Genetic mapping and transcriptomic characterization of a new fuzzless-tufted cottonseed mutant

Qian-Hao Zhu, Warwick Stiller, Philippe Moncuquet, Stuart Gordon, Yuman Yuan, Scott Barnes, and Iain Wilson

Abstract

Fiber mutants are unique and valuable resources for understanding the genetic and molecular mechanisms controlling initiation and development of cotton fibers that are extremely elongated single epidermal cells protruding from the seed coat of cottonseeds. In this study, we reported a new fuzzless-tufted cotton mutant (Gossypium hirsutum) and showed that fuzzless-tufted near-isogenic lines (NILs) had similar agronomic traits and a higher ginning efficiency compared to their recurrent parents with normal fuzzy seeds. Genetic analysis revealed that the mutant phenotype is determined by a single incomplete dominant locus, designated N₅. The mutation was fine mapped to an approximately 250-kb interval containing 33 annotated genes using a combination of bulked segregant sequencing, SNP chip genotyping, and fine mapping. Comparative transcriptomic analysis using 0–6 days post-anthesis (dpa) ovules from NILs segregating for the phenotypes of fuzzless-tufted (mutant) and normal fuzzy cottonseeds (wild-type) uncovered candidate genes responsible for the mutant phenotype. It also revealed that the flanking region of the N₅ locus is enriched with differentially expressed genes (DEGs) between the mutant and wild-type. Several of those DEGs are members of the gene families with demonstrated roles in cell initiation and elongation, such as calcium-dependent protein kinase and expansin. The transcriptome landscape of the mutant was significantly reprogrammed in the 6 dpa ovules and, to a less extent, in the 0 dpa ovules, but not in the 2 and 4 dpa ovules. At both 0 and 6 dpa, the reprogrammed mutant transcriptome was mainly associated with cell wall modifications and transmembrane transportation, while transcription factor activity was significantly altered in the 6 dpa mutant ovules. These results imply a similar molecular basis for initiation of lint and fuzz fibers despite certain differences.

Keywords: fuzzless-tufted cottonseed, fiber mutant, genetic mapping, near-isogenetic line, RNA sequencing

Introduction

Cotton fibers are elongated epidermal cells of cottonseeds. Based on their length and attachment to the seed coat after ginning, cotton fibers are classified into two categories: long lint fibers and short fuzz fibers. Lint fibers can be easily pulled off from the seed coat during the ginning process, whereas fuzz fibers remain attached to the seed coat after ginning. In Upland cotton (Gossypium hirsutum L.), lint fibers differentiate before anthesis, initiate at the day of anthesis and can elongate to an average length of ~30 mm, while fuzz fibers differentiate and initiate after initiation of lint fibers, usually at approximately 5 days post-anthesis (dpa), and can elongate only up to ~5 mm (Stewart 1975; Zhang et al. 2007). Fuzz fibers have low economic value and compete with lint fibers for energy and carbon source during their development.

Cotton gin energy costs have risen more than other gin operating costs. Energy audits conducted in US cotton gins representing a range of capacities between 2009 and 2013 revealed the average participating saw gin used between 34 and 40 kWh to process one cotton bale, i.e. 227 kg (Funk and Hardin 2012; Funk et al. 2013). While gins have become larger and more efficient again in the nearly 10 years since those audits, energy costs as a proportion of gin’s operating costs have become larger and now exceed >20% of the total cost of ginning cf. 6–10% in the 1970s (Funk and Hardin 2012). Ginning of seed cotton without fuzz fibers, or fuzzless, generally consumes less power and causes less damage to the lint fibers as much less force is required to remove the lint from the seed coat than fuzzy-seeded cotton (Bechere et al. 2009, 2012). Breeding cotton cultivars with fuzzless, or naked, seeds are one of the goals of many cotton breeding programs. To this end, it is essential to have fuzzless seed germplasm and to understand the genetic and molecular mechanisms controlling fuzz differentiation, initiation, and development.

Several natural and induced fuzzless cotton mutants have been reported in tetraploid cotton (G. hirsutum and G. barbadense). Genetic analysis has identified and assigned four fuzzless loci, including the dominant locus N₁, and the recessive loci n₂, n₃, and n₄ (for a review, see Fang et al. 2018). The N₁ mutation was first
reported in 1927 (Kearney and Harrison 1927). Both homozygous \(N_1N_1\) and heterozygous \(N_1n_1\) genotypes produce completely fuzzless seeds that lack any tuft at the micropylar tip of the seed and have a significantly reduced lint percentage (Ware 1940; Turley and Kloth 2002; Turley et al. 2007). The \(n_2\) mutation was documented in 1947 and also reported to have negative effect on lint percentage (Ware et al. 1947). The \(N_1\) and \(n_2\) mutations were assigned to a pair of homoeologous chromosomes, chromosome 12 (A12 of the A genomic subgenome) and chromosome 26 (D12 of the D subgenome), respectively (Endrizzi and Ramsay 1980). The recessive \(n_3\) locus was identified based on genetic analysis of the progeny derived from a cross between the \(N_1N_1\) and \(n_3n_3\) mutants as it is required for full expression of the fuzzless trait in accessions with a genotype of \(n_3n_3\) (Turley and Kloth 2002). Cotton plants with a genotype homozygous for all the three fuzzless loci, i.e. \(N_1N_1n_3n_3\), produce seeds that lack both lint and fuzz fibers, or fiberless, implying a role of these loci in the development of both lint and fuzz fibers. The fourth reported fuzzless mutation, \(n_4\), was induced by ethyl methanesulfonate (EMS) treatment. Seeds of the homozygous \(n_1n_1n_2\) mutant are partially naked with a tuft at the micropyre (Bechere et al. 2009, 2012). Unlike \(N_1\) and \(n_2\), the \(n_3\) mutation has been reported to have a less negative effect on lint percentage and is thus potentially more valuable in breeding fuzzless commercial cultivars (Bechere et al. 2009, 2012). The genomic locations of the \(n_3\) and \(n_4\) loci remain to be determined.

Thanks to the publication of cotton genomes and advances in high-throughput genotyping technologies, the gene responsible for the well characterized dominant \(N_1\) mutation has been determined to be the \(A_t\)-subgenome homoeolog of MYB25-like (i.e. MYB25L_A12). A noncoding transcript antisense to MYB25L_A12 was found at its 3’ end, forming double-stranded RNA to produce small-interfering RNAs (siRNAs) to post-transcriptionally silence MYB25L_A12. The siRNAs generated from the MYB25L_A12 locus are expected to be able to silence MYB25L_D12 as well given the high-level sequence similarity between the two homoeologs. Consequently, the expression level of MYB25-like (both MYB25L_A12 and MYB25L_D12) is very low in the developing ovules of the \(N_1\) mutant (Wan et al. 2016). MYB25-like is a master transcription regulator involved in initiation and development of both lint and fuzz fibers as transgenic plants with silenced MYB25-like produce fiberless cotton seeds, i.e. lacking both lint and fuzz fibers (Walford et al. 2011). Genetic and gene expression analyses of MYB25L_A12 and MYB25L_D12 in cotton accessions with variable lint and fuzz phenotypes imply the identity of \(n_2\) being a dysfunctional MYB25L_D12 allele (Zhu et al. 2018). It has been proposed that initiation and development of lint and fuzz fibers are regulated temporally by the two homoeologs of MYB25-like, with lint and fuzz fibers being mainly determined by MYB25L_A12 and MYB25L_D12, respectively (Zhu et al. 2018). In diploid \(G. \ arboreum\), MYB25-like (i.e. GaMYB25-like) was also found to be associated with fuzz development; however, in contrast to its major role in relation to fuzz development in tetraploid cotton, the role of GaMYB25-like in fuzz development in \(G. \ arboreum\) seems to be not as significant as GaGIR1 that has been proposed to be associated with fuzz development in \(G. \ arboreum\) based on investigations of several different fuzzless mutants (Feng et al. 2019; Liu et al. 2020). Identification of an additional \(n_3\) locus underlying the fuzzless phenotype in the \(n_3n_2\) genetic background (Turley and Kloth 2002) and the finding of the same genes involved in development of both lint and fuzz fibers (Zhu et al. 2018) hint the complexity of the genetic and molecular mechanisms underlying fiber development in tetraploid cotton. Consistent with this notion, it has been suggested that the fuzzless phenotype in Pima S-7 (\(G. \ barbadense\)) is associated with at least five loci with the \(n_2\) mutation being the major contributor (Zhu et al. 2018).

Forward and reverse genetics studies have generated a large body of information regarding genes and gene networks associated with cotton fiber development, but we still know little about why less than one-third of the ovule epidermal cells are differentiated to becoming lint fibers and what determines the fate of lint and fuzz fibers. To answer these questions and to enhance our understanding on the genetic networks responsible for initiation and development of fuzz fibers, and to provide tools for genetic improvement of lint fibers, it is essential to generate and characterize as many novel fiber mutants as possible. In this study, we reported a new fuzzless-tufted Upland cotton mutant with a lint percentage higher than that of the historic \(N_1\) and \(n_2\) mutants, investigated the genetics of the mutant and identified candidate genes responsible for the mutant phenotype. Comparative transcriptomic analysis uncovered biological and cellular functions associated with fuzz development and identified genes, with a known or unknown function in fiber development, potentially regulated by the fuzzless-tufted mutation.

**Materials and methods**

**Plant materials, mapping populations, and field experiments**

The fuzzless-tufted mutant (CSS386 or \(N_3\), \(G. \ hirsutum\)) with an unknown cause of the mutation is one of the germplasm in the CSIRO cotton collection. Sica V-2 (conventional) and Sicot 71BRF (with insect and herbicide-resistant transgenes) are commercial cultivars developed by the CSIRO cotton breeding program. MCU-5 is an Upland cotton variety introduced from India for its Fusarium wilt resistance (Lopez-Lavalle et al. 2012). Sicala V-2 and MCU-5 were used to generate the \(F_2\) and backcross populations used in genetic analyses, SNP genotyping, and bulked segregant sequencing. Sicot 71BRF was used as recurrent parent for generation of insect and herbicide-resistant fuzzless-tufted introgression line to enable it to be grown in a commercial field to produce seed cotton for the experiment of ginning efficiency comparison.

The fuzzless-tufted trait has been introgressed into the genetic background of Sicala V-2 using seven backcrosses to develop near-isogenic lines (NILs) with the mutant or normal fuzzy seed phenotype (Figure 1A). A population of 24 BC\(_{2F_4}\) and 16 BC\(_{2F_5}\) fuzzless-tufted NILs derived from Sicala V-2, together with 50 BC\(_{2F_3}\) or BC\(_{2F_4}\) fuzzless NILs derived from Pima S-7 were compared to the three parents (CSS386, Sicala V-2, and Pima S-7) in field experiments at the Australian Cotton Research Institute (Narrabri, New South Wales, Australia), following standard commercial growing practices. A high-volume instrument (HVI) system was used to determine the fiber properties. Three populations were used in genetic mapping: an \(F_2\) population (NM) with 86 individuals from CSS386 (\(N_3\) × MCU-5, an \(F_2\) population (NS1) with 1800 individuals from CSS386 (\(N_3\) × Sicala V-2, and a BC\(_{2F_2}\) population (NS) with 90 individuals from a cross between a fuzzless-tufted NIL line and Sicala V-2. These mapping populations were generated, grown, and phenotyped in a glasshouse in Canberra, Australia, at 28°C ± 2°C with natural lighting.

**Comparison of ginning efficiency**

The trials to compare the energy required to gin the fuzzless-tufted (one of the BC\(_{2F_5}\) NILs) and normal fuzzy (\(G. \ hirsutum\), cv. Sicot 71BRF) seed cotton were conducted at CSIRO...
Manufacturing (Geelong, Victoria, Australia). Trials were done using a single stand, cut down (119 to 40 saws) industrial scale Continental Eagle saw gin stand with a Continental/Moss Gordin Galaxie feeder with a maximum output of just under two bales/hr (400 kg/h). Two gin saw motor speeds (slow: 35 Hz/500 rpm and fast: 43 Hz/600 rpm) were assessed. Speeds were modulated using an ABB variable frequency drive attached to the 50 kW gin saw motor. No lint cleaning was used. For each genotype, three ginning runs (50 kg/run) were conducted under each motor speed. Current (ampere) was measured using a clamp on ammeter to one of the three-phase lines to the gin motor. Data was recorded using a HOBO data-logger at a frequency of one Hz. Measured amperes were converted to power values (kW) using the peak voltage (415 V) for a single phase and assuming a power factor of one. The kilowatt hours (kWh) for each run were determined by multiplying the time to gin 50 kg seed cotton by the average kW value for the run. All ginning runs were carried out under warm, dry ambient conditions (18–23°C and 30–50% relative humidity) with the fiber moisture content <5.5%.

CottonSNP63K array genotyping and bulked segregant sequencing

Four fuzzless-tufted, 11 intermediate and 9 normal fuzzy seed BC7F2:3 near-isogenic segregants were genotyped using the CottonSNP63K array as previously described (Hulse-Kemp et al. 2015; Zhu et al. 2016). To visualize the distribution of SNP frequency across chromosomes, the N5, Sicala V-2 and heterozygous alleles of each SNP were designed 1, 0, and 0.5, respectively, and plotted based on their chromosomal coordinates in the TM-1 reference genome (Zhang et al. 2015). Because only four fuzzless-tufted NILs were used, we combined them with the NILs with intermediate mutant phenotype when plotting the SNP frequency.
The region with a SNP frequency of 0.75 or close to 0.75 in the fuzzless-tufted and intermediate pool, and a SNP frequency of 0 or close to 0 (assuming all these NILs having the Sicala V-2 background) in the normal fuzzy seed pool was considered to be the candidate region containing the gene(s) responsible for the mutation. For bulked segregant sequencing, bulked DNA pool was prepared using equal amount of DNA isolated from 26 fuzzless-tufted or 22 normal fuzzy seed F3 individuals. Each DNA pool was sequenced at ~20x coverage of the G. hirsutum genome size. Short-reads mapping, SNP call, and plotting were performed as previously described (Zhu et al. 2018), except the reference genome, which was Sicala V-2 generated by mapping Sicala V-2 short-reads to TM-1 (Zhang et al. 2015).

**Fine genetic mapping**

We first genotyped 1800 F2 individuals from CSS386 (N1) × Sicala V-2 using KASP (Kompetitive allele-specific PCR) markers AM3 and AMS (Supplementary Table S1) located at the boundaries of the region identified based on SNP genotyping and bulked segregant sequencing. The analysis identified 276 segregants containing the region from the mutant. More KASP makers were then designed and used to genotype the identified 276 individuals. The results were used to construct a genetic linkage map and to perform QTL mapping using the software inclusive composite interval mapping (ICIM) (Li et al. 2017). The Kosambi mapping function was selected to convert a recombination frequency to genomic distance (cM). Linkage group and marker order were determined by using a logarithm of the odds (LOD) score of 15. QTL mapping was performed by using the ICIM-Add method of the ICIM software with 1000 permutation tests. Graphical representation was generated using MapChart (Voorrips 2002). KASP markers span the peak region were further used in genotyping of the NM and NS populations for co-segregation analysis.

**Kompetitive allele specific PCR genotyping**

The KASP reaction was performed in an 8 μl volume, comprising 1 μl DNA (15 ng), 4 μl 2× KASP master mix (LGCG Group), 0.11 μl primer mix (12 μM of each allele-specific primer and 30 μM of common primer) and 2.89 μl H2O. PCR cycling was performed on an Eppendorf Mastercycler ep384 using the following program: hot-start at 95°C for 15”, followed by 10 touchdown cycles (95°C for 20”, touchdown at 65–57°C, 0.8°C per cycle, 60”), and followed by 31 cycles of amplification at 94°C for 20” and 57°C for 60”. Plates were read on the Viia7 Real-Time PCR System (Life Technologies) at ambient temperature and analyzed using the Applied Biosystems software. If discriminating genotyping clusters had not formed after the above procedure, an additional three to six cycles of amplification were conducted and the plate was read and re-analyzed. KASP primers were designed based on polymorphic SNPs between the parental accessions (CSS386 and Sicala V-2) and are shown in Supplementary Table S1.

**DNA and RNA extraction**

Genomic DNA used in SNP genotyping (CottonSNP63K array and KASP) and re-sequencing was extracted from young true leaves using a DNeasy Plant Mini kit (Qiagen) according to the manufacturer’s instructions or the cetyl trimethylammonium bromide (CTAB) method with home-made solutions. Nanodrop 1000 (Thermo Scientific) was used in the quantification of DNA concentration. DNA samples used in next-generation sequencing were also quantified using Qubit (Invitrogen) and checked for integrity by gel electrophoresis.

Total RNAs were extracted from 0, 1, 2, 3, 4, 5, and 6 dpa ovules of fuzzless-tufted and normal fuzzy seed NILs using the Maxwell RSC Instrument (Promega) with a Maxwell RSC Plant RNA Kit. RNA quality was checked using the Agilent 2100 Bioanalyzer (Agilent Technologies) and Qubit (Invitrogen). RNA samples from 0, 2, 4, to 6 dpa with an RNA integrity number (RIN) above 7.0 were used in RNA-sequencing, while samples from all seven time points were used in qRT-PCR analysis. Each sample was analyzed with three biological replicates.

**Transcriptome sequencing and identification of differentially expressed genes**

RNA-sequencing was done using the paired-end (150 bp) configuration on an Illumina HiSeq 2000 instrument according to the manufacturer's instructions (Illumina). Raw reads were first processed using Trimmmomatic v0.39 (Bolger et al. 2014) to remove low-quality sequences and adaptors. The quality of trimmed FASTQ files was evaluated using FastQC v0.11.8. The clean reads were mapped to the TM-1 reference genome (Zhang et al. 2015) using Biokanga version 3.9.8 (https://github.com/csiro-crop-infor matics/biokanga). The expression levels of annotated genes were estimated using Transcript Per Million (TPM) mapped reads. TPM and unnormalized reads were created using in-house python scripts and HTSeq count (Anders et al. 2015), respectively. Differentially expressed genes (DEGs) were defined using DESeq2 (Love et al. 2014). Genes with an adjust P-value < 0.05 and an absolute value of Log2 (fold-change) ≥ 1 were considered as DEGs. Gene ontology (GO) enrichment analysis was performed using agriGO v2.0 based on the default settings (Tian et al. 2017). Heatmaps were generated using ggplot2 of the R package (https://www.r-project.org/).

**Quantitative real-time reverse transcription PCR (qRT-PCR)**

Total RNAs of 0 to 6 dpa whole ovules were used in the analysis of gene expression levels by qRT-PCR. Reverse transcription was performed using 2 μg of RQ1 DNase (Promega) treated total RNA, random hexamer, and SuperScript III reverse transcriptase (Invitrogen) in a 20 μl reaction. The first-strand cDNA reaction was diluted 10 folds and 4 μl of the diluted cDNA was then used as template in a 10-μl reaction with the FasterStart Universal SYBR Green Master Mix (ROX) (Roche). Negative controls with total RNA were performed to eliminate the possibility of genomic DNA contamination of the RNA samples. Relative gene expression levels were determined based on three biological replicates, each with three technical replicates using the 2^{-ΔΔCt} approach and tested by Student’s t-test for significance. The cotton ubiquitin gene (GenBank accession no. EU604080) was used as the reference gene for normalization. The reactions were run on the Viia7 Real-Time PCR System (Life Technologies). The primers used in qRT-PCR are shown in Supplementary Table S1.

**Analysis of promoter cis-elements**

One thousand base-pair sequence upstream of the start codon (ATG) of each gene was retrieved and submitted to the online database PLACE to identify cis-regulatory elements (Higo et al. 1999). The number of cis-elements that were found in the promoters of Gh_D13G0996 and/or Gh_D13G0998 was tallied separately for the regions of −1 to −500 bp (representing the core promoter) and −501 to −1000 bp.
**Gene cloning and analysis**

Total RNAs (2 μg) of TM-1 and N5 isolated from 5 dpa ovule using a Maxwell RSC Plant RNA Kit were reverse transcribed by oligo(dT) using the Superscript III first-strand synthesis system (Invitrogen) in a 20 μl reaction according to the manufacturer’s instruction. An aliquot of 2 μl cDNA reaction was then used in the amplification of the cDNAs using a Phusion™ High-Fidelity PCR Kit (New England Biolabs) with primers matching the start and end of Gh_D13G0096 or Gh_D13G0098. PCR products were cut from gel, purified using the QIAquick Gel Extraction Kit, and ligated to the blunt DNA cloning vector of the Zero Blunt™ PCR cloning kit (Invitrogen). Ten clones were sequenced to deduce the cDNA sequence of each gene.

**Data availability**

All RNA-sequencing file and metadata information can be found at https://doi.org/10.25919/5ee196cce855a.

Supplementary material is available at figshare DOI: https://doi.org/10.25387/g3.13345268.

**Results**

**Characterization of the fuzzless-tufted cottonseed mutant**

The mutant line (CSS386) used in this study produces seeds that are fuzzless but with a tuft grown on the micropylar end (Figure 1B). Compared to the recurrent parent Sicala V-2, a conventional commercial cotton cultivar used previously in Australia cotton production, the mutant has a lower lint percentage and poorer fiber quality. After seven backcrosses, lint yield, lint percentage, and fiber quality of the advanced fuzzless-tufted NILs were equivalent to or even better than that of the recurrent parent. Compared to the fuzzless NILs derived from Pima S-7 x Sicala V-2, in which the fuzzless trait is determined by a recessive mutation, the advanced fuzzless-tufted NILs also had a higher lint yield and lint percentage although with shorter fibers (Table 1). The ginning efficiency of a fuzzless-tufted introgression line was compared with that of the recurrent parent Sicot 71BRF. It showed that the fuzzless-tufted introgression line used ~10% and 14% less power under the slow and fast gin saw speed, respectively (Table 2). These results suggest that the fuzzless-tufted trait by itself has no negative effect on the performance of lint yield and fiber quality, and that it has the potential to save energy required for ginning seed cotton. The fuzzless-tufted trait is thus a valuable resource and is superior to the fuzzless trait from Pima S-7 for developing fuzzless commercial cotton cultivars.

**Inheritance of the fuzzless-tufted cottonseed trait**

To investigate the genetics underlying the fuzzless-tufted mutant trait, we generated two populations, one F2 population from CSS386 × MCU-5 (i.e. NM) and a BC2F2 (i.e. NS) population derived from a cross between a near-isogenic mutant line and the recurrent parent Sicala V-2. In both populations, F2:3 seeds were all tufted on the micropylar end but showed variable amount of fuzz on other parts of the seed coat. Based on the amount of fuzz, we separated the seeds into five groups, and considered groups 2 and 3 as intermediate and groups 4 and 5 as fuzzless (Figure 1B). For genetic analysis, segregants showing groups 2–5 seed phenotype were designated fuzzless-tufted. Using this criterion, the segregation ratio of fuzzless-tufted and fuzzy in both populations fits a single segregating locus model (Table 3). Due to the

| Genotype | No. of NILs | Lint yield (kg/ha) | Lint (%) | Fiber length (mm) | Uniformity (%) | Short fiber index | Fiber strength (g/tex) | Micronaire |
|----------|-------------|-------------------|---------|------------------|---------------|------------------|----------------------|-----------|
| Sicala V-2 (recurrent parent)×CSS386 (fuzzless-tufted mutant) | 24 | 2303 (1786-2615) | 41.20 (39.40–42.75) | 30.23 (28.96–32.77) | 84.15 (81.20–86.25) | 9.27 (8.10–11.10) | 32.25 (30.85–34.19) | 4.01 (3.15–4.52) |
| Fuzzless-tufted NILs derived from CSS386 (BC1F2) | 16 | 2551 (2292–2738) | 40.48 (39.00–41.70) | 30.23 (29.46–31.10) | 84.19 (82.30–86.63) | 7.94 (6.63–9.50) | 33.47 (32.02–34.63) | 4.09 (3.53–4.55) |
| Fuzzless-tufted NILs derived from CSS386 (BC2F2) | 47 | 2094 (1660–2547) | 38.31 (36.83–41.10) | 31.01 (29.21–32.51) | 84.65 (83.10–86.13) | 7.88 (7.00–9.70) | 32.91 (30.57–35.70) | 4.36 (3.83–4.67) |
Table 2 Comparison of mean current trace statistics and power consumption between ginning the fuzzless-tufted and fuzzy seed cotton at two different motor speeds

| Cotton seed phenotype | Seed cotton ginned/run (kg) | Motor speed of gin saw (Hz/rpm) | Current (ampere) | Power value (kW) | Ginning time (second) | Power consumed (kWh) |
|-----------------------|----------------------------|--------------------------------|-----------------|-----------------|-----------------------|----------------------|
| Fuzzless-tufted       | 50                         | 35/500                         | 14.91           | 6.19            | 255                   | 0.438 (–10.06%)      |
| Fuzzy                 | 50                         | 35/500                         | 14.06           | 5.92            | 297                   | 0.487                |
| Fuzzless-tufted       | 50                         | 43/600                         | 15.37           | 6.38            | 267                   | 0.472 (–14.03%)      |
| Fuzzy                 | 50                         | 43/600                         | 16.41           | 6.81            | 290                   | 0.549                |

Table 3 Genetic analysis of the fuzzless-tufted cottonseed trait

| Population         | Female                   | Male                     | Fuzzless-tufted | Fuzzy | \(\chi^2\) (3:1, 1:3:4) | P-value |
|--------------------|--------------------------|--------------------------|-----------------|-------|--------------------------|---------|
| NM (F2)            | CSS386 (Ns)              | MCU-5                    | 62              | 24    | 0.42                     | 0.51    |
| NS (BC3F2)         | CSS386 (Ns) - NIL        | Sicala V-2               | 64              | 26    | 0.72                     | 0.39    |

Presence of intermediate mutant phenotypes, we propose that the fuzzless-tufted mutant trait is controlled by a single locus with incomplete dominance and designate the locus (or gene) responsible for the mutation \(N_5\).

Identification of genetic region associated with the fuzzless-tufted cottonseed trait

Due to the nature of the incomplete dominance of the mutation, we reasoned that NILs showing the mutant or normal fuzzy seed phenotype would be the best choice for genetic mapping as their genetic background should have almost the same effect on the mutant phenotype. To this end, we introgressed the mutation into Sicala V-2 by seven backcrosses, and genotyped 4, 11, and 9 BC3F3 NILs showing fuzzless-tufted, intermediate, and normal fuzzy seed phenotype, respectively, using the CottonSNP63K array (Figure 1A). Of the ~63 K SNPs, 5456 are polymorphic between the mutant and Sicala V-2, 4802 of them are distributed across the 26 cotton chromosomes (Supplementary Table S2). As expected, all chromosomes, except an ~1.5 Mbp region on chromosome D13, of the fuzzless-tufted and intermediate NILs have been replaced by those from Sicala V-2 (Figures 1C and Supplementary Figure S1). The D13 region is thus where the mutation is most likely located. This was confirmed by bulked segregant sequencing of two DNA pools prepared using fuzzless-tufted or normal fuzzy seed segregants of a BC3F2 population (Figures 1D and Supplementary Figure S2).

Fine mapping the genetic locus responsible for the fuzzless-tufted cottonseed trait

To fine map the mutation responsible for the fuzzless-tufted trait, we developed an F2 population \([N5 = CSS386 (Ns) \times Sicala V-2]\) with 1800 individuals. We first genotyped all 1800 F2 individuals using two KASP markers (AM3 and AM5) located at the boundary of the D13 region to identify 276 segregants that showed recombination in the D13 region, i.e. the two markers had different origin (CSS386 or Sicala V-2). The identified segregants were then genotyped using another 13 KASP markers located between AM3 and AM5. Quantitative trait locus (QTL) analysis was done using the genotyping results and the pheno-typing data of F2,3 seeds, which identified four markers (AM13, AM83, AM74, and AM76) to be closely linked with the mutant phenotype. The genetic interval defined by AM13 and AM76 is ~275 kb (Figure 2A). These four markers and several flanking markers were then used to genotype the two populations (i.e. NM and NS) used in genetic mapping. Together, the genotyping and phenotyping results allowed us to further narrow down the causative region to an ~249 kb region, between markers AM83 and AM76 (Figure 2B). We noticed that, in all three segregating populations, some heterozygous plants, i.e. with a genotype of \(N_5N_5\), produced fuzzless-tufted seeds, but all plants with intermediate fiber phenotype were heterozygous for the \(N_5\) locus, consistent with the nature of incomplete dominance of the mutation.

Candidate genes underlying the fuzzless-tufted seed trait

The genetic interval between AM83 and AM76 has 33 annotated genes in the TM-1 genome (Zhang et al. 2015). To identify candidate gene(s) responsible for the fuzzless-tufted mutant phenotype, we did RNA-seq using RNAs isolated from 0, 2, 4, to 6 dpa ovules of the mutant and normal fuzzy seed NILs (hereafter, WT). Of the 33 genes located at the candidate region, 24 were expressed but only two \(Gh_D13G0096\) and \(Gh_D13G0098\) were differentially expressed between the mutant and WT in at least two of the four time points (Table 4), making them the best candidate genes contributing to fuzz initiation and development. Both \(Gh_D13G0096\) and \(Gh_D13G0098\) were upregulated in the mutant ovules, which was confirmed by qRT-PCR (Figure 3, A and B). In TM-1, these two genes are mainly expressed in ovule and fibers. \(Gh_D13G0096\) has a much higher expression level in 10–25 dpa ovules and 25 dpa fibers than in other tissues. \(Gh_D13G0098\) is mainly expressed in ~3 to 10 dpa ovules (Supplementary Figure S3). \(Gh_D13G0096\) encodes a hypothetical protein without known conserved domain, while \(Gh_D13G0098\) is closely related to cutonic amino acid transporter 7 (CAT7) and CAT6 of Arabidopsis thaliana, members of the amino acid-polyamine-choline (APC) transporter family.

No genomic sequence difference was found between the mutant and TM-1 in both \(Gh_D13G0096\) and \(Gh_D13G0098\). For \(Gh_D13G0098\), its flanking 1-kb regions were also identical between TM-1 and the mutant. Compared to TM-1, the mutant has three indels in the promoter of \(Gh_D13G0096\) (a single base insert between positions 215 and 216 counting from the start codon) or downstream of the gene (a single base insert between positions 653 and 654, and a deletion of two bases between positions 807 and 808, counting from the stop codon) (Supplementary Figure S4). For both genes, it is unclear what causes their up-regulation in the mutant, although for \(Gh_D13G0096\) the possible contribution of the three indels located at up- or down-stream the coding sequence could not be ruled out.
annotated first intron (71 bp) is retained in the cDNAs of both TM-1 and the mutant, which introduces a pre-mature stop codon at the beginning of the second exon. As a result, the cloned cDNAs contain two open reading frames, ORF1 and ORF2 that are equivalent to Ghir_D13G001110 and Ghir_D13G001100, respectively, in the TM-1 reference genome reported by Huazhong Agricultural University (Gh_HAU; Wang et al. 2019). ORF1 is identical to Ghir_D13G001110 but ORF2 and Ghir_D13G001100 have different starts and ends (Supplementary Figures S5 and S6A).

Despite the difference observed between the annotated Gh_D13G0098 (Gh_NAU) and its counterparts in Gh_HAU, Gh_ZJU, Gh_WHU, and Gh_JGI, all 10 cDNA clones from TM-1 or the mutant were identical, i.e. with the retained 1st intron, implying that the alternative splicing event is unlikely to be the cause of the mutation.

The annotation difference observed between Gh_D13G0098 (Gh_NAU) and its counterparts in Gh_HAU prompted us to compare Gh_D13G0096 and Gh_D13G0098 as well as their flanking genes with their counterparts in the five published TM-1 genomes (Zhang et al. 2015; Hu et al. 2019; Wang et al. 2019; Chen et al. 2020; Huang et al. 2020). For Gh_D13G0098 (Gh_NAU), except Gh_HAU (Wang et al. 2019), an identically annotated gene was found in Gh_ZJU (Hu et al. 2019), Gh_WHU (Huang et al. 2020), and Gh_JGI (Chen et al. 2020). Gh_D13G0096 and its counterparts were found in all five TM-1 genomes, but interestingly two and three tandem copies of the gene were annotated in the Gh_WHU and Gh_JGI versions of the TM-1 genome (Supplementary Figure S6B; Huang et al. 2020; Chen et al. 2020). Most genes flanking Gh_D13G0096 and Gh_D13G0098 have the same annotation in different versions of the TM-1 genome. The discrepancy in annotation of Gh_D13G0096 and a few other genes in the N5 locus implies the difficulty involved in assembly and annotation of the region possibly due to copy number variation and/or other unknown structural variation(s). We re-sequenced the mutant genome using the second-generation sequencing technology, but it is currently impossible to know if the mutant is different from TM-1 in genome structure and/or gene copy number at the N5 locus.

We also checked coding sequence variations for all other annotated genes at the N5 locus. Nucleotide variations (SNPs and indels) causing amino acid changes in the N5 mutant were found only in two genes (Gh_D13G0089 and Gh_D13G0090) that were not differentially expressed between the mutant and WT. Two and one amino acid substitution was observed in Gh_D13G0089 and Gh_D13G0090, respectively (Supplementary Figure S7).

The flanking region of the N5 locus is enriched with differentially expressed genes

Overall, transcriptome analysis identified 740, with 649 non-redundant, genes differentially expressed between the mutant and WT in 0–6 dpa ovules. Of which, 158 (133 non-redundant) and 582 (516 non-redundant) were up- and down-regulated in the mutant, respectively (Supplementary Table S3). The majority DEGs were found at 6 dpa (63%), followed by 0 (25%), 4 (8%), and 2 (4%) dpa (Figure 4A). Most DEGs (90%) were unique to a single

![Figure 2](https://example.com/figure2.png)
Table 4 Expression changes [\(\log_{2}(M/N)\)] of the annotated genes in the \(N_{5}\) locus

| Gene_ID        | Start | End   | Functional annotation                                  | 0 dpa | 2 dpa | 4 dpa | 6 dpa |
|----------------|-------|-------|--------------------------------------------------------|-------|-------|-------|-------|
| D13G0078       | 847797| 849821| DC1 domain-containing protein                           | 0.29  | 0.39  | 0.22  | 0.3   |
| D13G0080       | 853400| 854761|                                                        | 0.19  | 0.19  | 0.15  | 0.1   |
| D13G0081       | 861848| 863553| Cytochrome b–c1 complex                                 | -0.07 | -0.15 | 0     | 0.17  |
| D13G0082       | 865034| 867466| Protein serine/threonine kinases                        | -0.92 | 1.17  | -1.87 | 1.12  |
| D13G0083       | 867592| 868265| Protein of unknown function (DUF1118)                   | -0.43 | 0.04  | 0.27  | -0.73 |
| D13G0085       | 873770| 877019| GroES-like zinc-binding alcohol dehydrogenase family protein | 0.18  | -0.1  | -0.38 | -0.35 |
| D13G0086       | 877418| 879560| Alpha/beta-Hydrolases superfamily protein              | -0.08 | 0.06  | -0.01 | -0.15 |
| D13G0087       | 881435| 884262| Transducin/WD40 repeat-like superfamily protein         | -0.03 | 0.08  | -0.05 | -0.07 |
| D13G0088       | 887905| 891110| Tetratricopeptide repeat (TPR)-like superfamily protein | 0     | 0.05  | 0.03  | 0.03  |
| D13G0089       | 894002| 901520|                                                        | -0.18 | -0.01 | -0.06 | -0.16 |
| D13G0090       | 902473| 913155| Sec23/Sec24 protein transport family protein            | 0.02  | 0.18  | 0.26  | 0.24  |
| D13G0091       | 918387| 924132| DNAJ heat shock N-terminal domain-containing protein    | -0.15 | -0.12 | -0.16 | -0.07 |
| D13G0092       | 927208| 927526|                                                        | -0.06 | -0.38 | 0.21  | -0.33 |
| D13G0093       | 928840| 932595| Nuclear transport factor 2 family protein              | -0.22 | 0.58  | 0.58  | -2.08 |
| D13G0094       | 944433| 947011| Double Clp-N motif-containing P-loop nucleoside triphosphate hydrolases superfamily protein | -0.21 | 0.31  | -0.14 | -0.26 |
| D13G0095       | 959019| 967550| Pentatricopeptide repeat (PPR) superfamily protein      | -0.02 | 0.13  | 0.07  | 0.05  |
| D13G0096       | 971547| 972211|                                                        | 1.96a | 2.56a | 3.23a | 2.72a |
| D13G0098       | 987476| 990611| Cationic amino acid transporter 7                        | 0.96  | 1.09a | 0.85  | 1.11a |
| D13G0101       | 999015| 1002839| Translocon at the outer envelope membrane of chloroplasts | 0.06  | 0.22  | 0.12  | 0.04  |
| D13G0102       | 1008601| 1012317| Translocon at the outer envelope membrane of chloroplasts | 0.12  | 0.03  | 0.03  | 0.06  |
| D13G0103       | 1018928| 1022750| ACT domain-containing protein                            | -0.14 | -0.19 | -0.07 | -0.13 |
| D13G0104       | 1024398| 1025657| Trichome birefringence-like 36                          | -0.16 | -0.08 | -0.39 | -0.52 |
| D13G0105       | 1029310| 1032443| Regulatory particle triple-A ATPase 3                   | -0.04 | 0.12  | 0.1   | 0.13  |
| D13G0106       | 1033701| 1042658|                                                        | -0.35 | 0.42  | -1.15 | -1.74 |

* Significantly differentially expressed between the mutant and WT.

time point and only 46, 9, and 9 nonredundant DEGs were overlapping between 2, 3, and all 4 time points, respectively (Figure 4, B and C). Of the 649 non-redundant DEGs, 29 were on chromosome D13 and 14 of the 29 D13 DEGs were located at the \(N_{5}\) locus and its 10-Mbp flanking regions (Figure 4D), including 5 (\(Gh\_D13G0033, Gh\_D13G0037, Gh\_D13G0096, Gh\_D13G0139,\) and \(Gh\_D13G0152\)) of the nine nonredundant genes that were differentially expressed at all four time points, and 4 (\(Gh\_D13G0060, Gh\_D13G0098, Gh\_D13G0114,\) and \(Gh\_D13G0199\)) of the 46 non-redundant genes that were differentially expressed at two of the four time points. Amongst them were two (\(D13G0033\) and \(D13G0114\)) related to \(Ca^{2+}\) signaling and one (\(D13G0060\)) being a member of the expansin family. \(Gh\_D13G0033\) or \(Gh\_CPK94\) (Gao et al. 2018) is closely related to calcium dependent protein kinase 1 (AtCPD1K) of \(A\). \(thaliana\) and was significantly down-regulated at all four time points. Similarly, \(D13G0114\), an IQ-domain containing gene, was also down-regulated at all four time points, particularly at the 4 dpa mutant ovules. In WT ovule, the expression level of \(Gh\_D13G0060\) or \(Gh\_EXPASI\) (Lv et al. 2020), a homolog of \(A\). \(thaliana\) Expansin A8 (AtEXP8), increased almost 30 times in ovules from 0 dpa (~10 TPM) to 6 dpa (~300 TPM), with a surge observed from 4 dpa (~90 TPM) to 6 dpa. By contrast, the expression level of \(Gh\_D13G0060\) was very low (~6 TPM) in the mutant ovules, particularly in the 4 and 6 dpa ovular ovules (Supplementary Table S3, Figure 4D). Intriguingly, the expression level of \(Gh\_D13G0139\), a homolog of Arabidopsis senescence associated gene 20 (AtSAG20), was quite low (~40–70 TPM) in the 0–6 dpa WT ovules, but its expression level was extremely high (6600–13,000 TPM) in the mutant ovules, consequently, \(Gh\_D13G0139\) was the most significantly up-regulated gene at all four time points in the mutant ovules (Figure 4D).

Enrichment of DEGs at the \(N_{5}\) locus and its flanking regions indicates the existence of a cluster of co-regulated and functionally related genes associated with the mutant phenotype. A similar phenomenon has been reported for the genomic region with the Lign intless-2 mutation (Patel et al. 2020). One possibility for co-regulation of functionally related genes is the presence of shared cis-regulatory elements in their promoters. Indeed, a number of cis-regulatory elements were found to be common in the promoters of these co-regulated genes (Supplementary Table S4). Alternatively, these co-regulated genes may belong to a large chromosomal domain regulated by the same chromatin conformation (Zhan et al. 2006).

Transcriptome reprogramming in the mutant

To investigate the biological function of the DEGs, we searched enriched gene ontology (GO) terms for the up- or down-regulated DEGs identified at each time point. For the 35 and 13 upregulated DEGs identified at 0 and 2 dpa, respectively, no significantly enriched GO term was found. The 15 upregulated DEGs at 4 dpa were enriched for a single GO term, multi-organism process. For the 95 up-regulated DEGs identified at 6 dpa, enriched GO terms included cell wall modification, single-organism cellular process, and ion transmembrane transport (Supplementary Table S5). For the 149, 19, 43, and 371 down-regulated DEGs identified at 0, 2, 4,
and 6 dpa ovules, respectively, no significantly enriched biological process GO term was found for the 2 and 4 dpa DEGs. At 0 dpa, the down-regulated DEGs were enriched with genes related to cellular responses to multiple stimuli, including lipid, oxygen-containing compound, and abscisic acid (ABA). The down-regulated DEGs identified at 6 dpa were not only enriched with genes related to responses to ABA, oxidative reaction, and lipid metabolism, but also with genes responses to water deprivation, ethylene (ET), and jasmonic acid (JA). Regarding molecular function, the down-regulated DEGs found at 0 dpa were enriched with genes involved in transmembrane transporter activity, such as nitrate transporter, ABC-2 type transporter, and sodium/calcium exchanger, while at 6 dpa the down-regulated DEGs were enriched with genes related to transcription factor activity. Only three transcription factors (TFs) were up-regulated in the mutant ovules at 6 dpa, but a total of 53 TFs, including multiple members of the MYB, NAC, bHLH, ERF, HD-ZIP, and LBD families, showed a significantly lower expression level in the mutant ovules than in the ovules of the fuzzy seed NILs (Figure 5). Of the 10 members of the R2R3-MYB subgroup MYBMIXTA-like, or MML, 4 were significantly down-regulated in the 6 dpa mutant ovules, including MYB25-like (GhMML3), GhMML4, GhMML8, and GhMML9 (Figures 3C and 5). MYB25L-D12 that has been shown to be the major gene determining fuzz development in G. barbadense (Zhu et al. 2018) had a lower expression level in the 1–6 dpa mutant ovules (Figure 3D). Other genes reported to be related to fiber initiation and elongation, such as GhHOX1, GhHOX2, GhHOX3, GhVIN1, GhJA22, and GhCaM7 all were not differentially expressed between the mutant and WT, although GhHD1 was marginally significantly down-regulated in the 6 dpa mutant ovules.

The down-regulated DEGs identified at 0 and 6 dpa have a cellular functionality related to cell periphery and cell wall, respectively (Supplementary Table S5). In addition to GhEXP8i (Gh_D13G0060), several genes encoding xyloglucan endotransglycosylase/hydrolase (XTH) and exordium (EXO) were significantly down-regulated in the 6 dpa mutant ovules. Among the up-regulated DEGs found in the 6 dpa mutant ovules were genes encoding plasma membrane intrinsic proteins or involved in potassium transport (Supplementary Table S3). The functions of all these genes in cotton fiber initiation remain to be explored.

Together, these results imply that the transcriptome landscape of the mutant is relatively similar to that of WT in the 2 and 4 dpa ovules, but has reprogrammed in the 0 and 6 dpa ovules, particularly at 6 dpa when fuzz fibers start to initiate and/or development, for genes involved in transmembrane transportation, cell wall modifications, and TF activities.

**Discussion**

Numerous fuzzless cotton mutants with a variable amount of lint fibers have been reported. The historic dominant N₅ mutant has a very low lint percentage and consequently significant penalty on lint yield. From a breeding perspective, such a mutant has little utility, although it is valuable for understanding fiber...
initiation and development. But some EMS induced fuzzless mutants were reported to have comparable lint percentage compared to the cultivar from which the mutant was induced, and they have been used in breeding programs to develop fuzzless genotypes as the fuzzless trait has the potential to reduce short fiber content and to improve both ginning quality and efficiency (Bechere et al. 2009, 2012; Hendon et al. 2019). The fuzzless-tufted mutant reported in this study has a lint percentage of C24 35%.

After several backcrosses, lint yield, lint percentage, and fiber quality of the fuzzless-tufted NILs were similar to or higher than that of the recurrent parent (Table 1). The commercial case for fuzzless seed cotton cultivar also lies in improved ginning efficiency. Comparison of a fuzzless-tufted introgression line with its recurrent parent (Sicot 71BRF) illustrated the extent of the benefit (Table 2). These results demonstrate the breeding value of the fuzzless-tufted mutant. To facilitate introgressing the trait into elite commercial cultivars and to understand the molecular mechanism underlying the mutant phenotype, we genetically characterized the mutant, determined the mutant phenotype to be controlled by a single genetic locus (N5) with incomplete dominance, and mapped the mutant locus to a chromosomal location that has not previously been associated with fiber development (Figure 2). Two genes, Gh_D13G0096 and Gh_D13G0098, were considered to be the candidates related to fuzz initiation and development based on their location in the genetic interval containing the fuzzless-tufted mutation and their differential expression profile in the developing ovules of the mutant and WT (Table 4). We are currently unsure about the nature of the mutations in these two candidate genes that causes their up-regulation in the mutant, as it is complicated by the variable copy numbers of Gh_D13G0096 annotated in a different version of the same TM-1 genome and the alternative splicing events of Gh_D13G0098 (Supplementary Figure S6B).

Up-regulation of Gh_D13G0096 and Gh_D13G0098 in the mutant suggest that they might be repressors of fuzz initiation and/or development. The function of Gh_D13G0096 cannot be inferred due to lack of characterized homologous gene in other plants, but given its expression pattern in various tissues of TM-1 (Supplementary Figure S3), Gh_D13G0096 might play a more important role in fiber elongation or seed development than in fuzz initiation. Gh_D13G0098 encodes an amino acid-polyamine-choline (APC) transporter. Amino acids are indispensable for cell biology and plant development, and act as regulators of metabolic pathways. Their translocation within plants and transport across intracellular membranes are mediated by multiple families of amino acid transporters, including APC transporters. APC transporters have been identified in almost all organisms, implying their functional conversion and fundamental roles in cell biology (Wipf et al. 2002). The Arabidopsis genome contains 14 APC-type transporters, including 9 cationic amino acid transporters (AtCAT1 to AtCAT9; Su et al. 2004). Gh_D13G0098 is closely related to AtCAT7 and AtCAT6. The physiological and molecular function...
of AtCAT7 remains to be investigated but it has been suggested to act as a compatible solute transporter because the expression level of AtCAT7 is low under standard growth conditions but induced by treatment of NaCl (Su et al. 2004). AtCAT6 is preferentially expressed in sink tissues, including the outer epidermal cell layer of embryo, and is localized primarily in the plasma membranes. AtCAT6 transports both cationic and large, neutral amino acids (Hammes et al. 2006). Another two members of the Arabidopsis APC transporter family, AtCAT2 and AtCAT4, are tonoplast localized with AtCAT2 being important for regulation of the concentration of leaf amino acids (Hammes et al. 2006). Another two members of the Arabidopsis APC transporter family, AtCAT2 and AtCAT4, are tonoplast localized with AtCAT2 being important for regulation of the concentration of leaf amino acids (Yang et al. 2014). It is thus of interest to investigate the cellular localization and substrate specificity of Gh_D13G0098. We speculate that up-regulation of Gh_D13G0098 might have altered the homeostasis of certain amino acids and/or their analogs in the epidermal cells, leading to failure of the epidermal cells to be initiated as fuzz fibers via a gene network regulating Ca^{2+}-signaling, reactive oxygen species signaling, and cell wall loosening and transportation.

Calcium (Ca^{2+}) is an essential macronutrient in plants. It also serves as an important secondary messenger in plant development and has been shown to regulate diverse physiological processes and organ development, such as pollen tube growth and root hair elongation (Brost et al. 2019; Zheng et al. 2019). Comparative transcriptome analysis in cotton has suggested the importance of Ca^{2+}-signaling in fiber initiation as many down-regulated genes at the fiber initiation stage in the fiberless mutant are involved in calcium-mediated signal transduction pathways (Padmalatha et al. 2012). The essentialness of Ca^{2+} in fiber initiation has been further demonstrated by in vitro ovule culture experiments, in which no fiber initiation was observed on 1 dpa ovules cultured for 72 h on the medium without Ca^{2+} and fiber initiation could be reinitiated by supplying Ca^{2+} (Zhang et al. 2017a). Plants have several classes of Ca^{2+} sensors recognizing specific Ca^{2+} signals and transducing them into downstream effects, including calmodulins (CaMs), calmodulin-like proteins (CMLs), calcineurin B-like proteins (CBLs), and calcium-dependent protein kinases (CDPKs or CPKs) (Shi et al. 2018). In cotton, early studies have found that genes encoding different types of Ca^{2+} sensors are among those that are preferentially expressed in elongating fibers, suggesting an important role of Ca^{2+}-mediated signal transduction in fiber elongation (Gao et al. 2007, Huang et al. 2008). A recent study showed that, in cotton
accession with normal fiber, the expression level of GhCaM7 gradually increases in 0–15 dpa ovules and correlates with the Ca\(^{2+}\)- influx rate from the extracellular to the intracellular environment at the fiber tips. By contrast, fiberless accessions have a much lower expression level of GhCaM7 in 0 dpa ovules, implying a role of GhCaM7 in not only fiber elongation but also fiber initiation, which is supported by transgenic cottons with overexpressed or down-regulated expression level of GhCaM7 (Tang et al. 2014). In this study, we found a CPK-encoding gene D13G0033 (or GhCPK94; Gao et al. 2018) in the flanking region of the N\(_5\) locus. The expression level of D13G0033 in the 0 - 6 dpa ovules was constantly significantly lower in the mutant than in the WT (Figure 4), suggesting that Ca\(^{2+}\)- influx and/or signaling is compromised in the mutant, consistent with a positive role of Ca\(^{2+}\) in fiber initiation. In addition, a gene (D13G0114) encoding an IQ-domain-containing protein that mediates calmodulin (Bahler and Rhoads 2002), an archetypal Ca\(^{2+}\)-sensor, was found in the flanking region of the N\(_5\) locus. Together, these findings suggest that multiple Ca\(^{2+}\)-signaling pathways could have been affected by the fuzzless-tufted mutation.

Expansins are cell-wall loosening proteins involved in cell enlargement by local sliding of wall polymers without hydrolytic breakage of major structural components of the cell wall. They are involved in almost all physiological processes during plant development (Marow et al. 2016). Several expansin genes preferentially expressed in developing cotton fibers have been identified (Harmer et al. 2002; Li et al. 2016; Lv et al. 2020). For instance, GhEXPATR is an atypical \(\alpha\)-expansin that lacks domain 2 and is expressed in developing fibers of G. barbadense accessions but not in those of G. hirsutum accessions, implying a role of GHEXPATR in fiber elongation as G. barbadense fibers are much longer than G. hirsutum ones. Supporting this was the observation that transgenic plants overexpressing GHEXPATR had longer, stronger, and finer fibers (Li et al. 2016). Longer fibers were also observed in transgenic cotton plants overexpressing another \(\alpha\)-expansin gene, GhEXPA8 (Bajwa et al. 2015). In WT ovules, the expression level of D13G0060 or GHEXPA8, one of the 10 GHEXPA8 genes, gradually increases from 0 to 7 dpa, significantly decreases after 10 dpa and remains low afterwards (this study and Lv et al. 2020).

The time period with the highest expression level of D13G0060 in ovules co-occurs with the time of fuzz initiation (Stewart 1975; Zhang et al. 2007). The expression level of D13G0060 was down-regulated in 0–6 dpa ovules in the fuzzless-tufted mutant, particularly in 6 and 4 dpa mutant ovules (Figure 4D). Given the expression pattern of this gene in developing ovules and fibers, and the observation of a very low expression level in the 4 and 6 dpa mutant ovules (Supplementary Table S3), we propose that D13G0060 or GHEXPA8 may have a specific role in initiation of fuzz fibers.

One interesting finding from the transcriptome analysis is the up-regulation of a senescence associated gene (D13G0139) located in the flanking region of the N\(_5\) locus. D13G0139 is the most significantly up-regulated gene at all four time points in the mutant ovules (Supplementary Table S3). D13G0139 is homologous to Arabidopsis AtSAG20, a gene that has not been functionally characterized. Despite it is currently unable to speculate if D13G0139 play a role in fiber cell differentiation and initiation, the expression level of another senescence associated gene, AtSAG21, has been reported to play a role in root hair elongation as Arabidopsis plants overexpressing AtSAG21 have a higher proportion of longer root hairs whereas the root hairs of the plants with down-regulated AtSAG21 are shorter. And AtSAG21 was found to be located to mitochondria (Salleh et al. 2012). As the expression level of AtSAG21 was up-regulated in response to oxidants, it was proposed that AtSAG21 might function as a cofactor maintaining the stability of the mitochondria proteins involved in production and/or signaling of reactive oxygen species (Salleh et al. 2012). Up-regulation of D13G0139 in the mutant ovules suggests it being a negative regulator of fuzz initiation, different from the positive role of AtSAG21 in root hair initiation and elongation. The role of D13G0139 in fiber initiation and elongation warrants further investigation.

Based on the number of identified DEGs and enriched GO terms at each time point, it seems that the effect of the N\(_5\) mutation is mainly at 0 and 6 dpa, particularly at 6 dpa, consistent with the fuzzless mutant phenotype, as fuzz fibers are believed to be initiated at 5–10 dpa (Stewart 1975; Zhang et al. 2007). Compared to the published transcriptome studies based on fiberless (lintless and fuzzless) mutants (Padmalatha et al. 2012; Wan et al. 2014), our transcriptome analysis, using fuzzless mutant, should work better in the identification of genes related to fuzz development. Indeed, many genes that have not been previously associated with fiber development were uncovered to have a potential role in fuzz development, including many TFs and genes encoding proteins involved in cell wall extensibility, such as XTH and EXO. The primary cell walls of plants consists of cellulose microfibrils tethered by pectins and xyloglucans. XTH functions as cell wall loosening enzymes by cleaving the links between xyloglucans and cellulose microfibrils (Park and Cosgrove 2015). Overexpressing GhXTH1 was able to increase fiber length by 15–20% (Lee et al. 2010). EXO has been shown to be required for leaf cell expansion and a potential negative regulator of cell division (Farrar et al. 2003; Schroder et al. 2009). Significant down-regulation of XTH and EXO genes observed at 6 dpa in the fuzzless-tufted mutant implies an association of these genes with initiation of fuzz fibers (Figure 4). Furthermore, it has been suggested that xyloglucans tend to intertwine with cellulose microfibrils at the sites where \(\alpha\)-expansins act to loosen cell wall (Park and Cosgrove 2015), suggesting that initiation of fuzz fibers may require cell wall loosening coordinated by XTH and expansins.

As elongated epidermal cells, the formation of lint and fuzz fibers undergoes cell specification, initiation, and elongation three phases. We know almost nothing about the biological process specifying the fate of fiber cells, although auxin accumulation in fiber-to-be facilitated by polarized transport of auxin from neighboring non-fiber epidermal cells seems to be crucial for the fiber-to-be becoming a fiber initial (Zhang et al. 2017b; Zeng et al. 2019). We are unable to infer if the N\(_5\) mutation has a role in specification of epidermal cells to becoming fuzz fibers. Once the fate of fuzz fiber cells being determined, localized, and/or polarized cell wall loosening may be a prerequisite for protrusion of fuzz cells from the epidermal cell layer of seed coat. The DEGs identified in this study based on transcriptome sequencing suggest fuzz initiation involves diverse pathways related to cell wall modifications and integrity that are important for amino acid transport, Ca\(^{2+}\)-signaling, and polarized cell wall outgrowth. Many TFs, cell wall loosening enzymes, and genes with an unknown function seem to be crucial components of these pathways. Overall, the results presented in this study indicate that the N\(_5\) locus plays an indispensable role in fuzz development, possibly by regulating accumulation of amino acids, Ca\(^{2+}\)-signaling, and synthesis and deposition of cell wall loosening enzymes. The N\(_5\) locus may achieve these regulatory roles directly or indirectly through TFs regulating these biological processes.
**Acknowledgments**

We thank Drs Shiming Liu and Jiafu Tan for their valuable comments on the manuscript during the CSIRO’s internal review process.

**Funding**

This work was funded through the Cotton Breeding Australia, a joint venture between Cotton Seed Distributors Ltd. Australia and CSIRO Agriculture and Food (CBA04).

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

**Literature cited**

Anders S, Pyl PT, Huber W. 2015. HTSeq—a python framework to work with high-throughput sequencing data. Bioinformatics. 31: 166–169.

Bahler M, Rhoads A. 2002. Calmodulin signaling via the IQ motif. FEBS Lett. 513:107–113.

Bajwa KS, Shahid AA, Rao AQ, Bashir A, Aftab A, et al. 2015. Stable transformation and expression of GhEXPA8 fiber expansin gene to improve fiber length and micronaire value in cotton. Front Plant Sci. 6:638.

Bechere E, Auld DL, Hequet E. 2009. Development of ‘naked-tufted’ seed coat mutants for potential use in cotton production. Euphytica. 167:333–339.

Bechere E, Turley RB, Auld DL, Zeng L. 2012. A new fuzzless seed locus in an Upland cotton (Gossypium hirsutum L.) mutant. Am J Plant Sci. 3:799–804.

Boalger AM, Loseh M, Usadel B. 2014. Trimomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Brost C, Studtrucker T, Reimann R, Denninger P, Czekalla J, et al. 2019. Multiple cyclic nucleotide-gated channels coordinate calcium oscillations and polar growth of root hairs. Plant J. 99: 910–923.

Chen ZJ, Sreedasyam A, Ando A, Song QX, De Santiago LM, et al. 2020. Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. Nat Genet. 52: 525–533.

Endrizzi JE, Ramsay G. 1980. Identification of ten chromosome deficiencies in cotton. J Hered. 71:45–48.

Fang DD, Naoumkina M, Kim HJ. 2018. Unraveling cotton fiber development using fiber mutants in the post-genomic era. Crop Sci. 58:2214–2228.

Farrar K, Evans IM, Topping JF, Souter MA, Nielsen JE, et al. 2003. EXORDIUM—a gene expressed in proliferating cells and with a role in meristem function, identified by promoter trapping in Arabidopsis. Plant J. 33:61–73.

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

Funk PA, Hardin R, Funk PA, Hardin R, Hughes SE, Boykin JC. 2013. Changes in cotton gin energy consumption apportioned by 10 functions. J Cotton Sci. 17:174–183.

Gao P, Zhao PM, Wang J, Wang HY, Wu XM, et al. 2007. Identification of genes preferentially expressed in cotton fibers: a possible role of calcium signaling in cotton fiber elongation. Plant Sci. 173: 61–69.

Gao W, Xu FC, Guo DD, Zhao JR, Liu J, et al. 2018. Calcium-dependent protein kinases in cotton: insights into early plant responses to salt stress. BMC Plant Biol. 18:15.

Harmer SE, Orford SJ, Timmis JN. 2002. Characterisation of six alpha-expansin genes in Gossypium hirsutum (upland cotton). Mol Gen Genomics. 268:1–9.

Hamnes UZ, Nielsen E, Honaas LA, Taylor CG, Schachtman DP. 2006. AtCAT6, a sink-tissue-localized transporter for essential amino acids in Arabidopsis. Plant J. 48:414–426.

Hendon BR, Bechere E, Witt TW, Kelly BR, Mishra D, et al. 2019. Genetic improvement of naked-tufted seed mutants in upland cotton (Gossypium hirsutum L.). Euphytica. 215:Article number: 81.

Higo K, Uwagai Y, Iwamoto M, Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res. 27:297–300.

Hu Y, Chen J, Fang L, Zhang Z, Ma W, et al. 2019. Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet. 51:739–748.

Huang G, Wu Z, Percy RG, Bai M, Li Y, et al. 2020. Genome sequence of Gossypium herbaceum and genome updates of Gossypium arbor- eum and Gossypium hirsutum provide insights into cotton A-genome evolution. Nat Genet. 52:516–524.

Huang QS, Wang HY, Gao P, Wang GY, Xia GQ. 2008. Cloning and characterization of a calcium dependent protein kinase gene associated with cotton fiber development. Plant Cell Rep. 27: 1869–1875.

Hulse-Kemp AM, Lemm J, Flieske J, Ashrafi H, Buryarapu R, et al. 2015. Development of a 63K SNP array for cotton and high-density mapping of intraspecific and interspecific populations of Gossypium spp. G3 (Bethesda). 5:1187–1209.

Lee J, Burns TH, Light G, Sun Y, Fokar M, et al. 2010. Xyloglucan endothallase/hydrolase genes in cotton and their role in fiber elongation. Planta. 232:1191–1205.

Li H, Ye G, Wang J. 2007. A modified algorithm for the improvement of composite interval mapping. Genetics. 175:361–374.

Li Y, Tu L, Pettolino FA, Ji S, Hao J, et al. 2016. GhEXPATR, a species-specific expansin, enhances cotton fibre elongation through cell wall restructuring. Plant Biotechnol J. 14:951–963.

Liu X, Moncuquet P, Zhu QH, Stiller W, Zhang Z, et al. 2020. Genetic identification and transcriptome analysis of lintless and fuzzless traits in Gossypium arboreum L. Int J Mol Sci. 21:1675.

Lopez-Lavalle LAB, Gillespie VJ, Tate WA, Ellis MH, Stiller WN, et al. 2012. Molecular mapping of a new source of Fusarium wilt resistance in tetraploid cotton (Gossypium hirsutum L.). Mol Breeding. 30:1181–1191.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Lv LM, Zuo DY, Wang XF, Cheng HL, Zhang YP, et al. 2020. Genome-wide identification of the expansin gene family reveals that expansin genes are involved in fibre cell growth in cotton. BMC Plant Biol. 20:223.

Marowa P, Ding A, Kong Y. 2016. Expansins: roles in plant growth and potential applications in crop improvement. Plant Cell Rep. 35:949–965.

Padmalatha KV, Patil DP, Kumar K, Dhandapani G, Kanakachari M, et al. 2012. Functional genomics of fuzzless-lintless mutant of Gossypium hirsutum L. cv. MCUS reveal key genes and pathways involved in cotton fibre initiation and elongation. BMC Genomics. 13:624.

Park YB, Cosgrove DJ. 2015. Xyloglucan and its interactions with other components of the growing cell wall. Plant Cell Physiol. 56: 180–194.
Patel JD, Huang X, Lin L, Das S, Chandnani R, et al. 2020. The Ligon lintless -2 short fiber mutation is located within a terminal deletion of chromosome 18 in cotton. Plant Physiol. 183:277–288.

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, et al. 2012. Repeated polyploidization of Gossypium genomes and the evolution of spinable cotton fibres. Nature. 492:423–427.

Salleh FM, Evans K, Goodall B, Machin H, Mowla SB, et al. 2012. A novel function for a redox-related LEA protein (SAG21/AtLEA5) in root development and biotic stress responses. Plant Cell Environ. 35:418–429.

Schröder F, Lissow J, Lange P, Mussig C. 2009. The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. BMC Plant Biol. 9:20.

Shi S, Li S, Asim M, Mao J, Xu D, et al. 2018. The Arabidopsis Calcium-Dependent Protein Kinases (CDPKs) and their roles in plant growth regulation and abiotic stress responses. Int J Mol Sci. 19:1900.

Stewart J. 1975. Fiber initiation on the cotton ovule (Gossypium hirsutum). Am J Bot. 62:723–730.

Su YH, Frommer WB, Ludewig U. 2004. Molecular and functional characterization of a family of amino acid transporters from Arabidopsis. Plant Physiol. 136:3104–3113.

Tang W, Tu L, Yang X, Tan J, Deng F, et al. 2014. The calcium sensor GhCaM7 promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New Phytol. 202:509–520.

Tian T, Liu Y, Yan H, You Q, Yi X, et al. 2017. agrigo v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 45:W122–W129.

Turley RB, Kloth RH. 2002. Identification of a third fuzzless seed locus in upland cotton (Gossypium hirsutum L.). J Hered. 93:359–364.

Turley RB, Vaughn KC, Scheffler JA. 2007. Lint development and properties of fifteen fuzzless seed lines of Upland cotton (Gossypium hirsutum L.). Euphytica. 156:57–65.

Voorrips RE. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 93:77–78.

Walford SA, Wu Y, Llewellyn DJ, Dennis ES. 2011. GhMYB25-like: a key factor in early cotton fibre development. Plant J. 65:785–797.

Wan Q, Guan X, Yang N, Wu H, Pan M, et al. 2016. Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. New Phytol. 210:1298–1310.

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:e97313.

Wang M, Tu L, Yuan D, Zhu D, Shen C, et al. 2019. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. Nat Genet. 51:224–229.

Ware JO. 1940. Relation of fuzz pattern to lint in an upland cotton cross. J Hered. 31:489–496.

Ware JO, Benedict LI, Rolfe WH. 1947. A recessive naked-seed character in upland cotton. J Hered. 38:313–320.

Wipf D, Ludewig U, Tegeder M, Rentsch D, Koch W, et al. 2002. Conservation of amino acid transporters in fungi, plants and animals. Trends Biochem Sci. 27:139–147.

Yang H, Krebs M, Stierhof YD, Ludewig U. 2014. Characterization of the putative amino acid transporter genes AtCAT2, 3 & 4: the tonoplast localized AtCAT2 regulates soluble leaf amino acids. J Plant Physiol. 171:594–601.

Zeng J, Zhang M, Hou L, Bai W, Yan X, et al. 2019. Cytokinin inhibits cotton fiber initiation by disrupting PIN3a-mediated asymmetric accumulation of auxin in the ovule epidermis. J Exp Bot. 70:3139–3151.

Zhan S, Horrocks J, Lukens LN. 2006. Islands of co-expressed neighbouring genes in Arabidopsis thaliana suggest higher-order chromosomal domains. Plant J. 45:347–357.

Zhang DY, Zhang TZ, Sang ZQ, Guo WZ. 2007. Comparative development of lint and fuzz using different cotton fiber-specific developmental mutants in Gossypium hirsutum. J Integr Plant Bio. 49:1038–1046.

Zhang M, Cao HZ, Hou L, Song SQ, Zeng JY, et al. 2017a. In vivo imaging of Ca2+ accumulation during cotton fiber initiation using fluorescent indicator YC3.60. Plant Cell Rep. 36:911–918.

Zhang M, Zeng JY, Long H, Xiao YH, Yan XY, et al. 2017b. Auxin regulates cotton fiber initiation via GhPIN-mediated auxin transport. Plant Cell Physiol. 58:385–397.

Zhang T, Hu Y, Jiang W, Fang L, Guan X, et al. 2015. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 33:531–537.

Zheng RH, Su S, Xiao H, Tian HQ. 2019. Calcium: a critical factor in pollen germination and tube elongation. Int J Mol Sci. 20:420.

Zhu QH, Yuan Y, Stiller W, Jia Y, Wang P, et al. 2018. Genetic dissection of the fuzzless seed trait in Gossypium barbadense. J Exp Bot. 69:997–1009.

Zhu QH, Zhang J, Liu D, Stiller W, Liu D, et al. 2016. Integrated mapping and characterization of the gene underlying the okra leaf trait in Gossypium hirsutum L. J Exp Bot. 67:763–774.

Communicating editor: J. Wendel