Molecular Genetics and Genomics (2021) 296:1041–1049
https://doi.org/10.1007/s00438-021-01802-0

ORIGINAL ARTICLE

Mapping-by-sequencing the locus of EMS-induced mutation responsible for tufted-fuzzless seed phenotype in cotton

Marina Naoumkina[1] · Gregory N. Thyssen[1,2] · David D. Fang[1] · Efrem Bechere[3] · Ping Li[1] · Christopher B. Florane[1]

Received: 1 March 2021 / Accepted: 2 June 2021 / Published online: 10 June 2021

Abstract

Cotton fiber mutants are valuable resources for studying functions of altered genes and their roles in fiber development. The n_4^t is a recessive tufted-fuzzless seed mutant created through chemical mutagenesis with ethyl methanesulfonate. Genetic analysis indicated that the tufted-fuzzless phenotype is controlled by a single recessive locus. In this study, we developed an F_2 population of 602 progeny plants and sequenced the genomes of the parents and two DNA bulks from F_2 progenies showing the mutant phenotype. We identified DNA sequence variants between the tufted-fuzzless mutant and wild type by aligning the sequence reads to the reference TM-1 genome and designed subgenome-specific SNP markers. We mapped the n_4^t locus on chromosome D04 within a genomic interval of about 411 kb. In this region, seven genes showed significant differential expression between the tufted-fuzzless mutant and wild type. Possible candidate genes are discussed in this study. The utilization of the n_4^t mutant along with other fiber mutants will facilitate our understanding of the molecular mechanisms of cotton fiber cell growth and development.

Keywords Cotton · n_4^t mutation · Mapping-by-sequencing · Single nucleotide polymorphism · Tufted-fuzzless phenotype

Introduction

Cotton (*Gossypium hirsutum* L.) fibers are single-cell unbranched trichomes developed from epidermal cells of a seed. Approximately one in four epidermal cells develop into fiber (Steward 1975). There are four stages observed during cotton fiber development, including initiation, elongation, secondary cell wall synthesis, and maturation (Kim and Triplett 2001). Most commercially grown cotton cultivars produce two types of fibers: lint fiber that is easily detached during the ginning process and fuzz fiber that strongly adheres to the seed coat. Lint fiber cells grow approximately 30 mm long and are relatively fine, spinnable, and used for textile production. The fuzz fibers are short (about 5 mm length) and are considered an undesirable trait since they hinder seed germination and increases energy consumption during the ginning process (Bechere et al. 2011). There are differences in the time when lint and fuzz fiber cells emerge from the epidermal layer: lint fibers initiate around 3 days before to the day of anthesis (DOA), whereas fuzz fibers delay initiation until 3–5 days post-anthesis (DPA) (Steward 1975; Zhang et al. 2007).

Cotton fiber mutants are valuable resources for studying functions of altered genes and their roles in fiber development. A number of spontaneous fiber mutants, as well as artificially mutagenized lines, are known to cotton scientists. The single dominant N_1 mutation is responsible for the 'fuzzless seed' phenotype that has no fuzz and different degrees of lint on the seeds (Kearney and Harrison 1927). The recessive n_2 mutant produces a fuzzless seed phenotype with normal lint fibers (Ware et al. 1947). Recently a new partially dominant N_5 mutant has been reported that produces a tufted-fuzzless seed phenotype (Zhu et al. 2020). A recessive tufted-fuzzless seed n_4^t mutant was developed through chemical mutagenesis (Bechere et al. 2012). The n_4^t

Communicated by Stefan Hohmann.

✉ Marina Naoumkina
marina.naoumkina@usda.gov

[1] Cotton Fiber Bioscience Research Unit, USDA-ARS-SRRC, New Orleans, LA 70124, USA

[2] Cotton Chemistry and Utilization Research Unit, USDA-ARS-SRRC, New Orleans, LA 70124, USA

[3] Crop Genetics Research Unit, USDA-ARS, Stoneville, MS 38776, USA

seeds are partially fuzzless (tufted) with fuzz grown around the micropyle and chalazal ends of the ovule. The n_4^t locus is not allelic to N_1 or n_2 (Bechere et al. 2012). Phenotypic evaluation of fiber characteristics determined that the lint yield and percentage of the n_4^t were not significantly different than that of the wild-type, whereas the short fiber content and amount of coat neps were 15% and 45% lower in the mutant's fiber; the n_4^t mutant line required less net energy for ginning compared with fuzzy cotton cultivars (Bechere et al. 2012; Bechere and Auld 2014).

Using a map-based cloning approach the N_1 mutation was linked to a gene on the chromosome (Chr.) A12 that was annotated as a MYB-MIXTA-like (MML) transcription factor *GhMML3_A12* (Wan et al. 2016). Small interfering RNAs originating from bidirectional transcripts of *GhMML3_A12* reduced its transcript level and therefore caused the fuzzless seed phenotype (Wan et al. 2016). RNA interference (RNAi) suppression of the *GhMML3_A12* gene resulted in fiberless seeds but normal trichomes elsewhere which demonstrated a critical role of this gene in early cotton fiber development (Walford et al. 2011). Mutations responsible for the n_2 fuzzless phenotype were linked to two loci, on Chr. A12 and D12 (Turley and Kloth 2002; Song et al. 2010). Using a map-based cloning approach the gene from Chr. D12 locus has been isolated. The gene is annotated as another MML transcription factor *GhMML4_D12*, which was proposed to be responsible for lint fiber development (Wu et al. 2017). The most recent survey study has identified two mutations in the n_2 line, including a non-synonymous SNP (nsSNP) in the third exon of the *MML3-A12* gene and an nsSNP in the *MML4-D12* gene that removes stop codon, leading to an extension of the coding region of this gene (Naoumkina et al. 2020). The genetic locus responsible for the N_5 tufted-fuzzless seed phenotype had been identified on Chr. D13 of *G. hirsutum* (Zhu et al. 2020).

Progress has been made in understanding the genetic control of lint and fuzz fiber development. Also, differences have been found in pathways of fiber development between diploid and tetraploid cotton species. In tetraploid *G. hirsutum* it has been proposed that both *MML3_A12* and *MML3_D12* are involved in lint development, whereas fuzz formation is mainly determined by the expression level of *MML3_D12* at about 3 DPA (Zhu et al. 2018). In diploid *G. arboreum* the *GaMML3* gene was also associated with fuzz fiber development, however its role was not as significant as *GaGIR1*, which has been proposed to be the main factor associated with fuzz development in *G. arboreum* according to investigations of several different fuzzless mutants (Feng et al. 2019; Liu et al. 2020). Besides the known key factors, such as *GhMML3* and *GaGIR1*, there are many more yet unidentified genes that are also important for cotton fiber development. For example, two candidate genes from the N_5 locus (Chr. D13) have been proposed to be causative genes, which might trigger transcriptional reprograming of fiber cells leading to a tufted-fuzzless phenotype in *G. hirsutum* (Zhu et al. 2020).

The aim of our study was to identify the n_4^t locus and potential candidate genes that cause the tufted-fuzzless seed phenotype. We developed an F_2 mapping population consisting of 602 individuals from a cross between the n_4^t mutant and a wild type (WT). We sequenced the genomes of DNA bulks from F_2 progenies showing the mutant phenotype along with their parents. The n_4^t locus was mapped within a genomic region of about 411 kb on Chr. D04. The n_4^t locus contains 45 genes according to a *G. hirsutum* TM-1 genome assembly (Wang et al. 2019). The putative candidate genes that might cause the n_4^t fiber phenotype are discussed in this study.

Materials and methods

Plant materials and development of the F_2 population

The n_4^t tufted-fuzzless cotton line was developed through EMS chemical mutagenesis of the cultivar SC9023 (PI 590933) (Bechere et al. 2012; Bechere and Auld 2014). This mutant has reduced density of short fibers (fuzz) that remain attached to the seed after spinnable lint fibers are removed and have only a small amount of fuzz attached to the micropyle and chalazal ends of the seed. Crosses were made between SC9023 and homozygous n_4^t mutant plants in the summer of 2018 in a field in New Orleans, LA, USA. Four F_1 plants were grown in New Orleans's greenhouse during the winter to produce F_2 seeds. F_2 plants (602 individuals) were grown during the summer of 2019 in a field in New Orleans. Standard field management practices were applied.

Fiber quality testing and statistical analysis

Ten naturally opened bolls were manually harvested from each individual plant for fiber quality testing. The cotton samples were ginned using a laboratory roller gin. Fiber quality traits of F_2 populations were measured using a High Volume Instrument (HVI, Uster Technologies Inc.). The fiber testing was conducted by the Cotton Fiber Testing Lab, USDA-ARS-SRRC, New Orleans, LA. A *t*-test (JMP Genomics 9.0; SAS, Cary, NC, USA) was used to compare fiber characteristics between the two groups of samples, including WT and n_4^t.

DNA isolation and sequencing of super bulked segregants

Young leaves were collected from each individual F_2 plant within the segregating population. DNA was isolated as previously described (Fang et al. 2010). The concentration of DNA samples was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

F_2 plants were selected for sequencing by a bulked segregant analysis approach (Michelmore et al. 1991; Takagi et al. 2013; Thyssen et al. 2014). We selected six plants of category-1 for the first bulk and 41 plants of category-2 for the second bulk (description of seed categorization is described in the Results section). DNAs from the two mutant bulks along with the n_4^t mutant and WT (SC9023) parental DNAs were sent to Novogene Corporation (Chula Vista, CA, USA) for library preparation, and whole genome sequencing using Illumina HiSeq 2500 with paired-end 150 bp reads. Each was sequenced at 25 × coverage (about 65 Gb).

Identification of the genomic locus harboring the n_4^t mutation

Sequence reads of the two bulked DNAs and two parents were aligned to the reference *Gossypium hirsutum* cv. Texas Marker-1 (TM-1) genome (Wang et al. 2019) using BWA-MEM software (Zhu et al. 2018). Variants were identified with samtools mpileup (with "-Ego" flags) and bcftools ("call–vm") software (Dettmer et al. 2007; Obata and Fernie 2012). A histogram was generated by counting the number of single nucleotide polymorphisms (SNP) and insertion/deletions (indel) in 1 Mb intervals. Primers were designed to analyze the segregation of the SNP markers in the F_2 population according to our previously published protocol (Thyssen et al. 2014). The primer sequences are listed in Table S1.

Association analyses between the n_4^t phenotype and genetic markers were performed using the composite interval mapping (CIM) model in JMP Genomics 9.0. The seed phenotype trait was numerically annotated for association analysis: fuzzy WT—0; category-3—0.7; category-2—0.8 and category-1—0.9. The numbers reflect the degree of the nakedness of seeds: the less fuzz the higher number. Therefore, the seeds of category-1 with a small amount of fuzz attached to both ends were annotated as 0.9, while fully fuzzy WT was annotated as 0. The SNP data was also numerically annotated: homozygous WT—0; heterozygous—1; and homozygous mutant—2. The genotype probability data set for the 602 F_2 individuals was built using the same software. The forward regression method was used for control marker selection with a window size of 10.0 cM and the control marker number was 5. The expectation–maximization (EM) algorithm, at a threshold of 10 (LOD > 10), was used to identify the significant association associated locus.

RNA isolation and RT-qPCR

Cotton bolls were harvested at DOA, and 3 and 5 DPA. Epithelial cells were separated from ovules using a glass bead shearing technique (Taliercio and Boykin 2007). Total RNA was isolated using a Sigma SpectrumTM Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO). The cDNA synthesis reactions were performed using the iScriptTM cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA) according to the manufacturer's instructions with 1 µg of total RNA per reaction. The qPCR reactions were performed with iTaq™ SYBR® Green Supermix (Bio-Rad Laboratories Inc., Hercules, CA) in a Bio-Rad CFX96 real time PCR detection system in three biological replicates with two technical replicates. Amplification of the endogenous 18S rRNA (Genbank accession U42827) was used as a reference for normalization. The relative expression levels were calculated using $2^{-\Delta\Delta C_t}$ (Pfaffl 2001). Primers are provided in Table S2. A two-way ANOVA statistical test (GraphPad Prism 7) was applied to identify significantly differentially expressed samples.

Results

The n_4^t tufted-fuzzless phenotype and its genetic control

The n_4^t mutation affected only the seed phenotype (Fig. 1). There were no morphological differences observed between the WT and the mutant in the plant's vegetative organs including stems, leaves and flowers. It was

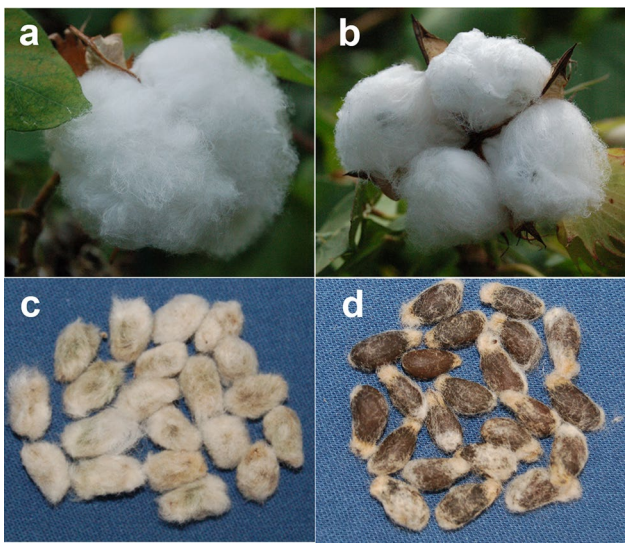

Fig. 1 Bolls and seeds of wild type and tufted-fuzzless n_4^t mutant. **a** Wild type boll; **b** n_4^t boll; **c** seeds of wild type; **d** seeds of n_4^t mutant

previously reported that no significant differences in lint yield and lint percentage were observed between the WT and the mutant lines [see also details in the Introduction] (Bechere et al. 2012; Bechere and Auld 2014). The homozygous n_4^t seeds are partially fuzzless, with tuft grown around the micropyle and chalazal ends of the seed (Fig. 1d). F_1 seeds are fuzzy, indicating that the n_4^t mutant phenotype is recessive (Fig. 2a). Seeds from F_2 segregating plants showed different degrees of fuzz (Fig. 2b–d). We grouped the tufted seed fiber phenotypes of F_2 segregating plants into three categories: category-1 (six plants) with clearly tufted seeds (Fig. 2b); category-2 (41 plants) with a mixture of clearly tufted seeds and seeds that showed sparsely distributed short fuzz fiber cells around the seed (Fig. 2c); and category-3 (103 plants) where all seeds in the boll showed varying degrees of fuzz fiber cells that were nevertheless clearly less than normal for WT seeds (Fig. 2d).

We observed 452 plants with wild-type fuzzy seeds and 150 plants with n_4^t tufted-fuzzless seeds among 602 individuals of the F_2 segregating population. Therefore, the segregation of seed fiber phenotypes in this population fits a three wild type to one mutant ratio, suggesting one recessive Mendelian locus.

Fiber quality analysis

To determine whether or not the n_4^t phenotype affects the fiber quality traits we measured the fiber properties of every individual plant from the F_2 population using an HVI. HVI-measured fiber properties of WT and n_4^t phenotype-groups from the F_2 segregating population were subjected to statistical analysis to identify significant differences between the two groups of samples. Among the fiber traits measured by HVI, only fiber length showed a significant difference between the WT and the n_4^t samples. Analysis of the upper half mean length (UHML) of fiber showed that samples from fuzzless mutant were significantly ($p < 0.0001$) longer than wild-type samples (Fig. 3).

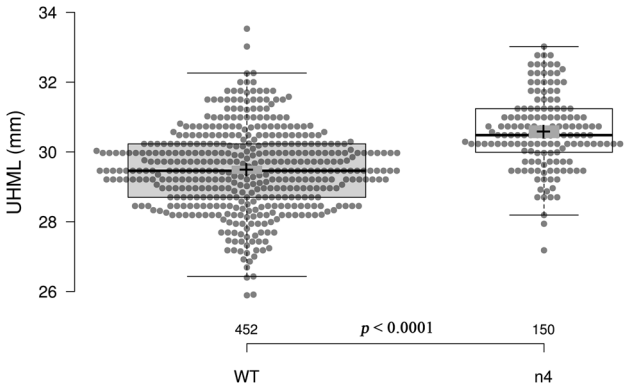

Fig. 3 Statistical analysis of fiber length trait between WT and n_4^t groups of samples in segregating F_2 population. The box plot was built by using R software (http://shiny.chemgrid.org/boxplotr/). Upper half mean length (UHML) of cotton lint fiber is presented in mm. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; crosses represent sample means; bars indicate 95% confidence intervals of the means; width of the boxes is proportional to the square root of the sample size; p value is determined by t test; data points are plotted as grey circles; $n = 452$, 150 sample points

The n_4^t locus is positioned on chromosome D04

Since the n_4^t tufted-fuzzless trait was observed to be recessive, we did not make a DNA bulk from F_2 plants showing the wild type phenotype (because those plants may have a heterozygous genotype). The two DNA pools, including category-1 (six samples) and category-2 (41 samples), from the F_2 segregating population that showed mutant phenotypes along with the parental DNAs were used for whole-genome sequencing. The two categories of bulked samples with different degrees of fuzz on their seeds were sequenced separately to enable the detection of differences in genomic sequences. However, we did not detect noticeable differences in sequences at the candidate loci between two bulks. We identified SNPs and indels by aligning the sequence reads to the reference TM-1 genome (Wang et al. 2019). To build the histogram of the distribution of polymorphisms our initial

Fig. 2 Seeds of F_1 plants and three categories of seeds from F_2 segregating population. **a** F_1 seeds; **b** seeds of category-1; **c** seeds of category-2; **d** seeds of category-3

criteria was that the SNPs and indels were homozygous for the reference (TM-1) type in the parent WT (SC9023), homozygous for an alternative allele type in the parent n_4^t mutant and both homozygous and heterozygous for an alternative allele in F_2 bulks. This criteria of polymorphisms selection showed two major peaks of SNPs and indels on Chr. D07 (3381) and Chr. D04 (673) [Figure S1]. When we used more stringent parameters by selecting only homozygous SNPs and indels for the alternative allele in both F_2 bulks. This approach allowed us to identify only one major locus, containing 164 SNPs and indels on Chr. D04 (Fig. 4). Both initially identified regions of significant polymorphism on D04 and D07 were considered as possible genomic locations of the n_4^t locus. Therefore, we developed PCR-based SNP markers (Table S1) from Chr. D04 and Chr. D07 to test the association of markers with phenotype in the segregating F_2 population.

We ran 14 SNP markers from Chr. D04 and six SNP markers from Chr. D07 on the segregating F_2 population of 602 individuals. The n_4^t tufted-fuzzless phenotype was significantly associated with the Chr. D04 locus and the highest LOD score was associated with the D4_549 marker (Fig. 5a). There was no significant association between SNPs from Chr. D07 and the n_4^t phenotype (Fig. 5b). The flanking region of n_4^t locus was between the D4_546 and D4_550 markers, which is about 440 kb according to the *G. hirsutum* TM-1 reference genome assembly (Wang et al. 2019).

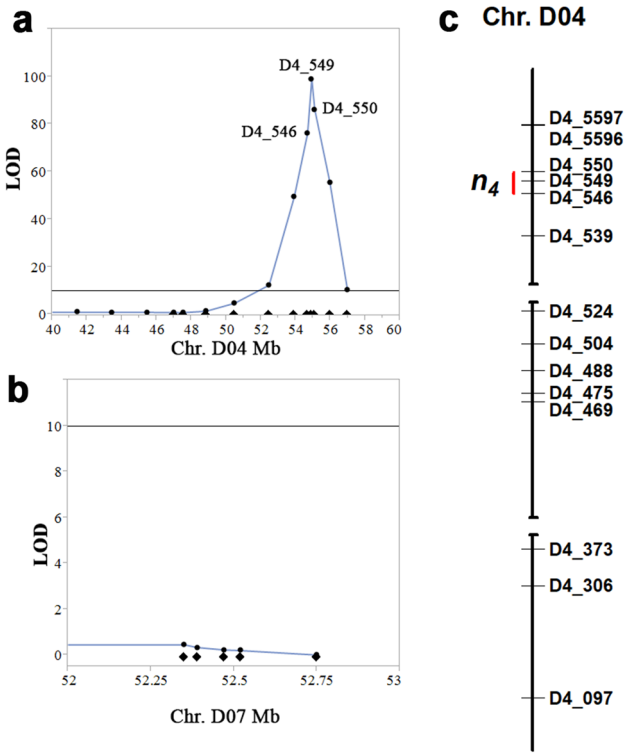

Fig. 5 Association mapping of n_4^t locus. Composite interval mapping analysis of association of the n_4^t tufted phenotype with genetic markers on Chr. D04 (**a**) and Chr. D07 (**b**). **c** Physical map of the n_4^t locus on Chr. D04 *G. hirsutum* TM-1 reference genome. Table S1 provides the sequences of SNP markers, polymorphism and positions on chromosome

Putative candidate genes at the n_4^t locus

The genomic region flanked by our D4_546 and D4_550 SNP markers contained 45 annotated genes according to the TM-1 genome assembly (Wang et al. 2019). To find possible candidate genes we initially evaluated their expression in different cotton tissues by using the publicly available ccNET gene expression database (You et al. 2017), and then the expressed genes were tested by RT-qPCR analysis in early developing fiber cells (DOA, 3 and 5 DPA) of WT and the n_4^t mutant. Of the 45 genes, 42 were available in the ccNET database and among those only 22 genes were expressed above the detection level (Fig. 6). Though, out of 22 expressed genes, only seven genes showed significantly ($p < 0.05$) different expression between WT and the n_4^t mutant in at least one fiber development time point (Fig. 7).

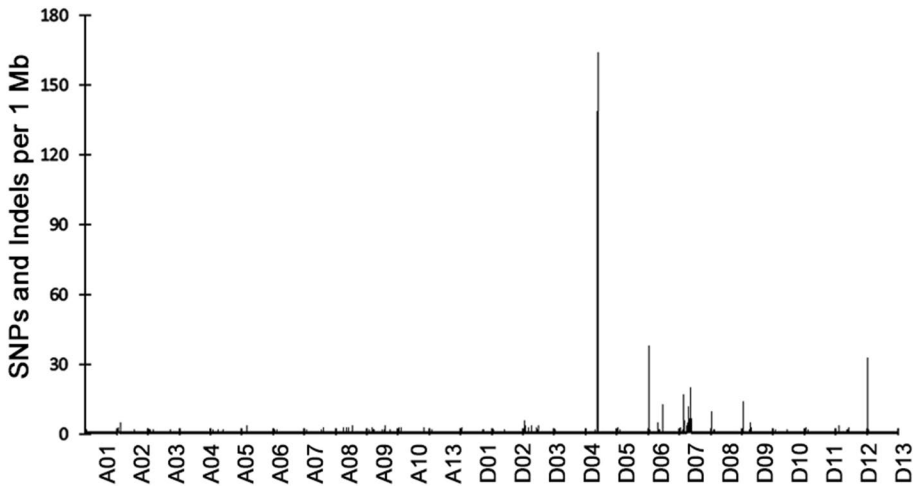

Fig. 4 Histogram of SNPs and indel distribution between the WT and the n_4^t tufted-fuzzless mutant across 26 chromosomes of *G. hirsutum*. The major peak on Chr. D04 represents the SNPs and indels homozygous for an alternative allele type in two F_2 bulks showing the n_4^t phenotype

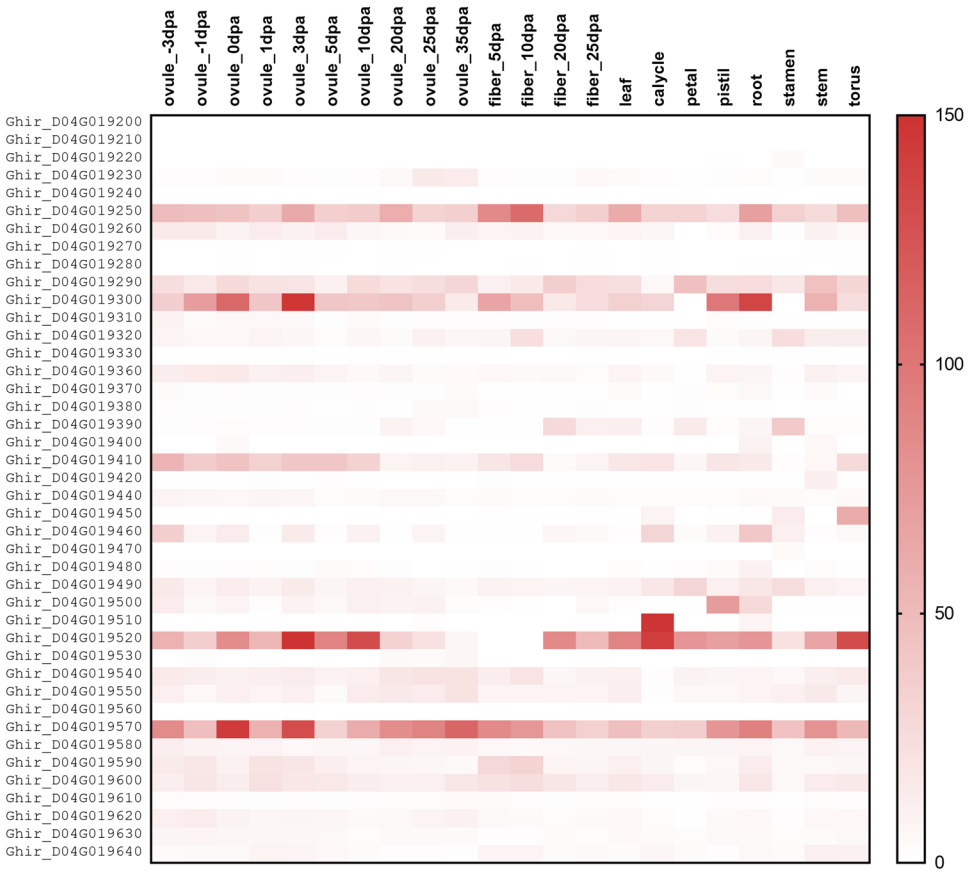

Fig. 6 Heat map of expression of the genes from the n_4^t locus. Expression data for different cotton tissues were obtained from ccNET database (http://structuralbiology.cau.edu.cn/gossypium/)

All seven genes were downregulated in developing n_4^t fiber cells (Fig. 7). *Ghir_D04G019250,* annotated as ribosomal protein L3B, had reduced expression at 3 DPA. *Ghir_D04G019300,* annotated as cytochrome c2, was downregulated at DOA. *Ghir_D04G019410,* annotated as NHL domain transcription factor, was downregulated at 3 DPA. *Ghir_D04G019490,* annotated as lumazine synthase, was downregulated at DOA and 3 DPA. The most highly expressed gene *Ghir_D04G019570,* annotated as ribosomal protein L18e/L15, was downregulated at DOA. *Ghir_D04G019600* and *Ghir_D04G019640,* annotated as o-fucosyltransferase and cellulase were downregulated at 3 DPA.

We compared genomic sequences of differentially expressed genes between WT and the n_4^t mutant. *Ghir_D04G019490* has a SNP in its fifth intron and three SNPs and three indels in its promoter region (Table S3). *Ghir_D04G019640* has an indel in its second intron and a missense substitution in its third exon, changing isoleucine to methionine. The other expressed genes have identical genomic sequence between SC9023 and the n_4^t mutant.

Also, we tested expression levels of the *MML3_A12* and *MML3_D12* genes, which are known to be key regulators of fiber development (Walford et al. 2011; Wan et al. 2016; Zhu et al. 2018). The *MML3_A12* gene was significantly down-regulated in the mutant at DOA, 3 and 5 DPA, whereas the *MML3_D12* was significantly downregulated at 3 and 5 DPA (Fig. 7).

Discussion

The ratio of seed fiber phenotypes in the F_2 population was three (fuzzy) to one (tufted-fuzzless), indicating that the n_4^t tufted-fuzzless trait is controlled by a single recessive locus; this finding was consistent with the segregation ratio reported before (Bechere et al. 2012). The previous study did not detect a significant difference in lint yield and percentage between the n_4^t and WT type cultivar, which with the improved ginning efficiency of fuzzless phenotype makes the n_4^t trait desirable to use in breeding programs (Bechere et al. 2012). In the current study, we analyzed major fiber properties, as measured by HVI, in the F_2 population and found that the fiber length of the fuzzless seeds was significantly longer than in the WT cultivar SC9023 (Fig. 3).

Identification of the n_4^t locus and the gene responsible for tufted-fuzzless trait will provide the means for genetic manipulations to improve ginning efficiency and lint length of modern cotton cultivars. We have identified the candidate locus through genetic mapping of the F_2 population

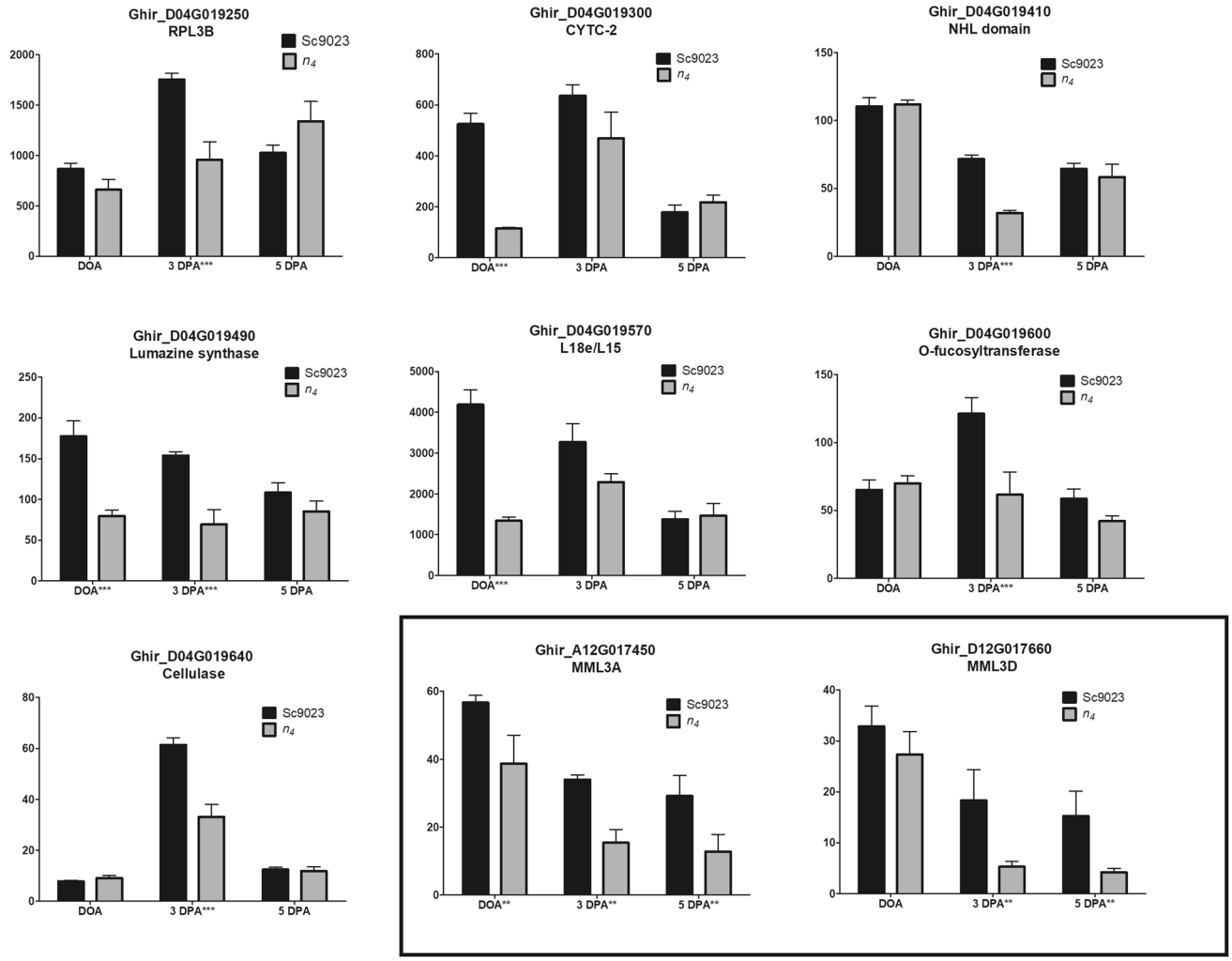

Fig. 7 RT-qPCR analysis of seven selected genes from the n_4^t locus plus two previously reported fiber-related MML3 genes. Error bars indicate standard deviation from three biological replicates with two technical replicates. Asterisks on *x*-axis indicate significantly different expressions between the wild type (Sc9024) and the n_4^t samples. The number of asterisks reflects the level of significance (**$p < 0.001$ and ***$p < 0.0001$). Table S2 provides primers sequences and annotation of the genes used in RT-qPCR

(Fig. 5). The n_4^t locus has been mapped to Chr. D04 where previously no genes have been associated with fiber development. Seven genes from the n_4^t locus showed significant downregulation in developing fibers of the fuzzless mutant compared to the WT. *Ghir_D04G019570,* annotated as ribosomal protein *L18e/L15,* was the most highly expressed gene at the n_4^t locus and was significantly downregulated in the tufted-fuzzless mutant at DOA (Fig. 7). The ribosomal *L18e/L15* gene has been associated with fiber strength QTL in an independent study and showed high expression in the superior fiber MD52ne cultivar compared to a weaker fiber line MD90ne (Islam et al. 2016).

We have detected polymorphisms in the genomic sequence of the n_4^t fuzzless mutant in two genes. *Ghir_D04G019640,* annotated as cellulase (glycosyl hydrolase family 5), has a substitution of isoleucine to methionine that can lead to changes in protein structure or function. Cellulases catalyze the breaking of β-1, 4 glycosidic bonds in the interior of longer β-1, 4 glucose chains; genes from this family are suggested to possess multiple functions, including accelerating cell growth through cell wall loosening, and wall reconstruction during cell expansion (Cosgrove 2005; Yu et al. 2013). Antisense suppression of poplar cellulase (PopCel1) reduced leaf growth (Ohmiya et al. 2003). Higher expression of PopCel1 in *Arabidopsis* increases the size of the rosettes through promoting an increase in cell size (Park et al. 2003). Research in *Arabidopsis* and poplar showed the involvement of cellulases in leaf growth; however, the role of the *Ghir_D04G019640* in cotton fiber development has not been studied yet. Given the expression pattern of the *Ghir_D04G019640* in developing fiber cells (Fig. 7) and the observation of

downregulation of this gene at 3 DPA in the n_4^t mutant we suggest this gene as a candidate for further studies to understand its role in fuzz fiber development.

We detected SNPs and indels in the promoter region of the *Ghir_D04G019490* gene, annotated as lumazine synthase. Lumazine synthase catalyzes the penultimate step of riboflavin biosynthesis, which is essential for critical cellular processes such as the citric acid cycle, fatty acid oxidation, photosynthesis, mitochondrial electron transport, and de novo pyrimidine biosynthesis (Jordan et al. 1999). The *Ghir_D04G019490* gene is our favored candidate since it is located close to SNP marker D04_549, which showed the highest LOD score association with the fuzzless phenotype (Fig. 5). It is unclear what causes downregulation of this gene in the mutant; though, polymorphism in the promoter region of the *Ghir_D04G019490* should not be ruled out as a causative factor of expression change. The function of the *Ghir_D04G019490* gene is not characterized yet in cotton. Results of this study indicate this gene might be involved in fuzz fiber development and should be considered for future research.

What causes different degrees of fuzz fiber cell development on the n_4^t seeds is unclear. The genomic sequences of the two studied categories of seeds with different degrees of fuzz were similar, therefore could not explain why such differences occur. However, factors such as small RNAs, epigenetic modifications or other mechanisms that modulate penetrance should not be ruled out. For example, the N_1 mutant has a wide range of fiber densities, even on the bolls of a single plant (Wan et al. 2016). Wan et al. (2016) suggested that the randomness of the small RNA distribution pattern and the epigenetic modifications might be associated with the randomness of the fiberless phenotype in the N_1 mutant.

The current model of fuzz fiber development in tetraploid *Gossypium* species proposes that it is mainly determined by the expression level of *MML3-D* around 3 DPA (Zhu et al. 2018). Since we have detected significant downregulation of both *MML3-A* and *MML3-D* genes in the n_4^t fuzzless mutant at 3 and 5 DPA (Fig. 7), we do not challenge the key role of MML3 family genes. But, we have identified a new n_4^t locus that has not previously been associated with fuzz fiber formation, which suggests that there are more unknown genes involved in fiber development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00438-021-01802-0.

Acknowledgements Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U. S. Department of Agriculture, which is an equal opportunity provider and employer.

Author contributions MN conceived the research, designed experiments, and drafted the manuscript. GNT performed bioinformatics analysis. DDF oversaw the mapping experiment. EB developed the n_4^t mutant. LP and CBF conducted the experiments. All authors read and approved the manuscript.

Funding This research was funded by the United States Department of Agriculture-Agricultural Research Service CRIS project 6054-21000-018-00D.

Declaration

Conflict of Interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Bechere E, Auld DL (2014) Registration of a tufted-naked seed upland cotton germplasm, 9023n_4^t. J Plant Regist 8(1):63–67

Bechere E, Boykin JC et al (2011) Evaluation of cotton genotypes for ginning energy and ginning rate. J Cotton Sci 15(1):11–21

Bechere E, Turley RB et al (2012) A new fuzzless seed locus in an upland cotton (*Gossypium hirsutum* L.) mutant. Am J Plant Sci 3(06):799

Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol 6(11):850–861. https://doi.org/10.1038/nrm1746

Dettmer K, Aronov PA et al (2007) Mass spectrometry-based metabolomics. Mass Spectrom Rev 26(1):51–78

Fang DD, Xiao J et al (2010) A new SNP haplotype associated with blue disease resistance gene in cotton (*Gossypium hirsutum* L.). Theor Appl Genet 120(5):943–953. https://doi.org/10.1007/s00122-009-1223-y

Feng X, Cheng H et al (2019) Fine mapping and identification of the fuzzless gene GaFzl in DPL972 (*Gossypium arboreum*). Theor Appl Genet 132(8):2169–2179

Islam MS, Fang DD et al (2016) Comparative fiber property and transcriptome analyses reveal key genes potentially related to high fiber strength in cotton (*Gossypium hirsutum* L.) line MD52ne. BMC Plant Biol 16(1):1–19

Jordan DB, Bacot KO et al (1999) Plant riboflavin biosynthesis: cloning, chloroplast localization, expression, purification, and partial characterization of spinach lumazine synthase. J Biol Chem 274(31):22114–22121

Kearney TH, Harrison GJ (1927) Inheritance of smooth seeds in cotton. J Agric Res 35:193–217

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

Liu X, Moncuquet P et al (2020) Genetic identification and transcriptome analysis of lintless and fuzzless traits in *Gossypium arboreum* L. Int J Mol Sci 21(5):1675

Michelmore RW, Paran I et al (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci 88(21):9828–9832

Naoumkina M, Thyssen GN et al (2020) Elucidation of sequence polymorphism in fuzzless-seed cotton lines. Mol Genet Genom. https://doi.org/10.1007/s00438-020-01736-z

Obata T, Fernie AR (2012) The use of metabolomics to dissect plant responses to abiotic stresses. Cell Mol Life Sci 69(19):3225–3243. https://doi.org/10.1007/s00018-012-1091-5

Ohmiya Y, Nakai T et al (2003) The role of PopCel1 and PopCel2 in poplar leaf growth and cellulose biosynthesis. Plant J 33(6):1087–1097

Park YW, Tominaga R et al (2003) Enhancement of growth by expression of poplar cellulase in *Arabidopsis thaliana*. Plant J 33(6):1099–1106

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res 29(9):e45–e45

Song L, Guo WZ et al (2010) Genetic analysis and molecular validation of chromosome assignment for fuzzless genes N1 and n2 in cotton. J Nanjing Agric Univ 33:21–26

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

Takagi H, Abe A et al (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74(1):174–183

Taliercio EW, Boykin D (2007) Analysis of gene expression in cotton fiber initials. BMC Plant Biol 7(1):22

Thyssen GN, Fang DD et al (2014) Next generation genetic mapping of the Ligon-lintless-2 (Li_2) locus in upland cotton (*Gossypium hirsutum* L.). Theor Appl Genet 127(10):2183–2192. https://doi.org/10.1007/s00122-014-2372-1

Turley R, Kloth R (2002) Identification of a third fuzzless seed locus in upland cotton (*Gossypium hirsutum* L.). J Hered 93(5):359–364

Walford SA, Wu Y et al (2011) GhMYB25-like: a key factor in early cotton fibre development. Plant J 65(5):785–797. https://doi.org/10.1111/j.1365-313X.2010.04464.x

Wan Q, Guan X et al (2016) Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. New Phytol 210(4):1298–1310

Wang M, Tu L et al (2019) Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet 51(2):224–229

Ware JO, Benedict LI et al (1947) A recessive naked-seed character in upland cotton. J Hered 38(10):313–320

Wu H, Tian Y et al (2017) Genetics and evolution of MIXTA genes regulating cotton lint fiber development. New Phytol 217(2):883–895. https://doi.org/10.1111/nph.14844

You Q, Xu W et al (2017) ccNET: database of co-expression networks with functional modules for diploid and polyploid Gossypium. Nucleic Acids Res 45(D1):D1090–D1099

Yu L, Sun J et al (2013) PtrCel9A6, an endo-1, 4-β-glucanase, is required for cell wall formation during xylem differentiation in Populus. Mol Plant 6(6):1904–1917

Zhang D, Zhang T et al (2007) Comparative development of lint and fuzz using different cotton fiber-specific developmental mutants in *Gossypium hirsutum*. J Integr Plant Biol 49(7):1038–1046. https://doi.org/10.1111/j.1672-9072.2007.00454.x

Zhu Q-H, Yuan Y et al (2018) Genetic dissection of the fuzzless seed trait in *Gossypium barbadense*. J Exp Bot 69(5):997–1009. https://doi.org/10.1093/jxb/erx459

Zhu Q-H, Stiller W et al (2020) Genetic mapping and transcriptomic characterisation of a new fuzzless-tufted cottonseed mutant. G3 Genes Genomes Genet 11:1–14