

RESEARCH

Open Access



# Mapping of dwarfing QTL of Ari1327, a semi-dwarf mutant of upland cotton

Chenhui Ma<sup>1</sup>, Abdul Rehman<sup>2,3</sup>, Hong Ge Li<sup>1</sup>, Zi Bo Zhao<sup>2</sup>, Gaofei Sun<sup>4</sup> and Xiong Ming Du<sup>1\*</sup>

## Abstract

**Background:** Upland Cotton (*Gossypium hirsutum* L.) has few cotton varieties suitable for mechanical harvesting. The plant height of the cultivar is one of the key features that need to modify. Hence, this study was planned to locate the QTL for plant height in a <sup>60</sup>Co γ treated upland cotton semi-dwarf mutant Ari1327.

**Results:** Interestingly, bulk segregant analysis (BSA) and genotyping by sequencing (GBS) methods exhibited that candidate QTL was co-located in the region of 5.80–9.66 Mb at D01 chromosome in two F<sub>2</sub> populations. Using three InDel markers to genotype a population of 1241 individuals confirmed that the offspring's phenotype is consistent with the genotype. Comparative analysis of RNA-seq between the mutant and wild variety exhibited that *Gh\_D01G0592* was identified as the source of dwarfness from 200 genes. In addition, it was also revealed that the appropriate use of partial separation markers in QTL mapping can escalate linkage information.

**Conclusions:** Overwhelmingly, the results will provide the basis to reveal the function of candidate genes and the utilization of excellent dwarf genetic resources in the future.

**Keywords:** Cotton, Semi-dwarf, QTL, Segregation distortion

## Introduction

Cotton is an important cash crop worldwide, accounting for about 35% of the world's fiber production [1]. Mechanical picking of cotton is a new trend of modern agriculture, but plant height (PH) is one of the most important factors which impede mechanical harvesting [2]. Nowadays, cotton plant height is controlled chemically, which adds additional cost and also leads to chemical contamination [3, 4]. Alternatively, the most reliable and feasible method is the use of natural semi-dwarf plants. The semi-dwarf traits can be used for mechanized management and increase lodging resistance and photosynthesis [5]. In contrast, the semi-dwarf natural resources of cotton are still limited due to fewer reports

on the discovery of dwarf mutants and QTL (Quantitative Trait Loci) locations in related mutants.

Quisenberry et al. found a Lubbock dwarf mutant with poor fiber quality [6]. Abzalo et al. found a homozygous dominant lethal dwarf mutant line L-691, in cotton [7]. Moreover, some researchers also revealed that the phenotype of cotton mutants is closely linked with plant height [8, 9]. Others dwarf mutants showed extreme phenotypes without application potential. An ultra-dwarf mutant of cotton controlled by a *du* gene, and its dwarfing gene was located on chromosome A06 [10]. *GhLi-1* was a classical dwarf mutant of upland cotton, with short fibers. An SNP variant G65V affected its typical polymerization and biological processes based on the actin cytoskeleton (such as intracellular transport), which ultimately restrict cell elongation [11]. In cotton, *AS98* is a super-dwarf mutant with a plant height of 25.6 cm, and the candidate QTL was located between markers GH537 and E4M13 [12]. Recently, dwarf QTL was located in the F<sub>2</sub> population

\*Correspondence: duxiongming@caas.cn

<sup>1</sup> State Key Laboratory of cotton Biology, Institute of Cotton Research of the Chinese Academy of Agricultural Sciences, Anyang 455000, China  
Full list of author information is available at the end of the article



of AS98 through BSA and revealed that dwarfness was caused by the copy number variation of *GhDREB1B* [13].

Plant height was provoked by multiple developmental factors, i.e., the number of phytomeres, stem length, and plant hormones [14]. Dwarf mutant *sd1* of rice was regulated by GA (Gibberellic Acid), which exhibited that plant hormone like GA plays a vital role in the green revolution and plant height control [15]. Negatively regulated *GA-DELLA* protein can interact with many transcription factors and ultimately affect cell elongation. Rice *SLR1* (*DELLA*) protein can directly interact with transcription factors *NAC29* and *NAC31*, which mediate *MYB61* and *cellulose synthase* genes [16]. In Arabidopsis stem tips, the *DELLA* protein inhibited several TCP transcription factors involved in the cell cycle [17]. Proteomic analysis of upland cotton revealed that its dwarfness might be regulated by GA pathways [18]. BR (Brassinosteroid) is another plant hormone that controls plant height by regulating cell elongation. Moreover, it was also noticed that disturbance in BR biosynthesis or signaling pathway reduce plant height significantly in rice and Arabidopsis mutants [19, 20]. IAA (Indole Acetic Acid) can control young leaves and roots apical meristem, promote root elongation, stem growth and flower differentiation. The IAA efflux transporter (PIN) family play a key role in plant height. Overexpression of both *ZmPIN1a* and *OsPIN2* can reduce plant height [21, 22]. Thiamin (vitamin B1) is involved in chloroplast photosynthesis and mitochondrial tricarboxylic acid cycle [23–25]. Research on various crop plants has shown that changes in thiamine levels affect plant height, yellowing of leaves, delay

growth and development, delay flowering, and ultimately reduce yield [26–28].

With the development of sequencing technology, multiple high-throughput sequencing methods are widely used in QTL identification. Sun et al. successfully mapped 12 salt-tolerance-related genes using 196 upland cotton species from GWAS (Genome-Wide Association Studies) results and RNA-seq data [29]. 12 QTL related to lint percentage were located in 355 cotton genotypes by SLAF-GWAS in four environments and exhibited *Gh\_A02G1268* responsible for lint percentage [30]. Map-based cloning is a research method to study the QTL of separated populations, but phenotypic values of the permanently separate population are essential to locate accurately in multiple environments [1, 31].

This study mapped semi-dwarf QTL of mutant cotton genotype Ari1327 precisely by GBS, BSA and RNA-seq database to locate the candidate gene area on the D01 chromosome. This work will provide a foundation to verify the dwarfing function of candidate genes and ultimately realize the use of excellent dwarfing resources.

## Results

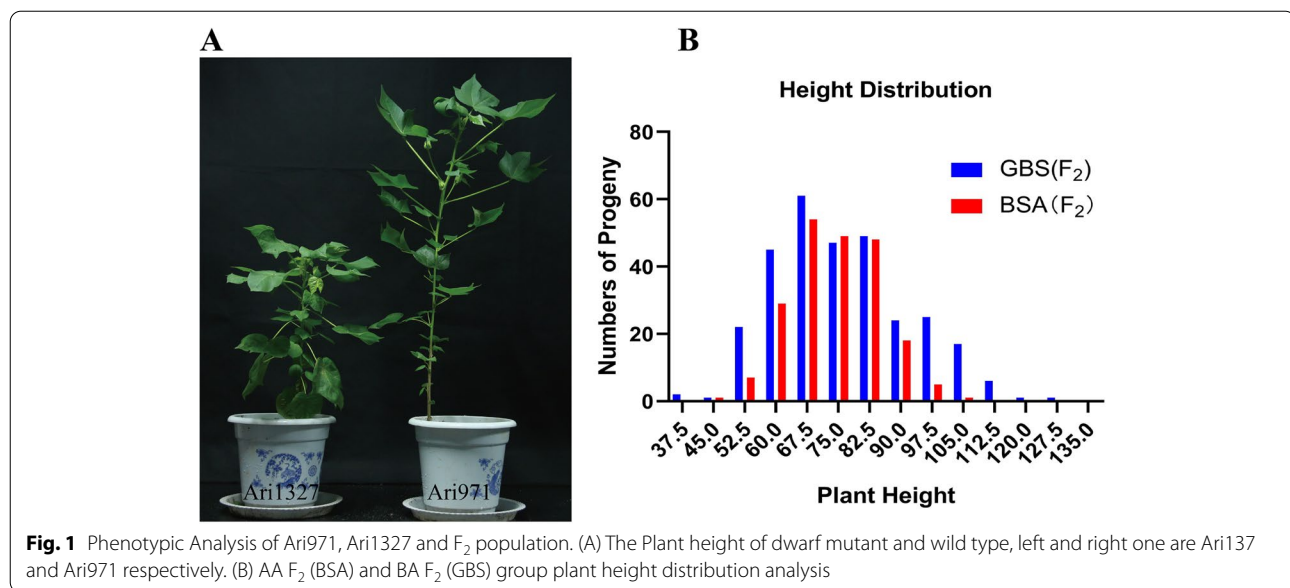
### Phenotypic analysis

Ari1327 was a semi-dwarf mutant as compared to wild-type Ari971. Significant differences were observed between 16 agronomic and fiber traits for mutant and wild types in filed conditions. The results exhibited that mutant and wild type had plant height of 48.60 cm and 76.15 cm, respectively (Table 1 and Fig. 1A). The plant height of Ari1327 was only 68% of Ari971. Ari971 has 13

**Table 1** Statistical analysis of significance of morphological traits of cotton

Traits	Ari1327	Ari971	Significance
Average length of upper fibers (mm)	28.10 ± 0.95	30.07 ± 1.86	**
Boll Number	11.00 ± 2.75	11.45 ± 4.37	
Fiber Elongation	9.37 ± 1.59	9.46 ± 1.55	
Fiber Maturity	0.86 ± 0.02	0.84 ± 0.03	
Fiber Strength(cN/tex)	28.49 ± 2.68	30.85 ± 2.13	*
Fiber Uniformity	85.02 ± 1.28	85.99 ± 1.29	*
Fruit Branch Length	29.77 ± 10.97	33.80 ± 12.98	
Fruit branch number	10.50 ± 2.04	13.00 ± 1.45	**
Leaf area	145.10 ± 54.43	145.71 ± 24.96	
Leaf number	5.63 ± 2.58	5.79 ± 3.06	
Lint Percentage	36% ± 0%	40% ± 1%	*
Lint weight (30 Bolls)	52.75 ± 2.47	70.05 ± 1.91	*
Micronaire	4.97 ± 0.51	4.36 ± 0.84	*
Plant height	48.60 ± 11.54	76.15 ± 10.12	**
Seed Cotton weight (30 Bolls)	148.60 ± 8.34	176.55 ± 10.68	
Short Fiber (%)	5.92 ± 1.6	5.31 ± 0.86	

\* mean significance, \*\* mean the differences are highly significant at the 0.05 and 0.01 level respectively



sympodial branches, 2.5 more than Ari1327. Among 11 fiber quality traits, only the average length of upper fibers (FL) revealed significant difference, while in mutant 7% reduction was also observed ( $P < 0.01$ ). Among the five morphological data related to cotton growth, only the number of sympodial branches and plant height showed significant differences ( $P < 0.01$ ).

FL (average length of upper fibers), FU (fiber uniformity), FS (fiber strength), MC (micronaire), FE (fiber elongation) and PH (plant height) in the BA F<sub>2</sub> population also revealed that fiber quality traits and plant height lack significant correlation (Additional file 1). Results also depicted that the correlation coefficient between PH and the other five fiber quality traits was  $\pm 0.15$ , with no significant correlation. While MC and FL, FU, FS showed a negative correlation. Moreover, the correlation between FL, FU, FS, and FE was higher than 0.45. Hence, it concluded that PH does not correlate with fiber quality traits hence, inherited independently.

Ari3697 (Group AA) and BL-Y10 (BA group) were used as two female parents to construct two F<sub>2</sub> populations to identify the dwarfing QTL in Ari1327. Their PH values were  $105.85 \pm 7.81$  cm and  $93.70 \pm 6.48$  cm, respectively. The PH of male parent Ari1327 was  $46.43 \pm 5.57$  and  $67.36 \pm 6.04$  in the two populations. The number of individuals in the group AA and group BA were 214 and 303, respectively. The skewness of plant height of two populations was greater than 0, biased towards semi-dwarf parents. When the kurtosis is equal to 0, it predicts normal distribution characteristics. After the Shapiro-Wilk normality test of R software, the  $W$  values and  $p$ -values of the BA and AA groups were 0.99419, 0.97534 and 0.5844,

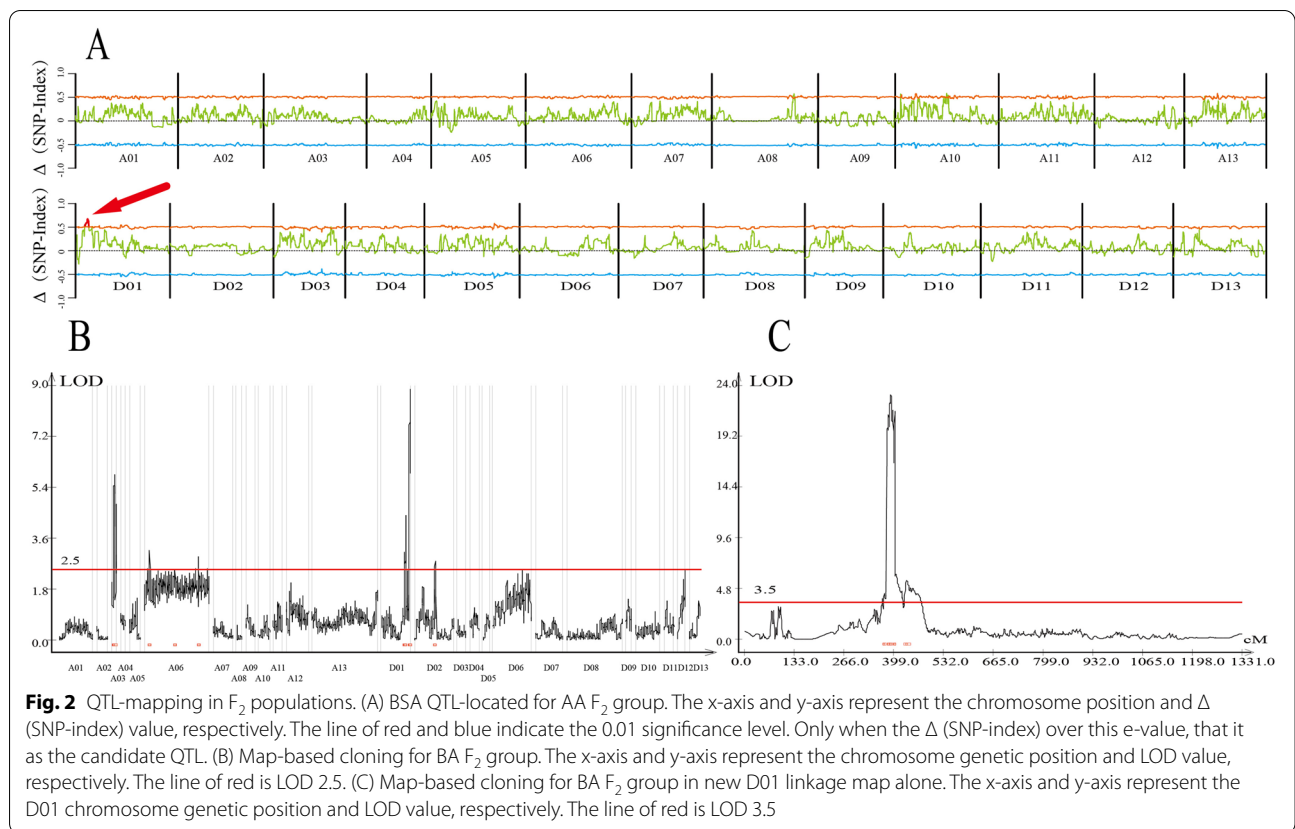
4.779e-05, respectively. It shows that the BA group confirmed the normal distribution (Fig. 1B, Additional file 9 and Additional file 10).

#### BSA QTL locate in AA F<sub>2</sub> population

BSA method was used to identify QTL of group AA with an  $e$ -value greater than 99%. A total of 10 QTLs were detected on five chromosomes. According to the QTL nomenclature, we named the QTL as qPH-A08-(BSA)01, qPH-A08-(BSA)02, up to qPH-A08-(BSA)10 (Fig. 2A and Additional file 9). Although a total of 8 QTL were identified on chromosomes A08, A10, A13, and D03, and  $\Delta$  (SNP-index) of each QTL just over the threshold line. Thus, we focus on two QTLs that have higher values on chromosome D01, qPH-D01-(BSA)01 and qPH-D01-(BSA)02. The physical position of qPH-D01-(BSA)01 and qPH-D01-(BSA)02 were 6.07–9.66 Mb and 10.69–12.43 Mb, respectively.

#### GBS QTL locate in BA F<sub>2</sub> population

To verify the AA group mapping results, we used another F<sub>2</sub> population with GBS method to locate the QTL of PH. SLAF sequencing identified a total of 1,543,942 SNPs in two parents. After filtering the SNP site, 7743 marker sites were screened with completeness (75%) and SD ( $p > 0.001$ ). The linkage map was constructed by Joinmap4.0 with 5046 SNP markers in the end. The number of SNP markers and the total genetic distance of each linkage group are shown in Additional file 2 and Additional file 3. The total genetic distance is 5954.87 cM, the longest linkage group length is 713.94 cM, and the maximum gap is 16.94 cM. SNP markers were used from 26



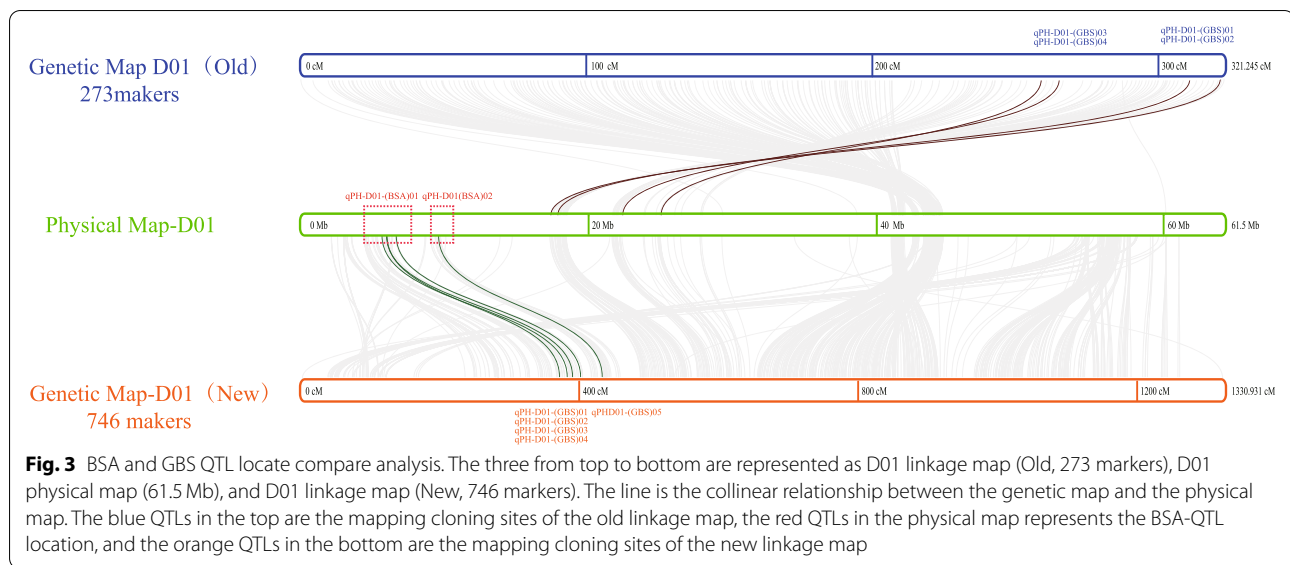
chromosomes and divided into 26 linkage groups, hence, the chromosome number is used instead of the linkage group number. In total, 11 QTL for PH were detected on four chromosomes. Among all QTLs, 6th and 5th QTL

were identified on the At and Dt chromosomes respectively (Table 2 and Fig. 2B). On ChrA03, QTL exhibited 15.46% of PVE and 4.8 of LOD between marker mk626\_A03 and mk7354\_A03. Only *Gh\_A03G0832* and

**Table 2** QTL information for plant height in AA F<sub>2</sub> group

QTL Name	Chr	Position (genetic)	Position (physical)	LOD	Additive effect	Dominant effect	R <sup>2</sup>
qPH-A03-(GBS)01	A03	26.5	93,727,847	5.4587	-9.5239	-5.2898	0.0004
qPH-A03-(GBS)02	A03	34.2	93,585,329	5.861	-9.8883	-4.5461	0.001
qPH-A03-(GBS)03	A03	50.1	27,371,589	4.8108	8.5959	-5.3265	0.1546
qPH-A06-(GBS)01	A6	61	93,928,717	3.1788	-4.0743	-3.2641	0.0137
qPH-A06-(GBS)02	A6	338.6	70,513,686	2.5121	-3.585	-2.5447	0.0131
qPH-A06-(GBS)03	A6	597.7	112,971	2.9496	-4.0087	-3.0578	0.0133
qPH-D01-(GBS)01	D01	256.6	25,086,049	3.1118	7.8675	-2.9575	0.0845
qPH-D01-(GBS)02	D01	263.8	22,413,590	4.4199	10.505	-2.2176	0.0976
qPH-D01-(GBS)03	D01	310.1	17,756,478	8.563	-14.693	-3.784	0.0341
qPH-D01-(GBS)04	D01	318.6	17,407,200	8.879	-15.2109	-3.6284	0.0338
qPH-D02-(GBS)01	D02	223.3	44,349,602	2.7979	-4.0764	-2.8479	0.0131
<b>qPH-D01-(GBS)01</b>	<b>D01</b>	<b>373.5</b>	<b>5,887,414</b>	<b>4.3594</b>	<b>3.0558</b>	<b>-7.2987</b>	<b>0</b>
<b>qPH-D01-(GBS)02</b>	<b>D01</b>	<b>385.6</b>	<b>6,056,815</b>	<b>21.2673</b>	<b>12.8405</b>	<b>-7.5648</b>	<b>0.3529</b>
<b>qPH-D01-(GBS)03</b>	<b>D01</b>	<b>391.3</b>	<b>6,023,764</b>	<b>23.1031</b>	<b>13.061</b>	<b>-7.7263</b>	<b>0.3611</b>
<b>qPH-D01-(GBS)04</b>	<b>D01</b>	<b>402.7</b>	<b>6,698,093</b>	<b>5.967</b>	<b>11.7556</b>	<b>-8.0992</b>	<b>0.15</b>
<b>qPH-D01-(GBS)05</b>	<b>D01</b>	<b>433.9</b>	<b>9,645,343</b>	<b>5.587</b>	<b>-6.5948</b>	<b>-7.2828</b>	<b>0.014</b>

The blue background reflect the old linkage map QTLs results. The italic and bold font highlight the new D01 linkage map QTLs results



*Gh\_A03G0833* existed with annotation IPMS2 and RPD1 but showed no differences in expression.

#### Comparison analysis of BSA and GBS result

QTL located in chromosome D01 were focused because the first group AA (BSA method) QTL were exactly located on it with a higher e-value. Four QTL qPH-D01-(GBS)01, qPH-D01-(GBS)02, qPH-D01-(GBS)03 and qPH-D01-(GBS)04 were found on D01 with explaining the phenotypic variation of 8.45, 9.76, 3.41 and 3.38% respectively, which can be divided into two QTL clusters. The additive effect of the first QTL cluster is positive, revealing the predominance of male parent Ari1327 on the short stature of the plant.

GBS scanned the physical positions of QTL to visualize the overlap area between the two groups. However, D01 chromosome contains 273 SNP markers with a total genetic distance of 321.25 cM and an average genetic distance of 1.18 cM. Moreover, uneven distribution of D01 markers was observed mainly in the front end (Fig. 3). SNPs deficient area was exactly the candidate area identified by group AA (BSA method). However, the physical position in the BA QTL location was still closest to the AA QTL location despite the lack of good markers (Fig. 3 middle, green). So, we re-examined the GBS-D01 markers. It was found that when we performed the chi-square test on the separation ratio of 1:2:1 ( $p < 0.1$ ), after correction ( $P < 0.0001$ ), only 273 markers were obtained, which was the deep orange block on line “marker choose” (Additional file 4). Whether we checked the chi-square value on Additional file 12 or the genotype distribution on Additional file 4 of the AA population, we can find that many markers have been deleted due to statistical

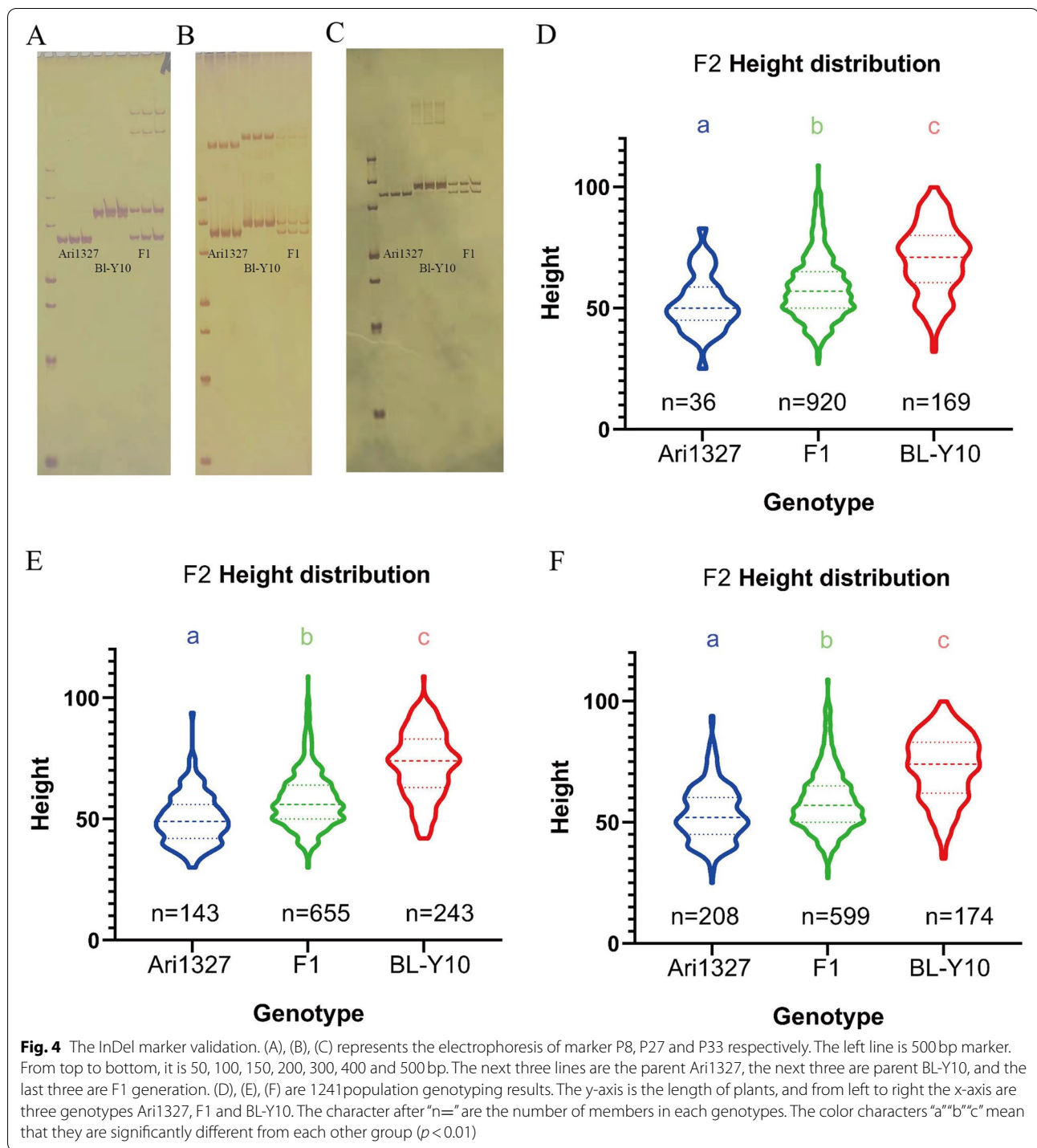
differences. So we decided to increase the density of the linkage map by using SDM (segregation distortion markers) to cover the complete chromosome, especially in the front end (Fig. 3 below, orange).

The new linkage map of a single D01 chromosome consists of 746 markers with a total genetic distance of 1330.931 cM. The new linkage map can cover the entire D01 chromosome front and back end as well. Five QTL were found during re-locating with a higher threshold  $LOD > 3.5$ . Among them, qPH-D01-(GBS)01, qPH-D01-(GBS)02, qPH-D01-(GBS)03 and qPH-D01-(GBS)04 were responsible for contributing dwarf allele from the male parent Ari1327. These four QTLs range from 0.81 Mb along from 5.89 Mb to 6.7 Mb, thus considered a QTL cluster. The last three QTLs explained the phenotypic variation at 35.29, 36.11 and 15% respectively. qPH-D01-(GBS)05 was located at 9.65 Mb region, which is the second QTL or QTL cluster.

After increasing the linkage map density by using the SDM, the QTL positioning results of BSA and GBS had overlapped. The max candidate area was 5.80 to 9.66 MB on D01 from 200 genes, including *Gh\_D01G0486* to *Gh\_D01G0685*.

#### QTL validation by InDel marker

Three InDel markers were used to verify the QTLs, namely p8, p27 and p33 on D01:5818220, 5,930,069 and 6,004,746 bp (Additional file 11). In all 1241 individuals, the ratio of  $F_1$ -like genotype was all more than 50%. The three group's genotype distribution was beyond the segregation ratio of 1:2:1 ( $p < 0.01$ ). For primer P8, the number of H genotypes was 920, and only 36



individuals were homozygous from the dwarf parent, showing a more severe SD.

The estimates of Ari1327 genotype on populations were  $52.08 \pm 12.03$ ,  $50.08 \pm 10.23$  and  $53.58 \pm 11.42$ , significantly less than those had BL-Y10 genotype. They had a higher value of  $70.07 \pm 14.37$ ,  $72.67 \pm 13.60$  and

$72.6 \pm 13.62$  respectively ( $p < 0.01$ ) as shown in Fig. 4. The main QTL that causes Ari1327 to become shorter existed in the candidate area.

### RNA-seq analysis between mutant Ari1327 and wild type Ari971

Although the two QTL mapping methods can successfully co-locate the dwarf QTL on a small interval, still, it contains many candidate genes. To screen of desired gene, gene expression analyses of wild-type and mutant were performed. RNA-seq analysis was carried out in various time intervals for PH of wild-type and mutant plants. Results revealed that both genotypes differed from each at two true leaf stages. The results were more significant at the four true leaves stage, while obvious and highly significant differences in plant height were observed at the six-leaf stage (Fig. S4). Hence, 222, 491 and 1909 DEGs (Differential Expressed Genes) were observed at 2nd, 4th, and 6th leaf stages. 29 DEG existed in the three times but no one existed in the D01 candidate gene area. DEGs at the 2nd leaf stage were not enriched in any pathway, and DEGs at the 4th leaf stage were enriched in four GO pathways and one KEGG pathway. The 6th leaf stage had many DEGs, so all of them can be enriched into 19 GO and 19 KEGG pathways. Four pathways related to the cell cycle and signal pathways like M00693, ko04110, ko04712 and ko04010 were found (Additional file 12). Since, the DEGs shared in the first three time intervals were not in the candidate area, we re-analyzed the expression of all genes (Additional file 12). Genes were selected with  $FC > |1.5|$ ,  $FPKM > 0.5$ , and  $P < 0.01$  to construct a heat map. In order to better display the results, the heat map was calculated through log values.

Among the 200 identified genes in the candidate area, 21 genes showed significant differences at least in one time interval. The expression of *Gh\_D01G0503*, *Gh\_D01G0585*, *Gh\_D01G0586* and *Gh\_D01G0592* in the mutant was higher than that of wild type during 4th and 6th leaf stages. *Gh\_D01G0531* expressed lower at 6th leaf stage in mutant than wild type. Except for these five genes with huge differences, the other 16 genes can be grouped into three clusters according to their expression pattern. The first type contained genes i.e., *Gh\_D01G0666*, *Gh\_D01G0534*, *Gh\_D01G0642*, *Gh\_D01G0545*, *Gh\_D01G0500* and *Gh\_D01G0646* exhibited lower expression values. The second class included *Gh\_D01G0522*, *Gh\_D01G0520*, *Gh\_D01G0514*, *Gh\_D01G0498*, *Gh\_D01G0502*, *Gh\_D01G0547* and *Gh\_D01G0608*, all of which are expressed higher than wild type at six leaves stage. The other three genes i.e., *Gh\_D01G0528*, *Gh\_D01G0589* and *Gh\_D01G0551* were grouped into third class. Only *Gh\_D01G0551* has a notably higher expression at 4th leaf stage in the mutant (Fig. 5).

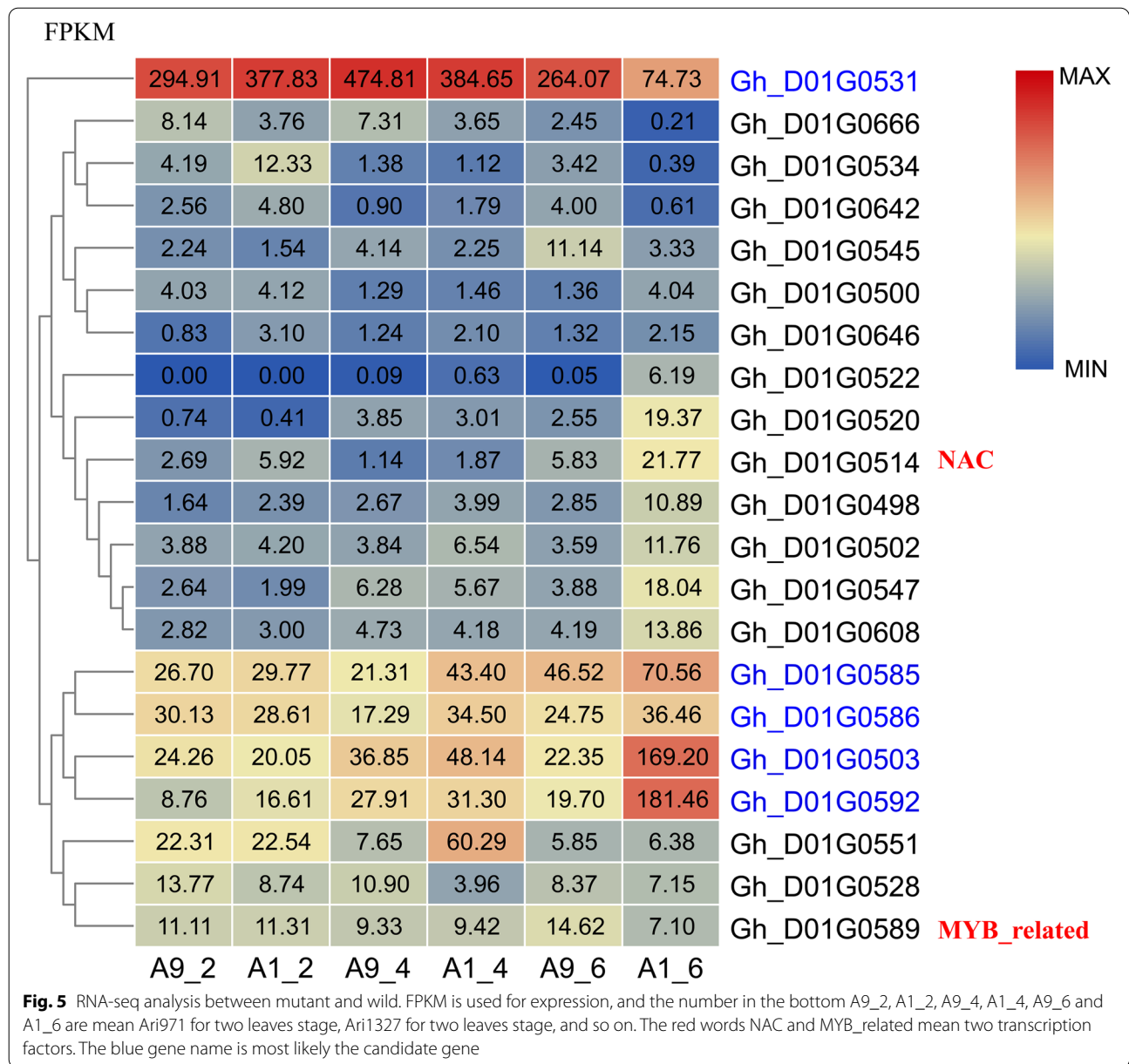
qRT-PCR was used to verify the expression of the five candidates. These genes all showed significant expression differences in the six leaves stage,  $p < 0.01$ . The expression trends of *Gh\_D01G0503*, *Gh\_D01G0531*, *Gh\_D01G0585* and *Gh\_D01G0592* were consistent with RNA-seq data. *Gh\_D01G0503* and *Gh\_D01G0592* were highly expressed in the mutant with three to four-fold differences than wild. Furthermore, *Gh\_D01G0531* showed almost threefold expression of the wild type. The expression differences of *Gh\_D01G0585* and *Gh\_D01G0586* were showed 1.5 folds and the expression trend of *Gh\_D01G0586* was opposite to RNA-seq.

### Discussion

Semi-dwarf is an essential trait for cotton mechanical picking and yield management [32, 33]. But the discovery and utilization of natural semi-dwarf mutants are still insufficient. Previous studies from our lab have found a semi-dwarf mutant Ari1327 different only in plant height from wild-type Ari971. Preliminary physiological studies exhibited that the length of the mutant's second and fourth internodes was shorter than wild. However, the number of internodes remain same [34]. In the current study, re-evaluation of the performance of wild-type and mutant were performed based on agronomic traits. Significant differences were observed in FL, the number of sympodial branches, PH, whereas mutant showed short fiber length compared to wild. As we all know, the number of sympodial branches, number of bolls, lint weight (30 bolls) and seed cotton weight (30 bolls) are directly proportional to plant height. Although the yield per plant may decrease with short nature cotton plants, the dwarf population can increase the planting density and the cotton yield per unit area [1].

Moreover, GBS population results in short heighted Ari1327 reinforce no correlation between PH with FL, FU, FS, FE, and MC. Very few dwarf cotton mutants are reported but correlated with fiber or other agronomic traits [11]. Ari1327 is a possible dwarfing germplasm because of the lack of a relationship between dwarfing and fibre quality.

Forward genetics are commonly used to identify the underlying molecular mechanism of a mutant, and QTL-mapping was used to separate populations [35]. However, the grouping of phenotypic data in various environments is a pre-requisite to constructing QTL map. Using RIL (Recombinant Inbred Line), NIL (Near Isogenic Line) and back-cross techniques are great QTL positioning methods, but often take many years to produce multiple generations [31, 36]. Next-generation sequencing (NGS) technology combined with BSA has been proven to be a cost-effective and straightforward rapid QTL detection method. Zhang et al. [37] only used  $F_1$ ,  $F_2$  and  $BC_1F_1$



populations to successfully identify maize QTL *qPH7* for plant height. With the advancement in sequencing technology and high-density markers like SNP and tiny-InDel were replacing traditional markers such as RFLP, STS, and SSR in Map-based cloning [38, 39]. In this case, we used F<sub>2</sub> populations separately to co-locate semi-dwarf genes. Within one year, two populations were produced from the same dwarf parent Ari1327 and different tall parent. High-density genetic linkage map based on SNP marker in another F<sub>2</sub> population BL-Y10 × Ari1327 was constructed by using enzyme digestion SLAF-seq (Pacific-Locus Amplified Fragment Sequencing). 11 loci

are located on four chromosomes, QTL physical position was compared in order to cross-validate BSA-QTL results. However, there is no significant GBS-QTL signal in the BSA-QTL area. The segregation filter process lost a large number of markers at the front end and back end of the chromosome which caused the deviation of QTL position. We re-picked all these SDM and constructed the single D01 chromosome linkage map with higher density. The results of new D01 linkage map provide the better insight and good overlap.

SD is a phenomenon of significant deviations between the observed segregation rate and the expected



Mendelian ratio [40, 41], due to gametic selection and interspecific or intraspecific hybridization [42, 43]. SD was related to the proportion of heterozygous alleles in the population or the frequency of heterozygous chromosomes, which affects sex differentiation and gamete selection. It is considered to be an important driving force for biological evolution [44, 45]. SDM was usually deleted in map construction because it will affect the genetic distance and QTL for accuracy [46, 47]. Deleting it may cause potential QTL to be missed, or a change in QTL location [48, 49]. Hackett et al. believed that when the marker density is less than 10 cM, the segregation marker had no significant effect on linkage groups [50]. Zuo et al. used the orthogonal and reciprocal RIL populations of soybean to study the effects of SDM on linkage groups and QTL localization in multiple environments. It is believed that SDM can be added to improve the density and quality of linkage groups [51]. After reconsidering the SDM, repacked all the SDM and had a good QTL overlap with another population. Our results follow the above-mentioned key point that the removal of SDM will reduce the density of linkage groups, even change the position of QTL.

RNA-seq was used to narrow down the selection criteria from 200 candidate genes to 21 genes. Finally, focused on six genes i.e., *Gh\_D01G0503*, *Gh\_D01G0514*, *Gh\_D01G0585*, *Gh\_D01G0586*, *Gh\_D01G0503* and *Gh\_D01G0592* due to their different expression patterns in three time periods, and the FPKM values are greater than five. Except for *Gh\_D01G0586*, the expression trends of the other four genes are consistent with the RNA-seq in the six leaves stage. Among them, *Gh\_D01G0585* and *Gh\_D01G0586* were annotated as homologous genes of the *Arabidopsis* ubiquitin10. Ubiquitin was involved in regulating many physiological processes of plants including growth and development [52]. However the ubiquitination of proteins mainly depended on E3 ubiquitin ligase to regulate specific signaling pathways [53]. *Gh\_D01G0585* and *Gh\_D01G0586* genes encode poly-ubiquitin, and their expression changes should influence the many phenotypes not only PH. The annotation of *Gh\_D01G0503* was stem-specific protein. Our laboratory's previous research on Ari1327 physiology found that internodal shortening was the cause of its dwarfness [34], so stem apex was selected as RNA-seq sampling material. The selection of this gene may be related to our sampling choice, but the possibility of being the target gene was not excluded. *Gh\_D01G0531* was annotated as *CAB13*, Chlorophyll a/b-binding protein, or the *LHCB3* protein in the PSII system [54]. In the barley, correlation was elucidated with gene diversity of LHCP protein and the agronomic traits [55]. But in general, there were few related studies about the effect of LHCP protein on dwarfing plant.

*Gh\_D01G0514* was one of NAC gene family, which the biggest TF families regulate both development and stress responses [56]. One of AtNACs *JUB1* participates in the negative regulation of GA/BR, to short hypocotyls, dwarfism, late flowering and male sterility [57]. Knockdown of *OsY37* expression caused dwarfism and high accumulation of chlorophyll during the vegetative phase [58]. Wu et al. found *ATAF1*-overexpression lines displayed many altered phenotypes, including dwarfism and short primary roots [59]. In our case, the expression between Ari1327 and Ari971 showed a difference only at the six leaves stage, but it still can be the reason of dwarfing.

Thiamin, also known as vitamin B1, is an essential vitamin for metabolic processes such as the citric acid cycle, glycolysis and pentose phosphate pathway. Thiamine consists of pyrimidine ring and thiazole ring. *THI1* was the first enzymes of thiazole ring synthesis [60]. Thiamine and its active form TPP is an essential cofactor for enzymes involved in a number of critical metabolic processes, such as the production of acetyl-CoA, the tricarboxylic acid cycle (TCA), the pentose phosphate pathway (PPP) and C3 cycle etc. [61]. In *Arabidopsis thaliana*, overexpression of *THI1* appears to increase plant drought resistance and regulate ion channels [62]. In Maize, *THI2* and thiamine played a vital role in proliferation of stem cells [63]. At the same time, thiamine is also reported in DNA repair, metabolism, photosynthesis, and respiration [64, 65]. *Gh\_D01G0592* is a kind of *THI1*, in Ari1327, and it exhibited more expression level than wild type at six leaves stage. The accumulation of excessive thiamine could cause the disturbance of the above-mentioned metabolic cycle, resulting in chlorosis, growth retardation [66, 67]. Thus, it was prioritized as a dwarfing gene in Ari1327.

## Conclusion

Overwhelmingly, in present study two F2 populations were used to locate the major QTL of the semi-dwarf mutant Ari1327 individually, and successfully co-localizing it to a smaller range on the D01 chromosome. InDel marker phenotyping 1241 population can also confirm the existence of the QTLs. Through RNA-seq and gene annotation, candidate genes were reduced to three from 21 selected genes, i.e., *Gh\_D01G0503*, *Gh\_D01G0531* and *Gh\_D01G0592*. Among them, *Gh\_D01G0592* was the key gene, regulating *THI1* in the process of thiamine synthesis, and we believed that it is most likely to be the dwarf gene. Therefore, it is suggested that this information must be substantiated by another genetic experiment another genetic experiment be substantiate this information to verify the gene action. This work will pave

the way for understanding the dwarfing mechanism of Ari1327 and finally using its excellent dwarfing genes.

## Methods

### Plant material, mapping population construction and phenotype analysis

The semi-dwarf mutant Ari1327 was produced by  $^{60}\text{Co}$   $\gamma$ -ray irradiation treatment of Ari971 [68]. The two tall parental lines Ari3697 and BL-Y10 were crossed as female parents with mutant Ari1327 to generate two  $F_2$  mapping populations, AA (Ari3697  $\times$  Ari1327) and BA (BL-Y10  $\times$  Ari1327) in 2018. Ari3697 was the tall height mutant from Ari971. In comparison, BL-Y10 was selected just only by plant height character and not developed from Ari971. Two tall parents have close and distant genetic relationships, respectively. Ari3697 and BL-Y10 were selected as tall parents crossing with Ari1327 in order to observe in generations easily. The AA group of 215 individuals were planted in the experimental area, CRICAAS (Cotton Research Institute, Chinese Academy of Agricultural Sciences), Anyang (Henan Province, China 36.06°N, 114.49°E). While the BA group of 303 individuals were planted in Sanya (Hainan Province, China, 18.41°N, 109.20°E). In addition, we used the same  $F_1$  generation of BA group to generate a huge  $F_2$  validation population of 1241 individual plants in CRICAAS, 2020. All recommended production practices and plant protection measures were adopted during the experiment. Row to row 70 cm and plant to plant distance of 30 cm was maintained. The first dose of fertilizer (DAP, urea) was applied at the germination stage. The second dose of fertilizer (Urea) was applied at the flowering stage. The last fertilizer (Urea) dose was applied at the boll formation stage. Two weeding practices were performed at the cotton field, the first one was performed one month after sowing and the second weeding practice was done after one month of first weeding. The tinning was done with first weeding practices. When the vertical growth ceased, the plant's final height was measured with a measuring rod from the first cotyledonary node to the apical bud. The sympodial branches or fruit branches are the direct fruit-bearing branches. At maturity, the sympodial branches from each genotype were counted from all selected plants. Also, measure the length of the fruit branches. The number of effective mature bolls from all the picks was counted, and the cumulative record was maintained for each plant separately. Lint percentage or ginning outturn is the weight of lint obtained from a given weight of seed cotton and expressed in percentage. Samples of seed cotton obtained from individual plants were weighed and ginned separately with an electrical ginning machine in the cotton laboratory. Lint was

weighed, and values were used for the determination of GOT percentage.

$$\text{GOT\%} = \text{Weight of lint in a sample} / \text{Weight of seed cotton in a sample} \times 100$$

Seed cotton weight is the 30 bolls weight. 30 cotton bolls were picked and seeds were taken at random for each sample from each genotype and weighed. Similarly, lint of picked 30 bolls was also weighed separately for lint weight. The fiber characters were measured using high volume instrument 900 (HVI-900, Uster technologies Ltd., Switzerland). It is a computerized instrument that provides a comprehensive profile of raw fiber. Agronomic and fiber traits between Ari1327 and wild species were recorded by a core collection of upland cotton project. This two-year experiment was planned in six environments with three blocks and 16 agronomic traits (Table 1) [69]. Ari1327 and wild species were also grown in the greenhouse for RNA-seq. Genomic DNA was extracted using CTAB (Cetyltrimethylammonium bromide) method [70]. The stem tips of wild-type and mutants were used to extract RNA with RNAprep pure Plant Kit DP441 (Tiangen Biotech, Beijing, China). Statistical analysis for plant height was conducted using SPSS (Version 23.0, SPSS, Chicago). In this project, all the seeds Ari1327, Ari971, BL-Y10 and Ari3697 were provided by Germplasm Repository of Institute of Cotton Research, Chinese Academy of Agricultural Sciences (CRI of CAAS, Anyang, Henan province, China) only for scientific research purposes.

### QTL mapping by BSA

The two DNA pools were developed by mixing an equal amount of DNA from the 20-maximum height (Higher than 87.05 cm) and 20 dwarfs (lower than 60 cm) lines of AA  $F_2$  populations which were screened by BSA (bulked segregant analysis). Both DNA pools were randomly broken into fragments at the size of around 350 bp by the method of ultrasonication and sequenced by Illumina HiSeq 2500 (Biomarker Technology, Qingdao, China). The experimental process was carried out according to the standard protocol provided by Illumina [71]. The raw reads were trimmed and filtered, then aligned to the TM-1 (*G. hirsutum*) [72] reference genome using the BWA software [73]. Aligned files were converted to BAM format and detected SNPs using SAMtools [74] and GATK software [75]. ANNOVAR software [76] was used to annotate SNPs. After identifying polymorphic SNP markers between parents, the two lines have already been re-sequenced in the core collection of the upland cotton project [69]. SNP-index and  $\Delta$  (SNP-index) were calculated by sliding window method with default settings of 1 Mb and step size of 10 kb. Only the area where

$\Delta$  (SNP-index) above the threshold level of 0.01 significance were considered candidate loci [77, 78].

### QTL mapping by GBS

The BA group was screened using GBS (Genotyping-by-sequencing) to construct an ultra-high-density genetic map. MseI and NlaIII were used to double digest the genomic DNA of  $F_2$  populations and their parents. Furthermore, the libraries were constructed and sequenced on Illumina HiSeq PE150 platform according to the manufacturer's standard protocols. The method mentioned in section 2.2 was used to reads filter, alignment and SNPs detection. The SNPs sites where no homozygous genotype is detected in any of the two parents, firstly filtered out, and then further screened those sites on the basis of mark integrity (75%) and SD (segregation distortion) ( $p < 0.001$ ). Joinmap4.0 [79] was used to construct a linkage map (Maximum likelihood) and WinQTLCart2.5 [80] was employed to map QTL (CIM mapping method, window size set at 10 cM, LOD > 2.5, 1000 permutation, 1 cM walking speed) [81, 82]. The genotype was coded as two from the female parent. So that when the additive effect is positive, female (BL-Y10) in the parental allele provides the effect of increasing plant height, and male (Ari1327) provides the effect of reducing plant height [83]. In addition, using the marker for un-screened by SD to construct the single linkage map of D01 chromosome.

### QTL validation

InDel markers in the candidate area were used for QTL validation. The principle of InDel selection was first homozygous and different between two parents, second the sequence difference was between 5 to 100 bp and finally the sequencing depth must be greater than 10. Once the sites were determined, the 250 bp of its upstream and downstream flanking sequence was used to design primer on NCBI Primer-BLAST tools (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>). PCR reactions contained 2  $\mu$ l of genomic DNA template (50 ng/ $\mu$ l), 5  $\mu$ l of 3G Taq Master Mix for PAGE (Vazyme, Nanjing, China), 0.25  $\mu$ l primer (100  $\mu$ M/ $\mu$ l)  $\times$  2 and 2.5  $\mu$ l water. PCR steps are followed as: 95  $^{\circ}$ C for 5 min; 35 cycles of 95  $^{\circ}$ C for 15 s, 50–65  $^{\circ}$ C for 15 s, 72  $^{\circ}$ C for 30 s; and a final extension at 72  $^{\circ}$ C for 5 min. The PCR amplification products were separated on 8% polyacrylamide gel by electrophoresis of 200 V for 2.0 h. The gel was stained in 0.1% AgNO<sub>3</sub> solution (Sinopharm Chemical Reagent, Shanghai, China), and revealed the DNA bands in 1.5% sodium hydroxide and 0.4% formaldehyde solutions (Sinopharm Chemical Reagent, Shanghai, China) [84]. According to the genotype consistent with the parents or F<sub>1</sub>, the plant height data of 1241 populations were divided into three

groups. ANOVA-test and Chi-square were performed at ( $p < 0.01$ ).

### RNA-seq and qRT-PCR validation between wild and mutant

The critical period in which the mutant showed a high degree of difference was measured and set a sampling point before and after it. RNA extraction was carried out using an RNAPrep Pure Plant Kit (TIANGEN Biotech, Beijing, China). Then samples were sent to the company (Biomarker Technology, Qingdao, China) for RNA-sequencing and data analysis was performed in Biocloud online platform (Biomarker Technology, Qingdao, China). HISAT2 [85] and StringTie [86] aligned and assembled the fileted reads to TM-1 reference genome. The FPKM (fragments per kilobase of transcript per million fragments mapped) was calculated to find the area of candidate DEGs (differential expressed genes). The condition for differential genes was  $p$ -value < 0.01, FC (fold change)  $\geq$  |1.5|. Enrichment analysis was conducted on CottonFGD website (<https://cottonfgd.org/>), the significance level was 0.0001, and the minimum gene number for each analyzed term was five. The extracted RNA was also used for qRT-PCR, 1  $\mu$ g of total RNA was used to be transcribed into cDNA by using One-Step RT-PCR Kit (Novoprotein Scientific, Beijing, China). The qPCR was performed by using ChamQ Universal SYBR qPCR Master Mix (Vazyme Biotech co., Ltd, Nanjing, China). LightCycler 480 by Roche Diagnostics GmbH, Mannheim, Germany was used to perform PCR amplification. Detailed Primer information used in the current study is given in Table S3.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-021-03359-x>.

- Additional file 1.
- Additional file 2.
- Additional file 3.
- Additional file 4.
- Additional file 5.
- Additional file 6.
- Additional file 7.
- Additional file 8.
- Additional file 9.
- Additional file 10.
- Additional file 11.
- Additional file 12.
- Additional file 13.

### Contributions

XMD and CHM conceived the project and formulated the scientific objectives; CHM set up the experimental design, performed the data analysis and wrote the draft manuscript; and AR, HGL, GFS and ZBZ revised the paper. All authors discussed the results and read and approved the final manuscript for publication.

### Funding

This work was financially supported by the National Natural Science Foundation of China [grant number No. 31671746]; and the National Key R&D Program of China [grant number 2017YFD0101600]. The funders had no role in the experimental design, data analysis, decision to publish, or preparation of the manuscript.

### Availability of data and materials

All raw data was submitted to NCBI (<https://www.ncbi.nlm.nih.gov/sra>). The accession number of GBS data is PRJNA734430. The original data for parent BL-Y10 of BSA is part of project PRJNA634606, and the internal number is GH0543. The raw data of parent Ari1327 and extreme pools of BSA is PRJNA734972. The accession number of RNA-seq between Ari1327 and Ari971 is PRJNA733290.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>State Key Laboratory of cotton Biology, Institute of Cotton Research of the Chinese Academy of Agricultural Sciences, Anyang 455000, China. <sup>2</sup>Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou 450000, China. <sup>3</sup>Department of Plant Breeding and Genetics, Bahauddin Zakariya University, Multan 66000, Pakistan. <sup>4</sup>State Key Laboratory of Cotton Biology, Research Base, Anyang Institute of Technology, Anyang, China.

Received: 15 April 2021 Accepted: 24 November 2021

Published online: 03 January 2022

### References

- Zhang JF, Fang H, Zhou HP, Sanogo S, Ma ZY. Genetics, Breeding, and Marker-Assisted Selection for Verticillium Wilt Resistance in Cotton. *Crop Sci.* 2014;54(4):1289–303. <https://doi.org/10.2135/cropsci2013.08.0550>.
- Ren XM, Zhang LZ, Du MW, Evers JB, van der Werf W, Tian XL, et al. Managing mepiquat chloride and plant density for optimal yield and quality of cotton. *Field Crop Res.* 2013;149:1–10. <https://doi.org/10.1016/j.fcr.2013.04.014>.
- Liu DX, Teng ZH, Kong J, Liu XY, Wang WW, Zhang X, et al. Natural variation in a CENTRORADIALIS homolog contributed to cluster fruiting and early maturity in cotton. *Bmc Plant Biol.* 2018;18. <https://doi.org/10.1186/s12870-018-1518-8>.
- Gu S, Evers JB, Zhang L, Mao L, Zhang S, Zhao X, et al. Modelling the structural response of cotton plants to mepiquat chloride and population density. *Ann Bot.* 2014;114(4):877–87. <https://doi.org/10.1093/aob/mct309>.
- Li S, Tian Y, Wu K, Ye Y, Yu J, Zhang J, et al. Modulating plant growth-metabolism coordination for sustainable agriculture. *Nature.* 2018;560(7720):595–600. <https://doi.org/10.1038/s41586-018-0415-5>.
- Quisenberry JE. Inheritance of Plant Height in Cotton. I. A Cross Between Lubbock Dwarf and Texas Marker-1. *Crop Sci.* 1975;15(2):197–9. <https://doi.org/10.2135/cropsci1975.0011183x001500020016x>.
- Abzalov MF, Zhumaev FK. Revelation in the cotton plant of chromosomes responsible for the regulation of plant growth. *J Chem Nat Compd.* 1995;31(1):111–3. <https://doi.org/10.1007/bf01167585>.
- Hutchinson JB, Ghose RLM. On the occurrence of “Crinkled Dwarf” in *Gossypium hirsutum* L. *J Genet Genomics.* 1937;34(3):437. <https://doi.org/10.1007/BF02982305>.
- Harland SC. The genetics of cotton. *J Genet.* 1932;25(3):261–70. <https://doi.org/10.1007/BF02984590>.
- Jing C, Ma XJ, Di JC, Chen XS. Gene mapping of an ultra-dwarf mutant in upland cotton. *Hereditas (Beijing).* 2011;33(12):1393–7. <https://doi.org/10.3724/sp.j.1005.2011.01393>.
- Sun YW, Liang WH, Shen WJ, Feng H, Chen JD, Si ZF, et al. G65V Substitution in Actin Disturbs Polymerization Leading to Inhibited Cell Elongation in Cotton. *Front Plant Sci.* 2019;10:1486. <https://doi.org/10.3389/fpls.2019.01486>.
- Zhang C, Sun JL, Jia YH, Zhou ZL, Pan ZE, He SP, Xu ZJ, Du XM. Traits Analysis and Molecular Mapping of a Cotton Supper-dwarf Mutant AS98. *Acta Agric Nucleatae Sin.* 2014;28(2):186–92. <https://doi.org/10.11869/j.issn.100-8551.2014.02.0186>.
- Ji G, Liang C, Cai Y, Pan Z, Meng Z, Li Y, Jia Y, Miao Y, Pei X, Gong W et al. A copy number variant at the HPDA-D12 locus confers compact plant architecture in cotton. 2020. <https://doi.org/10.1111/nph.17059>
- Wang B, Smith SM, Li JY. Genetic Regulation of Shoot Architecture. *Annu Rev Plant Biol.* 2018;69:437–68. <https://doi.org/10.1146/annurev-arpla-042817-040422>.
- Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, et al. Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature.* 2002;416(6882):701–2. <https://doi.org/10.1038/416701a>.
- Huang D, Wang S, Zhang B, Shang-Guan K, Shi Y, Zhang D, et al. A Gibberellin-Mediated DELLA-NAC Signaling Cascade Regulates Cellulose Synthesis in Rice. *Plant cell.* 2015;27(6):1681–96. <https://doi.org/10.1105/tpc.15.00015>.
- Daviere JM, Wild M, Regnault T, Baumberger N, Eisler H, Genschik P, Achard P. Class I TCP-DELLA interactions in inflorescence shoot apex determine plant height. *Curr Biol.* 2014;24(16):1923–8. <https://doi.org/10.1016/j.cub.2014.07.012>
- Tu XJ, Wang Q, Li J, Liu AY. Quantitative proteomic analysis of upland cotton stem terminal buds reveal phytohormone-related pathways associated with dwarfism. *Biol Plant.* 2016;61:1. <https://doi.org/10.1007/s10535-016-0644-0>
- Guo HQ, Li L, Aluru M, Aluru S, Yin YH. Mechanisms and networks for brassinosteroid regulated gene expression. *Curr Opin Plant Biol.* 2013;16(5):545–53. <https://doi.org/10.1016/j.pbi.2013.08.002>.
- Wang ZY, Bai MY, Oh E, Zhu JY. Brassinosteroid Signaling Network and Regulation of Photomorphogenesis. *Annu Rev Genet.* 2012;46:701–24. <https://doi.org/10.1146/annurev-genet-102.209-163.450>.
- Du XM, Liu SY, Sun JL, Zhang GY, Jia YH, Pan Z, et al. Dissection of complex genetic architecture and breeding perspective of cottonseed traits by genome-wide association study. *Bmc Genomics.* 2018;19. <https://doi.org/10.1186/s12864-018-4837-0>.
- Chen YN, Fan XR, Song WJ, Zhang YL, Xu GH. Over-expression of OsPIN2 leads to increased tiller numbers, angle and shorter plant height through suppression of OsLAZY1. *Plant Biotechnol J.* 2012;10(2):139–49. <https://doi.org/10.1111/j.1467-7652.2011.00637.x>.
- Sayed SA, Gadallah MAA. Effects of shoot and root application of thiamin on salt-stressed sunflower plants. *Plant Growth Regul.* 2002;36(1):71–80. <https://doi.org/10.1023/A:1014784831387>.
- Tunc-Ozdemir M, Miller G, Song LH, Kim J, Sodek A, Koussevitzky S, et al. Thiamin Confers Enhanced Tolerance to Oxidative Stress in Arabidopsis. *Plant Physiol.* 2009;151(1):421–32. <https://doi.org/10.1104/pp.109.140046>.
- Rosado-Souza L, Proost S, Moulin M, Bergmann S, Bocobza SE, Aharoni A, et al. Appropriate Thiamin Pyrophosphate Levels Are Required for Acclimation to Changes in Photoperiod. *Plant Physiol.* 2019;180(1):185–97. <https://doi.org/10.1104/pp.18.01346>.
- Khozaei M, Fisk S, Lawson T, Gibon Y, Sulpice R, Stitt M, et al. Overexpression of plastid transketolase in tobacco results in a thiamine auxotrophic phenotype (vol 27, pg 432, 2015). *Plant cell.* 2016;28(7):1752–4. <https://doi.org/10.1105/tpc.16.00527>.
- Hsieh WY, Liao JC, Wang HT, Hung TH, Tseng CC, Chung TY, et al. The Arabidopsis thiamin-deficient mutant pale green1 lacks thiamin monophosphate phosphatase of the vitamin B1 biosynthesis pathway. *Plant J.* 2017;91(1):145–57. <https://doi.org/10.1111/tpj.13552>.

28. Zhang JC, Li B, Yang YP, Mu PY, Qian WQ, Dong LL, et al. A novel allele of L-galactono-1,4-lactone dehydrogenase is associated with enhanced drought tolerance through affecting stomatal aperture in common wheat. *Sci Rep-Uk*. 2016;6. <https://doi.org/10.1038/srep30177>.
29. Yuan YC, Xing HX, Zeng WG, Xu JL, Mao LL, Wang LY, et al. Genome-wide association and differential expression analysis of salt tolerance in *Gossypium hirsutum* L at the germination stage. *BMC Plant Biol*. 2019;19(1):394. <https://doi.org/10.1186/s12870-019-1989-2>.
30. Zhang C, Li LB, Liu QB, Gu LJ, Huang JQ, Wei HL, et al. Identification of Loci and Candidate Genes Responsible for Fiber Length in Upland Cotton (*Gossypium hirsutum* L.) via Association Mapping and Linkage Analyses. *Front Plant Sci*. 2019;10:53–3. <https://doi.org/10.3389/fpls.2019.00053>.
31. Reiter RS, Williams JG, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA. Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc Natl Acad Sci U S A*. 1992;89(4):1477–81. <https://doi.org/10.1073/pnas.89.4.1477>.
32. Craig W, Bednarz, W D, Shurley, W S, Robert L, Nichols.: Yield, quality, and profitability of cotton produced at varying plant densities. *Agron J*. 2005;97(1):235–40. <https://doi.org/10.2134/agronj2005.0235a>.
33. Dai JL, Li WJ, Tang W, Zhang DM, Li ZH, Lu HQ, et al. Manipulation of dry matter accumulation and partitioning with plant density in relation to yield stability of cotton under intensive management. *Field Crop Res*. 2015;180:207–15. <https://doi.org/10.1016/j.fcr.2015.06.008>.
34. An WY, Gong WF, He SP, Pan ZE, Sun JL, Du XM. MicroRNA and mRNA expression profiling analysis revealed the regulation of plant height in *Gossypium hirsutum*. *Bmc Genomics*. 2015;16. <https://doi.org/10.1186/s12864-015-2071-6>.
35. Bazakos C, Hanemian M, Trontin C, Jiménez-Gómez JM, Loudet O. New Strategies and Tools in Quantitative Genetics: How to Go from the Phenotype to the Genotype. *Annua Rev Plant Biol*. 2017;68:435–55. <https://doi.org/10.1146/annurev-arplant-042916-040820>.
36. Rowan BA, Patel V, Weigel D, Schneeberger K. Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3 (Bethesda, Md)* 2015;5(3):385–98. <https://doi.org/10.1534/g3.114.016501>.
37. Zhang H, Wang X, Pan Q, Li P, Liu Y, Lu X, et al. QTG-Seq Accelerates QTL Fine Mapping through QTL Partitioning and Whole-Genome Sequencing of Bulk Segregant Samples. *Mol Plant*. 2019;12(3):426–37. <https://doi.org/10.1016/j.molp.2018.12.018>.
38. Ganai MW, Altmann T, Röder MS. SNP identification in crop plants. *Curr Opin Plant Biol*. 2009;12(2):211–7. <https://doi.org/10.1016/j.pbi.2008.12.009>.
39. Almasly L, Blangero J. Human QTL linkage mapping. *Genetica*. 2009;136(2):333–40. <https://doi.org/10.1007/s10709-008-9305-3>.
40. Wang C, Zhu C, Zhai H, Wan J. Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). *Genet Res*. 2005;86(2):97–106. <https://doi.org/10.1017/s0016672305007799>.
41. Gardner KA, Wittern LM, Mackay IJ. A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnol J*. 2016;14(6):1406–17. <https://doi.org/10.1111/pbi.12504>.
42. Faris JD, Haen KM, Gill BS. Saturation mapping of a gene-rich recombination hot spot region in wheat. *Genetics*. 2000;154(2):823–35. <https://doi.org/10.1017/S0016672399004061>.
43. Silver LM: The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends Genet*. 1993;9(7):250–4. [https://doi.org/10.1016/0168-9525\(93\)90090-5](https://doi.org/10.1016/0168-9525(93)90090-5).
44. Taylor DR, Ingvarsson PK. Common features of segregation distortion in plants and animals. *Genetica*. 2003;117(1):27–35. <https://doi.org/10.1023/A:1022308414864>.
45. Xu Y, Zhu L, Xiao J, Huang N, McCouch SR. Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). *Mol Gen Genet*. 1997;253(5):535–45. <https://doi.org/10.1007/s004380050355>.
46. Zhang L, Wang S, Li H, Deng Q, Zheng A, Li S, et al. Effects of missing marker and segregation distortion on QTL mapping in F2 populations. *Theor Appl Genet*. 2010;121(6):1071–82. <https://doi.org/10.1007/s00122-010-1372-z>.
47. Zhao J, Han D, Shi K, Wang L, Gao J, Yang R. Influence of epistatic segregation distortion loci on genetic marker linkages in Japanese flounder. *Genomics*. 2018;110(1):59–66. <https://doi.org/10.1016/j.jygeno.2017.08.006>.
48. Xu S. Quantitative trait locus mapping can benefit from segregation distortion. *Genetics*. 2008;180(4):2201–8. <https://doi.org/10.1534/genetics.108.090688>.
49. Shah R, Cavanagh CR, Huang BE. Computationally efficient map construction in the presence of segregation distortion. *Theor Appl Genet*. 2014;127(12):2585–97. <https://doi.org/10.1007/s00122-014-2401-0>.
50. Hackett CA, Broadfoot LB. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity*. 2003;90(1):33–8. <https://doi.org/10.1038/sj.hdy.6800173>.
51. Zuo J-F, Niu Y, Cheng P, Feng J-Y, Han S-F, Zhang Y-H, et al. Effect of marker segregation distortion on high density linkage map construction and QTL mapping in Soybean (*Glycine max* L.). *Heredity*. 2019;123(5):579–92. <https://doi.org/10.1038/s41437-019-0238-7>.
52. Hellmann H, Estelle M: Plant development: regulation by protein degradation. *Science*. 2002;297(5582):793–7. <https://doi.org/10.1126/science.1072831>.
53. Yamao F. Ubiquitin system: selectivity and timing of protein destruction. *J Biochem*. 1999;125(2):223–9. <https://doi.org/10.1093/oxfordjournals.jbchem.a022277>.
54. Crepin A, Caffarri S. Functions and Evolution of Lhcb Isoforms Composing LHCII, the Major Light Harvesting Complex of Photosystem II of Green Eukaryotic Organisms. *Curr Protein Pept Sci*. 2018;19(7):699–713. <https://doi.org/10.2174/1389203719666180222101534>.
55. Xia Y, Ning Z, Bai G, Li R, Yan G, Siddique KH, et al. Allelic variations of a light harvesting chlorophyll a/b-binding protein gene (*Lhcb1*) associated with agronomic traits in barley. *Plos ONE*. 2012;7(5):e37573. <https://doi.org/10.1371/journal.pone.0037573>.
56. Mathew IE, Agarwal P. May the Fittest Protein Evolve: Favoring the Plant-Specific Origin and Expansion of NAC Transcription Factors. *BioEssays*. 2018;40(8):e1800018. <https://doi.org/10.1002/bies.201800018>.
57. Shahnejat-Bushehri S, Tarkowska D, Sakuraba Y, Balazadeh S. Arabidopsis NAC transcription factor JUB1 regulates GA/BR metabolism and signaling. *Nat Plants*. 2016;2:16013. <https://doi.org/10.1038/nplants.2016.13>.
58. El Mannai Y, Akabane K, Hiratsu K, Satoh-Nagasawa N, Wabiko H: The NAC Transcription Factor Gene *OsY37* (*ONAC011*) Promotes Leaf Senescence and Accelerates Heading Time in Rice. *Int J Mol Sci*. 2017;18(10). <https://doi.org/10.3390/ijms18102165>.
59. Wu Y, Deng Z, Lai J, Zhang Y, Yang C, Yin B, et al. Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. *Cell Res*. 2009;19(11):1279–90. <https://doi.org/10.1038/cr.2009.108>.
60. Goyer A. Thiamine in plants: Aspects of its metabolism and functions. *Phytochemistry*. 2010;71(14–15):1615–24. <https://doi.org/10.1016/j.phytochem.2010.06.022>.
61. Hohmann S, Meacock PA. Thiamin metabolism and thiamin diphosphate-dependent enzymes in the yeast *Saccharomyces cerevisiae*: genetic regulation. *Biochim Biophys Acta*. 1998;1385(2):201–19. [https://doi.org/10.1016/s0167-4838\(98\)00069-7](https://doi.org/10.1016/s0167-4838(98)00069-7).
62. Li CL, Wang M, Wu XM, Chen DH, Lv HJ, Shen JL, et al. *THI1*, a Thiamine Thiazole Synthase, Interacts with Ca<sup>2+</sup>—Dependent Protein Kinase CPK33 and Modulates the S-Type Anion Channels and Stomatal Closure in Arabidopsis. *Plant Physiol*. 2016;170(2):1090–104. <https://doi.org/10.1104/pp.15.01649>.
63. Woodward JB, Abeydeera ND, Paul D, Phillips K, Rapala-Kozik M, Freeling M, et al. A maize thiamine auxotroph is defective in shoot meristem maintenance. *Plant Cell*. 2010;22(10):3305–17. <https://doi.org/10.1105/tpc.110.07776>.
64. Machado CR, de Oliveira RL, Boiteux S, Praekelt UM, Meacock PA, Menck CF. *Thi1*, a thiamine biosynthetic gene in Arabidopsis thaliana, complements bacterial defects in DNA repair. *Plant Mol Biol*. 1996;31(3):585–93. <https://doi.org/10.1007/bf00042231>.
65. Rosado-Souza L, Fernie AR, Aarabi F: Ascorbate and Thiamin: Metabolic Modulators in Plant Acclimation Responses. *Plants*. 2020;9(1). <https://doi.org/10.3390/plants9010101>.
66. Bocabza SE, Malitsky S, Araújo WL, Nunes-Nesi A, Meir S, Shapira M, et al. Orchestration of thiamin biosynthesis and central metabolism by combined action of the thiamin pyrophosphate riboswitch and the circadian

- clock in Arabidopsis. *Plant Cell*. 2013;25(1):288–307. <https://doi.org/10.1105/tpc.112.106385>.
67. Wang L, Ye, Liu H, Liu X, Wei C, Huang Y, Liu Y, Tu J. Both overexpression and suppression of an *Oryza sativa* NB-LRR-like gene OsLSR result in auto-activation of immune response and thiamine accumulation. *Sci Rep-Uk*. 2016;6(1):24079. <https://doi.org/10.1038/srep24079>.
  68. Wu B, Hu W, Xing YZ. The history and prospect of rice genetic breeding in China. *Yi chuan = Hereditas* 2018;40(10):841–57. <https://doi.org/10.16288/j.ycz.18-213>.
  69. Ma ZY, He SP, Wang XF, Sun JL, Zhang Y, Zhang GY, Wu LQ, Li ZK, Liu ZH, Sun GF et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat Genet*. 2018;50(6):803–13. <https://doi.org/10.1038/s41588-018-0119-7>.
  70. Paterson AH, Brubaker CL, Wendel JF. A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol Biol Rep*. 1993;11(2):122–7. <https://doi.org/10.1007/BF02670470>.
  71. Watson M. Quality assessment and control of high-throughput sequencing data. *Front Genet*. 2014;5. <https://doi.org/10.3389/fgene.2014.00235>.
  72. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allo-tetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015;33(5):531–7. <https://doi.org/10.1038/nbt.3207>.
  73. Yang ZE, Ge XY, Yang ZR, Qin WQ, Sun GF, Wang Z, et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat Commun*. 2019;10. <https://doi.org/10.1038/s41467-019-10,820-x>.
  74. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>
  75. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/g.107524.110>.
  76. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16). <https://doi.org/10.1093/nar/gkq603>
  77. Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H, Matsumura H et al. MutMap plus: Genetic Mapping Plos ONE. 2013;8(7). <https://doi.org/10.1371/journal.pone.0068529>
  78. Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, et al. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol*. 2012;30(2):174–8. <https://doi.org/10.1038/nbt.2095>.
  79. Stam P. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J*. 1993;3(5):739–44. <https://doi.org/10.1111/j.1365-313x.1993.00739.x>.
  80. S. W, Basten CJ, Zeng ZB. Windows QTL Cartographer 2.5. In. Raleigh, NC.: Department of Statistics, North Carolina State University; 2012. <https://brcwebportal.cos.ncsu.edu/qtlcart/WQTLCart.htm>.
  81. Ma JJ, Pei WF, Ma QF, Geng YH, Liu GY, Liu J, et al. QTL analysis and candidate gene identification for plant height in cotton based on an interspecific backcross inbred line population of *Gossypium hirsutum* x *Gossypium barbadense*. *Theor Appl Genet*. 2019;132(9):2663–76. <https://doi.org/10.1007/s00122-019-03380-7>.
  82. Diouf L, Magwanga RO, Gong WF, He SP, Pan ZE, Jia YH, Kirungu JN, Du XM: QTL Mapping of Fiber Quality and Yield-Related Traits in an Intra-Specific Upland Cotton Using Genotype by Sequencing (GBS). *Int J Mol Sci*. 2018;19(2). <https://doi.org/10.3390/ijms19020441>
  83. Li HH, Ribaut JM, Li ZL, Wang JK. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor Appl Genet*. 2008;116(2):243–60. <https://doi.org/10.1007/s00122-007-0663-5>.
  84. Bassam BJ, Caetano-Anollés G, Gresshoff PM. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem*. 1991;196(1):80–3. [https://doi.org/10.1016/0003-2697\(91\)90120-i](https://doi.org/10.1016/0003-2697(91)90120-i).
  85. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60. <https://doi.org/10.1038/nmeth.3317>.
  86. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5. <https://doi.org/10.1038/nbt.3122>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

