LIÈGE université
Gembloux
Agro-Bio Tech

# Identification and functional characterization of a fuzzless gene in *Gossypium arboreum*

**FENG Xiaoxu**

**Promotors: Prof. Song Guoli**

**Prof. Hervé Vanderschuren**

**Année civile: 2022**

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

# Identification and functional characterization of a fuzzless gene in *Gossypium arboreum*

**FENG Xiaoxu**

Identification et caractérisation fonctionnelle d'un gène *fuzzless* sur le *Gossypium arboreum*

Promoteurs: Prof. SONG Guoli & Prof. Hervé Vanderschuren

Année civile: 2022

**Copyright.** Cette œuvre est sous licence Creative Commons. Vous êtes libre de reproduire, de modifier, de distribuer et de communiquer cette création au public selon les conditions suivantes:

- paternité (BY): vous devez citer le nom de l'auteur original de la manie re indiquée par l'auteur de l'œuvre ou le titulaire des droits qui vous confère cette autorisation (mais pas d'une manière qui suggérerait qu'ils vous soutiennent ou approuvent votre utilisation de l'œuvre);

- pas d'utilisation commerciale (NC): vous n'avez pas le droit d'utiliser cette création à des fins commerciales;

- partage des conditions initiales à l'identique (SA): si vous modifiez, transformez ou adaptez cette création, vous n'avez le droit de distribuer la création qui en résulte que sous un contrat identique à celui-ci. À chaque réutilisation ou distribution de cette création, vous devez faire apparaitre clairement au public les conditions contractuelles de sa mise à disposition. Chacune de ces conditions peut être levée si vous obtenez l'autorisation du titulaire des droits sur cette œuvre. Rien dans ce contrat ne diminue ou ne restreint le droit moral de l'auteur.

# Résumé

**FENG Xiaoxu. (2022). Identification et caractérisation fonctionnelle d'un gène *fuzzless* sur le *Gossypium arboreum* (Thèse de doctorat en anglais).** Gembloux, Belgique, Gembloux Agro-Bio Tech, Université de Liège, 172 p., 24 tableaux, 39 fig.

**Résumé** — Le coton est largement cultivé dans le monde entier et conitent le matériau fibreux le plus essentiel et le plus naturel pour l'industrie textile moderne. Le clonage et l'analyse génétique d'un plus grand nombre de gènes associés aux caractères des fibres devraient permettre de mieux comprendre le mécanisme de développement des fibres et peuvent constituer la base de l'amélioration génétique permettant un meilleur rendement et une meilleure qualité des fibres. Le stade d'initiation des fibres détermine le rendement final en fibres. Les fibres de coton sont divisées en deux types en fonction du temps d'émergence et de la longueur de la dernière fibre: la fibre pelucheuse et la fibre duveteuse. Comparé à la fibre pelucheuse, le duvet n'est pas facile à égrener, il entrave l'absorption d'eau des graines et facilite la propagation des agents pathogènes. C'est pourquoi, le caractère «duvet» tend à être abandonnédans la production industrielle. Les mutants *fuzzless* (« sans duvet »)fournissent des ressources précieuses en matériel génétique pour explorer les mécanismes moléculaires sous-jacents à l'initiation et au développement des fibres et du duvet. Bien que la plupart des chercheurs se soient concentrés sur l'initiation et l'allongement des fibres ces dernières années, les gènes fonctionnels et les mécanismes moléculaires contribuant à la formation du duvet restent largement inconnus. La présente étude a mené une analyse génétique et un clonage cartographique du gène *fuzzless* dans des lignées de coton diploïdes en utilisant le mutant *fuzzless* DPL972 et le de type sauvage DPL971 (*G. arobreum*). De plus, nous avons identifié les membres de la famille de *GaGIR1* et analysé les modèles d'expression aux premiers stades de développement de la fibre et du duvet. Nous avons également exploré le réseau de co-expression impliqué dans le développement du duvet et identifié les gènes de module/pivots qui agissent potentiellement en synergie avec *GaGIR1* pour réguler la formation du duvet. Cette étude a permis une meilleure compréhension du mécanisme d'initiation du duvet du coton diploïde et apportent des informations importantes pour les ressources génétiques du coton afin de faciliter l'amélioration synchrone de la qualité et de la quantité des fibres.

Les principaux résultats ont été les suivants :

**1)** Dans la première partie, le gène *fuzzless* dans le coton diploïde DPL972 a été identifié comme un gène dominant monogénique nommé *GaFzl*. En combinant le clonage traditionnel basé sur carte et le séquençage de nouvelle génération, nous avons développé des marqueurs moléculaires étroitement liés au phénotype *fuzzless*. Nous avons cartographié *GaFzl* dans une région de 70 kb contenant sept gènes annotés du chromosome A08. Le séquençage de l'ARN et l'analyse de re-séquençage ont réduit ces candidats à deux gènes exprimés différentiellement (DEG), Coton_A_11941 et Coton_A_11942. Coton_A_11941 a montré une régulation à la hausse notable chez le mutant *fuzzless* et a été détecté des polymorphismes de séquence entre deux lignes de coton. Il a codé une protéine homologue d'AtGIR1 (répresseur interagissant avec GLABRA2) qui a

été impliquée dans le développement des poils racinaires chez *Arabidopsis thaliana* et est donc le gène candidat le plus probablement responsable du trait *fuzzless* dans DPL972.

**2)** Dans cette partie, nous avons d'abord confirmé que le *GaFzl* code pour une protéine de colocation membrane nucléaire et a une activité d'activation transcriptionnelle robuste. Ensuite, vingt et un gènes de la famille GIR ont été identifiés à partir de l'arboreum diploïde gossypium de coton, et ils ont été caractérisés pour leurs relations phylogénétiques, la structure des gènes, la distribution des chromosomes et la dynamique évolutive. Les éléments agissant en cis dans les régions protomères devraient être sensibles à la lumière, aux phytohormones, à la défense et au stress. En outre, les sites de liaison de plusieurs facteur de transcription tels que ERF, BBR-BPC, MYB, AP2/B3, DOF pourraient être prédits à partir des séquences promotrices. La plupart des gènes de la famille GIR n'étaient pas exprimés différemment entre un type sauvage et un mutant *fuzzless* du coton, et 14 des 21 membres de la famille présentaient des niveaux d'expression relativement élevés, ce qui indique que ces gènes peuvent jouer un rôle essentiel dans le développement des fibres et la formation du duvet. De plus, l'expression des orthologues des principaux régulateurs du développement des fibres chez *G.hirsutum* n'a pas été affectée par *GaFzl* dans *G.arboreum*, ce qui suggère que *GaFzl* pourrait être dans une nouvelle voie de développement des fibres.

**3)** Les ovules au stade initiatique (1, 3 et 5 DPA) de DPL971 et DPL972 ont été collectés pour une analyse RNA-Seq. L'analyse pondérée du réseau de co-expression génique a mis au point le module magenta ME fortement associé à un trait *fuzz*/*fuzzless*, qui comprenait un total de 50 gènes pivot exprimés différemment entre deux matériaux. Sur la base du changement de niveau d'expression, les gènes pivot ont été divisés en deux cluster1 (régulé à la hausse dans DPL972) et cluster2 (régulé à la baisse dans DPL972). *GaFzl*, qui régule négativement la formation de trichomes et du duvet, ainsi que 27 autres gènes ont été trouvés impliqués dans le cluster1 magenta ME. Il est à noter que Ga04G1219 et Ga04G1240, respectivement, codent la protéine arabinogalactane de type fascicline 18 (FLA18) et la protéine de transport, ont montré des différences remarquables de niveau d'expression,ils pourraient être impliqués dans la glycosylation des protéines pour réguler la formation et le développement du duvet.

En résumé, les résultats des trois expériences ont montré que *GaFzl* régulait négativement la formation de trichomes et du duvet dans les lignées de coton diploïdes et pourrait subir une voie nouvellement découverte différente des autres gènes publiés liés au développement de la fibre.

**Mots-clés**: initiation de fibre, mutant *fuzzless*, clonage cartographique, RNA-Seq, WGCNA.

# Abstract

**FENG Xiaoxu. (2021). Identification and functional characterization of a fuzzless gene in *Gossypium arboreum* (PhD Dissertation in English).** Gembloux, Belgium, Gembloux Agro-Bio Tech, University of Liège, 172 p., 24 tables, 39 figures.

**Abstract** — Cotton is extensively grown throughout the globe and comprises the most essential and natural fiber material for the modern textile industry. Cloning and genetic analysis of more genes associated with the fiber trait will illuminate fiber developmental mechanism and lay the foundation for genetic improvement of fiber yield and quality. The fiber initiation stage determines the final fiber yield. Cotton fibers are divided into two types based on the emerging time and last length: lint and fuzz. Compared with lint fiber, fuzz is not easy to gin, hinders the water absorption of seeds, and is easy to spread pathogens, and thus the fuzz trait tends to be abandoned in the agronomic study and actual production. The fuzzless mutants provide valuable germplasm resources to explore the molecular mechanism underlying fiber and fuzz initiation and development. Though most researchers focused on fiber initiation and elongation in recent years, the crucial functional genes and the molecular mechanisms contributing to fuzz formation remain largely unknown. This study conducted a genetic analysis and map-based cloning of the fuzzless gene in diploid cotton lines using the fuzzless mutant DPL972 and the wild-type DPL971 (*Gossypium arobreum*). In addition, we identified the family members of *GaGIR1* and analyzed the expression patterns at the early developmental stages of fiber and fuzz. We also explored the co-expression network involved in fuzz development and identified the crucial module/ hub genes potentially working together with *GaGIR1* to regulate fuzz formation. This study will promote a better understanding of the mechanism of fuzz initiation of diploid cotton and support gene resources to facilitate the cotton molecular design breeding and the synchronous improvement breeding of fiber quality and quantity.

The main results were as follows:

**1)** In the first part, the fuzzless gene in diploid cotton DPL972 was identified as a monogenic dominant gene named *GaFzl*. By combining traditional map-based cloning and next-generation sequencing, we developed molecular markers tightly linked with the fuzzless phenotype. We mapped *GaFzl* to a 70-kb region containing seven annotated genes of Chromosome A08. RNA-Sequencing and re-sequencing analysis narrowed these candidates to two differentially expressed genes (DEGs), Cotton_A_11941 and Cotton_A_11942. Cotton_A_11941 exhibited a notable upregulation in the fuzzless mutant and was detected sequence polymorphisms between two cotton lines. It encoded a homologous protein of AtGIR1 (GLABRA2-interacting repressor) which was involved in the root hair development in *Arabidopsis thaliana* and is thus the candidate gene most likely responsible for the fuzzless trait in DPL972.

**2)** In this part, we first confirmed that the *GaFzl* encodes for a nuclear-membrane colocation protein and has a robust transcriptional activation activity. Then twenty-one *GIR* family genes were identified from diploid cotton *Gossypium arboreum*, and they were characterized for phylogenetic relationships, gene structure, chromosome distribution, and evolutionary dynamics. *Cis*-acting

elements in protomer regions were predicted to be responsive to light, phytohormone, defense, and stress. In addition, the binding sites of multiple TFs such as ERF, BBR-BPC, MYB, AP2/B3, DOF could be predicted from the promoter sequences. Most *GIR* family genes were not differentially expressed between a wild type and a fuzzless mutant of cotton, and 14 out of 21 family members exhibited relatively high expression levels, indicating these genes may play essential roles in fiber development and fuzz formation. In addition, the expression of orthologs of key regulators of fiber development in *G.hirsutum* was not affected by *GaFzl* in *G.arboreum*, suggesting *GaFzl* might be in a new pathway of fiber development.

**3)** The initiation stage ovules (1, 3, and 5 DPA) of DPL971 and DPL972 were collected for RNA-Seq. Weighted gene co-expression network analysis discerned the ME magenta module highly associated with a fuzz/fuzzless trait, which included a total of 50 hub genes differentially expressed between two materials. Based on the expression level change, the hub genes were devided into two cluster1 (up-regulated in DPL972) and cluster2 (down-regulated in DPL972). *GaFzl*, which negatively regulates trichome and fuzz formation, along with other 27 genes were found involved in ME magenta cluster1. It is noteworthy that Ga04G1219 and Ga04G1240, respectively, encode Fasciclin-like arabinogalactan protein 18 (FLA18) and transport protein, showed remarkable differences of expression level and implied that they might be involved in protein glycosylation to regulate fuzz formation and development.

In summary, the results of the three experiments showed that *GaFzl* negatively regulated trichome and fuzz formation in diploid cotton lines and may undergo a newly discovered pathway different from other “star” fiber-related genes.

**Key words:** fiber initiation, fuzzless mutant, map-based cloning, RNA-Seq, WGCNA

# Acknowledgments

Time flies, and this moment is finally coming to the end of my doctoral program. During the past six years, there are pains and happiness, making me grow up and harvest a lot. The doctoral time will become a valuable experience in my life. Without the help and support of my teachers, schoolmates, and friends, I couldn't finish my doctoral thesis. Hence, I would like to express my sincerest thanks to them.

First of all, I would like to express my heartfelt thanks to my academic committee members, and special appreciation to my supervisors, Professor Song Guoli from the Institute of Cotton Research of Chinese Academy of Agricultural Sciences and Professor Hervé Vanderschuren from University of Liège. Professor Song Guoli gives me a lot of valuable guidance in experiments and free play space in scientific research. His prudent and meticulous work style and sincere personality are benchmarks in my future work and life. Thanks to Professor Hervé Vanderschuren. Although I have only stayed in Gembloux for one year, his rigorous scientific research spirit, open scientific research thinking, top scientific research team, and pursuit have deeply impressed me. Thanks to Zuo Dongyun for his valuable suggestions on my experiments in the ICR of CAAS. I want to thank Ludivine Lassois for her friendly communication for resolving the problems or obstacles in scientific research and for actively promoting the progress of my experiment at Gembloux. Besides, thanks to her for becoming a member of my academic jury.

Thanks to the research group, teacher Cheng Hailiang, Wang Qiaolian, Lv Limin and Zhang Youping, for the care of my experiment and life. Thanks to Dr. Li Shuyan, Wu Chaofeng, Wu Cuicui, Xiao Shuiping, Liu Shang and MS. Yang Qiuhong, Li Simin, Li Wenqian, Dai Yuanli, and so on. Thank you for your support and help in my experiments and daily life so that I have had happy and warm doctoral years.

Thanks to my classmates Wang Xiaoge and Li Keqi, thank you for accompanying me to chat and decompress in my life and making my life so colorful! May the friendship last forever!

Our year in Gembloux is a special one, during which we went through the global COVID-19 pandemic. It was a unique and challenging experience, and I will not forget the special times we faced together. Here I appreciate to Song Zhaoxin, Liu Shangwu, Song Jiao, Hou Wanying, Zhang Yuping, and other friends, for your enthusiastic communication and the thoughtful company at my disappointment time. Your always being there make my life full of moving and happy in the foreign land.

At the same time, I also want to thank the laboratory team members in Gembloux for our academic discussion and for overcoming difficulties together. Thanks to Gao Xiaoming, Chang Aixia, Sok Lay Him, Emanoella Lima Soares, Leonard Shumbe, Syed Shan, Sara Shakir, Kumar Vasudevan, Ivan Jauregui, Izargiaida Vega, and Briue Lecart. I sincerely wish you with health and happiness, and everything goes well.

I am also very grateful to the Education project of a doctoral degree in Agronomy and Bioengineering cooperatively run by the Graduate School of the Chinese academy of agriculture science (GSCAAS) and University of Liège (ULg). Thanks for giving me this opportunity to participate in international exchanges and cooperation. Thanks

to Zhang Mingjun's encouragement and supervision of our academic progress. Thanks to the State Key Laboratory of Cotton Biology for providing a good scientific research platform for the successfully completing of my doctoral studies.

Finally, I would like to thank my husband Cheng Hailiang, who is my spiritual pillar and constantly encourages me to move forward bravely both in scientific research and in life. Thanks to my parents and family for their tolerance and support in our life, lightening our burden and removing our worries.

Thanks for every meeting in my life, and may the future be better for all of us.

FENG Xiaoxu
2022
Anyang, China

# Tables of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| At | A sub genome |
| bp | Base pair |
| BSA-Seq | Bulked segregant analysis sequencing |
| cDNA | complementary deoxyribonucleic acid |
| ChIP-Seq | Chromatin Immunoprecipitation sequencing |
| cM | Centi-Morgan(s) |
| Co-IP | Co-Immunoprecipitation |
| DPA | Days post-anthesis |
| Dt | D sub genome |
| EMSA | Electrophoretic mobility shift assay |
| GAs | Gibberellins |
| GFP | Green fluorescence protein |
| GIR | GL2-interacting-repressors |
| GWAS | Genome wide association sequence |
| HTS | High-throughput sequencing |
| InDel | Insertion and Deletion |
| JA | Jasmonic acid |
| Kb | Kilo base pair |
| MBW | MYB-bHLH-WD40 |
| PCR | Polymerase chain reaction |
| RNA-Seq | RNA sequencing |
| ROS | Reactive oxygen species |
| qRT-PCR | Real-time Quantitative PCR |
| SA | Salicylic acid |
| SCW | Secondary cell wall |
| SD | Synthetic Dropout. |
| SSR | Simple sequence repeats |
| TF | Transcription factors |
| VLCFA | Very long chain fatty acids |
| WGCNA | Weighted gene co-expression network analysis |
| Y2H | Yeast two-hybrid |

# 1

# General introduction

Cotton (*Gossypium spp.*) is grown in more than 80 countries around the globe and comprises the most essential and natural fiber material for the global textile industry (Chen et al. 2007). In addition, cotton is also one of the primary crops of seed oil and protein, supplying good nutrition sources for feed, foodstuff, and oil (Alford et al. 1996). As an entire commercial crop, the cotton-processing industry plays an important role in the national economy in China. Since the late 20$^{th}$ century, China has been beyond America and become the largest producer and consumer of raw cotton, with approximately 27.8% contribution toward global cotton production, up to 30.3% contribution to global imports of natural cotton, about 37.5% for global cotton consumption in textile factories, and 28.6% towards worldwide textile and garment export trade (USDA, Cotton: World Markets and Trade, www.usda.gov). Although the production per unit area increased in the past eight years, the national planting area declined significantly. Furthermore, modern industrial equipment is advanced and precision-engineered, the raw cotton produced in China is difficult to meet the quality requirements of the textile industry for medium and high-end spinning products, leading to the emergence of a large number of chemical fiber substitutes and more import of superior quality of cotton (Huang et al. 2017a). Due to the increasing population base and economic level, our demand for cotton products is in short supply, and we still have to import large quantities of raw cotton from other cotton-producing countries.

The evaluation of cotton fiber quality is mainly based on fiber length, fiber fracture-specific strength, and fiber fineness, namely, the Micronaire value. The requirement of fiber length and fiber fracture-specific strength is more urgent. Cultivating new varieties of cotton with higher yield and fiber quality can meet the increasing cotton consumption and the pursuit of high quality and improve the utilization rate of cotton, reduce the cultivated area, and actively respond to the country's supply reform. Therefore, it is the goal of cotton breeders to achieve simultaneous improvement of cotton yield and fiber quality.

# 1. Systematic classification of cotton genus

According to botanical classification, cotton is a genus of the Angiosperm, Dicotyledoneae, Malvales, Malvaceae, Gossypium. The classification of the cotton genus mainly experiences three periods: based on morphological characteristics, chromosome numbers, and genome numbers, respectively (Endrizzi et al. 1985). The morphology of Gossypium species differentiated diversely, ranging from small shrubs to tall trees up to 15 m. According to the further study of the natural History of Cotton published in 1979, the genus Gossypium was first divided into four subgenera, eight groups, nine subgroups, and fifty species (Fryxell 1992, 1979). *G. ekmanianumsi* and *G.stephensii* are new species identified in 2015 and 2017, respectively. (Grover et al. 2015; Gallagher et al. 2017). Till now, the genus Gossypium includes 52 species, 45 diploids (2n = 2x = 26) and 7 allotetraploids (2n = 4x = 52) (Table 1-1) (Beasley 1940, 1941; Endrizzi et al. 1985; Wendel et al. 2010).

**Table 1-1** Classification for *Gossypium*. (Wendel & Grove, 2015)

| Genome | Recognized species |
|---|---|
| A (2) | *G. arboreum, G. herbaceum* |
| B (3) | *G. anomalum, G. triphyllum, G. capitis-viridis* |
| C (3) | *G. sturtianum, G. nandewarense, G. robinsonii* |
| D (13) | *G. thurberi, G. armourianum, G. harknessii, G. davidsonii, G. klotzschianum, G. aridum, G. raimondii, G. gossypioides, G. lobatum, G. trilobum, G. laxum, G.turneri, G. schwendimanii* |
| E (8) | *G. stocksii Mast., G. somalense, G. areysianum, G. incanum, G. benadirense, G. bricchettii, G. vollesenii, G. trifurcatum* |
| F (1) | *G. longicalyx* |
| G (3) | *G. bickii, G.australe, G. nelsonii* |
| K (12) | *G. anapoides, G. costulatum, G. cunninghamii, G. enthyle, G. exiguum, G. londonderriense, G. marchantii, G. nobile, G. pilosum, G. populifolium, G. pulchellum, G. rotundifolium* |
| AD (7) | *G. hirsutum, G. barbadense, G. tomentosum, G. mustelinum, G. darwinii, G. ekmanianum, G. stephensii* |

Cultivated cotton indicates these species, which underwent long period of cultivation and domestication, afterward were fine selected and cultured to meet human demands. After long-term natural and artificial selections, cultivated cotton was divided into distinct types and led to abundant species diversities (Applequist et al. 2001). There are currently four varieties of cultivated cotton planted worldwide: two diploids *Gossypium herbaceum* L. and *Gossypium arboreum* L., and two allotetraploids, *Gossypium hirsutum* and *Gossypium barbadense* (Wendel and Cronn 2003).

Sea island cotton species have the finest and longest fibers, and their excellent fiber qualities make them ideal materials for high-end fabrics. Unfortunately, there are several shortcomings narrowing their adaptabilities: sensitivity to illumination, long growth and bolling, low yield and lint percentage. Upland cotton is characterized by large boll, long fiber, good quality, high yield, insensitivity to sunshine, and wide adaptability. It thus accounts for more than 95% of the cultivated areas of cotton (Wendel et al. 1992; Constable and Bange 2015). Diploid cultivars have the disadvantages of coarse fiber and low yield and thus have been basically replaced by tetraploid cultivars (Gandhi et al. 2015; Khaleequr et al. 2012; Kulkarni et al. 2009; Kantartzi and Stewart 2008). However, due to their great resistances, diploid cultivars are necessary genetic resources for cotton breeding and improvement (Maqbool et al. 2010; Huang et al. 2020).

Furthermore, with the release of high-quality genomes, researchers systematically revealed the evolutionary relationships and the genetic diversity of cotton genomes from the genomic level (Liu et al. 2015b; Wang et al. 2015b; Yuan et al. 2015; Zaidi et al. 2018). Tetraploid cotton lines have been produced by the fusion of diploid $A_1$

and $D_5$, and experienced polyploidization through whole-genome replication (Wendel and Grover 2015). Genes of tetraploid cotton present the subgroup-biased expression patterns and perform different functions due to the variations of *cis*-acting elements (Wang et al. 2017). And the traits of cotton fiber quality may be more inherited from subgroup A (Wang et al. 2015b; Fang et al. 2017). Therefore, *G. arboreum* provides superior germplasm resources and is a perfect research object to decipher the controlling genes and regulatory networks contributing to fiber development and fiber quality improvement.

# 2. Morphological and physiological characteristics of cotton fiber in different developmental stages

Cotton fibers is a single-cell of elongated structures derived from the ovule epidermis, and it grows in a specialized and highly polarized manner (Basra and Malik 1984; Oosterhuis 1990). Fiber structure includes three parts from out to in: primary cell wall, secondary cell wall, and the central cavity (Kim and Triplett 2001). The primary wall is the original cell wall of fibrous cells with about 0.1~0.2 um thick, composed of pectin and cellulose. The secondary wall is the major part of the mature fiber, accounting for about 90% of the total weight of mature fiber cells, and its main component is cellulose. The central cavity is the innermost gap in fiber structure in which contains various inclusions (Yu et al. 2019). Fibers will exhibit a colorful character when a higher concentration of pigment is enriched and accumulated in inclusions.

Fiber develops through four distinct but overlapping stages (Figure 1-1): initiation, elongation, secondary cell wall synthesis, and maturation, respectively (Tiwari and Wilkins 1995; Lee et al. 2007).

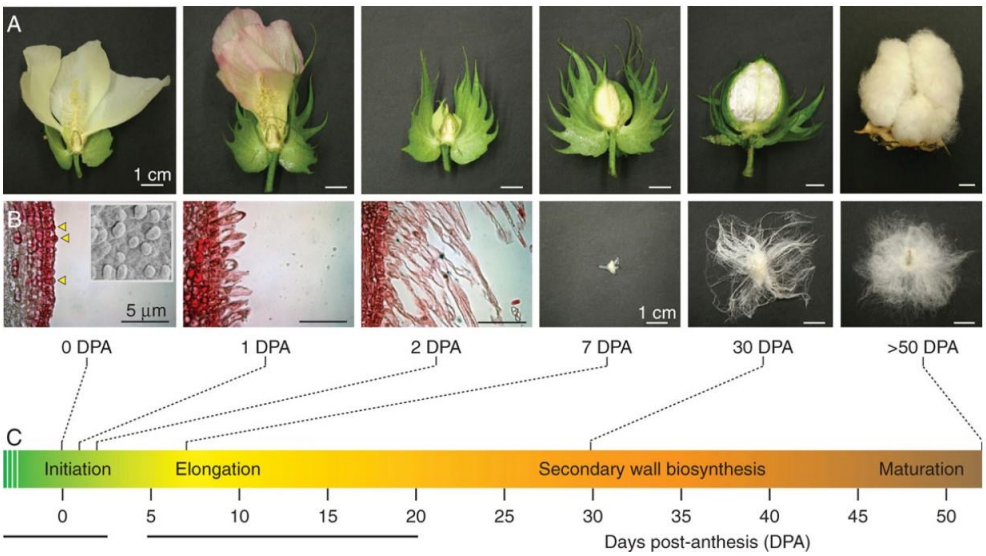



**Figure 1-1:** Fiber developmental stages. (Lee et al. 2007)

## *2.1 Fiber initiation stage*

Fiber initial development can divide into original cell differentiation and fiber cell processes. The fiber initiation period lasts for a short time, generally from the day of flowering to the end of the third day. Although all epidermal cells are potential fibroblasts except of stomatal guard cells and micropyle cells, not all epidermal cells can differentiate into fibroblasts, and only about 25% of ovule epidermal cells eventually develop into fibroblasts (Stewart 1975; Ryser 1999; Mansoor and Paterson 2012). The time that fibroblasts begin to differentiate is often challenging to determine morphologically because a series of physiological and biochemical changes have taken place within the cell before the change happens in cell morphology. From the pre-flowering to the day of flowering, fibroblasts have been differentiated and formed, and then the fibrous cells continue to protrude under the stimulation of post-pollination substances. The length and strength of mature cotton fibers are directly determined by the time of differentiation of fibroblasts. The peak time of the fibrous blast process was diverse in different cotton species and varieties of the same cotton species. Generally, there are two peak times of fibroblast protuberance. The first is the day of flowering or earlier when the fibroblasts differentiate and finally develop into lint. The second is 3-5 days after flowering when the fibroblasts differentiate and produce into fuzz with a final length of less than 15 mm.

Regarding the position of fibroblast protuberance starting, the previous study revealed that fibroblast protuberance firstly occurs at the chalaza end of the ovule and then at the micropyle end. However, according to the observation results by the scanning electron microscope, fibroblast protuberance was firstly observed at the ridge of the funiculus and then spread around. Till several hours later, the protuberance was detected at the chalaza end and micropyle end. There was no protuberant cells near micropyle in several days after flowering. Therefore, fibers on the supper side of ridge of the funiculus were the longest and thickest. Fibers on chalaza end and middle part of the ovule were shorter and more sparse. Fibers close to micropyle ends were the shortest and the lightest. This sequence of fibrous development was caused by the uneven distribution of endogenous hormones or different sensitivities of ovule epidermal cells to hormones (Stewart 1975).

## *2.2 The rapid elongation period*

Cotton fiber rapid elongation period is the primary wall synthesis period, which begins with cell protuberance and lasts till 20-25 days after flowering. In this period, the fastest elongation rate can reach about 2 mm/day, and the length of the fiber in this period can reach 90% of the final length of the fiber. When the elongation period completed, the length of the fiber cell can get 3000 times its diameter (Singh et al. 2009). The rapid elongation of the fiber is accompanied by the formation of a large central vacuole. The vacuole occupies most of the intracellular volume, and the cytoplasm compresses into a thin layer. The direction and polarity of fiber extension

are determined by both turgor pressure and cell wall structure. Turgor pressure is generated by the accumulation of osmotic regulation solute in vacuoles. And the elongation of the fiber is driven by osmotic pressure caused by the high concentration of potassium malate in the vacuole (Basra and Malik 1984). During fiber elongation, the other significant change is the synthesis of the primary cell wall, which mainly contains cellulose, hemicellulose, pectin, and a small amount of wax. Fiber cells can rapidly extend polar under the drive of turgor pressure because of the strong extension ability of the primary wall. Meanwhile, the cell activity is very vigorous, and genes related to the cell wall, cytoskeleton, and carbohydrate metabolism form to be significantly transcribed (Haigler et al. 2005; Qin and Zhu 2011; Gou et al. 2007).

## *2.3 The secondary cell wall (SCW) thickening stage*

Cotton fiber elongation and thickening are controlled by two different systems, primary cell wall elongation and secondary cell wall (SCW) thickening, which overlap for about 5 to 10 days. Cotton fiber SCW thickening begins 16~19 days after flowering, when the elongation of fiber cells and synthesis of primary cell wall taper off (Gou et al. 2007). At the same time, a large amount of cellulose is synthesized and deposited, indicative of the beginning of secondary wall thickening, and cellulose deposition lasts until 40-50 days after anthesis (Hu et al. 2019). In addition to the changes in cellulose content, the neutral sugar content in the cell wall also alters, among which the biggest changes are non-cellulose glucans such as xyloglucan and *β*-1, 4-glucan. Xyloglucan is a kind of polysaccharides that exists in the primary cel wall and plays a vital role in cell growth. It is synthesized and accumulated only in the primary wall synthesis period but not in the secondary wall thickening period. *β*-1, 4-glucan chains accumulate in large quantities to finish cellulose accumulation and form 20-30 layers of "growth day wheel". During the development of cotton fiber, the protein content in the fiber cell wall increases first and then decreases, especially after the secondary wall thickening starts. The protein content in the cell wall continuously decreases until the fiber is mature and no more protein residues in the cell wall (Liu 2013). This suggests the importance of cell wall proteins in forming the cell wall structure and realizing its function.

## *2.4 Fiber dehydration and maturation stage*

From 45~60 days post-anthesis, the secondary wall thickening period is almost over, cotton fiber is into the dehydration and maturity period. It begins from cotton boll cracking and ends until the completion of boll opening. Fiber cells are displayed in tubular before cracking, with a high water content inside the central cavity and cellulose arranged spirally. An astounding amount of cellulose constitutes the main ingredient of mature fibers, which accounts for approximately 95% of the dry weight. Besides, hemicellulose, colloid, wax, mineral substance, and trace amounts of pigment, sugar and organic acid also accumulate in mature fiber (Haigler et al. 2005).

During the maturation stage, about 5~7 days after boll cracking, fiber is exposed to atmospheric conditions, and fiber cells start to dehydrate. The forces of hydrogen bond between the cellulose in the secondary wall inequality, which therefore causes a surface shrinkage crimping (Ryser and Holloway 1985). The twisted level of each fiber differs and depends on the size of the helix angle and secondary wall thickness. Fibers with excellent ripeness exhibit thicker cell walls and more twists, producing a greater cohesive force during spinning process.

# 3. Multiple factors are involved in fiber developmental process

Plant trichomes are giant cells that differentiate from the epidermis of leaves, stems, petals, and other organs (Pattanaik et al. 2014; Hulskamp 2004; Balkunde et al. 2010). According to characteristics of morphology and nature, trichomes can divide into various types: unicellular trichome or multicellular trichome; glandular trichome or non-glandular trichome; with branch or non-branch (Yang and Ye 2013; Tissier et al. 2017). These hair-like cells play essential roles along with plant development. For example, trichomes could increase the tolerance to abiotic stresses such as UV irradiation, extreme temperatures, drought, and salt. They would also protect against pathogens, insects and even herbivore attacks (Bac-Molenaar et al. 2019). Besides, glandular trichomes may synthesize and secret various metabolite compounds to provide direct or indirect chemical protections in response to stress and diseases (Huchelmann et al. 2017; Andreuzza 2020; Chalvin et al. 2020).

Furthermore, trichomes or the contents generated and stored in trichomes may also be essential and valuable resources for human living and medical research. For instance, fiber is a single-cell trichome derived from the seed coat of cotton and is regarded as a crucial and natural source and raw material for the textile industry because of its high cellulose contents (Rinehart et al. 1996). Artemisinin, a famous and effective drug for antimalaria, could be produced and generated in the trichomes of *Artemisia annua* (Singh et al. 2016; Yan et al. 2017). Hence, it is of great importance to explore differentiation patterns and regulatory mechanisms underlying the formation of trichomes in nearly all plants.

## *3.1 Developmental pattern of trichomes in Arabidopsis*

The molecular basis, as well as regulatory genes and networks of trichome development, have been typically and detailedly investigated on rosette leaves of *Arabidopsis thaliana*. Trichome development undergoes an immensely complicated and intricate process (Wang et al. 2019d; Hulskamp 2004). Specified single protodermal cells at the base of leaf form into incipient trichomes and then initiate and enter into four endoreduplication cycles, resulting in increased nuclear size and new direction of growth and eventually forming into 2-3 branches (Akhtar et al. 2017; Haigler et al. 2012).

Numerous available shreds of evidences have proved that a sophisticated gene

regulatory network and various core transcription factors are involved in Arabidopsis trichome patterning. The R2R3 MYB transcription factor GLABROUS1 (GL1) interacts with the WD-40 repeat protein TRANSPARENT TESTA GLABRA 1 (TTG1) and the bHLH transcription factor GLABRA 3 (GL3)/ENHANCER OF GLABRA 3 (EGL3) (Larkin et al. 1999; Johnson et al. 2002; Payne et al. 2000). These proteins compose a trimeric activation complex (MYB-bHLH-WDR (MBW)) to activate the expression of homeodomain-leucine zipper (HD-Zip) GLABRA2 (GL2) and a WRKY transcription factor TRANSPARENT TESTA GLABRA2 (TTG2) and thus stimulate the differentiation of epidermal cells and turn on trichome initiation (Shen et al. 2006). At the same time, the trimeric complex can also trigger the R3-MYB transcription factors, which share redundant functions and negatively regulate trichome formation (Kirik et al. 2004; Larkin et al. 2003). TRIPTYCHON (TRY), CAPRICE (CPC), and ENHANCER OF TRY AND CPC (ETCs) would transfer to neighboring cells and compete with GL1 for interacting with GL3 and TTG1, leading to a new MBW complex without the activation ability of GL2 expression and thus disturbing trichome differentiation and development in the neighboring cells (Zhang et al. 2003; Dai et al. 2015; Song et al. 2015). Except for this known and fundamental activator complex, several other transcription factors have also been identified in the trichome development (Kong et al. 2021; Zhang et al. 2019; Wang et al. 2019a; Cox and Smith 2019). A NAC transcription factor NTL8 and SQUAMOSA PROMOTER BINDING PROTEIN LIKE (SPL) transcription factors (miR156-targeted genes) have been respectively reported to trigger TCL1 and TRY directly and regulate trichome formation negatively (Tian et al. 2017; Yu et al. 2010; Xu et al. 2016). Besides, the CIN–TCP transcription factor TCP4, which also refers as the miR319-targeted genes, acts as a negative regulator to suppress trichome initiation by directly activating the transcription of *TCL1* and *TCL2* (Vadde et al. 2018; Vadde et al. 2019; Challa et al. 2019; Challa et al. 2016). However, experimental evidence has shown no interaction within NTL8, SPL9 and TCP4 both in vivo and in vitro, suggesting that these upstream genes of *TCL1/2* and *TRY* work independently during trichome development (Tian et al. 2017; Vadde et al. 2019). The C2H2-zinc finger transcription factors, e.g. GIS, GIS2, GIS3, ZFP8, ZFP5, and ZFP6, have been proved to act upstream of the MBW complex and positively activate the expression of *GL1* according to the loss-of-function experiments and expression pattern analysis (Gan et al. 2006; Gan et al. 2007b; Gan et al. 2007a; Ishida et al. 2008; Zhou et al. 2013; Zhou et al. 2011; Sun et al. 2015; An et al. 2012).

Phytohormones, such as auxin (IAA), gibberellic acids (GAs), cytokines (CKs), ethylene (ETH), jasmonic acid (JA), salicylic acid (SA) and brassinosteroids (BRs), play important roles in regulating plant growth and different developmental processes, independently or overlappingly (Perazza et al. 1998; Maes et al. 2008; Traw and Bergelson 2003). Recently, numerous efforts have unveiled the network of signaling and transcriptional regulators involved in trichome formation and proliferation (Pattanaik et al. 2014; Matias-Hernandez et al. 2016; Yan et al. 2017; Liu et al. 2018). GAI and SPINDLY (SPY) respectively function as the GA synthesis factor and GA signaling repressor to significantly affect trichome development by regulating the level of GAs (Dill and Sun 2001). GISs and ZFPs could integrate the crosstalk of GA and CK to regulate trichome formation (Gan et al. 2007a). The CORONATINE

INSENTIVE1 (COI1) interacts with $ASK_1$/$ASK_2$, Cullin1, and Rbx1 and form the $SCF^{COI1}$ complex to mediate JA induction (Xie et al. 1998; Devoto et al. 2002). In the knockout line *coi1-2*, JA signal transduction attenuated, and the number of trichomes dramatically declined (Qi et al. 2011). Furthermore, when JA is deficient, JAZs, the negative JA signaling regulators, could bind to GL3, EGL3, and GL1, thus restraining the normal formation of the MBW complex, resulting in a fewer production of trichomes (Qi et al. 2014). The *bls1* mutant hindered the BR signal transduction and produced fewer leaf trichomes (Laxmi et al. 2004). On the contrary, SA exhibited negative regulatory effects on trichome formation, and the application of SA could notably decrease the density and the total number of trichomes (Traw and Bergelson 2003).

## *3.2 Homologous genes of trichome formation in Arabidopsis also contribute to fiber development*

Trichomes in Arabidopsis derive from a single-cell and have two or three branches whereas the cotton fiber is a unicellular trichome without any branches of the seed coat. In addition, the length of cotton fibers are much longer than the trichome length in Arabidopsis leaf (Wang et al. 2019d). These dissimilarities in morphology indicate that the trichomes in Arabidopsis leaf and cotton fiber may evolve in different underlying mechanisms. However, due to the short growth cycle, Arabidopsis remains an attractive model plant to study several key pathways also present in cotton and possibly relevant to fiber development. In the past decades, numerous studies have shown that the molecular regulatory network underlying the differentiation and development of the Arabidopsis trichome may also contribute to cotton fiber formation and growth (Du et al. 2018; Huang et al. 2017b; Ma et al. 2018a). Thus, detailed studies in Arabidopsis have provided plenty of references and lay a solid background foundation for studying cotton fiber initiation and elongation.

In cotton, MYB109 and MYB2 belonged to R2R3 MYB and were identified as homologs of GL1. *MYB109* exhibited a high expression level during the early stage of fiber development (Suo et al. 2003; Pu et al. 2008), and ectopic expression of *MYB2* in *gl1* mutant in Arabidopsis could fine rescue the glabrous phenotype (Guan et al. 2011; Wang et al. 2019c). Compared with *GL1*, *MYB25* and *MYB25-like* were *MIXTA* genes belonging to different kinds of R2R3 MYB and proved to participate in fiber initiation and elongation. Silencing of *MYB25* leads to delayed fiber initiation and a reduced number of leaf trichomes (Machado et al. 2009). Besides, *MYB25-like* functions upstream of *MYB25* and *MYB109* to regulate fiber development (Walford et al. 2011; Wan et al. 2016; Wu et al. 2017). While, overexpression of R3 MYB, such as *GhCPC* and *GhTRY*, may suppress the fiber initiation and elongation by their interaction with *GhMYC1* (*GL3*) (Shangguan et al. 2010; Wang et al. 2013a; Liu et al. 2015a). *GhDEL65*, identified and cloned as the homolog of *GL3*, could partly complement the smooth phenotype in *gl3/egl3* mutant in Arabidopsis and positively regulate fiber development in cotton (Shangguan et al. 2016). Another bHLH

transcription factor, *GhbHLH1*, specifically expressed in the early developmental stage and was involved in quick fiber elongation (Meng et al. 2009). Family members of WD40 transcription factors, e.g., *GhTTG1*, *GhTTG3*, shared high sequence similarity with *AtTTG1* and could rescue the phenotype of *ttg* mutant in Arabidopsis (Humphries et al. 2005; Tian and Zhang 2021).

Moreover, several HD-Zip transcription factor genes were cloned and identified in cotton, and these homologs of *AtGL2* comprise *ML1* (MERISTEM LAYER1), *HD1*, and *HOXs*. Ectopic expression of *GbML1* in Arabidopsis could increase the number and density of the trichome of leaves and stems (Zhang et al. 2010). According to the overexpression and knockdown experiments in *G. arboreum* and *G. hirsutum*, *HD-1* has been proved to significantly trigger downstream genes and prompt fiber and trichome formation by regulating the calcium ion signaling pathway and the accumulation of ethylene and ROS (Walford et al. 2012; Liu et al. 2020a). In addition, genetic complementary experiments showed that could fully complement the hairless phenotype of *gl2* in Arabidopsis, and *in situ* hybridization revealed a relatively high expression level in fiber-attached ovules at 0 and +1 DPA in wild-type cotton and rarely expressed in fiber mutant (Guan et al. 2008). Also, genetic tests showed that the elevated expression level of *GhHOX3* in tetraploid cotton highly correlated with the ideal fiber length, and *vice versa* (Shan et al. 2014a; Pei 2015). With the participation of the gibberellin pathway, *GhHOX3* could interact with *GhHD1*, another homeobox gene and regulates fiber development, to activate the expression of downstream genes (e.g., cell wall loosening protein genes, *GhRDL1* and *GhEXPA1*) to affect fiber elongation thus (Cao et al. 2020; Wang et al. 2021).

Except for the core genes of the MBW complex, other types of transcription factors also have influence on regulating fiber development. *GhTCP4*, the atypical bHLH transcription factor, antagonistically interacts with *GhHOX3* to balance the process of fiber elongation and cell wall synthesis through miR319-mediated expression pattern (Cao et al. 2020). *GhPRE1* also contains a noncanonical conserved DNA binding domain and is clustered into TCP TFs. In tetraploid cotton, *GhPRE1* in $A_t$-subgenome specifically expressed in fiber cells during elongation stage, whereas *GhPRE1* in $D_t$-subgenome failed to trigger its transcription due to a *cis*-element deletion of TATA-box in promoter sequence (Zhao et al. 2018). *GhFSN1* was identified as an NAC-type transcription factor and exhibited a high expression level at fiber secondary cell wall synthesis stages. Ectopic expression of *GhFSN1* and its known upstream genes such as *GhMYBL1* and *GhKNL1* presented shorter fiber (Gong et al. 2014). Still, it increased cell wall thickness, indicating *GhFSN1* and its regulatory network participated in the process of fiber elongation transition to second cell wall synthesis stage (Zhang et al. 2018; Sun et al. 2020). Recently, a plant-specific transcription factor, GhWRKY16, has been verified to be phosphorylated by *GhMPK3-1* and positively regulate fiber initiation and elongation through directly activating the downstream expressions of *GhHOX3*, *GhMYB109*, *GhCesA6D-D11*, and *GhMYB25* (Wang et al. 2021).

## *3.3 Phytohormones and their signaling pathways play important roles during fiber developmental stages.*

Phytohormones and growth regulators play essential roles in plant development, and

cotton fiber is no exception. Many studies have shown that hormones highly contribute to fiber development. The establishment of an in vitro culture system for cotton ovules particularly allowed researchers to directly observe the single effect or crosstalk of multiple phytohormones on fiber development (Beasley and Ting 1973, 1974). Besides, the determination of endogenous hormone concentration in fiber cells in vivo can deepen the understanding of the role of the hormone in the process of fiber initiation and elongation to a certain extent (Kim and Triplett 2001; Kim et al. 2015).

Auxin is reported to play crucial roles during fiber initiation and elongation phase. The endogenous auxin in fiber undergoes a dynamic variation process: it begins to accumulate from the flowering day, peaks at 2 to 3 DPA, and gradually descends to the initial level till 10 DPA. In vitro ovule culture proved that exogenous treatment of IAA would result in a dominantly increasing number of fiber initials and longer fiber length (Seagull and Giavalis 2004; Zhang et al. 2017b). Using the auxin monoclonal antibody, the auxin concentration was detected. A high level of IAA accumulation was found in initiated fiber cells rather than non-initiated cells on the flowering day. While in vivo application of auxin transport inhibitor, e.g., 1-N-naphthylphthalamic acid (NPA), can impede IAA accumulation and simultaneously suppress the fiber initiation process. Zhang et al. utilized the specific promoter of cotton ovule epidermal to trigger the overexpression of *iaaM*, an important enzyme involved in the auxin synthetic pathway (Zhang et al. 2011). Consequently, the auxin concentration was lifted directionally and moderately during fiber initiation stage, and thus promoted the fiber cell initiation and significantly improved fiber yield and quality at the same time. Genes involved in auxin signaling pathway have also identified to play important roles in fiber initiation and development. *GhARF2* and *GhARF18* exhibited higher expression levels in the fiber initial stage. The elevating expression of these two genes in Arabidopsis would promote trichome initiation, suggesting that *GhARF2* and *GhARF18* may positively regulate fiber initiation and development (Xiao et al. 2018). In contrast, *GhIAA16* is thought to be a negative regulatory factor for fiber initiation, and its expression is lower in wild-type cotton ovules and higher in fiberless mutants (FL) (Jin-Feng et al. 2004; Xiao et al. 2019; Han et al. 2012). Treating exogenous IAA could stimulate the expression level of *GhTCP14* at initiation and elongation phages of fiber cells (Wang et al. 2013b; Cai et al. 2018). And *GhTCP14* directly binds to the promoters of downstream genes *AUX1* (auxin uptake carrier), *IAA3* (auxin response protein), and *PIN2* (auxin efflux carrier). It triggers them to highly express and thus regulate fiber initiation and elongation. Studies have shown that auxin in fibroblasts mainly derives from the outer epidermis of ovules. Polar auxin transporters (GhPINs) can transport auxin from ovules to mature fiber. The specific suppression of *GhPINs* (especially *GhPIN1c* and *GhPIN3a*) in the ovule of transgenic cotton leads to a hindered fiber initiation and elongation (Zhang et al. 2017b; Mei and Zhang 2019). However, cytokinin (CK) was reported to have an antagonistic effect against auxin during fiber initiation. Abundant cytokinin can down-regulate *GhPIN3a* and disturb the polar localization of *GhPIN3a* protein in non-fiber cells, thus undermining the asymmetric auxin accumulation in epidermal cells of cotton ovule and restraining the fiber initiation and elongation (Zeng et al. 2019). In addition, feeding experiments have revealed that elevated CK level (>5 μM) inhibits the growth of fiber cells in vitro, whereas low concentration (<0.5 μM) promotes fiber initiation and increases

production. What's more, suppression of CK dehydrogenase (*GhCKX3*) would result in an increasing level of the endogenous CK concentration and thus impede fiber initiation, suggesting CK played a negative role in fiber development (Xu et al. 2019; Zeng et al. 2019).

In addition to auxin and cytokinin, gibberellins (GAs), jasmonic acid (JA), ethylene (ETH), and brassinosteroids (BRs) also participate in the regulation of fiber initiation and development. Exogenous gibberellin significantly promotes cotton fiber elongation (Kim and Triplett 2001; Samuel Yang et al. 2006), while ovules treated with gibberellin inhibitors in vitro have less and shorter fibers than the control (David et al. 2016; Seagull and Giavalis 2004). The content of endogenous gibberellin in fibers soars to the highest during the rapid elongating phase (10DPA) and decreases rapidly. And long fiber varieties also exhibit a significantly higher concentration of endogenous $GA_3$ than short fiber varieties. Genetic tests have revealed that ectopic expression of *GhGA20ox1* boosted the accumulation of $GA_3$ and $GA_4$ in cotton fibers and promoted the production of long fibers (Xiao et al. 2010). Previous studies have shown that GhHOX3 interacts with DELLA and GhSLR1 proteins in lower GA environment. While GhHOX3 interacts directly with GhHD1 to trigger the expression of downstream *GhRDL1* and *GhEXPA1* through ubiquitination of *GhSLR1* to regulate fiber elongation in a GA-dependent signaling pathway (Shan et al. 2014b; Shangguan et al. 2016; Cao et al. 2020).

Jasmonic acid is associated with biotic and abiotic stresses, but it has also been reported to be involved in trichome development (Roberts 2016). Transcriptome analysis suggests that the metabolic pathway of jasmonic acid may be related to fiber initiation (Wang et al. 2015a). Further genetic and biochemical evidence revealed that GhJAZ2 could interact with the MBW complex to inhibit the initiation of cotton fibers (Li et al. 2014a; Hu et al. 2016; He et al. 2017).

Ethylene makes excellent contributions to plant growth and development, such as root hair development, hypocotyl growth, and apical hook formation (Dubois et al. 2018). Shi et al. first found that ethylene biosynthesis is one of the most significant upregulated pathways in the process of fiber elongation through the analysis of gene microarray (Shi et al. 2006). Exogenous treatment of ethylene *in vitro* culture of ovule can significantly promote fiber elongation, while the application of an inhibitor of ethylene biosynthesis AVG can notably inhibit it. *GhACO* (ACC oxidase, ACO), an important gene of the rate-limiting enzyme in the ethylene synthesis pathway, had the highest expression level at 10-15 DPA during fiber rapid elongation. The accumulation level of ACO in fiber was significantly higher than that in the ovule. However, the expression levels of ACO1 and ACO3 were also relatively high in *G. ramondii* (with a characteristic of short lint), suggesting that excessive ethylene accumulation may cause premature maturation of fiber and the failure for fiber to develop into lint (Yang et al. 2015; Li et al. 2010). Besides, previous studies revealed that abundant exogenous ethylene application in ovule culture system could cause the accumulation of reactive oxygen species and thus to regulate fiber elongation. Exogenous application of very long-chain fatty acids (VLCFAs) can promote ethylene synthesis and fiber development. Ethylene could rescue the impeded fiber development caused by the VLCFAs-deficient treatment through the relevant inhibitors. The above results imply that ROS may act downstream, whereas VLCFAs

act upstream in the ethylene signaling pathway to regulate fiber elongation and development (Qin and Zhu 2011; Qin et al. 2007).

Brassinolide is another crucial hormone that positively regulates fiber initiation and elongation *in vitro* ovule culture (Sun et al. 2005). Low concentrations of brassinolide promote fiber development, while application of brassinazole (BRZ), a biosynthetic inhibitor of brassinolide, suppresses this process. *GhDET2* encodes a steroid 5α-reductase in brassinolide biosynthetic (Luo et al. 2007). *GhFP1*, identified as the key regulator of *GhDET2*, could bind with the E-box elements at the promoter regions of *GhDWF4* and *GhCPD*, which are also related to BR synthesis (Liu et al. 2020b). Both ectopic expressions of *GhDET2* and *GhFP1* exhibited more and longer fiber, whereas silencing experiments revealed the opposite pattern. *GhPAG1* is homologous to Arabidopsis *CYP734A1* and inactivates endogenous BRs through C-26 hydroxylation, thus negatively regulating fiber development by inhibiting gene expression patterns involved in VLCFAs biosynthesis, ethylene-mediated signaling, calcium signal response, and cell growth regulation (Wu et al. 2021; Yang et al. 2014). The BR receptor *GhBRI1* in cotton is also highly expressed during fiber elongation and can completely complement the phenotype of Arabidopsis *bri1-5* mutants (Sun et al. 2004). *GhBZR1* is a critical transcription factor in response to brassinolide signals in cotton. Although the function of *GhBZR1* independently contributed to fiber development has not been reported yet, it can interact with Gh14-3-3L and bind to the promoter region of *GhXTH1* and *GhEXP* during fiber growth. Overexpression of *Gh14-3-3L* alone will also facilitate the elongation process of mature fibers. In contrast, silencing will cause the reverse, and the short fiber phenotype of RNAi transgenic plants could be partially compensated by exogenous application of brassinolide (Zhou et al. 2015).

Abscisic acid (ABA) plays a negative regulatory role in fiber elongation. Additional ABA treatment *in vitro* culture of ovules inhibited the elongation of fiber cells and the initial numbers of fiber cells (Nayyar et al. 1989). As reported, dynamic change of ABA accumulation occurred during fiber initiation to maturation. The ABA content is high at the initial stage and decreases from the elongation stage until cell dehydration and maturation stage, in which phase the ABA accumulates again. Furthermore, the endogenous content of ABA in ovules of *Ligon-lintless2* (*Li*$_2$) mutant, which exhibited an extremely short fiber characteristic, was much higher than that in wild type cotton with regular fiber, implying ABA is an inhibitor of fiber elongation and development (Gilbert et al. 2013).

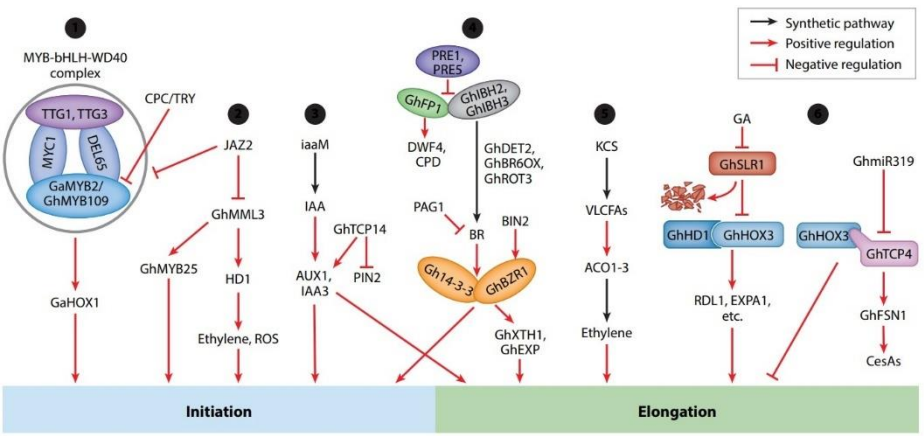


**Figure 1-2:** Partial regulatory networks during cotton fiber initiation and elongation. (Huang et al., 2020)

## *3.4 Other factors regulate fiber development*

Fiber development undergoes quite intricate process. In addition to the regulation of core TFs and phytohormones, other factors such as metabolites, turgor pressure, cytoskeleton, and signal transduction are also essential for fiber formation and development (Gou et al. 2007; Ruan 2007).

Fatty acids and sugars are crucial substrates for maintaining turgor pressure, thus promoting fiber elongation and serving as energy sources for life activities, cellulose synthesis and cell wall formation during fiber development (Ruan 2007). As we know, the accumulation of long-chain fatty acids can regulate ethylene synthesis and promote fiber elongation. However, the expression of enzyme FDH in the long-chain fatty acid synthesis pathway was inhibited in the fuzzless mutant $N_1$ (Lee et al. 2006). Phosphatidylinositol monophosphate transporter *GhLTPG1* is reported to regulate fiber elongation positively (Deng et al. 2016). Sucrose synthase (Sus) could degrade sucrose into UDP-glucose and fructose and play an important role in fiber initiation and secondary cell walls synthesis (Haigler et al. 2007). However, varieties with higher levels of sugar in fibers are always accompanied shorter fibers (Xu et al. 2012). Immunohistochemistry showed that SUS protein exhibited the highest expression in epidermal cells of wild-type (WT) ovules of 5 DPA but not in ovules with same ages from lintless mutant FLS (*SL1-7-1*), which has short fibers instead of long fibers. Elevating the expression level of *GhSUSA1* can promote fiber elongation (Jiang et al. 2012). The transcription factor *GhMYB212* acted as a key regulator to control fiber elongation by promoting sucrose transport from ovule to fiber (Sun et al. 2019a). Recent studies have shown that ectopic expression of *SUTs* (SUCROSE TRANSPORT PROTEINS) or *GhSWEETs* (SUGARS WILL EVENTUALLY BE EXPORTED TRANSPORTERs) driving by a seed coat-specific promotor in cotton enhanced the sugar contents in fiber-attached ovules, causing dominantly improved initial density but decreased length of fibers (Cox et al. 2017). They further confirmed that the ROS accumulation was induced by the increased sugar content and thus triggered the fiber initial process and generated much more fibers. Besides, the high ROS level also governed the SCW biosynthesis and, in turn, impeded the fiber elongation and accordingly produced thinner and shorter fibers.

Silencing the steroid transporter gene *GhSCP2D*, which is responsible for sucrose transport, will lead to the closure of plasmodesmata and thus hinder fiber development (Zhang et al. 2017c).

The cell morphology undergoes a highly polarized change along with fiber development, and the actin cytoskeleton plays crucial role in this process. Previous studies have shown that interfering with *GhACTIN1* would cause a notably reduced fiber length but not affect the initial process by disorganizing the regulatory network of the actin cytoskeleton (Iqbal et al. 2021; Qin and Zhu 2011). Conversely,

downregulation of *GhADF1* that encodes an actin-depolymerizing factor could promote both fiber length and strength (Wang et al. 2009; Nan et al. 2009). *GhFIM2* encodes an actin-bundling protein and positively facilitates the formation of the actin bundle to accelerate fiber growth (Zhang et al. 2017a). Ectopic expression of the fiber-preferential actin-binding protein *GhPFN2* generates thicker and ampler *F*-actin bundles. It thus leads to the premature termination of cell elongation as well as the shortened fiber length. Besides, *GhTUB1*, which encodes tubulin, also affects cytoskeleton assembly and fiber elongation (Li et al. 2002). The LIM-domain protein GhWLIM1a can promote the synthesis of the secondary cell wall by binding to tubulin (Han et al. 2013). The BURP-containing protein GhRDL1 interacts with the cell expansion protein GhEXPA1 to participate in the cell wall loosening during fiber elongation and cell wall synthesis (Li et al. 2002; Li et al. 2005).

Major signaling factors, such as $Ca^{2+}$ and ROS, also play essential roles in fiber initiation and elongation (Qin and Zhu 2011; Tang et al. 2014; Yang et al. 2016). $Ca^{2+}$ influx significantly increased during fiber elongation, peaked at 10 DPA to 15 DPA, and attenuated from the cell wall synthesis phase. *GhCaM7*, which encodes the Calmodulin protein involving $Ca^{2+}$ signal transduction, predominately expressed in the initial and elongating stages. Overexpression of *GhCaM7* can expedite rapid elongation of fiber cells and whereas RNAi interference lines exhibited delayed and impeded fiber initiation and elongation (Tang et al. 2014).

# 4. Research progress on fiber mutants

Cotton fiber mutants are the materials that mutate in various fiber phenotypes compared with the wild type. They are important resources and ideal materials to study fiber differentiation and development's molecular and genetic mechanism. According to their fiber characteristics, fiber mutants can be divided into three types (Figure 1-3):

1. ultra-short fiber mutants, with only ultra-fuzz attached on the epidermis of ovules.
2. fuzzless mutants, with lint but not fuzz on the seed coats.
3. fiberless mutants, without both fuzz and lint on the surface of seeds.

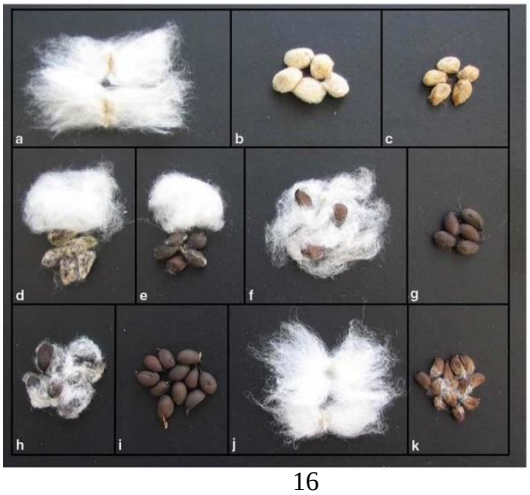


**Figure 1-3:** Wild and fiber mutant seeds. a Wild type (tetraploid cotton); b-c ultra-short fiber mutants, $Li_1$ and $Li_2$; d-i fuzzless mutants (tetraploid cotton); j wild type (diploid cotton); k fiberless mutant, *sma-4*. (Rong et al., 2005)

With the release of high-quality sequencing data from multiple cotton species, conjoint analyses through the traditional fine-mapping and bioinformatics-based tools have also facilitated the exploration of functional gene mapping and network regulation (Wang et al. 2012; Li et al. 2014b; Liu et al. 2015b; Zhang et al. 2015; Du et al. 2018; Ma et al. 2018b; Wang et al. 2019b). So far, plenty of research progress has been made in both tetraploid and diploid fiber mutant lines or species.

## *4.1 Research progress on ultra-short fiber mutants*

The fiber elongation process is the vital determinant of fiber length and thus affects fiber quality desired in the textile industry. Understanding the genetic mechanisms and molecular basis underlying fiber elongating can provide valuable gene resources to facilitate the improvement of fiber quality and yield. Ultra-short fiber mutants were used frequently as excellent model systems to study the molecular mechanisms and regulatory networks associated with fiber elongation and development.

*Ligon lintless1* was the first reported ultra-short fiber mutant which presented extremely short fibers (~5 mm), twisted and dwarf plants, and curled leaves and stems (Rong et al. 2005; Thyssen et al. 2014). Population genetic analysis revealed a single dominant gene controlled the ultra-short fiber. Previous studies have mapped this gene to chromosome 22 (D4) and narrowed the candidate mapping region to a genetic interval of 0.3 cM using molecular markers and genetic populations (Gilbert et al. 2013; Gilbert et al. 2014; Naoumkina et al. 2015). Based on the mapping-by-sequencing and the subsequent function verification by VIGS, Thyssen et al. proposed that the amino acid substitution of a single Gly65Val on GhACT_LI1 (Gh_D04G0865) would be the causative mutation governing the dominant short fiber phenotype in *Ligon-lintless* $Li_1$ mutant (Thyssen et al. 2017). Coincidently, Sun et al. and Cao et al. also obtained this candidate gene through comprehensive strategies including map-based cloning as well as comparative sequencing, and further proved this substitution mutation disturbed the polymerization of *F*-actin and the establishment of actin cytoskeleton-based biological processes, which resulted in dominant-negative effects of on fiber cell elongation and plant morphogenesis (Cao et al. 2021; Sun et al. 2019b). *Ligon lintless x* mutant has similar deformed leaves and stem as well as short fiber with $Li_1$. Cai et al. mapped the candidate region, which is tightly linked with the $li_x$ phenotype on chromosome 4 (A4), but the crucial gene remains unclear (Cai et al. 2013). *Ligon lintless y* controlled by a monogenic recessive locus was EMS-mutagenized cotton line with short fibers and the stunted plant length (Naoumkina et al. 2017a). According to the genetic analysis and GWAS results, a novel QTL hotspot containing Gh_A08G0716 and GH_A08G0783 revealed a quite high correlation with lint percentage (LP) and seed index (SI). In addition, the Gh_A07G1389, a tetratricopeptide repeat (TPR)-like encoding gene, was involved in the short fiber formation process in $Li_y$ mutant, implying important roles in lint percentage and fiber

yield (Naoumkina et al. 2017a). However, based on BSA-Sequencing and map-based cloning, a threonine/isoleucine substitution mutation in the coding gene (Ghir_A12G008870) of the other one tetratricopeptide repeat-like superfamily protein was detected between $li_y$ and its wild type and was considered as the only one promising candidate answerable for the phenotype of $li_y$ mutant (Fang et al. 2020). Compared with $Li_1$, $Li_x$ and $Li_y$, $Li_2$ exhibited average vegetable growth. Multiple previous studies in this mutant have revealed the coarse linkage map and genetic positions, differentially expressed genes, and metabolite changes (Thyssen et al. 2014; He et al. 2021). Naoumkina et al. found genes related to xyloglucan biosynthesis showed significantly differential expression patterns between $Li_2$ and its near-isogenic wild type, suggesting accumulation of xyloglucan may affect the elongation of fiber cells of $Li_2$ (Naoumkina et al. 2017b). In a recent study, a 176- to 221-kb segment deletion on chromosome 18 was reportedly responsible for the extremely short fiber phenotype of $Li_2$. Within the 35 candidate genes, *GhUGT87A1* and *GhUGT87A2* showed the dominant significance of expression levels between the two materials (Patel et al. 2020). Besides, the fiber length showed a notable reduction after performing the VIGS experiment, suggesting that these two homolog genes may be the most likely causable factors contributing to the $Li_2$ phenotype.

## *4.2 Research progress on fiberless mutants*

The number of initiated fiber cells determines the ultimate cumulative yield. Hence it is of great significance to explore the orchestrating network governing fiber cell differentiation and fiber formation. Fiberless and fuzzless mutants are valuable germplasm resources to study the genetic and molecular mechanism corresponding to the initiation and development of fiber or fuzz cells.

Till now, multiple fiberless (*fl*) mutants have been found and widely used, e.g., cv. XZ142FLM, MCU5, XinWX, MD17, and *SL-1-7-1* (Rong et al. 2005; Padmalatha et al. 2012; Turley 2002; Turley and Kloth 2003). Since polygenic loci controlled and regulated both fiber and fuzz development of *fl* mutants, it is much more intricate to dig out the distinct genes which contribute to fiber or fuzz formation. In former research, based on the multi-omics analyses, predecessors suggested phytohormonal signaling networks, miRNAs and protein acetylation may play important roles in fiber development and failed to obtain the precise genes directly responsible for fiber and fuzz initiation (Wang et al. 2015a; Sun et al. 2017; Singh et al. 2020). Wu et al. constructed an excellent segregating population derived from the cross of fuzzless line n2NSM and fiberless line XZ142FLM. They then identified the tandemly arranged *GhMML4_D12* ($li_3$) and *GhMML3_D12* ($n_2$) as the candidate genes contributing to fiber and fuzz formation in XZ142, respectively (Wu et al. 2017). However, Chen et al. proposed that the loss-of-function allele *GhMML3_D12* ($li_3$) containing a retrotransposon insertion should be the best candidate responsible for the fiberless trait in XZ142FLM using map-based cloning (Chen et al. 2020). At the same time, the $n_2$ locus was possibly allelic to $li_3$ and co-located with $li_3$ to participate in fuzz initiation and formation.

Compared with tetraploid cotton, diploid *G. arboreum*, which has a relatively simple genome, is potentially highly suitable for digging candidate genes corresponding to mutant traits and exploring the mechanism and regulatory network underlying these

genes. *GaHD1*, located on chromosome A06, has been reported to be responsible for fiberless seeds and glabrous stems in a recessive fiberless mutant *sma-4* (*ha*) by combining traditional map-based cloning with transcriptome analysis (Ding et al. 2020). Transcript splicing mistakes occurred because of a critical point mutation in *GaHD1* and further affects fiber and trichome initiation and retards elongation by cellular $H_2O_2$ and $Ca^{2+}$ signals. Liu et al. also established *GaHD-1* as the candidate gene responsible for the lintless trait in diploid fiberless mutant *A2-106* (Liu et al. 2020a).

## *4.3 Research progress on fuzzless mutants*

Compared with lint fiber, fuzz tightly adheres to seed coats and is difficult to gin, thus causing more energy costs and spreading pathogenic bacteria such as *Colletotrichum gossypii* Southw, *Fusarium monihforme*, *Rhizoctomia solani* Kühn, and others (Afanador-Kafuri et al. 2014; Varo et al. 2016a,b; Baird et al. 2000). Therefore, it is necessary to study or create fuzzless seed germplasms and find out the formation and regulatory mechanism underlying fuzz differentiation, initiation, and development. Though most research focused on fiber initiation and elongation, the key functional genes and the fuzz initiation and development mechanisms remain largely elusive.

Several loci have been identified and assigned by genetic and comprehensive analysis in tetraploid fuzzless mutant lines (Rong et al. 2005; Wan et al. 2014; Ma et al. 2016). The $N_1$ locus was the first reported mutation to dominantly and monogenetically control fuzz fiber development. In previous studies, the $N_1$ locus has been determined to be located in At-subgenome A12 by genetic linkage analysis and traditional mapping strategy. Thanks to the available release of cotton genomes and advances in bioinformatics and high-throughput genotyping technologies, *GhMYB25-like*, which encodes a MYBMIXTA-like transcription factor 3 (GhMML3_A12), has been identified to be associated with fuzz fiber initiation and development. In the $N_1$ mutant, due to the 3' antisense promoter, *GhMML3_A12* bidirectionally transcribed and produced natural antisense transcripts (NAT), thus forming the double-stranded RNA to generate small interfering RNAs (siRNAs) and reducing the gene expression level. The *silenced-MYB25-like* transgenic lines resulted in fuzzless or even fiberless cotton seeds (Wan et al. 2016). The $n_2$ locus has been proposed and documented as a recessive fuzzless mutation derived from *G. barbadense* species. A recent study using SNP63K array and mapping-by-sequencing strategy has found multiple loci controlling the recessive fuzzless trait in *G. barbadense* and further identified *GhMYB25like_D12* as $n_2$ candidate locus, which mainly determines fuzz initiation and development (Zhu et al. 2018). The genomic location of the recessive $n_3$ remains to be determined. Recently, Chen et al. proposed that the recessive fuzzless loci $n_2$ and $n_3$ in *Gossypium hirsutum* were identified to be overlapped with $li_3$ and $N_1$, respectively, indicating that they are likely the multiple alleles of the known fuzzless genes, respectively (Chen et al. 2020). And thus, it has been suspected that the homologs of *MYB25-like* may tempo-spatially regulate the initiation and development of both lint and fuzz.

An EMS-induced $n_4^t$ mutant was reported as a single recessive tufted-fuzzless mutant with fuzz only appearing near the micropyle and chalazal of cotton seeds. The

gene controlling $n_4^t$ phenotype is nonallelic to $N_1$, $n_2$ or $n_3$ (Bechere et al. 2012; Percy et al. 2015). Naoumkina et al. developed the segregation population and mapped the $n_4^t$ gene to a 411 kb-interval of chromosome D04, with seven candidate genes significantly expressing between the mutant and the wild type (Naoumkina et al. 2021b; Naoumkina et al. 2021a). Recently, a new tufted-fuzzless mutant has been reported and designated $N_5$, controlled by a monogenic and incomplete dominant gene. Using high-throughput sequencing and traditional fine-mapping, the mutation was determined to be an approximately 250-kb genomic region on chromosome D13, including 33 annotated candidates (Zhu et al. 2021). The $n_4^t$ and $N_5$ mutations negatively affect lint development and produce higher lint percentages than $N_1$ and $n_2$ mutants. Thus, they are potentially more usable and valuable in breeding and generating new fuzzless cultivars. However, the precise genes governing phenotypes in these two mutants remain to be further determined.

Relatively rare research progress has been reported about fuzzless mutants in diploid cotton lines. In previous studies, Feng et al. first proposed that the fuzzless phenotype was controlled by a single dominant gene in *G. arboreum* DPL972 and closely linked with 11 SSR markers distributed on chromosome 08. Subsequently, this candidate region was identified to a 600-kb interval on Chrs08 through genome-wide association analysis and genetic mapping from 215 fuzzy and fuzzless accessions (Du et al. 2018). Meanwhile, combined BSA-Seq, RNA-Seq with genetic linkage analysis, Feng et al. fine-mapped this gene to a 70-kb genomic interval and proposed *GaGIR1* (GLABRA2-interacting repressor, Ga08G0121) as the most likely candidate gene due to the sequence variations and the notable differential expression levels between the fuzzless mutant and the wild type (Feng et al. 2019). Through correlation analysis and genetic linkage analyses, Liu et al. demonstrated *GaGIR1* and *GaMYB25-like* are crucial candidate genes controlling fuzz development in seven fuzzless *G. arboreum* accessions. However, in contrast to the pivotal roles in fiber and fuzz initiation and growth in upland cotton, *GaMYB25-like* seems to have more negligible role in relation to fuzz formation than *GaGIR1*. In other words, *GaGIR1* exhibited significant epistasis to *GaMYB25-like* during fuzz development in diploid cotton (Liu et al. 2020a). A recent study has suggested a 6.2-kb insertion located about 18 kb upstream of *GaFZ* (Ga08G0121) was responsible for the fuzzless phenotype in DPL972 based on a genome-wide association study (GWAS). The large InDel acted as an enhancer to thus elevate the expression level of *GaFZ* and negatively regulate fuzz and trichome formation. In addition, *GaFZ* may function by directly repressing the VLCFA pathway to regulate fuzz initiation and elongation, indicating the regulatory network independently distinguished from the MBW-GL2 system (Wang et al. 2020). Although much progress has been achieved, the exact mechanism and regulatory network still need further validate and study.

# References

Afanador-Kafuri L, González A, Gañán L, Mejía JF, Cardona N, Alvarez E (2014) Characterization of the Colletotrichum Species Causing Anthracnose in Andean Blackberry in Colombia. Plant Dis. Nov;98(11):1503-1513. doi: 10.1094/PDIS-07-13-0752-RE. PMID: 30699787.

Akhtar MQ, Qamar N, Yadav P, Kulkarni P, Kumar A, Shasany AK (2017) Comparative glandular trichome transcriptome-based gene characterization reveals reasons for differential (−)-menthol biosynthesis in Mentha species. Physiologia plantarum 160 (2):128-141

Alford BB, Liepa GU, Vanbeber AD (1996) Cottonseed protein: What does the future hold? Plant Foods for Human Nutrition 49 (1):1-11. doi:10.1007/BF01092517

An L, Zhou Z, Sun L, Yan A, Xi W, Yu N, Cai W, Chen X, Yu H, Schiefelbein J, Gan Y (2012) A zinc finger protein gene *ZFP5* integrates phytohormone signaling to control root hair development in Arabidopsis. Plant J 72 (3):474-490. doi:10.1111/j.1365-313X.2012.05094.x

Andreuzza S (2020) UnMixta-ing trichome development in tomato. Plant Cell 32 (5):1342-1343. doi:10.1105/tpc.20.00224

Applequist WL, Cronn R, Wendel JF (2001) Comparative development of fiber in wild and cultivated cotton. Evolution & development 3 (1):3-17

Bac-Molenaar JA, Mol S, Verlaan MG, van Elven J, Kim HK, Klinkhamer PGL, Leiss KA, Vrieling K (2019) Trichome independent resistance against western flower thrips in tomato. Plant Cell Physiol 60 (5):1011-1024. doi:10.1093/pcp/pcz018

Baird R, Batson W, Carling D, Scruggs M (2000) First Report of Rhizoctonia solani AG-7 on Cotton in Mississippi. Plant Dis. Oct;84(10):1156. doi: 10.1094/PDIS.2000.84.10.1156B. PMID: 30831928.

Balkunde R, Pesch M, Hülskamp M (2010) Chapter ten - trichome patterning in arabidopsis thaliana: from genetic to molecular models. In: Marja CPT (ed) Current Topics in Developmental Biology, vol Volume 91. Academic Press, pp 299-321. doi:http://dx.doi.org/10.1016/S0070-2153 (10)91010-7

Basra AS, Malik C (1984) Development of the cotton fiber. International review of cytology 89:65-113

Beasley C, Ting IP (1973) The effects of plant growth substances on *in vitro* fiber development from fertilized cotton ovules. American Journal of Botany 60 (2):130-139

Beasley C, Ting IP (1974) Effects of plant growth substances on in vitro fiber development from unfertilized cotton ovules. American Journal of Botany 61 (2):188-194

Beasley J (1940) The origin of American tetraploid *Gossypium* species. The American Naturalist 74 (752):285-286

Beasley J (1941) Hybridization, cytology and polyploidy of *Gossypium*. Chron Bot 6:394-395

Bechere E, Turley RB, Auld DL, Zeng L (2012) A New Fuzzless Seed Locus in an Upland Cotton (*Gossypium hirsutum* L.) Mutant. American Journal of Plant Sciences 03 (06):799-804. doi:10.4236/ajps.2012.36096

Cai C, Tong X, Liu F, Lv F, Wang H, Zhang T, Guo W (2013) Discovery and identification of a novel Ligon lintless-like mutant ($Li_x$) similar to the Ligon lintless ($Li_1$) in allotetraploid cotton. Theoretical and Applied Genetics 126 (4):963-970

Cai Y, Zheng K, Wang Y, He H, Qu Y, Ni Z, Chen Q (2018) Cloning and expression

analysis of *GbTCP14* gene in *Gossypium barbadense*. Acta Botanica Boreali-Occidentalia Sinica 38 (3):401-408

Cao JF, Zhao B, Huang CC, Chen ZW, Zhao T, Liu HR, Hu GJ, Shangguan XX, Shan CM, Wang LJ, Zhang TZ, Wendel JF, Guan XY, Chen XY (2020) The mir319-targeted *GhTcp4* promotes the transition from cell elongation to wall thickening in cotton fiber. Mol Plant 13 (7):1063-1077. doi:10.1016/j.molp.2020.05.006

Cao Y, Huang H, Yu Y, Dai H, Hao H, Zhang H, Jiang Y, Ding M, Li F, Tu L, Kong Z, Rong J (2021) A Modified Actin (Gly65Val Substitution) Expressed in Cotton Disrupts Polymerization of Actin Filaments Leading to the Phenotype of Ligon Lintless-1 (Li1) Mutant. Int J Mol Sci 22 (6). doi:10.3390/ijms22063000

Challa KR, Aggarwal P, Nath U (2016) Activation of *YUCCA5* by the transcription factor TCP4 integrates developmental and environmental signals to promote hypocotyl elongation in Arabidopsis. Plant Cell 28 (9):2117-2130. doi:10.1105/tpc.16.00360

Challa KR, Rath M, Nath U (2019) The CIN-TCP transcription factors promote commitment to differentiation in Arabidopsis leaf pavement cells *via* both auxin-dependent and independent pathways. PLoS Genet 15 (2):e1007988. doi:10.1371/journal.pgen.1007988

Chalvin C, Drevensek S, Dron M, Bendahmane A, Boualem A (2020) Genetic control of glandular trichome development. Trends Plant Sci 25 (5):477-487. doi:10.1016/j.tplants.2019.12.025

Chen W, Li Y, Zhu S, Fang S, Zhao L, Guo Y, Wang J, Yuan L, Lu Y, Liu F, Yao J, Zhang Y (2020) A retrotransposon insertion in *GhMML3_D12* is likely responsible for the Lintless locus $li_3$ of tetraploid cotton. Front Plant Sci 11:593679. doi:10.3389/fpls.2020.593679

Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD, Arioli T, Brubaker C, Cantrell RG, Lacape JM, Ulloa M, Chee P, Gingle AR, Haigler CH, Percy R, Saha S, Wilkins T, Wright RJ, Van Deynze A, Zhu Y, Yu S, Abdurakhmonov I, Katageri I, Kumar PA, Mehboob Ur R, Zafar Y, Yu JZ, Kohel RJ, Wendel JF, Paterson AH (2007) Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol 145 (4):1303-1310. doi:10.1104/pp.107.107672

Constable G, Bange M (2015) The yield potential of cotton (*Gossypium hirsutum* L.). Field Crops Research 182:98-106

Cox KL, Meng F, Wilkins KE, Li F, Wang P, Booher NJ, Carpenter SC, Chen L-Q, Zheng H, Gao X (2017) TAL effector driven induction of a *SWEET* gene confers susceptibility to bacterial blight of cotton. Nature communications 8 (1):1-14

Cox N, Smith LM (2019) A novel upstream regulator of trichome development inhibitors. Plant Physiol 181 (4):1398-1400. doi:10.1104/pp.19.01269

Dai X, Zhou L, Zhang W, Cai L, Guo H, Tian H, Schiefelbein J, Wang S (2015) A single amino acid substitution in the R3 domain of GLABRA1 leads to inhibition of trichome formation in Arabidopsis without affecting its interaction with GLABRA3. Plant Cell Environ. doi:10.1111/pce.12695

David LC, Berquin P, Kanno Y, Seo M, Daniel-Vedele F, Ferrario-Méry S (2016)

N availability modulates the role of NPF3.1, a gibberellin transporter, in GA-mediated phenotypes in Arabidopsis. Planta 244 (6):1315-1328

Deng T, Yao H, Wang J, Wang J, Xue H, Zuo K (2016) GhLTPG1, a cotton GPI-anchored lipid transfer protein, regulates the transport of phosphatidylinositol monophosphates and cotton fiber elongation. Scientific Reports 6 (1):1-12

Devoto A, Nieto-Rostro M, Xie D, Ellis C, Harmston R, Patrick E, Davis J, Sherratt L, Coleman M, Turner JG (2002) COI1 links jasmonate signalling and fertility to the SCF ubiquitin–ligase complex in Arabidopsis. The Plant Journal 32 (4):457-466

Dill A, Sun T-p (2001) Synergistic derepression of gibberellin signaling by removing RGA and GAI function in *Arabidopsis thaliana*. Genetics 159 (2):777-785

Ding M, Cao Y, He S, Sun J, Dai H, Zhang H, Sun C, Jiang Y, Paterson AH, Rong J (2020) *GaHD1*, a candidate gene for the *Gossypium arboreum SMA-4* mutant, promotes trichome and fiber initiation by cellular $H_2O_2$ and $Ca^{2+}$ signals. Plant Molecular Biology:1-15

Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M, Jia Y, Pan Z, Gong W, Liu Z, Zhu H, Ma L, Liu F, Yang D, Wang F, Fan W, Gong Q, Peng Z, Wang L, Wang X, Xu S, Shang H, Lu C, Zheng H, Huang S, Lin T, Zhu Y, Li F (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat Genet 50 (6):796-802. doi:10.1038/s41588-018-0116-x

Dubois M, Van den Broeck L, Inzé D (2018) The pivotal role of ethylene in plant growth. Trends in plant science 23 (4):311-323

Endrizzi J, Turcotte E, Kohel R (1985) Genetics, cytology, and evolution of *Gossypium*. Advances in genetics 23:271-375

Fang DD, Naoumkina M, Thyssen GN, Bechere E, Li P, Florane CB (2020) An EMS-induced mutation in a tetratricopeptide repeat-like superfamily protein gene (*Ghir_A12G008870*) on chromosome A12 is responsible for the *li$_y$* short fiber phenotype in cotton. Theor Appl Genet 133 (1):271-282. doi:10.1007/s00122-019-03456-4

Fang L, Gong H, Hu Y, Liu C, Zhou B, Huang T, Wang Y, Chen S, Fang DD, Du X, Chen H, Chen J, Wang S, Wang Q, Wan Q, Liu B, Pan M, Chang L, Wu H, Mei G, Xiang D, Li X, Cai C, Zhu X, Chen ZJ, Han B, Chen X, Guo W, Zhang T, Huang X (2017) Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. Genome Biol 18 (1):33. doi:10.1186/s13059-017-1167-5

Feng X, Cheng H, Zuo D, Zhang Y, Wang Q, Liu K, Ashraf J, Yang Q, Li S, Chen X, Song G (2019) Fine mapping and identification of the fuzzless gene *GaFzl* in DPL972 (*Gossypium arboreum*). Theor Appl Genet 132 (8):2169-2179. doi:10.1007/s00122-019-03330-3

Fryxell PA (1979) The natural history of the cotton tribe (Malvaceae, tribe Gossypieae). Texas A & M University Press.,

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2 (2):108-165

Gallagher JP, Grover CE, Rex K, Moran M, Wendel JF (2017) A new species of

cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). Systematic Botany 42 (1):115-123

Gan Y, Kumimoto R, Liu C, Ratcliffe O, Yu H, Broun P (2006) GLABROUS INFLORESCENCE STEMS modulates the regulation by gibberellins of epidermal differentiation and shoot maturation in Arabidopsis. The Plant Cell 18 (6):1383-1395

Gan Y, Liu C, Yu H, Broun P (2007a) Integration of cytokinin and gibberellin signalling by Arabidopsis transcription factors *GIS, ZFP8* and *GIS2* in the regulation of epidermal cell fate.

Gan Y, Yu H, Peng J, Broun P (2007b) Genetic and molecular regulation by DELLA proteins of trichome development in Arabidopsis. Plant physiology 145 (3):1031-1042

Gandhi K, Chaudhary N, Litoriya N, Patel N, Talati J (2015) Identification of cotton (*Gossypium herbaceum* L.) Genotypes using Electrophoretic Techniques. Indian Journal of Agricultural Biochemistry 28 (2):122-127

Gilbert MK, Bland JM, Shockey JM, Cao H, Hinchliffe DJ, Fang DD, Naoumkina M (2013) A transcript profiling approach reveals an abscisic acid-specific glycosyltransferase (UGT73C14) induced in developing fiber of *Ligon lintless-2* mutant of cotton (*Gossypium hirsutum* L.). PloS one 8 (9):e75268

Gilbert MK, Kim HJ, Tang Y, Naoumkina M, Fang DD (2014) Comparative transcriptome analysis of short fiber mutants L*igon-lintless 1* and *2* reveals common mechanisms pertinent to fiber elongation in cotton (Gossypium hirsutum L.). PLoS One 9 (4)

Gong S-Y, Huang G-Q, Sun X, Qin L-X, Li Y, Zhou L, Li X-B (2014) Cotton *KNL1*, encoding a class II KNOX transcription factor, is involved in regulation of fibre development. Journal of experimental botany 65 (15):4133-4147

Gou J-Y, Wang L-J, Chen S-P, Hu W-L, Chen X-Y (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. Cell research 17 (5):422-434

Grover C, Gallagher J, Jareczek J, Page J, Udall J, Gore M, Wendel J (2015) Re-evaluating the phylogeny of allopolyploid *Gossypium* L. Molecular phylogenetics and evolution 92:45-52. doi:10.1016/j.ympev.2015.05.023

Guan X, Lee JJ, Pang M, Shi X, Stelly DM, Chen ZJ (2011) Activation of Arabidopsis seed hair development by cotton fiber-related genes. PloS one 6 (7):e21301

Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, Chen XY (2008) The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of the Arabidopsis *GLABRA2*. Physiologia Plantarum 134 (1):174-182

Haigler CH, Betancur L, Stiff MR, Tuttle JR (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Frontiers in plant science 3:104

Haigler CH, Singh B, Zhang D, Hwang S, Wu C, Cai WX, Hozain M, Kang W, Kiedaisch B, Strauss RE (2007) Transgenic cotton over-producing spinach sucrose phosphate synthase showed enhanced leaf sucrose synthesis and improved fiber quality under controlled environmental conditions. Plant molecular biology 63

(6):815-832

Haigler CH, Zhang D, Wilkerson CG (2005) Biotechnological improvement of cotton fibre maturity. Physiologia plantarum 124 (3):285-294

Han L-B, Li Y-B, Wang H-Y, Wu X-M, Li C-L, Luo M, Wu S-J, Kong Z-S, Pei Y, Jiao G-L (2013) The dual functions of *WLIM1a* in cell elongation and secondary wall formation in developing cotton fibers. The Plant Cell 25 (11):4421-4438

Han X, Xu X, Fang DD, Zhang T, Guo W (2012) Cloning and expression analysis of novel *Aux/IAA* family genes in *Gossypium hirsutum*. Gene 503 (1):83-91

He X, Zhu L, Wassan GM, Wang Y, Miao Y, Shaban M, Hu H, Sun H, Zhang X (2017) *GhJAZ2* attenuates cotton resistance to biotic stresses *via* inhibiting the transcriptional activity of *GhbHLH171*. Molecular Plant Pathology. doi:10.1111/mpp.12575

He Z, Nam S, Fang DD, Cheng HN, He J (2021) Surface and thermal characterization of cotton fibers of phenotypes differing in fiber length. Polymers (Basel) 13 (7). doi:10.3390/polym13070994

Hu H, He X, Tu L, Zhu L, Zhu S, Ge Z, Zhang X (2016) *GhJAZ2* negatively regulates cotton fiber initiation by interacting with the R2R3-MYB transcription factor *GhMYB25-like*. The Plant Journal 88 (6):921-935. doi:10.1111/tpj.13273

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 YL, Rahman MU, 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, Zhang T (2019) *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet 51 (4):739-748. doi:10.1038/s41588-019-0371-5

Huang C, Nie X, Shen C, You C, Li W, Zhao W, Zhang X, Lin Z (2017a) Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. Plant Biotechnology Journal 15 (11):1374-1386. doi:10.1111/pbi.12722

Huang C, Nie X, Shen C, You C, Li W, Zhao W, Zhang X, Lin Z (2017b) Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNP s. Plant biotechnology journal 15 (11):1374-1386

Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, John ZY, Zhu Y (2020) Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. Nature genetics 52 (5):516-524

Huchelmann A, Boutry M, Hachez C (2017) Plant glandular trichomes: natural cell factories of high biotechnological interest. Plant Physiology 175 (1):6-22. doi:10.1104/pp.17.00727

Hulskamp M (2004) Plant trichomes: a model for cell differentiation. Nat Rev Mol Cell Biol 5 (6):471-480. doi:10.1038/nrm1404

Humphries JA, Walker AR, Timmis JN, Orford SJ (2005) Two WD-repeat genes

from cotton are functional homologues of the *Arabidopsis thaliana TRANSPARENT TESTA GLABRA1* (*TTG1*) gene. Plant molecular biology 57 (1):67-81

Iqbal A, Latif A, Galbraith DW, Jabbar B, Ali MA, Ahmed M, Gul A, Rao AQ, Shahid AA, Husnain T (2021) Structure-based prediction of protein–protein interactions between GhWlim5 Domain1 and GhACTIN-1 proteins: a practical evidence with improved fibre strength. Journal of Plant Biochemistry and Biotechnology 30 (2):373-386

Ishida T, Kurata T, Okada K, Wada T (2008) A genetic regulatory network in the development of trichomes and root hairs. Annu Rev Plant Biol 59:365-386

Jiang Y, Guo W, Zhu H, Ruan YL, Zhang T (2012) Overexpression of *GhSusA1* increases plant biomass and improves cotton fiber yield and quality. Plant biotechnology journal 10 (3):301-312

Jin-Feng S, Li P, Xiao-e L, Yong-Biao X (2004) Identification of an endothelium-specific gene *GhIAA16* in cotton (*Gossypium hirsutum*). Journal of Integrative Plant Biology 46 (4):472

Johnson CS, Kolevski B, Smyth DR (2002) *TRANSPARENT TESTA GLABRA2*, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. The Plant Cell 14 (6):1359-1375

Kantartzi S, Stewart JM (2008) Association analysis of fibre traits in *Gossypium arboreum* accessions. Plant Breeding 127 (2):173-179

Khaleequr R, Arshiya S, Shafeequr R (2012) *Gossypium herbaceum* Linn: an ethnopharmacological review. Journal of Pharmaceutical and Scientific Innovation (JPSI) 1 (5):1-5

Kim HJ, Hinchliffe DJ, Triplett BA, Chen ZJ, Stelly DM, Yeater KM, Moon HS, Gilbert MK, Thyssen GN, Turley RB (2015) Phytohormonal networks promote differentiation of fiber initials on pre-anthesis cotton ovules grown in vitro and in planta. PLoS One 10 (4)

Kim HJ, Triplett BA (2001) Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. Plant physiology 127 (4):1361-1366

Kirik V, Simon M, Huelskamp M, Schiefelbein J (2004) The *ENHANCER OF TRY AND CPC1* gene acts redundantly with *TRIPTYCHON* and *CAPRICE* in trichome and root hair cell patterning in Arabidopsis. Developmental biology 268 (2):506-513

Kong D, Pan X, Jing Y, Zhao Y, Duan Y, Yang J, Wang B, Liu Y, Shen R, Cao Y, Wu H, Wei H, Wang H (2021) *ZmSPL10/14/26* are required for epidermal hair cell fate specification on maize leaf. New Phytol. doi:10.1111/nph.17293

Kulkarni VN, Khadi BM, Maralappanavar MS, Deshapande LA, Narayanan S (2009) The worldwide gene pools of *Gossypium arboreum* L. and *G. herbaceum* L., and their improvement. In: Genetics and genomics of cotton. Springer, pp 69-97

Larkin JC, Brown ML, Schiefelbein J (2003) How do cells know what they want to be when they grow up? Lessons from epidermal patterning in Arabidopsis. Annual Review of Plant Biology 54 (1):403-430

Larkin JC, Walker JD, Bolognesi-Winfield AC, Gray JC, Walker AR (1999) Allele-specific interactions between *ttg* and *gl1* during trichome development in *Arabidopsis*

*thaliana*. Genetics 151 (4):1591-1604

Laxmi A, Paul LK, Peters JL, Khurana JP (2004) Arabidopsis constitutive photomorphogenic mutant, *bls1*, displays altered brassinosteroid response and sugar sensitivity. Plant molecular biology 56 (2):185-201

Lee JJ, Hassan OS, Gao W, Wei NE, Kohel RJ, Chen X-Y, Payton P, Sze S-H, Stelly DM, Chen ZJ (2006) Developmental and gene expression analyses of a cotton naked seed mutant. Planta 223 (3):418-432

Lee JJ, Woodward AW, Chen ZJ (2007) Gene expression changes and early events in cotton fibre development. Annals of botany 100 (7):1391-1401

Li C, He X, Luo X, Xu L, Liu L, Min L, Jin L, Zhu L, Zhang X (2014a) Cotton *WRKY1* mediates the plant defense-to-development transition during infection of cotton by *Verticillium dahliae* by activating *JASMONATE ZIM-DOMAIN1* expression. Plant Physiol 166 (4):2179-2194. doi:10.1104/pp.114.246694

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, Chen W, Liang X, Shang H, Liu W, Shi C, Xiao G, Gou C, Ye W, Xu X, Zhang X, Wei H, Li Z, Zhang G, Wang J, Liu K, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S (2014b) Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46 (6):567-572. doi:10.1038/ng.2987

Li X-B, Cai L, Cheng N-H, Liu J-W (2002) Molecular characterization of the cotton *GhTUB1* gene that is preferentially expressed in fiber. Plant physiology 130 (2):666-674

Li X-B, Fan X-P, Wang X-L, Cai L, Yang W-C (2005) The cotton *ACTIN1* gene is functionally expressed in fibers and participates in fiber elongation. The Plant Cell 17 (3):859-875

Li Y, Liu D, Tu L, Zhang X, Wang L, Zhu L, Tan J, Deng F (2010) Suppression of *GhAGP4* gene expression repressed the initiation and elongation of cotton fiber. Plant cell reports 29 (2):193-202

Liu B, Zhu Y, Zhang T (2015a) The R3-MYB gene *GhCPC* negatively regulates cotton fiber elongation. PloS one 10 (2):e0116272

Liu X, Moncuquet P, Zhu QH, Stiller W, Zhang Z, Wilson I (2020a) Genetic identification and transcriptome analysis of lintless and fuzzless traits in *Gossypium arboreum* L. Int J Mol Sci 21 (5). doi:10.3390/ijms21051675

Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L, Zhou Y, Wang LJ, Guo WZ, Bai YL, Ruan JX, Shangguan XX, Mao YB, Shan CM, Jiang JP, Zhu YQ, Jin L, Kang H, Chen ST, He XL, Wang R, Wang YZ, Chen J, Wang LJ, Yu ST, Wang BY, Wei J, Song SC, Lu XY, Gao ZC, Gu WY, Deng X, Ma D, Wang S, Liang WH, Fang L, Cai CP, Zhu XF, Zhou BL, Jeffrey Chen Z, Xu SH, Zhang YG, Wang SY, Zhang TZ, Zhao GP, Chen XY (2015b) *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Sci Rep 5:14139. doi:10.1038/srep14139

Liu Y (2013) Recent progress in fourier transform infrared (FTIR) spectroscopy study of compositional, structural and physical attributes of developmental cotton fibers. Materials 6 (1):299-313

Liu Y, Liu D, Khan AR, Liu B, Wu M, Huang L, Wu J, Song G, Ni H, Ying H, Yu H, Gan Y (2018) *NbGIS* regulates glandular trichome initiation through GA signaling in tobacco. Plant Mol Biol 98 (1-2):153-167. doi:10.1007/s11103-018-0772-3

Liu ZH, Chen Y, Wang NN, Chen YH, Wei N, Lu R, Li Y, Li XB (2020b) A basic helix–loop–helix protein (GhFP1) promotes fibre elongation of cotton (*Gossypium hirsutum*) by modulating brassinosteroid biosynthesis and signalling. New Phytologist 225 (6):2439-2452

Luo M, Xiao Y, Li X, Lu X, Deng W, Li D, Hou L, Hu M, Li Y, Pei Y (2007) GhDET2, a steroid 5α-reductase, plays an important role in cotton fiber cell initiation and elongation. The Plant Journal 51 (3):419-430

Ma QF, Wu CH, Wu M, Pei WF, Li XL, Wang WK, Zhang J, Yu JW, Yu SX (2016) Integrative transcriptome, proteome, phosphoproteome and genetic mapping reveals new aspects in a fiberless mutant of cotton. Sci Rep 6:24485. doi:10.1038/srep24485

Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G (2018a) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nature genetics 50 (6):803-813

Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G, Yan Y, Jia Y, Yang J, Pan Z, Gu Q, Li X, Sun Z, Dai P, Liu Z, Gong W, Wu J, Wang M, Liu H, Feng K, Ke H, Wang J, Lan H, Wang G, Peng J, Wang N, Wang L, Pang B, Peng Z, Li R, Tian S, Du X (2018b) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat Genet 50 (6):803-813. doi:10.1038/s41588-018-0119-7

Machado A, Wu Y, Yang Y, Llewellyn DJ, Dennis ES (2009) The MYB transcription factor *GhMYB25* regulates early fibre and trichome development. The Plant Journal 59 (1):52-62

Maes L, Inzé D, Goossens A (2008) Functional specialization of the *TRANSPARENT TESTA GLABRA1* network allows differential hormonal control of laminal and marginal trichome initiation in Arabidopsis rosette leaves. Plant physiology 148 (3):1453-1464

Mansoor S, Paterson AH (2012) Genomes for jeans: cotton genomics for engineering superior fiber. Trends in biotechnology 30 (10):521-527

Maqbool A, Abbas W, Rao AQ, Irfan M, Zahur M, Bakhsh A, Riazuddin S, Husnain T (2010) *Gossypium arboreum GHSP26* enhances drought tolerance in *Gossypium hirsutum*. Biotechnology progress 26 (1):21-25

Matias-Hernandez L, Aguilar-Jaramillo AE, Cigliano RA, Sanseverino W, Pelaz S (2016) Flowering and trichome development share hormonal and transcription factor regulation. J Exp Bot 67 (5):1209-1219. doi:10.1093/jxb/erv534

Mei G, Zhang Z (2019) Optimization of polar distribution of *GhPIN3a* in the ovule epidermis improves cotton fiber development. Journal of experimental botany 70 (12):3021

Meng C-M, Zhang T-Z, Guo W-Z (2009) Molecular cloning and characterization of a novel *Gossypium hirsutum* L. *bHLH* gene in response to ABA and drought stresses. Plant molecular biology reporter 27 (3):381-387

Nan Z, Li H, Zhang H, Li J, Li J, Wang J, Jiao G (2009) Effects of transforming antisense *GhADF1* gene on the fiber quality in upland cotton (*Gossypium hirsutum* L.). Scientia Agricultura Sinica 42 (12):4139-4144

Naoumkina M, Bechere E, Fang DD, Thyssen GN, Florane CB (2017a) Genome-wide analysis of gene expression of EMS-induced short fiber mutant *Ligon lintless-y* ($li_y$) in cotton (*Gossypium hirsutum* L.). Genomics 109 (3-4):320-329. doi:10.1016/j.ygeno.2017.05.007

Naoumkina M, Hinchliffe DJ, Fang DD, Florane CB, Thyssen GN (2017b) Role of xyloglucan in cotton (*Gossypium hirsutum* L.) fiber elongation of the short fiber mutant *Ligon lintless-2* ($Li_2$). Gene 626:227-233. doi:10.1016/j.gene.2017.05.042

Naoumkina M, Thyssen GN, Fang DD (2015) RNA-seq analysis of short fiber mutants *Ligon-lintless-1* ($Li_1$) and *–2* ($Li_2$) revealed important role of aquaporins in cotton (*Gossypium hirsutum* L.) fiber elongation. BMC Plant Biology 15 (1). doi:10.1186/s12870-015-0454-0

Naoumkina M, Thyssen GN, Fang DD, Bechere E, Li P, Florane CB (2021a) Mapping-by-sequencing the locus of EMS-induced mutation responsible for tufted-fuzzless seed phenotype in cotton. Mol Genet Genomics 296 (5):1041-1049. doi:10.1007/s00438-021-01802-0

Naoumkina M, Thyssen GN, Fang DD, Li P, Florane CB (2021b) Elucidation of sequence polymorphism in fuzzless-seed cotton lines. Mol Genet Genomics 296 (1):193-206. doi:10.1007/s00438-020-01736-z

Nayyar H, Kaur K, Basra A, Malik C (1989) Hormonal regulation of cotton fibre elongation in *Gossypium arboreum* L. in vitro and in vivo. Biochemie und Physiologie der Pflanzen 185 (5-6):415-421

Oosterhuis DM (1990) Growth and development of a cotton plant. Nitrogen nutrition of cotton: Practical issues:1-24

Padmalatha KV, Patil DP, Kumar K, Dhandapani G, Kanakachari M, Phanindra ML, Kumar S, Mohan T, Jain N, Prakash AH (2012) Functional genomics of fuzzless-lintless mutant of *Gossypium hirsutum* L. cv. MCU5 reveal key genes and pathways involved in cotton fibre initiation and elongation. BMC genomics 13 (1):1-15

Patel JD, Huang X, Lin L, Das S, Chandnani R, Khanal S, Adhikari J, Shehzad T, Guo H, Roy-Zokan EM, Rong J, Paterson AH (2020) The *Ligon lintless -2* short fiber mutation is located within a terminal deletion of chromosome 18 in cotton. Plant Physiol 183 (1):277-288. doi:10.1104/pp.19.01531

Pattanaik S, Patra B, Singh SK, Yuan L (2014) An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Front Plant Sci 5:259. doi:10.3389/fpls.2014.00259

Payne CT, Zhang F, Lloyd AM (2000) *GL3* encodes a bHLH protein that regulates trichome development in Arabidopsis through interaction with GL1 and TTG1. Genetics 156 (3):1349-1362

Pei Y (2015) The homeodomain-containing transcription factor, GhHOX3, is a key regulator of cotton fiber elongation. Science China Life Sciences 58 (3):309-310. doi:10.1007/s11427-015-4813-8

Perazza D, Vachon G, Herzog M (1998) Gibberellins promote trichome formation by up-regulating *GLABROUS1* in Arabidopsis. Plant physiology 117 (2):375-383

Percy R, Hendon B, Bechere E, Auld D (2015) Qualitative genetics and utilization of mutants. doi:10.2134/agronmonogr57.2013.0042

Pu L, Li Q, Fan X, Yang W, Xue Y (2008) The R2R3 MYB transcription factor *GhMYB109* is required for cotton fiber development. Genetics 180 (2):811-820

Qi T, Huang H, Wu D, Yan J, Qi Y, Song S, Xie D (2014) Arabidopsis DELLA and JAZ proteins bind the WD-repeat/bHLH/MYB complex to modulate gibberellin and jasmonate signaling synergy. The Plant Cell 26 (3):1118-1133

Qi T, Song S, Ren Q, Wu D, Huang H, Chen Y, Fan M, Peng W, Ren C, Xie D (2011) The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. The Plant Cell 23 (5):1795-1814

Qin Y-M, Pujol FM, Hu C-Y, Feng J-X, Kastaniotis AJ, Hiltunen JK, Zhu Y-X (2007) Genetic and biochemical studies in yeast reveal that the cotton fibre-specific *GhCER6* gene functions in fatty acid elongation. Journal of Experimental Botany 58 (3):473-481

Qin Y-M, Zhu Y-X (2011) How cotton fibers elongate: a tale of linear cell-growth mode. Current opinion in plant biology 14 (1):106-111

Rinehart JA, Petersen MW, John ME (1996) Tissue-specific and developmental regulation of cotton gene *FbL2A* (demonstration of promoter activity in transgenic plants). Plant Physiology 112 (3):1331-1341

Roberts MR (2016) Jasmonic Acid Signalling. In: eLS. John Wiley & Sons, Ltd. doi:10.1002/9780470015902.a0023721

Rong J, Pierce GJ, Waghmare VN, Rogers CJ, Desai A, Chee PW, May OL, Gannaway JR, Wendel JF, Wilkins TA, Paterson AH (2005) Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theor Appl Genet 111 (6):1137-1146. doi:10.1007/s00122-005-0041-0

Ruan Y-L (2007) Rapid cell expansion and cellulose synthesis regulated by plasmodesmata and sugar: insights from the single-celled cotton fibre. Functional Plant Biology 34 (1):1-10

Ryser U (1999) Cotton fiber initiation and histodifferentiation. Cotton fibers: developmental biology, quality improvement, and textile processing. Haworth Press, Binghamton, New York:21-29

Ryser U, Holloway P (1985) Ultrastructure and chemistry of soluble and polymeric lipids in cell walls from seed coats and fibres of *Gossypium* species. Planta 163 (2):151-163

Samuel Yang S, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, Stelly DM, Thaxton P, Triplett B, Town CD (2006) Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. The Plant Journal 47 (5):761-775

Seagull RW, Giavalis S (2004) Pre-and post-anthesis application of exogenous

hormones alters fiber production in *Gossypium hirsutum* L. cultivar Maxxa GTO. Journal of cotton science

Shan C-M, Shangguan X-X, Zhao B, Zhang X-F, Chao L-m, Yang C-Q, Wang L-J, Zhu H-Y, Zeng Y-D, Guo W-Z (2014a) Control of cotton fibre elongation by a homeodomain transcription factor *GhHOX3*. Nature communications 5:5519

Shangguan X, Yu N, Wang L, Chen X (2010) Recent advances in molecular biology research on cotton fiber development. Cotton:161-175

Shangguan XX, Yang CQ, Zhang XF, Wang LJ (2016) Functional characterization of a basic helix-loop-helix (bHLH) transcription factor *GhDEL65* from cotton (*Gossypium hirsutum*). Physiol Plant 158 (2):200-212. doi:10.1111/ppl.12450

Shen B, Sinkevicius KW, Selinger DA, Tarczynski MC (2006) The homeobox gene *GLABRA2* affects seed oil content in Arabidopsis. Plant molecular biology 60 (3):377-387

Shi Y-H, Zhu S-W, Mao X-Z, Feng J-X, Qin Y-M, Zhang L, Cheng J, Wei L-P, Wang Z-Y, Zhu Y-X (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. The plant cell 18 (3):651-664

Singh B, Avci U, Eichler Inwood SE, Grimson MJ, Landgraf J, Mohnen D, Sørensen I, Wilkerson CG, Willats WG, Haigler CH (2009) A specialized outer layer of the primary cell wall joins elongating cotton fibers into tissue-like bundles. Plant physiology 150 (2):684-699

Singh ND, Kumar S, Daniell H (2016) Expression of beta-glucosidase increases trichome density and artemisinin content in transgenic *Artemisia annua* plants. Plant Biotechnol J 14 (3):1034-1045. doi:10.1111/pbi.12476

Singh PK, Gao W, Liao P, Li Y, Xu FC, Ma XN, Long L, Song CP (2020) Comparative acetylome analysis of wild-type and fuzzless-lintless mutant ovules of upland cotton (*Gossypium hirsutum* Cv. Xu142) unveils differential protein acetylation may regulate fiber development. Plant Physiol Biochem 150:56-70. doi:10.1016/j.plaphy.2020.02.031

Song SK, Kwak SH, Chang SC, Schiefelbein J, Lee MM (2015) *WEREWOLF* and *ENHANCER of GLABRA3* are interdependent regulators of the spatial expression pattern of *GLABRA2* in Arabidopsis. Biochem Biophys Res Commun 467 (1):94-100. doi:10.1016/j.bbrc.2015.09.115

Stewart JM (1975) Fiber initiation on the cotton ovule (*Gossypium hirsutum*). American Journal of Botany 62 (7):723-730

Sun L, Zhang A, Zhou Z, Zhao Y, Yan A, Bao S, Yu H, Gan Y (2015) *GLABROUS INFLORESCENCE STEMS3* (*GIS3*) regulates trichome initiation and development in Arabidopsis. New Phytol 206 (1):220-230. doi:10.1111/nph.13218

Sun Q, Huang J, Guo Y, Yang M, Guo Y, Li J, Zhang J, Xu W (2020) A cotton NAC domain transcription factor, *GhFSN5*, negatively regulates secondary cell wall biosynthesis and anther development in transgenic Arabidopsis. Plant Physiol Biochem 146:303-314. doi:10.1016/j.plaphy.2019.11.030

Sun R, Li C, Zhang J, Li F, Ma L, Tan Y, Wang Q, Zhang B (2017) Differential

expression of microRNAs during fiber development between fuzzless-lintless mutant and its wild-type allotetraploid cotton. Scientific Reports 7 (1). doi:10.1038/s41598-017-00038-6

Sun W, Gao Z, Wang J, Huang Y, Chen Y, Li J, Lv M, Wang J, Luo M, Zuo K (2019a) Cotton fiber elongation requires the transcription factor *GhMYB212* to regulate sucrose transportation into expanding fibers. New Phytologist 222 (2):864-881

Sun Y, Fokar M, Asami T, Yoshida S, Allen RD (2004) Characterization of the *brassinosteroid insensitive1* genes of cotton. Plant molecular biology 54 (2):221-232

Sun Y, Liang W, Shen W, Feng H, Chen J, Si Z, Hu Y, Zhang T (2019b) G65V substitution in actin disturbs polymerization leading to inhibited cell elongation in cotton. Front Plant Sci 10:1486. doi:10.3389/fpls.2019.01486

Sun Y, Veerabomma S, Abdel-Mageed HA, Fokar M, Asami T, Yoshida S, Allen RD (2005) Brassinosteroid regulates fiber development on cultured cotton ovules. Plant and Cell Physiology 46 (8):1384-1391

Suo J, Liang X, Pu L, Zhang Y, Xue Y (2003) Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1630 (1):25-34

Tang W, Tu L, Yang X, Tan J, Deng F, Hao J, Guo K, Lindsey K, Zhang X (2014) The calcium sensor *GhCaM7* promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New phytologist 202 (2):509-520

Thyssen GN, Fang DD, Turley RB, Florane C, Li P, Naoumkina M (2014) Next generation genetic mapping of the *Ligon-lintless-2* ($Li_2$) locus in upland cotton (*Gossypium hirsutum* L.). Theoretical and applied genetics 127 (10):2183-2192

Thyssen GN, Fang DD, Turley RB, Florane CB, Li P, Mattison CP, Naoumkina M (2017) A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and *F*-actin organization resulting in dwarf, lintless cotton plants. The Plant Journal 90 (1):111-121

Tian H, Wang X, Guo H, Cheng Y, Hou C, Chen JG, Wang S (2017) *NTL8* regulates trichome formation in Arabidopsis by directly activating R3 MYB genes *TRY* and *TCL1*. Plant Physiol 174 (4):2363-2375. doi:10.1104/pp.17.00510

Tian Y, Zhang T (2021) MIXTAs and phytohormones orchestrate cotton fiber development. Current Opinion in Plant Biology 59:101975

Tissier A, Morgan JA, Dudareva N (2017) Plant volatiles: going 'In' but not 'Out' of trichome cavities. Trends Plant Sci 22 (11):930-938. doi:10.1016/j.tplants.2017.09.001

Tiwari SC, Wilkins TA (1995) Cotton (*Gossypium hirsutum*) seed trichomes expand via diffuse growing mechanism. Canadian Journal of Botany 73 (5):746-757

Traw MB, Bergelson J (2003) Interactive effects of jasmonic acid, salicylic acid, and gibberellin on induction of trichomes in Arabidopsis. Plant physiology 133 (3):1367-1375

Turley R (2002) Registration of MD17 fiberless upland cotton as a genetic stock.

(Registrations of genetic stocks). Crop science 42 (3):994-996

Turley R, Kloth R Inheritance model for fiberless upland cotton (*Gossypium hirsutum* L.) line SL1-7-1: variation on a theme. In: Proceedings of the Beltwide Cotton Conferences, 2003.

Vadde BVL, Challa KR, Nath U (2018) The *TCP4* transcription factor regulates trichome cell differentiation by directly activating *GLABROUS INFLORESCENCE STEMS* in *Arabidopsis thaliana*. Plant J 93 (2):259-269. doi:10.1111/tpj.13772

Vadde BVL, Challa KR, Sunkara P, Hegde AS, Nath U (2019) The TCP4 transcription factor directly activates *TRICHOMELESS1* and *2* and suppresses trichome initiation. Plant Physiol 181 (4):1587-1599. doi:10.1104/pp.19.00197

Varo, A., Moral, J., Lozano-Tóvar, M. D., and Trapero, A (2016a) Development and validation of an inoculation method to assess the efficacy of biological treatments against Verticillium wilt in olive trees. Biol. Control 61, 283–292. doi: 10.1007/s10526-015-9710-3

Varo, A., Raya-Ortega, M. C., and Trapero, A (2016b) Selection and evaluation of micro-organisms for biocontrol of Verticillium dahliae in olive. J. Appl. Microbiol. 121, 767–777. doi: 10.1111/jam.13199

Walford SA, Wu Y, Llewellyn DJ, Dennis ES (2011) *GhMYB25-like*: a key factor in early cotton fibre development. The Plant Journal 65 (5):785-797

Walford SA, Wu Y, Llewellyn DJ, Dennis ES (2012) Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, *GhHD-1*. Plant J 71 (3):464-478. doi:10.1111/j.1365-313X.2012.05003.x

Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, Fang L, Yang S, Hu Y, Ye W, Zhang H, Ma P, Chen J, Wang Q, Mei G, Cai C, Yang D, Wang J, Guo W, Zhang W, Chen X, Zhang T (2016) Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. New Phytol 210 (4):1298-1310. doi:10.1111/nph.13860

Wan Q, Zhang H, Ye W, Wu H, Zhang T (2014) Genome-wide transcriptome profiling revealed cotton fuzz fiber development having a similar molecular model as Arabidopsis trichome. PLoS One 9 (5):e97313. doi:10.1371/journal.pone.0097313

Wang G, Zhao GH, Jia YH, Du XM (2013a) Identification and characterization of cotton genes involved in fuzz-fiber development. Journal of Integrative Plant Biology 55 (7):619-630

Wang HY, Wang J, Gao P, Jiao GL, Zhao PM, Li Y, Wang GL, Xia GX (2009) Down-regulation of *GhADF1* gene expression affects cotton fibre properties. Plant biotechnology journal 7 (1):13-23

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet 44 (10):1098-1103. doi:10.1038/ng.2371

Wang L, Zhou CM, Mai YX, Li LZ, Gao J, Shang GD, Lian H, Han L, Zhang TQ, Tang HB, Ren H, Wang FX, Wu LY, Liu XL, Wang CS, Chen EW, Zhang XN, Liu

C, Wang JW (2019a) A spatiotemporally regulated transcriptional complex underlies heteroblastic development of leaf hairs in *Arabidopsis thaliana*. EMBO J 38 (8). doi:10.15252/embj.2018100063

Wang L, Zhu Y, Hu W, Zhang X, Cai C, Guo W (2015a) Comparative transcriptomics reveals jasmonic acid-associated metabolism related to cotton fiber initiation. PLoS One 10 (6):e0129854. doi:10.1371/journal.pone.0129854

Wang M-Y, Zhao P-M, Cheng H-Q, Han L-B, Wu X-M, Gao P, Wang H-Y, Yang C-L, Zhong N-Q, Zuo J-R (2013b) The cotton transcription factor *TCP14* functions in auxin-mediated epidermal cell differentiation and elongation. Plant physiology 162 (3):1669-1680

Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L, Zhou X, Nie X, Li Z, Guo K, Ma Y, Huang C, Jin S, Zhu L, Yang X, Min L, Yuan D, Zhang Q, Lindsey K, Zhang X (2017) Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. Nat Genet 49 (4):579-587. doi:10.1038/ng.3807

Wang M, Tu L, Yuan D, Zhu, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q, Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA, Zhang X (2019b) Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet 51 (2):224-229. doi:10.1038/s41588-018-0282-x

Wang N, Ma Q, Ma J, Pei W, Liu G, Cui Y, Wu M, Zang X, Zhang J, Yu S (2019c) A comparative genome-wide analysis of the R2R3-MYB gene family among four Gossypium species and their sequence variation and association with fiber quality traits in an interspecific *G. hirsutum*× *G. barbadense* population. Frontiers in genetics 10:741

Wang NN, Li Y, Chen YH, Lu R, Zhou L, Wang Y, Zheng Y, Li XB (2021) Phosphorylation of WRKY16 by MPK3-1 is essential for its transcriptional activity during fiber initiation and elongation in cotton (*Gossypium hirsutum*). Plant Cell 33 (8):2736-2752. doi:10.1093/plcell/koab153

Wang S, Chen J, Zhang W, Hu Y, Chang L, Fang L, Wang Q, Lv F, Wu H, Si Z, Chen S, Cai C, Zhu X, Zhou B, Guo W, Zhang T (2015b) Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. Genome Biology 16 (1). doi:10.1186/s13059-015-0678-1

Wang X, Miao Y, Cai Y, Sun G, Jia Y, Song S, Pan Z, Zhang Y, Wang L, Fu G, Gao Q, Ji G, Wang P, Chen B, Peng Z, Zhang X, Wang X, Ding Y, Hu D, Geng X, Wang L, Pang B, Gong W, He S, Du X (2020) Large-fragment insertion activates gene *GaFZ* (*Ga08G0121*) and is associated with the fuzz and trichome reduction in cotton (*Gossypium arboreum*). Plant Biotechnol J. doi:10.1111/pbi.13532

Wang Z, Yang Z, Li F (2019d) Updates on molecular mechanisms in the development of branched trichome in Arabidopsis and nonbranched in cotton. Plant Biotechnol J 17 (9):1706-1722. doi:10.1111/pbi.13167

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. American Journal of Botany 79 (11):1291-

1310

Wendel JF, Brubaker CL, Seelanan T (2010) The origin and evolution of *Gossypium*. In: Physiology of cotton. Springer, pp 1-18

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Advances in agronomy 78:139

Wendel JF, Grover CE (2015) Taxonomy and evolution of the cotton genus, *Gossypium*. Cotton 57:25-44

Wu H, Ren Z, Zheng L, Guo M, Yang J, Hou L, Qanmber G, Li F, Yang Z (2021) The bHLH transcription factor *GhPAS1* mediates BR signaling to regulate plant development and architecture in cotton. The Crop Journal

Wu H, Tian Y, Wan Q, Fang L, Guan X, Chen J, Hu Y, Ye W, Zhang H, Guo W, Chen X, Zhang T (2017) Genetics and evolution of *MIXTA* genes regulating cotton lint fiber development. New Phytologist. doi:10.1111/nph.14844

Xiao G, He P, Zhao P, Liu H, Zhang L, Pang C, Yu J (2018) Genome-wide identification of the *GhARF* gene family reveals that *GhARF2* and *GhARF18* are involved in cotton fibre cell initiation. J Exp Bot 69 (18):4323-4337. doi:10.1093/jxb/ery219

Xiao G, Zhao P, Zhang Y (2019) A pivotal role of hormones in regulating cotton fiber development. Frontiers in plant science 10:87

Xiao Y-H, Li D-M, Yin M-H, Li X-B, Zhang M, Wang Y-J, Dong J, Zhao J, Luo M, Luo X-Y (2010) Gibberellin 20-oxidase promotes initiation and elongation of cotton fibers by regulating gibberellin synthesis. Journal of plant physiology 167 (10):829-837

Xie D-X, Feys BF, James S, Nieto-Rostro M, Turner JG (1998) *COI1*: an Arabidopsis gene required for jasmonate-regulated defense and fertility. Science 280 (5366):1091-1094

Xu J, Chen L, Sun H, Wusiman N, Sun W, Li B, Gao Y, Kong J, Zhang D, Zhang X (2019) Crosstalk between cytokinin and ethylene signaling pathways regulates leaf abscission in cotton in response to chemical defoliants. Journal of experimental botany 70 (5):1525-1538

Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, Yang L, Poethig RS (2016) Developmental functions of miR156-regulated *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes in *Arabidopsis thaliana*. PLoS Genet 12 (8):e1006263. doi:10.1371/journal.pgen.1006263

Xu S-M, Brill E, Llewellyn DJ, Furbank RT, Ruan Y-L (2012) Overexpression of a potato sucrose synthase gene in cotton accelerates leaf expansion, reduces seed abortion, and enhances fiber production. Molecular Plant 5 (2):430-441

Yan T, Chen M, Shen Q, Li L, Fu X, Pan Q, Tang Y, Shi P, Lv Z, Jiang W, Ma Y-n, Hao X, Sun X, Tang K (2017) *HOMEODOMAIN PROTEIN 1* is required for jasmonate-mediated glandular trichome initiation in *Artemisia annua*. New Phytologist 213 (3):1145-1155. doi:10.1111/nph.14205

Yang C, Lu X, Ma B, Chen S-Y, Zhang J-S (2015) Ethylene signaling in rice and Arabidopsis: conserved and diverged aspects. Molecular plant 8 (4):495-505

Yang C, Ye Z (2013) Trichomes as models for studying plant cell differentiation. Cell Mol Life Sci 70 (11):1937-1948. doi:10.1007/s00018-012-1147-6

Yang J, Hu W, Zhao W, Chen B, Wang Y, Zhou Z, Meng Y (2016) Fruiting branch $K^+$ level affects cotton fiber elongation through osmoregulation. Frontiers in plant science 7:13

Yang Z, Zhang C, Yang X, Liu K, Wu Z, Zhang X, Zheng W, Xun Q, Liu C, Lu L (2014) *PAG1*, a cotton brassinosteroid catabolism gene, modulates fiber elongation. New phytologist 203 (2):437-448

Yu N, Cai WJ, Wang S, Shan CM, Wang LJ, Chen XY (2010) Temporal control of trichome distribution by microRNA156-targeted *SPL* genes in *Arabidopsis thaliana*. Plant Cell 22 (7):2322-2335. doi:10.1105/tpc.109.072579

Yu Y, Wu S, Nowak J, Wang G, Han L, Feng Z, Mendrinna A, Ma Y, Wang H, Zhang X (2019) Live-cell imaging of the cytoskeleton in elongating cotton fibres. Nature plants 5 (5):498-504

Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L, Li Y, Liang Q, Lin Z, Yang X, Liu N, Jin S, Lei Y, Ding Y, Li G, Ruan X, Ruan Y, Zhang X (2015) The genome sequence of Sea-island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. Scientific Reports 5 (1). doi:10.1038/srep17662

Zaidi SS, Mansoor S, Paterson A (2018) The rise of cotton genomics. Trends Plant Sci 23 (11):953-955. doi:10.1016/j.tplants.2018.08.009

Zeng J, Zhang M, Hou L, Bai W, Yan X, Hou N, Wang H, Huang J, Zhao J, Pei Y (2019) Cytokinin inhibits cotton fiber initiation by disrupting *PIN3a*-mediated asymmetric accumulation of auxin in the ovule epidermis. J Exp Bot 70 (12):3139-3151. doi:10.1093/jxb/erz162

Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis.

Zhang F, Zuo K, Zhang J, Liu X, Zhang L, Sun X, Tang K (2010) An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. Journal of experimental botany 61 (13):3599-3613

Zhang J, Huang GQ, Zou D, Yan JQ, Li Y, Hu S, Li XB (2018) The cotton (*Gossypium hirsutum*) NAC transcription factor (*FSN1*) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. New Phytologist 217 (2):625-640

Zhang M, Han LB, Wang WY, Wu SJ, Jiao GL, Su L, Xia GX, Wang HY (2017a) Overexpression of *GhFIM2* propels cotton fiber development by enhancing actin bundle formation. Journal of integrative plant biology 59 (8):531-534

Zhang M, Zeng JY, Long H, Xiao YH, Yan XY, Pei Y (2017b) Auxin regulates cotton fiber initiation via *GhPIN*-mediated auxin transport. Plant Cell Physiol 58 (2):385-397. doi:10.1093/pcp/pcw203

Zhang M, Zheng X, Song S, Zeng Q, Hou L, Li D, Zhao J, Wei Y, Li X, Luo M (2011) Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. Nature biotechnology 29 (5):453-458

Zhang S, Feng M, Chen W, Zhou X, Lu J, Wang Y, Li Y, Jiang CZ, Gan SS, Ma N, Gao J (2019) In rose, transcription factor PTM balances growth and drought survival via PIP2;1 aquaporin. Nat Plants 5 (3):290-299. doi:10.1038/s41477-019-0376-1

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, Hulse-Kemp AM, Wan Q, Liu B, Liu C, Wang S, Pan M, Wang Y, Wang D, Ye W, Chang L, Zhang W, Song Q, Kirkbride RC, Chen X, Dennis E, Llewellyn DJ, Peterson DG, Thaxton P, Jones DC, Wang Q, Xu X, Zhang H, Wu H, Zhou L, Mei G, Chen S, Tian Y, Xiang D, Li X, Ding J, Zuo Q, Tao L, Liu Y, Li J, Lin Y, Hui Y, Cao Z, Cai C, Zhu X, Jiang Z, Zhou B, Guo W, Li R, Chen ZJ (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nature Biotechnology 33 (5):531-537. doi:10.1038/nbt.3207

Zhang Z, Ruan Y-L, Zhou N, Wang F, Guan X, Fang L, Shang X, Guo W, Zhu S, Zhang T (2017c) Suppressing a putative sterol carrier gene reduces plasmodesmal permeability and activates sucrose transporter genes during cotton fiber elongation. The Plant Cell 29 (8):2027-2046

Zhao B, Cao JF, Hu GJ, Chen ZW, Wang LY, Shangguan XX, Wang LJ, Mao YB, Zhang TZ, Wendel JF (2018) Core *cis*-element variation confers subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. New Phytologist 218 (3):1061-1075

Zhou Y, Zhang ZT, Li M, Wei XZ, Li XJ, Li BY, Li XB (2015) Cotton (*Gossypium hirsutum*) 14-3-3 proteins participate in regulation of fibre initiation and elongation by modulating brassinosteroid signalling. Plant biotechnology journal 13 (2):269-280

Zhou Z, An L, Sun L, Zhu S, Xi W, Broun P, Yu H, Gan Y (2011) Zinc finger protein5 is required for the control of trichome initiation by acting upstream of zinc finger protein8 in Arabidopsis. Plant Physiol 157 (2):673-682. doi:10.1104/pp.111.180281

Zhou Z, Sun L, Zhao Y, An L, Yan A, Meng X, Gan Y (2013) Z inc Finger Protein 6 (ZFP 6) regulates trichome initiation by integrating gibberellin and cytokinin signaling in *Arabidopsis thaliana*. New Phytologist 198 (3):699-708

Zhu Q-H, Yuan Y, Stiller W, Jia Y, Wang P, Pan Z, Du X, Llewellyn D, Wilson I (2018) Genetic dissection of the fuzzless seed trait in *Gossypium barbadense*. Journal of Experimental Botany 69 (5):997-1009. doi:10.1093/jxb/erx459

Zhu QH, Stiller W, Moncuquet P, Gordon S, Yuan Y, Barnes S, Wilson I (2021) Genetic mapping and transcriptomic characterization of a new fuzzless-tufted cottonseed mutant. G3 (Bethesda) 11 (1):1-14. doi:10.1093/g3journal/jkaa042

# 2

# Objectives and thesis structure

Cotton fibers are single-celled tubular outgrowths that originate and arise from the epidermis of seed coats. According to the principle differences of initiating time and final length, hairs are divided into two different types: long lint and short fuzz. Long lint fibers can be easily removed by ginning process, and thus, they are of great economic importance to be extensively used in the production of textiles and fabrics. In contrast, fuzz fibers form a dense and tangled mat adhered to the seed surface and cannot be removed by ginning. Besides, fuzz fibers impede seeds from absorbing water and germinating, disseminating pathogens. Hence, to some extent, fuzz is a kind of desirable agronomic trait in cotton breeding. And thus, fuzzless mutants are an essential and crucial material to study the regulatory mechanism of fiber and fuzz differentiation and even provide important gene resources to create novel commercial fuzzless cultivars and genetic germplasms. With the rapid development of high-throughput sequencing technology, multiple versions of the cotton genome data have been released and have provided momentous and fundamental references for the bioinformatics-aided analysis. Currently, the combination of segregation populations and the HTS-based mapping and cloning has been the most common method in identifying and validating functional genes.

# 1 Objectives of the Thesis

The main objective of this dissertation was to obtain the precise candidate gene contributing to fuzzless phenotype in diploid fiber mutant and to decipher the molecular mechanisms and regulatory networks underlying fuzz initiation and development. We hope this research may provide valuable gene resources and novel insights into mechanisms of fuzz-associated processes to create and generate new fuzzless germplasms.

# 2 Structures of the Ph.D. thesis

The thesis mainly contains the scientific papers that have been or are being published. The structures of this PhD thesis are divided into the following six parts:

Chapter 1 is a general introduction to cotton fibers and a brief review focused on the research progress of key regulatory factors and relevant fiber mutants.

Chapter 2 is an introduction to the objective and outline of this thesis.

Chapter 3 demonstrates the fine mapping of fuzzless gene *GaFzl* in diploid cotton DPL972 through the combination strategy of high throughput sequencing and traditional map-based cloning. This part of the work has been published in "Theoretical and Applied Genetics".

Chapter 4 introduces the whole genome-wide identification of the *GIR* gene family in cotton and the expression patterns of family members during fiber and fuzz initiation stages. This part of the work has been published in "Planta".

Chapter 5 reveals the co-expressed module with fuzz development and the potential co-expression genes and networks involved in fuzz initiation and formation. This part of work has been published in "Genes".

Chapter 6 summarizes the main conclusion, general discussion, and perspectives of the thesis.

# 3

# Fine mapping and identification of the fuzzless gene *GaFzl* in DPL972 (*Gossypium arboreum*)

*The fuzzless gene GaFzl was fine mapped to a 70-kb region containing a GIR1 gene, Cotton_A_11941, responsible for the fuzzless trait in Gossypium arboreum DPL972.*

From Feng X, Cheng H, Zuo D, *et al.* Fine mapping and identification of the fuzzless gene *GaFzl* in DPL972 (*Gossypium arboreum*). *Theor Appl Genet*, 2019, 132: 2169-2179.

**Abstract**: Cotton fiber is the most important natural textile resource. The fuzzless mutant DPL972 (*Gossypium arboreum*) provides a useful germplasm resource to explore the molecular mechanism underlying fiber and fuzz initiation and development. In our previous research, the fuzzless gene in DPL972 was identified as a single dominant gene and named *GaFzl*. In present study, we fine mapped this gene using $F_2$ and $BC_1$ populations. By combining traditional map-based cloning and next generation sequencing, we mapped *GaFzl* to a 70-kb region containing seven annotated genes. RNA-Sequencing (RNA-Seq) and re-sequencing analysis narrowed these candidates to two differentially expressed genes (DEGs), Cotton_A_11941 and Cotton_A_11942. Sequence alignment uncovered no variation in coding or promoter regions of Cotton_A_11942 between DPL971 and DPL972, whereas, two single-base-mutations in the promoter region and a TTG insertion in the coding region were detected in Cotton_A_11941 in DPL972. Cotton_A_11941 encoding a homologous gene of *GIR1* (GLABRA2-interacting repressor) in *Arabidopsis thaliana* is thus the candidate gene most likely responsible for the fuzzless trait in DPL972. Our findings should lead to a better understanding of cotton fuzz formation, thereby accelerating marker-assisted selection during cotton breeding.

**Key words**: Gossypium arboreum, fuzzless, map-based cloning, bulked segregant analysis, RNA-Seq

# 1.Introduction

In the history of development economics, cotton has been thought as a great commercial crop due to its edible oil and especially fiber which regarded as an important natural textile resource. Similar to trichomes in Arabidopsis, cotton fibers are single cells derived from the ovule epidermis (Ishida et al. 2008, Wan et al. 2014). Cotton fiber is categorized into two types according to the initiation stage: lint, developing before or at 0 days post-anthesis (DPA), and fuzz, developing approximately 3 to 5 DPA (Lee et al. 2007). Compared with lint, fuzz impedes seed germination and therefore leads to cause a low fiber production in the next growing season. Fuzzless breeds are thus increasingly being considered as important resources in cotton genetics and breeding.

Fuzzless mutants have been a major focus of cotton research since 1927 (Jiang et al. 2015, Liang et al. 2015). Two fuzzless loci in *G. hirsutum*, $N_1$ and $n_2$, are currently known to inhibit fuzz development. $N_1$ is a single dominant gene and controls a fuzzless-lint trait (Rong et al. 2005, Wan et al. 2014, Wan et al. 2016). According to map-based cloning, $N_1$ is located on chromosomes 12 (A12) and has been annotated as a MIXTA-like gene, *MYB25-like* (also known as *GhMML3*). In the mutant $N_1$, *GhMML3* also shows reverse transcriptional activity and generates natural antisense transcripts (NATs) (Wan et al. 2016). The bidirectional transcripts of *GhMML3_A12* form double-stranded RNAs and small RNAs. Small RNAs may mediate self-degradation and self-cleavage of mRNA and then result in a naked seed phenotype. In addition, data from virus-induced gene silencing (VIGS) suggest that disruption of *GhMML3* in TM-1 leads to a fuzzless or reduced-fuzz phenotype. The other fuzzless gene, $n_2$, is a single recessive locus, located on chromosome26 (D12) (Zhu et al. 2018). But the candidate gene for $n_2$ has not yet been identified. Recent research has shown that $n_2$ and $Li_3$ are respectively responsible for the production of fuzz fiber and lint fiber in the fuzzless-lintless mutant XuZhou142 (Wu et al. 2018). Using a map-based cloning strategy, the authors of cited study identified $Li_3$ as another MML transcription factor (*GhMML4_D12*) regulating lint development. Nevertheless, the $n_2$ gene has remained elusive. The inheritance of fuzz and lint fiber initiation and development is complex (Naoumkina et al. 2016, Sun et al. 2017). These studies on fuzzless mutants and fiber-related genes have contributed to our understanding of the mechanism of fiber/fuzz initiation and regulation.

In contrast to research on $N_1$ and $n_2$, few systematic studies have been carried out on the diploid *G. arboreum* fuzzless mutant DPL972. A comparison of molecular mechanisms between diploids and tetraploids is thus needed (Parekh et al. 2016). DPL972 is a near-isogenic line (NIL) of the wild-type DPL971. They carry the same genetic background with the exception of traits for fuzz and flower color. In a whole-genome expression analysis, *MYB23*, *MYB5* and *TTG1* were significantly differentially expressed between DPL971 and DPL972. Wang et al. (2013) subsequently identified *MYB23* as a homologous gene of GLABRA1 involved in the regulation of Arabidopsis trichome development and cotton fiber initiation. Whether MYB23 controls the fuzzless trait in the mutant DPL972 is still unknown. In a previous study, we developed SSR markers linked to *GaFzl* and mapped this gene to chromosome A08 (Feng et al. 2016), but were unable to identify the exact gene that

controls fuzz development.

In the present study, we explored and identified *GaFzl* to better understand fiber initiation and fuzz development. To fine map the target gene, we integrated bulked segregant analysis sequencing (BSA-Seq), map-based cloning and RNA-Seq strategies. We found a *GIR1* gene significantly differentially expressed between the mutant and wild type in the candidate region. Our finding should contribute to a deeper understanding of fuzz development.

# 2. Materials and Methods

## *2.1 Materials and population construction*

In this study, we used *G. arboreum* DPL971 and its fuzzless mutant DPL972, which had been self-pollinated for more than 10 generations in our laboratory. Plants were grown annually in accordance with standard agronomic practices at the farm of the Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, China.

To fine map the target gene, we consecutively constructed $F_2$ and $BC_1$ populations. The $F_2$ population comprised 4,010 individuals derived from a cross between DPL971 and DPL972 in 2014. $F_1$ plants were then back-crossed with DPL971 in 2017 to produce a $BC_1$ population comprising 607 individuals. These two populations were used to fine-map the fuzzless gene in DPL972. The $F_2$ population was also used as a source of individuals with extreme traits for construction of sequencing pools. The phenotypic segregation ratios of the two populations were analyzed by a $\chi^2$ goodness-of-fit test in SAS.

## *2.2 Genomic DNA and RNA extraction*

Young leaves from the parents and F2 and BC1 individuals were collected and stored at -80 °C. Young leaf tissues were frozen in liquid nitrogen and ground into a fine powder using a hybrid grinding machine (Retsch MM400, Vereder Scientific, Germany). Genomic DNA was then extracted using the cetyl trimethylammonium bromide method and stored at -20 °C.

Fiber-bearing ovules were excised from developing cotton bolls at 2-day collecting intervals (i.e. at 1, 3 and 5 days post-anthesis (DPA)). The ovule samples were wrapped in tinfoil, frozen directly in liquid nitrogen and stored at -80 °C for subsequent experiments. Total RNA was isolated from the collected ovules using an EASY-spin Plus Plant RNA kit (Aidlab, China) and an RNA Prep Pure Plant kit (Polysaccharides & Polyphenolics-rich) (Tiangen, China) according to the kit manuals. Total RNA integrity, purity and concentration were examined using an Agilent 2100 RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA, USA). Suitable RNA samples (A260/A280 ratio = 1.8-2.0, A260/A230 ratio > 1.5, and RNA integrity number > 8) were stored at -80 °C for subsequent deep sequencing and quantitative real-time PCR (qRT-PCR) analysis.

## *2.3 BSA-Seq and re-sequencing analyses*

Genomic DNA of 30 extremely fuzzy and 30 extremely fuzzless individuals from the F2 progenies were selected to generate two bulked pools. The DNAs of parents DPL971 and DPL972 were also extracted for BSA-Seq and re-sequencing on an Illumina NovaSeq 6000 platform. Raw reads from the four DNA pools were filtered and then aligned to the cotton genome sequence using the Burrows–Wheeler alignment tool (BWA) (Li and Durbin 2009). GATK and Picard were used to detect InDels and single-nucleotide polymorphisms (SNPs) (Mckenna et al. 2010). Euclidean distance (ED) and ΔSNP-index values were calculated to identify candidate genomic regions associated with the fuzzy trait (Fekih et al. 2013). In regards to the ED strategy, ED values of regions other than target gene-related ones tend to be consistent and trend towards 0. ΔSNP-index values were determined as the difference in the SNP-index between fuzzy and fuzzless pools with values of genomic regions including the target gene expected to approach 1. By examining ED and ΔSNP-index values between the two bulked pools, the plot peak regions above the threshold value were defined as candidate regions for association with the fuzzless trait.

## *2.4 Fine mapping of the fuzzless gene*

In our previous study, we developed genome-wide SSR markers, and identified 13 markers linked to the fuzzless gene. In the present study, we used these same markers to further fine map the target gene. To perform the subsequent mapping, additional SSR and InDel marker primers were designed with software SnapGene and Primer5 software based on the results of re-sequencing. Those markers with suitable levels of polymorphism and high specificity between the two mapping parents and bulked pools were used to construct the linkage map and narrow the candidate region.

## *2.5 RNA-Seq analysis*

To explore the molecular mechanism underlying fuzz initiation and development in *G. arboreum*, fiber bearing ovules were selected at three developmental stages (1, 3 and 5 DPA). Two biological replicates of three stages of ovules of the DPL972 fuzzless mutant and its wild-type DPL971 were used to create 12 independent RNA libraries. These libraries were subjected to 101-cycle paired-end sequencing on an Illumina HiSeq4000 platform at Berry Genomics (Beijing, China).

Raw data from the 12 sample libraries were assessed and filtered to remove adaptors and low-quality reads. The remaining high-quality reads, referred to as clean data, were aligned to the reference genomes of *G. arboreum* (Li et al. 2014), *G. raimondii* (Wang et al. 2012) and *G. hirsutum* (Li et al. 2015, Zhang et al. 2015) using TopHat2 (Kim et al. 2013). The mapped sequences were then annotated against the above-mentioned reference genomes, with a maximum of two bases allowed per mismatch in our alignments. In addition, the number of fragments per kilobase of transcript per million mapped reads (FPKM) of each transcript in the fuzzless mutant DPL972 and wild-type DPL971 was calculated using Cuffdiff (v2.2.1) software (Trapnell et al. 2016). Based on these FPKM values, a false discovery rate (FDR) $< 0.01$ and $|\log2(\text{ratio})| \geq 1$ were set as criteria to identify differentially expressed transcripts between the mutant and the wild type. P-values of the statistical t-test were adjusted and applied

tocontrol FDR using multiple testing procedures. HemI 1.0 was used for clustering analysis of DEGs (Deng et al. 2014).

## *2.6 Gene cloning and multiple sequence alignment*

Gene- and promoter-amplification primers were designed using Primer5 and SnapGene software. The full-length sequences of genes and promoters were respectively amplified using cDNA and genomic DNA from DPL971 and DPL972. PCR products were purified using a QIAquick Gel Extraction kit and sequenced by Sangon Biotech (Shanghai, China). The resulting sequences were aligned with DNAMAN software.

## *2.7 qRT-PCR*

To verify the accuracy of RNA-Seq and identify expression levels of candidate genes, qRT-PCR was also performed. Fiber-bearing ovules were selected at two developmental stages (1 and 3 DPA). RNA samples (1-2 μg) were subjected to reverse transcription using TransScript All-in-one First-strand cDNA Synthesis SuperMix (TransGen Biotech). QRT-PCR amplifications were conducted using TransStart TOP Green qPCR SuperMix (TransGen Biotech) on an ABI Prism7500 Fast Real-time PCR System (Applied Biosystems, USA). DEG expressions were normalized using *GaHis3* as a reference gene. Primers for amplifying DEGs were designed with NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and BatchPrimer3 (http://wheat.pw.usda.gov/demos/BatchPrimer3/) and synthesized by Genewiz (Beijing, China). Primer sequences are provided in Table S1. Each qRT-PCR reaction included three biological replicates and three technical replicates. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001).

# 3. Results

## *3.1 Phenotypes and segregation analysis of genetic populations*

In the current study we used two mapping parents: wild type DPL971 and the fuzzless mutant DPL972 (Fig. 3-1). We successively constructed $F_2$ and $BC_1$ populations in 2014 and 2016. Phenotypes of $F_1$ individuals derived from the cross between DPL971 and DPL972 were always consistent with those of DPL972, which indicates that the fuzzless trait in *G. arboreum* was dominantly inherited. As can be seen from Table 3-1, the observed segregations in $F_2$ and $BC_1$ generations fit the expected phenotypic segregation ratios of 3:1 and 1:1, respectively. Consistent with our previously published data (Feng et al. 2016), the phenotypes of the $BC_1$ population further validated our finding that a single dominant gene controls the fuzzless phenotype in DPL972. The new $BC_1$ population and the previous $F_2$ one were both used for subsequent mapping of *GaFzl*.

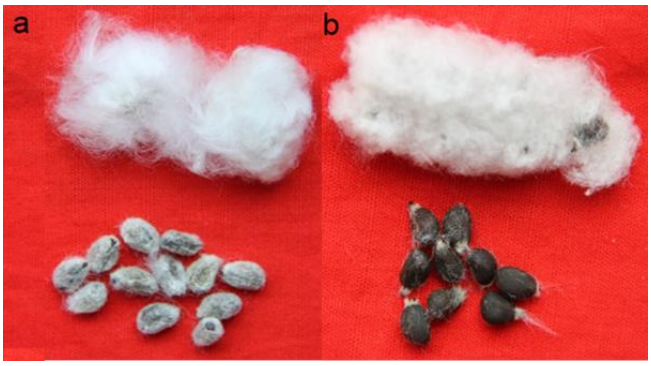



**Figure 3-1**: Phenotypes of fuzzy seeds DPL971 (**a**) and the fuzzless mutant DPL972 (**b**)

**Table 3-1**: Genetic analysis of fibre trait in parents and two segregating populations

| Material | Total | Phenotype | | Observed ratio | Expected ratio | $\chi^2$ | *P* |
|---|---|---|---|---|---|---|---|
| | | Fuzzless | Fuzzy | | | | |
| DPL971 | 20 | 0 | 20 | | | | |
| DPL972 | 20 | 20 | 0 | | | | |
| $F_1$ | 300 | 300 | 0 | | | | |
| $F_2$ Population | 4,010 | 3,004 | 1,006 | 2.99:1 | 3:1 | 0.02 | 0.9 |
| $BC_1$ Population | 607 | 315 | 292 | 1.08:1 | 1:1 | 0.87 | 0.35 |

## *3.2 Mapping of GaFzl gene to chromosome A08 by BSA-Seq*

To rapidly map and identify the genomic region contributing to the fuzzless trait in *G. arboreum* DPL972, we used BSA-Seq to fine map the mutation. A total of 228.9 Gb clean data were obtained by re-sequencing with an average depths of ~33× (details in Table 3-S1). These re-sequencing reads were then aligned to the preferred TM-1 reference genome using BWA software (Table 3-S1). Details of all SNPs and InDels between DPL971 and DPL972 as well as the two bulked pools were listed in Tables 3-S2 and 3-S3. A total of 353,611 high-quality SNPs and 51,244 small InDels were filtered out and identified between the two bulked pools. Applying the ED method to InDels and SNPs, we chose 0.35 (median+3SD of all fitted values of points) as the threshold value and identified a 2.23Mb region from 76,294 to 2,305,623 bp on chromosome A08 (Figs. 3-2a, 3-3a). Using the ΔSNP-index strategy, we also identified a 2.22Mb region from 68,670 to 2,195,304 bp on A08 (Figs. 3-2b, 3-3b). These regions overlapped. The merged region from 76,294 to 2,195,304 bp is thus likely the candidate region of the GaFzl locus controlling the fuzzless trait in *G. arboreum*.

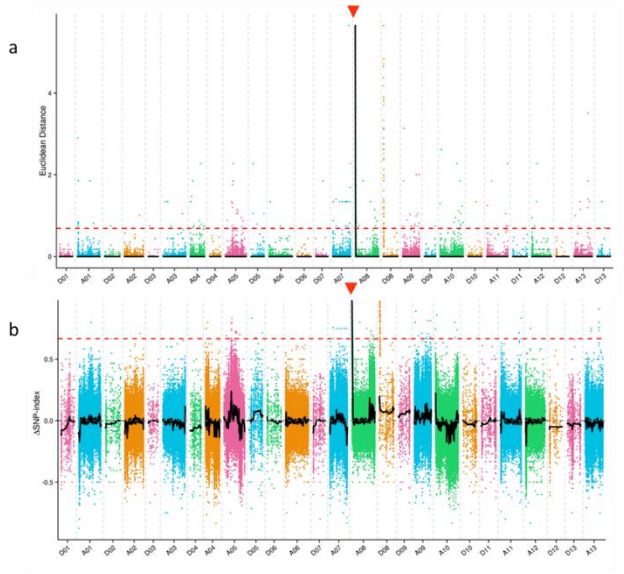



**Figure 3-2**: Candidate region of *GaFzl* based on application of a combined ED/ΔSNP-index strategy to InDels. **a.** ED graph of InDels between DPL971, DPL972 and two bulked pools. **b.** ΔSNP-index graph of InDels between DPL971, DPL972 and two bulked pools.

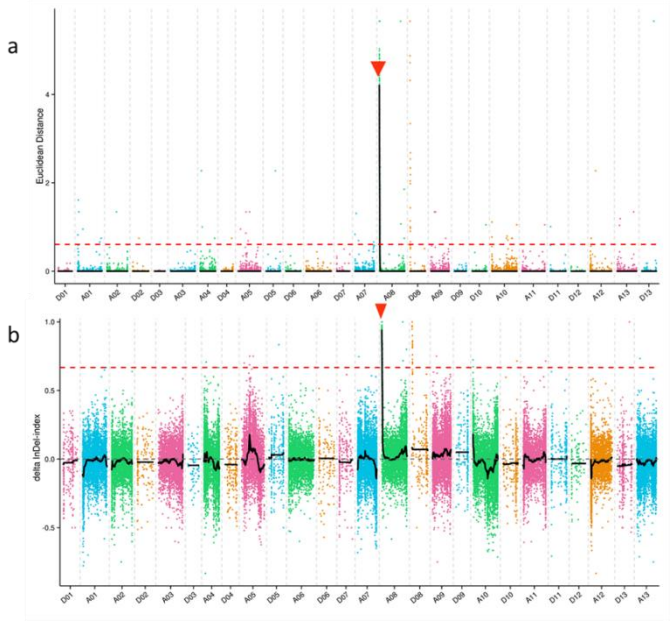



**Figure 3-3**: Candidate region of *GaFzl* based on application of a combined ED/ΔSNP-index strategy to SNPs. **a.** ED graph of SNPs between DPL971, DPL972 and two bulked pools. **b.** ΔSNP-index graph of SNPs between DPL971, DPL972 and two bulked pools.

## *3.3 Narrowing of the GaFzl gene to a 70-kb region using SSR and InDel markers*

In our previous study, we identified 62 polymorphic SSR markers between DPL971 and DPL972 by screening the whole genome. Only one marker on chromosome A08 showed a tight linkage with the fuzzless trait. Because the exact region was undetermined, we developed 652 SSR markers on A08, of which 13 polymorphic SSR markers showed a linkage relationship with the fuzzless trait (Table 3-2). The location of these SSR markers overlapped with the region obtained by BSA-Seq.

To rapidly shorten the physical mapping interval and validate the target gene, we developed 11 InDels and 98 SSR markers (detailed in Tables 3-2 and 3-S4) based on the results of the re-sequencing of DPL971, DPL972 and the bulked pools. GaFzl was narrowed down to a 70-kb region between SSR82 and InDel9 on chromosome A08. According to cotton genome annotation information, only seven ORFs exist in the corresponding genomic region (Fig. 3-4). Details on ORF position are given in Table 3-3.

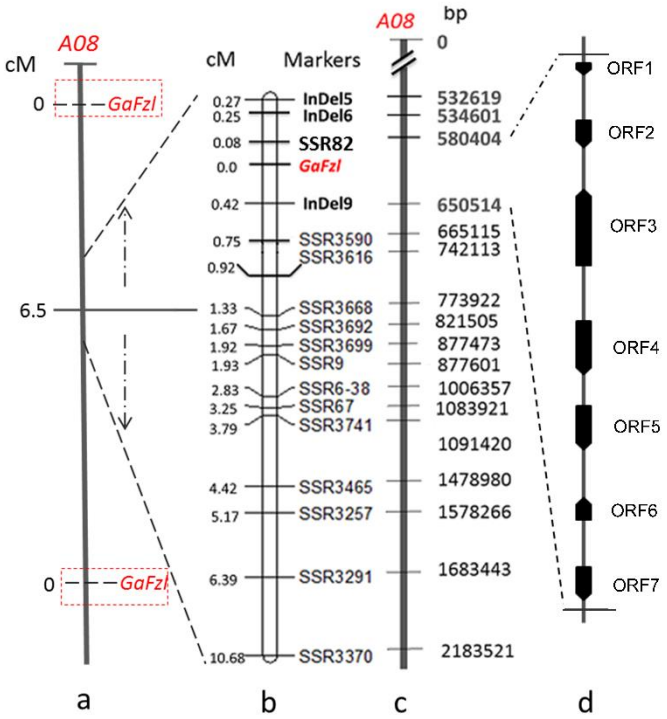



**Figure 3-4**: Map position of the fuzzless gene GaFzl in Gossypium arboreum DPL972 on chromosome A08. a. Primary mapping for the fuzzless trait. Map distances (cM) are on the left. b. Linkage map based on SSR and InDel markers. Map distances (cM) are on the left with marker names on the right. c. A physical map of the candidate region for the *GaFzl* gene. Numbers on the right indicate loci of the markers. d. ORFs in the candidate region.

**Table 3-2**: Details of linked markers on A08 used for fine mapping

| Primer Name | ForwardPrimer (5'→3') | ReversePrimer (5'→3') | Product Length (bp) | Start position (bp) |
|---|---|---|---|---|
| InDel5 | CGAATCCTGAACCCCAAACCTAAAC | CATCACCATGGCAACAACTCC | 376 | 532619 |
| InDel6 | CGCGAGGACTAAAATTTGAAAGTTTGGA | GGTAAGGATTGGGGCATTAACTGG | 178 | 534601 |
| SSR82 | CCTTCCATGCATATTGGAAA | CAAAGCACCCAATTTCAAGG | 283 | 580404 |
| InDel9 | GATTTCCCTTTCATCAAATATTCTATGTTAGCG | ACTCAAATACTTAGTATACCATATTAGCCTCTTTC | 536 | 650514 |
| SSR3590 | AAACCCAATATATTCTGAGTTAAATGT | TCACTATAACTAGCGGGTGGAAA | 184 | 665115 |
| SSR3616 | GGGATACCTGCAAACATTGTG | TTCATGGCCTTCCTCTCTGT | 154 | 742113 |
| SSR3668 | TGGATCGGTAATGGTAGAAAGC | AGCAAGGTCTTAGATGGCAA | 240 | 773922 |
| SSR3692 | GCTGTGAGGACATGAAACGA | TTGGTCTCCCTTTAGCAACG | 222 | 821505 |
| SSR3699 | TCGAGTTCGGTTAACTCATAACA | CCGAACACAAACTTAATTGGAA | 151 | 877473 |
| SSR9 | AAATATAACGATGTGGGTGGAAA | TGTCATGACTTAACCGAACACA | 292 | 877601 |
| SSR6-38 | CGTCGTTGGGTACTGATCCA | CCAACCAAGCCTTTCACACG | 329 | 1006357 |
| SSR67 | AAGTGGGATATTGCCATCCA | GCTATGTTAATAGTGTCATCAAATGAA | 274 | 1083921 |
| SSR3741 | AACATGGTCAAGATAATTGCACTA | AAAGTGCATACAGATGCCAAA | 244 | 1091420 |
| SSR3465 | GCTTAGGACGGATTTGGTAAA | TGCAAGTTTGAAGGAATATAATGAA | 229 | 1478980 |
| SSR3257 | GAATACTCCCTCATCCCAATAAA | TGATCGACACTTCTTCTGTCTCA | 246 | 1578266 |
| SSR3291 | ATCCTTGTTATGCTCCGCTC | TTCAATGGACTGTGAGGGTAAA | 236 | 1683443 |
| SSR3370 | TTTGGATCGGATTTGGGTTA | GCAATCAAATCCTTGAAGCC | 188 | 2183521 |

**Table 3-3**: Position details of seven ORFs

| Gene | Start (bp) | End (bp) | Strand | CDS (bp) | Annotation |
|---|---|---|---|---|---|
| ORF1 | 580856 | 581116 | + | 261 | *GLABRA2* interacting repressor |
| ORF2 | 605735 | 606576 | + | 813 | Probable CCR4-associated factor 1 homolog 11 |
| ORF3 | 607704 | 618632 | - | 2865 | Serine/threonine-protein phosphatase BSL2 |
| ORF4 | 626811 | 632509 | + | 1770 | Outer envelope protein 61 |
| ORF5 | 635236 | 638304 | + | 1158 | E3 ubiquitin-protein ligase RHF2A |
| ORF6 | 639890 | 642641 | - | 441 | Succinate dehydrogenase subunit 6, mitochondrial |
| ORF7 | 645188 | 647657 | + | 651 | Small heat-shock protein 1 |

## *3.4 Application of RNA-Seq to filter DEGs*

To further identify the candidate gene related to fiber initiation and fuzz development, RNA-Seq was performed on 12 independent ovule RNA libraries derived from two replicates each of the fuzzless mutant DPL972 and WT DPL971 at three fiber developmental stages (1, 3 and 5DPA). After removal of adaptors and low quality reads, 151,811,659 and 149,115,836 clean reads (44.73 and 45.54 Gb of clean data) were generated from DPL971 and DPL972, respectively. On average, we obtained 25,077,291 high quality reads (7.52 Gb of clean data) from each library (Table 3-S5). More than 91.00% of the total reads were then aligned and mapped to the published *G. arboreum* genome. After calculating FPKM values, a FDR normalized to a p-value <0.01 and a $|\log2\ (\text{ratio})| \geq 1$ was set as appropriate thresholds to distinguish significant differences in gene expression. The numbers of DEGs identified by RNA-Seq are statistically summarized in Fig. 3-5a as a Venn diagram. A total of 383 DEGs were identified as differentially expressed between the fuzzless mutant DPL972 and the wild type DPL971 at three developmental stages (Fig. 3-5b). Of them, 370 were differentially expressed at 1 DPA, while 13 and 22 were detected at 3 DPA and 5 DPA, respectively. Although fuzz emerged from the ovule epidermis at 3 to 5 DPA, this result indicates that the genes controlling fuzz development may come into play at an earlier stage.

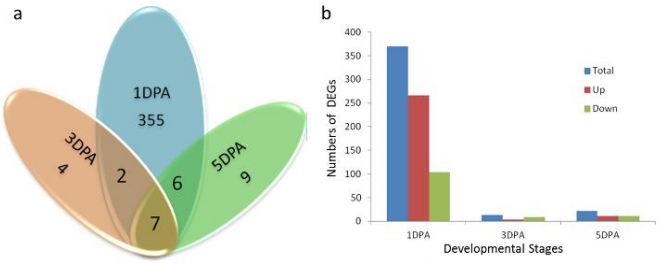



**Figure 3-5**: Differentially expressed genes detected by RNA-Seq. a. Venn graph of DEGs during different ovule developmental stages. b. Numbers of up-regulated/down-regulated DEGs at different stages.

## *3.5 Identification and sequence analysis of candidate genes*

According to the genome comparison and RNA-Seq data annotation, only two DEGs (ORF1 and ORF2) were tandemly located in the mapping region and both were notably up-regulated in DPL972 at 1 DPA. The other five genes exhibited no differences in expression levels. In addition, no InDels or SNPs were observed in the target region that could lead to a point mutation or frame-shift mutation of those five genes based on the re-sequencing results. We therefore finally identified ORF1 and ORF2 as fuzzless-trait candidate genes. ORF2 (Cotton_A_11942), annotated as CAF1, encodes a transcription factor annotated as a component of the CCR4 complex. In contrast, no annotation information is available in cotton for ORF1 (Cotton_A_11941), it could only be annotated on the basis of sequence similarity as a GIR1 gene (AT5G06270), which interacts with the GLABRA2 (GL2) repressor and is a regulator of root hair development.

Based on the re-sequencing data and amplification results, we compared the sequences of Cotton_A_11941 and Cotton_A_11942 in DPL971 and DPL972. No differences in the genomic or promoter sequence of Cotton_A_11942 were observed between the two parents. In Cotton_A_11941 of DPL972, however, we detected a TTG insertion that resulted in the insertion of a leucine rather than causing a shift or non-sense mutation (Fig. 3-6). In addition, BSA and re-sequencing data from the two parents and bulked pools revealed single G→A and C→A mutations in the promoter region of Cotton_A_11941 (Fig. 3-6). These mutations in the promoter may lead to differences in expression levels. Cotton_A_11941 is thus the most likely gene corresponding to *GaFzl*.

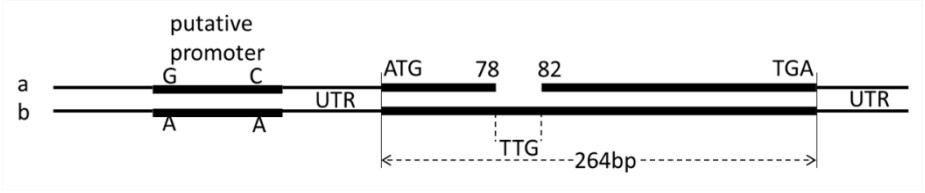



**Figure 3-6**: Sequence comparison of Cotton_A_11941 between DPL971 and DPL972. a. Gene sequence in DPL971. b. Gene sequence in DPL972

## *3.6 Expression profiling to check candidate genes*

To verify the accuracy of the RNA-Seq results, we examined the expression levels of candidate genes in ovules at different stages by qRT-PCR. As shown in Fig. 3-7, expression levels of Cotton_A_11941 at 1 and 3 DPA in DPL972 were much higher than those in DPL971, consistent with the RNA-Seq data. Interestingly, Cotton_A_11941 was annotated and identified as a regulator of root and trichome development, thus indicating that elevated expression of Cotton_A_11941 is more likely to result in a repression of fuzz development. This result further suggests that Cotton_A_11941 is the candidate gene controlling fuzz development.

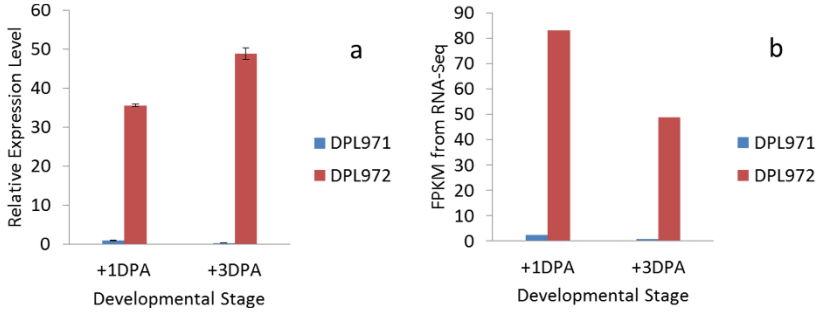



**Figure 3-7**: Expressions profiling of Cotton_A_11941 in DPL971 and DPL972. a. Relative expression level. b. FPKMs from RNA-Seq data. The x-axis represents different developmental stages. The y-axis corresponds to relative expression level. Error bars indicate standard deviations.

# 4. Discussion

## *4.1 A sequencing-assisted strategy is an efficient method for gene fine mapping*

Identification of target genes based on natural or mutagenic mutants is the central idea of forward genetics (Chang et al. 2016, Andres et al. 2017, Liu et al. 2018, Qureshi et al. 2018, Yang et al. 2018). Map-based cloning, the most efficient and visual strategy for gene identification, has been used to obtain numerous genes (Dou et al. 2018, Han et al. 2018, Jeong et al. 2018, Karthikeyan et al. 2018). For instance, a previous study used this technique to map and characterize a dominant single gene Gl2e that controls the development of gland pigment (Cheng et al. 2016). However, this strategy has apparent drawbacks as it is time-consuming and laborious. To guarantee accuracy and pinpoint recombinants, we always expend considerable time to develop sufficiently large populations and copious markers (Zhu et al. 2017). Otherwise, the experiment may be hindered or even fail to yield final results because variants are absent or the mapping parents possess low polymorphism.

Recently, BSA and re-sequencing have been of tremendous assistance in gene fine mapping and cloning (Yang et al. 2017). For instance, Wang et al. (2017) narrowed the genetic locus of citrus polyembryony to an 80-kb region through the strategy including de novo sequencing and BSA-Seq and identified a *CiRWP* gene responsible for this apomictic phenotype. As another example, Huang et al. (2017) performed BSA and deep sequencing in *Setaria viridis* to fine map the sparse panicle gene to a 1Mb interval and characterized *SvAUX1* as a regulator of inflorescence development and root gravitropism. In cotton, Zhu et al. (2017) combined the strategy of BSA and VIGS strategies to rapidly map and identify the target *v1* locus, an approach that proved to be an efficient and reliable method to complete map-based gene cloning.

In the current study, our combined application of BSA-Seq and traditional fine mapping uncovered a quite very narrow candidate region of 70 kb. We also performed a RNA-Seq analysis to assist filtering of candidate genes. We eventually confirmed

that only one candidate gene existed in the 70-kb region. High-throughput sequencing has greatly improved and accelerated fine mapping and cloning of genes with important agronomical traits.

## *4.2 Transcription factors and other genes regulate fiber and fuzz development*

As mentioned earlier, cotton fiber initiation and development share similarity with Arabidopsis trichome development (Hulskamp 2004, Yoshida et al. 2009). Previous studies have noted that homologs of the MYB-bHLH-WD40 complex also influence fiber initiation in cotton (Lu et al. 2017). *GhMYB2 (GL1)*, *GhMYB25/MIXTA*, *GhTTG1* and *GhHOX3 (GL2)* are essential for fiber initiation and development (Walford et al. 2012, Wang et al. 2014, Salih et al. 2016). In contrast, *GhCPC* functions as a negative regulator for fiber early initiation and elongation (Zhang et al. 2015). Recently Zhang proposed the existence of a new MYB-MIXTA-WD complex, including the *MIXTA* and *WD40* genes that controls fiber and fuzz formation.

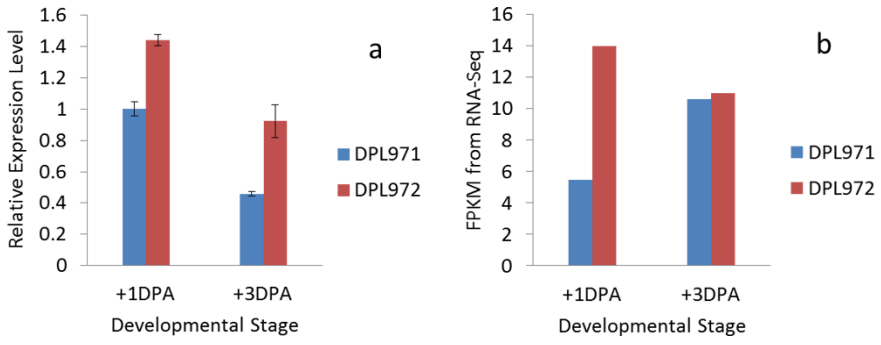



**Figure 3-8**: Expressions profiling of Cotton_A_11942 in DPL971 and DPL972. a. Relative expression level. b. FPKMs from RNA-Seq data. The x-axis represents different developmental stages. The y-axis corresponds to relative expression level. Error bars indicate standard deviations.

In this study, we found that Cotton_A_11942 is differentially expressed in DPL971 and DPL972, with qRT-PCR results consistent with the RNA-Seq data (Fig. 3-8). This gene encodes a cinnamoyl-CoA reductase (CCR)-associated factor (CAF1). Previous investigations have demonstrated that high expression of *GhCCR4* in fiber can cause an obvious thickening of fiber cell wall, which results in a pre-termination and a shorter fiber length. Although we failed to detect variable sites in Cotton_A_11942 between DPL971 and DPL972, distinct expression differences were definitely observed. Cotton_A_11942 may therefore act during fiber elongation and secondary wall thickening or function downstream of genes controlling fuzz development.

We found that the most likely candidate gene, Cotton_A_11941, shared a high sequence similarity with *GIR1* in Arabidopsis (Wu and Citovsky 2017). Using loss- and gain-of-function strategies, Wu and Citovsky (2017) discovered that GIR1 and GIR2 interact with the repressor of GL2 to repress root hair development. We checked

the expression levels of GaGL2 in our study, and detected no significant differences between the wild type and the mutant (Fig. S1), which suggests that GaGIR1 may interact with GaGL2 and repress the transcriptional activity of GaGL2 to regulate the downstream gene and fuzz formation. The function of GIR1 in cotton is still unclear, however, and we suspect that GIR1 also influences fiber/fuzz initiation and formation. Cotton_A_11941, as the homolog of *GIR1*, is thus most likely the gene that regulates fuzz development. According to RNA-Seq data, in addition, expression levels of MYB-MIXTA-WD complex genes were not significantly different between DPL971 and DPL972, thus suggesting that Cotton_A_11941 may function at the downstream of the MYB-MIXTA-WD complex to participate in the control of fuzz development.

**Author contribution statement** GS managed the project. GS, XF, HC and DZ designed the research. XF, HC, DZ, YZ, QW, KL, JA, QY, SL and XC performed the experiments. XF wrote the paper. All authors read and approved the final manuscript.

**Acknowledgments** This work was supported by grants from the National Natural Science Foundation of China (No. 31621005) and the National Key R and D Plan of China (No. 2018YFD0100402). We thank Hervé Vanderschuren (Gembloux Agro BioTech, University of Liège), for revising the draft of this manuscript.

**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical standard** The experiments comply with the ethical standards in the country in which they were performed.

# References

Andres R J, Coneva V, Frank M H, Tuttle J R, Samayoa L F, Han S W, Kaur B, Zhu L, Fang H, Bowman D T, Rojas-Pierce M, Haigler C H, Jones D C, Holland J B, Chitwood D H and Kuraparthy V (2017) Modifications to a *LATE MERISTEM IDENTITY1* gene are responsible for the major leaf shapes of Upland cotton (*Gossypium hirsutum L.*). Proc Natl Acad Sci U S A 114 (1): E57-E66.

Chang L, Fang L, Zhu Y, Wu H, Zhang Z, Liu C, Li X and Zhang T (2016) Insights into interspecific hybridization events in allotetraploid cotton formation from characterization of a gene-regulating leaf shape. Genetics 204 (2): 799-806.

Cheng H, Lu C, Yu J Z, Zou C, Zhang Y, Wang Q, Huang J, Feng X, Jiang P and Yang W (2016) Fine mapping and candidate gene analysis of the dominant glandless gene $Gl_2^e$ in cotton (*Gossypium spp.*). Theor Appl Genet 129 (7): 1347-1355.

Deng W, Wang Y, Liu Z, Cheng H and Xue Y (2014) HemI: A Toolkit for Illustrating Heatmaps. PLoS One 9 (11): e111988.

Dou J, Zhao S, Lu X, He N, Zhang L, Ali A, Kuang H and Liu W (2018) Genetic mapping reveals a candidate gene (*ClFS1*) for fruit shape in watermelon (*Citrullus lanatus* L.). Theor Appl Genet 131 (4): 947-958.

Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H and Matsumura H (2013) MutMap+: Genetic mapping and mutant identification without crossing in rice. PLoS One 8 (7): e68529.

Feng X, Zou C, Lu C, Cheng H and Zhang Y (2016) Simple sequence repeat markers closely linked with a new fuzzless gene in fuzzless mutant DPL972 (*Gossypium arboreum*). Cotton Science 28 (4): 392-398.

Han Y, Zhao F, Gao S, Wang X, Wei A, Chen Z, Liu N, Tong X, Fu X, Wen C, Zhang Z, Wang N and Du S (2018) Fine mapping of a male sterility gene *ms-3* in a novel cucumber (*Cucumis sativus* L.) mutant. Theor Appl Genet 131 (2): 449-460.

Huang P, Jiang H, Zhu C, Barry K, Jenkins J, Sandor L, Schmutz J, Box M S, Kellogg E A and Brutnell T P (2017) *Sparse panicle1* is required for inflorescence development in *Setaria viridis* and maize. Nat Plants 3: 17054.

Hulskamp M (2004) Plant trichomes: a model for cell differentiation. Nat Rev Mol Cell Biol 5 (6): 471-480.

Ishida T, Kurata T, Okada K and Wada T (2008) A genetic regulatory network in the development of trichomes and root hairs. Annual Review of Plant Biology 59 (1): 365-386.

Jeong K, Choi D and Lee J (2018) Fine mapping of the genic male-sterile *ms1* gene in *Capsicum annuum* L. Theor Appl Genet 131 (1): 183-191.

Jiang Y, Ding M, Cao Y, Yang F, Zhang H, He S, Dai H, Hao H and Rong J (2015) Genetic fine mapping and candidate gene analysis of the *Gossypium hirsutum Ligon lintless-1 ($Li_1$)* mutant on chromosome 22 (D). Mol Genet Genomics 290 (6): 2199-2211.

Karthikeyan A, Li K, Li C, Yin J, Li N, Yang Y, Song Y, Ren R, Zhi H and Gai J (2018) Fine-mapping and identifying candidate genes conferring resistance to Soybean mosaic virus strain *SC20* in soybean. Theor Appl Genet 131 (2): 461-476.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R and Salzberg S L (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14 (4): R36.

Lee J J, Woodward A W and Chen Z J (2007) Gene expression changes and early events in cotton fiber development. Ann Bot 100 (7): 1391-1401.

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel R J, Ma Z, Shang H, Ma X, Wu J, Liang X, Huang G, Percy R G, Liu K, Yang W, Chen W, Du X, Shi C, Yuan Y, Ye W, Liu X, Zhang X, Liu W, Wei H, Wei S, Huang G, Zhang X, Zhu S, Zhang H, Sun F, Wang X, Liang J, Wang J, He Q, Huang L, Wang J, Cui J, Song G, Wang K, Xu X, Yu J Z, Zhu Y and Yu S (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol 33 (5): 524-530.

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, Chen W, Liang X, Shang H, Liu W, Shi C, Xiao G, Gou C, Ye W, Xu X, Zhang X, Wei H, Li Z, Zhang G, Wang J, Liu K, Kohel R J, Percy R G, Yu J Z, Zhu Y X, Wang J and Yu S (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46 (6): 567-572.

Li H and Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform, Oxford University Press.

Liang W, Fang L, Xiang D, Hu Y, Feng H, Chang L and Zhang T (2015)

Transcriptome analysis of short fiber mutant *Ligon lintless-1* ($li_1$) reveals critical genes and key pathways in cotton fiber elongation and leaf development. PLoS One 10 (11): e0143503.

Liu K, Xu H, Liu G, Guan P, Zhou X, Peng H, Yao Y, Ni Z, Sun Q and Du J (2018) QTL mapping of flag leaf-related traits in wheat (*Triticum aestivum* L.). Theor Appl Genet 131 (4): 839-849.

Livak K J and Schmittgen T D (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25 (4): 402-408.

Lu N, Roldan M and Dixon R A (2017) Characterization of two TT2-type MYB transcription factors regulating proanthocyanidin biosynthesis in tetraploid cotton, *Gossypium hirsutum*. Planta 246 (2): 323-335.

Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S and Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20 (9): 1297-1303.

Naoumkina M, Thyssen G N, Fang D D, Hinchliffe D J, Florane C B and Jenkins J N (2016) Small RNA sequencing and degradome analysis of developing fibers of short fiber mutants *Ligon-lintles-1* ($Li_1$) and *-2* ($Li_2$) revealed a role for miRNAs and their targets in cotton fiber elongation. BMC Genomics 17: 360.

Parekh M J, Kumar S, Zala H N, Fougat R S, Patel C B, Bosamia T C, Kulkarni K S and Parihar A (2016) Development and validation of novel fiber relevant dbEST–SSR markers and their utility in revealing genetic diversity in diploid cotton (*Gossypium herbaceum* and *G. arboreum*). Industrial Crops and Products 83: 620-629.

Qureshi N, Bariana H, Kumran V V, Muruga S, Forrest K L, Hayden M J and Bansal U (2018) A new leaf rust resistance gene *Lr79* mapped in chromosome 3BL from the durum wheat landrace Aus26582. Theor Appl Genet 131 (2): 1-8.

Rong J, Pierce G J, Waghmare V N, Rogers C J, Desai A, Chee P W, May O L, Gannaway J R, Wendel J F, Wilkins T A and Paterson A H (2005) Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theor Appl Genet 111 (6): 1137-1146.

Salih H, Gong W, He S, Sun G, Sun J and Du X (2016) Genome-wide characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*. BMC Genet 17 (1): 129.

Sun R, Li C, Zhang J, Li F, Ma L, Tan Y, Wang Q and Zhang B (2017) Differential expression of microRNAs during fiber development between fuzzless-lintless mutant and its wild-type allotetraploid cotton. Sci Rep 7 (1): 3.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley D R, Pimentel H, Salzberg S L, Rinn J L and Pachter L (2016) Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. Nature Protocols 7 (3): 562.

Walford S-A, Wu Y, Llewellyn D J and Dennis E S (2012) Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, *GhHD-1*. The Plant Journal 71 (3): 464-478 .

Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, Fang L, Yang S, Hu Y, Ye W, Zhang

H, Ma P, Chen J, Wang Q, Mei G, Cai C, Yang D, Wang J, Guo W, Zhang W, Chen X and Zhang T (2016) Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. New Phytol 210 (4): 1298-1310.

Wan Q, Zhang H, Ye W, Wu H and Zhang T (2014) Genome-wide transcriptome profiling revealed cotton fuzz fiber development having a similar molecular model as Arabidopsis trichome. PLoS One 9 (5): e97313.

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel R J, Percy R G, Yu J Z, Zhu Y X, Wang J and Yu S (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet 44 (10): 1098-1103.

Wang L, Cook A, Patrick J W, Chen X Y and Ruan Y L (2014) Silencing the vacuolar invertase gene *GhVIN1* blocks cotton fiber initiation from the ovule epidermis, probably by suppressing a cohort of regulatory genes via sugar signaling. Plant J 78 (4): 686-696.

Wang X, Xu Y, Zhang S, Cao L, Huang Y, Cheng J, Wu G, Tian S, Chen C, Liu Y, Yu H, Yang X, Lan H, Wang N, Wang L, Xu J, Jiang X, Xie Z, Tan M, Larkin R M, Chen L L, Ma B G, Ruan Y, Deng X and Xu Q (2017) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet 49 (5): 765-772.

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 Phytol 217 (2): 883-895.

Wu R and Citovsky V (2017) Adaptor proteins GIR1 and GIR2. I. Interaction with the repressor GLABRA2 and regulation of root hair development. Biochem Biophys Res Commun 488 (3): 547-553.

Yang M, Chen L, Wu X, Gao X, Li C, Song Y, Zhang D, Shi Y, Li Y, Li Y X and Wang T (2018) Characterization and fine mapping of *qkc7.03*: a major locus for kernel cracking in maize. Theor Appl Genet 131 (2): 437-448.

Yang X, Xia X, Zhang Z, Nong B, Zeng Y, Xiong F, Wu Y, Gao J, Deng G and Li D (2017) QTL mapping by whole genome re-sequencing and analysis of candidate genes for nitrogen use efficiency in rice. Front Plant Sci 8: 1634.

Yoshida Y, Sano R, Wada T, Takabayashi J and Okada K (2009) Jasmonic acid control of GLABRA3 links inducible defense and trichome patterning in Arabidopsis. Development 136 (6): 1039-1048.

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski C A, Scheffler B E, Stelly D M, Hulse-Kemp A M, Wan Q, Liu B, Liu C, Wang S, Pan M, Wang Y, Wang D, Ye W, Chang L, Zhang W, Song Q, Kirkbride R C, Chen X, Dennis E, Llewellyn D J, Peterson D G, Thaxton P, Jones D C, Wang Q, Xu X, Zhang H, Wu H, Zhou L, Mei G, Chen S, Tian Y, Xiang D, Li X, Ding J, Zuo Q, Tao L, Liu Y, Li J, Lin Y, Hui Y, Cao Z, Cai C, Zhu X, Jiang Z, Zhou B, Guo W, Li R and Chen Z J (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol 33 (5): 531-537.

Zhang X, Liu B, Zhu Y and Zhang T (2015) The R3-MYB gene *GhCPC* negatively regulates cotton fiber elongation. PLoS One 10 (2): e0116272.

Zhu J, Chen J, Gao F, Xu C, Wu H, Chen K, Si Z, Yan H and Zhang T (2017) Rapid mapping and cloning of the *virescent-1* gene in cotton by bulked segregant analysis-next generation sequencing and virus-induced gene silencing strategies. J Exp Bot 68 (15): 4125-4135.

Zhu Q H, Yuan Y, Stiller W, Jia Y, Wang P, Pan Z, Du X, Llewellyn D and Wilson I (2018) Genetic dissection of the fuzzless seed trait in *Gossypium barbadense*. J Exp Bot 69 (5): 997-1009.

# *Supplementary Materials*

**Table S3-1**: Information of BSA-seq data

| Sample ID | Clean-Base | Q30(%) | Total_reads | Mapped (%) | Properly mapped(%) | Ave depth | Cov ratio1X (%) |
|---|---|---|---|---|---|---|---|
| DPL971P1 | 53,736,668,700 | 90.96 | 358244458 | 97.98 | 81.8 | 31 | 64.59 |
| DPL972P2 | 60,905,420,400 | 91.22 | 406036136 | 97.79 | 82.27 | 35 | 64.75 |
| FuzzyPool1 | 60,910,934,100 | 90.82 | 406072894 | 98 | 81.43 | 35 | 64.85 |
| FuzzlessPool2 | 53,419,518,900 | 91.23 | 356130126 | 97.7 | 81.56 | 30 | 64.54 |

**Table S3-2**: Number of SNP detected by the BSA-seq

| Type | R04vsR03 | R06vsR05 |
|---|---|---|
| INTERGENIC | 598,506 | 455,600 |
| INTRON | 22,934 | 8,937 |
| UPSTREAM | 51,420 | 20,206 |
| DOWNSTREAM | 39,499 | 16,212 |
| SPLICE_SITE_ACCEPTOR | 23 | 10 |
| SPLICE_SITE_DONOR | 22 | 3 |
| SPLICE_SITE_REGION | 260 | 51 |
| START_LOST | 12 | 3 |
| SYNONYMOUS_CODING | 3,034 | 1,010 |
| NON_SYNONYMOUS_CODING | 4,314 | 1,353 |
| SYNONYMOUS_STOP | 4 | 2 |
| STOP_GAINED | 115 | 47 |
| STOP_LOST | 7 | 2 |
| Other | 142 | 115 |
| Total | 720,292 | 503,551 |

Note: R3: Fuzzy parent DPL971.
R4: Fuzzless parent DPL972.
R5: Bulked pool of 30 extreme fuzzy progenies.
R6: Bulked pool of 30 extreme fuzzless progenies.

**Table S3-3**: Number of InDel detected by the BSA-seq

| Type | R04vsR03 | R06vsR05 |
|---|---|---|
| INTERGENIC | 81,013 | 60,307 |
| INTRON | 7,319 | 3,238 |
| UPSTREAM | 16,392 | 6,654 |
| DOWNSTREAM | 12,189 | 5,036 |
| SPLICE_SITE_ACCEPTOR | 8 | 5 |
| SPLICE_SITE_DONOR | 23 | 11 |
| SPLICE_SITE_REGION | 83 | 51 |
| START_LOST | 8 | 1 |
| FRAME_SHIFT | 372 | 180 |
| CODON_DELETION | 57 | 17 |
| CODON_INSERTION | 82 | 29 |
| CODON_CHANGE_PLUS_CODON_DELETION | 41 | 16 |
| CODON_CHANGE_PLUS_CODON_INSERTION | 18 | 16 |
| STOP_GAINED | 11 | 3 |
| STOP_LOST | 6 | 2 |
| Other | 26 | 22 |
| Total | 117,648 | 75,588 |

Note: R3: Fuzzy parent DPL971.

R4: Fuzzless parent DPL972.

R5: Bulked pool of 30 extreme fuzzy progenies.

R6: Bulked pool of 30 extreme fuzzless progenies.

**Table S3-4**: Details of all designed markers in the fine mapping region of A08

| Primer Name | Forward Primer (5'→3') | Reverse Primer (5'→3') | Product Length(bp) | Start position(bp) |
|---|---|---|---|---|
| SSR1 | ATCACCTTGATTGGCACACA | TAAACAAGCCCATGCAAACA | 209 | 7267 |
| SSR2 | GCATCAATTCCATGGTGTCTT | TGAAATTTCCACTAACTAGGATGAA | 235 | 7906 |
| SSR3 | AGGGAAAGAAGCTGAGAGGG | TTAACTTGGGCCAGATACAAA | 198 | 8899 |
| SSR4 | TTTGCTTAAGTCTATTTCGGTTGA | GGCTTGAGCTTAACCCGTTT | 286 | 35886 |
| SSR5 | AATGACACTCTCCCGACGAC | TGAAGTGATGATATGGCACTTTG | 169 | 36414 |
| SSR6 | TTCTCTGAGCTTGCTGGGAT | GGGACGTCAAGTAAATTAATCCT | 278 | 61694 |
| SSR7 | AAACAAAGGTTGATGGCAAA | TTGGGTCGCTATCAATCTAAA | 302 | 68265 |
| SSR8 | TTAAAGGTCCAACATCGCGT | TGCATGTATAGGTGAATCTTGAGG | 251 | 78387 |
| SSR9 | GACCATGTTAATGACAACGATGA | TGCGCACTATCACATACCGT | 182 | 79010 |
| SSR10 | ACCACATGACACATGGTAATTAGA | CCAATCAATGCCAGTCAACA | 133 | 81310 |
| SSR11 | GAATTCCCTCACAGCTCCAA | GTCTCGCTGCGATTTCTAGG | 329 | 82272 |
| SSR12 | GGGCCATTAAGCAATAGGGA | GGGAACCAACGAACTCAGTC | 350 | 88280 |

| | | | | |
|---|---|---|---|---|
| SSR13 | AAGCATAATAGCCCGTATGACAA | TTTAACGGCTTACTTCATGTTGA | 181 | 88901 |
| SSR14 | CAATCAAGATCTGAACAAAGTCG | GCTCAAGCTCGACCTGATCT | 259 | 103185 |
| SSR15 | CTCTCCCATTGCCATCACTT | TGGTTGGTATATGTAGTTGAGGGA | 225 | 104037 |
| SSR16 | ATATGCATGGACGGATGGAT | CCGAGTTCGAAGCTGAAAGT | 190 | 112969 |
| SSR17 | GTGCCAAGAACAAGGAGGAG | GTGCAATATTCATTGGACCG | 296 | 114472 |
| SSR18 | CACGAGTGCTAATGGGACAA | ATCCAGGCAAATTGCGTACT | 190 | 114727 |
| SSR19 | GCTAACACGCTCAGCAAGTG | TACTGGTGTCAACGATCCGA | 293 | 115407 |
| SSR20 | TCGGATCGTTGACACCAGTA | ATCCTCCTCCTCCTCCTCCT | 195 | 115718 |
| SSR21 | ACGCATAAACGAACTCGACA | AAATTCCAGGAAACTGCAAA | 322 | 115733 |
| SSR22 | ATTCGGATATGCTCAGTGCC | TTGATGGTTAGTGCAATCCG | 312 | 130308 |
| SSR23 | CCAATTGGAGAACACAAATTAAA | CCTTGGGATTTAGGCTTTCC | 251 | 133736 |
| SSR24 | GTTTGGTGGCTTTGTTGGTT | CGAAGTTCAAACCTATATTCTCCAA | 224 | 135323 |
| SSR25 | TTTGCGAGTACAATTATGTGAAA | CAAATCATGCTAGTAGTCCCTTTG | 239 | 139953 |
| SSR26 | CTGTCGTATCCCTCCGTGTT | GATTTCAATGCCTGATCGCT | 313 | 141729 |
| SSR27 | TCGATGTTATATCAATTAGGTCATTC | CGACTTTAGATCTCATATGGCGT | 273 | 144282 |
| SSR28 | AAATGTCTCCGTTTCCATCG | CCTTCACCTGATTTCCCAAA | 267 | 178822 |
| SSR29 | TTTGGGAAATCAGGTGAAGG | AGCTGAATGAGCAGCAAATG | 264 | 179193 |
| SSR30 | TGCAAACCCACGTTATGTGT | GGTTAAAGGAGATGGGAGGG | 250 | 189112 |
| SSR31 | GCGAATTTCTCTACCTCTGCAA | CACACATCAAACTTCTAATTGGAATAC | 251 | 201766 |
| SSR32 | CTCATCCTCTCCATTGGCAT | ATCAATTCCCACATCAAGCC | 283 | 216626 |
| SSR33 | AAGACGATGAAGAGCACGGT | GAAAGTCCACAAGCTGCTCC | 205 | 217094 |
| SSR34 | AAATTGTTAAGTTTGCCCGC | CACTCTTGCTCCACCATCAA | 241 | 217621 |
| SSR35 | TGTCATTTCTATCACAGCAGCA | TTGTTAGCAAGCATTCGTCG | 282 | 219094 |
| SSR36 | TTGATAAATTCGATTGTTTGATTG | CCCTAGATTTGCATTGCATGA | 313 | 234864 |
| SSR37 | GGGAGGGACTATGAGCAGTTC | AAGTTTAAGCCTCATCACATTCAA | 126 | 235763 |
| SSR38 | ATGCTTGAACCCTGCCATTA | TCCCTTCAGCAACAATTGAA | 262 | 251584 |
| SSR39 | AAGATTCAAAGTTTAGGACGTATGG | AAATCTTAAACTCTCGAAATGATTG | 239 | 258303 |
| SSR40 | AGGGAAAGCGTTTCCAGTAA | ATGAATTCGGCCTGTCACTC | 325 | 266237 |
| SSR41 | CGAAGAGGCTGAAACCCATA | ATGTTGCAGTTGGTGCTGAG | 341 | 268367 |
| SSR42 | TGCATGCCTGATAATAAGCAA | CAGGGTCTGTTTGCAGAATTT | 202 | 268972 |
| SSR43 | AGTGAAGAGTCGAATTTGCCA | AGCTGAATAGTTTGATGCATTATTT | 298 | 276049 |
| SSR44 | TCACCCACTGTAATCAGTGCTC | TGGGTTTGAGTGTGCTGAAG | 173 | 278986 |
| SSR45 | GTGCCAAACAACAATCTCCC | TGCATGTGGAGTTGAACCAT | 187 | 280568 |
| SSR46 | TGTTTGATGCGAGTTCACCT | CATGTGAACCTTCTCTATTGCTTT | 254 | 282267 |
| SSR47 | CCCAACATTAGCGGTGTTAG | GTGAAACGGAGGGAACTTCA | 137 | 291474 |
| SSR48 | AGCCCAATTAGCCCACAAC | GATTTCAGGCTTTGTAGCCG | 194 | 292801 |
| SSR49 | GGCTGAAATGGGTTAGGTCA | TTGAAAGTATTTACATCGGATAATGT | 187 | 298359 |
| SSR50 | GGCTGAAATGGGTTAGGTCA | TTGAAAGTATTTACATGGGATAATGT | 187 | 299425 |
| SSR51 | CACGGAAGTACACCCAACCT | GAGAAGCTACGAATGTCGGC | 286 | 301593 |
| SSR52 | GGACTAAGACCATTATCTACTCCATAG | TTGAAAGTATTTACATGGGATAATGT | 273 | 302988 |
| SSR53 | CGGCCAGTTGAATAGGAAAC | GGCTAAGGCTACTCAATCGC | 298 | 303757 |

| | | | | |
|---|---|---|---|---|
| SSR54 | GCCTTAATCCTCAAATTTCCC | CCCTAAGTGCCATGTGTGTG | 250 | 317941 |
| SSR55 | CGAACACACAAAGACCCAAA | TGAAATTCAAATATGAACACCAAA | 296 | 320488 |
| SSR56 | TGACCTTGACCCGGAAATAA | AAATCAAAGTTTATTATGCCCAAA | 149 | 321626 |
| SSR57 | TGATATTGGCGTGGGAACTT | ATATTTGGCCCTCAACCTTT | 287 | 322245 |
| SSR58 | TTGCAACTGTGGTTACTTACAGC | GGCATGAAGGAATGCAGATT | 165 | 329336 |
| SSR59 | CAGTCAGATTGATTGGACTGAGA | TTGTGCACACCATCACTTGT | 322 | 341123 |
| SSR60 | TTCCTGTCTGGGTTTGCTTT | CCGGAAGTTTAAGGTTGTTGC | 281 | 342137 |
| SSR61 | CCGAACCACGTAAATCCTTG | TGAGCTACTCAATACTCAAACTCG | 227 | 349626 |
| SSR62 | AGCCCATTTGATGCAATCTG | GCCAAATCGAAATCGAGAAC | 180 | 357760 |
| SSR63 | CCAAGCTCAAGAGGGAATTG | CGAATGGAATCGATGAAGGA | 339 | 361419 |
| SSR64 | TAGCTCATCGGTCCCAGTTT | GCACGCATGCATACCTACAT | 228 | 366943 |
| SSR65 | CACTCGGCCCATAACTTGAT | CCGCCAACTTGGAAATCTAC | 289 | 372200 |
| SSR66 | CGGGCCATGAACATCTCTAA | ACAGTAACAGGCGGTATCGG | 344 | 374080 |
| SSR67 | ATGCCAGTAGGTGTTGGTCC | GTCCAGGATACGGTGTTGCT | 337 | 375142 |
| SSR68 | ATGAGGAGGACGAGCAAGAA | CAGCCTGCTGAATCTGATGA | 255 | 377194 |
| SSR69 | ATGAGGAGGACGAGCAAGAA | TGCCTTCTTCAGAATTGCCT | 336 | 377214 |
| SSR70 | TTTATGATCCCTTTCACCCAA | GCATACAGCTGAATATATTAACACTCA | 216 | 398847 |
| SSR71 | TGAAGGAAGAAGGGATTAATTTCA | AGAGCCAAAGGGTAGTGGGT | 154 | 419165 |
| SSR72 | TCGATGGTGTATTTGGCTCA | ATGCTTTGTTCCACAGCACA | 299 | 424052 |
| SSR73 | ATTGTCGGTACAGGAGGACG | TCCTTGTGCAAGAAGACGTG | 221 | 424456 |
| SSR74 | AAATCCAAGAGAATTGAGTGATTAAA | GTTGCTACTTGCCACGTTGA | 197 | 425765 |
| SSR75 | CATGTTAGTCGGTTGGTGACA | TGCAGATCACAGAGAATGGC | 302 | 437510 |
| SSR76 | TGCCATTCTCTGTGATCTGC | ATGATTCCAACCCGATTCAA | 176 | 437770 |
| SSR77 | CTGATTCTCAAGTGGCACGA | TTGCAAATTCATCTCCCTTTG | 156 | 440278 |
| SSR78 | GGCACTTGGCATTGTTTCTT | ACGGCAAAGTTGACGGAATA | 223 | 444245 |
| SSR79 | CCACAATCAAAGTTTAATGCGTA | GAGCTCTAGTTTAGTCGGGTGG | 225 | 446768 |
| SSR80 | TCAACTTCAATGCACGAGGA | GCAAACAGCTCTGTGAACGA | 125 | 455339 |
| SSR81 | CTGAAACCCTCATTTGCAGC | GGGTCTCTTCGCCGTAGTC | 157 | 459808 |
| SSR82 | CCTTCCATGCATATTGGAAA | CAAAGCACCCAATTTCAAGG | 283 | 469010 |
| SSR83 | CACCTTGAAATTGGGTGCTT | GCTTGGTGACCCTCTAGCTG | 191 | 469261 |
| SSR84 | AGCTAGAGGGTCACCAAGCA | CATCCCACTAGCAGCATTGA | 162 | 469435 |
| SSR85 | TGTTAATCATGTGATTTATTGAGTGTG | CATGAACTACCACGCGGTTT | 257 | 470082 |
| SSR86 | TGTTGTTAAATTGCTTACTTGGGA | TCGATACCTACCAAGGTTTGAA | 310 | 485966 |
| SSR87 | TTATTATCCGCCCAAACAGC | AACTTGGCAGGTGGATTGAT | 325 | 486744 |
| SSR88 | TCCTATTTGCTGGCGAGATT | AAGTGGCCTGACCCATACAC | 316 | 490174 |
| SSR89 | TGAACGTATACGAAGATTTGGTT | AAGTGGCCTGACCCATACAC | 253 | 490296 |
| SSR90 | AATGCCTTACCTTTGAAATGC | TGGACTTGTTCTTTCTTGTTGC | 136 | 491476 |
| SSR91 | AACAGGAGCATTGATGCAAA | GCTGCTGAAGCTGAGGCTAT | 306 | 498723 |
| SSR92 | GTATCTGCAAGGGTGGAGGA | AAGATACGCAATGGCAAAGC | 203 | 520481 |
| SSR93 | TTTGAGTTTAAATTGAACGGATTT | TCTTGGTGGTGGAGTTCACA | 160 | 544219 |
| SSR94 | TGTGAACTCCACCACCAAGA | TTCACTACCATCATTCGGGA | 259 | 544357 |

| | | | | |
|---|---|---|---|---|
| SSR95 | TGTGAACTCCACCACCAAGA | TTCACTACCATCATTCGGGA | 259 | 544421 |
| SSR96 | GGGCAAACTGCAATACGAGT | AACCAATCAGCATAGCCCAC | 275 | 546417 |
| SSR97 | TTGTTTCGCAACGAATTAAATG | TTAACCTTACTCCTTCGAACCAT | 237 | 547865 |
| SSR98 | AATTGTGGGATCTGCTGCAT | TGGCGTGGACATGTACCTTA | 323 | 551918 |
| InDel1 | GGCTGCCACCCCCC | CCTCACATTTAACGTCAAGCTCAC | 356 | 76294 |
| InDel2 | CCAAACATCCATGTGGACCAATTTG | GGTAATAACTGTTCTTTATGAGCCCTTTTTC | 360 | 501736 |
| InDel3 | CATCAACGTGGCAAGTAGCAACA | CCCAGGGGCAAAAAAAATTTGGG | 372 | 525615 |
| InDel4 | GCCCCTCCAGATTTGCCTC | GGTCCAAATTGTCTAAAATAAATCCGTAGAGAC | 565 | 531453 |
| InDel5 | CGAATCCTGAACCCCAAACCTAAAC | CATCACCATGGCAACAACTCC | 376 | 532619 |
| InDel6 | CGCGAGGACTAAAATTTGAAAGTTTGGA | GGTAAGGATTGGGGCATTAACTGG | 178 | 534201 |
| InDel7 | CCATAATTAGAGTTTCACGTGTATAATTGTACC | GAGTTTTGTGTCACCACTCAATCG | 256 | 534829 |
| InDel8 | GGCTTGTGCAGCCATTGTG | TCGTGCATTAAAAAGAACATGCTAGG | 377 | 640157 |
| InDel9 | GATTTCCCTTTCATCAAATATTCTATGTTAGCG | ACTCAAATACTTAGTATACCATATTAGCCTCTTTC | 536 | 650514 |
| InDel10 | AGTAGGTGCTACAAAATCAGCTCC | CTGTTTTAGCATTTTGGGTGCTCTTTTC | 327 | 664027 |
| InDel11 | CACAACATTTAGTAAATAGAAATTTTGGGCAATCC | GGTTATTGACTTGAAGAGGGGGG | 252 | 666443 |

Note: markers in highlight are polymorphic between two mapping parents.

**Table S3-5**: Details of all designed markers in the fine mapping region of A08

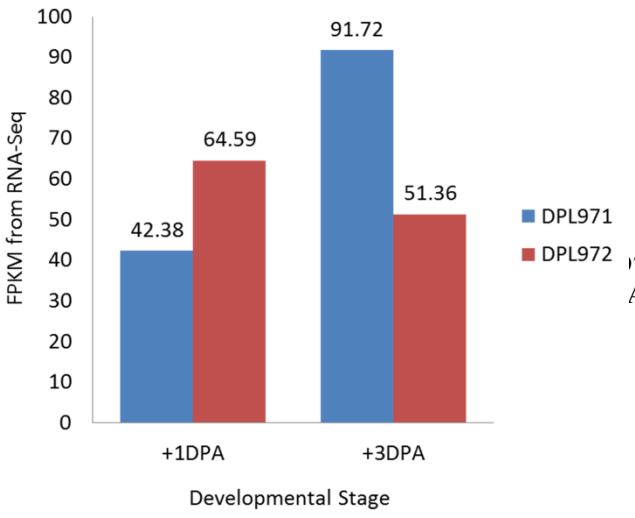
| Sample | Clean Bases (bp) | Clean Q30(%) | Total reads | Read mapping rate | Multiple mapping rate |
|---|---|---|---|---|---|
| DPL971+1R1 | 7,168,622,700 | 95.38 | 23895409 | 91% | 5% |
| DPL971+1R2 | 6,412,138,500 | 95.4 | 25321633 | 91% | 9% |
| DPL971+3R1 | 7,571,760,300 | 95.31 | 25239201 | 91% | 6% |
| DPL971+3R2 | 7,265,123,100 | 95.46 | 23930032 | 91% | 5% |
| DPL971+5R1 | 7,042,378,500 | 95.27 | 23474595 | 91% | 5% |
| DPL971+5R2 | 8,017,725,900 | 95.28 | 27254966 | 91% | 6% |
| DPL972+1R1 | 7,596,489,900 | 95.41 | 21373795 | 91% | 7% |
| DPL972+1R2 | 8,608,806,300 | 95.46 | 28696021 | 91% | 7% |
| DPL972+3R1 | 7,179,009,600 | 95.25 | 24217077 | 91% | 4% |
| DPL972+3R2 | 7,619,923,500 | 95.24 | 25399745 | 91% | 5% |
| DPL972+5R1 | 8,176,489,800 | 95.51 | 26725753 | 92% | 6% |
| DPL972+5R2 | 7,619,780,400 | 95.35 | 25399268 | 91% | 7% |

**Table S3-6**: Sequence information about amplification primers of genes and promoters

**Table S3-7**: Sequence information of qRT-PCR primers

| Primer Name | Product Length(bp) | ForwardPrimer (5'→3') | ReversePrimer (5'→3') |
|---|---|---|---|
| 11941 | 264 | ATGTCAGTTTCTCCCCTG GAAATATCG | TCAATTGCTTGCCCATT TAGCAGTC |
| 11941P | 2000 | TCCCAGGTCACTAACAA TATTGTCG | GACATGCTTGGTGACC CTCTAG |
| 11942 | 843 | ATGGGTTTTGAAGATTC AGATAGGGTATC | TTAATAAACCTCTAAT CCAAATAATACTCCGG CA |
| 11942P | 2000 | CTATGACCATGATTACG CCAAGCTTGTCTATTTTG GTCATTAACGGGCTAAC | ACCACCCGGGGATCCT CTAGATATTGGAAAAT CCAAGAAATTACAACA AAAAAAAAGAGA |

| Primer Name | Product Length | ForwardPrimer (5'→3') | ReversePrimer (5'→3') |
|---|---|---|---|
| q11941-2 | 137 | GTCTTCGGAACCCGGAGATG | TCCCACTAGCAGCATTGACG |
| q11941-3 | 114 | TGGGTGCTTTGAAGGCTTGT | CTGACATGCTTGGTGACCCT |
| q11942-1 | 126 | GGAGTTCGAGTTGATCAGCCA | TTTGAGTACCGTTGTACGGC |
| q11942-2 | 133 | TCCTGACCCGACGAGATTTG | CCCACCATACAAGCTGTTACAA |




**Figu** '72 during different A-Seq.

# 4

# Genome-wide identification and expression analysis of GL2-interacting-repressor (GIR) genes during cotton fiber and fuzz development

*GL2-interacting-repressor (GIR) family members may contribute to fiber/fuzz formation via a newly discovered unique pathway in Gossypium arboreum.*

From **Feng, X.,** Cheng, H., Zuo, D. et al. Genome-wide identification and expression analysis of GL2-interacting-repressor (GIR) genes during cotton fiber and fuzz development. Planta 255, 23 (2022).

**Abstract:** There are similarities between cotton fiber development and the formation of trichomes and root hairs. The GL2-interacting-repressors (GIRs) are crucial regulators of root hair and trichome formation. The *GaFzl* gene, annotated as *GaGIR1*, is negatively associated with trichome development and fuzz initiation. However, there is relatively little available information regarding the other *GIR* genes in cotton, especially regarding their effects on cotton fiber development. In this study, 21 *GIR* family genes were identified in the diploid cotton species *Gossypium arboreum*; these genes were divided into three groups. The *GIR* genes were characterized in terms of their phylogenetic relationships, structures, chromosomal distribution and evolutionary dynamics. These *GIR* genes were revealed to be unequally distributed on 12 chromosomes in the diploid cotton genome, with no *GIR* gene detected on Ga06. The *cis*-acting elements in the protomer regions were predicted to be responsive to light, phytohormones, defense activities and stress. The transcriptomic data and qRT-PCR results revealed that most *GIR* genes were not differentially expressed between the wild-type control and the fuzzless mutant line. Moreover, 14 of 21 family genes were expressed at high levels, indicating these genes may play important roles during fiber development and fuzz formation. Furthermore, *Ga01G0231* was predominantly expressed in root samples, suggestive of a role in root hair formation rather than in fuzz initiation and development. The results of this study have enhanced our understanding of the *GIR* genes and their potential utility for improving cotton fiber through breeding.

**Keywords** *Cis*-acting elements. Family genes. *Gossypium arboreum. GaGIR.* Expression patterns

# 1. Introduction

Cotton is the most important natural fiber crop because it provides the textile industry with raw materials (Fang et al. 2017). Cotton fibers, which differentiate from the epidermal cells of the ovule, undergo the following four distinct, but overlapping, stages to reach final maturity: initiation, elongation, secondary cell wall (SCW) synthesis, and dehydrated maturation (Padmalatha et al. 2012; Wang et al. 2015; Hu et al. 2016; Sun et al. 2017). The initiation and early elongation stages are essential and highly correlated with fiber characteristics related to yield and quality, including fiber density, length, and uniformity (Rong et al. 2005; Kim et al. 2015; Zhu et al. 2018). Environmental conditions and genetic factors significantly influence these developmental stages, thereby affecting the final fiber quality (Hinchliffe et al. 2011; Gilbert et al. 2014; Liang et al. 2015; Chen et al. 2019). Therefore, the genetic mechanisms underlying fiber initiation should be investigated so that the associated genes may be used to improve cotton fiber quality via molecular breeding.

The recent development and application of high-throughput DNA sequencing technology has resulted in the publication of increasing amounts of cotton genome data, which have been compiled in databases that used as resources for predicting and screening functional genes (Wan et al. 2016; Cheng et al. 2016; Zhu et al. 2017; Wu et al. 2018; Fang et al. 2020). Many genes involved in fiber development have been identified, cloned and functionally characterized (Jiang et al. 2015; Wan et al. 2016; Hu et al. 2016; Thyssen et al. 2017; Wu et al. 2018; Patel et al. 2020; Sun et al. 2020a). For example, *GhbHLH18* is negatively associated with fiber quality because it encodes a protein that strongly binds to the E-box element of the *GhPER8* promoter to activate expression and modulate peroxidase-mediated lignin metabolism during the fiber elongating stage (Gao et al. 2019). Additionally, *GhFSN5*, which is a NAC domain transcription factor gene, is reportedly preferentially expressed during the SCW synthesis stage to negatively regulate the expression patterns of SCW-associated genes related to cellulose, xylan, lignin and several transcription factors mediating SCW formation (Sun et al. 2020b). In earlier studies, *GhPIN3a* expression is down-regulated by cytokinin, and the resulting change to the polar distribution of *GhPIN3a* disrupts the asymmetric accumulation of auxin in the ovule epidermis to inhibit cotton fiber initiation (Zeng et al. 2019; Mei and Zhang 2019). Moreover, some loci or candidate genomic regions associated with fuzz fiber initiation and formation in tetraploid cotton lines were detected by multiple combined analyses, including *GhMML3_A12* for the $N_1$ locus (Wan et al. 2016), *GhMML3_D12* for the $n_2$ locus (Zhu et al. 2018; Chen et al. 2020), the *GhMML3_A12* allele for the $n_3$ locus (Chen et al. 2020), the 411-kb genomic interval on chromosome D04 for the $n_4^t$ locus (Naoumkina et al. 2021a; Naoumkina et al. 2021b), and the 250-kb candidate region on chromosome D13 for the $N_5$ locus (Zhu et al. 2021).

Cotton fibers share many similarities with *Arabidopsis thaliana* trichomes and root hairs (Lee et al. 2007). Studies on the regulatory network involved in determining Arabidopsis cell fates have provided researchers with a framework for investigating cotton fiber initiation and elongation (Balkunde et al. 2010; Yang and Ye 2013). The multimeric complex that activates trichome initiation and development includes the R2R3 MYB protein GLABROUS1 (GL1), the WD40 repeat-containing protein

TRANSPARENT TESTA GLABRA 1 (TTG1), and the bHLH proteins GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) (Pattanaik et al. 2014; Matías-Hernández et al. 2016; Dai et al. 2016). Researchers have identified and verified the homolog of the complex-encoding genes as well as other regulatory genes in cotton (e.g., *GhMYB25-like*, *GhTTG1/4*, *GhHOX3* and *GhHD-1*) and subsequently confirmed their effects on fiber development (Walford et al. 2012; Huang et al. 2013; Shan et al. 2014; Liu et al. 2015; Wan et al. 2016; Ioannidi et al. 2016; Ding et al. 2020).

Most genes contributing to trichome development are also involved in root hair patterning, suggesting the underlying models are similar, but the resulting phenotypes are different (Matías-Hernández et al. 2016; Wei and Li 2018). The overexpression of *GLABRA2* (*GL2*) in the shoot eventually activates trichome formation. Root epidermal cells expressing *GL2* are prevented from forming root hairs (Rerie et al. 1994; Di Cristina et al. 1996; Ohashi et al. 2002; Ohashi et al. 2003; Hülskamp 2004). The GL2-interacting-repressors (GIRs) negatively regulate root hair development by interacting with GL2 and controlling the interaction network (Wu and Citovsky 2017a, b). Several recent studies indicated that *GaFzl*, which is annotated as *GIR1* in diploid cotton, adversely affects fuzz development (Du et al. 2018; Feng et al. 2019; Liu et al. 2020b; Wang et al. 2020c). However, the characteristics and underlying mechanisms of the GIRs functions and regulatory network in cotton remain unknown.

In this study, we revealed the basic features of *GaFzl* by determining the subcellular localization and analyzing its transcriptional activation of the encoded protein. Furthermore, on the basis of the published cotton genomes, we screened and identified 21, 21 and 40 *GIR* gene family members in *Gossypium arboreum*, *Gossypium raimondii* and *Gossypium hirsutum*, respectively, and then characterized their phylogenetic relationships, gene structures, chromosomal distribution and evolutionary dynamics. We also investigated the expression patterns of *GaGIR* genes and other fiber-related genes in cotton fiber. The results presented herein will be useful for clarifying the molecular mechanisms associated with the *GIR* genes in diploid cotton.

# 2. Materials and methods

## *2.1 Plant materials and bacterial strains*

The plant materials used in this study included *G. arboreum* lines DPL971 (wild-type), DPL972 (near isogenic fuzzless mutant), *G. hirsutum* lines XZ142 (wild-type), fuzzless mutant $N_1$, and fuzzless-lintlesss mutant XZ142FLM, which were obtained from Germplasm Repository of Institute of Cotton Research, Chinese Academy of Agricultural Sciences (CRI of CAAS, Anyang, Henan province, China) only for scientific research purpose. Plants were grown and self-pollinated for conservation annually in accordance with standard agronomic practices at the farm of CRI. *Nicotiana benthamiana* was grown in a 26 °C incubator with 16 h light/ 8 h dark.

*Escherichia coli DH5α* competent cells were used for gene cloning. *Agrobacterium tumefaciens* strain GV3101 was used for the transformation of tobacco (*Nicotiana benthamiana)*.

## *2.2 Subcellular localization*

Subcellular localization was performed using tobacco mesophyll cells. The CDS of *GaFzl* was cloned into binary vector pBI121-GFP to construct the GaFzl-GFP fusion protein downstream of CaMV 35S promoter. The recombinant constructs, 35S: GaFzl-GFP and the control vector pBI121- GFP, were respectively transformed into the *A. tumefaciens* strain GV3101 and inoculated into young leaves of tobacco. About 48h later, the injected tobacco leaves were stained with 1-[3-(Triethylaminio)propyl]-4-[6-[4-(diethylamino)phenyl]-1,3,5-hexatrienyl]pyridinium (FM4-64, a membrane-specific dye) and 4′6-diamidino-2-phenylindole (DAPI, a nucleus-specific dye) for 5 min, respectively, and afterward the fluorescence signals of fusion proteins in tobacco mesophyll cells were observed using a confocal fluorescence microscope (Leica). Primers used for gene cloning and vector construction are listed in Supplemental Table S4-1.

## *2.3 The transcriptional activation analysis*

The transcriptional activation assay of GaFzl was performed in yeast. The CDS of *GaFzl* from DPL971 and DPL972 were cloned into the bait vector pGBKT7 to construct the GaFzl_BD fusion protein, respectively. Both the recombinant constructs were transformed into yeast strain Y2HGold. The transformants were further cultivated on Synthetic Dropout (SD) medium (SD/-Trp, SD/-Trp/X-α-Gal, SD/-Trp/X-α-Gal /AbA and SD/-Trp/-Ade/-His/X-α-Gal /AbA).

## *2.4 Identification and characterization of GaGIR genes*

The protein sequence of AtGIR1 was downloaded from NCBI and subsequently used as a query to identify the GIR members in cotton genomes and search the hidden Markov model (HMM) profile from the Phytozome v12.1 database (https://phytozome.jgi.doe.gov/pz/portal.html#) and the PANTHER Classfication System (http://pantherdb.org/). We downloaded the latest versions of genome annotation data of Arabidopsis from the Arabidopsis Information Resource (TAIR; available online: https://www.arabidopsis.org/index.jsp). The genome data of *G. hirsutum* (version: AD1_ZJU v2.1), *G. aboreum* (version: A2_CRI) and *G. raimondii* (version: D5_JGI) were downloaded from Cotton Functional Genomic Database (CottonFGD; https://cottonfgd.org/). The HMM profile was used to screen and identify the *GIR* genes in *Gossypium* by the HMMSEARCH program from HMMER 3 software. Furthermore, BLAST was also used to confirm the family members.

## *2.5 Phylogenetic, structure and conserved motif analysis*

The conserved region of the 21 GIR protein sequences from *G. arboreum* and 6 from *A. thaliana* were used for phylogenetic tree construction. Multiple sequence alignments were conducted using ClustalX program with default parameters. The phylogenetic tree was constructed in MEGA10 by using the neighbor-jointing algorithm with the *P*-distance model and the pairwise deletion option with 1000 bootstrap replicates.

The Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn/) was employed to graphically visualize the exon-intron structure of *GIR* genes. The MEME (Multiple

Expectation Maximization for Motif Elicitation) (http://meme-suite.org/tools/meme) program was used to predict both conserved and potential motifs of GaGIR protein sequences using the parameter settings: the minimum motif width = 6, the maximum motif width = 50, and the maximum number of motifs = 10.

### *2.6 Chromosomal locations and synteny analysis*

MapChart software (http://www.earthatlas.mapchart.com/) was used to visualize the distribution of *GIR* genes on cotton chromosomes. MCScan was utilized to analyze the synteny pairs of *GIR* genes of different cotton species. The synonymous substitution (Ks) and nonsynonymous substitution (Ka) rate of *GIR* gene pairs were calculated using the DnaSP software.

### *2.7 Cis-acting elements analysis and identification of TF binding sites*

The 2000 bp sequence upstream of the start codons of *GaGIR* genes were downloaded from Cotton Functional Genomics Database (https://cottonfgd.org/). The cis-acting elements were screened and predicted using the Plant *Cis*-Acting Regulatory Element website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) database, and the diagram was generated by TBtools. The TF binding sites in the promoter regions of *GaGIR* genes were analyzed and identified using Plant Transcription Factor Database (http://planttfdb.gao-lab.org/).

### *2.8 RNA isolation and qRT-PCR analysis*

For samples collection, cotton lines DPL971, DPL972, XZ142, XZ142FLM and $N_1$ were grown in the farm of CRI of CAAS with standard field management in Anyang, Henan province (China). Fiber-bearing ovules were harvested at -1, 0, 1, 3 and 5 days post-anthesis (DPA). The ovule samples were wrapped in foil and frozen directly in liquid nitrogen. Total RNA was extracted using RNA Prep Pure Plant kit (Tiangen) according to the kit manuals and cDNA was reverse-transcribed from 1 μg total RNA using TransScript all-in-one first-strand cDNA Synthesis SuperMix (TransGen Biotech). The qRT-PCR amplifications were conducted using TransStart TOP Green qPCR SuperMix (TransGen Biotech) on ABI Prism7500 Fast Real-time PCR System (Applied Biosystems). Gene expression levels were normalized and calculated using *GaHis3* as a reference gene. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method(Livak and Schmittgen 2001). Primers used in this study were designed using NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and qPrimerDB (https://biodb.swu.edu.cn/qprimerdb/) and synthesized by Sangon Biotech company. Primer sequences are provided in Suppl. Table S1.

# 3. Results

## *3.1 Subcellular localization analysis of GaFzl*

In our previous study, *GaFzl* was fine-mapped and identified as the candidate gene controlling fuzz development in *G. arboreum*. To further explore the *GaFzl* function related to cotton fuzz initiation, its subcellular localization was analyzed using tobacco leaves (*Nicotiana benthamiana*). The GaFzl-GFP fusion protein was transiently expressed in young tobacco leaves, which were then stained using a membrane-specific dye (FM4-64) and a nucleus-specific dye (DAPI) to ascertain the localization of the GaFzl-GFP signal by confocal microscopy. Consistent with the findings of an earlier investigation by Wang et al. (Wang et al. 2020c), the strong fluorescence of the GaFzl-GFP fusion protein was detected in the membrane and nucleus, which were stained with FM4-64 (red fluorescence) and DAPI (blue fluorescence), respectively (Fig. 4-1). Accordingly, GaFzl appears to be co-localized in the nuclear and membrane of cells.

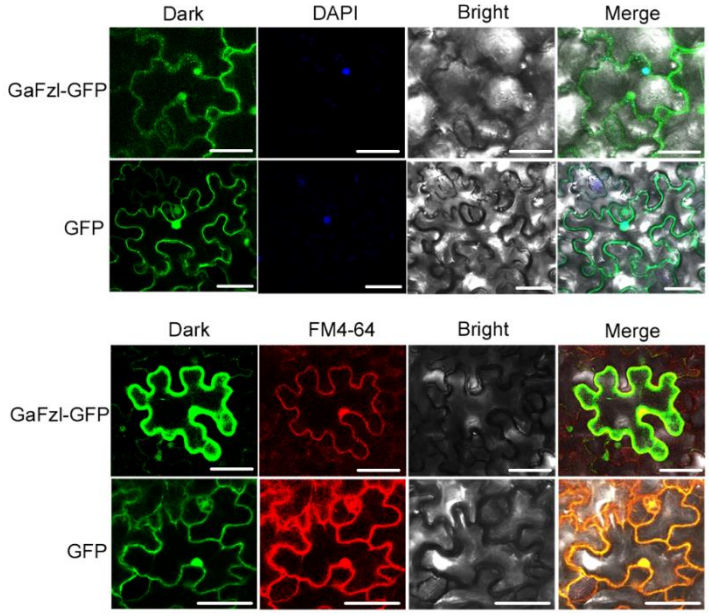



**Figure 4-1:** Subcellular localization of GaFzl. The names of constructs are shown on the left. The scale bar is 50 μm.

## *3.2 Transcriptional activation analysis of GaFzl*

Because AtGIR1 is a transcription factor that regulates root hair formation as, we assessed whether GaFzl has transcriptional activation activity by conducting an autoactivation analysis of GaFzl in yeast. The GaFzl- BD (GAL4 binding domain) fusion protein in yeast activated the expression of the reporter genes MEL1(X-α-Gal) and AUR1-C (AbA), suggesting that GaFzl has transcriptional activation activity and may function as a transcriptional regulator (Fig. 4-2). To examine whether the transcriptional activation activity of the fusion protein could be repressed, SD agar

medium (SD/-Trp/-Ade/-His/X-α-Gal /AbA) was inoculated with Y2HGold cells expressing the fusion proteins. In yeast cells, the fusion protein with GaFzl from DPL972 (fuzzless isogenic line) had a higher transcriptional activation activity level than the fusion protein with GaFzl from DPL971 (wild-type), and it activated the expression of all four reporter genes. The observed difference in the transcriptional activation activity may be related to the GaFzl sequence variation between DPL971 and DPL972.

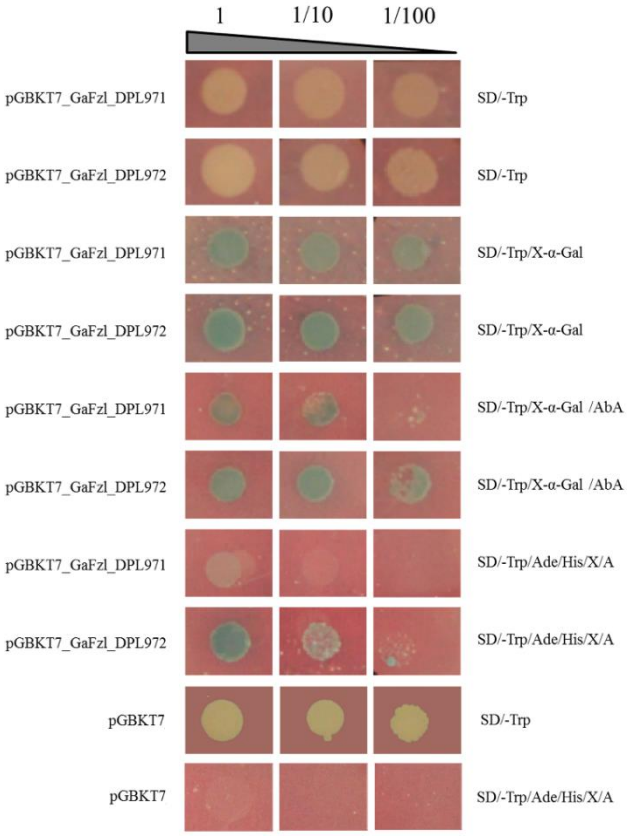



**Figure 4-2:** Transcriptional autoactivation of GaFzl was detected in yeast. GaFzl amplified from DPL971 and DPL972 were inserted into pGBKT7 to construct expression vectors indicated on the left; Y2Hgold stains containing the expression constructs were diluted and inoculated on the SD medium indicated on the right. The number of 1, 1/10 and 1/100 indicate yeast stain without dilution, 10-fold dilution and 100-fold dilution.

## *3.3 Genome-wide identification of GIR genes in G. arboreum*

To thoroughly characterize the *GIR* gene family members in cotton, the whole-genome sequences of cotton species (*G. arboreum* (version: A2_CRI), *G. hirsutum* (version: AD1_ZJU v2.1) and *G. raimondii* (version: D5_JGI)) were analyzed to

identify *GIR* genes. The Arabidopsis GIR1 protein sequence was used as the query to screen three cotton reference genomes for candidate GIR proteins, with an E-value threshold of 0.01. On the basis of the hidden Markov model (HMM) of PTHR33177, the HMMER 3 program was used to identify 21 GIR proteins in *G. arboreum*. Additionally, 21, 40 and 6 GIR proteins were detected in *G. raimondii*, *G. hirsutum*, and *A. thaliana*, respectively (Suppl. Table S4-2, S4-3). Specific information of *GaGIR* genes, such as gene ID, chromosomal location, protein size (AA), molecular weight (kDa) were listed in Table 1. The encoded proteins comprised 87 amino acids (Ga08G0121) to 243 amino acids (Ga07G1259), suggesting these are small proteins that may function as adapters or small molecular peptides that interact with other proteins to perform their functions.

**Table 4-1**: Genome-wide identification of *GIR* family genes in *G. arboreum*

| No. | Gene ID | Chr. | Strand | Start | End | AA | kDa | Isoelectric point | Grand average of hydropathy |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Ga01G1554 | 1 | - | 58,027,883 | 58,028,242 | 119 | 12.989 | 8.209 | -0.516 |
| 2 | Ga02G1535 | 2 | - | 95,660,733 | 95,661,080 | 115 | 12.651 | 8.785 | -0.463 |
| 3 | Ga11G0145 | 11 | + | 1,272,908 | 1,273,225 | 105 | 11.146 | 4.552 | -0.242 |
| 4 | Ga13G0044 | 13 | - | 460,349 | 460,690 | 113 | 12.482 | 6.441 | -0.581 |
| 5 | Ga05G0849 | 5 | - | 7,401,337 | 7,401,666 | 109 | 11.739 | 8.495 | -0.341 |
| 6 | Ga07G1212 | 7 | + | 19,410,537 | 19,410,872 | 111 | 12.345 | 5.828 | -0.51 |
| 7 | Ga09G1549 | 9 | - | 72,639,829 | 72,641,385 | 113 | 12.266 | 8.114 | -0.181 |
| 8 | Ga08G0121 | 8 | + | 905,296 | 905,559 | 87 | 9.834 | 4.041 | -0.28 |
| 9 | Ga10G2569 | 10 | - | 124,689,192 | 124,690,513 | 237 | 25.162 | 9.781 | -0.435 |
| 10 | Ga07G1259 | 7 | - | 20,894,052 | 20,895,016 | 243 | 26.101 | 8.577 | -0.472 |
| 11 | Ga13G1545 | 13 | - | 99,450,521 | 99,452,131 | 313 | 34.876 | 10.271 | -0.493 |
| 12 | Ga04G1030 | 4 | + | 41,854,005 | 41,854,336 | 96 | 10.241 | 8.332 | -0.406 |
| 13 | Ga13G0100 | 13 | - | 1,004,877 | 1,005,634 | 175 | 19.189 | 6.118 | -0.714 |
| 14 | Ga13G0104 | 13 | - | 1,077,386 | 1,078,137 | 171 | 19.158 | 5.058 | -0.494 |
| 15 | Ga13G0101 | 13 | - | 1,013,106 | 1,013,903 | 184 | 20.43 | 4.635 | -0.719 |
| 16 | Ga01G0231 | 1 | - | 1,753,020 | 1,753,839 | 181 | 19.811 | 4.318 | -0.872 |
| 17 | Ga12G1438 | 12 | - | 20,700,695 | 20,702,020 | 203 | 22.877 | 4.237 | -1.005 |
| 18 | Ga13G0103 | 13 | - | 1,075,955 | 1,076,629 | 135 | 15.119 | 8.121 | -0.479 |
| 19 | Ga05G2409 | 5 | - | 22,655,865 | 22,658,507 | 172 | 19.27 | 9.172 | -0.653 |
| 20 | Ga03G0898 | 3 | + | 17,560,035 | 17,560,783 | 209 | 23.684 | 7.319 | -0.669 |
| 21 | Ga03G0159 | 3 | - | 1,384,990 | 1,385,763 | 193 | 21.889 | 6.495 | -0.594 |

## *3.4 Phylogenetic, gene structure, and motif analyses of GIR proteins in G. arboreum*

To classify the GIR proteins and clarify their evolutionary relationships, a multiple sequence alignment analysis (Suppl. Fig. S4-1) was performed using 21 and 6 homologous protein sequences from *G. arboreum* and *A. thaliana*, respectively. Additionally, an unrooted phylogenetic tree was constructed according to the neighbor-joining algorithm of the MEGA10.0 software, with 1000 bootstraps replicates. In the constructed phylogenetic tree (Fig. 4-3), the 27 proteins were divided into three groups. Within the GaGIR family, 7 members were clustered in Group 1, 4 members belonged to Group 2 and 10 members were included in Group 3. Group 1 included GaGIR1 and two AtGIR proteins (AtGIR1 and AtGIR2), implying that these members may be associated with similar functions.

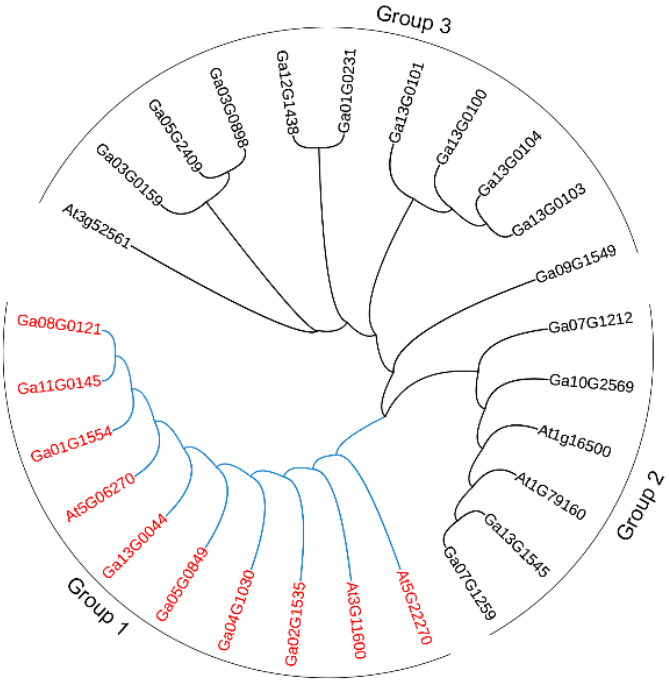



**Figure 4-3:** Phylogenetic analysis of the *GIR* gene family. The phylogenetic tree was constructed using the full length GIR protein amino acid sequences from *G. arboreum* and *A. thaliana*. The genes highlighted in red fonts and blue lines were classified and named as Group 1. At, *Arabidopsis thaliana*; Ga, *Gossypium arboreum*.

Gene structures and organization may provide additional evidence of the evolutionary relationships among *GIR* family members. A comparative analysis of gene structures and motifs detected 0-2 introns in the *GaGIR* genes. More specifically, seven genes had no intron, whereas one gene had two introns, and the remaining genes comprised only one intron (Fig. 4-4). Ten conserved motifs in the GaGIR proteins were identified using the MEME online software (i.e., motifs 1 to 10). Almost all

GaGIR proteins had motifs 1, 2 and 3, implying these motifs may function as a highly conserved domain. Most of the Group 1 GaGIR proteins contained motif 4. In contrast, most of the Group 2 proteins had motifs 5, 6, 8, 9, and 10, whereas the Group 3 proteins had motifs 3, 5, and 7.

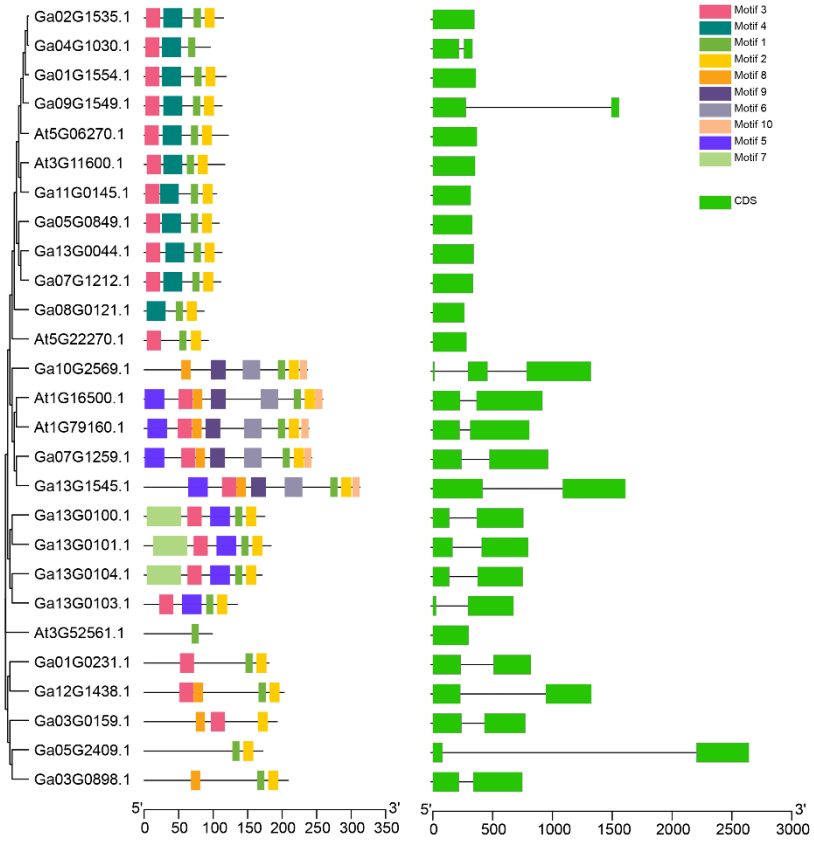



**Figure 4-4:** Phylogenetic relationship, conserved motif and gene structure analysis of GaGIR proteins. The left part shows the conserved motifs involved in GaGIR proteins. The number and order of motifs in each GaGIR proteins are presented. Gene structure (exon-intron organization) of *GaGIRs* is displayed on the right. Exons and introns are represented by green boxes and black lines, respectively. The scale bar is shown at the bottom.

## *3.5 Chromosomal distribution of GaGIR genes*

The *GaGIR* genes were mapped on cotton chromosomes to gain new insights into the organization of *GaGIR* genes in the cotton genome. Using the available *G. arboreum* genome as a reference, we determined that the 21 *GaGIR* genes were unequally distributed among the chromosomes (Fig. 4-5). Seven chromosomes (chr. 2, 4, 8, 9, 10 and 12) contained only one *GaGIR* gene, whereas four chromosomes (chr. 1, 3, 5 and 7) had two *GaGIR* genes. Chromosome 13 included six *GaGIR* genes,

four of which were tandemly distributed. However, the chromosomal distribution of the *GIR* genes varied among *G. arboreum*, *G. raimondii*, and *G. hirsutum* (At and Dt-subgenomes), indicative of gain and loss events during whole-genome duplications (Suppl. Fig. S4-2, S4-3, S4-4). To clarify the collinearity of the *GIR* gene families between two diploid cotton ancestors and *G. hirsutum*, linked gene pairs were identified using MCScan. A total of 34 collinear gene pairs were identified between *G. hirsutum* and *G. arboreum*, 18 of which involved the A-subgenome of *G. hirsutum* (Fig. 4-6). Additionally, 39 collinear gene pairs were detected between *G. hirsutum* and *G. raimondii*, with 20 genes from the D-subgenome. To elucidate the evolutionary dynamics and selection pressures, we calculated the nonsynonymous (Ka) and synonymous (Ks) substitution rates of the *GIR* gene pairs. Most of the Ka/Ks ratios were less than 1.0 in the intergenomic and intragenomic analyses, implying that most *GIR* genes underwent a purifying selection in the diploid and allotetraploid cotton species (Suppl. Table S4-4, S4-5, S4-6).

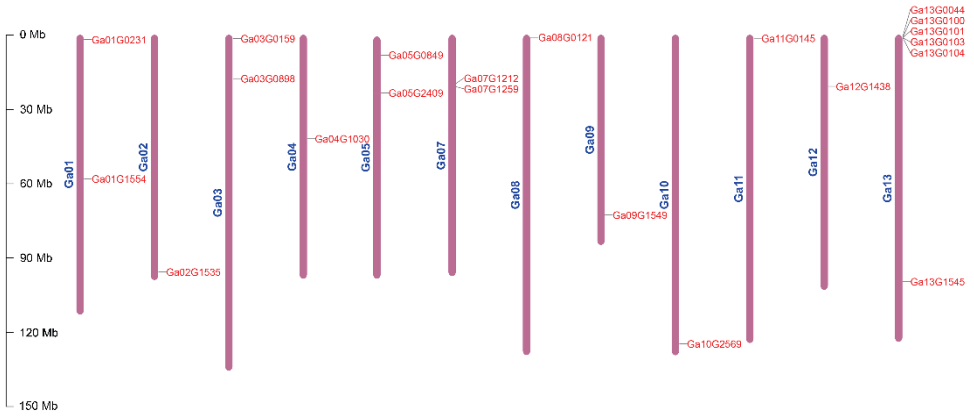



**Figure 4-5:** Chromosome distribution of *GIR* family genes in *G.arboreum*. The chromosome name is presented on the left side of the graph, and the gene ID is on the right. The vertical scale shows the size of chromosomes and black lines indicate the corresponding position of the genes. The scale bar indicates the chromosome length in base pair (bp).

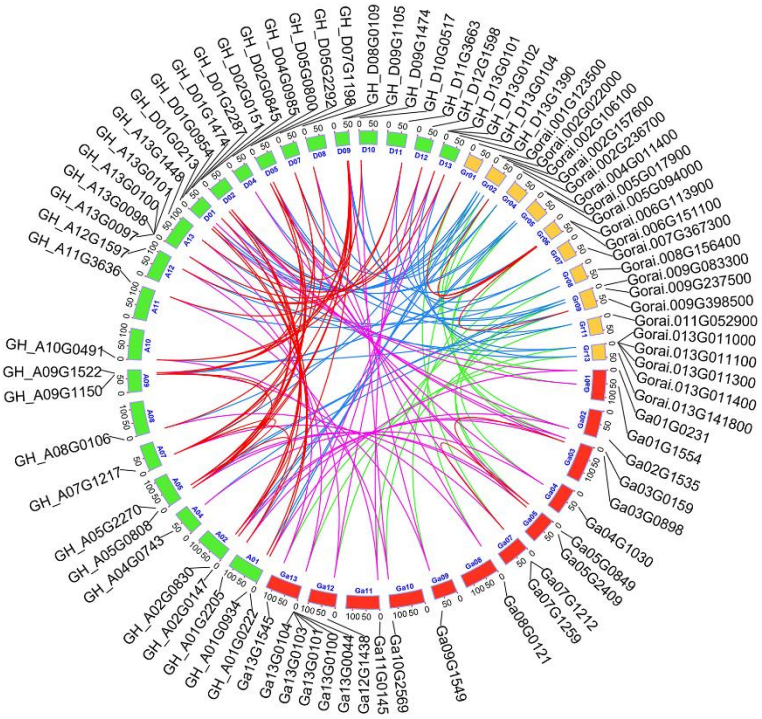



**Figure 4-6:** Collinearity analyses of *GIR* genes among *G. hirsutum*, *G. arboreum* and *G. raimondii*. Red lines indicate the intergenomic collinearity; other three lines represent the intragenomic collinearity.

## *3.6 Cis-acting element analysis*

Because most *GIRs* remain unannotated, the putative functions of *GIRs* were first assessed by identifying the *cis*-acting elements in the 2-kb promoter region of the corresponding genes. In addition to the general elements (i.e., TATA-box and CAAT-box), the detected *cis*-acting elements were mainly enhancer activity elements and elements responsive to light, phytohormones, defense activities and abiotic stress (i.e., drought, low-temperature) (Fig. 4-7). Interestingly, 20 of the 21 *GaGIR* genes had *cis*-acting elements responsive to 3 - 6 phytohormones (ABA, followed by ethylene, MeJA, SA, GA and auxin). An analysis of the transcription factor-binding site in the promoter regions indicated the binding sites of ERF, BBR-BPC, MYB, AP2/B3, DOF and C2H2 transcription factors were extensively distributed in the *GaGIR* promoters. These results suggest that *GaGIR* genes likely contribute to plant growth and development through various signal transductions pathways, especially those related to phytohormone, light, and abiotic stress responses.

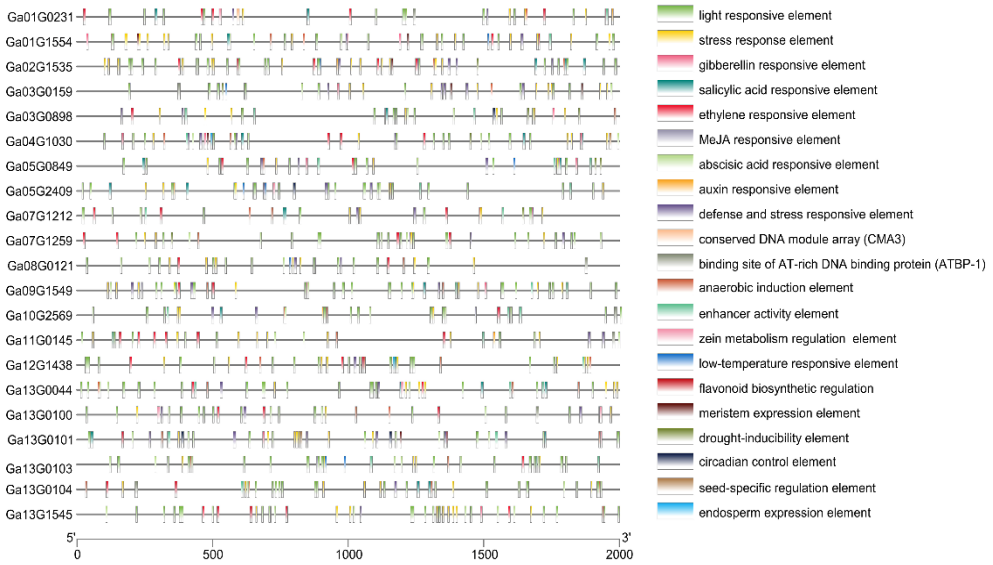



**Figure 4-7:** Predicted *cis*-acting elements in promoter regions of *GaGIRs*.

## *3.7 Expression patterns of GaGIR genes at different stages of fiber development*

To investigate the potential functions of *GIR* genes in *G. arboreum*, we analyzed their expression patterns using the transcriptome data for different stages of fiber development (1, 3 and 5 DPA). According to their expression levels, the *GaGIR* genes were roughly classified as strongly expressed or barely expressed. The heatmap representation of expression revealed that the *GaGIRs* genes were not significantly differentially expressed between DPL971 (wild-type) and DPL972 (fuzzless isogenic line), with the exception of *GaFzl* (Fig. 4-8). Consistent with previous studies, the *GaFzl* expression level was much higher at 1, 3 and 5 DPA in DPL972 than in DPL971, implying that the up-regulated expression of *GaFzl* likely suppresses fuzz development. This gene is reportedly significantly up-regulated by a 6.2 kb insertion in fuzzless mutant and thus leads to a fuzzless phenotype (Wang et al. 2020c). Hence, other genes which normally expressed at low levels in wild-type cotton might also alter fuzz development if they are expressed at higher levels. To verify the accuracy and reliability of the RNA-seq results, we also examined the relative expression levels of *GIR* family genes in ovules at different stages by qRT-PCR. The results were consistent with the RNA-seq data, confirming the transcriptome data were reliable (Fig. 4-9). Previous studies proved AtGIRs are involved in Arabidopsis root hair formation (Wu and Citovsky 2017a, b). To explore the functions of *GaGIR* genes, we examined the root phenotypes of two diploid cotton species and then compared the *GaGIR* expression patterns between the two materials by qRT-PCR. There were no obvious differences in the root hair phenotypes (Suppl. Fig. S4-5) and gene expression levels (Fig. 4-9). Moreover, most of the *GIR* genes were expressed at much lower levels in the root than in the fiber. However, *Ga01G0231* was highly expressed in the

root (Fig. 4-9), indicative of an important role in root hair formation and development.

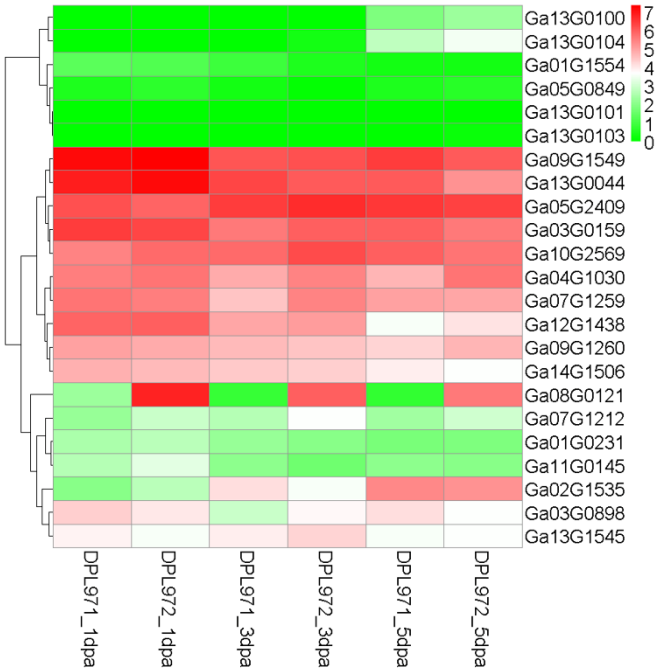



**Figure 4-8:** RNA-seq data heat map of *GaGIR* gene expression levels at different stages of fiber development. The differences in gene expression are shown in different colors. DPA, day post anthesis.

## *3.8 Regulatory relationship between GaFzl and other fiber-associated genes*

To further investigate the potential regulatory network of *GaGIR* genes involved in fuzz development, we analyzed the expression level of *GaFzl* in various fiber mutants as well as the expression of other fiber-related genes in *G. arboreum* by qRT-PCR. Because gene expression levels may be influenced by allele dosage effects in polyploids, we designed sequence-specific primers to examine the expression level of the *GaFzl* homolog (*GhFzl*) in tetraploid cotton lines. Unlike the expression pattern in diploid cotton, *GhFzl_A* and *GhFzl_D* were expressed at very low levels (Fig. 4-10), with no major differences among wild-type XZ142, fuzzless mutant $N_1$, and fiberless mutant XZ142FLM. Additionally, the genes encoding well-known fiber-related regulatory factors were similarly expressed in DPL971 and DPL972 (Fig. 4-11). For example, the *MML3*, *TTG1*, and *GL2* expression levels were unaffected by changes to *GaFzl* expression, indicating that the altered expression of *GaFzl* had no effect on the regulatory network of the MIXTA-WD40 complex and other fiber-related genes.

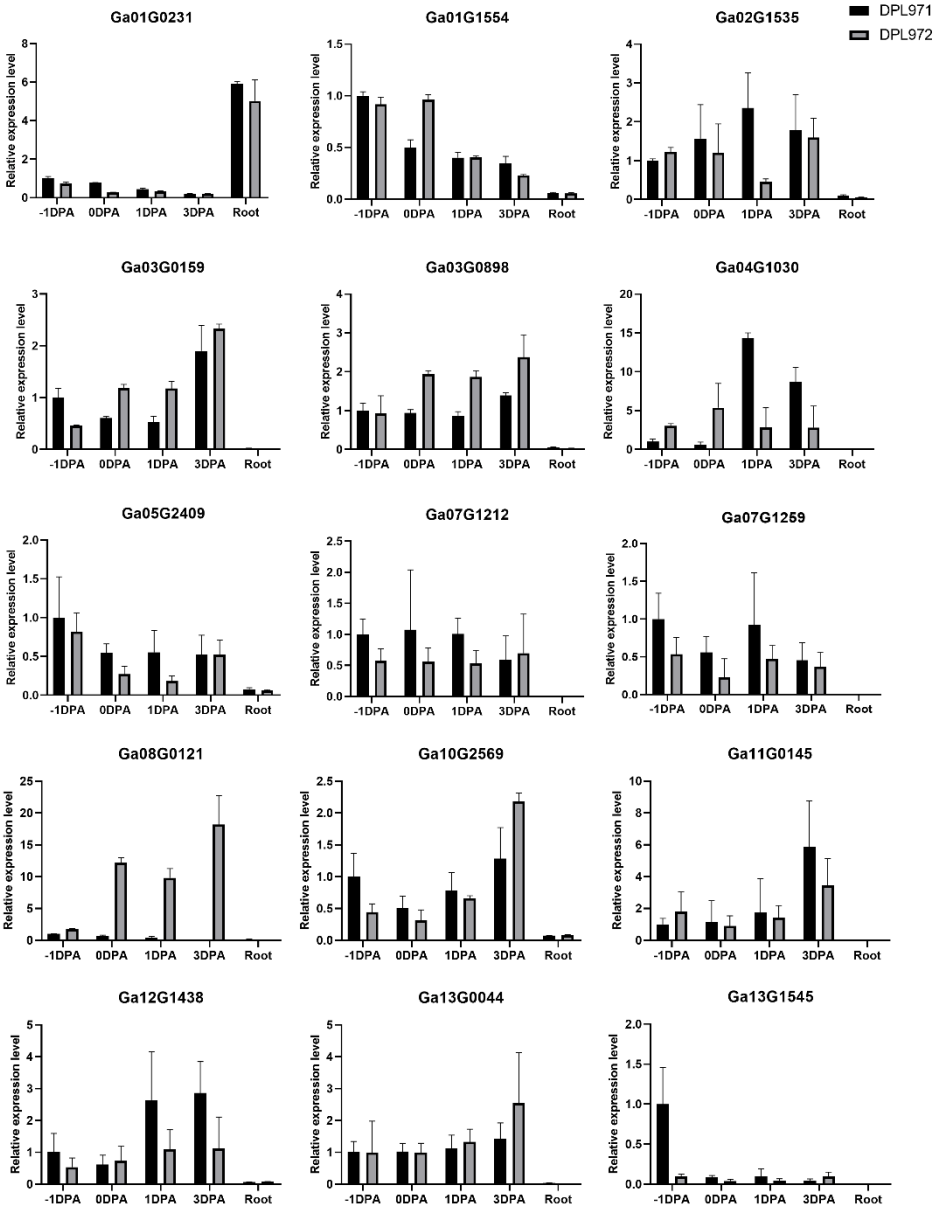



**Figure 4-9:** Relative expression levels of *GaGIR* genes in *G. arboreum* wild-type DPL971 and fuzzless isogenic mutant DPL972 at different stages of cotton fiber development. *GhHis3* was applied as the internal control. The expression value of fiber samples in DPL971 at -1 DPA was set as 1. Data are presented as mean ± standard deviation (*n=3*).

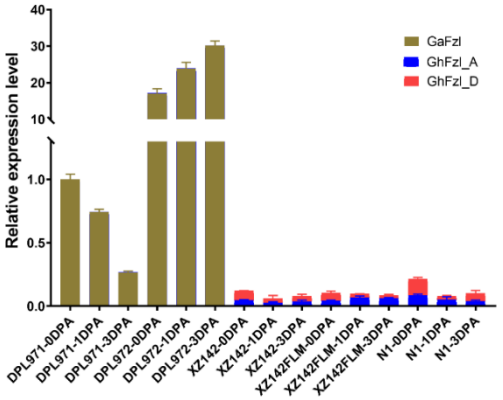



**Figure 4-10:** Expression patterns of *GaFzl* in *G. arboreum* DPL971 and DPL972, in *G. hirsutum* $N_1$, XZ142 and XZ142FLM. *GhHis3* was applied as the internal control. The expression value of fiber samples in DPL971 at -1 DPA was set as 1. Data are presented as mean ± standard deviation (*n=3*).

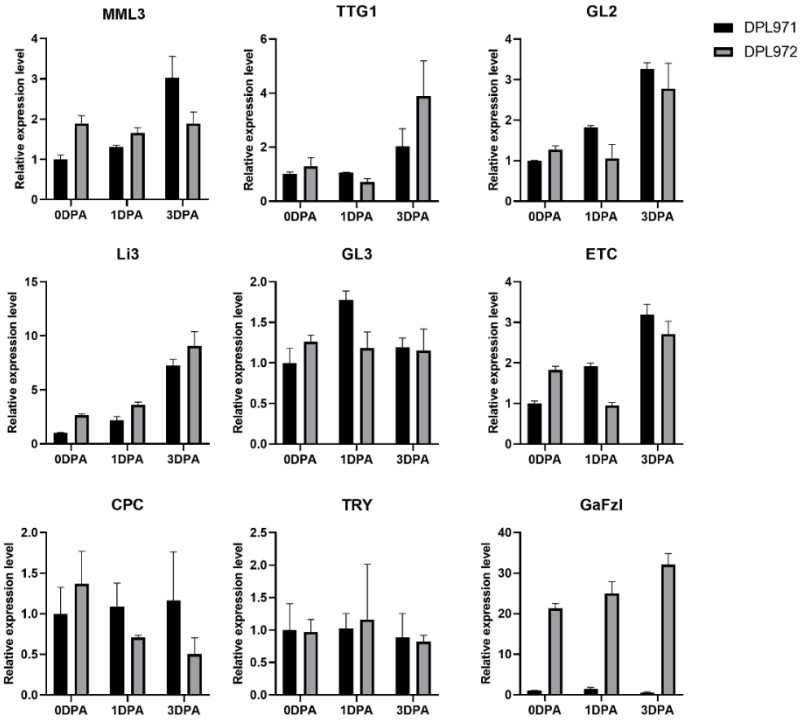



**Figure 4-11:** Expression analysis of several core genes in *G. arboreum* DPL971 and DPL972 in upland cotton fiber development. *GhHis3* was applied as the internal control. The expression value of fiber samples in DPL971 at -1 DPA was set as 1. Data are presented as mean ± standard deviation (*n=3*).

# 4. Discussion

Arabidopsis is a model plant system for studying cell fate determination and differentiation (Balkunde et al. 2010; Yang and Ye 2013). Because of the considerable similarity in the underlying process, studies on Arabidopsis trichome and root hair formation may provide useful insights into cotton fiber initiation and elongation (Lee et al. 2007; Ding et al. 2014; Wan et al. 2014; Wang et al. 2019). The *GIR* genes have been identified in diverse monocotyledonous and dicotyledonous plant species. The *GIR* genes in Arabidopsis are involved in the GL2-mediated control of root hair development (Wu and Citovsky 2017a). In tomato, *SlCycB2*, which encodes a GIR1/2 homolog reportedly interacts with Wo, which is GL2 homolog, to regulate trichome formation and development (Yang et al. 2011). Several recent studies demonstrated that *GaFzl,* annotated as *GaGIR1*, is closely associated with the fuzzless phenotype of diploid cotton. It encodes a negative regulator of trichome formation and fuzz development, but it does not affect root hair formation (Du et al. 2018; Feng et al. 2019; Liu et al. 2020b; Wang et al. 2020c). These studies suggest that the functions of these homologs may have diverged slightly as the different species evolved. However, very few of the *GIR* genes in cotton have been identified and functionally annotated. Thus, it is important that these genes in cotton are identified so their expression patterns and functions related to cotton fiber development may be determined.

On the basis of a whole-genome analysis, we identified 21, 21, and 40 *GIR* family genes in *G. arboreum*, *G. raimondii* and *G. hirsutum*, respectively. The *GaGIRs* genes, which were divided into three groups, were unevenly distributed on 12 of 13 chromosomes. Additionally, the encoded proteins varied substantially in terms of amino acid sequence length and molecular weight. Analyses of the phylogenetic relationships, structures, and motifs revealed that most of the GIR proteins in the same subgroup were similar regarding the organization of exons and conserved motifs, whereas obvious differences were detected between subgroups. These differences may reflect some functional diversification during evolution. Moreover, some conserved motifs were identified in all GIR proteins, whereas others were subgroup-specific, suggesting that motif diversity might be related to functional diversity. The gain or loss of key motifs as species evolved may have resulted in changes to protein functions that altered plant development. Phytohormones and abiotic stresses might affect fiber development through signal transduction pathways (Chen et al. 2019; He et al. 2019; Cheng et al. 2020; Wang et al. 2020a; Wang et al. 2020b; Zhang et al. 2020; Tian and Zhang 2021; Wu et al. 2021). We identified numerous *cis*-acting elements responsive to light, phytohormones, defense processes, and stress as well as multiple transcription factor-binding sites upstream of transcriptional start sites. Accordingly, GIR proteins may participate in signal transduction pathways that help regulate fiber and plant development.

We compared the *GaGIRs* expression patterns between wild-type and fuzzless mutant *G. arboreum* at different stages of fiber development. Some of the genes were highly expressed at multiple fiber developmental stages, indicating they may positively regulate fiber formation. In contrast, some *GaGIR* genes were expressed at relatively low levels, suggesting that they are not directly involved in fiber development or they may negatively regulate fiber formation. There were no obvious

differences in the expression of most of these genes between the wild-type and mutant cotton. Interestingly, *GaFzl* was significantly expressed in the fuzzless mutant but not in the wild-type control, during the fiber and fuzz initiation stages (0, 1, 3 DPA) (Fig. 4-10). This is in accordance with the results of previous studies (Feng et al. 2019; Liu et al. 2020a). This gene was annotated as *GaGIR1*, and its upregulated expression is associated with the fuzzless phenotype in *G. arboreum*. The significant difference in the expression of this gene between the wild-type and mutant cotton plants may be caused by a ~6.2-kb insertion (Copy Number Variation) in the upstream region. Our analyses of phylogenetic relationships and gene expression patterns indicated that five Group 1 members that were clustered with *GaFzl* were expressed at low levels during the fiber and fuzz initiation stages. We hypothesize that overexpressing the genes classified in the same group as *GaFzl* with the same expression patterns may similarly affect the fuzz initiation stage, ultimately resulting in the fuzzless phenotype. Because the functions of most of these genes are unknown, they will need to be more thoroughly investigated in future studies.

There were no obvious differences in *GhFzl* expression among *G. hirsutum* XZ142, the fuzzless mutant $N_1$, and the lintless-fuzzless mutant XZ142FLM, which is in contrast to the corresponding expression patterns in *G. arboreum*. Furthermore, there was no significant difference between *G. arboreum* DPL971 and DPL972 regarding the expression patterns of the core genes involved in fiber and fuzz development (i.e., *MML3*, *TTG1*, and *CPC*). Thus, the mechanism regulating *GaFzl* expression in *G. arboreum* may not be associated with the well-known MYB-bHLH-WD40 complex involved in the fiber/fuzz initiation and development of *G. hirsutum*. This finding may lead to the development of new ways to specifically regulate fiber/fuzz formation in the initial steps of this process in *G. arboreum*.

In previous studies, *GaFzl* was identified as a candidate gene controlling fuzz initiation and development in *G. arboreum*. Liu et al. suggested that the variability in the sequences of the *GaGIR1* haplotypes in diverse fuzzless mutants may be related to the changes in expression patterns or gene functions (Liu et al. 2020b). While Wang et al. identified an enhancer as the essential element for controlling gene expression and fuzz development (Wang et al. 2020c). Interestingly, to explore the potential function of *GaFzl* during fuzz initiation, we conducted a transcriptional activation analysis in yeast, which demonstrated that GaFzl has strong transcriptional activation activity and may act as a transcriptional activator. However, the fusion proteins constructed from two materials revealed minor differences in the transcriptional activation activity in yeast cells. Because of the differences in the *GaFzl* coding sequence, we speculated that sequence variations between two parental lines may be responsible for the differences in transcriptional activation activities or even gene functions. This possibility will need to be experimentally verified in future studies.

**Abbreviations**

GIRs: GL2-interacting-repressors; SCW: Secondary cell wall; DPA: Day post anthesis; GFP: Green fluorescence protein; SD: Synthetic Dropout.

**Author contribution statement** Research conception and design, GLS and JY; Data collection, XXF; Data analysis, XXF, HLC, DYZ, YPZ, QLW, LML and SYL;

Research management, GLS; Writing – initial draft, XXF; Writing – revised manuscript, JY. All authors read and approved the final manuscript.

**Funding**

This work was funded by grants from the National Natural Science Foundation of China (No. 31621005), the National Natural Science Foundation of China (No. 31901581), the National Key R & D Plan of China (No. 2018YFD0100402) and the United States Department of Agriculture - Agricultural Research Service (USDA-ARS Project No. 3091-21000044-00D). The funders had no role in the experimental design, data collection and analysis, or in writing work.

**Acknowledgements**

The authors thank Lima Soares Emanoella from Plant Genetics Laboratory of Gembloux Agro Bio-Tech, University of Liège for kindly revising the draft.

**Declarations**

**Conflict of interest**

The authors declare that they have no conflicts of interest.

**Data availability statements**

All related datasets supporting the results of this study are available within the manuscript and its supplementary files.

# References

Balkunde R, Pesch M, Hülskamp M (2010) Trichome patterning in *Arabidopsis thaliana*: From genetic to molecular models. Curr Top Dev Biol 91: 299-321

Chen Y, Chen B, Wang H, Hu W, Wang S, Zhou Z (2019) Combined elevated temperature and soil waterlogging stresses limit fibre biomass accumulation and fibre quality formation by disrupting protein activity during cotton fibre development. Funct Plant Biol 46 (8):715-724

Chen W, Li Y, Zhu S, Fang S, Zhao L, Guo Y, Wang J, Yuan L, Lu Y, Liu F, Yao J, Zhang Y (2020) A retrotransposon insertion in *GhMML3_D12* is likely responsible for the lintless locus *li3* of tetraploid cotton. Front Plant Sci 11:593679. doi:10.3389/fpls.2020.593679

Cheng G, Zhang L, Wei H, Wang H, Lu J, Yu S (2020) Transcriptome analysis reveals a gene expression pattern associated with fuzz fiber initiation induced by high temperature in *Gossypium barbadense*. Genes (Basel) 11 (9):1066

Cheng H, Lu C, John ZY, Zou C, Zhang Y, Wang Q, Huang J, Feng X, Jiang P, Yang W (2016) Fine mapping and candidate gene analysis of the dominant glandless gene $Gl_2^e$ in cotton (*Gossypium* spp.). Theor Appl Genet 129 (7):1347-1355

Dai X, Zhou L, Zhang W, Cai L, Guo H, Tian H, Schiefelbein J, Wang S (2016) A single amino acid substitution in the R3 domain of GLABRA1 leads to inhibition of trichome formation in Arabidopsis without affecting its interaction with GLABRA3. Plant Cell Environ 39 (4):897-907

Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G (1996) The Arabidopsis Athb-10 (GLABRA2) is an HD-Zip protein required for regulation

of root hair development. Plant J 10 (3):393-402

Ding M, Cao Y, He S, Sun J, Dai H, Zhang H, Sun C, Jiang Y, Paterson AH, Rong J (2020) *GaHD1*, a candidate gene for the *Gossypium arboreum* SMA-4 mutant, promotes trichome and fiber initiation by cellular $H_2O_2$ and $Ca^{2+}$ signals. Plant Mol Biol 103(4-5):409-423

Ding M, Jiang Y, Cao Y, Lin L, He S, Zhou W, Rong J (2014) Gene expression profile analysis of Ligon lintless-1 (Li1) mutant reveals important genes and pathways in cotton leaf and fiber development. Gene 535 (2):273-285

Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M, Jia Y, Pan Z, Gong W, Liu Z, Zhu H, Ma L, Liu F, Yang D, Wang F, Fan W, Gong Q, Peng Z, Wang L, Wang X, Xu S, Shang H, Lu C, Zheng H, Huang S, Lin T, Zhu Y, Li F (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat Genet 50 (6):796-802

Fang DD, Naoumkina M, Thyssen GN, Bechere E, Li P, Florane CB (2020) An EMS-induced mutation in a tetratricopeptide repeat-like superfamily protein gene (*Ghir_A12G008870*) on chromosome A12 is responsible for the *li y* short fiber phenotype in cotton. Theor Appl Genet 133 (1):271-282

Fang L, Gong H, Hu Y, Liu C, Zhou B, Huang T, Wang Y, Chen S, Fang DD, Du X (2017) Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. Genome Biol 18 (1):33

Feng X, Cheng H, Zuo D, Zhang Y, Wang Q, Liu K, Ashraf J, Yang Q, Li S, Chen X (2019a) Fine mapping and identification of the fuzzless gene *GaFzl* in DPL972 (*Gossypium arboreum*). Theor Appl Genet 132 (8):2169-2179

Gao Z, Sun W, Wang J, Zhao C, Zuo K (2019) *GhbHLH18* negatively regulates fiber strength and length by enhancing lignin biosynthesis in cotton fibers. Plant Sci 286:7-16

Gilbert MK, Kim HJ, Tang Y, Naoumkina M, Fang DD (2014) Comparative transcriptome analysis of short fiber mutants Ligon-lintless 1 and 2 reveals common mechanisms pertinent to fiber elongation in cotton (*Gossypium hirsutum* L.). PLoS One 9 (4)

He P, Zhang Y, Liu H, Yuan Y, Wang C, Yu J, Xiao G (2019) Comprehensive analysis of WOX genes uncovers that *WOX13* is involved in phytohormone-mediated fiber development in cotton. BMC Plant Biol 19 (1):312

Hinchliffe DJ, Turley RB, Naoumkina M, Kim HJ, Tang Y, Yeater KM, Li P, Fang DD (2011) A combined functional and structural genomics approach identified an EST-SSR marker with complete linkage to the Ligon lintless-2 genetic locus in cotton (*Gossypium hirsutum* L.). BMC Genom 12 (1):445

Hu H, He X, Tu L, Zhu L, Zhu S, Ge Z, Zhang X (2016) GhJAZ2 negatively regulates cotton fiber initiation by interacting with the R2R3-MYB transcription factor GhMYB25-like. Plant J 88 (6):921-935

Huang Y, Liu X, Tang K, Zuo K (2013) Functional analysis of the seed coat-specific gene *GbMYB2* from cotton. Plant Physiol Bioch 73:16-22

Hülskamp M (2004) Plant trichomes: a model for cell differentiation. Nat Rev Mol

Cell Biol 5 (6):471-480

Ioannidi E, Rigas S, Tsitsekian D, Daras G, Alatzas A, Makris A, Tanou G, Argiriou A, Alexandrou D, Poethig S (2016) Trichome patterning control involves TTG1 interaction with SPL transcription factors. Plant Mol Biol 92 (6):675-687

Jiang Y, Ding M, Cao Y, Yang F, Zhang H, He S, Dai H, Hao H, Rong J (2015) Genetic fine mapping and candidate gene analysis of the *Gossypium hirsutum* Ligon lintless-1 (Li1) mutant on chromosome 22 (D). Mol Genet Genomics 290 (6):2199-2211

Kim HJ, Hinchliffe DJ, Triplett BA, Chen ZJ, Stelly DM, Yeater KM, Moon HS, Gilbert MK, Thyssen GN, Turley RB (2015) Phytohormonal networks promote differentiation of fiber initials on pre-anthesis cotton ovules grown in vitro and in planta. PLoS One 10 (4)

Lee JJ, Woodward AW, Chen ZJ (2007) Gene expression changes and early events in cotton fibre development. Ann Bot 100 (7):1391-1401

Liang W, Fang L, Xiang D, Hu Y, Feng H, Chang L, Zhang T (2015) Transcriptome analysis of short fiber mutant Ligon lintless-1 (Li1) reveals critical genes and key pathways in cotton fiber elongation and leaf development. PLoS One 10 (11)

Liu B, Zhu Y, Zhang T (2015) The R3-MYB gene *GhCPC* negatively regulates cotton fiber elongation. PloS One 10 (2)

Liu X, Moncuquet P, Zhu Q-H, Stiller W, Zhang Z, Wilson I (2020a) Genetic identification and transcriptome analysis of lintless and fuzzless traits in *Gossypium arboreum* L. Int J Mol Sci 21 (5):1675

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25 (4):402-408

Matías-Hernández L, Aguilar-Jaramillo AE, Cigliano RA, Sanseverino W, Pelaz S (2016) Flowering and trichome development share hormonal and transcription factor regulation. J Exp Bot 67 (5):1209-1219

Mei G, Zhang Z (2019) Optimization of polar distribution of *GhPIN3a* in the ovule epidermis improves cotton fiber development. J Exp Bot 70 (12):3021

Naoumkina M, Thyssen GN, Fang DD, Bechere E, Li P, Florane CB (2021a) Mapping-by-sequencing the locus of EMS-induced mutation responsible for tufted-fuzzless seed phenotype in cotton. Mol Genet Genomics 296 (5):1041-1049

Naoumkina M, Thyssen GN, Fang DD, Li P, Florane CB (2021b) Elucidation of sequence polymorphism in fuzzless-seed cotton lines. Mol Genet Genomics 296 (1):193-206

Ohashi Y, Oka A, Rodrigues-Pousada R, Possenti M, Ruberti I, Morelli G, Aoyama T (2003) Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. Science 300 (5624):1427-1430

Ohashi Y, Oka A, Ruberti I, Morelli G, Aoyama T (2002) Entopically additive expression of *GLABRA2* alters the frequency and spacing of trichome initiation. Plant J 29 (3):359-369

Padmalatha KV, Patil DP, Kumar K, Dhandapani G, Kanakachari M, Phanindra ML, Kumar S, Mohan T, Jain N, Prakash AH (2012) Functional genomics of fuzzless-

lintless mutant of *Gossypium hirsutum* L. cv. MCU5 reveal key genes and pathways involved in cotton fibre initiation and elongation. BMC Genom 13 (1):624

Patel JD, Huang X, Lin L, Das S, Chandnani R, Khanal S, Adhikari J, Shehzad T, Guo H, Roy-Zokan EM (2020) The Ligon lintless-2 short fiber mutation is located within a terminal deletion of chromosome 18 in cotton. Plant Physiol 183(1):277-288

Pattanaik S, Patra B, Singh SK, Yuan L (2014) An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Front Plant Sci 5:259

Rerie WG, Feldmann KA, Marks MD (1994) The *GLABRA2* gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. Gene Dev 8 (12):1388-1399

Rong J, Pierce GJ, Waghmare VN, Rogers CJ, Desai A, Chee PW, May OL, Gannaway JR, Wendel JF, Wilkins TA (2005) Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theor Appl Genet 111 (6):1137-1146

Shan C, Shangguan X, Zhao B, Zhang X, Chao L, Yang C, Wang L, Zhu H, Zeng Y, Guo W (2014) Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. Nat Commun 5:5519

Sun H, Hao P, Gu L, Cheng S, Wang H, Wu A, Ma L, Wei H, Yu S (2020a) Pectate lyase-like gene *GhPEL76* regulates organ elongation in Arabidopsis and fiber elongation in cotton. Plant Sci 293:110395

Sun Q, Huang J, Guo Y, Yang M, Guo Y, Li J, Zhang J, Xu W (2020b) A cotton NAC domain transcription factor, GhFSN5, negatively regulates secondary cell wall biosynthesis and anther development in transgenic Arabidopsis. Plant Physiol Bioch 146:303-314

Sun R, Li C, Zhang J, Li F, Ma L, Tan Y, Wang Q, Zhang B (2017) Differential expression of microRNAs during fiber development between fuzzless-lintless mutant and its wild-type allotetraploid cotton. Sci Rep 7 (1):1-10

Thyssen GN, Fang DD, Turley RB, Florane CB, Li P, Mattison CP, Naoumkina M (2017) A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. Plant J 90 (1):111-121

Tian Y, Zhang T (2021) MIXTAs and phytohormones orchestrate cotton fiber development. Curr Opin Plant Biol 59:101975

Walford SA, Wu Y, Llewellyn DJ, Dennis ES (2012) Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, *GhHD-1*. Plant J 71 (3):464-478

Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, Fang L, Yang S, Hu Y, Ye W (2016) Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. New Phytol 210 (4):1298-1310

Wan Q, Zhang H, Ye W, Wu H, Zhang T (2014) Genome-wide transcriptome profiling revealed cotton fuzz fiber development having a similar molecular model as Arabidopsis trichome. PLoS One 9 (5)

Wang L, Cheng H, Xiong F, Ma S, Zheng L, Song Y, Deng K, Wu H, Li F, Yang Z

(2020a) Comparative phosphoproteomic analysis of BR-defective mutant reveals a key role of GhSK13 in regulating cotton fiber development. Sci China Life Sci 63 (12):1905-1917

Wang L, Wang G, Long L, Altunok S, Feng Z, Wang D, Khawar KM, Mujtaba M (2020b) Understanding the role of phytohormones in cotton fiber development through omic approaches; recent advances and future directions. Int J Biol Macromol 163:1301-1313

Wang L, Zhu Y, Hu W, Zhang X, Cai C, Guo W (2015) Comparative transcriptomics reveals jasmonic acid-associated metabolism related to cotton fiber initiation. PloS One 10 (6)

Wang X, Miao Y, Cai Y, Sun G, Jia Y, Song S, Pan Z, Zhang Y, Wang L, Fu G, Gao Q, Ji G, Wang P, Chen B, Peng Z, Zhang X, Wang X, Ding Y, Hu D, Geng X, Wang L, Pang B, Gong W, He S, Du X (2020c) Large-fragment insertion activates gene *GaFZ* (*Ga08G0121*) and is associated with the fuzz and trichome reduction in cotton (*Gossypium arboreum*). Plant Biotechnol J 19(6):1110-1124

Wang Y, Yu Y, Chen Q, Bai G, Gao W, Qu Y, Ni Z (2019) Heterologous expression of *GbTCP4*, a Class II TCP transcription factor, regulates trichome formation and root hair development in Arabidopsis. Genes 10 (9):726

Wei Z, Li J (2018) Receptor-like protein kinases: Key regulators controlling root hair development in *Arabidopsis thaliana*. J Integr Plant Biol 60 (9):841-850

Wu H, Tian Y, Wan Q, Fang L, Guan X, Chen J, Hu Y, Ye W, Zhang H, Guo W (2018) Genetics and evolution of *MIXTA* genes regulating cotton lint fiber development. New Phytol 217 (2):883-895

Wu H, Zheng L, Qanmber G, Guo M, Wang Z, Yang Z (2021) Response of phytohormone mediated plant homeodomain (PHD) family to abiotic stress in upland cotton (*Gossypium hirsutum* spp.). BMC Plant Biol 21 (1):13

Wu R, Citovsky V (2017a) Adaptor proteins GIR1 and GIR2. I. Interaction with the repressor GLABRA2 and regulation of root hair development. Biochem Biophys Res Commun 488 (3):547-553

Wu R, Citovsky V (2017b) Adaptor proteins GIR1 and GIR2. II. Interaction with the co-repressor TOPLESS and promotion of histone deacetylation of target chromatin. Biochem Biophys Res Commun 488 (4):609-613

Yang C, Li H, Zhang J, Luo Z, Gong P, Zhang C, Li J, Wang T, Zhang Y, Ye Z (2011) A regulatory gene induces trichome formation and embryo lethality in tomato. P Natl Acad Sci 108 (29):11836-11841

Yang C, Ye Z (2013) Trichomes as models for studying plant cell differentiation. Cell Mol Life Sci 70 (11):1937-1948

Zeng J, Zhang M, Hou L, Bai W, Yan X, Hou N, Wang H, Huang J, Zhao J, Pei Y (2019) Cytokinin inhibits cotton fiber initiation by disrupting PIN3a-mediated asymmetric accumulation of auxin in the ovule epidermis. J Exp Bot 70 (12):3139-3151

Zhang G, Yue C, Lu T, Sun L, Hao F (2020) Genome-wide identification and expression analysis of NADPH oxidase genes in response to ABA and abiotic stresses,

and in fibre formation in *Gossypium*. Peer J 8:e8404

Zhu J, Chen J, Gao F, Xu C, Wu H, Chen K, Si Z, Yan H, Zhang T (2017) Rapid mapping and cloning of the *virescent-1* gene in cotton by bulked segregant analysis–next generation sequencing and virus-induced gene silencing strategies. J Exp Bot 68 (15):4125-4135

Zhu Q, Yuan Y, Stiller W, Jia Y, Wang P, Pan Z, Du X, Llewellyn D, Wilson I (2018) Genetic dissection of the fuzzless seed trait in *Gossypium barbadense*. J Exp Bot 69 (5):997-1009

Zhu Q, Stiller W, Moncuquet P, Gordon S, Yuan Y, Barnes S, Wilson I (2021) Genetic mapping and transcriptomic characterization of a new fuzzless-tufted cottonseed mutant. G3 (Bethesda) 11 (1):1-14

# *Supplementary Materials*

**Table S4-1**: Primers for qRT-PCR, subcellular localization and autoactivation.

| Gene | Sequence (5' to 3') |
|---|---|
| Ga01G0231_F | CACCGGCGTCGGCTATTTAGAA |
| Ga01G0231_R | GAGAGATGGGGGTTGAGGTCCA |
| Ga01G1554_F | GACGTCGTCGTCAGCAATGGTA |
| Ga01G1554_R | AAACGGTGCTTTTGCATTGGGG |
| Ga02G1535_F | GGTTCTGGTCGGATGTCCAAGG |
| Ga02G1535_R | CTCTTCCTTGCCTTGGTGGTGG |
| Ga03G0159_F | TCTCACCAACAAGCTCCAACGG |
| Ga03G0159_R | GACAAGCAGGGGAGGATTTGGG |
| Ga03G0898_F | TGTGCAGGCAGTGCCATATGTT |
| Ga03G0898_R | TAGTTTTGGAGGGCCTGGGTCT |
| Ga04G1030_F | GACGTCGTCGTCAGCAATGGTA |
| Ga04G1030_R | AAACGGTGCTTTTGCATTGGGG |
| Ga05G0849_F | ACCGATGACTCGAGGACGATGA |
| Ga05G0849_R | CCTTTTTGGCGTTGTGTTGGCT |
| Ga05G2409_F | CGAGGGCCGGTTCAGTTGATAA |
| Ga05G2409_R | TAGTTTCGGAGGGCCTTGGTCA |
| Ga07G1212_F | GCTAGCGCAAGTGAATCCGA |
| Ga07G1212_R | TCAGTTGGATGGCATTTTGGA |
| Ga07G1259_F | GCGCCTAGACTTGAAATCCGGT |
| Ga07G1259_R | CTCTTTCTCGGCACGTTCAAGC |
| Ga08G0121_F | TGTCTTCGGAACCCGGAGATGA |
| Ga08G0121_R | GGCATCCCACTAGCAGCATTGA |
| Ga10G2569_F | GTCTTGTTCAACCCCAAGCCCT |
| Ga10G2569_R | GGTGGTGACGATGAACGCTTCT |
| Ga11G0145_F | GAATCGTTGACGACGACGGAGA |
| Ga11G0145_R | GCCGCAGTATCGGTGGTAGTTT |
| Ga12G1438_F | CTGGAGGAACGGGATGAAAGCA |
| Ga12G1438_R | GTTGCTGCTGCTGCTGTTGTTA |
| Ga13G0044_F | GGGTCACCGAGCATGTCAGTTT |
| Ga13G0044_R | ACCGACTCCGACTCCTCCTTTT |
| Ga13G1545_F | TAGAGACCTTCTTGGCGTCGGA |

| | |
|---|---|
| Ga13G1545_R | TCAAGTCAAGGCGCTTTTCCCA |
| GhFzl_A_F | ATGTCAGTTTCTCCCCTGGAA |
| GhFzl_A_R | ATCGTCACCACTGACTCCTC |
| GhFzl_D_F | ATGTCAGTTTCTCCCCAGGAA |
| GhFzl_D_R | ATCGTCACCACCGACTCCGA |
| GhHis3_F | CGGTGGTGTGAAGAAGCCTCAT |
| GhHis3_R | AATTTCACGAACAAGCCTCTGGAA |
| CPC_F | TGATTGCTGGGAGAATCCGTGG |
| CPC_R | CATTTTCCGACGGCGAGTTGTG |
| ETC_F | CGGACTCCAACTCCATAGCGGA |
| ETC_R | TTCCAGGGATTCTCCCAGCGAT |
| GL2_F | GGCCTTGCTCCAAGACAAGTCA |
| GL2_R | AGTGGTAGCCATGCCACAGTTG |
| GL3_F | GGGTTGCGGAATTGGAATCTTGC |
| GL3_R | TCGCCATGGACACTTGAACTCG |
| Li3_F | AATCTTACCACCCGCCCCACTA |
| Li3_R | TGCATCATCAGCATCGCTGGAA |
| MYB25-like_F | ACTCCACGCCCTTCTTGGAAAC |
| MYB25-like_R | AGGGTCGATCCCTATCGTCGTC |
| TRY_F | CGTTGGTGACAAGTGGGCCTTA |
| TRY_R | GGCGAATCCTTCTCCGTGTCTC |
| TTG1_F | GCCGTCACCCATCTCAACTACC |
| TTG1_R | GATTTCCGGTGGGGTTGGAACA |
| GaFzl_GFP_F | GAGAACACGGGGGACTCTAGAATGTCAGTTTCTCCCCTGGAAATATC |
| GaFzl_GFP_R | CTTCTCCTTTACTCATACTAGTATTGCTTGCCCATTTAGCAGTC |
| GaFzl_GAL4_F | ATGGCCATGGAGGCCGAATTCATGTCAGTTTCTCCCCTGGAAATATC |
| GaFzl_GAL4_R | GGTTATGCTAGTTATGCGGCCGCTCAATTGCTTGCCCATTTAGCAGTC |

**Table S4-2**: Genome-wide identification of *GIR* family genes in *Gossypium raimondii*.

| No. | Gene ID | Chr. | Strand | Chr_Start | Chr_End | Size (AA) | Molecular Weight | Isoelectric Point | Grand Average of |
|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | (kDa) | | Hydropathy |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Gorai.006G151100 | Chr06 | - | 41,023,672 | 41,026,517 | 116 | 12.527 | 8.77 | -0.387 |
| 2 | Gorai.002G236700 | Chr02 | - | 59,818,045 | 59,819,961 | 115 | 12.536 | 8.505 | -0.474 |
| 3 | Gorai.009G398500 | Chr09 | - | 56,469,742 | 56,470,793 | 110 | 11.799 | 8.205 | -0.487 |
| 4 | Gorai.002G157600 | Chr02 | - | 34,177,429 | 34,178,436 | 119 | 12.972 | 8.773 | -0.565 |
| 5 | Gorai.007G367300 | Chr07 | - | 60,000,021 | 60,001,025 | 108 | 11.638 | 4.777 | -0.287 |
| 6 | Gorai.006G113900 | Chr06 | + | 36,199,084 | 36,200,233 | 113 | 12.476 | 5.138 | -0.415 |
| 7 | Gorai.009G083300 | Chr09 | - | 6,086,245 | 6,087,253 | 109 | 11.726 | 8.768 | -0.32 |
| 8 | Gorai.001G123500 | Chr01 | + | 15,425,854 | 15,426,780 | 110 | 12.31 | 7.084 | -0.477 |
| 9 | Gorai.004G011400 | Chr04 | + | 801,806 | 802,797 | 159 | 17.626 | 7.303 | -0.127 |
| 10 | Gorai.011G052900 | Chr11 | + | 4,162,581 | 4,164,306 | 237 | 25.176 | 9.926 | -0.436 |
| 11 | Gorai.002G106100 | Chr02 | - | 14,103,854 | 14,105,165 | 243 | 26.113 | 8.577 | -0.442 |
| 12 | Gorai.013G141800 | Chr13 | - | 38,630,018 | 38,631,879 | 284 | 30.984 | 9.707 | -0.524 |
| 13 | Gorai.013G011000 | Chr13 | - | 751,743 | 752,821 | 175 | 19.265 | 5.085 | -0.778 |
| 14 | Gorai.013G011300 | Chr13 | - | 772,808 | 773,188 | 126 | 13.932 | 4.229 | -0.244 |
| 15 | Gorai.013G011400 | Chr13 | - | 774,784 | 775,770 | 167 | 18.598 | 4.855 | -0.423 |
| 16 | Gorai.013G011100 | Chr13 | - | 759,971 | 760,906 | 181 | 19.995 | 4.157 | -0.713 |
| 17 | Gorai.002G022000 | Chr02 | - | 1,537,097 | 1,538,090 | 181 | 19.859 | 4.284 | -0.914 |
| 18 | Gorai.008G156400 | Chr08 | + | 41,718,629 | 41,721,106 | 202 | 22.748 | 4.266 | -0.993 |
| 19 | Gorai.005G094000 | Chr05 | + | 13,202,244 | 13,203,274 | 209 | 23.712 | 6.753 | -0.681 |
| 20 | Gorai.005G017900 | Chr05 | - | 1,345,616 | 1,346,921 | 194 | 22.144 | 6.497 | -0.617 |
| 21 | Gorai.009G237500 | Chr09 | - | 18,791,821 | 18,793,770 | 221 | 24.987 | 7.224 | -0.757 |

**Table S4-3**: Genome-wide identification of *GIR* family genes in *Gossypium hirsutum*.

| No. | Gene ID | Chr. | Strand | Chr_Start | Chr_End | Size (AA) | Molecular Weight (kDa) | Isoelectric Point | Grand Average of Hydropathy |
|---|---|---|---|---|---|---|---|---|---|
| 1 | GH_A09G1522 | A09 | - | 71,450,560 | 71,450,910 | 116 | 12.527 | 8.77 | -0.387 |
| 2 | GH_D09G1474 | D09 | - | 41,597,231 | 41,597,581 | 116 | 12.527 | 8.77 | -0.387 |
| 3 | GH_D01G2287 | D01 | - | 61,619,657 | 61,620,004 | 115 | 12.635 | 8.785 | -0.51 |
| 4 | GH_A01G2205 | A01 | - | 114,481,891 | 114,482,238 | 115 | 12.621 | 8.785 | -0.51 |
| 5 | GH_D04G0985 | D04 | + | 29,820,176 | 29,820,508 | 110 | 11.848 | 8.493 | -0.467 |
| 6 | GH_D01G1474 | D01 | - | 31,329,606 | 31,329,965 | 119 | 12.988 | 8.773 | -0.519 |
| 7 | GH_A04G0743 | A04 | + | 47,571,378 | 47,571,710 | 110 | 11.85 | 8.109 | -0.466 |
| 8 | GH_A09G1150 | A09 | + | 64,762,594 | 64,762,935 | 113 | 12.482 | 6.441 | -0.581 |
| 9 | GH_D11G3663 | D11 | - | 70,346,800 | 70,347,126 | 108 | 11.614 | 4.777 | -0.301 |
| 10 | GH_A11G3636 | A11 | - | 120,250,765 | 120,251,079 | 104 | 10.992 | 4.386 | -0.265 |
| 11 | GH_D09G1105 | D09 | + | 36,699,034 | 36,699,375 | 113 | 12.448 | 5.138 | -0.436 |
| 12 | GH_A07G1217 | A07 | + | 19,915,316 | 19,915,651 | 111 | 12.317 | 5.828 | -0.553 |
| 13 | GH_D05G0800 | D05 | - | 6,489,904 | 6,490,233 | 109 | 11.716 | 8.766 | -0.255 |
| 14 | GH_A05G0808 | A05 | - | 7,280,291 | 7,280,620 | 109 | 11.839 | 8.773 | -0.399 |
| 15 | GH_D07G1198 | D07 | + | 15,561,433 | 15,561,765 | 110 | 12.31 | 7.084 | -0.477 |
| 16 | GH_A08G0106 | A08 | + | 769,283 | 769,543 | 86 | 9.721 | 4.041 | -0.328 |
| 17 | GH_D08G0109 | D08 | + | 783,557 | 783,811 | 84 | 9.346 | 4.336 | -0.281 |
| 18 | GH_D10G0517 | D10 | + | 4,427,556 | 4,428,580 | 231 | 24.461 | 9.797 | -0.444 |
| 19 | GH_D01G0954 | D01 | - | 12,949,289 | 12,950,262 | 243 | 26.2 | 8.791 | -0.489 |
| 20 | GH_A10G0491 | A10 | + | 4,678,067 | 4,679,080 | 228 | 24.172 | 9.611 | -0.438 |
| 21 | GH_A01G0934 | A01 | - | 16,081,661 | 16,082,629 | 243 | 26.048 | 8.271 | -0.443 |
| 22 | GH_D13G1390 | D13 | - | 44,517,865 | 44,519,277 | 250 | 27.078 | 9.034 | -0.487 |
| 23 | GH_A13G1448 | A13 | - | 88,403,944 | 88,405,364 | 250 | 27.005 | 8.856 | -0.446 |
| 24 | GH_A13G0097 | A13 | - | 1,003,415 | 1,004,174 | 175 | 19.179 | 5.085 | -0.762 |
| 25 | GH_D13G0101 | D13 | - | 864,468 | 865,222 | 175 | 19.265 | 5.085 | -0.778 |
| 26 | GH_A13G0101 | A13 | - | 1,030,064 | 1,030,815 | 171 | 19.194 | 5.058 | -0.52 |
| 27 | GH_A13G0098 | A13 | - | 1,010,017 | 1,010,813 | 184 | 20.428 | 4.397 | -0.727 |
| 28 | GH_D13G0104 | D13 | - | 885,627 | 886,367 | 167 | 18.61 | 4.855 | -0.459 |
| 29 | GH_D13G0102 | D13 | - | 872,504 | 873,293 | 181 | 19.965 | 4.157 | -0.727 |
| 30 | GH_D01G0213 | D01 | - | 1,653,289 | 1,654,069 | 181 | 19.829 | 4.236 | -0.926 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 31 | GH_A01G0222 | A01 | - | 1,934,429 | 1,935,234 | 181 | 19.96 | 4.284 | -0.892 |
| 32 | GH_A12G1597 | A12 | + | 87,954,972 | 87,956,297 | 203 | 22.877 | 4.237 | -1.005 |
| 33 | GH_D12G1598 | D12 | + | 45,643,769 | 45,645,088 | 200 | 22.476 | 4.245 | -0.968 |
| 34 | GH_A13G0100 | A13 | - | 1,028,642 | 1,029,349 | 158 | 17.71 | 7.752 | -0.456 |
| 35 | GH_D02G0845 | D02 | + | 13,717,361 | 13,718,109 | 209 | 23.696 | 7.361 | -0.657 |
| 36 | GH_A02G0830 | A02 | + | 15,608,051 | 15,608,799 | 209 | 23.714 | 6.966 | -0.652 |
| 37 | GH_D02G0151 | D02 | - | 1,491,385 | 1,492,160 | 195 | 22.099 | 6.497 | -0.65 |
| 38 | GH_A02G0147 | A02 | - | 1,382,319 | 1,383,092 | 193 | 21.862 | 6.495 | -0.618 |
| 39 | GH_D05G2292 | D05 | - | 20,058,628 | 20,059,425 | 221 | 25.002 | 6.982 | -0.745 |
| 40 | GH_A05G2270 | A05 | - | 22,043,515 | 22,044,321 | 221 | 24.927 | 7.224 | -0.708 |

**Table S4-4**: Ka and Ks calculations of orthologous *GIR* gene pairs between A2 and AtDt.

| Genes 1 | Genes 2 | Ka | Ks | Ka_Ks |
|---|---|---|---|---|
| GH_A01G0222 | Ga01G0231 | 0.004767597 | 0 | 0 |
| GH_A01G2205 | Ga02G1535 | 0.007571042 | 0 | 0 |
| GH_A02G0147 | Ga03G0159 | 0.002174703 | 0 | 0 |
| GH_A02G0830 | Ga03G0898 | 0.002030458 | 0 | 0 |
| GH_A04G0743 | Ga04G1030 | 0.0046368 | 0.028439426 | 0.163041262 |
| GH_A05G0808 | Ga05G0849 | 0.012105172 | 0 | 0 |
| GH_A05G2270 | Ga05G2409 | 0.097526973 | 0.116267746 | 0.838813656 |
| GH_A07G1217 | Ga07G1212 | 0.011639419 | 0.013777655 | 0.844804061 |
| GH_A08G0106 | Ga08G0121 | 0 | 0 | 0 |
| GH_A09G1522 | Ga09G1549 | 0.104791847 | 0.105066595 | 0.997385008 |
| GH_A10G0491 | Ga10G2569 | 0.004361886 | 0.010574193 | 0.412502984 |
| GH_A11G3636 | Ga11G0145 | 0.004210537 | 0 | 0 |
| GH_A12G1597 | Ga12G1438 | 0 | 0 | 0 |
| GH_A13G0098 | Ga13G0101 | 0.011751122 | 0.008163346 | 1.439498198 |
| GH_A13G0100 | Ga13G0103 | 0.015962294 | 0.011396231 | 1.400664388 |
| GH_A13G0101 | Ga13G0104 | 0.006211216 | 0.013910754 | 0.44650459 |
| GH_A13G1448 | Ga13G1545 | 0.001752849 | 0.011267818 | 0.155562438 |
| GH_D01G0213 | Ga01G0231 | 0.014414858 | 0.068180343 | 0.211422493 |
| GH_D01G1474 | Ga01G1554 | 0.003750008 | 0.011236165 | 0.333744453 |
| GH_D01G2287 | Ga02G1535 | 0.011385418 | 0.065687472 | 0.173327083 |
| GH_D02G0151 | Ga03G0159 | 0.015390793 | 0.034242752 | 0.449461335 |
| GH_D02G0845 | Ga03G0898 | 0.024817015 | 0.022487631 | 1.103585134 |

| | | | | |
|---|---|---|---|---|
| GH_D04G0985 | Ga04G1030 | 0.009327753 | 0.088012952 | 0.105981598 |
| GH_D05G0800 | Ga05G0849 | 0.020354031 | 0.026089587 | 0.780159174 |
| GH_D05G2292 | Ga05G2409 | 0.107633894 | 0.164767801 | 0.653245924 |
| GH_D07G1198 | Ga07G1212 | 0.019634629 | 0.170837948 | 0.11493131 |
| GH_D08G0109 | Ga08G0121 | 0.036930537 | 0.053834762 | 0.685997966 |
| GH_D09G1474 | Ga09G1549 | 0.104715783 | 0.164712386 | 0.635749293 |
| GH_D10G0517 | Ga10G2569 | 0.003828981 | 0.048797695 | 0.078466429 |
| GH_D11G3663 | Ga11G0145 | 0.0340484 | 0.100148544 | 0.339978977 |
| GH_D12G1598 | Ga12G1438 | 0.002119394 | 0.040278134 | 0.052618968 |
| GH_D13G1390 | Ga13G1545 | 0.015929802 | 0.034291687 | 0.464538311 |
| GH_D13G0102 | Ga13G0101 | 0.048847372 | 0.097698896 | 0.499978747 |
| GH_D13G0104 | Ga13G0104 | 0.069537522 | 0.076867215 | 0.904644749 |

**Table S4-5**: Ka and Ks calculations of orthologous GIR gene pairs between D5 and AtDt.

| Genes 1 | Genes 2 | Ka | Ks | Ka_Ks |
|---|---|---|---|---|
| GH_A07G1217 | Gorai.001G123500 | 0.023647279 | 0.103740921 | 0.227945529 |
| GH_D07G1198 | Gorai.001G123500 | 0 | 0.043281238 | 0 |
| GH_A01G0934 | Gorai.002G106100 | 0.007210511 | 0.035809885 | 0.201355318 |
| GH_A01G0222 | Gorai.002G022000 | 0.011969135 | 0.042175552 | 0.283793202 |
| GH_A01G2205 | Gorai.002G236700 | 0.007561501 | 0.052308775 | 0.144555109 |
| GH_D01G0954 | Gorai.002G106100 | 0.007217016 | 0 | 0 |
| GH_D01G2287 | Gorai.002G236700 | 0.003771221 | 0.038885598 | 0.096982461 |
| GH_D01G0213 | Gorai.002G022000 | 0.007168513 | 0.024898552 | 0.287908845 |
| GH_D01G1474 | Gorai.002G157600 | 0.003750008 | 0 | 0 |
| GH_A08G0106 | Gorai.004G011400 | 0.037709917 | 0.090470991 | 0.416817777 |
| GH_D08G0109 | Gorai.004G011400 | 0.048731616 | 0.035145522 | 1.386566861 |
| GH_A02G0147 | Gorai.005G017900 | 0.017583223 | 0.025716805 | 0.683725013 |
| GH_A02G0830 | Gorai.005G094000 | 0.018465628 | 0.022800996 | 0.809860568 |
| GH_D02G0151 | Gorai.005G017900 | 0.00652178 | 0.008379975 | 0.778257677 |
| GH_D02G0845 | Gorai.005G094000 | 0.008177253 | 0 | 0 |
| GH_A09G1150 | Gorai.006G113900 | 0.019615371 | 0.038103432 | 0.514792767 |
| GH_A09G1522 | Gorai.006G151100 | 0 | 0.061133622 | 0 |
| GH_D09G1105 | Gorai.006G113900 | 0.003878483 | 0 | 0 |
| GH_D09G1474 | Gorai.006G151100 | 0 | 0.023906406 | 0 |
| GH_A11G3636 | Gorai.007G367300 | 0.047635054 | 0.117311765 | 0.406055218 |
| GH_D11G3663 | Gorai.007G367300 | 0.008108187 | 0.013274683 | 0.610800814 |

| | | | | |
|---|---|---|---|---|
| GH_A12G1597 | Gorai.008G156400 | 0 | 0.031919711 | 0 |
| GH_D12G1598 | Gorai.008G156400 | 0.002119394 | 0.007884435 | 0.268807311 |
| GH_A04G0743 | Gorai.009G398500 | 0.008040278 | 0.038511133 | 0.208778017 |
| GH_A05G0808 | Gorai.009G083300 | 0.032725027 | 0.026376345 | 1.240696083 |
| GH_A05G2270 | Gorai.009G237500 | 0.017437014 | 0.014337354 | 1.21619469 |
| GH_D04G0985 | Gorai.009G398500 | 0.004018764 | 0.012526387 | 0.320823852 |
| GH_D05G0800 | Gorai.009G083300 | 0.008075448 | 0 | 0 |
| GH_D05G2292 | Gorai.009G237500 | 0.005763717 | 0 | 0 |
| GH_A10G0491 | Gorai.011G052900 | 0.006308675 | 0.048003997 | 0.131419788 |
| GH_D10G0517 | Gorai.011G052900 | 0.001912047 | 0.011905012 | 0.160608569 |
| GH_A13G1448 | Gorai.013G141800 | 0.021334772 | 0.028385602 | 0.751605413 |
| GH_A13G0098 | Gorai.013G011000 | 0.143518462 | 0.108051673 | 1.328239141 |
| GH_A13G0100 | Gorai.013G011300 | 0.140277893 | 0.215597207 | 0.650648006 |
| GH_A13G0101 | Gorai.013G011400 | 0.073489986 | 0.062636702 | 1.173273541 |
| GH_D13G0101 | Gorai.013G011000 | 0 | 0 | 0 |
| GH_D13G0102 | Gorai.013G011000 | 0.159639423 | 0.133995846 | 1.191375915 |
| GH_D13G0104 | Gorai.013G011400 | 0.007768736 | 0.00891541 | 0.871382942 |
| GH_D13G1390 | Gorai.013G141800 | 0.003514945 | 0.005586618 | 0.629172232 |

**Table S4-6**: Ka and Ks calculations of orthologous GIR gene pairs between A2 and D5.

| Genes 1 | Genes 2 | Ka | Ks | Ka_Ks |
|---|---|---|---|---|
| Ga07G1212 | Gorai.001G123500 | 0.019634629 | 0.120256988 | 0.163272252 |
| Ga01G1554 | Gorai.002G157600 | 0.00751886 | 0.011236165 | 0.669166021 |
| Ga01G0231 | Gorai.002G022000 | 0.011978693 | 0.042057268 | 0.284818627 |
| Ga02G1535 | Gorai.002G236700 | 0.015209647 | 0.052194969 | 0.291400628 |
| Ga08G0121 | Gorai.004G011400 | 0.037709917 | 0.090470991 | 0.416817777 |
| Ga03G0159 | Gorai.005G017900 | 0.019803131 | 0.025753603 | 0.768946019 |
| Ga03G0898 | Gorai.005G094000 | 0.020573894 | 0.02268604 | 0.906896691 |
| Ga09G1549 | Gorai.006G151100 | 0.104791847 | 0.164332298 | 0.637682597 |
| Ga11G0145 | Gorai.007G367300 | 0.042807623 | 0.11561301 | 0.37026649 |
| Ga12G1438 | Gorai.008G156400 | 0 | 0.031919711 | 0 |
| Ga04G1030 | Gorai.009G398500 | 0.004642181 | 0.072961496 | 0.063625082 |
| Ga05G2409 | Gorai.009G237500 | 0.111816095 | 0.159685083 | 0.700228808 |
| Ga05G0849 | Gorai.009G083300 | 0.020298939 | 0.02631849 | 0.771280523 |
| Ga10G2569 | Gorai.011G052900 | 0.001857586 | 0.048016394 | 0.038686497 |
| Ga13G0101 | Gorai.013G011000 | 0.152386498 | 0.118157616 | 1.289688328 |
| Ga13G0103 | Gorai.013G011300 | 0.144531197 | 0.247253452 | 0.584546733 |
| Ga13G1545 | Gorai.013G141800 | 0.067134871 | 0.05266811 | 1.274677817 |

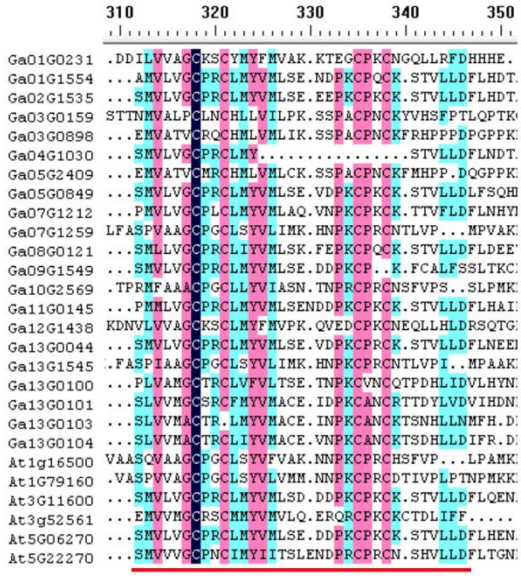



**Figure S4-1**: The multiple sequence alignment of the conserved regions of GaGIR proteins.

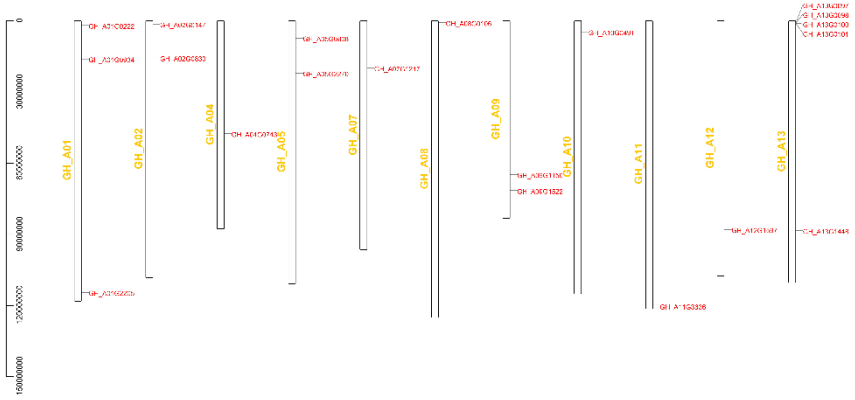



**Figure S4-2**: Chromosome distribution of *GIR* family genes in *G.hirsutum*_At.

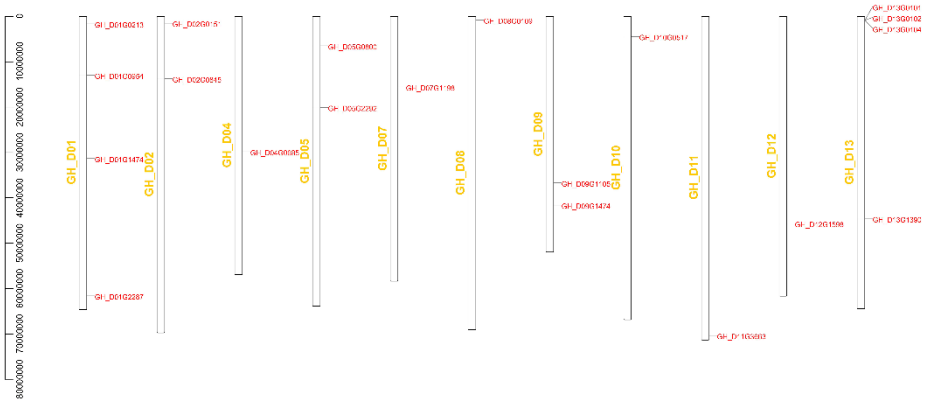



**Figure S4-3**: Chromosome distribution of *GIR* family genes in *G.hirsutum*_Dt.

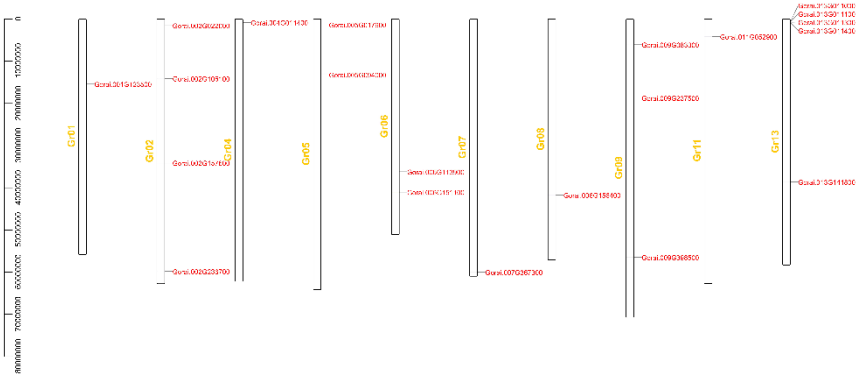



**Figure S4-4**: Chromosome distribution of *GIR* family genes in *G.raimondi*

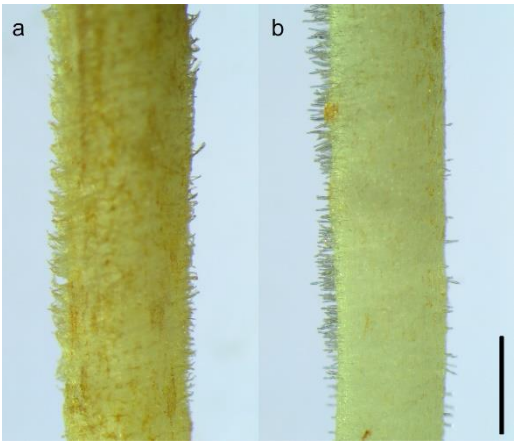



**Figure S4-5**: The root phenotypes of DPL971 and DPL972. **a** root phenotype of DPL971; **b** root phenotype of DPL972. The scale bar is 1 mm.

# 5

# Weighted Gene Co-expression Network Analysis Reveals Hub Genes Contributing to Fuzz Development in *Gossypium arboreum*

*In this chapter, the module highly associated with fuzz/fuzzless phenotype in Gossypium arboreum was identified and further investigated using WGCNA. And hub genes were screened and discerned as the crucial co-expression genes during fuzz development. This study provides technical advances and gene resources to explore the mechanism underlying fuzz initiation and formation.*

From Feng X, Liu S, Cheng H *et al.* Weighted gene co-expression network analysis reveals hub genes contributing to fuzz development in *Gossypium arboreum*. *Genes*. 2021; 12(5):753.

**Abstract**: Fuzzless mutants are ideal materials to decipher the regulatory network and mechanism underlying fuzz initiation and formation. In this study, we utilized two *Gossypium arboreum* accessions differing in fuzz characteristics to explore expression pattern differences and discriminate genes involved in fuzz development using RNA sequencing. Gene ontology (GO) analysis was conducted and found that DEGs were mainly enriched in the regulation of transcription, metabolic processes and oxidation–reduction-related processes. Weighted gene co-expression network analysis discerned the MEmagenta module highly associated with a fuzz/fuzzless trait, which included a total of 50 hub genes differentially expressed between two materials. *GaFZ*, which negatively regulates trichome and fuzz formation, was found involved in MEmagenta cluster1. In addition, twenty-eight hub genes in MEmagenta cluster1 were significantly up-regulated and expressed in fuzzless mutant DPL972. It is noteworthy that Ga04G1219 and Ga04G1240, which, respectively, encode Fasciclin-like arabinogalactan protein 18(FLA18) and transport protein, showed remarkable differences of expression level and implied that they may be involved in protein glycosylation to regulate fuzz formation and development. This module and hub genes identified in this study will provide new insights on fiber and fuzz formation and be useful for the molecular design breeding of cotton genetic improvement.

**Keywords:** transcriptome analysis; WGCNA; fuzz; module; hub genes

# 1. Introduction

Trichomes are highly specialized cells originating and extending from the epidermal surface of leaves, stems, petal bases and seeds in plants [1,2]. They contribute to a variety of unessential but important functions for plant growth, including protections against pathogens, insects or even herbivores, water regulation and increasing the tolerance to extreme high temperature and UV irradiation [3,4]. Trichomes could be characterized and classified into various types by their morphology and nature, e.g., uni-/multi-cellular, non-/branched or non-/glandular [3,5,6]. Numerous molecular genetic studies have provided much insight into the molecular genetic mechanisms and regulatory networks associated with trichome formation. Plenty of functional genes were identified to regulate the mode of epidermal cell development in Arabidopsis, rice, maize, tomato, *Artemisia annua*, such as *MYB-bHLH-WD40* (MBW) complex [7,8], *OsHL6* [9], *OsWOX3* [10], *ZmOCL4* [11], *ZmSPL10/14/26* [12], *SlMYC1* [13], *SlWo* [14], *AaHD1* [15].

Cotton fibers are single-celled and unbranched trichomes that outgrow from the seed epidermal cells and successively undergo a highly polarized elongation similar to leaf trichomes [16,17]. Therefore, cell fate determination and elongation patterns unveiled from Arabidopsis trichomes may provide a useful framework involved in cotton fiber initiation and elongation. It has been reported that the orthologue members in the Arabidopsis MBW complex also evolved crucial roles in cotton fiber formation [18]. Besides, recent studies have uncovered various genes involved in multiple pathways to govern fiber initiation and development. *GhPIN3a* [19,20], *GhARF2/18* [21] and *GhHD1* [22,23] were reported as crucial members to regulate fiber initiation in cotton by responding to various phytohormones. Furthermore, *GaDEL65* was predominantly associated with fuzz initiation and may interact with *GhMYB2/3* and *GhTTG3* to regulate cotton fiber elongation [24]. According to previous surveys, *GhPEL76* [25], *GhSK13* [26], *GhAP2L* [26], *GhHDZ5* [26] and *Gh_DNF_YB19* [27], have been identified to extensively regulate fiber elongation responsive to the phytohormone signaling pathway, respectively. AKR2A modulates the biosynthesis pathway of very long-chain fatty acids and participates in fiber elongation regulation [28]. More recently, a set of transcription factors, such as *GhFSN1* [29], *GhHOX3* [30] and the miR319-targeted *GhTCP4* [31], were identified to remarkably playing positive or negative roles in the second cell wall deposition stage.

Fiber initiation stage determines the final fiber yield. Hence it is essential and necessary to highly explore the orchestrating network governing fiber cell differentiation and fiber formation [32]. According to the initiating time and final length, cotton fibers can be divided into lint and fuzz. Though most research studies focused on fiber initiation and elongation, the essential functional genes and the mechanisms underlying fuzz initiation and development remain largely unknown.

Fiber mutant and natural species exhibit diverse or significant fiber characteristics and are great resources to decipher the mechanisms underlying fiber and fuzz formation. With the release of high-quality sequencing data from multiple cotton species, bioinformatics-based analyses have also facilitated functional gene mapping and net-work regulation exploration. *GhMML3* and *GhMML4* have been proposed to be responsible for the fuzzless or fiberless phenotypes in tetraploid fiber mutant lines

[32–37]. Compared with tetraploid cotton, diploid *G. arboreum*, a relatively simple genome, is potentially highly suitable for digging candidate genes corresponding to mutant traits. It has been reported that transcript splicing mistakes occurred because of a critical point mutation in *GaHD1* in a recessive fiber mutant *sma-4 (ha)* and further affects fiber and trichome initiation and retards elongation by cellular $H_2O_2$ and $Ca^{2+}$ signals [38]. Meanwhile, through whole genome re-sequencing and genetic linkage analyses, Liu et al. proposed *GaGIR1* and *GaMYB25-like* as crucial candidate genes controlling fuzz development and established *GaHD-1* to be the candidate gene responsible for the lintless trait in seven fuzzless accessions and one lintless accession of *G. arboreum* [39].

In a previous study, we fine-mapped a dominant fuzzless gene to a 70-kb region and speculated *GaGIR1* (GLABRA2-interacting repressor, Ga08G0121) as the likely candidate [40]. Recently, based on a genome-wide association study (GWAS), a 6.2 kb insertion, larINDEL$_{FZ}$, located ~18 kb upstream of *GaFZ* (Ga08G0121) has been identified to be an enhancer to elevate the expression level of *GaFZ* and negatively regulate fuzz and trichome formation [41]. Although much progress has been achieved, the mechanism and regulation network remains mysterious and elusive.

In this study, we utilized RNA-Seq for screening genes related to fuzz initiation development in two *G. arboreum* accessions and validated our data with qPCR. We also performed weighted co-expression gene network analysis to identify the modules and hub genes co-expressing with *GaFZ* and contributing to the fuzz development. Our results provide valuable gene resources and novel insights into mechanisms and regulatory networks of fuzz-associated genes.

# 2. Materials and methods

## *2.1 Plant materials*

Plants of *G. arboreum* DPL971 with normal fuzz and lint fibers and its isogenic fuzzless mutant line DPL972 were grown under standard agronomic field conditions at the Institute of Cotton Research, Chinese Academy of Agricultural Sciences (CAAS), Anyang (E 114°48′, N 36°06′), China. All lines were strictly self-pollinated. Flowers were firstly tagged at 0 DPA, and cotton bolls were collected in morning (9:00–11:00) at 1, 3 and 5 DPA, respectively; the ovules were dissected and frozen in liquid nitrogen for RNA extraction.

## *2.2 RNA extraction and RNA-Seq*

All samples were finely ground before RNA isolation. Total RNA was extracted following the manufacturer's instructions (TIANGEN BIOTECH Beijing, China). Two biological replicates from each sample were used for this experiment. The quantity and quality of total RNA were monitored by NanoDrop 2000 spectrophotometer and 1% agarose gel electrophoresis. RNA samples of good quality were used for subsequent deep sequencing and quantitative real-time PCR (qRT-PCR) analysis, and cDNA libraries were constructed and subjected to 101-cycle paired-end sequencing on an Illumina HiSeq4000 platform at Berry Genomics (Beijing, China). All raw data were de-posited to the NCBI's Sequence Read Archive (SRA) and

available with the accession number PRJNA729876.

## *2.3 RNA-Seq data analysis*

Raw sequences obtained from the 12 sample libraries were assessed and filtered by Fastp (version = 0.20.1). The GC content and Q30 were calculated by FastQC. Hisat2 (version = 2.2.1) was used to align the remaining clean data to reference genome of *G. arboreum* [42]. The number of fragments per kilobase of transcript per million mapped reads (FPKM) of each transcript in the fuzzless mutant DPL972 and wild-type DPL971 was calculated using StringTie (version = 2.1.1) software. Based on read counts obtained from StringTie, differentially expressed genes (DEGs) between the mutant and the wild type were identified by using the DESeq2 with false discovery rate (FDR) < 0.05 and |log2 (ratio)| ≥ 1. The KMeans-clustering method was used for clustering DEGs. KMeans method was performed by python package Sklearn (version = 0.19.2). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were conducted via Cotton Functional Genomics Database (https://cottonfgd.org/ , accessed on 9th, March, 2021) to determine the main biological functions and path-ways of the significant DEGs.

## *2.3 qRT-PCR*

To verify the reproducibility and accuracy of RNA-Seq, qRT-PCR was also performed in this study. The extracted total RNA samples with 1 μg concentration were used as template to reverse transcription using TransScript All-in-one First-strand cDNA Synthesis SuperMix (TransGen Biotech, China) according to the manufacturer's instructions. QRT-PCR amplifications were conducted using TransStart TOP Green qPCR SuperMix (TransGen Biotech, China) on an ABI QuantStudio5 Real-time PCR System (Applied Biosystems, USA). DEG expressions were normalized using *GaHis3* as reference gene. A total of eleven pairs of specific primers were designed from with NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/ , accessed on 9th, March, 2021) and qPCR Primer Database (https://biodb.swu.edu.cn/qprimerdb/, accessed on 9th, March, 2021) . Primer sequences were provided in Supplementary Table S5-1. Each qRT-PCR reaction was performed in triplicate and relative expression levels were calculated using the $2^{-\Delta\Delta ct}$ method.

## *2.3 Construction of co-expression network and network analyses*

To construct the co-expression network underlying two accessions, transcripts with FPKM count ≥ 5 count were used and all the outliers were removed. The WGCNA (weighted gene co-expression network analysis) R package was implemented to analyze co-expression networks and select hub genes which highly correlated with fuzz characteristic [43–45]. The dendrogram was generated using the cutreeDynamicTree algorithm, with a limits of minimum module size of 30 genes. A weighted correlation threshold of ≥0.85 was set to screen modules correlated with fuzz/fuzzless traits. The Sankey diagram was performed by Pyecharts (version = 1.9.0). The interaction networks among genes in selected modules were constructed using Cytoscape and visualized by Pyecharts. The membership (KME) values threshold of ≥0.80 was set

to screen the module hub genes. Heatmap for visualization of expression was generated by python package Seaborn (version = 0.9.0).

# 3. Results

## *3.1 Data processing and analysis*

To capture a global view and the temporal gene expression changes during fiber and fuzz initiation development, a transcriptome profiling experiment was performed from 12 cotton ovule libraries including wild type DPL971 and the fuzzless mutant DPL972 during the fuzz and fiber initiation stage (+1DPA, +3DPA and +5 DPA). Two biological replicates were adopted for each material in each time point. After removing adaptor and low-quality reads, a total of 90.27 Gb of cleaning data was generated, including 30.1 million reads, averaging 7.52 G and 25,077,291 reads per sample (Supplementary Table S5-2). For all samples, the GC content varied from 44% to 45%, and Q30 altered from 95.24% to 95.51%. All clean reads were aligned to the newly published reference *G. arboreum* genome, and 89.00% to 93.02% of the sequenced reads from each library could be perfectly mapped. These results indicated that our transcriptome sequencing data were of high quality and good enough to further identify DEGs and construct the regulatory network.

## *3.2 Differential gene expression in DPL971 and DPL972 during fuzz initiation*

A total of 1076 DEGs were detected using the criteria of false discovery rate (FDR) ≤0.05 and |log2 Ratio| ≥ 1. The number of DEGs between the WT and mutant during fuzz initiation stages (+1DPA, +3 DPA and +5 DPA) was 428, 330 and 614, respectively (Figure 5-1). Besides, eighty-two genes shared overlapping and differentially co-expression across the three different time points, suggesting that the transcriptome variation between the two accessions maintains a stable state through fuzz determination and the formation stage.

To verify the reproducibility and accuracy of the RNA-seq data, qRT-PCR was performed on selected 11 DEGs with gene-specific primers (Supplementary Table S2). As shown in Figure 5-2, all 11 genes showed a consistent expression pattern between the RNA-seq data and the qRT-PCR results. Genes significantly up-regulated in RNA-Seq data also exhibited an up-regulation in qPCR, and vice versa. These results also sup-ported the reproducibility and accuracy of transcriptomic data, which were utilized in downstream co-expression network and functional enrichment analysis.

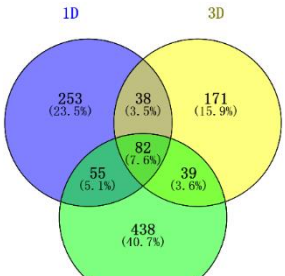


**Figure 5-1**: Venn diagram of DEGs identified from two diploid cotton accessions; 1D, 3D, 5D represents 1 Days post-anthesis (DPA), 3 DPA, 5 DPA.

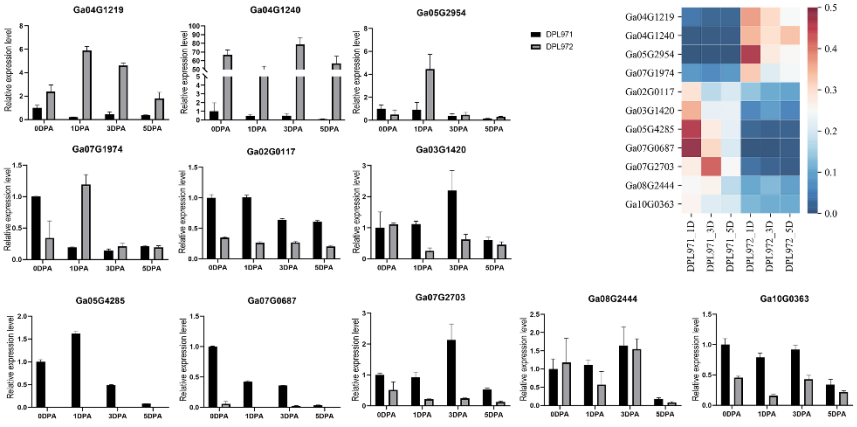



**Figure 5-2**: Validation of DEGs identified from transcriptome analysis with qRT-PCR. The heatmap on the top right of figure represents the relative transcription abundances (based on FPKM).

## *3.3 Functional enrichment of DEGs*

To estimate the main biological functions and regulations of identified DEGs, Gene ontology enrichment analysis and KEGG were performed on gene IDs that involved in each developmental points. As shown in Figure 5-3, DEGs generated from 1D were enriched in the following Molecular Function terms "transcription factor activity, sequence-specific DNA-binding" and "DNA-binding". They were also enriched in terms of "photosystem II stabilization" and "regulation of transcription, DNA-templated" from the biological process category. The enrichment of cellular component annotations included the terms of photosystem II reaction and extracellular region. While DEGs identified from 3D were mainly enriched in ATP synthesis and metabolic process and proton transporting associated terms, and DEGs identified from 5D were clustered into terms of "oxidoreductase activity" and "oxidation-reduction process". Further, the pathway enrichment analysis revealed that cell cycle, oxidative phosphorylation, biosynthesis of secondary metabolites, and glycolysis/gluconeogenesis were enriched among the DEGs and might be involved in fuzz initiation in cotton epidermal cell differentiation and development (Supplementary Table S5-3).

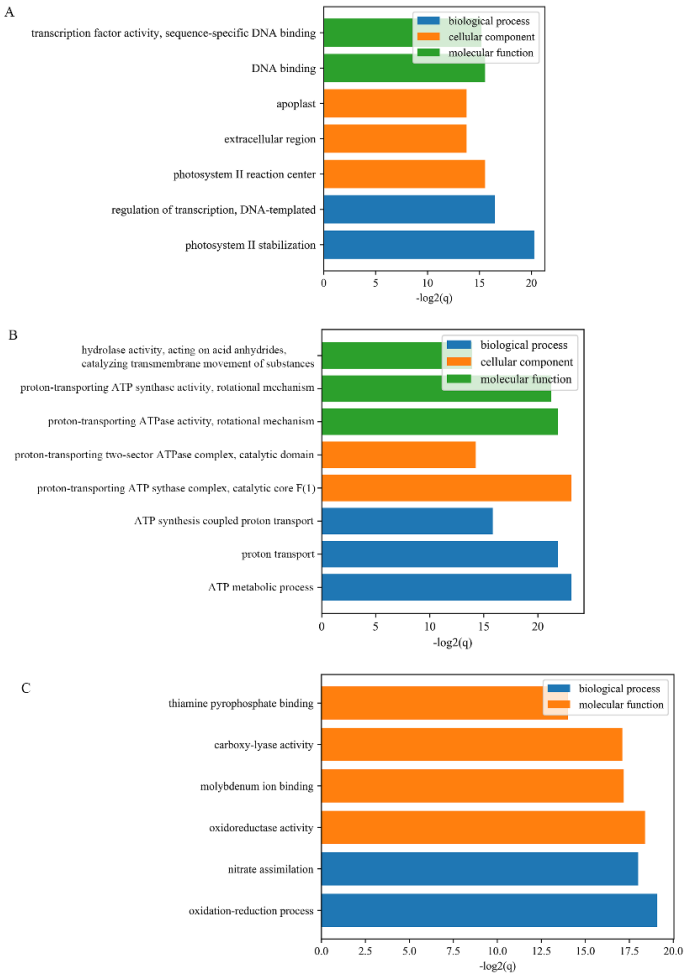



**Figure 5-3**: GO enrichment analysis of DEGs at three fiber developmental stages; (**A**), DEGs detected from 1DPA; (**B**), DEGs detected from 1DPA; (**C**) DEGs detected from 1DPA.

## *3.4 Co-expression network construction and module mining*

To uncover important genes and regulation pathways involved in fuzz and fiber initiation stage, we performed the R WGCNA package on all 12 samples. A soft-threshold power of 25 was introduced into the network topology to reveal the scale independence and mean connectivity of the network (Figure 5-4A,B). Based on the fragments per kilobase per million mapped (FPKM) expression matrix and phenotypic traits including fuzz or fuzzless characteristics and three developmental points, a total of 20,601 genes in 12 samples were clustered and divided into 10 modules which decorated with diacritical colors (Figure 5-4C). The outliers and genes whose FPKM values <5 in all samples were excluded in this analysis. Heatmap was graphed to unveil the correlation between modules and traits, p-values were presented in

Supplementary Table S5-4.

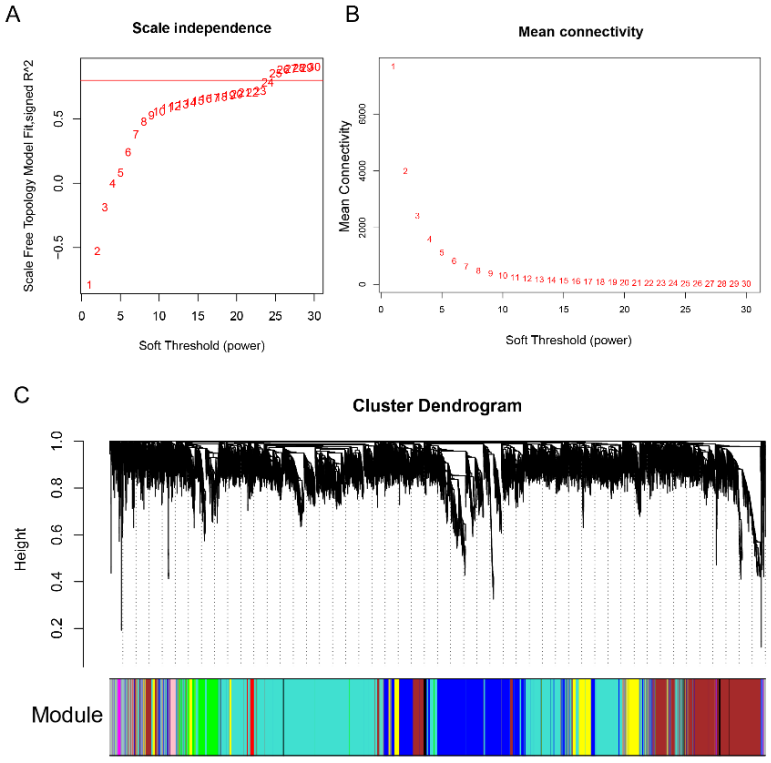



**Figure 5-4**: Module identification by weighted gene co-expression network analysis (WGCNA). (**A**,**B**) represent the soft threshold with scale independence and mean connectivity. (**C**), Hierarchical dendrogram reveals co-expression modules identified by WGCNA. Each leaf represents one gene. Ten modules were identified based on calculation of eigengenes; each module was decorated with a different color.

After calculating the correlation coefficient, one module was strongly associated with fuzz and fuzzless characteristics, indicating it might be highly contributed to fuzz initiation and formation (Figure 5-5). The MEmagenta module exhibited extremely exclusive correlations with fuzz/fuzzless traits (r = 0.996, p-value = 6.27×10-12 ). This module was analyzed further. In addition, we also identified red, brown and green modules significantly and substantially correlated with fiber development of 1, 3 and 5 DPA, respectively, suggesting genes involved in these modules most probably participated in fiber initiation and elongation during fiber developmental stages. Besides, the Sankey diagram was generated to detect the distribution of DEGs in each module, revealing that DEGs from MEmagenta distributed evenly at 1, 3, 5 DPA (Figure 5-6).

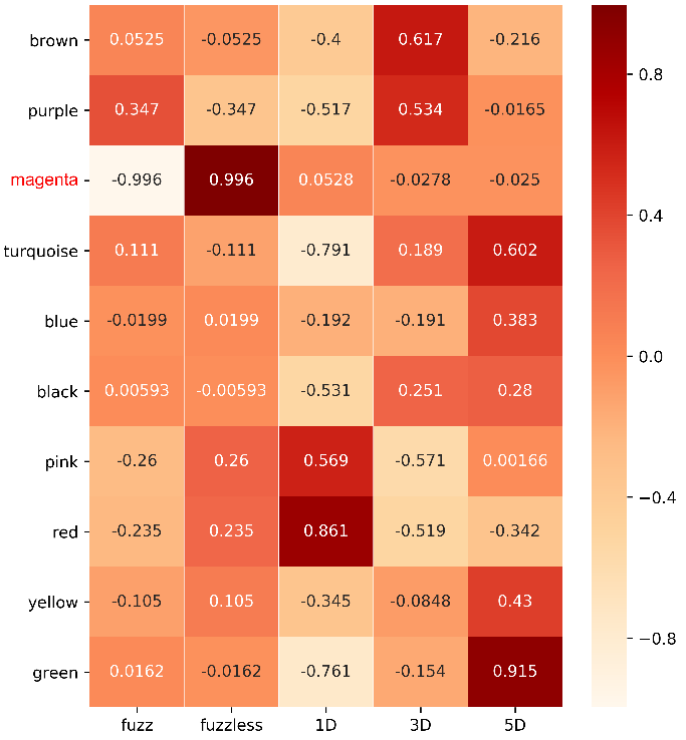



**Figure 5-5**: The WGCNA showed the MEmagenta module is significantly associated with fuzz formation. Each row means a module, and the correlation coefficient are shown in each square and the *p*-value was list in Table S5-4. The names in red in left represent the module highly associated with fuzz development.

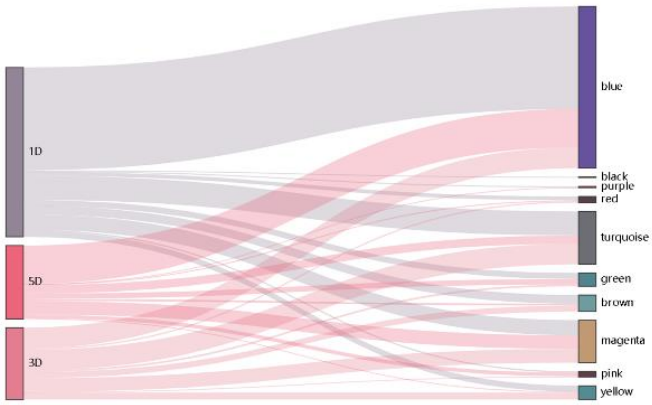



**Figure 5-6**: The Sankey diagram represents the distributions of DEGs in each module.

## *3.5 Functional enrichment and expression analysis of fuzz-associated hub genes*

To obtain more key genes involving the process of fuzz initiation, we did further analysis on the MEmagenta module which comprises 98 genes. To better understand gene functions and regulatory networks, GO and KEGG enrichment analyses were also used to analyze the fuzz-correlated module. The top terms were, respectively, presented in Supplementary Table S5-5 and Supplementary Figure S5-1. The GO results demonstrate genes in the MEmagenta module were enriched in transferase activity and transport protein, and the KEGG revealed genes in this module were best matched with the NADH dehydrogenase (ubiquinone) 1 α sub-complex item.

According to different expression profiles, but similar functions, genes in the MEmagenta module could be divided into two clusters using the K-means clustering method (Figure 5-7). Based on the kME value, the intramodular connectivity ranks and genes with kME value >0.8 were selected as hub genes that would determine the net-work centrality and potential crucial roles in fuzz/fuzzless characteristics. These hub genes were listed in Supplementary Table S5-6 and complex interaction networks in the co-expression module were respectively shown in Supplementary Figure S5-2 to high-light the potential networks underlying fuzz development.

To uncover the expression patterns of hub genes, we compared DEGs and genes in the selected module. Interestingly, a total of 50 out of 56 identified hub genes exhibited highly differential expression levels between DPL971 and DPL972 (Figure 5-8), indicating these hub genes play important roles in regulating fuzz initiation and development. In addition, previous studies reported *GaFZ* (Ga08G0121) differentially expressed be-tween DPL971 and DPL972 due to a large insertion and thus negatively regulates fuzz formation [40,41]. The gene was also identified and involved in MEmagenta module cluster1, which we selected as main cluster negatively correlating with fuzz trait (positively correlating with fuzzless trait), suggesting the reliability of the interested module and network identified from WGCNA. What is more, similar to *GaFZ*, Ga04G1219 encoding Fasciclin-like arabinogalactan protein 18 (FLA18) and Ga04G1240 encoding transport protein exhibited dramatic up-regulated expression in fuzzless DPL972 compared to DPL971 (Figure 5-2). It suggested that they and their interaction network may be involved in fuzz formation and development via controlling protein-glycosylation and responding to phytohormones.

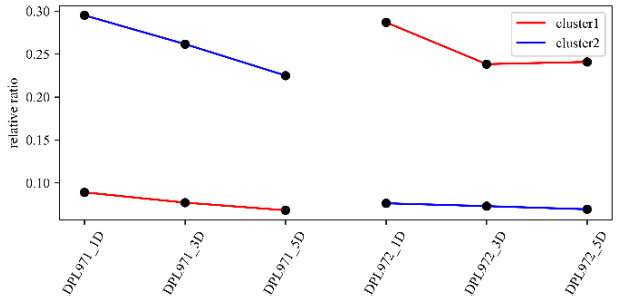



**Figure 5-7**: K-means clustering of DEGs in the MEmagenta module.

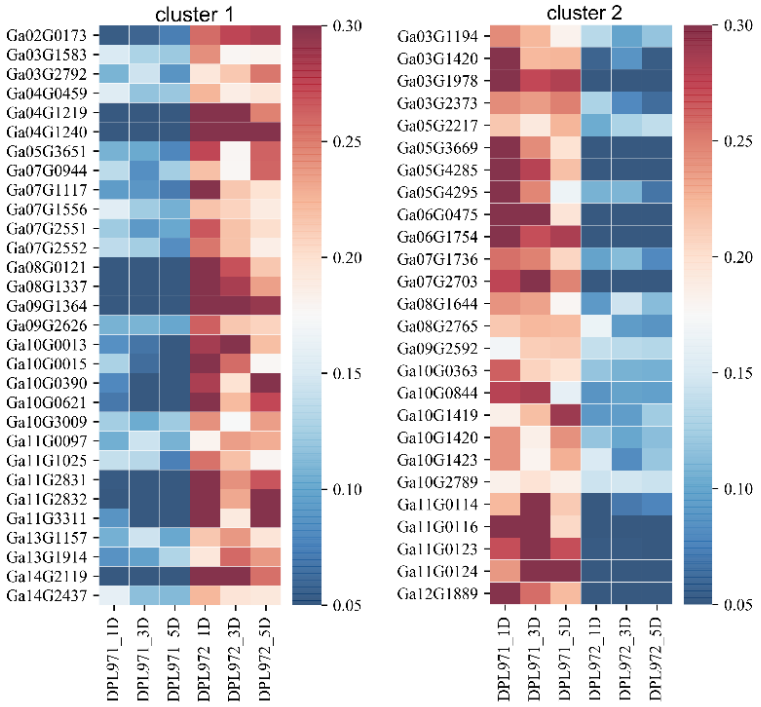



**Figure 5-8**: The expression heatmaps of hub genes in the MEmagenta module. The right side of figure represents the relative transcription abundances (based on FPKM). The gene IDs were listed on the left and sample names were displayed on the bottom.

# 4. Discussion

In the present study, we investigated gene expressions and regulatory networks in two *G. arboreum* accessions with a significant difference in fuzz characteristic. By per-forming transcriptomic analysis of fiber-attached ovules collected from 1, 3 and 5 DPA, DEGs were identified, respectively. GO enrichment analysis presented that DEGs in 1 DPA were mainly enriched in transcription factor activity, and DEGs in 3 and 5 DPA were clustered into items of ATP synthesis and proton transporting, oxidoreductase activity, respectively. Further, we established a weighted gene co-expression network analysis underlying fuzz initiation and formation. WGCNA revealed that only one co-expression module, MEmagenta, had a dominantly high correlation with fuzz trait, and the genes in these modules were significantly enriched in items of transferase and transport protein based on the GO enrichment analysis.

## *4.1 Hub genes in MEgrey60 cluster might be involved in regulatory network responsible to fuzz initiation*

Fiber and fuzz initiation undergo rather complex and obscure processes. Until now, multiple loci and factors have been identified to control or regulate fiber and fuzz initiation. In previous studies, larINDEL$_{FZ}$ was screened out to be responsible for the

fuzzless phenotype in various diploid accessions through genome wide association study (GWAS) and enhanced the expression level of *GaFZ* [41]. Overexpression of *GaFZ* in tetraploid cotton presented a reduction of trichomes and fuzz, suggesting *GaFZ* might be contributed to fuzz initiation and formation. In our study, *GaFZ* significantly differentially expressed in two accessions and was involved in MEmagenta module cluster1with a correlation of 0.816. By conducting the prediction and comparison of protein structure, we found this protein shared a high similarity with DNA-directed RNA polymerase II subunit RPB12, which functions as part of activated transcription complex (Supplementary Figure S5-2). It implies *GaFZ* might act in this way to regulate other genes to affect fuzz initiation. Besides, a comparison of the results of WGCNA and DEGs analysis indicated that 93 out of 98 genes in the ME-magenta module exhibited differential expression levels. In addition, 50 out of 56 hub genes, which were identified as critical genes to be significantly correlated with fuzz initiation and formation, presented notably differential expression patterns. In previous studies, glycosyltransferase31 could influence fiber initiation and elongation via controlling the glycosylation of FLAs and regulating pectin biosynthesis [46–53]. Interestingly, genes encoding FLA, UGT, RLK, transport protein were also identified in hub genes in MEmagenta cluster1. Together with *GaFZ*, these differentially expressed hub genes may be the key regulators underlying fuzz initiation and development in *G. arboreum* by protein glycosylation and phytohormone response. These results also indicated that the combination of transcriptomic analysis and WGCNA is highly efficient and superior in detecting and discerning functionally associated genes as well as gene regulatory networks.

## *4.2 Multiple protein kinase genes may regulate fuzz and plant development in G. arboreum*

Cotton is a great model for detecting the mechanism of cell differentiation and elongation. Understanding the molecular pathway regulating fiber and fuzz initiation can provide essential and valuable information for improving the fiber production and prompting the process of the molecular design breeding. Various protein kinases are important and essential in response to biotic and abiotic stresses as well as regulating plant and fiber development [54–59]. Mitogen-activated protein kinase (MAPK) cascades have been reported to alter pathogen resistance through gibberellins (GA) signaling pathway and participate in cotton fiber elongation by phosphorylating its downstream proteins [60,61]. Ectopic expression of receptor-like kinases (RLKs) improved cotton tolerance and resistance to fungal pathogen infections [62–66]. Calcineurin B-like protein and CBL-interacting protein kinases (CIPKs) were proved to be associated with cotton oil and sugar content and stress responses [59,67–70]. Glycosyl-transferases (UGT) were identified to participate and function in biosynthesis process of hemicellulose glucuronoxylan xylan during fiber and plant development whereas xyloglucan endo-transglycosylases /hydrolases (XTH) would limit the elongation of cotton fiber [56,71,72]. Glutathione S-transferase genes had an important role in resistance of *Verticillium wilt*, and sulfotransferases (SOTs) genes might be involved in cotton fiber development [73,74]. In the present study, CBL-interacting protein kinase 9 (CIPK9), UDP-glycosyltransferase 83A1 (UGT83A1), MDIS1-interacting receptor-like kinase 2 (MIK2), probable inactive histone-lysine N-

methyltransferase (SUVR1) and Inositol-tetrakisphosphate 1-kinase 2 (ITPK2) were identified as hub genes in the ME-magenta module, indicating that these protein kinase-related genes might be crucial and potential targets for the regulation of fiber/fuzz initiation and development in *G. arboreum*, and they are well worth further and deeper understanding and research.

# 5. Conclusions

In this study, we collected the fiber-attached ovules samples at three fiber developmental stages (1, 3 and 5 DPA) for RNA-Seq from a pair of *G. arboreum* accessions with different fuzz traits, and through transcriptome comparisons, 428 DEGs at 1DPA, 330 DEGs at 3 DPA and 614 DEGs at 5 DPA were identified, respectively. WGCNA identified a MEmagenta module highly associated with fuzz/fuzzless development. Within this module, a total of 93 genes exhibited significantly differential expression patterns between two accessions. *GaFZ*, which has been reported to negatively regulate trichome and fuzz formation, was found involved in MEmagenta cluster1. Besides, the expression levels of twenty-eight hub genes in MEmagenta cluster1 were notably up-regulated and twenty-two hub genes in cluster2 were dominantly down-regulated in fuzzless mutant DPL972. Furthermore, the coding genes of FLA18, UGT and transport protein showed remarkable differences of expression profiles and implied that they may be involved in protein glycosylation to regulate fuzz development. This study thus unveils the regulatory mechanisms underlying fuzz initiation and the development stage in *G. arboreum*, and may facilitate the molecular design breeding of cotton with improved fiber characteristics.

**Author Contributions:** Conceptualization, X.F. and G.S.; data curation, X.F.; formal analysis, X.F., S.L., H.C., D.Z., Y.Z. and L.L.; funding acquisition, G.S.; investigation, G.S.; methodology, H.C.; project administration, G.S.; resources, H.C., D.Z. and Q.W.; software, S.L.; supervision, G.S.; visualization, S.L. and H.C.; writing—original draft, X.F., S.L.; writing—review and editing, S.L. and H.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by grants from the National Natural Science Foundation of Chi-na (No. 31621005), the National Natural Science Foundation of China (No. 31901581) and the National Key R & D Plan of China (No. 2018YFD0100402).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in the article and supplementary material.

**Conflicts of Interest:** The authors declare no conflict of interest.

# References

1. Dai, X.; Zhou, L.; Zhang, W.; Cai, L.; Guo, H.; Tian, H.; Schiefelbein, J.; Wang, S. A single amino acid substitution in the R3 domain of GLABRA1

leads to inhibition of trichome formation in Arabidopsis without affecting its interaction with GLABRA3. Plant Cell Environ. 2015, 39, 897–907.

2. Tan, J.; Walford, S.A.; Dennis, E.S.; Llewellyn, D.J. Trichomes at the base of the petal are regulated by the same transcription factors as cotton seed fibers. Plant Cell Physiol. 2020, 61, 1590–1599.
3. Matias-Hernandez, L.; Aguilar-Jaramillo, A.E.; Cigliano, R.A.; Sanseverino, W.; Pelaz, S. Flowering and trichome development share hormonal and transcription factor regulation. J. Exp. Bot. 2016, 67, 1209–1219, doi:10.1093/jxb/erv534.
4. Yang, C.; Ye, Z. Trichomes as models for studying plant cell differentiation. Cell Mol. Life Sci. 2013, 70, 1937–1948, doi:10.1007/s00018-012-1147-6.
5. Hulskamp, M. Plant trichomes: a model for cell differentiation. Nat. Rev. Mol. Cell Biol. 2004, 5, 471–480, doi:10.1038/nrm1404.
6. Balkunde, R.; Pesch, M.; Hülskamp, M. Chapter Ten-trichome patterning in *Arabidopsis thaliana*: from genetic to molecular models. In Current Topics in Developmental Biology; Marja, C.P.T., Ed.; Academic Press: Cambridge, MA, USA, 2010; Volume 91, pp. 299–321.
7. Pattanaik, S.; Patra, B.; Singh, S.K.; Yuan, L. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Front. Plant Sci. 2014, 5, 259, doi:10.3389/fpls.2014.00259.
8. Yoshida, Y.; Sano, R.; Wada, T.; Takabayashi, J.; Okada, K. Jasmonic acid control of *GLABRA3* links inducible defense and trichome patterning in Arabidopsis. Development 2009, 136, 1039–1048, doi:10.1242/dev.030585.
9. Sun, W.; Gao, D.; Xiong, Y.; Tang, X.; Xiao, X.; Wang, C.; Yu, S. *Hairy Leaf 6*, an AP2/ERF transcription factor, interacts with *OsWOX3B* and regulates trichome formation in rice. Mol. Plant 2017, 10, 1417–1433, doi:10.1016/j.molp.2017.09.015.
10. Zhang, H.W., K.; Wang, Y.; Peng, Y.; Hu, F.; Wen, L.; Han, B.; Qian, Q.; Teng, S. A *WUSCHEL-like* homeobox gene, *OsWOX3B* responses to NUDA/GL-1 locus in rice. Rice J. 2012, 5, 30.
11. Vernoud, V.; Laigle, G.; Rozier, F.; Meeley, R.B.; Perez, P.; Rogowsky, P.M. The HD-ZIP IV transcription factor *OCL4* is necessary for trichome patterning and anther development in maize. Plant J. Cell Mol. Biol. 2009, 59, 883–894, doi:10.1111/j.1365-313X.2009.03916.x.
12. Kong, D.; Pan, X.; Jing, Y.; Zhao, Y.; Duan, Y.; Yang, J.; Wang, B.; Liu, Y.; Shen, R.; Cao, Y.; et al. *ZmSPL10/14/26* are required for epidermal hair cell fate specification on maize leaf. New Phytol. 2021, 230, 1533–1549, doi:10.1111/nph.17293.
13. Xu, J.; van Herwijnen, Z.O.; Drager, D.B.; Sui, C.; Haring, M.A.; Schuurink, R.C. *SlMYC1* regulates type VI glandular trichome formation and terpene biosynthesis in tomato glandular cells. Plant Cell 2018, 30, 2988–3005, doi:10.1105/tpc.18.00571.
14. Yang, C.; Li, H.; Zhang, J.; Luo, Z.; Gong, P.; Zhang, C.; Li, J.; Wang, T.;

Zhang, Y.; Lu, Y.; et al. A regulatory gene induces trichome formation and embryo lethality in tomato. Proc. Natl. Acad. Sci. USA 2011, 108, 11836–11841, doi:10.1073/pnas.1100532108.

15. Yan, T.; Chen, M.; Shen, Q.; Li, L.; Fu, X.; Pan, Q.; Tang, Y.; Shi, P.; Lv, Z.; Jiang, W.; et al. HOMEODOMAIN PROTEIN 1 is required for jasmonate-mediated glandular trichome initiation in *Artemisia annua*. New Phytol. 2017, 213, 1145–1155, doi:10.1111/nph.14205.
16. Wang, L.K.D.; Ruan, Y. Looking into 'hair tonics' for cotton fiber initiation. New Phytol. 2020, 229, 1844–1851.
17. Yu, Y.; Wu, S.; Nowak, J.; Wang, G.; Han, L.; Feng, Z.; Mendrinna, A.; Ma, Y.; Wang, H.; Zhang, X.; et al. Live-cell imaging of the cytoskeleton in elongating cotton fibres. Nat. Plants 2019, 5, 498–504, doi:10.1038/s41477-019-0418-8.
18. Wang, L.; Wang, G.; Long, L.; Altunok, S.; Feng, Z.; Wang, D.; Khawar, K.M.; Mujtaba, M. Understanding the role of phytohormones in cotton fiber development through omic approaches; recent advances and future directions. Int. J. Biol. Macromol. 2020, 163, 1301–1313, doi:10.1016/j.ijbiomac.2020.07.104.
19. Mei, G.; Zhang, Z. Optimization of polar distribution of *GhPIN3a* in the ovule epidermis improves cotton fiber development. J. Exp. Bot. 2019, 70, 3021–3023, doi:10.1093/jxb/erz183.
20. Zeng, J.; Zhang, M.; Hou, L.; Bai, W.; Yan, X.; Hou, N.; Wang, H.; Huang, J.; Zhao, J.; Pei, Y. Cytokinin inhibits cotton fiber initiation by disrupting *PIN3a*-mediated asymmetric accumulation of auxin in the ovule epidermis. J. Exp. Bot. 2019, 70, 3139–3151, doi:10.1093/jxb/erz162.
21. Xiao, G.; He, P.; Zhao, P.; Liu, H.; Zhang, L.; Pang, C.; Yu, J. Genome-wide identification of the *GhARF* gene family reveals that *GhARF2* and *GhARF18* are involved in cotton fibre cell initiation. J. Exp. Bot. 2018, 69, 4323–4337, doi:10.1093/jxb/ery219.
22. Zhang, M.; Zeng, J.Y.; Long, H.; Xiao, Y.H.; Yan, X.Y.; Pei, Y. Auxin regulates cotton fiber initiation via *GhPIN*-Mediated auxin transport. Plant Cell Physiol. 2017, 58, 385–397, doi:10.1093/pcp/pcw203.
23. Walford, S.A.; Wu, Y.; Llewellyn, D.J.; Dennis, E.S. Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, *GhHD-1*. Plant J. Cell Mol. Biol. 2012, 71, 464–478, doi:10.1111/j.1365-313X.2012.05003.x.
24. Shangguan, X.X.; Yang, C.Q.; Zhang, X.F.; Wang, L.J. Functional characterization of a basic helix-loop-helix (bHLH) transcription factor *GhDEL65* from cotton (*Gossypium hirsutum*). Physiol. Plant 2016, 158, 200–212, doi:10.1111/ppl.12450.
25. Sun, H.; Hao, P.; Gu, L.; Cheng, S.; Wang, H.; Wu, A.; Ma, L.; Wei, H.; Yu, S. Pectate lyase-like Gene *GhPEL76* regulates organ elongation in Arabidopsis and fiber elongation in cotton. Plant Sci. Int. J. Exp. Plant Biol. 2020, 293, 110395, doi:10.1016/j.plantsci.2019.110395.

26. Wang, L.; Cheng, H.; Xiong, F.; Ma, S.; Zheng, L.; Song, Y.; Deng, K.; Wu, H.; Li, F.; Yang, Z. Comparative phosphoproteomic analysis of BR-defective mutant reveals a key role of *GhSK13* in regulating cotton fiber development. Sci. China Life Sci. 2020, 63, 1905–1917, doi:10.1007/s11427-020-1728-9.
27. Chen, Y.; Yang, Z.; Xiao, Y.; Wang, P.; Wang, Y.; Ge, X.; Zhang, C.; Zhang, X.; Li, F. Genome-wide analysis of the NF-YB gene family in *Gossypium hirsutum* L. and characterization of the role of *GhDNF-YB22* in Embryogenesis. Int. J. Mol. Sci. 2018, 19, 483, doi:10.3390/ijms19020483.
28. Hu, W.; Chen, L.; Qiu, X.; Wei, J.; Lu, H.; Sun, G.; Ma, X.; Yang, Z.; Zhu, C.; Hou, Y.; et al. AKR2A participates in the regulation of cotton fibre development by modulating biosynthesis of very-long-chain fatty acids. Plant Biotechnol. J. 2020, 18, 526–539, doi:10.1111/pbi.13221.
29. Zhang, J.; Huang, G.-Q.; Zou, D.; Yan, J.-Q.; Li, Y.; Hu, S.; Li, X.-B. The cotton (*Gossypium hirsutum*) NAC transcription factor (*FSN1*) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. New Phytol. 2017, 217, 625–640, doi:10.1111/nph.14864.
30. Shan, C.-M.; Shangguan, X.-X.; Zhao, B.; Zhang, X.-F.; Chao, L.-m.; Yang, C.-Q.; Wang, L.-J.; Zhu, H.-Y.; Zeng, Y.-D.; Guo, W.-Z.; et al. Control of cotton fibre elongation by a homeodomain transcription factor *GhHOX3*. Nat. Commun. 2014, 5, 5519, doi:10.1038/ncomms6519.
31. Cao, J.F.; Zhao, B.; Huang, C.C.; Chen, Z.W.; Zhao, T.; Liu, H.R.; Hu, G.J.; Shangguan, X.X.; Shan, C.M.; Wang, L.J.; et al. The miR319-targeted *GhTCP4* promotes the transition from cell elongation to wall thickening in cotton fiber. Mol. Plant 2020, 13, 1063–1077, doi:10.1016/j.molp.2020.05.006.
32. Tian, Y.; Zhang, T. MIXTAs and phytohormones orchestrate cotton fiber development. Curr. Opin. Plant Biol. 2021, 59, 101975, doi:10.1016/j.pbi.2020.10.007.
33. Wan, Q.; Guan, X.; Yang, N.; Wu, H.; Pan, M.; Liu, B.; Fang, L.; Yang, S.; Hu, Y.; Ye, W.; et al. Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. New Phytol. 2016, 210, 1298–1310, doi:10.1111/nph.13860.
34. Chen, W.; Li, Y.; Zhu, S.; Fang, S.; Zhao, L.; Guo, Y.; Wang, J.; Yuan, L.; Lu, Y.; Liu, F.; et al. A retrotransposon insertion in *GhMML3_D12* is likely responsible for the lintless locus *li3* of tetraploid cotton. Front. Plant Sci. 2020, 11, 593679, doi:10.3389/fpls.2020.593679.
35. Wu, H.; Tian, Y.; Wan, Q.; Fang, L.; Guan, X.; Chen, J.; Hu, Y.; Ye, W.; Zhang, H.; Guo, W.; et al. Genetics and evolution of *MIXTA* genes regulating cotton lint fiber development. New Phytol. 2017, 217, 883–895, doi:10.1111/nph.14844.
36. Zhu, Q.-H.; Yuan, Y.; Stiller, W.; Jia, Y.; Wang, P.; Pan, Z.; Du, X.; Llewellyn, D.; Wilson, I. Genetic dissection of the fuzzless seed trait in *Gossypium barbadense*. J. Exp. Bot. 2018, 69, 997–1009, doi:10.1093/jxb/erx459.
37. Naoumkina, M.; Thyssen, G.N.; Fang, D.D.; Li, P.; Florane, C.B. Elucidation

of sequence polymorphism in fuzzless-seed cotton lines. Mol. Genet. Genom. MGG 2021, 296, 193–206, doi:10.1007/s00438-020-01736-z.

38. Ding, M.; Cao, Y.; He, S.; Sun, J.; Dai, H.; Zhang, H.; Sun, C.; Jiang, Y.; Paterson, A.H.; Rong, J. *GaHD1*, a candidate gene for the *Gossypium arboreum SMA-4* mutant, promotes trichome and fiber initiation by cellular $H_2O_2$ and $Ca^{2+}$ signals. Plant Mol. Biol. 2020, 103, 409–423, doi:10.1007/s11103-020-01000-3.
39. Liu, X.; Moncuquet, P.; Zhu, Q.H.; Stiller, W.; Zhang, Z.; Wilson, I. Genetic identification and transcriptome analysis of lintless and fuzzless traits in *Gossypium arboreum* L. Int. J. Mol. Sci. 2020, 21, 1675, doi:10.3390/ijms21051675.
40. Feng, X.; Cheng, H.; Zuo, D.; Zhang, Y.; Wang, Q.; Liu, K.; Ashraf, J.; Yang, Q.; Li, S.; Chen, X.; et al. Fine mapping and identification of the fuzzless gene *GaFzl* in DPL972 (*Gossypium arboreum*). TAG Theor. Appl. Genet. Theor. Angew. Genet. 2019, 132, 2169–2179, doi:10.1007/s00122-019-03330-3.
41. Wang, X.; Miao, Y.; Cai, Y.; Sun, G.; Jia, Y.; Song, S.; Pan, Z.; Zhang, Y.; Wang, L.; Fu, G.; et al. Large-fragment insertion activates gene *GaFZ* (*Ga08G0121*) and is associated with the fuzz and trichome reduction in cotton (*Gossypium arboreum*). Plant Biotechnol. J. 2020, 1-15. doi:10.1111/pbi.13532.
42. Du, X.; Huang, G.; He, S.; Yang, Z.; Sun, G.; Ma, X.; Li, N.; Zhang, X.; Sun, J.; Liu, M.; et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat. Genet. 2018, 50, 796–802, doi:10.1038/s41588-018-0116-x.
43. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinform. 2008, 9, 559, doi:10.1186/1471-2105-9-559.
44. Kuang, J.F.; Wu, C.J.; Guo, Y.F.; Walther, D.; Shan, W.; Chen, J.Y.; Chen, L.; Lu, W.J. Deciphering transcriptional regulators of banana fruit ripening by regulatory network analysis. Plant Biotechnol. J. 2021, 19, 477–489, doi:10.1111/pbi.13477.
45. Zaidi, S.S.; Naqvi, R.Z.; Asif, M.; Strickler, S.; Shakir, S.; Shafiq, M.; Khan, A.M.; Amin, I.; Mishra, B.; Mukhtar, M.S.; et al. Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (*Gossypium hirsutum*). Plant Biotechnol. J. 2020, 18, 691–706, doi:10.1111/pbi.13236.
46. Huang, G.Q.; Xu, W.L.; Gong, S.Y.; Li, B.; Wang, X.L.; Xu, D.; Li, X.B. Characterization of 19 novel cotton *FLA* genes and their expression profiling in fiber development and in response to phytohormones and salt stress. Physiol. Plant 2008, 134, 348–359, doi:10.1111/j.1399-3054.2008.01139.x.
47. Bowling, A.J.; Vaughn, K.C.; Turley, R.B. Polysaccharide and glycoprotein distribution in the epidermis of cotton ovules during early fiber initiation and growth. Protoplasma 2011, 248, 579–590, doi:10.1007/s00709-010-0212-y.
48. Huang, G.Q.; Gong, S.Y.; Xu, W.L.; Li, W.; Li, P.; Zhang, C.J.; Li, D.D.; Zheng, Y.; Li, F.G.; Li, X.B. A fasciclin-like arabinogalactan protein, GhFLA1,

is involved in fiber initiation and elongation of cotton. Plant Physiol. 2013, 161, 1278–1290, doi:10.1104/pp.112.203760.

49. Kumar, S.; Kumar, K.; Pandey, P.; Rajamani, V.; Padmalatha, K.V.; Dhandapani, G.; Kanakachari, M.; Leelavathi, S.; Kumar, P.A.; Reddy, V.S. Glycoproteome of elongating cotton fiber cells. Mol. Cell Proteom. 2013, 12, 3677–3689, doi:10.1074/mcp.M113.030726.
50. Liu, H.; Shi, R.; Wang, X.; Pan, Y.; Li, Z.; Yang, X.; Zhang, G.; Ma, Z. Characterization and expression analysis of a fiber differentially expressed Fasciclin-like arabinogalactan protein gene in Sea Island cotton fibers. PLoS ONE 2013, 8, e70185, doi:10.1371/journal.pone.0070185.
51. Qin, L.X.; Rao, Y.; Li, L.; Huang, J.F.; Xu, W.L.; Li, X.B. Cotton *GalT1* encoding a putative glycosyltransferase is involved in regulation of cell wall pectin biosynthesis during plant development. PLoS ONE 2013, 8, e59115, doi:10.1371/journal.pone.0059115.
52. Hande, A.S.; Katageri, I.S.; Jadhav, M.P.; Adiger, S.; Gamanagatti, S.; Padmalatha, K.V.; Dhandapani, G.; Kanakachari, M.; Kumar, P.A.; Reddy, V.S. Transcript profiling of genes expressed during fibre development in diploid cotton (*Gossypium arboreum* L.). BMC Genom. 2017, 18, 675, doi:10.1186/s12864-017-4066-y.
53. Qin, L.X.; Chen, Y.; Zeng, W.; Li, Y.; Gao, L.; Li, D.D.; Bacic, A.; Xu, W.L.; Li, X.B. The cotton beta-galactosyltransferase 1 (GalT1) that galactosylates arabinogalactan proteins participates in controlling fiber development. Plant J. Cell Mol. Biol. 2017, 89, 957–971, doi:10.1111/tpj.13434.
54. Yang, C.-Q.; Wu, X.-M.; Ruan, J.-X.; Hu, W.-L.; Mao, Y.-B.; Chen, X.-Y.; Wang, L.-J. Isolation and characterization of terpene synthases in cotton (*Gossypium hirsutum*). Phytochemistry 2013, 96, 46–56, doi:10.1016/j.phytochem.2013.09.009.
55. Zou, C.; Lu, C.; Shang, H.; Jing, X.; Cheng, H.; Zhang, Y.; Song, G. Genome-wide analysis of the *Sus* gene family in cotton. J. Integr. Plant Biol. 2013, 55, 643–653, doi:10.1111/jipb.12068.
56. Li, L.; Huang, J.; Qin, L.; Huang, Y.; Zeng, W.; Rao, Y.; Li, J.; Li, X.; Xu, W. Two cotton fiber-associated glycosyltransferases, GhGT43A1 and GhGT43C1, function in hemicellulose glucuronoxylan biosynthesis during plant development. Physiol. Plant 2014, 152, 367–379, doi:10.1111/ppl.12190.
57. Zhang, B.; Liu, J.-Y. Cotton cytosolic pyruvate kinase GhPK6 participates in fast fiber elongation regulation in a ROS-mediated manner. Planta 2016, 244, 915–926, doi:10.1007/s00425-016-2557-8.
58. Wang, J.J.; Lu, X.K.; Yin, Z.J.; Mu, M.; Zhao, X.J.; Wang, D.L.; Wang, S.; Fan, W.L.; Guo, L.X.; Ye, W.W.; et al. Genome-wide identification and expression analysis of *CIPK* genes in diploid cottons. Genet. Mol. Res. 2016, 15, 10-4238, doi:10.4238/gmr15048852.
59. Cui, Y.; Su, Y.; Wang, J.; Jia, B.; Wu, M.; Pei, W.; Zhang, J.; Yu, J. Genome-wide characterization and analysis of CIPK gene family in two cultivated allopolyploid cotton species: sequence variation, association with seed oil

content, and the role of *GhCIPK6*. Int. J. Mol. Sci. 2020, 21, 863, doi:10.3390/ijms21030863.

60. Wang, C.; He, X.; Wang, X.; Zhang, S.; Guo, X.; Denby, K. Ghr-miR5272a-mediated regulation of *GhMKK6* gene transcription contributes to the immune response in cotton. J. Exp. Bot. 2017, 68, 5895–5906, doi:10.1093/jxb/erx373.
61. Chen, L.; Sun, H.; Wang, F.; Yue, D.; Shen, X.; Sun, W.; Zhang, X.; Yang, X. Genome-wide identification of MAPK cascade genes reveals the *GhMAP3K14-GhMKK11-GhMPK31* pathway is involved in the drought response in cotton. Plant Mol. Biol. 2020, 103, 211–223, doi:10.1007/s11103-020-00986-0.
62. Jun, Z.; Zhang, Z.; Gao, Y.; Zhou, L.; Fang, L.; Chen, X.; Ning, Z.; Chen, T.; Guo, W.; Zhang, T. Overexpression of *GbRLK*, a putative receptor-like kinase gene, improved cotton tolerance to *Verticillium wilt*. Sci. Rep. 2015, 5, 15048, doi:10.1038/srep15048.
63. Yuan, N.; Rai, K.M.; Balasubramanian, V.K.; Upadhyay, S.K.; Luo, H.; Mendu, V. Genome-wide identification and characterization of LRR-RLKs reveal functional conservation of the *SIF* subfamily in cotton (*Gossypium hirsutum*). BMC Plant Biol. 2018, 18, 185, doi:10.1186/s12870-018-1395-1.
64. Babilonia, K.; Wang, P.; Liu, Z.; Jamieson, P.; Mormile, B.; Rodrigues, O.; Zhang, L.; Lin, W.; Danmaigona Clement, C.; Menezes de Moura, S.; et al. A nonproteinaceous Fusarium cell wall extract triggers receptor-like protein-dependent immune responses in Arabidopsis and cotton. New Phytol. 2021, 230, 275–289, doi:10.1111/nph.17146.
65. Wang, P.; Zhou, L.; Jamieson, P.; Zhang, L.; Zhao, Z.; Babilonia, K.; Shao, W.; Wu, L.; Mustafa, R.; Amin, I.; et al. The cotton wall-associated kinase GhWAK7A mediates responses to fungal wilt pathogens by complexing with the chitin sensory receptors. Plant Cell 2020, 32, 3978–4001, doi:10.1105/tpc.19.00950.
66. Feng, H.; Li, C.; Zhou, J.; Yuan, Y.; Feng, Z.; Shi, Y.; Zhao, L.; Zhang, Y.; Wei, F.; Zhu, H. A cotton WAKL protein interacted with a DnaJ protein and was involved in defense against *Verticillium* dahliae. Int. J. Biol. Macromol. 2021, 167, 633–643, doi:10.1016/j.ijbiomac.2020.11.191.
67. Gao, P.; Zhao, P.M.; Wang, J.; Wang, H.Y.; Du, X.M.; Wang, G.L.; Xia, G.X. Co-expression and preferential interaction between two calcineurin B-like proteins and a CBL-interacting protein kinase from cotton. Plant Physiol. Biochem. PPB Soc. Fr. Physiol. Veg. 2008, 46, 935–940, doi:10.1016/j.plaphy.2008.05.001.
68. Lu, T.; Zhang, G.; Sun, L.; Wang, J.; Hao, F. Genome-wide identification of *CBL* family and expression analysis of *CBLs* in response to potassium deficiency in cotton. PeerJ 2017, 5, e3653, doi:10.7717/peerj.3653.
69. Deng, J.; Yang, X.; Sun, W.; Miao, Y.; He, L.; Zhang, X. The calcium sensor CBL2 and its interacting kinase CIPK6 are involved in plant sugar homeostasis via interacting with tonoplast sugar transporter TST2. Plant Physiol. 2020, 183, 236–249, doi:10.1104/pp.19.01368.
70. Sun, W.; Zhang, B.; Deng, J.; Chen, L.; Ullah, A.; Yang, X. Genome-wide

analysis of *CBL* and *CIPK* family genes in cotton: conserved structures with divergent interactions and expression. Physiol. Mol. Biol. Plants 2021, 27, 359–368, doi:10.1007/s12298-021-00943-1.

71. Huang, J.; Pang, C.; Fan, S.; Song, M.; Yu, J.; Wei, H.; Ma, Q.; Li, L.; Zhang, C.; Yu, S. Genome-wide analysis of the family 1 glycosyltransferases in cotton. Mol. Genet. Genom. MGG 2015, 290, 1805–1818, doi:10.1007/s00438-015-1040-8.
72. Lee, J.; Burns, T.H.; Light, G.; Sun, Y.; Fokar, M.; Kasukabe, Y.; Fujisawa, K.; Maekawa, Y.; Allen, R.D. Xyloglucan endotransglycosylase/hydrolase genes in cotton and their role in fiber elongation. Planta 2010, 232, 1191–1205, doi:10.1007/s00425-010-1246-2.
73. Li, Z.K.; Chen, B.; Li, X.X.; Wang, J.P.; Zhang, Y.; Wang, X.F.; Yan, Y.Y.; Ke, H.F.; Yang, J.; Wu, J.H.; et al. A newly identified cluster of glutathione S-transferase genes provides *Verticillium wilt* resistance in cotton. Plant J. Cell Mol. Biol. 2019, 98, 213–227, doi:10.1111/tpj.14206.
74. Wang, L.; Liu, X.; Wang, X.; Pan, Z.; Geng, X.; Chen, B.; Liu, B.; Du, X.; Song, X. Identification and characterization analysis of *sulfotransferases* (*SOTs*) gene family in cotton (*Gossypium*) and its involvement in fiber development. BMC Plant Biol. 2019, 19, 595, doi:10.1186/s12870-019-2190-3.

# *Supplementary Materials*

**Table S5-1**: Primers used in qRT-PCR

| Primer Name | Sequence (5' to 3') |
|---|---|
| qGa02G0117-F | AGACCCAAAGCTTCGAAACACGA |
| qGa02G0117-R | CCCAGGTCCGATTTCCAAGACG |
| qGa03G1420-F | CATCCGTAAGAACGTGGAAGT |
| qGa03G1420-R | TCACAGTTGTGAGAGGAGGGA |
| qGa04G1219-F | GGCCATGAACGCATCTACCCTT |
| qGa04G1219-R | ATGAACCCCGTAAGCACAGACG |
| qGa04G1240-F | TGATCCGCTAAGAGAGTTCGC |
| qGa04G1240-R | AGCTGTACGGACCGCTACTT |
| qGa05G2954-F | GGTGAAGGTCGTCAAGCTCA |
| qGa05G2954-R | TGTTTGAGCACCCTCTGTCC |
| qGa05G4285-F | TCGTTGGCCAGCATTGAGAACT |
| qGa05G4285-R | GGCGGCTATTTCCCACGTTACT |
| qGa07G2703-F | GGATGACCTCCACAGCAGGAAC |
| qGa07G2703-R | AATAGCCCGTTGCACATGCTGA |
| qGa08G2444-F | GCCTGGAACAACACAGAAGTGGA |
| qGa08G2444-R | GCCACTTGACCATCTGCGGTAT |
| qGa10G0363-F | CGTATTGCGCCCTAGCCATA |
| qGa10G0363-R | TCCCTATCCATGCCACTCCA |
| qGa10G0784-F | CGACGTTGTCACCAAATCCG |
| qGa10G0784-R | GAGGAGAGAGCACAATGAATGC |
| qGa07G0687-F | GCCAAAAGCACAAAAAGGGCT |
| qGa07G0687-R | TCAGCCGATCTGCGATCAAG |

**Table S5-2**: Statistics and alignments of RNA-seq reads mapped to reference genome

| Sample | Clean Bases (bp) | Total reads | GC Content (%) | CleanQ30 (%) | Read mapping rate (%) |
|---|---|---|---|---|---|
| DPL971_1DR1 | 7,168,622,700 | 23,895,409 | 45 | 95.38 | 89 |
| DPL971_1DR2 | 6,412,138,500 | 25,321,633 | 45 | 95.4 | 92.75 |
| DPL971_3DR1 | 7,571,760,300 | 25,239,201 | 45 | 95.31 | 92.3 |
| DPL971_3DR2 | 7,265,123,100 | 23,930,032 | 45 | 95.46 | 92.31 |
| DPL971_5DR1 | 7,042,378,500 | 23,474,595 | 45 | 95.27 | 92.39 |
| DPL971_5DR2 | 8,017,725,900 | 27,254,966 | 45 | 95.28 | 91.92 |
| DPL972_1DR1 | 7,596,489,900 | 21,373,795 | 45 | 95.41 | 91.44 |
| DPL972_1DR2 | 8,608,806,300 | 28,696,021 | 45 | 95.46 | 90.57 |
| DPL972_3DR1 | 7,179,009,600 | 24,217,077 | 45 | 95.25 | 93.02 |
| DPL972_3DR2 | 7,619,923,500 | 25,399,745 | 45 | 95.24 | 91.9 |
| DPL972_5DR1 | 8,176,489,800 | 26,725,753 | 45 | 95.51 | 91.14 |
| DPL972_5DR2 | 7,619,780,400 | 25,399,268 | 45 | 95.35 | 90.48 |

**Table S5-3**: KEGG enrichment analysis of DEGs obtained from two accessions

| Time Points | Accession | Name | Q-value |
|---|---|---|---|
| 1DPA | M00692 | Cell cycle - G1/S transition | 0.0037 |
| | M00130 | Inositol phosphate metabolism, PI=> PIP2 => Ins(1,4,5)P3 => Ins(1,3,4,5)P4 | 0.0037 |
| | ko05204 | Chemical carcinogenesis | 0.0068 |
| | ko00982 | Drug metabolism - cytochrome P450 | 0.0068 |
| | ko00980 | Metabolism of xenobiotics by cytochrome P450 | 0.0068 |
| | ko00940 | Phenylpropanoid biosynthesis | 0.0068 |
| 3DPA | ko05012 | Parkinson's disease | 0.0024 |
| | ko00982 | Drug metabolism - cytochrome P450 | 0.0024 |
| | ko04626 | Plant-pathogen interaction | 0.0076 |
| | ko00190 | Oxidative phosphorylation | 0.0079 |
| | ko05204 | Chemical carcinogenesis | 0.008 |
| | ko00980 | Metabolism of xenobiotics by cytochrome P450 | 0.008 |
| 5DPA | ko00430 | Taurine and hypotaurine metabolism | 0.00089 |
| | ko01100 | Metabolic pathways | 0.00089 |
| | ko01110 | Biosynthesis of secondary metabolites | 0.0013 |
| | ko00010 | Glycolysis / Gluconeogenesis | 0.0021 |
| | ko00982 | Drug metabolism - cytochrome P450 | 0.0043 |
| | M00146 | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex | 0.0043 |
| | M00368 | Ethylene biosynthesis, methionine => ethylene | 0.0052 |

**Table S5-4**: p-values of relationships between module and sample

| module | trait_fuzz | trait_fuzzless | trait_1D | trait_3D | trait_5D |
|---|---|---|---|---|---|
| brown | 0.871198562 | 0.871198562 | 0.197183459 | 0.032640165 | 0.499165124 |
| purple | 0.269830545 | 0.269830545 | 0.085091745 | 0.073966131 | 0.959520132 |
| magenta | 6.27E-12 | 6.27E-12 | 0.870651309 | 0.931772311 | 0.938519767 |
| turquoise | 0.732052642 | 0.732052642 | 0.002200962 | 0.556418105 | 3.85E-02 |
| blue | 0.951095794 | 0.951095794 | 0.549820958 | 0.551504548 | 0.218673854 |
| black | 0.98541025 | 0.98541025 | 0.07557608 | 0.430418871 | 0.378681232 |
| pink | 0.415147589 | 0.415147589 | 0.053357971 | 0.052508323 | 0.9959061 |
| red | 0.462596392 | 0.462596392 | 0.000320438 | 0.083933915 | 0.276041091 |
| yellow | 0.746168953 | 0.746168953 | 0.27145751 | 0.793268688 | 0.162695776 |
| green | 0.960124496 | 0.960124496 | 0.00403728 | 0.633341916 | 3.05E-05 |

**Table S5-5**: GO enrichment analysis of hub genes in MEmagenta module

| Accession | Name | GO Type | Q-value |
|---|---|---|---|
| GO:0016758 | transferase activity, transferring hexosyl groups | molecular_function | 0.0043 |
| GO:0015031 | protein transport | biological_process | 0.0043 |

**Table S5-6**: Hub genes in MEmagenta module

| GeneID | Correlation Coefficient | GeneID | Correlation Coefficient | GeneID | Correlation Coefficient |
|---|---|---|---|---|---|
| Ga02G0173 | 0.927649 | Ga07G1556 | 0.848364 | Ga10G1420 | -0.928319 |
| Ga03G1194 | -0.807986 | Ga07G1736 | -0.856951 | Ga10G1423 | -0.840714 |
| Ga03G1420 | -0.898496 | Ga07G2551 | 0.833754 | Ga10G2789 | -0.803682 |
| Ga03G1583 | 0.856747 | Ga07G2552 | 0.845366 | Ga10G3009 | 0.856736 |
| Ga03G1978 | -0.935624 | Ga07G2703 | -0.867092 | Ga11G0097 | 0.827208 |
| Ga03G2373 | -0.84596 | Ga08G0121 | 0.812143 | Ga11G0114 | -0.868638 |
| Ga03G2792 | 0.810602 | Ga08G1337 | 0.906449 | Ga11G0116 | -0.918689 |
| Ga04G0459 | 0.838359 | Ga08G1644 | -0.834424 | Ga11G0123 | -0.956435 |
| Ga04G1219 | 0.941649 | Ga08G2765 | -0.829319 | Ga11G0124 | -0.949664 |
| Ga04G1240 | 0.973989 | Ga09G1364 | 0.964927 | Ga11G1025 | 0.853337 |
| Ga05G2217 | -0.886669 | Ga09G2592 | -0.848502 | Ga11G2831 | 0.951245 |
| Ga05G3651 | 0.811511 | Ga09G2626 | 0.954313 | Ga11G2832 | 0.945251 |

| | | | | | |
|---|---|---|---|---|---|
| Ga05G3669 | -0.853329 | Ga10G0013 | 0.958049 | Ga11G3311 | 0.859207 |
| Ga05G4285 | -0.826726 | Ga10G0015 | 0.857779 | Ga12G1889 | -0.869984 |
| Ga05G4295 | -0.85 | Ga10G0363 | -0.930322 | Ga13G1157 | 0.828302 |
| Ga06G0475 | -0.883333 | Ga10G0390 | 0.875471 | Ga13G1914 | 0.82948 |
| Ga06G1754 | -0.929948 | Ga10G0621 | 0.916853 | Ga14G2119 | 0.944428 |
| Ga07G0944 | 0.823556 | Ga10G0844 | -0.839038 | Ga14G2437 | 0.824119 |
| Ga07G1117 | 0.85672 | Ga10G1419 | -0.868817 | | |

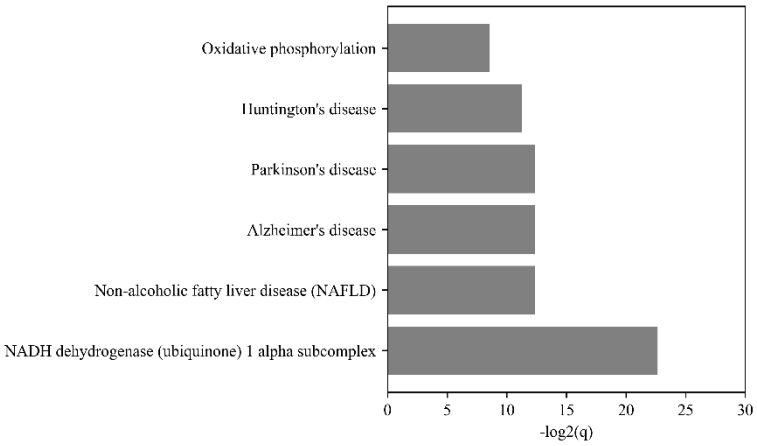



**Figure S5-1**: KEGG enrichment analysis of genes in the MEmagenta module.

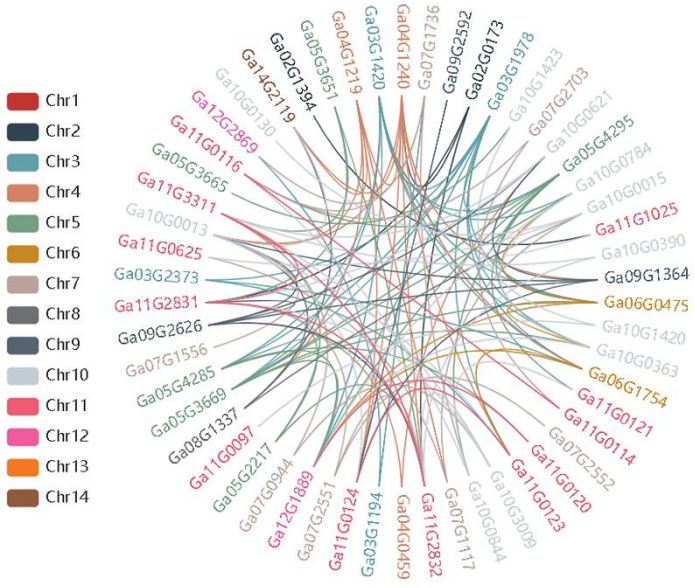



**Figure S5-2**: Diagram of interaction networks of genes in the MEmagenta module.

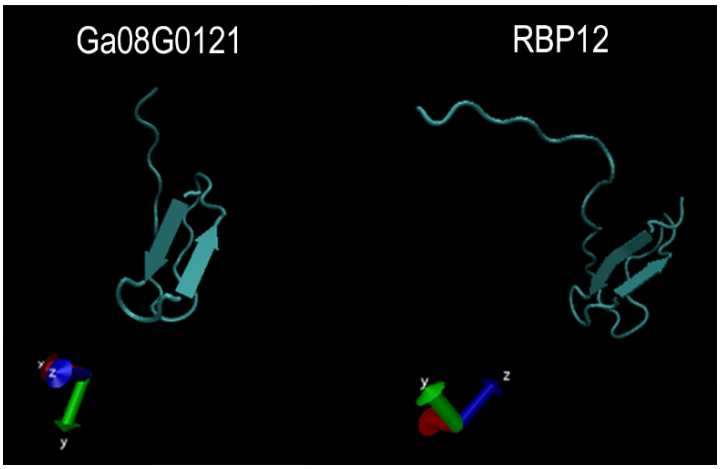



**Figure S5-3**: Comparison of protein structures between GaFZ and RPB12.

# 6

# Conclusion and future perspectives

# 1. General discussion

With the soaring development of high-throughput sequencing technology and the releases of multiple versions for cotton genome data, the strategy of combining the traditional mapping and high-throughput sequencing for map-based cloning has been widely used in forwarding genetics (Wang et al. 2012; Li et al. 2014; Li et al. 2015; Liu et al. 2015; Yuan et al. 2015; Zhang et al. 2015; Du et al. 2018; Ma et al. 2018; Hu et al. 2019; Wang et al. 2019). Several loci controlling fuzz development in tetraploid cotton lines have been identified or orientated to the small intervals, e.g., $N_1$, $n_2$, $n_3$, $n_4^t$, $N_5$ (Rong et al. 2005; Wan et al. 2016; Zhu et al. 2018; Chen et al. 2020; Wu et al. 2017; Naoumkina et al. 2021a; Naoumkina et al. 2021b; Zhu et al. 2021). In our study, we reported a newly discovered gene contributing to fuzz development in *Gossypium arboreum*. Based on the segregation ratio of two genetic populations ($F_2$ and $BC_1$) constructed by the wild-type DPL971 and its fuzzless mutant DPL972, we proved the fuzzless trait in *Gossypium arboreum* was controlled by a single dominant gene. Based on the genome data of *Gossypium arboreum* and *Gossypium hirsutum*, we developed and detected SSR and InDel markers tightly linked with the fuzzless gene and then narrowed the candidate gene to a small genomic interval. Compared to the wild-type, two genes in this candidate region showed notable upregulation in the growing fibers of the fuzzless mutant at the early fiber developing stages. In addition, polymorphisms that may lead to the changes of gene expression levels or the protein structure and functions in the promoter and genomic sequence were only detected in Cotton_A_11941 (showed as Ga08G0121 in the new version of genome data). Therefore, Ga08G0121 was annotated as the *GaGIR1* encoding gene and determined as the final candidate gene contributing to the fuzzless phenotype. However, it is important to note that our analysis of candidate genes within the mapping interval tightly depends on the quality and annotation level of the reference genomes. In the earlier mapping stage, we did not detect a clear linkage region based on the information of the draft A genome (*Gossypium arboreum* (version: A2_CGP-BGI_v2_a1)), but only after merging the AD genome data (version: AD1_NAU-NBI_v1.1_a1.1) from *Gossypium hirsutum* did we obtain the final mapping interval. There were differences of gene annotations between two genome resources: Ga08G0121 had no annotation in AD genome and was not predicted and identified as a gene. Since we detected the transcripts of this gene from our transcriptome data of diploid cotton lines DPL971 and DPL972, we believed that there is a gene missing in the genomic data of upland cotton. Therefore, combined with the annotated data of the two versions of the genome, we identified 7 candidate genes in the corresponding region. Both of these two versions of genome data were generated by paired-end sequencing through Illumina and Sanger, in addition, BAC end sequences and ultra-dense genetic maps were used to improve the assembly (Li et al. 2014; Zhang et al. 2015). The second generation sequencing has an obvious advantage of low-costing and a nonnegligible limitation of short reads. Currently, updated genomic data were obtained by the third generation sequencing such as PacBio SMRT (Du et al. 2018). Compared with the second generation sequencing, the third one demonstrated the remarkable advantages of ultra-long reads, non-GC-bias and the coverage uniformity (Rhoads and Au 2015; Xie et al. 2020). In addition, low-depth PacBio SMRT sequencing can also effectively detect the large-fragment genome-wide structural

variations (Wang et al. 2020). To a certain extent, it can correct the mis-assembly and improve the data integrity and accuracy. Besides, due to the big size and high complexity of the cotton genome, different versions of genome sequences and annotation are somewhat different, suggestive of the incomplete information of genome in each version. To obtain accurate results, it is better to compare multiple and high-quality genomic data during the fine mapping work and gene annotation so as to avoid the missing of important genes or structural variations.

In our study, we have detected SNPs in the promoter region of Ga08G0121 and speculated it would be the potential reason for the significantly different expression levels between the wild-type and the mutant (Feng et al. 2019). Liu et al. subsequently found some different SNPs or small InDels existing in the promoter region in diploid cotton, suggesting they would be allelic loci in these genomic sites (Liu et al. 2020). However, Wang et al. found and proposed that a distant and large insertion upstream of Ga08G0121 may act as an enhancer to regulate the gene expression. It would be the most direct factor rather than the SNPs and InDels detected from two materials to control the fuzz formation and develop (Wang et al. 2020). All these speculations need to be confirmed by subsequent transgenic experiments *in vivo*. Ga08G0121 was a newly discovered gene, and minor annotation or gene function has been elaborately described in cotton. Researches in Arabidopsis and tomato showed the orthologous genes of Ga08G0121 might be involved in root hair development and leaf trichome formation (Wu and Citovsky 2017a, b; Yang et al. 2011). Ectopic expression of Ga08G0121 in cotton showed that this gene could negatively regulate the formation and development of fuzz and trichome, with no obvious differences in root hair phenotype. We also found the glabrous leaves and stems in Ga08G0121_overexpressed lines. We observed and compared the root phenotypes of the wild type and the fuzzless mutant under the stereomicroscope and then found no significant difference between the two materials. Given the expression patterns of the family members of *GaGIRs*, only one gene showed preferential and differential expression in the developing root of two cotton lines. Most family members would be involved in fiber and fuzz development. These results indicated that the orthologs or the family members in diverse species might exhibit multiple functions during plants development. In previous studies, numerous genes have been reported to participate in fiber and fuzz development (Walford et al. 2011; Shan et al. 2014; Tian et al. 2020). However, we failed to detect the “star genes” differentially expressed in the diploid wild type and the fuzzless mutant. The studies conducted by Wang et al. found that the differentially expressed genes obtained according to results of the RNA-seq were mainly clustered and enriched in the process of VLCFA synthesis, suggesting that VLCFA may play important roles in the regulatory network of fiber and fuzz development in diploid cotton lines (Wang et al. 2020). But in our analysis, the hub genes, especially those differentially expressed hub genes, were predominately involved in the protein glycosylation pathway to control or regulate the fiber and fuzz initiation and development (Feng et al. 2021). It indicated the molecular mechanism of fuzz initiation and formation in diploid cotton lines may underly a newly discovered and unique pathway. As for the precise regulatory network, we need to further explore and verify it with in vitro and in vivo experiments in the future.

# 2. Main conclusion

## *2.1 The single dominant fuzzless gene GaFzl was fine-mapped on Chromosome A08 and GaGIR1 was identified as the candidate gene negatively responsible for the fuzzless phenotype in diploid cotton.*

In the present study, we combined the traditional map-based cloning and high throughput sequencing to fine map the fuzzless gene in diploid fuzzless mutant DPL972. Genetic analysis showed that the observed segregating ratio in each population fits the expected phenotypic segregation ratio, confirming the fuzzless phenotype in DPL972 was controlled by a single dominant gene. According to the association analysis of BSA-seq and genetic linkage analysis, polymorphic SSR and InDel markers were obtained and the candidate region was reduced to a 70-kb interval, including seven annotated genes. On the basis of re-sequencing and RNA-seq results, sequence variations of Cotton_A_11941, which encodes a homolog of GIR1 (GLABRA2-interacting repressor) in Arabidopsis thaliana, were detected in coding sequence and promoter region. Additionally, Cotton_A_11941 showed a notable upregulation in the fuzzless mutant DPL972. Therefore, Cotton_A_11941, namely *GaFzl*, was identified as the candidate gene responsible for the fuzzless phenotype in diploid cotton.

## *2.2 GL2-interacting-repressor (GIR) family members may contribute to fiber/fuzz formation via a newly discovered unique pathway in Gossypium arboreum.*

Here we revealed the basic features of *GaFzl* by determining the subcellular localization and analyzing its transcriptional activation of the encoded protein, suggesting *GaFzl* appeared to be co-localized in the nuclear and membrane of cells and had solid transcriptional activation activity in yeast. Based on the published cotton genomes, the *GIR* family genes were generally characterized in terms of their phylogenetic relationships, structures, chromosomal distribution, and evolutionary dynamics. Most members of the GaGIR family are tiny and may function as adapters or small molecular peptides to functionally interact with other proteins.

The *cis*-acting elements revealed that *GaGIR* genes likely have essential roles on plant growth and development via various signal transductions pathways, e.g., phytohormone, light, and abiotic stress responses. Expression pattern analysis discerned that most the family genes were expressed at high levels at fiber and fuzz developmental stages, whereas only Ga01G0231 was predominantly expressed in root samples. It suggested that *GaGIRs* mainly contributed to fuzz initiation and formation rather than root hair development. Furthermore, we detected the expression levels of crucial genes which were related to fiber and fuzz initiation both in diploid and tetraploid cotton lines. The results implied that *GaGIRs*, especially *GaFzl*, did not have any regulatory relationships with these genes and may regulate fiber and fuzz formation through an independent pathway from the known genes.

## *2.3 Fifty hub genes differentially expressed between two materials may contribute to fuzz initiation and development.*

In the present study, we re-analyzed gene expression profiles and regulatory networks by transcriptomic analysis in two *G. arboreum* accessions which differed in the fuzz characteristic. GO enrichment analysis presented that DEGs detected from the RNA-seq were mainly generated in terms of transcription factor activity, ATP synthesis and proton transport, and oxidoreductase activity. In addition, we identified a gene co-expression module through WGCNA to be highly correlated with the fuzz trait. According to the GO analysis, genes in this module were significantly enriched in transferase and transport protein items, indicating these proteins may play important roles in regulating fuzz initiation and development. Furthermore, twenty-eight hub genes in MEmagenta cluster1 and twenty-two hub genes in cluster2 exhibited significantly differential expression patterns between two accessions, suggestive of a pivotal role in the process of fuzz development. And the coding genes of FLA18, UGT and transport protein also presented remarkable differences in expression levels between two cotton lines, implying that they may be involved in regulating fuzz development by controlling the process of protein glycosylation.

In summary, our study revealed that the single dominant fuzzless gene *GaFzl* negatively regulated the trichome and fuzz formation in diploid cotton lines. *GaFzl* and its co-expression network were likely to participate in fiber and fuzz development via an independent pathway which remains to be discovered further. The present study unveiled and provided the crucial and potential genes underlying fuzz initiation and formation in *G. arboreum* and may facilitate the process of cotton molecular breeding to achieve the goal of fiber improvement.

# 3. Future perspectives

Fuzz fiber is an undesirable agronomic trait in cotton breeding. Therefore it is of great significance to clarify the regulatory mechanisms underlying fuzz formation and even create new fuzzless materials. This study demonstrates that the GaGIR1-encoding gene *GaFzl* negatively controls cotton leaf trichome and fuzz initiation and development. And the *GaFzl* may contribute to fiber/fuzz formation via a newly discovered unique pathway in *Gossypium arboreum*. Several genes co-expressed with *GaFzl* were putative to be involved in fuzz initiation and formation. However, the specific mechanism and regulatory network of *GaFzl* underlying the fuzz development process remains elusive. What we should mainly focus on in the future is to clarify the potential gene regulatory network contributing to fiber and fuzz development in diploid cotton species.

The yeast two-hybrid is the most commonly used method to identify interacting proteins. However, because of the strong self-activating of GaFzl, we failed to perform the regular library-screening in yeast. In our previous study, to clarify the function of *GaFzl* in fiber and fuzz development, we generated transgenic materials upregulated for *GaFzl* transcript levels through ectopic expression in tetraploid wild-type cotton line Hua1. The His-tag fused with *GaFzl* will provide us the possibility to explore the genome-wide identification of proteins interacting with GaFzl through CoIP-MS. Co-inmunoprecipitation (CoIP) is a classic technique to capture and enrich target proteins as well as interacting proteins or complexes in plant by using specific antibodies. Compared with yeast two-hybrid and pull down methods, one of its characteristics is

that the captured protein interactions occur in the specific tissue cells in plant, preserving the "endogenous" state of the interacting proteins or complexes. CoIP-MS is the combination of immunoprecipitation and mass spectrometry technology. The use of LC-MS/MS will be instrumental to identify proteins captured by Co-IP (Chen et al. 2016; Keller-Pinter et al. 2017; Jia et al. 2020).

Once the candidate protein(s) interacting with GaFzl is/are obtained, we could fuse them into the “bait” vectors and then construct the GaFzl into “prey” vector to verify the interactions between selected protein pairs by Y2H in vitro. To check the direct interaction of two proteins, GST pull-down assay could provide a useful solution in vitro. GST-GaFzl and flag-target fused protein could be obtained quickly in prokaryotic expression system. But, the protein from prokaryotic cells might be not proper represent the true characteristic of plant protein. In this case, BiFC could provide the way to confirm direct protein interaction in vivo. BiFC refers to that if two proteins have the interaction affinity, they will close their connected fluorescence protein fragment and form a complete functioning fluorescence protein, which is an effective method to examine the interaction of two known proteins in vivo as well. Combining these multiple approaches could obtain the accurate interaction proteins of GaFzl. (Bracha-Drori et al. 2004; Walter et al. 2004; Cao et al. 2020).

Additionally, we should also explore the upstream and downstream genes underlying the gene regulatory network of *GaFzl*, and strive to make clear how they function to participate in fiber and fuzz initiation and development. We can conduct a yeast one-hybrid to detect the potential proteins binding to the promoter region of *GaFzl*. The electrophoretic mobility shift assays (EMSA) and the dual-luciferase reporter assay are popular methods to detect the relationships between the DNA fragments and target proteins. And thus they could also be used to validate the upstream genes assisting with the Y1H.

Because we have generated the transgenic cotton lines for ectopic expression of *GaFzl* in Hua_1(*Gossypium hirsutum*, as the control line), we could design the whole transcriptome sequencing, including the mRNA and non-coding RNA. Once we get the differentially expressed genes, we may analyze the significantly up/down-regulated genes or the co-expressing genes with *GaFzl* in *GaFzl_OE* transgenic lines. This approach will likely provide the candidate genes downstream of *GaFzl*. And the Y1H and EMSA could be conducted for validation of interaction of GaFzl and target promoters (Muino et al. 2011; Kuang et al. 2017). Meanwhile, the non-coding RNAs have also reported to be involved in the regulation of fiber development (Wang et al. 2015a,b; Hu et al. 2018). Thus it is also worthy to deeply explore this potential important regulators.

Apart from the above, there is one more interesting point that deserves our attention. According to the previous study by Wang (Wang et al. 2020), a distant large insertion located about 18-kb upstream of *GaFzl* was uncovered and luciferase reporter assay has shown the complete and specific orientated sequence of this large insertion would activate the expression of reporter gene. In addition, they sequenced the transcriptomes of fuzzy and fuzzless accessions as well as the Arabidopsis lines overexpressing *GaFzl*. Based on the differentially expressed genes, they found most DEGs were enriched in the very long chain fatty acid pathway. The elevated

expression of *GaFzl* may repress the first step in fatty acid elongation and further impact the downstream steps of those pathways. In our previous studies, we did not notice and detect this enhancer, and we also failed to get the regulatory pathway of VLCFA from our transcriptome data of the same cotton lines. Therefore, we need to take this into account when we look at regulatory networks. We should better amplify the enhancer first in our cotton materials and validate its activating effect. And then, we could screen the *trans*-acting factors directly interacting with this enhancer to clarify how the enhancer regulate the expression of *GaFzl*. Meanwhile, we can re-analyze the whole transcriptome sequencing data and re-screening the DEGs and co-expressing genes with *GaFzl* to explore their potential regulatory network.

Cotton fiber and fuzz development undergo a quite intricate way, and lots of transcription factors are involved in this process. Assay for transposase accessible chromatin using sequencing (ATAC-Seq) is a relatively prominent and influential experimental approach for identifying the causable cis-acting elements that control the gene transcription process (Lu et al. 2017; Chen et al. 2016). And we have devised ATAC-seq to genome-widely detect the dynamic changes preferentially in open chromatin regions and thus explore the key transcription factors involved in the fuzz and fiber initiation and development program. Combining it with the whole transcriptome sequencing or the ChIP-Seq, we may better characterize the regulatory module of fiber and fuzz development.

A brief plan for the subsequent work could be as follows:

1) For the transgenic materials: We've obtained the transgenic lines with overexpressing *GaFzl* and observed the hairless phenotypes, but we haven't checked transgenic events and the expression levels at the protein level yet. Therefore, we need to analyze and validate the transgenic materials through approaches of southern blot (for the copy number) and proteomics (for the quantitative analysis) to ensure the successful process of subsequent experiments.

2), For the interaction proteins: We could collect samples of fiber-adhered ovules once the transgenic lines bloom to conduct the CoIP-MS experiment so as to screen the proteins or the complex interacting with GaFzl. And we have constructed the constructs of GaFzl for Y2H and BiFC. Based on the CoIP-MS results, candidate interaction proteins will be verified by Y2H and BiFC in vitro and in vivo, respectively.

3), For the upstream genes: We have segmentally amplified the promoter sequences (~2.5 kb before the transcriptional start sites) of *GaFzl*. A series of truncated *GaFzl* promoter luciferase reporter vectors will be constructed to find out the core promotor region. The region will be employed as bait of Y1H to screen the cDNA libraries derived from ovules at fiber initiation stage. Candidate upstream regulators of GaFzl could be obtained based on the Y1H results. And then, we could implement the EMSA to validate the upstream gene in vitro.

4), For the downstream genes: We have collected the samples of fiber-adhered ovules from the fuzzy DPL971 and fuzzless DPL972. We will also prepare the corresponding samples from the transgenic lines and the upland cotton Hua1 (WT). We plan to conduct the whole transcriptome sequencing and ATAC-Seq. After then, combining all the data together, we may find out the differentially expressed genes

and different chromatin accessibility in corresponding promoters. Open chromatin regions in promoter are likely to contain the motifs bound by regulators, which provide the information for predicting the interaction of GaFzl and target promoters.. By analyzing the genes co-expressing with *GaFzl*, we may predict the downstream genes and validate them by Y1H and EMSA. At the same time, we could identify new transcription factors and their target genes involved in fuzz initiation based on the RNA-seq and ATAC-seq results. Identification of more functional genes will improve the regulatory network underlying the fiber and fuzz development.

# References

Bracha-Drori K, Shichrur K, Katz A, Oliva M, Angelovici R, Yalovsky S, Ohad N (2004) Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. The Plant Journal 40 (3):419-427. doi:10.1111/j.1365-313X.2004.02206.x

Cao JF, Zhao B, Huang CC, Chen ZW, Zhao T, Liu HR, Hu GJ, Shangguan XX, Shan CM, Wang LJ, Zhang TZ, Wendel JF, Guan XY, Chen XY (2020) The miR319-targeted *GhTCP4* promotes the transition from cell elongation to wall thickening in cotton fiber. Mol Plant 13 (7):1063-1077. doi:10.1016/j.molp.2020.05.006

Chen R, Xiao M, Gao H, Chen Y, Li Y, Liu Y, Zhang N (2016) Identification of a novel mitochondrial interacting protein of C1QBP using subcellular fractionation coupled with CoIP-MS. Anal Bioanal Chem Feb;408(6):1557-64.

Chen W, Li Y, Zhu S, Fang S, Zhao L, Guo Y, Wang J, Yuan L, Lu Y, Liu F, Yao J, Zhang Y (2020) A retrotransposon insertion in *GhMML3_D12* is likely responsible for the lintless locus *li3* of tetraploid cotton. Front Plant Sci 11:593679. doi:10.3389/fpls.2020.593679

Chen X, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A, Doudna JA, Greenleaf WJ, Liphardt JT, Chang HY (2016) ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. Nature Methods 13 (12):1013-1020. doi:10.1038/nmeth.4031

Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M, Jia Y, Pan Z, Gong W, Liu Z, Zhu H, Ma L, Liu F, Yang D, Wang F, Fan W, Gong Q, Peng Z, Wang L, Wang X, Xu S, Shang H, Lu C, Zheng H, Huang S, Lin T, Zhu Y, Li F (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat Genet 50 (6):796-802. doi:10.1038/s41588-018-0116-x

Feng X, Cheng H, Zuo D, Zhang Y, Wang Q, Liu K, Ashraf J, Yang Q, Li S, Chen X, Song G (2019) Fine mapping and identification of the fuzzless gene *GaFzl* in DPL972 (*Gossypium arboreum*). Theor Appl Genet 132 (8):2169-2179. doi:10.1007/s00122-019-03330-3

Feng X, Liu S, Cheng H, Zuo D, Zhang Y, Wang Q, Lv L, Song G (2021) Weighted gene co-expression network analysis reveals hub genes contributing to fuzz development in *Gossypium arboreum*. Genes 12 (5):753

Hu H, Wang M, Ding Y, Zhu S, Zhao G, Tu L Zhang X (2018) Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (*Gossypium hirsutum L.*). Plant Biotechnol. J. 16, 1002–1012.

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 YL, Rahman MU, 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, Zhang T (2019) *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet 51 (4):739-748. doi:10.1038/s41588-019-0371-5

Jia J, Jin J, Chen Q, Yuan Z, Li H, Bian J, Gui L (2020) Eukaryotic expression, Co-IP and MS identify BMPR-1B protein-protein interaction network. Biol Res May 29;53(1):24.

Keller-Pinter A, Ughy B, Domoki M (2017) The phosphomimetic mutation of syndecan-4 binds and inhibits Tiam1 modulating Rac1 activity in PDZ interaction-dependent manner. PLoS One Nov 9;12(11):e0187094

Kuang J-F, Chen J-Y, Liu X-C, Han Y-C, Xiao Y-Y, Shan W, Tang Y, Wu K-Q, He J-X, Lu W-J (2017) The transcriptional regulatory network mediated by banana (*Musa acuminata*) dehydration-responsive element binding (*MaDREB*) transcription factors in fruit ripening. New Phytologist 214 (2):762-781. doi:10.1111/nph.14389

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J, Liang X, Huang G, Percy RG, Liu K, Yang W, Chen W, Du X, Shi C, Yuan Y, Ye W, Liu X, Zhang X, Liu W, Wei H, Wei S, Huang G, Zhang X, Zhu S, Zhang H, Sun F, Wang X, Liang J, Wang J, He Q, Huang L, Wang J, Cui J, Song G, Wang K, Xu X, Yu JZ, Zhu Y, Yu S (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol 33 (5):524-530. doi:10.1038/nbt.3208

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, Chen W, Liang X, Shang H, Liu W, Shi C, Xiao G, Gou C, Ye W, Xu X, Zhang X, Wei H, Li Z, Zhang G, Wang J, Liu K, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46 (6):567-572. doi:10.1038/ng.2987

Liu X, Moncuquet P, Zhu QH, Stiller W, Zhang Z, Wilson I (2020) Genetic identification and transcriptome analysis of lintless and fuzzless traits in *Gossypium arboreum* L. Int J Mol Sci 21 (5). doi:10.3390/ijms21051675

Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L, Zhou Y, Wang LJ, Guo WZ, Bai YL, Ruan JX, Shangguan XX, Mao YB, Shan CM, Jiang JP, Zhu YQ, Jin L, Kang H, Chen ST, He XL, Wang R, Wang YZ, Chen J, Wang LJ, Yu ST, Wang BY, Wei J, Song SC, Lu XY, Gao ZC, Gu WY, Deng X, Ma D, Wang S, Liang WH, Fang L, Cai CP, Zhu XF, Zhou BL, Jeffrey Chen Z, Xu SH, Zhang YG, Wang SY, Zhang TZ, Zhao GP, Chen XY (2015) *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Sci Rep 5:14139. doi:10.1038/srep14139

Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ (2017) Combining

ATAC-seq with nuclei sorting for discovery of *cis*-regulatory regions in plant genomes. Nucleic Acids Research 45 (6):e41-e41. doi:10.1093/nar/gkw1179

Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G, Yan Y, Jia Y, Yang J, Pan Z, Gu Q, Li X, Sun Z, Dai P, Liu Z, Gong W, Wu J, Wang M, Liu H, Feng K, Ke H, Wang J, Lan H, Wang G, Peng J, Wang N, Wang L, Pang B, Peng Z, Li R, Tian S, Du X (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat Genet 50 (6):803-813. doi:10.1038/s41588-018-0119-7

Muino JM, Hoogstraat M, van Ham RC, van Dijk AD (2011) PRI-CAT: a web-tool for the analysis, storage and visualization of plant ChIP-seq experiments. Nucleic Acids Res 39 (Web Server issue):W524-527. doi:10.1093/nar/gkr373

Naoumkina M, Thyssen GN, Fang DD, Bechere E, Li P, Florane CB (2021a) Mapping-by-sequencing the locus of EMS-induced mutation responsible for tufted-fuzzless seed phenotype in cotton. Mol Genet Genomics 296 (5):1041-1049. doi:10.1007/s00438-021-01802-0

Naoumkina M, Thyssen GN, Fang DD, Li P, Florane CB (2021b) Elucidation of sequence polymorphism in fuzzless-seed cotton lines. Mol Genet Genomics 296 (1):193-206. doi:10.1007/s00438-020-01736-z

Rhoads A, Au KF (2015) PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics Oct;13(5):278-89. doi: 10.1016/j.gpb.2015.08.002.

Rong J, Pierce GJ, Waghmare VN, Rogers CJ, Desai A, Chee PW, May OL, Gannaway JR, Wendel JF, Wilkins TA, Paterson AH (2005) Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theor Appl Genet 111 (6):1137-1146. doi:10.1007/s00122-005-0041-0

Shan C-M, Shangguan X-X, Zhao B, Zhang X-F, Chao L-m, Yang C-Q, Wang L-J, Zhu H-Y, Zeng Y-D, Guo W-Z, Zhou B-L, Hu G-J, Guan X-Y, Chen ZJ, Wendel JF, Zhang T-Z, Chen X-Y (2014) Control of cotton fibre elongation by a homeodomain transcription factor *GhHOX3*. Nature Communications 5:5519. doi:10.1038/ncomms6519

Tian Y, Du J, Wu H, Guan X, Chen W, Hu Y, Fang L, Ding L, Li M, Yang D, Yang Q, Zhang T (2020) The transcription factor MML4_D12 regulates fiber development through interplay with the WD40-repeat protein WDR in cotton. J Exp Bot 71 (12):3499-3511. doi:10.1093/jxb/eraa104

Walford SA, Wu Y, Llewellyn DJ, Dennis ES (2011) *GhMYB25-like*: a key factor in early cotton fibre development. Plant J 65 (5):785-797. doi:10.1111/j.1365-313X.2010.04464.x

Walter M, Chaban C, Schütze K, Batistic O, Weckermann K, Näke C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. The Plant Journal 40 (3):428-438. doi:10.1111/j.1365-313X.2004.02219.x

Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, Fang L, Yang S, Hu Y, Ye W, Zhang H, Ma P, Chen J, Wang Q, Mei G, Cai C, Yang D, Wang J, Guo W, Zhang W, Chen X, Zhang T (2016) Small interfering RNAs from bidirectional transcripts of

*GhMML3_A12* regulate cotton fiber development. New Phytol 210 (4):1298-1310. doi:10.1111/nph.13860

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet 44 (10):1098-1103. doi:10.1038/ng.2371

Wang M, Yuan D, Tu L, GaoW, He Y, Hu H, Wang P et al. (2015a) Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). New Phytol. 207, 1181–1197.

Wang M, Tu L, Yuan D, Zhu, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q, Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA, Zhang X (2019) Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet 51 (2):224-229. doi:10.1038/s41588-018-0282-x

Wang X, Zhao B, Roundtree, I.A, Lu Z, Han D, Ma H, Weng X et al. (2015b) N-6-methyladenosine modulates messenger RNA translation efficiency. Cell, 161, 1388–1399

Wang X, Miao Y, Cai Y, Sun G, Jia Y, Song S, Pan Z, Zhang Y, Wang L, Fu G, Gao Q, Ji G, Wang P, Chen B, Peng Z, Zhang X, Wang X, Ding Y, Hu D, Geng X, Wang L, Pang B, Gong W, He S, Du X (2020) Large-fragment insertion activates gene *GaFZ* (*Ga08G0121*) and is associated with the fuzz and trichome reduction in cotton (*Gossypium arboreum*). Plant Biotechnol J. doi:10.1111/pbi.13532

Wu H, Tian Y, Wan Q, Fang L, Guan X, Chen J, Hu Y, Ye W, Zhang H, Guo W, Chen X, Zhang T (2017) Genetics and evolution of *MIXTA* genes regulating cotton lint fiber development. New Phytologist. doi:10.1111/nph.14844

Wu R, Citovsky V (2017a) Adaptor proteins GIR1 and GIR2. I. Interaction with the repressor GLABRA2 and regulation of root hair development. Biochem Biophys Res Commun 488 (3):547-553. doi:10.1016/j.bbrc.2017.05.084

Wu R, Citovsky V (2017b) Adaptor proteins GIR1 and GIR2. II. Interaction with the co-repressor TOPLESS and promotion of histone deacetylation of target chromatin. Biochem Biophys Res Commun 488 (4):609-613. doi:10.1016/j.bbrc.2017.05.085

Xie H, Yang C, Sun Y, Igarashi Y, Jin T, Luo F (2020) PacBio Long Reads Improve Metagenomic Assemblies, Gene Catalogs, and Genome Binning. Front Genet Sep 8;11:516269.

Yang C, Li H, Zhang J, Luo Z, Gong P, Zhang C, Li J, Wang T, Zhang Y, Lu Y, Ye Z (2011) A regulatory gene induces trichome formation and embryo lethality in tomato. Proc Natl Acad Sci U S A 108 (29):11836-11841. doi:10.1073/pnas.1100532108

Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L, Li Y, Liang Q, Lin Z, Yang X, Liu N, Jin S, Lei Y, Ding Y, Li G, Ruan X, Ruan Y, Zhang X (2015) The genome sequence of Sea-Island cotton (*Gossypium barbadense*)

provides insights into the allopolyploidization and development of superior spinnable fibres. Scientific Reports 5 (1). doi:10.1038/srep17662

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, Hulse-Kemp AM, Wan Q, Liu B, Liu C, Wang S, Pan M, Wang Y, Wang D, Ye W, Chang L, Zhang W, Song Q, Kirkbride RC, Chen X, Dennis E, Llewellyn DJ, Peterson DG, Thaxton P, Jones DC, Wang Q, Xu X, Zhang H, Wu H, Zhou L, Mei G, Chen S, Tian Y, Xiang D, Li X, Ding J, Zuo Q, Tao L, Liu Y, Li J, Lin Y, Hui Y, Cao Z, Cai C, Zhu X, Jiang Z, Zhou B, Guo W, Li R, Chen ZJ (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nature Biotechnology 33 (5):531-537. doi:10.1038/nbt.3207

Zhu Q-H, Yuan Y, Stiller W, Jia Y, Wang P, Pan Z, Du X, Llewellyn D, Wilson I (2018) Genetic dissection of the fuzzless seed trait in *Gossypium barbadense*. Journal of Experimental Botany 69 (5):997-1009. doi:10.1093/jxb/erx459

Zhu QH, Stiller W, Moncuquet P, Gordon S, Yuan Y, Barnes S, Wilson I (2021) Genetic mapping and transcriptomic characterization of a new fuzzless-tufted cottonseed mutant. G3 (Bethesda) 11 (1):1-14. doi:10.1093/g3journal/jkaa042

# Appendix – Publications

1. **Feng, X.**; Cheng, H.; Zuo, D.; Zhang, Y.; Wang, Q.; Liu, K.; Ashraf, J.; Yang, Q.; Li, S.; Chen, X; Song, G. Fine mapping and identification of the fuzzless gene *GaFzl* in DPL972 (*Gossypium arboreum*). *Theoretical and Applied Genetics*. 2019, 132, 2169-2179.
2. **Feng X**, Liu S, Cheng H, Zuo D, Zhang Y, Wang Q, Lv L, Song G. Weighted gene co-expression network analysis reveals hub genes contributing to fuzz development in *Gossypium arboreum*. *Genes*. 2021; 12(5):753.
3. **Feng, X.,** Cheng, H., Zuo, D. et al. Genome-wide identification and expression analysis of GL2-interacting-repressor (GIR) genes during cotton fiber and fuzz development. Planta 255, 23 (2022).
4. Ali, M.; Cheng, H.; Soomro, M.; Shuyan, L.; Bilal Tufail, M.; Nazir, M.F.; **Feng, X.**; Zhang, Y.; Dongyun, Z.; Limin, L.; Wang, Q.; Song, G. Comparative transcriptomic analysis to identify the genes related to delayed gland morphogenesis in *Gossypium bickii*. *Genes* 2020, 11, 472.
5. Li, S., Zuo, D., Cheng, H., Ali, M., Wu, C., Ashraf, J., Zhang, Y., **Feng, X.**, Lin, Z., Wang, Q., Lv, L., Song, G.. Glutathione S-transferases *GhGSTF1* and *GhGSTF2* involved in the anthocyanin accumulation in *Gossypium hirsutum* L. *International journal of biological macromolecules*. 2020, 165(Pt B), 2565–2575.
6. Cheng H, **Feng X**, Zuo D, Zhang Y, Wang Q, Lv L, Wu C, Li S, Dai Y, Qu D, He M, Liu S, Song G. Gene expression correlation analysis reveals MYC-NAC regulatory network in cotton pigment gland development. *International Journal of Molecular Sciences*. 2021; 22(9): 5007.