibits fiber cell elongation in cotton, Plant physiology, 197(1): kiae637.
https://doi.org/10.1093/plphys/kiae637

Grover C., Jareczek J., Swaminathan S., Lee Y., Howell A., Rani H., Arick M., Leach A., Miller E., Yang P., Hu G., Xiong X., Mallery E., Peterson D., Xie J., Haigler C., Zabotina O., Szymanski D., and Wendel J., 2024, A high-resolution model of gene expression during *Gossypium hirsutum* (cotton) fiber development, BMC Genomics, 26(1): 221.
https://doi.org/10.1186/s12864-025-11360-z

Han J., López-Arredondo D., Yu G., Wang Y., Wang B., Wall S., Zhang X., Fang H., Barragán-Rosillo A., Pan X., Jiang Y., Chen J., Zhang H., Zhou B., Herrera-Estrella L., Zhang B., and Wang K., 2022, Genome-wide chromatin accessibility analysis unveils open chromatin convergent evolution during polyploidization in cotton, Proceedings of the National Academy of Sciences, 119(44): e2209743119.
https://doi.org/10.1073/pnas.2209743119

Hu Y., Chen J., Fang L., Zhang Z., Ma W., Niu Y., Ju L., Deng J., Zhao T., Lian J., Baruch K., Fang D., Liu X., Ruan Y., Rahman M., Han J., Wang K., Wang Q., Wu H., Mei G., Zang Y., Han Z., Xu C., Shen W., Yang D., Si Z., Dai F., Zou L., Huang F., Bai Y., Zhang Y., Brodt A., Ben-Hamo H., Zhu X., Zhou B., Guan X., Zhu S., Chen X., and Zhang T., 2019, *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton, Nature Genetics, 51(4): 739-748.
https://doi.org/10.1038/s41588-019-0371-5

Huang J., Chen F., Guo Y., Gan X., Yang M., Zeng W., Persson S., Li J., and Xu W., 2021, GhMYB7 promotes secondary wall cellulose deposition in cotton fibers by regulating *GhCesA* gene expression through three distinct *cis*-elements, New Phytologist, 232(4): 1718-1737.
https://doi.org/10.1111/nph.17612

Joshi B., Singh S., Tiwari G., Kumar H., Boopathi N., Jaiswal S., Adhikari D., Kumar D., Sawant S., Iquebal M., and Jena S., 2023, Genome-wide association study of fiber yield-related traits uncovers the novel genomic regions and candidate genes in Indian upland cotton (*Gossypium hirsutum* L.), Frontiers in Plant Science, 14: 1252746.
https://doi.org/10.3389/fpls.2023.1252746

Khan Z., Khan S., Ahmed A., Iqbal M., Mubarik M., Ghouri M., Ahmad F., Yaseen S., Ali Z., Khan A., and Azhar M., 2023, Genome editing in cotton: challenges and opportunities, Journal of Cotton Research, 6(1): 3.
https://doi.org/10.1186/s42397-023-00140-3

Kumar R., Das J., Puttaswamy R., Kumar M., Balasubramani G., and Prasad Y., 2024, Targeted genome editing for cotton improvement: prospects and challenges, The Nucleus, 67(1): 181-203.
https://doi.org/10.1007/s13237-024-00479-1

Kushanov F., Turaev O., Ernazarova D., Gapparov B., Oripova B., Kudratova M., Rafieva F., Khalikov K., Erjigitov D., Khidirov M., Kholova M., Khusenov N., Amanboyeva R., Saha, S., Yu J., and Abdurakhmonov I., 2021, Genetic diversity, QTL mapping, and marker-assisted selection technology in cotton (*Gossypium* spp.), Frontiers in Plant Science, 12: 779386.
https://doi.org/10.3389/fpls.2021.779386

Li J., Liu Z., You C., Qi Z., You J., Grover C., Long Y., Huang X., Lu S., Wang Y., Zhang S., Wang Y., Bai R., Zhang M., Jin S., Nie X., Wendel J., Zhang X., and Wang M., 2024, Convergence and divergence of diploid and tetraploid cotton genomes, Nature Genetics, 56(11): 2562-2573.
https://doi.org/10.1038/s41588-024-01964-8

Li J., Yuan D., Wang P., Wang Q., Sun M., Liu Z., Si H., Xu Z., Ma Y., 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 X., Huang G., Zhou Y., Wang K., and Zhu Y., 2024, GhATL68b regulates cotton fiber cell development by ubiquitinating the enzyme required for β-oxidation of polyunsaturated fatty acids, Plant Communications, 5(10): 101003.
https://doi.org/10.1016/j.xplc.2024.101003

Lv L., Zuo D., Wang X., Cheng H., Zhang Y., Wang Q., Song G., and Ma Z., 2020, Genome-wide identification of the expansin gene family reveals that *expansin* genes are involved in fibre cell growth in cotton, BMC Plant Biology, 20(1): 223.
https://doi.org/10.1186/s12870-020-02362-y

Malik W., Wang X., Wang X., Shu N., Cui R., Chen X., Wang D., Lu X., Yin Z., Wang J., and Ye W., 2020, Genome-wide expression analysis suggests glutaredoxin genes response to various stresses in cotton, International Journal of Biological Macromolecules, 153: 470-491.
https://doi.org/10.1016/j.ijbiomac.2020.03.021

Mangi N., Iqbal M., Shuli F., Iqbal M., Alharthi B., Jatoi G., Sarfraz Z., Ma Q., and Sun X., 2024, Empowering cotton breeding programs through the strategic exploration and exploitation of phenotypic diversity of genetic resources under climate change conditions, Plant Stress, 13: 100548.
https://doi.org/10.1016/j.stress.2024.100548

Pan Y., Meng F., and Wang X., 2020, Sequencing multiple cotton genomes reveals complex structures and lays foundation for breeding, Frontiers in Plant Science, 11: 560096.
https://doi.org/10.3389/fpls.2020.560096

Qiao X., Li Q., Yin H., Qi K., Li L., Wang R., Zhang S., and Paterson A., 2019, Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants, Genome Biology, 20(1): 38.
https://doi.org/10.1186/s13059-019-1650-2

Salama E., Farid M., El-Mahalawy Y., El-Akheder A., Aboshosha A., Fayed A., Yehia W., and Lamlom S., 2024, Exploring agro-morphological and fiber traits diversity in cotton (*G. barbadense* L.), BMC Plant Biology, 24(1): 403.
https://doi.org/10.1186/s12870-024-04912-0

Shahzad K., Mubeen I., Zhang M., Zhang X., Wu J., and Xing C., 2022, Progress and perspective on cotton breeding in Pakistan, Journal of Cotton Research, 5(1): 29.
https://doi.org/10.1186/s42397-022-00137-4

Shuya M., Le L., Huiyun S., Yu G., Yujun L., and Qanmber G., 2023, Genomic identification of cotton *SAC* genes branded ovule and stress-related key genes in *Gossypium hirsutum*, Frontiers in Plant Science, 14: 1123745.
https://doi.org/10.3389/fpls.2023.1123745

Suo Q., Fang N., Zeng J., Yan F., Zhu X., Wang Y., Yu W., Chen J., Liang A., Li Y., Kong J., and Xiao Y., 2024, R2R3 MYB transcription factor GhMYB201 promotes cotton fiber elongation via cell wall loosening and very-long-chain fatty acid synthesis, International Journal of Molecular Sciences, 25(17): 9559.
https://doi.org/10.3390/ijms25179559

Tao X., Li M., Zhao T., Feng S., Zhang H., Wang L., Han J., Gao M., Lu K., Chen Q., Zhou B., and Guan X., 2021, Neofunctionalization of a polyploidization-activated cotton long intergenic non-coding RNA *DAN1* during drought stress regulation, Plant Physiology, 186(4): 2152-2168.
https://doi.org/10.1093/plphys/kiab179

Thangaraj A., Kaul R., Sharda S., and Kaul T., 2024, Revolutionizing cotton cultivation: a comprehensive review of genome editing technologies and their impact on breeding and production, Biochemical and Biophysical Research Communications, 742: 151084.
https://doi.org/10.1016/j.bbrc.2024.151084

Wang D., Hu X., Ye H., Wang Y., Yang Q., Liang X., Wang Z., Zhou Y., Wen M., Yuan X., Zheng X., Ye W., Guo B., Yusuyin M., Russinova E., Zhou Y., and Wang K., 2023, Cell-specific clock-controlled gene expression program regulates rhythmic fiber cell growth in cotton, Genome Biology, 24(1): 49.
https://doi.org/10.1186/s13059-023-02886-0

Wang J., Wang X., Wang L., Nazir M., Fu G., Peng Z., Chen B., Xing A., Zhu M., Ma X., Wang X., Jia Y., Pan Z., Wang L., Xia Y., He S., and Du X., 2024, Exploring the regulatory role of non-coding RNAs in fiber development and direct regulation of *GhKCR2* in the fatty acid metabolic pathway in upland cotton, International Journal of Biological Macromolecules, 266: 131345.
https://doi.org/10.1016/j.ijbiomac.2024.131345

Wang M., Wang P., Lin M., Ye Z., Li G., Tu L., Shen C., Li J., Yang Q., and Zhang X., 2018, Evolutionary dynamics of 3D genome architecture following polyploidization in cotton, Nature Plants, 4(2): 90-97.
https://doi.org/10.1038/s41477-017-0096-3

Wang P., Abbas M., He J., Zhou L., Cheng H., and Guo H., 2024, Advances in genome sequencing and artificially induced mutation provides new avenues for cotton breeding, Frontiers in Plant Science, 15: 1400201.
https://doi.org/10.3389/fpls.2024.1400201

Wang P., Zhao H., Liu X., Li G., Zhang X., Zhang X., and Xing Y., 2023a, Co-expression of multi-gene in cotton promotes the aggregation of multi-resistance and yield traits, Industrial Crops and Products, 204: 117311.
https://doi.org/10.1016/j.indcrop.2023.117311

Wang Y., Li Y., He S., Xu S., Li L., Zheng Y., and Li X., 2023b, The transcription factor ERF108 interacts with AUXIN RESPONSE FACTORs to mediate cotton fiber secondary cell wall biosynthesis, The Plant Cell, 35(11): 4133-4154.
https://doi.org/10.1093/plcell/koad214

Wei Y., Li Z., Wedegaertner T., Jaconis S., Wan S., Zhao Z., Liu Z., Liu Y., Zheng J., Hake K., Peng R., and Zhang B., 2022, Conservation and divergence of phosphoenolpyruvate carboxylase gene family in cotton, Plants, 11(11): 1482.
https://doi.org/10.3390/plants11111482

Wen X., Zhai Y., Zhang L., Chen Y., Zhu Z., Chen G., Wang K., and Zhu Y., 2022, Molecular studies of cellulose synthase supercomplex from cotton fiber reveal its unique biochemical properties, Science China Life Sciences, 65(9): 1776-1793.
https://doi.org/10.1007/s11427-022-2083-9

Wu H., Tian Y., Wan Q., Fang L., Guan X., Chen J., Hu Y., Ye W., Zhang H., Guo W., Chen X., and Zhang T., 2018, Genetics and evolution of MIXTA genes regulating cotton lint fiber development, New Phytologist, 217(2): 883-895.
https://doi.org/10.1111/nph.14844

Wu Y., Sun R., Huan T., Zhao Y., Yu D., and Sun Y., 2024, An insight into the gene expression evolution in *Gossypium* species based on the leaf transcriptomes, BMC Genomics, 25(1): 179.
https://doi.org/10.1186/s12864-024-10091-x

Xiong X., Zhu D., Grover C., Wendel J., Ma X., and Hu G., 2024, Dynamics of duplicated gene regulatory networks governing cotton fiber development following polyploidy, bioRxiv, 2024: 6076274.
https://doi.org/10.1101/2024.08.12.607624

Yang Z., Gao C., Zhang Y., Yan Q., Hu W., Yang L., Wang Z., and Li F., 2022, Recent progression and future perspectives in cotton genomic breeding, Journal of Integrative Plant Biology, 65(2): 548-569.
https://doi.org/10.1111/jipb.13388

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, The Crop Journal, 10(6): 1545-1556.
https://doi.org/10.1016/j.cj.2022.03.002

You J., Liu Z., Qi Z., Ma Y., Sun M., Su L., Niu H., Peng Y., Luo X., Zhu M., Huang Y., Chang X., Hu X., Zhang Y., Pi R., Liu Y., Meng Q., Li J., Zhang Q., Zhu L., Lin Z., Min L., Yuan D., Grover C., Fang D., Lindsey K., Wendel J., Tu L., Zhang X., and Wang M., 2023, Regulatory controls of duplicated gene expression during fiber development in allotetraploid cotton, Nature Genetics, 55(11): 1987-1997.
https://doi.org/10.1038/s41588-023-01530-8

Yu S.Y., 2024, The physicochemical properties of hemp fibers and their applications in the textile industry, Bioscience Evidence, 14(6): 293-303.
https://doi.org/10.5376/be.2024.14.0030

Zhang L.J., and Yang X.P., 2024, Advances in sugarcane genomics: navigating through complex polyploid genomes, Molecular Plant Breeding, 15(4): 155-166.
https://doi.org/10.5376/mpb.2024.15.0016

Zhang S., Jiang Z., Chen J., Han Z., Chi J., Li X., Yu J., Xing C., Song M., Wu J., Liu F., Zhang X., Zhang J., and Zhang J., 2021, The cellulose synthase (*CesA*) gene family in four *Gossypium* species: phylogenetics, sequence variation and gene expression in relation to fiber quality in Upland cotton, Molecular Genetics and Genomics, 296(2): 355-368.
https://doi.org/10.1007/s00438-020-01758-7

Zhang T., Zhang N., Li W., Zhou X., Pei X., Liu Y., Ren Z., He K., Zhang W., Zhou K., Zhang F., Ma X., Yang D., and Li Z., 2020, Genetic structure, gene flow pattern, and association analysis of superior germplasm resources in domesticated upland cotton (*Gossypium hirsutum* L.), Plant Diversity, 42(3): 189-197.
https://doi.org/10.1016/j.pld.2020.03.001

Zhao R., Cheng H., Wang Q., Lv L., Zhang Y., Song G., and Zuo D., 2022, Identification of the CesA subfamily and functional analysis of GhMCesA35 in *Gossypium hirsutum* L., Genes, 13(2): 292.
https://doi.org/10.3390/genes13020292