

RESEARCH

Open Access



# Identification of QTLs and candidate genes for physiological traits associated with drought tolerance in cotton

MAGWANGA Richard Odongo<sup>1,2†</sup>, LU Pu<sup>1†</sup>, KIRUNGU Joy Nyangasi<sup>1†</sup>, CAI Xiaoyan<sup>1</sup>, ZHOU Zhongli<sup>1</sup>, AGONG Stephen Gaya<sup>2</sup>, WANG Kunbo<sup>1\*</sup> and LIU Fang<sup>1\*</sup> 

## Abstract

**Background:** Cotton is mainly grown for its natural fiber and edible oil. The fiber obtained from cotton is the indispensable raw material for the textile industries. The ever changing climatic condition, threatens cotton production due to a lack of sufficient water for its cultivation. Effects of drought stress are estimated to affect more than 50% of the cotton growing regions. To elucidate the drought tolerance phenomenon in cotton, a backcross population was developed from *G. tomentosum*, a drought tolerant donor parent and *G. hirsutum* which is highly susceptible to drought stress.

**Results:** A genetic map of 10 888 SNP markers was developed from 200 BC<sub>2</sub>F<sub>2</sub> populations. The map spanned 4 191.3 centi-Morgan (cM), with an average distance of 0.104 7 cM, covering 51% and 49% of At and Dt sub genomes, respectively. Thirty stable Quantitative trait loci (QTLs) were detected, in which more than a half were detected in the At subgenome. Eighty-nine candidate genes were mined within the QTL regions for three traits: cell membrane stability (CMS), saturated leaf weight (SLW) and chlorophyll content. The genes had varied physiochemical properties. A majority of the genes were interrupted by introns, and only 15 genes were intronless, accounting for 17% of the mined genes. The genes were found to be involved molecular function (MF), cellular component (CC) and biological process (BP), which are the main gene ontological (GO) functions. A number of miRNAs were detected, such as miR164, which is associated with *NAC* and *MYB* genes, with a profound role in enhancing drought tolerance in plants. Through RT-qPCR analysis, 5 genes were found to be the key genes involved in enhancing drought tolerance in cotton. Wild cotton harbors a number of favorable alleles, which can be exploited to aid in improving the narrow genetic base of the elite cotton cultivars. The detection of 30 stable QTLs and 89 candidate genes found to be contributed by the donor parent, *G. tomentosum*, showed the significant genes harbored by the wild progenitors which can be exploited in developing more robust cotton genotypes with diverse tolerance levels to various environmental stresses.

**Conclusion:** This was the first study involving genome wide association mapping for drought tolerance traits in semi wild cotton genotypes. It offers an opportunity for future exploration of these genes in developing highly tolerant cotton cultivars to boost cotton production.

**Keywords:** Cotton spp., Quantitative trait loci, Genetic map, Drought tolerance, miRNAs, Gene ontology

\* Correspondence: [wkbcri@163.com](mailto:wkbcri@163.com); [liufcri@163.com](mailto:liufcri@163.com)

<sup>†</sup>Magwanga RO, Lu P and Kirungu JN contributed equally to this work.

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



## Background

Upland cotton (*Gossypium hirsutum* L.) is the main global crop for natural fiber production, a key raw material for textile industries and producer of edible oil for more than half the world's population (Chakravarthy et al. 2012). Being a field crop, it is particularly susceptible to water stress, especially at the seedling stage (Argyrokastritis et al. 2015). It is estimated that more than 50% of the world's cotton producing regions are affected more or less by a number of abiotic stress factors such as drought, salinity, and extreme temperature variations (Dabbert and Gore 2014). Even though cotton is partially tolerant to drought stress, upland cotton, which is the main cotton genotype grown for its high fiber quality, demands a sufficient amount of fresh water during growth, which makes its production encounter a lot of challenges when drought suddenly occurs (Chapagain et al. 2006). Breeding for new drought tolerant cotton cultivars will not only save a great amount of water but also help to increase and stabilize cotton yields during periods of uncertain rainfall, and will also offer a reprieve in the face of the ever deteriorating global weather dynamics (Blum 2005). In the recent past, many plant breeders have explored conventional type of breeding, which to some extent has yielded little improvement. However, adoption of molecular and genetic engineering techniques, will accelerate the ultimate goal of producing plants that are more versatile and highly tolerant to various environmental stresses (Ashraf 2010). Genetic improvement for drought adaptation, addressed through a conventional approach by trait selection for yield and its stability over locations and years, have yielded some limited progress (Ashraf 2010). Such selection programs are slow because of the low heritability of yield under stress, the inherent variation in the field and time restraints (Nguyen et al. 1997). Alternatively, yield improvements in water-limited environments could be achieved by identifying secondary traits contributing to drought tolerance and be utilized in breeding programs (Liu et al. 2010). Breeding technique through marker-assisted selection (MAS) provides faster and a more accurate approach in the selection for the desired phenotypes in a breeding population (Tester and Langridge 2010). The use of advanced genetic approaches to detect and analyze the genetic variations linked to phenotypic traits have greatly enhanced the improvement of agronomic traits, in which most are quantitative (Swinnen et al. 2012). The recently developed molecular markers techniques, such as genomic selection (GS) and MAS, have made the mapping of quantitative trait loci (QTL) a reality. QTL identification is done through the linkage mapping method, where polymorphisms between two parents are detected in either a segregating or in a real/stable population, which is either developed through

interspecific or intraspecific methods and are associated with phenotypic traits (Deschamps et al. 2012).

Many genotyping methods through molecular markers have been developed, including sequence characterized amplified regions (SCARs) (Paran and Michelmore 1993), restriction fragment length polymorphisms (RFLPs) (Bernatzky and Tanksley 1986), simple sequence repeats (SSRs) (Litt and Luty 1989), amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995), random amplification of polymorphic DNAs (RAPDs) (Williams et al. 1990), cleaved amplified polymorphic sequences (CAPS) (Konieczny and Ausubel 1993), inter simple sequence repeats (ISSRs) (Salimath et al. 1995), and direct amplification of length polymorphisms (DALP) (Desmarais et al. 1998). However, these methods are too expensive, labor-intensive and time-consuming to be widely used and accessible for many studies. Therefore, the genotyping by sequence (GBS) method offers a realistic alternative. GBS is applicable to GS, which predicts complex, economically important quantitative traits using genome-wide molecular markers at a lower cost than what is achieved through other methods (Poland et al. 2012). The introduction of GBS has revolutionized the entire field owing to its specificity, simplicity, high reproducibility and increased speed owing to the simultaneous detection of single nucleotide polymorphisms (SNPs) and genotyping (Furuta et al. 2017). Thus, the significance of GBS is the reduced sequencing steps, reduced cost, reduced sample handling, fewer polymerase chain reactions (PCR) and purification steps. Other advantages are that it has no size fractionation, no reference sequence limits, efficient barcoding and a system that is easy to scale up (Davey et al. 2011).

Drought tolerance is a complex trait, which is controlled by multiple small effect QTLs, and improving water use efficiency always involves trade-offs with growth (Barnabás et al. 2008). QTL mapping has become an important tool for quantitative trait research and has been widely used to map a number of traits including drought tolerance traits in various crops (Azhar and McNeilly 1988). A number of QTLs associated with drought tolerance traits have been identified in plants such as barley (Fan et al. 2015), *Oryza sativa* (Mardani et al. 2013), *Zea mays* (Lu et al. 2010), and wheat (Fleury et al. 2010). QTL mapping for drought tolerance traits has been reported in which SSR markers have been used to develop the genetic map (Zheng et al. 2016). However, the use of genotypic data derived from GBS has not been reported yet in a segregating backcross population (BC<sub>2</sub>F<sub>2</sub>) derived from interspecific backcross between *Gossypium hirsutum* and *G. tomentosum*, though the same techniques have been used to explore salt tolerance in intraspecific an F<sub>2:3</sub> population in upland cotton (Qi et al. 2017).

The wild cotton species harbors significant traits, which are vital to improving the performance of elite cotton cultivars (Magwanga et al. 2018a). *G. tomentosum* is a wild tetraploid cotton endemic to Hawaiian island, which is dry and saline in nature. *G. tomentosum* is thus highly tolerant to salt and drought stress conditions (Zheng et al. 2016). The two parental lines have been explored widely in developing a mapping population, especially the  $F_{2:3}$  population, which has been used in QTL mapping for salt tolerance traits (Oluoch et al. 2016) and drought tolerance traits (Zheng et al. 2016). The use of the  $F_{2:3}$  population does not allow for the saturation of the donor alleles, thus the adoption of the backcross technique provides the opportunity to increase the donor parents' contribution to the mapping population (Swamy et al. 2018). The backcross method has been widely used in evaluating the performance of a number of plants. For example, backcross populations were used in mapping QTLs for the grain mineral elements, iron and zinc, in rice (Swamy et al. 2018) and studies on resistance to *Verticillium* wilt in cotton (Zhang et al. 2015a). Based on the wider research carried out on the backcross technique, especially on the backcross inbred lines, we applied the  $BC_2F_2$  generation in mapping for QTLs for drought stress tolerance traits, further carried out with in silico analysis and RT-qPCR validation of the candidate gene identified within the QTL regions.

## Materials and methods

### Development of plant materials

The segregating backcross population ( $BC_2F_2$ ) was developed using *G. hirsutum* CCRI-12 (G09091801-2), as the recurrent parent, and *G. tomentosum*-AD3-00 (P0601211) as the donor parent. *G. hirsutum* accession number CCRI-12 is an elite upland cotton, which was developed by Institute of Cotton Research, Chinese Academy of Agricultural Sciences, China, thus the code CCRI. The donor parent, *G. tomentosum* accession number AD3-00 (P0601211) was developed and maintained by the same institute, in their wild cotton germplasm nursery located in Sanya, Hainan Province, China. *G. hirsutum* is an upland cotton, mainly grown for its high fiber yielding ability, though it is negatively affected by drought stress (Chen et al. 2013). *G. tomentosum* is closely related to *G. hirsutum*, but of wild origin (Pleasant and Wendel 2010). A single line of the recurrent parent, *G. hirsutum*, was crossed with the donor parent, *G. tomentosum*, to obtain the  $F_1$  lines. The tagged flower of the recurrent plant was pollinated, then covered to prevent entry of any foreign pollen grains. Upon maturity, the boll was harvested and the seeds replanted. Thirty plants were considered for backcrossing with the donor male parent. In each line, 20 bolls were harvested to obtain seeds of  $BC_1F_1$  seeds. The  $BC_1F_1$  plants were evaluated and only 30 lines were finally chosen, which were then

backcrossed with the recurrent parent to obtain  $BC_2F_1$ . In each line, 30 bolls were again collected, each boll representing one line. The lines were then evaluated and a single line was chosen for selfing to obtain the  $BC_2F_1$  lines. Over 400 lines were eventually developed (Additional file 1: Figure S1). For this research, only 200  $BC_2F_2$  populations were selected for severe drought stress study in two environments. The selection was based on the seed amount and the heterogeneity of the  $BC_2F_2$  lines, determined through gel electrophoresis. The selection of the backcross population used for the study was purely based on seed counts. The development of the  $BC_2F_2$  lines was done in Sanya, within the 18°09' and 18°37' latitudes. Hainan Province has a tropical monsoon climate, making it hot and rainy. The annual mean temperature reaches 22–27 °C and the annual precipitation is between 1 500 and 2 600 mm.

### Drought stress treatments

In the simulated drought condition, drought susceptible *G. hirsutum* seeds, drought tolerant *G. tomentosum* seeds, and their segregating backcross inbred lines, the  $BC_2F_2$  seeds, were grown in planting boxes (45 cm length, 35 cm width, 25 cm depth), filled with peat moss growth media. In each line, three replications were maintained under optimal growth conditions. Plants were irrigated with tap water twice a week. The greenhouse conditions were set with the temperature at  $(23 \pm 1)^\circ\text{C}$  and a 14-h light/10-h dark photoperiod. At the emergence of the third true leaves, watering was totally withdrawn from the drought treated seedlings but not from plants under the control condition. Before the treatment, the soil water potential was maintained at -20 kPa because soil is well watered when the soil water potential is above -30 kPa (Parent et al. 2010). The soil water potential was monitored on a daily basis in both treatments by using of Em50, DECAGON soil moisture machine. On the 14th day post treatment, measurements for various physiological and morphological traits were carried out. The research was conducted from February to April 2017 (Environment 1) and from July to September 2017 (Environment 2). The adopted experimental design was a completely randomized block design (CRBD) set up in the greenhouse in Institute of Cotton Research (ICR), Anyang, Henan Province, China.

### Determination of the morphological traits for drought treated and non- drought treated plants

The growth performance of the  $BC_2F_2$  population and their parental lines were assessed for drought tolerance in terms of plant height (PH), fresh leaf weight (FLW), relative leaf water content (RLWC), total fresh biomass (TFB), excised leaf weight (ELW), dry root biomass (DRB), chlorophyll content (SPAD  $\text{mg}\cdot\text{g}^{-1}$  FW), saturated leaf weight (SLW), fresh shoot biomass-fresh root

biomass ratio (FSB/FRB), dry leaf weight (DLW), total dry biomass (TDB), cell membrane stability (CMS), fresh root biomass (FRB), dry shoot biomass (DSB), excised leaf water loss (ELWL) and dry shoot biomass-dry root biomass ratio (DSB/DRB).

#### Determination of physio-biochemical traits for drought treated and non- drought treated plants

##### Cell membrane stability (CMS)

Leaf discs weighing 0.5 g were taken from each genotype. Leaf samples were then washed with distilled water, then with deionized water, before being placed in sterilized test tubes. In each test tube, 9 mL of deionized water was added, then left for 24 h at room temperature. After 24 h, the test tubes were shaken before measuring the electrical conductivity (EC) of the water using a conductivity meter. Upon taking the measurements ( $T_1$ ), the leaves were then autoclaved at 70 °C for 20 min. The samples were then cooled to room temperature before final EC values were taken ( $T_2$ ).

The CMS was calculated using the following formula as described by Fokar (Fokar et al. 1998):

$$\text{Cell Membrane Stability (CMS)} = \left( \left( 1 - \frac{T_1}{T_2} \right) / \left( 1 - \frac{C_1}{C_2} \right) \right) \times 100$$

where, T is treatment and C is control, and 1 and 2 are the initial and final conductance measurements, respectively.

##### Relative leaf water content (RLWC)

Fresh leaves were obtained from each lines in three replicates, weighed to get the fresh weight (FW), then immediately put in distilled water for 24 h at room temperature. The leaves were then removed and quickly dried of any surface moisture with absorbent filter paper. Upon removal of surface moisture, the leaf samples were weighed to obtain fully saturated weights (SW). Samples were then oven dried at 80 °C for 24 h and weighed to determine their dry weights (DW) (Barrs and Weatherley 1962):

$$\text{Relative leaf water content (RLWC)} = \left( \frac{FW - DW}{SW - DW} \right) \times 100$$

where, FW is sample fresh weight, SW is sample saturated weight and DW is sample dry weight.

##### Excised leaf water loss (ELWL)

One leaf sample was taken from each plant. The samples were immediately weighed for their fresh weight (FW) using an electronic scale. The leaf samples were then left on a laboratory bench at room temperature overnight. After 24 hours the weights of the wilted leaf samples were recorded. The leaf samples were then oven dried at 80 °C to obtain their dry weights (DW). Excised leaf water loss

was calculated by the formula as described by Clarke and McCaig (1982):

$$\text{Excised leaf water loss (ELWL)} = \left( \frac{FW - WW}{DW} \right)$$

where FW is fresh weight, WW is wilted weight and DW is dry weight.

##### Microscopic examination of the number and stomatal pore size of the parental lines and their BC<sub>2</sub>F<sub>1</sub> generation under drought stress condition

Drought tolerance has been associated with either a reduced stomatal number or narrow stomatal pore (Haworth et al. 2016). To determine the relationship between drought tolerance and stomatal density, the parental lines, the drought susceptible *G. hirsutum*, the drought tolerant *G. tomentosum* and the BC<sub>2</sub>F<sub>1</sub> second backcross generation were used. Plants were grown in a growth chamber with day and night temperatures of approximately 28 °C and 25 °C, respectively, and relative humidity between 60% to 70%. Seeds were germinated in a peat: perlite mixture at a 2:1 ratio. After 3 days, the seedlings were transferred to hydroponic system with nutrients supplied through a Hoagland nutrient solution (Hoagland and Arnon 1950). Two weeks later, at the third leaf stage, drought treatment was started by adding a 20% (mass fraction) of polyethylene glycol-6 000 (PEG) concentration. A high concentration of PEG is suitable for imposing drought stress for a short period (Li et al. 2015b). In control plants no PEG was added. The leaf samples were then harvested for stomatal examination at 0, 1, 6, 12 h and 24 h, from each genotype. The numbers of stomata per view were scored, and stomatal lengths and widths were measured under a 40× objective lens of a photomicroscope fitted with objective and eyepiece micrometers (Olympus Corporation, Tokyo, Japan). For each leaf sample, 4 microscopic observations were made and averages of the four readings were used. Stomatal averages of 4 view areas ( $S = \pi r^2$ ,  $r$  = view radius) were calculated, and stomatal density was defined as  $N/S$  (number of stomata per square millimeter). Six stomata per view were randomly selected for measuring their lengths and widths which were then averaged as the value for each genotype. The relationships between the stomata density in control plants and the reduction in drought vs. control plants were investigated. The following formula was used for the calculations:

$$\text{Reduction in drought plants vs. control plants} = (\text{control} - \text{drought}) / \text{control} \times 100.$$

### DNA extraction, GBS library preparation, sequencing and SNP genotyping

Young tender leaves were obtained from the two parental lines and from each of the 200 individuals of the BC<sub>2</sub>F<sub>2</sub> population for simplicity; they are referred to as the segregating backcross population (BC<sub>2</sub>F<sub>2</sub>). The leaf samples were immediately frozen in liquid nitrogen upon collection and then stored at -80 °C until DNA extraction. DNA from the BC<sub>2</sub>F<sub>2</sub> populations of the 200 plants and 10 samples each for the parents were extracted using the CTAB method as described by Zhang et al. (2000). Then, DNA was diluted in 20 µL TE buffer (10 mmol·L<sup>-1</sup> Tris, pH 8, 1 mmol·L<sup>-1</sup> EDTA) (Krizman et al. 2006). The purity of DNA was determined using a Nano Photometer® spectrophotometer (IMPLEN, CA, USA). The ratio of absorbance at 260 nm and 280 nm was used to assess the purity of the DNA. The DNA samples with a ratio of ~1.8 were considered pure (Wilfinger et al. 1997). The DNA concentrations were determined with a Qubit fluorimeter (Thermo Fisher Scientific), and confirmed by gel electrophoresis on a 1% agarose gel. At least 100 ng·µL<sup>-1</sup> genomic DNA was used to prepare the libraries for each genotype. Library construction for GBS was conducted according to a previous report by Elshire et al. (2011). Briefly, genomic DNA from the female parent and each of 200 progenies were digested for 15 min at 37 °C in a 50-µL reaction with 20 units (U) of Taqa I (NEB, USA) and Mse I (NEB, USA). P1 adapter, a modified Illumina adapter, was ligated to the samples. After adapter ligation, sample were pooled and randomly sheared with a Bioruptor (Diagenode, Belgium) to an average size of 500 bp (base pair). DNA fragments of 300–500 bp were purified using the MinElute Gel Extraction Kit (Qiagen). The dsDNA ends were repaired using the Quick Blunting kit Enzyme Mix (NEB). Then, a modified Solexa P2 adapter was ligated to the obtained DNA fragments. Finally, purified and quantified DNA products were PCR-amplified using Phusion Master Mix (NEB, USA). PCR amplification was performed with the following cycle profile: 98 °C for 2 min, followed by 13 cycles at 98 °C for 30 s, 60 °C for 30 s and 72 °C for 15 s, and a final extension at 72 °C for 5 min. The prepared DNA libraries were sequenced using the Illumina HiSeq system at Shanghai Major Biological Medicine Technology Co., Ltd. (Illumina 2014). The high-quality FASTQ read sequences generated for each genotype were aligned to the reference *G. hirsutum* cotton genome using the Burrows–Wheeler aligner with the default parameters (Li and Durbin 2010). We applied SAM tools (Li et al. 2009) to produce BAM files for removing unmapped reads based on the mapping outputs. Variant call format (VCF) file version 4.1 v (Danecek et al. 2011) was then used to filter SNPs with mapping-quality scores of <30. The obtained high-quality SNPs

were reformatted and transferred to JoinMap 4.1 for linkage group determination. Because the population under this study was tetraploid cotton, 26 linkage groups were obtained.

### Data analysis, linkage map construction, QTL mapping and identification of candidate genes within the QTL regions

Analysis of variance (ANOVA) suitable for the specified experimental design was conducted with SAS to assess the genetic dissimilarity among the given BC<sub>2</sub>F<sub>2</sub> cotton genotypes at *P* = 0.05 (Henley 1983). The genetic advance (GA) at 5% selection intensity was calculated as described by Singh (Kalra 1998).

$$\text{Genetic advance (GA)} = \left( \frac{\text{Genotypic variance}}{\text{Phenotypic variance}} \right) \times 2.06.$$

Broad sense heritability (*H*<sup>2</sup>) was calculated using the formula described by Khan et al. (2010).

$$H^2 = \left( \frac{\text{Genotypic variance}}{\text{Phenotypic variance}} \right) \times 100.$$

In addition to genetic advance (GA), analysis of variance (ANOVA) and broad sense heritability (*H*<sup>2</sup>), we further estimated the phenotypic coefficient of variation (PCV), genotypic coefficient of variation (GCV) and coefficient of variability (CV). These were calculated using the formulas as outlined below. These were to determine the effects of the environment on the various measured traits.

$$\text{Phenotypic coefficient of variation (PCV)} = \left( \frac{\sqrt{\sigma_p^2}}{\bar{X}} \right) \times 100$$

$$\text{Genotypic coefficient of variation (GCV)} = \left( \frac{\sqrt{\sigma_g^2}}{\bar{X}} \right) \times 100$$

$$\text{Coefficient of variability (CV)} = \left( \frac{\text{Error mean square}}{\text{Population mean}} \right) \times 100$$

where:  $\sigma_p^2$  is phenotypic variance,  $\sigma_g^2$  genotypic variance and  $\bar{X}$  is the general mean of the character.

### Linkage map construction and QTL mapping

Markers were ordered based on their logarithm of odds (LOD) scores, pairwise recombination fractions and linkage group length (Reeder et al. 2016). Linkage analysis was conducted using JoinMap 4.1 (Van Ooijen and Voorrips 2001) with a recombination frequency set at 0.40 with an LOD score of 2.5 for the BC<sub>2</sub>F<sub>2</sub> population. An LOD of 2.0 and above have been adopted in evaluating various QTLs in a number of crops, such as ridgetail white prawn *Exopalaemon carinicauda* (Li et al. 2019). Moreover, QTLs with an LOD of at least 2.5 are considered common QTLs (Ma et al. 2017). The parameters

were to some degree a more stringent threshold than the value used for the relatively smaller genomes, and appropriate for cotton because cotton genome is estimated to be 4 500 centi-Morgan (cM) (Zhang et al. 2015a, 2015b). The Kosambi mapping function was used to convert the recombination frequencies to map distances (Kosambi 1943). Linkages at distances greater than 35 Kosambi cM were considered to be non-significant. Each data point represented the mean of three replications.

The physiological and morphological traits used to conduct QTL analysis were plant height (PH), leaf fresh weight (LFW), saturated leaf weight (SLW), excised leaf water loss (ELWL), leaf dry weight (LDW), shoot fresh weight (SFW), root fresh weight (RFW), shoot dry weight (SDW), root dry weight (RDW), cell membrane stability (CMS), chlorophyll content as determined by SPAD values (SPAD), ratio of shoot fresh weight and root fresh weight (SFW/RFW) and finally the ratio of shoot dry weight and root dry weight (SDW/RDW). QTLs were detected using composite interval mapping (CIM) (da Silva et al. 2016) by WinQTL Cartographer 2.5 (Wang et al. 2011).

In the CIM method, model 6, the forward–backward regression method with a 1 cM walking speed, a probability into and out of the model of 0.01 and a window size set at 10 cM were used. A stringent logarithm of odds (Civelek and Lysis 2014) threshold value was estimated by a 1 000 permutation test for all the traits and was used to declare the significant QTLs, with a significance level of 0.05. However, QTLs in two or more environments with an LOD threshold of at least 2.5 were considered common QTLs based on the description given by Lander and Kruglyak (1995). QTL nomenclature was carried out as per Liang et al. (Zhang et al. 2009). The observed phenotypic variance in each QTL was estimated by the coefficient of determination  $R^2$  (%) as a percentage. Modes of gene action for individual QTLs were calculated and categorized into various subsets depending on the values of additive (A) (0–0.20), partial dominant (PD) (0.21–0.80), dominant (D) (0.81–1.20) and over dominant (OD) > 1.20, as described by Paterson et al. (Stuber et al. 1987). The graphical presentation of the 23 marked linkage group and QTLs were done by Map Chart 2.2.

#### **Candidate gene identification, functional annotation, phylogenetic relation, gene structure and RNA Seq analysis**

The flanking marker regions were used to identify the various genes linked to QTLs for cell membrane stability (CMS), saturated leaf weight (SLW) and chlorophyll content as determined by the SPAD values using *G. hirsutum* as the reference genome. The marker positions were used as the query in the cotton functional genome

database (<https://cottonfgd.org>). Multiple sequence alignments of the deduced amino acid sequences of the key genes were performed using the default parameters of ClustalW, and a dendrogram was constructed using the neighbor joining (NJ) method and bootstrap analysis with 1 000 replications in the MEGA 7 program. We further undertook to confirm the subcellular localization prediction of these genes using online tool WoLF-PSORT ([http://www.genscript.com/psort/wolf\\_psort.html](http://www.genscript.com/psort/wolf_psort.html)). The results were then validated by re-analyzing the data through two online tools, using the TargetP1.1 (<http://www.cbs.dtu.dk/services/TargetP/>) server and Protein Prowler Subcellular Localisation Predictor version 1.2 ([http://bioinf.scmb.uq.edu.au/pprowler\\_webapp\\_1-2/](http://bioinf.scmb.uq.edu.au/pprowler_webapp_1-2/)). We carried out functional annotation and the expression levels of these key genes using Blast2GO pro-software version 4.1.1 (<https://www.blast2go.com>). Blast2GO annotation associates genes or transcripts with GO terms using hierarchical terms.

The mined genes were further analyzed by extracting their RNA sequences from the cotton genome database (<http://mascotton.njau.edu.cn>) in reference to salt and drought stress expression profiles at varying time intervals. The reads per kilobase of exon per million reads mapped (FPKM) data were then transformed into log<sub>10</sub> and a heatmap was constructed, the top 15 highly expressed key genes were later used for RT-qPCR validation under a drought stress condition. Finally we analyzed the gene structure to determine whether the mined genes were interrupted by introns or all were intronless, using the gene structure display server (<http://gsds.cbi.pku.edu.cn/>).

#### **miRNA target and Cis-regulatory elements analysis of the identified candidate genes within the three major QTL regions**

In order to determine if the mined genes were targeted by any known miRNAs, we predicted which miRNA could target the mined genes. The miRNA sequences were downloaded from miRBase (<http://www.mirbase.org>) and the plant miRNA database (<http://bioinformatics.cau.edu.cn/PMRD/>). The genes targeted by miRNAs were predicted by searching the 5' and 3' untranslated regions (UTRs) and the coding sequences (CDS) of all the mined genes for complementary sequences of cotton miRNAs using the psRNATarget server with default parameters (<http://plantgrn.noble.org/psRNATarget/function=3>). In addition, we carried out *cis* element analysis. The promoter sequences (2 kb upstream of the translation start site) of all the mined genes were obtained from the cotton genome project (<http://cgp.genomics.org.cn/page/species/index.jsp>). Transcriptional response elements of the mined gene promoters were predicted

using the online PLACE database (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>).

#### **RT-qPCR validation of the key functional genes identified within the QTL regions regulating SLW, CMS and chlorophyll content traits**

The samples for RNA extraction were collected on the 0, 7<sup>th</sup> and 14<sup>th</sup> day of drought stress treatment for plants under drought treatment and the controls. When soil is used as opposed to a hydroponic set up for carrying out drought stress tolerance screening in plants, longer stress exposure is always suitable for obtaining samples for conducting gene expression analysis (Magwanga et al. 2018b). Root, stem and leaf were the main organs used in this study to carry out RT-qPCR validation of the highly expressed mined genes as per the RNA sequence data. The RNA extraction kit, EASYspin plus plant RNA kit, by Aid Lab, China ([www.aidlab.cn](http://www.aidlab.cn)), was employed in the extraction of RNA from the samples. The concentration and quality of each extracted RNA sample were determined using a NanoDrop 2000 spectrophotometer and gel electrophoresis. The RNA samples that met the criteria of having a 260/280 ratio of 1.8–2.1, or 260/230 ratio  $\geq 2.0$ , were used for further analyses. The tetraploid cotton constitutive *Actin7* gene (forward 3'ATCCTCCGTCTTGACCTTG5' and reverse sequence 3'TGTCGGTCAGGCAACTCAT5') was used as a reference gene, and the 15 specific gene primers, were used for RT-qPCR validation. The first strand cDNA synthesis was done with TranScriptAll-in-One First-Strand cDNA Synthesis SuperMix for RT-qPCR, from TRAN company per the manufacturer's instructions. Primer Premier 5 (<http://www.premierbiosoft.com/primerdesign/>) was used to design the 15 gene specific primers with melting temperatures of 55–60 °C, primer lengths of 18–25 bp, and amplicon lengths of 101–221 bp. Details of the primers are shown in Additional file 4: Table S1. Fast Start Universal SYBR green Master (Rox) (Roche, Mannheim, Germany) was used to carry out RT-qPCR analysis in accordance with the manufacturer's instructions. The RT-qPCR reactions samples were prepared in a total volume of 20  $\mu$ L, containing 10  $\mu$ L of SYBR green master mix, 2  $\mu$ L of cDNA template, 6  $\mu$ L of ddH<sub>2</sub>O, and 2  $\mu$ L of each primer.

## **Results**

### **Phenotypic variation between parental lines, *G. hirsutum* and *G. tomentosum* with the BC<sub>2</sub>F<sub>2</sub> generation**

Significant differences were observed among the parental lines and the BC<sub>2</sub>F<sub>2</sub> generation ( $P < 0.0001$ ) for both physiological and morphological traits. In the BC<sub>2</sub>F<sub>2</sub> population, all the traits measured showed normal frequency distribution (Additional file 2: Figure S2) revealing quantitative inheritance, thus the traits were suitable for QTL analysis (Fang et al. 2014). The two parental

lines are diverse in phenotypic attributes. *G. tomentosum*, the donor parent, has small leaves, long roots and a shiny leaf surface, common features among the xerophytic plants (Li and Bao 2015) while *G. hirsutum*, the recurrent parent, has broad leaves, medium growth and relatively possessing characteristic of a mesophytic plant (Zhang et al. 2014). In addition to the two parental lines having diverse in morphological features, when subjected to drought treatment, *G. tomentosum* showed superior performance to *G. hirsutum*. This could be explained by the inherent genetic features of the two cultivars. *G. tomentosum* has superior traits towards drought tolerance while *G. hirsutum* is prone to drought stress, thus having less tolerance. Among the BC<sub>2</sub>F<sub>2</sub> population, there was a wide range of phenotypic variation in all the traits measured across the two environments: cell membrane stability (CMS), plant height (PH), chlorophyll content/ level (CHL), leaf fresh weight (LFW), excised leaf water loss (ELWL), saturated leaf weight (SLW), root fresh weight (RFW), shoot dry weight (SDW), root dry weight (RDW), and their ratios. All traits exhibited a typical segregation pattern, with normal distribution. Under a controlled environment, no water stress was imposed; therefore, no significant differences were noted except for plant height (PH) and chlorophyll content (CHL), however, the differences observed were not statistically significant. A contrast was observed under the drought treatment condition; all the traits had a significant reduction compared with the drought tolerant parent (Table 1 and Additional file 5: Table S2).

### **Microscopic examination of the parental lines and BC<sub>2</sub>F<sub>1</sub> generation**

*Gossypium hirsutum* (Gh) and *Gossypium tomentosum* (Gt) are closely related. Both are tetraploid cotton, but *G. tomentosum* is wild while *G. hirsutum* is domesticated. After the emergence of the A and D genome, polyploidization was then followed by radiation and divergence, with the evolution of distinct tetraploid species. *G. hirsutum* L. is now indigenous to Central America while *G. tomentosum* Nuttall ex Seemann is endemic to the Hawaiian islands (Fryxell 1982). The two parental lines have 26 gametic chromosomes, exhibit disomic pairing (Kimber 1961), and have similar genome sizes which have been estimated to range between 2.2 and 2.9 Gb (Wendel et al. 2002). Owing to high number of individuals in the BC<sub>2</sub>F<sub>2</sub> population used in this research, the stomatal pore examination was limited to the two parental lines and their BC<sub>2</sub>F<sub>1</sub> generation. In each lines, 10 individual plants were used, among which sequencing was done on 10 individuals of the parental lines. A higher stomatal density was observed on the upper leaf surface of *G. hirsutum*, followed by the BC<sub>2</sub>F<sub>1</sub> and the lowest number was detected in *G. tomentosum*.

**Table 1** Analysis of variance of the BC<sub>2</sub>F<sub>2</sub> population performance under drought the stress condition

Source	df	PH			FLW			RLWC			TFB			CH/SPAD			SLW		
		MS	F	Pr > F	MS	F	Pr > F	MS	F	Pr > F	MS	F	Pr > F	MS	F	Pr > F	MS	F	Pr > F
P	2	131.00	38.20	<0.0001	0.00	0.33	0.72	0.04	5.08	0.01	0.09	0.62	0.54	1.41	0.69	0.50	0.01	3.72	0.02
e	1	10 100.00	2 940.00	<0.0001	247.00	165 000.00	<0.0001	4.33	604.00	<0.0001	31 600.00	227 000.00	<0.0001	3 350.00	1 650.00	<0.0001	572.00	385 000.00	<0.0001
t	1	78.10	22.70	<0.0001	0.02	12.90	0.00	6.11	853.00	<0.0001	3.47	24.90	<0.0001	8 470.00	4 170.00	<0.0001	30.10	20 200.00	<0.0001
g	199	14.90	4.32	<0.0001	0.05	34.90	<0.0001	0.05	7.27	<0.0001	1.47	10.50	<0.0001	29.20	14.40	<0.0001	0.07	44.00	<0.0001
e*t	1	155.00	45.00	<0.0001	0.04	26.80	<0.0001	13.50	1 880.00	<0.0001	0.93	6.67	0.01	378.00	186.00	<0.0001	38.70	26 100.00	<0.0001
e*g	199	8.36	2.43	<0.0001	0.04	26.70	<0.0001	0.05	7.39	<0.0001	0.92	6.57	<0.0001	23.40	11.50	<0.0001	0.04	29.80	<0.0001
t*g	199	12.90	3.74	<0.0001	0.05	30.10	<0.0001	0.05	7.30	<0.0001	1.86	13.30	<0.0001	21.00	10.30	<0.0001	0.04	29.60	<0.0001
g*t*e	199	7.40	2.15	<0.0001	0.04	28.00	<0.0001	0.05	7.22	<0.0001	0.95	6.81	<0.0001	20.60	10.10	<0.0001	0.05	36.20	<0.0001
Source	df	FSB/FRB			CMS			DLW			FRB			DSB			ELW		
P	2	0.35	0.35	0.70	0.01	3.48	0.03	0.00	17.00	<0.0001	0.02	0.63	0.53	0.01	1.94	0.14	0.01	9.23	0.00
e	1	24 800.00	24 600.00	<0.0001	37.30	9 510.00	<0.0001	1.42	34 100.00	<0.0001	2 080.00	82 800.00	<0.0001	1 150.00	224 000.00	<0.0001	92.60	118 000.00	<0.0001
t	1	0.14	0.14	0.71	0.00	0.09	0.77	0.02	492.00	<0.0001	0.72	28.80	<0.0001	0.57	112.00	<0.0001	0.01	7.38	0.01
g	199	6.77	6.72	<0.0001	0.01	2.21	<0.0001	0.00	9.19	<0.0001	0.04	1.52	<0.0001	0.29	57.30	<0.0001	0.02	31.30	<0.0001
e*t	1	9.92	9.85	0.00	0.02	3.82	0.05	0.01	138.00	<0.0001	1.58	62.80	<0.0001	2.98	584.00	<0.0001	0.01	6.38	<0.0001
e*g	199	6.26	6.72	<0.0001	0.01	2.54	<0.0001	0.00	7.29	<0.0001	0.04	1.41	0.00	0.25	49.70	<0.0001	0.02	31.00	<0.0001
t*g	199	7.38	7.33	<0.0001	0.01	2.41	<0.0001	0.00	9.66	<0.0001	0.04	1.60	<0.0001	0.32	62.10	<0.0001	0.03	35.50	<0.0001
g*t*e	199	6.66	6.62	<0.0001	0.01	2.48	<0.0001	0.00	8.39	<0.0001	0.04	1.43	0.00	0.29	57.10	<0.0001	0.03	33.40	<0.0001
Source	df	ELWL			DSB/DRB			TDB			DRB			FSB					
P	2	0.42	1.42	0.24	0.09	0.07	0.94	0.01	1.19	0.31	0.00	0.51	0.60	0.04	0.31	0.74			
e	1	5 060.00	17 000.00	<0.0001	51 700.00	37 200.00	<0.0001	1 770.00	289 000.00	<0.0001	67.10	79 800.00	<0.0001	17 500.00	152 000.00	<0.0001			
t	1	0.38	1.28	0.26	376.00	271.00	<0.0001	1.36	223.00	<0.0001	0.17	203.00	<0.0001	1.03	8.94	0.00			
g	199	1.75	5.87	<0.0001	18.80	13.50	<0.0001	0.31	50.40	<0.0001	0.01	9.02	<0.0001	1.38	12.00	<0.0001			
e*t	1	0.13	0.45	<0.0001	62.50	45.00	<0.0001	4.68	764.00	<0.0001	0.19	226.00	<0.0001	0.08	0.74	0.39			
e*g	199	1.73	5.80	<0.0001	17.30	12.50	<0.0001	0.27	44.50	<0.0001	0.01	9.24	<0.0001	0.83	7.25	<0.0001			
t*g	199	1.97	6.60	<0.0001	17.50	12.60	<0.0001	0.33	53.10	<0.0001	0.01	7.49	<0.0001	1.71	14.90	<0.0001			
g*t*e	199	1.94	6.52	<0.0001	20.50	14.70	<0.0001	0.30	48.90	<0.0001	0.01	7.96	<0.0001	0.86	7.52	<0.0001			

P<sup>2</sup>: replication/block; e: environment; t: treatment; g: genotype; e\*t: environment-treatment interaction; e\*g: environment-genotype interaction; g\*t\*e: genotype-treatment-environment interaction; PH: Plant height (cm), FLW: Fresh leaf weight(g), RLWC: Relative leaf water content (%), TFB: Total fresh biomass (g), ELW: Excised leaf weight (g), DRB: Dry root biomass (g), CHI: Chlorophyll content (mg/g FW), SLW: Saturated leaf weight (g), FSB/FRB: Fresh shoot biomass-fresh root biomass ratio, DLW: Dry leaf weight (g), TDB: Total dry biomass (g), CMS: Cell membrane stability (%), FRB: Fresh root biomass (g), DSB: Dry shoot biomass (g), ELWL: Excised leaf water loss (%) and DSB/DRB: Dry shoot biomass-dry root biomass ratio

The proportion of the stomatal density on the lower leaf surface among the three cotton germplasms analysed were 31, 11 and 28 in *G. hirsutum*, *G. tomentosum* and the BC<sub>2</sub>F<sub>1</sub> generation, respectively (Fig. 1 I-III). The stomatal pore was relatively larger in *G. hirsutum* than in BC<sub>2</sub>F<sub>1</sub> and *G. tomentosum* (the donor parent). The reduction in stomatal number and pore size could be an adaptive mechanism employed by *G. tomentosum* to survive under limited water supply. Increased stomata and with larger pore, is disadvantageous to mesophytic plants; it is a trait for hydrophytic plants, to enable mesophytic plants to prevent losing excess water. A plant with high stomatal density has a higher rate of water dissipation through evapotranspiration making the plant highly susceptible to fluctuating water conditions.

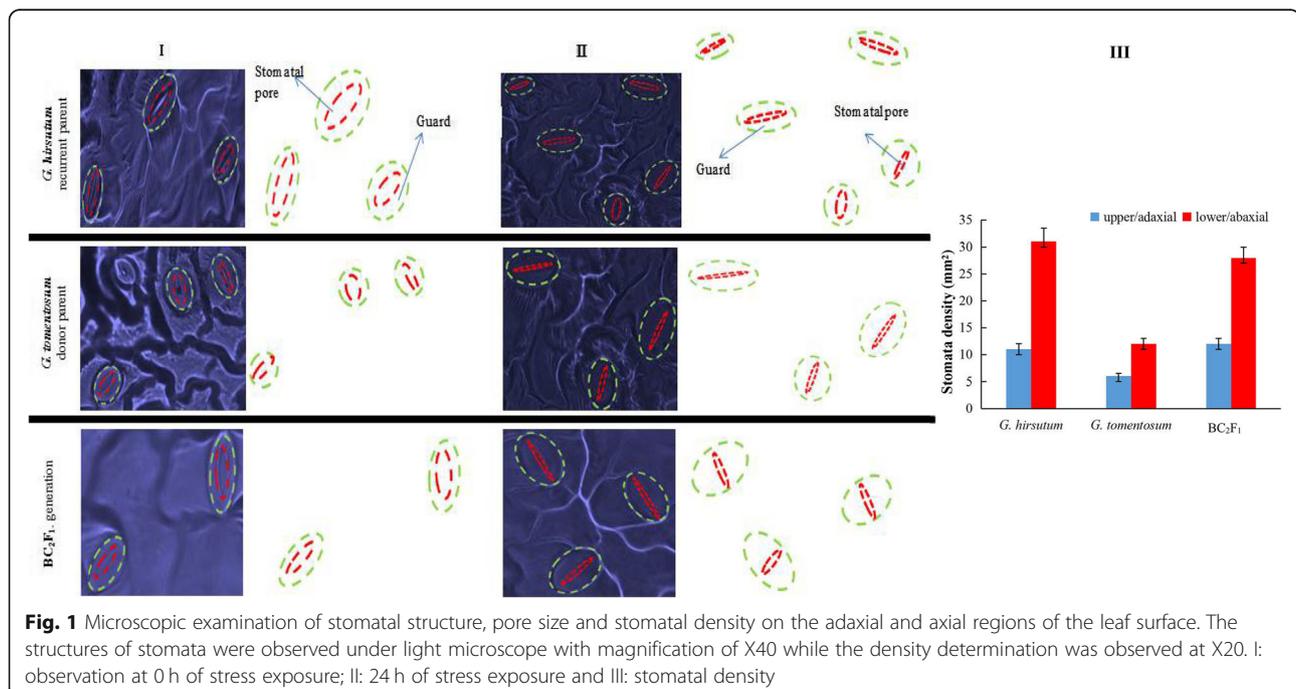
### Correlation analysis

To analyze the correlations among different traits, a Pearson's correlation coefficient on physiological and morphological traits was carried out. The analysis was performed by employing the statistical component of R software, version 3.4.2 "Performance Analytics" package with the Chart correlation function (R Development Core Team 2013). Significant positive correlations were noted between the following traits: PH with FLW, RLWC, FRB, DSB and TDB; Chl with RLWC, FRB and ELWL; FLW with SLW, RLWC, DSB, TDB, DSB/DRB, ELW and DLW; RLWC with FRB, TFB, DSB, DRB, TDB and FLW; FSB with TFB, DSB and DSB/DRB; FRB with TFB, DSB and TDB; TFB with FSB/FRB, DSB, TDB and DSB/DRB; DSB with DRB, TDB, DSB/DRB and DLW;

TBD with DSB/DRB and DLW; DSB/DRB with ELW and DLW. However, significant negative correlations were noted between the following traits: PH with FSB/FRB; Chl with SLW, FSB/FRB, DSB/DRB and ELW; SLW with RLWC, FRB, DRB and TDB; DLWS with RLWC, FSB, DRB and TDB; RLWC with DSB/DRB; finally ELW with ELWL (Additional file 6: Table S3). The result obtained correlated positively with previous findings, in which significant genotypic and phenotypic correlations have been detected for various physiological and morphological traits such as fresh root length, fresh shoot length, fresh root weight, fresh shoot weight, total fresh weight, dry root weight, dry shoot weight, total dry weight, photosynthetic rate, chlorophyll contents, leaf temperature and water use efficiency (Ali et al. 2015).

### GBS analysis and SNP generations in parental lines and BC<sub>2</sub>F<sub>2</sub> populations

The parental lines were sequenced using genotyping by sequencing (GBS) method with efficient sequencing depths. Regarding *G. hirsutum*-CRI-12 and *G. tomentosum*-AD3-00, average mapped reads of 10 individuals for each of the parental lines were mapped to the sequence of the cotton genome (<http://mascotton.njau.edu.cn>) and 13 695 154 and 13 496 550 reads were obtained, respectively. An average of 85 372 and 117 128 SNPs were identified for *G. hirsutum* and *G. tomentosum* respectively. The efficiency of enzyme digestion was 99% in both parental lines. The choice of enzyme is important in optimizing GBS for any given species, highlighting the importance of using in silico digests of



the genome of the target organism beforehand (Ariani et al. 2016).

For the BC<sub>2</sub>F<sub>2</sub> populations, the efficiency of enzyme digestion was relatively low compared with the efficiency levels of the two parental lines; the efficiency level for the BC<sub>2</sub>F<sub>2</sub> was 98.85%. A total of 1 507 193 217 mapped reads were produced, with an average of 5 074 724.636 mapped reads per individual, which corresponded to nearly 186.98 Gb of clean bases. The mapped reads obtained in the sequencing process were equivalent to 83.13-fold haploid genome coverage of raw paired-end Illumina reads by sequencing whole genome shotgun (WGS) libraries of homozygous cv. TM-1 compared with the results obtained by Li et al. (2015a). In their study, they generated 445.7 Gb of clean reads or 181-fold haploid genome coverage of raw paired-end Illumina reads by sequencing whole genome shotgun (WGS) libraries of homozygous cv. TM-1 with fragment lengths ranging from 250 to 40 000 bp. The average guanine cytosine (GC) content of the sequences was 38.25%, with a Q20 score of 94.66%. Base calling accuracy, measured by the Phred quality score (Q score), is the most common metric used to assess the accuracy of a sequencing platform. It indicates the probability that a given base is called either correctly or incorrectly by the sequencer. A lower base call accuracy of 90% (Q20) will have an incorrect base call probability of 1 in 100, meaning that every 100 bp sequenced read will likely to contain an error. When sequencing quality reaches Q30, virtually all of the reads will be perfect, having zero errors and ambiguities. High Q scores can decrease false-positive variant calls, and therefore result in accurate conclusions and lower costs for validation experiments (Salmela 2010). The parental lines, *G. hirsutum*-CRI-12 and *G. tomentosum*-AD3-00, were homozygous lines with “aa” and “bb” genotypes, respectively. The genotype “aa” × “bb”, consisting of 28 660 markers after removing duplicated markers was used for further analysis. All the SNPs generated were used because none fell below the threshold level and all had coverage of 75–100% of the entire BC<sub>2</sub>F<sub>2</sub> population.

Among the 28 660 SNP markers, the number of markers on the chromosomes ranged from 193 to 2 368 in the At\_sub-genome and 109 to 1918 in the Dt\_sub-genome. The markers covered 97.3%–100% of the length of the reference genome (Table 2). The highest marker locus was detected in Dt\_chr06 (38 markers/Mb), while the lowest level of marker locus density was noted in Dt\_chr05 (2 markers/Mb). The marker distribution was asymmetric. The highest number of markers was found on Dt\_chr06 with 2 419 markers while the lowest number of markers was detected on Dt\_chr05 with only 109 translating to only 0.38% of all the SNPs mapped.

We further compared the physical map sizes generated in this study with the A, D and AD genomes. In A genome, *Gossypium arboreum* was used. In D genome, we used *Gossypium raimondii*. In AD genome, we applied the physical map of *Gossypium hirsutum*. The genome coverage in the AD tetraploid cotton, *G. hirsutum*, ranged from 99% to 100%; almost all the chromosomes of the At-subgenome had 100% coverage except for At\_chr13, which had 98% coverage. In the Dt\_subgenome, Dt01\_chr14, Dt03\_chr17, Dt04\_chr22, Dt06\_chr25, Dt07\_chr16, Dt08\_chr24, Dt09\_chr20 and Dt12\_chr26 had 100% coverage while the remaining chromosomes had a coverage range between 97% and 99%. In addition, we checked if a similarly high percentage coverage observed in the tetraploid genome could be detected in A and D genomes when compared with the At and Dt subgenomes of the physical maps generated from this study. A huge variation was noted across the two genomes with the At\_subgenome physical map exhibiting the lowest coverage compared with the Dt-subgenome (Additional file 3: Figure S3).

#### High-density genetic linkage map with GBS markers

In the mapping of the BC<sub>2</sub>F<sub>2</sub> population, not all the 28 660 SNP markers generated were mapped. Several markers were duplicated within the same positions and with very high level of segregation distortion (SD). The repeated and highly distorted markers were filtered out. Finally, 10 888 markers were used and all were linked across the 26 linkage groups of the tetraploid cotton. The map generated from the 10 888 markers had a map size of 4 191.3 cM, with 2 149 cM and 2042.3 cM in At and Dt-subgenomes, respectively. The average marker distance was 0.384 9 cM, making the generated map to be the finest linkage map ever developed from the segregating a backcross population of the semi-wild-type cotton genotypes. The At-subgenome had the highest number of markers at 6 318 (58%) while the Dt-subgenome contained only 4 570 markers (42%). The results obtained could possibly be explained by the variation in sizes of the two tetraploid cotton sub-genome; the At\_subgenome is larger than the Dt\_subgenome.

The markers were unevenly distributed among the linkage groups (LGs). LG6\_chrD06 had the highest number of marker loci of 947 with a chromosome size of 158.72 cM, and an average marker distance of 0.168 cM. The LG1\_chrD01 had the lowest marker density loci, with only 45 markers, generating a map size of 151.78 cM with an average marker distance of 3.3728 cM. ChrA01, chrA02, chrA04, chrA07, chrA08, chrA11, chr18 (D13), chr20 (D10), chr24 (D08), chr25 (D06) and chr26 (D12) had more markers, as evident by the thick solid black regions within their chromosome strands (Fig. 2). Individual marker numbers were illustrated in

**Table 2** The GBS marker numbers per linkage group and their coverage on the 26 chromosomes of the AD cotton genome

Chromosomes	Number of markers	Cover length / Mb	Chr. Length / Mb	Coverage / %	Density/ (Number of markers-Mb <sup>-1</sup> )
At_chr01	2 306	99.850 26	99.884 7	100	23.10
At_chr02	2 368	83.276 69	83.447 906	99.8	28.40
At_chr03	221	100.212 2	100.263 05	99.9	2.21
At_chr04	2 057	62.761 93	62.913 772	99.8	32.80
At_chr05	778	91.982 2	92.047 023	99.9	8.46
At_chr06	193	102.953 2	103.170 44	99.8	1.87
At_chr07	1 572	78.155 38	78.251 018	99.9	20.10
At_chr08	1 664	103.608 9	103.626 34	100	16.10
At_chr09	1 054	74.861 52	74.999 931	99.8	14.10
At_chr10	1 139	100.69 26	100.866 6	99.8	11.30
At_chr11	1 650	93.305 73	93.316 192	100	17.70
At_chr12	353	87.396 59	87.484 866	99.9	4.04
At_chr13	1 285	78.020 31	79.961 121	97.6	16.50
At_sub genome	16 640	1 157.077	1 160.23	99.7	14.40
Dt_chr01	198	60.915 16	61.456 009	99.1	3.25
Dt_chr02	237	67.222 26	67.284 553	99.9	3.53
Dt_chr03	161	46.671 3	46.690 656	100	3.45
Dt_chr04	365	51.270 27	51.454 13	99.6	7.12
Dt_chr05	109	60.243 6	61.933 047	97.3	1.81
Dt_chr06	2 419	64.091 13	64.294 643	99.7	37.70
Dt_chr07	393	55.293 15	55.312 611	100	7.11
Dt_chr08	1 918	65.833 81	65.894 135	99.9	29.10
Dt_chr09	852	50.907 61	50.995 436	99.8	16.70
Dt_chr10	1 854	62.785 48	63.374 666	99.1	29.50
Dt_chr11	691	65.543 61	66.087 774	99.2	10.50
Dt_chr12	1 593	58.942 78	59.109 837	99.7	27.00
Dt_chr13	1 230	59.879 4	60.534 298	98.9	20.50
Dt_sub genome	12 020	769.599 5	774.421 8	99.4	15.60
Totals	28 660	1 926.677	1 934.65	99.6	14.90

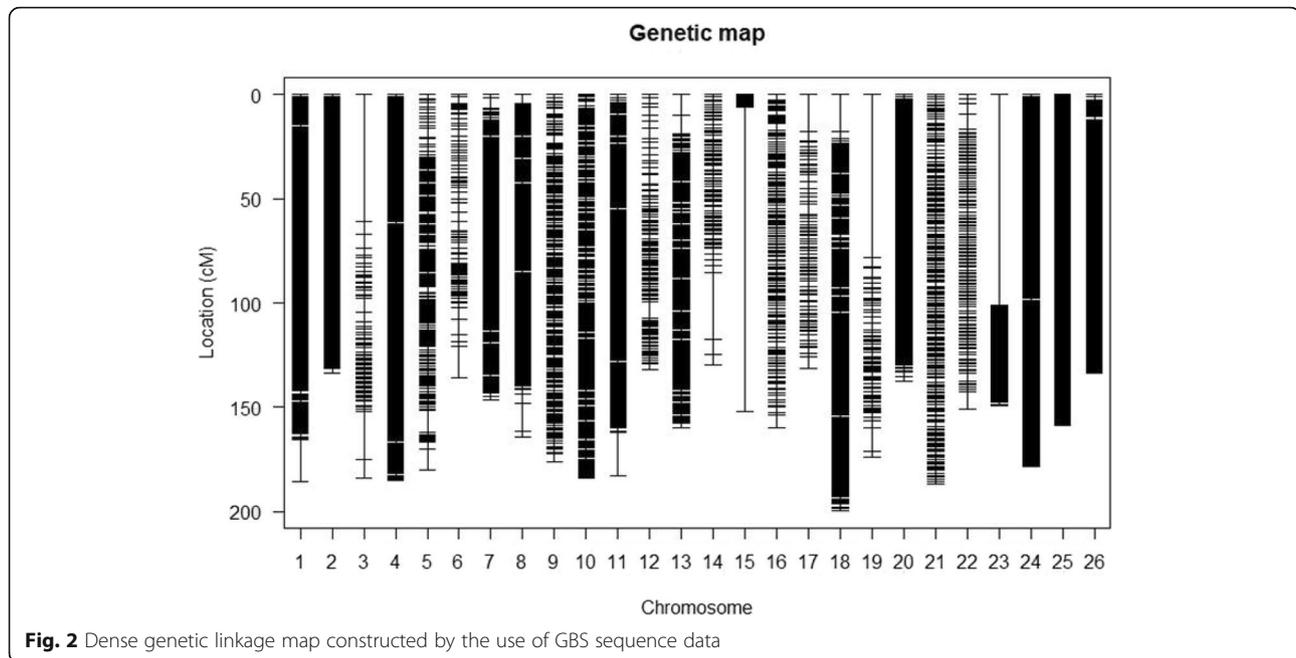
At: At-subgenome, Dt: Dt-subgenome, Mb: Mega base

Additional file 7: Table S4. Chromosome 15 (D01) had the lowest number of markers at 45, but had the smallest gap of 0.1047 cM among all 26 chromosomes. The marker file used for the construction of the genetic map, including the physical position in base pairs (bp) and centi-Morgan (cM), which also includes the allele scores for each of the 200 BC<sub>2</sub>F<sub>2</sub> individuals genotyped, are as shown in Additional file 8: Table S5.

#### Identification of Consistent and Clustered QTLs Region

The genetic variation of a quantitative trait is controlled by the collective effects of numerous genes, known as quantitative trait loci (QTLs), and therefore the identification of QTLs is of agronomic importance and its use in crop is significant for improving not only cotton but other plants as well. In this study, we identified 30 stable

QTLs among the 60 detected QTLs for 12 traits, which were cell membrane stability (CMS), chlorophyll content, evaluated through SPAD values (Chl), saturated leaf weight (SLW), leaf fresh weight (LFW), dry leaf weight (DLW), fresh shoot biomass (FSB), dry shoot biomass (DSB), total fresh biomass (TFB), the ratio between fresh shoot biomass and fresh root biomass (FSB/FRB), total dry biomass (TDB) and the ratio between dry shoot biomass and dry root biomass (DSB/DRB). The stable QTLs were detected in at least two environments, coded as E<sub>1</sub> (environment 1), E<sub>2</sub> (environment 2) and CA (combined analysis for E<sub>1</sub> and E<sub>2</sub>). In declaring the consistent QTLs, only environments 1 and 2 were considered. The distribution patterns of the stable QTLs were skewed toward the At-sub genome with 17 QTLs, while the remaining 13 QTLs were located in the Dt\_sub genome. This



supports the *At*\_sub genome being relatively larger in genome size compared with that of the *Dt*-sub genome. The stable QTLs were distributed in the following chromosomes: chrA01 (4 QTLs), chrA03 (1), chrA04 (1), chrA05 (5), chrA07 (1), chrA09 (2), chrA11 (1), chrA12 (1), chrA13 (1), chr15\_D01 (6), chr22\_D04 (1), chr19\_D05 (1), chr16\_D07 (1), chr23\_D09 (2), chr20\_D10 (1) and chr18\_D13 (1 QTL). QTL clusters are genome regions in which large quantities of QTLs are co-localized, also commonly referred to as QTL hot spot (Singh et al. 2017). Sixteen clusters for 11 traits were detected. The highest number of consistent QTLs mapped was six and all were identified in the marker interval of D01\_1 317 927–D01\_2 067 711 in cluster 10. This region was designated as Cluster 11, which ranged from 1 317 927 to 2 067 711 bp. The cluster harbored 6 QTLs for DSB, FSB, SLW, TDB, TFB and DSB/DRB, which explained the phenotypic variance range from 0.0435% to 24.3703%. The lowest numbers of major QTLs were identified in Clusters 2, 3, 5, 7, 8, 9, 11, 12, 13, 15 and 16, which harbored QTLs for FSB/FRB, DLW, TDB, SLW, SLW, TDB, SLW, DSB/DRB, SLW, SPAD and SPAD, respectively, with QTLs proportions per cluster of 3, 9, 3, 3, 3, 6, 2, 2, 3, 4 and 3, respectively (Table 3).

In determining the parental contributions towards the stable QTLs detected, *G. tomentosum*, used as the donor male parent, was found significantly contributed towards the following traits: saturated leaf weight (SLW), chlorophyll content (SPAD measured), total dry biomass (TDB), cell membrane stability (CMS), fresh shoot biomass (FSB) and total fresh biomass (TFB); the female parent, *G. hirsutum*, contributed towards dry shoot

biomass (DSB), fresh shoot biomass/fresh root biomass (FSB/FRB), leaf fresh weight (LFW) and dry shoot biomass/dry root biomass (DSB/RB). The stable QTLs were found to exhibit multiple duplications, 89 and 55 duplication events for QTLs contributed by *G. hirsutum* and *G. tomentosum*, respectively. Those which were duplicated within the same chromosomes were referred to as tandemly duplicated QTLs while those that exhibited duplication across different chromosomes were termed as segmentally duplicated QTLs.

Gene action is a vital indicator of the contribution of the QTLs detected on the overall performance of the plant under the stress condition being investigated (Lopes et al. 2014). The gene actions are described in four terms, namely as additive effect (Ae), dominant effect (De) partial dominance (PD) and over dominance (OD) as described by Paterson et al., (Stubber et al. 1987). In this study, we detected all four gene action attributes. A majority of the QTLs were found to exhibit over dominance gene action, regulating 10 QTLs. This was closely followed by dominance gene effect with 8 QTLs, then partial dominance with 7 QTLs and the additive gene effect with only 5 QTLs. Over dominance (OD) was observed in 10 traits, such as chlorophyll content as measured by SPAD values, cell membrane stability (CMS), saturated leaf weight (SLW), leaf fresh weight (LFW), fresh shoot biomass (FSB), dry shoot biomass (DSB), total fresh biomass (TFB), total dry biomass (TDB), fresh shoot biomass/fresh root biomass (TSB/FRB) and ratio of dry shoot biomass/ dry root biomass (DSB/DRB) (Additional file 9: Table S6).

Phenotypic variation, explained by a single QTL detected in this study, ranged from 0% to slightly above

**Table 3** Physiochemical properties and sub cellular localization prediction of the mined genes within the major clusters of the consistent QTLs

Gene ID	Protein Length / aa	Molecular Weight / kDa	Charge	Isoelectric Point	Grand Average of Hydropathy	Exon Number	Mean Exon Length / bp	Mean Intron Length / bp	Sub cellular localization		
									TargetP	pprowler	Wolfpsort
Gh_A01G1774	357	39.979	0	6.504	-0.128	5	214.8	430.5	S	SP	nucl
Gh_A01G1935	988	109.088	14	7.934	0.013	4	741.8	160.3	_	OTHER	nucl
Gh_A01G1936	352	39.524	-1.5	5.598	-0.034	3	353	95.5	_	OTHER	chlo
Gh_A01G1937	451	50.487	2.5	7.065	-0.067	3	452	93	_	OTHER	nucl
Gh_A01G1938	427	47.747	-1	6.265	-0.044	3	428	1 930.50	_	OTHER	chlo
Gh_A01G1939	779	86.296	12.5	8.292	-0.196	9	260	151.8	S	SP	nucl
Gh_A01G1940	306	33.661	15.5	10.036	0.895	4	230.3	817	S	SP	chlo
Gh_A01G1941	280	30.916	-20	4.234	-0.358	2	421.5	2 224.00	_	OTHER	nucl
Gh_A01G1942	227	25.397	10.5	9.383	-0.346	8	85.5	469.7	C	CTP	cyto
Gh_A01G1943	754	88.329	7	6.956	-0.743	19	119.2	432	_	OTHER	nucl
Gh_A01G1944	463	51.223	21.5	10.135	-0.833	1	1 392.00	No intron	C	OTHER	nucl
Gh_A01G1945	98	11.051	-7.5	4.26	-0.219	1	297	No intron	_	OTHER	cyto
Gh_A01G1946	695	78.499	10.5	7.85	-0.104	1	2088.00	No intron	_	mTP	chlo
Gh_A01G1947	256	29.155	-16.5	4.38	-0.622	2	385.5	347	C	CTP	chlo
Gh_A01G1948	558	61.373	10	7.821	0.067	5	335.4	546	S	SP	chlo
Gh_A05G2519	986	112.494	33.5	9.137	-0.295	1	2 961.00	No intron	C	CTP	chlo
Gh_A05G2520	854	93.522	14.5	8.815	-0.228	3	855	317.5	_	OTHER	cyto
Gh_A05G2521	255	27.771	1.5	7.435	-0.935	1	768	No intron	_	OTHER	nucl
Gh_A05G2522	232	25.714	-12	4.252	0.126	7	99.9	425	C	CTP	chlo
Gh_A05G2523	116	12.804	2.5	7.904	-0.213	2	175.5	158	_	OTHER	cyto
Gh_A05G3283	320	35.87	6	7.592	-1.076	3	321	246	_	OTHER	nucl
Gh_A05G3284	152	17.211	2	7.448	0.12	2	229.5	93	S	SP	extr
Gh_A05G3285	508	56.448	11	8.69	-0.318	7	218.1	237.8	_	SP	chlo
Gh_A05G3286	272	30.036	4.5	7.7	-0.785	4	204.8	562.7	_	OTHER	nucl
Gh_A05G3287	1 927	224.222	37	8.872	-0.107	42	137.7	529.6	C	mTP	plas
Gh_D01G0174	550	61.943	-10	5.405	-0.205	3	551	1 097.50	_	OTHER	cyto

**Table 3** Physiochemical properties and sub cellular localization prediction of the mined genes within the major clusters of the consistent QTLs (*Continued*)

Gene ID	Protein Length / aa	Molecular Weight / kDa	Charge	Isoelectric Point	Grand Average of Hydropathy	Exon Number	Mean Exon Length / bp	Mean Intron Length / bp	Sub cellular localization		
									TargetP	pprowler	Wolfpsort
Gh_D01G0175	700	80.06	4	6.882	-0.197	2	1 051.50	567	_	OTHER	cyto
Gh_D01G0176	401	45.366	-20	4.645	-0.789	7	172.3	14 444.20	_	OTHER	chlo
Gh_D01G0177	402	43.82	0.5	6.546	0.076	1	1 209.00	No intron	S	SP	chlo
Gh_D01G0178	718	81.905	22	9.613	-0.889	16	134.8	133.1	_	OTHER	nucl
Gh_D01G0179	115	12.729	-1.5	5.533	0.254	1	348	No intron	S	SP	extr
Gh_D01G0180	339	38.396	9.5	7.668	-0.782	1	1 020.00	No intron	_	OTHER	nucl
Gh_D01G0181	341	38.471	9	7.512	-0.769	1	1 026.00	No intron	_	OTHER	nucl
Gh_D01G0182	888	97.577	-24.5	4.905	-0.095	25	106.7	153.5	_	OTHER	plas
Gh_D01G0183	424	46.661	1	6.587	-0.69	6	212.5	170.6	_	OTHER	nucl
Gh_D01G0184	73	7.777	2	7.966	1.595	2	111	153	S	SP	extr
Gh_D01G0185	348	38.639	7.5	8.795	-0.395	3	349	469	_	OTHER	cysk
Gh_D01G0186	670	75.597	14	8.168	-0.235	13	154.8	309	M	mTP	nucl
Gh_D01G0187	672	75.797	19	8.641	-0.236	13	155.3	399.5	M	mTP	nucl
Gh_D01G0188	1 029	114.338	-2	6.384	-0.115	3	1 030.00	245	_	OTHER	nucl
Gh_D01G0189	558	62.6	14	7.929	-0.102	1	1 677.00	No intron	M	mTP	chlo
Gh_D01G0190	1 136	125.426	30.5	9.241	0.704	4	852.8	496.3	C	SP	plas
Gh_D01G0191	225	24.848	5.5	7.502	-0.988	4	169.5	417	_	OTHER	nucl
Gh_D01G0192	203	23.652	5.5	9.159	-0.758	4	153	725.7	_	OTHER	cyto
Gh_D01G0193	911	100.512	-9	5.919	-0.548	15	182.4	197.1	_	OTHER	nucl
Gh_D01G0194	363	39.691	-4.5	5.24	-0.699	3	364	732	_	OTHER	nucl
Gh_D01G0195	827	92.435	14	7.517	-0.473	19	130.7	259.8	_	OTHER	nucl
Gh_D01G0196	206	23.297	13	10.798	-0.618	5	124.2	100.8	_	OTHER	nucl
Gh_D01G0197	228	25.286	-6	5.286	-0.479	5	137.4	527.3	_	OTHER	chlo
Gh_D01G0198	181	19.829	-17.5	4.236	-0.926	2	273	235	C	OTHER	chlo
Gh_D01G0199	130	14.537	0	6.475	0.075	3	131	1 274.50	_	SP	extr
Gh_D01G0200	470	51.491	-4.5	5.821	-0.604	4	353.3	336.3	_	OTHER	nucl

**Table 3** Physiochemical properties and sub cellular localization prediction of the mined genes within the major clusters of the consistent QTLs (*Continued*)

Gene ID	Protein Length / aa	Molecular Weight / kDa	Charge	Isoelectric Point	Grand Average of Hydropathy	Exon Number	Mean Exon Length / bp	Mean Intron Length / bp	Sub cellular localization		
									TargetP	pprowler	Wolfpsort
Gh_D01G0201	362	39.258	-1.5	5.891	-0.043	14	77.8	227.1	M	mTP	mito
Gh_D01G0202	266	29.758	11	9.845	0.62	7	114.4	429.2	S	SP	chlo
Gh_D01G0203	463	50.97	3.5	6.989	-0.594	5	278.4	724	C	OTHER	nucl
Gh_D01G0204	463	51.217	5.5	7.863	-0.613	5	278.4	788.3	C	OTHER	nucl
Gh_D01G0205	481	53.146	-4.5	6.102	-0.05	10	144.6	377.6	_	OTHER	cyto
Gh_D01G0206	259	28.11	13	10.307	-0.397	7	111.4	302.8	C	cTP	chlo
Gh_D01G0207	450	50.452	2	6.841	-0.286	7	193.3	220.3	_	OTHER	cyto
Gh_D01G0208	598	67.75	5	7.185	-0.373	10	179.7	284.9	C	cTP	chlo
Gh_D01G0209	268	30.766	-13	4.653	-0.491	1	807	No intron	_	SP	E.R.
Gh_D01G0210	702	75.521	0	6.545	-0.06	4	527.3	196.3	C	cTP	chlo
Gh_D01G0211	751	84.485	3	6.889	-0.144	9	250.7	499.4	C	cTP	chlo
Gh_D01G0212	521	57.652	0.5	6.631	-0.122	9	174	404.9	S	SP	chlo
Gh_D01G0213	372	42.957	5.5	7.368	-0.469	6	186.5	1 091.20	M	mTP	nucl
Gh_D01G0214	691	76.102	3.5	6.837	-0.005	20	103.8	697.4	_	OTHER	extr
Gh_D01G0215	697	76.446	11.5	8.01	-0.082	23	91	278.4	_	OTHER	chlo
Gh_D01G0216	796	87.405	12	8.153	0.19	24	99.6	190.6	_	SP	plas
Gh_D01G0217	366	40.714	12.5	9.663	-0.562	11	100.1	316.3	_	OTHER	nucl
Gh_D01G0218	118	13.346	0.5	6.672	-0.161	3	119	768.5	_	OTHER	cyto
Gh_D01G0219	159	18.301	2.5	7.734	-0.451	1	480	No intron	_	OTHER	cyto
Gh_D01G0220	329	37.446	9	9.026	-0.553	8	123.8	344.4	_	OTHER	nucl
Gh_D01G0221	73	8.004	-1.5	5.584	0.064	2	111	133	_	OTHER	cyto
Gh_D01G0222	212	23.909	8	9.006	-0.018	4	159.8	425.7	S	SP	nucl
Gh_D01G0223	251	27.12	-1.5	6.022	-0.213	4	189	413.3	_	OTHER	chlo
Gh_D01G0224	445	50.884	13	9.42	-0.269	6	223	436	_	SP	chlo
Gh_D01G0225	386	42.709	-13	4.668	0.091	11	105.5	309.3	_	OTHER	cyto
Gh_D01G0226	1 076	119.125	-6	6.006	-0.647	13	248.5	321.1	_	OTHER	nucl

**Table 3** Physiochemical properties and sub cellular localization prediction of the mined genes within the major clusters of the consistent QTLs (*Continued*)

Gene ID	Protein Length / aa	Molecular Weight / kDa	Charge	Isoelectric Point	Grand Average of Hydropathy	Exon Number	Mean Exon Length / bp	Mean Intron Length / bp	Sub cellular localization		
									TargetP	pprowler	Wolfpsort
Gh_D01G0227	162	18.032	10	10.114	0.358	5	97.8	275	S	SP	plas
Gh_D01G0228	239	26.313	5	9.003	0.04	1	720	No intron	_	OTHER	chlo
Gh_D01G0229	239	26.218	6.5	9.285	-0.033	1	720	No intron	_	OTHER	mito
Gh_D01G0230	207	23.198	11.5	9.683	-0.968	1	624	No intron	_	OTHER	nucl
Gh_D01G0231	420	47.488	9.5	8.78	-0.308	8	157.9	368.3	C	cTP	chlo
Gh_D01G0232	420	47.471	11.5	8.886	-0.304	8	157.9	886.3	C	cTP	chlo
Gh_D01G0233	1 035	112.771	31.5	8.877	-0.45	9	345.3	284.6	_	OTHER	nucl
Gh_D01G0234	134	15.462	14	10.864	-0.41	5	81	296.3	_	OTHER	cyto
Gh_D01G0235	873	98.951	-0.5	6.475	-0.233	24	109.3	252.7	S	OTHER	nucl
Gh_D01G0236	1 039	116.109	11.5	7.216	-0.274	5	624	276.5	_	mTP	nucl
Gh_D01G0237	194	20.965	3.5	8.967	-1.206	4	146.3	518.3	_	OTHER	nucl

aa Amino acid, kDa kiloDalton, bp Base pair, chlo Chloroplast, cyto Cytoplasm, extr Extracellular part of the cell, nucl Nucleus, mito Mitochondrion, cysk Cytoskeleton, golg Golgi body, vacu Vacuole, plas Plasma membrane, ER Endoplasmic reticulum, SP Secretory pathway (presence of a signal peptide), mTP Mitochondrial targeting peptide, cTP Chloroplast transit peptide, Other (nucleus, cytoplasmic, or otherwise), C cytoplasm, S secretory pathway, M mitochondrion and -: others/other cell organelles

33.57%. A similar result has also been observed in the mapping of QTLs related to yield components and oligogenic control of the cap color of the button mushroom, *Agaricus bisporus*, in which the PPC1 locus, together with two additional genomic regions, were found to explain up to 90% of the phenotypic variation of the cap color, while the highest phenotypic variation explained by a single QTL was 84.5% (Foulongne-Oriol et al. 2012). The consistent QTL LODs ranged from 2.5038 to a maximum value of 6.71226, indicating that the QTLs detected were far above the noise regions and therefore harbored vital genes with greater effect on the performance of cotton under drought stress condition.

#### Identification of the candidate genes within the major QTLs clusters for CMS, SLW and chlorophyll content traits

The two parental lines used, *G. tomentosum* (donor male parent) and *G. hirsutum* (recurrent female parent) are phenotypically diverse genotypes. *G. hirsutum* is superior phenotypically compared with *G. tomentosum* except that *G. tomentosum* has greater tolerance towards salt and drought stress, being an endemic species of the dry and saline Hawaiian island (Oluoch et al. 2016). Therefore, morphologically related QTLs detected in this study, such as DSB, TDB, FSB/FRB, DLW, FSB, LFW, TFB and DSB/

DRB, were not considered in the determination of the candidate genes within the QTL regions. We considered the three main clusters, cluster 1 (4 QTLs), cluster 4 (5 QTLs) and 10 (6 QTLs), but with emphasis on the physiologically related QTLs contributed by the donor parent, *G. tomentosum*, which were cell membrane stability (CMS), chlorophyll content (SPAD determined) and saturated leaf weight (SLW). Clusters 1, 4 and 10 were located on chrA01, chrA05 and chr15 (D01), respectively. Eighty-nine genes were obtained, which could be critical in the regulation of CMS, SLW and chlorophyll content as evaluated through SPAD. For CMS, 10 genes were found, in which 5 were mined within 34 592 397–34 724 734 kb and the other 5 genes were obtained within the marker regions of 86 061 394–86 236 836 kb. For SLW we obtained 78 genes, 14 genes within the marker region 99 298 866–99 406 421 kb and 64 genes were obtained from the marker regions 1 317 927–2 067 711 kb, and finally a single gene was obtained for the trait chlorophyll content, as determined by SPAD values within the marker regions of 97 155 069–97 196 848 kb (Additional file 10: Table S7). Gene duplication is the mechanism underlying the evolution and expansion of genes (Magadam et al. 2013). Because of gene duplication, the overall numbers of genes were 110, a majority of duplication was detected for SLW, while only a single

duplication was detected for the gene controlling chlorophyll concentration, as determined through SPAD measurements. A unique observation was made among the determined genes within the QTL regions found to be regulating the three physiological traits. Eighteen genes were found to be uncharacterized genes, accounting for 20.22% of all the determined genes within the QTL regions. Detection of these uncharacterized genes could imply that new genes were evolving because the induction of genes are an adaptive feature adopted by plants to contain the deleterious effects caused by various abiotic stresses in which they are exposed. However, more research needs to be done to determine the exact roles of these uncharacterized genes.

#### **Physiochemical properties, gene structure analysis and GO functional annotation of the 89 mined genes within the three major clusters**

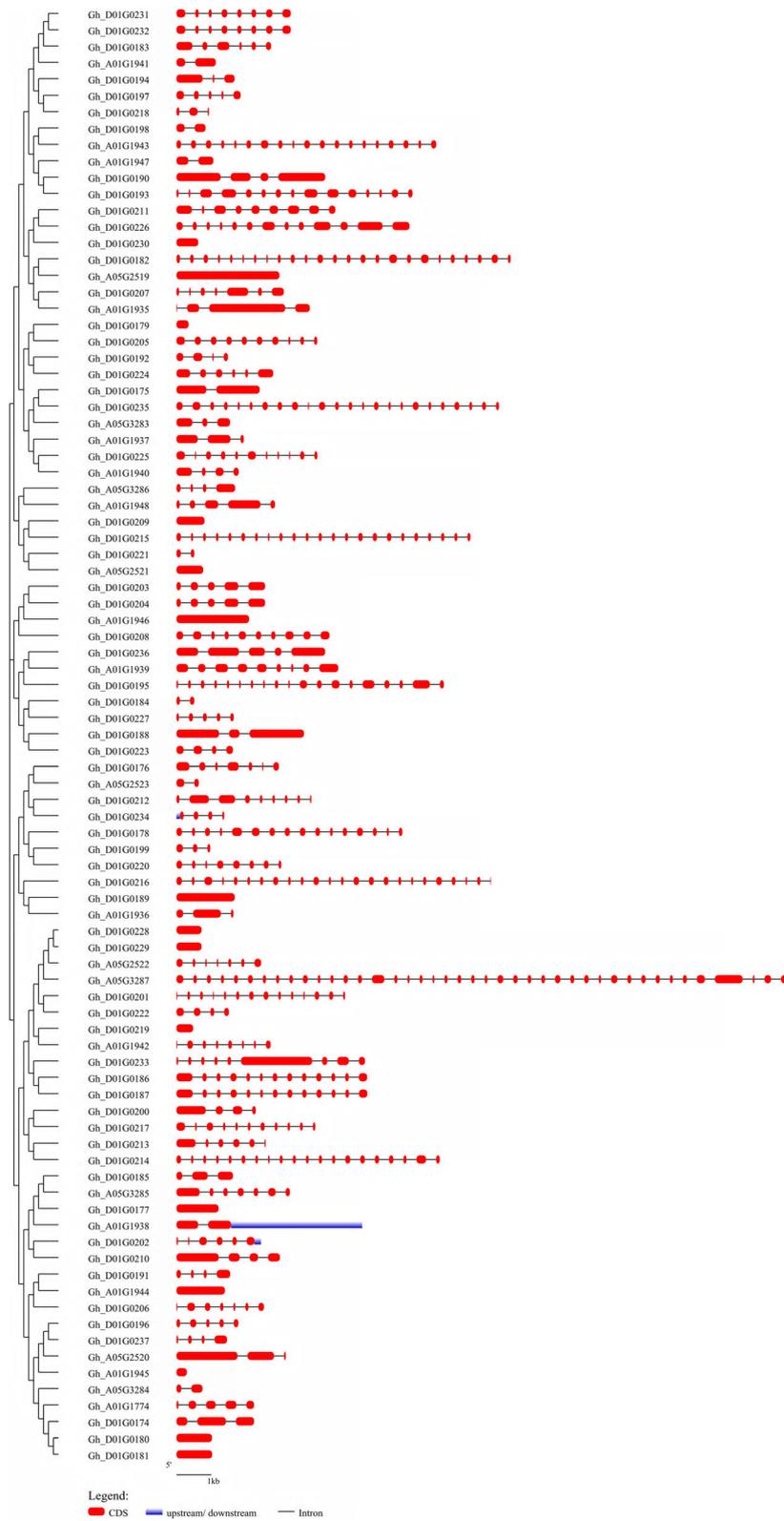
Gene physiochemical properties, such as molecular weights, grand average hydropathy values and isoelectric points, are important factors in determining the functionality of the genes. We sorted the 89 mined genes to determine their physiochemical properties. The protein lengths of the mined genes ranged from 73 to 1927 amino acids (aa), the molecular weights ranged from 7.777 to 224.222 kDa, the charge ranged from -24.5 to +37, and the GRAVY (Grand average of hydropathy) values ranged from -1.206 to 1.595, with 73 of the 89 genes having negative GRAVY. This is an indication that the mined genes had hydrophobic properties, a common feature for most abiotic stress related genes, such as *LEA* genes (Magwanga *et al.* 2018b). A majority of the genes were found to be interrupted by introns, while only 15 genes, *Gh\_A01G1944*, *Gh\_A01G1945*, *Gh\_A01G1946*, *Gh\_A05G2519*, *Gh\_A05G2521*, *Gh\_D01G0177*, *Gh\_D01G0179*, *Gh\_D01G0180*, *Gh\_D01G0181*, *Gh\_D01G0189*, *Gh\_D01G0209*, *Gh\_D01G0219*, *Gh\_D01G0228*, *Gh\_D01G0229* and *Gh\_D01G0230*, only 17% of all the mined genes, were intronless. (Fig. 3 and Table 4). Despite that a majority of the genes were interrupted by the introns, the intron numbers were relatively low, ranging from 2 to 43, implying that the burden occasioned by the intron interruptions was greatly reduced in these genes, an indication of their significant contribution in enhancing abiotic stress tolerance in upland cotton. The genes were located in various subcellular compartments. Twenty-six proteins encoding the candidate genes were embedded within the chloroplast, 14 cytoplasmic proteins, 1 cytoskeleton protein, 35 nucleic proteins, 5 plasma membranous proteins, 2 mitochondrion proteins, 5 extracellular structural proteins and finally only one endoplasmic reticulum (E.R) protein (Table 4 and Additional file 11: Table S8). The wider distribution of the proteins encoding the mined genes within the various cell structures provided an indication of their significant role within the cell. The

highest number of proteins encoding the mined genes was nucleic proteins. The nucleus regulates and coordinates vital cellular activities in order to minimize the deleterious effects of water stress within the cell (Fernández and Strand 2008).

Gene ontology (GO) provides fundamental information on which particular mechanism or part of the cell the genes play a role GO basically groups the genes into three categories (Dessimoz and Škunca 1984-2020). There are three fundamental processes describing gene ontology, namely, cellular component (CC), biological function (BF) and molecular process (MP) (Wood 2008). The three GO terms were detected for the mined genes. The highest level of GO annotation was observed for *Gh\_A01G1943* with 14 GO functional annotations, DNA ligase (ATP) activity (GO:0003910), mRNA guanylyltransferase activity (GO:0004484), polynucleotide 5'-phosphatase activity (GO:0004651), protein tyrosine phosphatase activity (GO:0004725), ATP binding (GO:0005524), nucleus (GO:0005634), DNA repair (GO:0006281), DNA recombination (GO:0006310), 7-methylguanosine mRNA capping (GO:0006370), mRNA processing (GO:0006397), protein dephosphorylation (GO:0006470), protein tyrosine/serine/threonine phosphatase activity (GO:0008138), dephosphorylation (GO:0016311) and phosphatase activity (GO:0016791). *Gh\_A01G1943* was mined within the QTL region for its saturated leaf weight (SLW) trait, within the marker regions of 99 298 866 to 99 406 421 bp. Leaf water saturation is a physiological process, mediated by a passive process known as osmosis, which occurs when the membrane integrity of the plant cell is maintained and not affected by water stress (Prado and Maurel 2013). The rest of the genes were found to be involved in one to a maximum of nine GO functional processes (Additional file 12: Table S9).

#### **Phylogenetic tree analysis of the mined genes**

The candidate genes were obtained from the QTL regions for CMS, SLW and chlorophyll content as determined through SPAD values. We investigated the evolutionary relationship of the mined genes to determine the orthologous gene pairs, if at all they could regulate similar trait or not, multiple sequence alignment of the mined genes was done. Based on phylogenetic tree analysis, the genes were classified into four groups. Members of group one were the majority, with 35 genes (39.33%), of all the genes obtained within the QTL regions. Moreover, two ortholog gene pairs were found to be controlling two traits. For example, *Gh\_D01G0223* and *Gh\_A01G1774* were obtained within the QTL regions controlling SLW and chlorophyll content traits, respectively. Similarly, *Gh\_D01G0201* and *Gh\_A05G3285* were ortholog pairs obtained within the QTL regions regulating SLW and CMS traits, respectively. In group 2, 3 and 4, three pairs of ortholog genes were found



**Fig. 3** Phylogenetic tree, and gene structure of the mined genes. The phylogenetic tree was constructed using MEGA 7.0. Exon/intron structures of the genes in upland cotton, exons introns and up/down-stream were represented by red boxes, black lines and blue boxes, respectively

**Table 4** Physiochemical properties and sub cellular localization prediction of the mined genes within the major clusters of the consistent QTLs

Gene ID	Protein Length / aa	Molecular Weight / kDa	Charge	pI	GRAVY	Exon Number	Mean Exon Length / bp	Mean Intron Length / bp	Sub cellular localization		
									TargetP	pprowler	Wolfpsort
Gh_A01G1774	357	39.979	0	6.504	-0.128	5	214.8	430.5	S	SP	nucl
Gh_A01G1935	988	109.088	14	7.934	0.013	4	741.8	160.3	-	Other	nucl
Gh_A01G1936	352	39.524	-1.5	5.598	-0.034	3	353	95.5	-	Other	chlo
Gh_A01G1937	451	50.487	2.5	7.065	-0.067	3	452	93	-	Other	nucl
Gh_A01G1938	427	47.747	-1	6.265	-0.044	3	428	1 930.50	-	Other	chlo
Gh_A01G1939	779	86.296	12.5	8.292	-0.196	9	260	151.8	S	SP	nucl
Gh_A01G1940	306	33.661	15.5	10.036	0.895	4	230.3	817	S	SP	chlo
Gh_A01G1941	280	30.916	-20	4.234	-0.358	2	421.5	2 224.00	-	Other	nucl
Gh_A01G1942	227	25.397	10.5	9.383	-0.346	8	85.5	469.7	C	cTP	cyto
Gh_A01G1943	754	88.329	7	6.956	-0.743	19	119.2	432	-	Other	nucl
Gh_A01G1944	463	51.223	21.5	10.135	-0.833	1	1 392.00	No intron	C	Other	nucl
Gh_A01G1945	98	11.051	-7.5	4.26	-0.219	1	297	No intron	-	Other	cyto
Gh_A01G1946	695	78.499	10.5	7.85	-0.104	1	2088.00	No intron	-	mTP	chlo
Gh_A01G1947	256	29.155	-16.5	4.38	-0.622	2	385.5	347	C	cTP	chlo
Gh_A01G1948	558	61.373	10	7.821	0.067	5	335.4	546	S	SP	chlo
Gh_A05G2519	986	112.494	33.5	9.137	-0.295	1	2 961.00	No intron	C	cTP	chlo
Gh_A05G2520	854	93.522	14.5	8.815	-0.228	3	855	317.5	-	Other	cyto
Gh_A05G2521	255	27.771	1.5	7.435	-0.935	1	768	No intron	-	Other	nucl
Gh_A05G2522	232	25.714	-12	4.252	0.126	7	99.9	425	C	cTP	chlo
Gh_A05G2523	116	12.804	2.5	7.904	-0.213	2	175.5	158	-	Other	cyto
Gh_A05G3283	320	35.87	6	7.592	-1.076	3	321	246	-	Other	nucl
Gh_A05G3284	152	17.211	2	7.448	0.12	2	229.5	93	S	SP	extr
Gh_A05G3285	508	56.448	11	8.69	-0.318	7	218.1	237.8	-	SP	chlo
Gh_A05G3286	272	30.036	4.5	7.7	-0.785	4	204.8	562.7	-	Other	nucl
Gh_A05G3287	1927	224.222	37	8.872	-0.107	42	137.7	529.6	C	mTP	plas
Gh_D01G0174	550	61.943	-10	5.405	-0.205	3	551	1 097.50	-	Other	cyto
Gh_D01G0175	700	80.06	4	6.882	-0.197	2	1 051.50	567	-	Other	cyto
Gh_D01G0176	401	45.366	-20	4.645	-0.789	7	172.3	14 444.20	-	Other	chlo
Gh_D01G0177	402	43.82	0.5	6.546	0.076	1	1 209.00	No intron	S	SP	chlo
Gh_D01G0178	718	81.905	22	9.613	-0.889	16	134.8	133.1	-	Other	nucl
Gh_D01G0179	115	12.729	-1.5	5.533	0.254	1	348	No intron	S	SP	extr
Gh_D01G0180	339	38.396	9.5	7.668	-0.782	1	1 020.00	No intron	-	Other	nucl
Gh_D01G0181	341	38.471	9	7.512	-0.769	1	1 026.00	No intron	-	Other	nucl
Gh_D01G0182	888	97.577	-24.5	4.905	-0.095	25	106.7	153.5	-	Other	plas
Gh_D01G0183	424	46.661	1	6.587	-0.69	6	212.5	170.6	-	Other	nucl
Gh_D01G0184	73	7.777	2	7.966	1.595	2	111	153	S	SP	extr
Gh_D01G0185	348	38.639	7.5	8.795	-0.395	3	349	469	-	Other	cysk
Gh_D01G0186	670	75.597	14	8.168	-0.235	13	154.8	309	M	mTP	nucl
Gh_D01G0187	672	75.797	19	8.641	-0.236	13	155.3	399.5	M	mTP	nucl
Gh_D01G0188	1 029	114.338	-2	6.384	-0.115	3	1 030.00	245	-	Other	nucl
Gh_D01G0189	558	62.6	14	7.929	-0.102	1	1 677.00	No intron	M	mTP	chlo
Gh_D01G0190	1 136	125.426	30.5	9.241	0.704	4	852.8	496.3	C	SP	plas

**Table 4** Physiochemical properties and sub cellular localization prediction of the mined genes within the major clusters of the consistent QTLs (*Continued*)

Gene ID	Protein Length / aa	Molecular Weight / kDa	Charge	pI	GRAVY	Exon Number	Mean Exon Length / bp	Mean Intron Length / bp	Sub cellular localization		
									TargetP	pprowler	Wolfpsort
Gh_D01G0191	225	24.848	5.5	7.502	-0.988	4	169.5	417	-	Other	nucl
Gh_D01G0192	203	23.652	5.5	9.159	-0.758	4	153	725.7	-	Other	cyto
Gh_D01G0193	911	100.512	-9	5.919	-0.548	15	182.4	197.1	-	Other	nucl
Gh_D01G0194	363	39.691	-4.5	5.24	-0.699	3	364	732	-	Other	nucl
Gh_D01G0195	827	92.435	14	7.517	-0.473	19	130.7	259.8	-	Other	nucl
Gh_D01G0196	206	23.297	13	10.798	-0.618	5	124.2	100.8	-	Other	nucl
Gh_D01G0197	228	25.286	-6	5.286	-0.479	5	137.4	527.3	-	Other	chlo
Gh_D01G0198	181	19.829	-17.5	4.236	-0.926	2	273	235	C	Other	chlo
Gh_D01G0199	130	14.537	0	6.475	0.075	3	131	1 274.50	-	SP	extr
Gh_D01G0200	470	51.491	-4.5	5.821	-0.604	4	353.3	336.3	-	Other	nucl
Gh_D01G0201	362	39.258	-1.5	5.891	-0.043	14	77.8	227.1	M	mTP	mito
Gh_D01G0202	266	29.758	11	9.845	0.62	7	114.4	429.2	S	SP	chlo
Gh_D01G0203	463	50.97	3.5	6.989	-0.594	5	278.4	724	C	Other	nucl
Gh_D01G0204	463	51.217	5.5	7.863	-0.613	5	278.4	788.3	C	Other	nucl
Gh_D01G0205	481	53.146	-4.5	6.102	-0.05	10	144.6	377.6	-	Other	cyto
Gh_D01G0206	259	28.11	13	10.307	-0.397	7	111.4	302.8	C	cTP	chlo
Gh_D01G0207	450	50.452	2	6.841	-0.286	7	193.3	220.3	-	Other	cyto
Gh_D01G0208	598	67.75	5	7.185	-0.373	10	179.7	284.9	C	cTP	chlo
Gh_D01G0209	268	30.766	-13	4.653	-0.491	1	807	No intron	-	SP	E.R.
Gh_D01G0210	702	75.521	0	6.545	-0.06	4	527.3	196.3	C	cTP	chlo
Gh_D01G0211	751	84.485	3	6.889	-0.144	9	250.7	499.4	C	cTP	chlo
Gh_D01G0212	521	57.652	0.5	6.631	-0.122	9	174	404.9	S	SP	chlo
Gh_D01G0213	372	42.957	5.5	7.368	-0.469	6	186.5	1 091.20	M	mTP	nucl
Gh_D01G0214	691	76.102	3.5	6.837	-0.005	20	103.8	697.4	-	Other	extr
Gh_D01G0215	697	76.446	11.5	8.01	-0.082	23	91	278.4	-	Other	chlo
Gh_D01G0216	796	87.405	12	8.153	0.19	24	99.6	190.6	-	SP	plas
Gh_D01G0217	366	40.714	12.5	9.663	-0.562	11	100.1	316.3	-	Other	nucl
Gh_D01G0218	118	13.346	0.5	6.672	-0.161	3	119	768.5	-	Other	cyto
Gh_D01G0219	159	18.301	2.5	7.734	-0.451	1	480	No intron	-	Other	cyto
Gh_D01G0220	329	37.446	9	9.026	-0.553	8	123.8	344.4	-	Other	nucl
Gh_D01G0221	73	8.004	-1.5	5.584	0.064	2	111	133	-	Other	cyto
Gh_D01G0222	212	23.909	8	9.006	-0.018	4	159.8	425.7	S	SP	nucl
Gh_D01G0223	251	27.12	-1.5	6.022	-0.213	4	189	413.3	-	Other	chlo
Gh_D01G0224	445	50.884	13	9.42	-0.269	6	223	436	-	SP	chlo
Gh_D01G0225	386	42.709	-13	4.668	0.091	11	105.5	309.3	-	Other	cyto
Gh_D01G0226	1 076	119.125	-6	6.006	-0.647	13	248.5	321.1	-	Other	nucl
Gh_D01G0227	162	18.032	10	10.114	0.358	5	97.8	275	S	SP	plas
Gh_D01G0228	239	26.313	5	9.003	0.04	1	720	No intron	-	Other	chlo
Gh_D01G0229	239	26.218	6.5	9.285	-0.033	1	720	No intron	-	Other	mito
Gh_D01G0230	207	23.198	11.5	9.683	-0.968	1	624	No intron	-	Other	nucl
Gh_D01G0231	420	47.488	9.5	8.78	-0.308	8	157.9	368.3	C	cTP	chlo
Gh_D01G0232	420	47.471	11.5	8.886	-0.304	8	157.9	886.3	C	cTP	chlo

**Table 4** Physiochemical properties and sub cellular localization prediction of the mined genes within the major clusters of the consistent QTLs (*Continued*)

Gene ID	Protein Length / aa	Molecular Weight / kDa	Charge	pI	GRAVY	Exon Number	Mean Exon Length / bp	Mean Intron Length / bp	Sub cellular localization		
									TargetP	pprowler	Wolfpsort
Gh_D01G0233	1 035	112.771	31.5	8.877	-0.45	9	345.3	284.6	-	Other	nucl
Gh_D01G0234	134	15.462	14	10.864	-0.41	5	81	296.3	-	Other	cyto
Gh_D01G0235	873	98.951	-0.5	6.475	-0.233	24	109.3	252.7	S	Other	nucl
Gh_D01G0236	1 039	116.109	11.5	7.216	-0.274	5	624	276.5	-	mTP	nucl
Gh_D01G0237	194	20.965	3.5	8.967	-1.206	4	146.3	518.3	-	Other	nucl

aa Amino acid, kDa KiloDalton, bp Base pair, *chlo* Chloroplast, *cyto* Cytoplasm, *extr* Extracellular part of the cell, *nucl* Nucleus, *mito* Mitochondrion, *cysk* Cytoskeleton, *golg* Golgi body, *vacu* Vacuole, *plas* Plasma membrane, *ER* Endoplasmic reticulum, *SP* Secretory pathway (presence of a signal peptide), *mTP* Mitochondrial targeting peptide, *cTP* Chloroplast transit peptide, Other (nucleus, cytoplasmic, or otherwise), C Cytoplasm, S Secretory pathway, M Mitochondrion and -: others/other cell organelles

to have overlapping roles. In the maintenance of SLW and CMS, the ortholog genes were *Gh\_D01G0179* (SLW)-*Gh\_A05G3286* (CMS), *Gh\_A01G1948* (SLW)-*Gh\_A05G3284* (CMS) and the third pair was *Gh\_D01G0219* (SLW)-*Gh\_A05G2520* (CMS). A pair of ortholog gene pairs was also detected in groups 3 and 4, with similar traits attributes (Fig. 4 and Additional file 13).

#### miRNA target and cis- regulatory element analysis of the mined genes

The small RNAs (miRNAs) regulate gene expression via translational inhibition and have been highly correlated to abiotic stress tolerance in plants (Sunkar et al. 2007). We analyzed the mined genes to determine the possible miRNA targets, 36 genes were found to be targeted by 75 miRNAs. The miRNAs targeted the genes by either translation or cleavage (Rhoades et al. 2002). In all the miRNAs detected, 36 miRNAs targeted various genes through cleavage and 39 miRNAs targeted genes via translation. The highest level of targeting was observed for the following genes: *Gh\_A01G1939* (targeted by 5 miRNAs), *Gh\_D01G0190* (4 miRNAs), *Gh\_D01G0208* (4 miRNAs), *Gh\_D01G0210* (4 miRNAs), *Gh\_D01G0223* (5 miRNAs) and *Gh\_D01G0235* (4 miRNAs). The rest of the genes were targeted by 1 to 3 miRNAs. In relation to miRNAs, ghr-miR2949a-3p was the only miRNA that targeted two genes, *Gh\_D01G0190* and *Gh\_D01G0233* (Additional file 14: Table S10). Among the miRNAs targeting the various genes, ghr-miR156a, ghr-miR156b and ghr-miR156d targeted *Gh\_A05G3285*, and ghr-miR156c targeted *Gh\_D01G0187*. The same miRNAs have been investigated intensively and have been found to confer drought and salt stress tolerance in cotton (Xie et al. 2015). Similarly, ghr-miR166b targeted *Gh\_A01G1943*, ghr-miR172 targeted *Gh\_D01G0210*, ghr-miR396a and ghr-miR396b targeted *Gh\_A01G1939*. Two miRNAs, ghr-miR156 and ghr-miR396, targeted the NAC, MYB, and MAPK families, the top ranked promoters related to drought and salt stress (Xie et al. 2015). This provided a strong indication of the vital roles

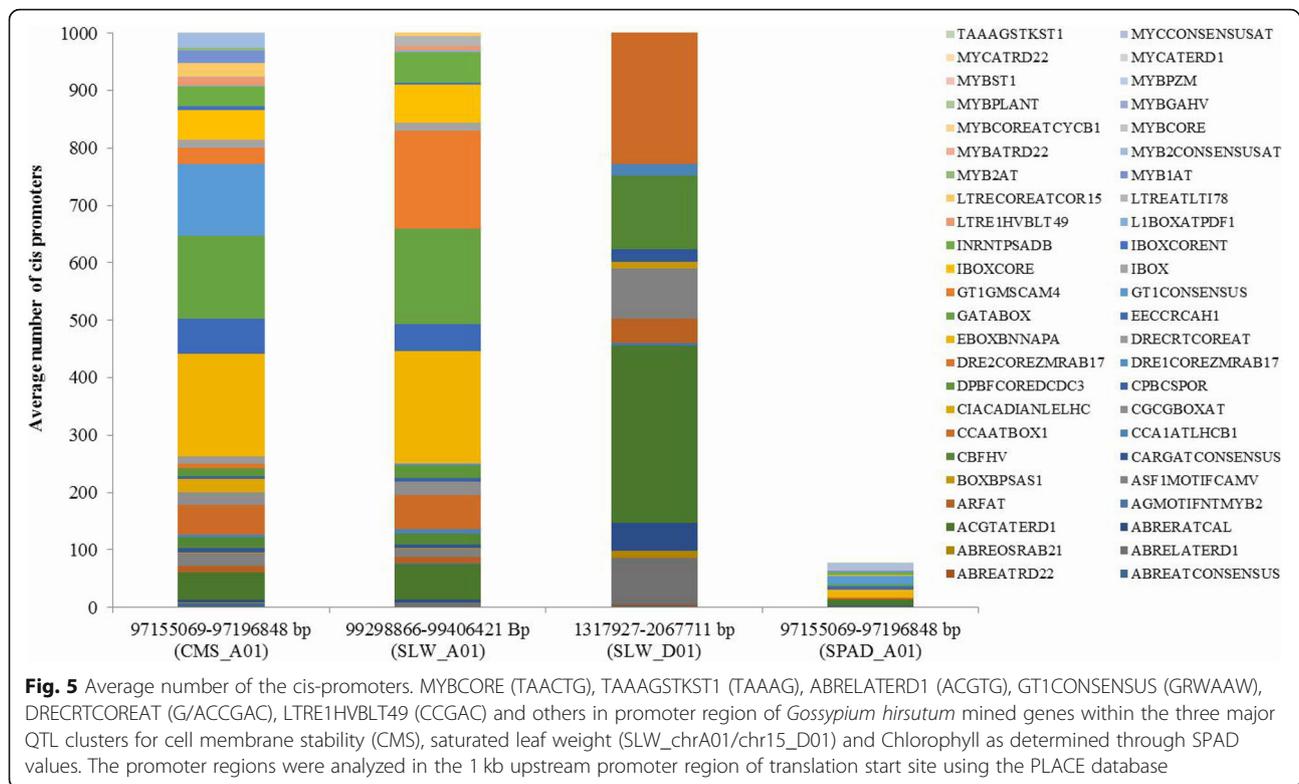
played by these genes in plants under the drought stress. It is interesting that *Gh\_D01G0210* exhibited significant up regulation as per the RNA sequence expression profile under salt and drought stress conditions.

*Cis* elements such as NAC, ABRE, MYB have been strongly associated with various abiotic stress factors in plants (Nakashima et al. 2014). In all the candidate genes identified within the QTL regions, we were able to detect various transcriptomes with direct roles in abiotic stress tolerance in plants. For example, the following categories of Myb related transcriptome factors were detected: MYB1AT (WAACCA); MYB2AT (TAACTG); MYB2CONSENSUSAT (YAACKG); MYBATRD22 (CTAACCA) and MYBCORE (CNGTTR), all with the dominant role of being responsive to dehydration and or being induced by a water deficit condition (Fig. 5). This provided stronger evidence of the possibility of the mined genes being involved in various physiological and or biological processes within the plants, aimed at reducing the effects of drought stress, thus enhancing their ability to tolerate drought stress and sustain their productivity under the stress condition.

#### RNA Seq. expression analysis of the mined genes under drought and salt stress conditions

We undertook to investigate if the identified candidate genes within the QTL regions had any functional connotation towards enhancing drought stress tolerance in upland cotton. We downloaded the RNA sequence data profiled for the roots, leaves, stem, calyx and petal from the cotton functional genome database (<https://cottonfgd.org/analyze/>) to determine the distribution and expression levels of the mined genes in various tissues. The RNA sequence data obtained were then transformed into log 10. The expression pattern of all the mined genes based on the heatmap analysis, were clustered into three groups. The RNA sequence data used for drought stress were profiled at 0 h, 1 h, 3 h, 6 h and 12 h of stress exposure. Group 1 members were significantly highly up-regulated, with 12 genes under drought



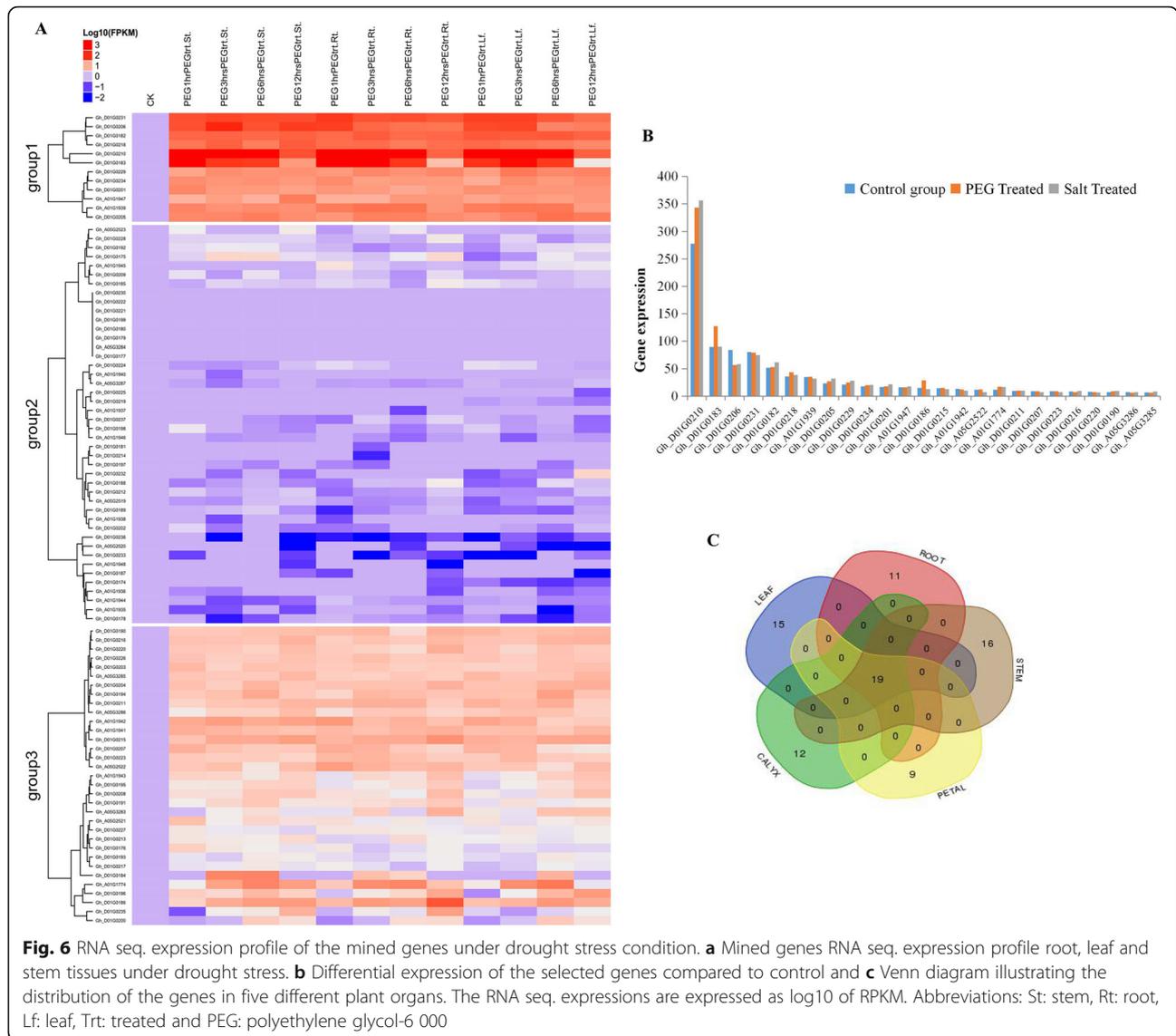


interest that genes that were highly up-regulated under drought stress, such as *Gh\_D01G0210* (ATP-dependent zinc metalloprotease FTSH 2, chloroplastic), *Gh\_D01G0183* (transcription activator GLK1), *Gh\_D01G0182* (26S proteasome non-ATPase regulatory subunit 2 homolog A), *Gh\_D01G0218*, *Gh\_D01G0205* (5'-deoxyadenosine deaminase), *Gh\_D01G0229* (enoyl-CoA delta isomerase 3) and *Gh\_D01G0234* (60S ribosomal protein L14-2), were also up-regulated under salt stress. Genes such as *Gh\_D01G0210* (ATP-dependent zinc metalloprotease FTSH 2, chloroplastic) play a vital role in the chloroplast, and chloroplast proteome changes has been found to confer drought stress tolerance in plants (Watson et al. 2018). By constructing the Venn diagram for the RNA seq data obtained for the five main tissues, root, leaf, calyx, petal and stem, 19 genes were found to exhibit common expression pattern among the five organs. The proportions of genes predominant to organ specificity were as follows: root (11 genes), leaf (15 genes), stem (16 genes), calyx (12 genes) and petal (9 genes). The leaf and the stem were the organs with the highest expression of the mined genes as compared with other tissues (Fig. 6c).

#### RT-qPCR validation of the key genes by use of the CT method

We carried out the validation of the highly expressed genes determined from the RNA seq. data. This was to confirm the expression levels of these genes on three

vital plant tissues, the leaf, root and stem, of the two parental lines used in this study. We used 15 genes out of 89 candidate genes obtained from the QTL regions. The selection of the genes was based on the RNA sequence data, type of QTLs from which the genes were obtained and phylogenetic tree analysis. The RT-qPCR results revealed three clusters of gene expression patterns. The Cluster 1 were highly up-regulated in the tissues of the two cotton species tested. Among Cluster 1 were *Gh\_D01G0182*, *Gh\_D01G0218*, *Gh\_D01G0183*, *Gh\_D01G0205* and *Gh\_A01G1774*. The genes in Cluster 2 were mainly down-regulated, except *Gh\_A01G1944* and *Gh\_A05G2521*, which were up-regulated on the leaf tissues of the tolerant donor parental line, *G. tomentosum*. The genes in Cluster 3 showed differential expression, with a high number being induced in various tissues of the tolerant parent, *G. tomentosum*, compared with the recurrent parental line, *G. hirsutum* (Fig. 7a). The level of gene induction between the two parental lines showed significant variation. A high number of genes were up-regulated in various tissues of the tolerant donor parent, *G. tomentosum* (Fig. 7b), indicating that the tolerant cultivars have increased capacity to mobilize genes under stress conditions. Expression levels of the various genes in the tissues of susceptible recurrent parent, *G. hirsutum*, were generally low compared with the tolerant donor parent (Fig. 7c). However, in both cases, a majority of the highly up-regulated genes were those

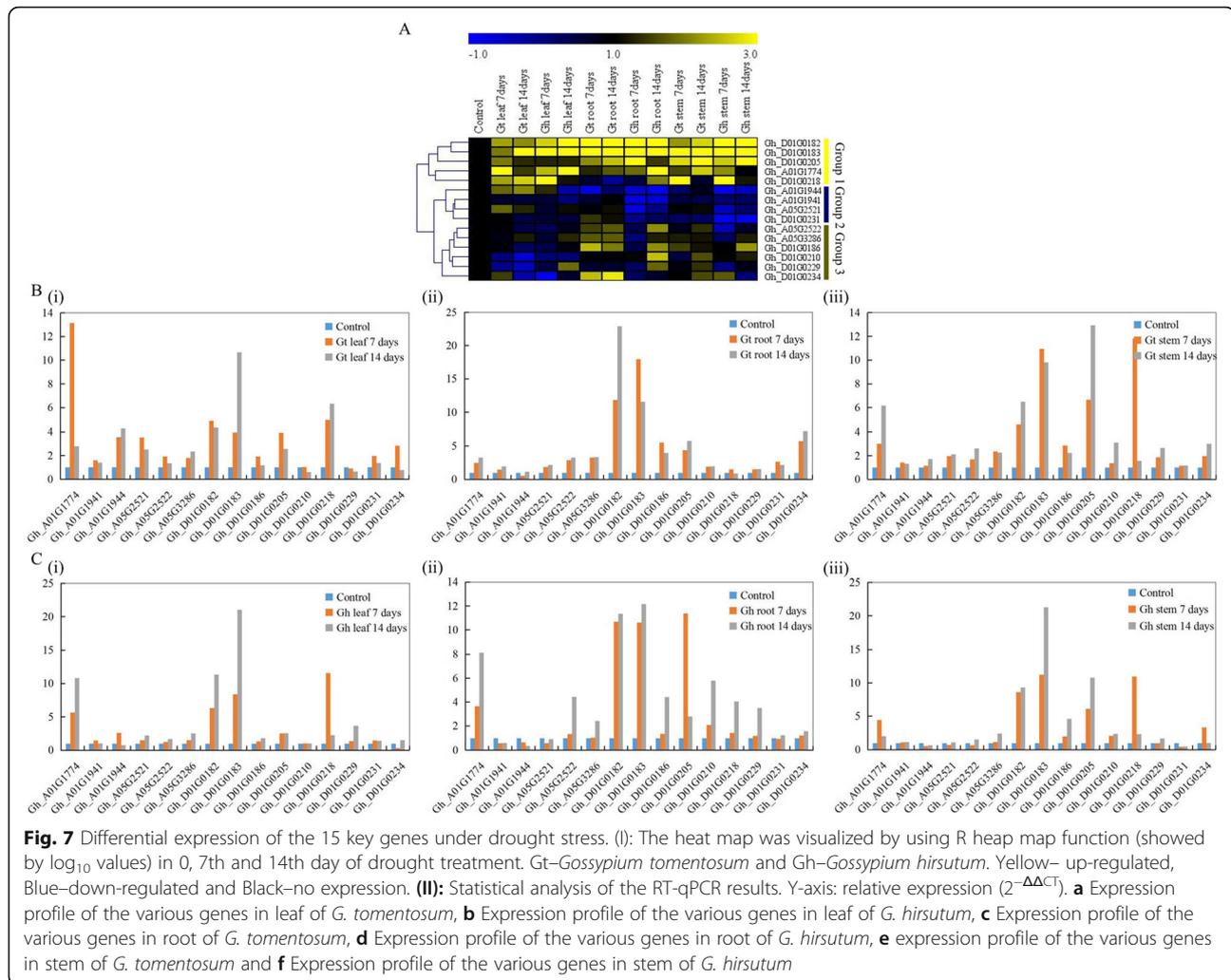


of the D-type. The genes derived from the Dt sub-genome were significantly up-regulated as opposed to those obtained from At sub-genome. The upregulation of these genes in *G. tomentosum* as opposed to *G. hirsutum* at 14 days of stress exposure indicated that the tolerant genotypes had the ability to induce more stress related genes than stress susceptible cultivars. Similar results have also been obtained in the expression profiling of two maize cultivars in which the drought tolerant genotypes were found to induce more genes than the susceptible cultivar when exposed to drought stress (Hayano-Kanashiro et al. 2009). The results obtained reaffirm the significant contribution of the D-genome in the development of tetraploid cotton, vital genes with profound functional role on fiber, abiotic and biotic stress tolerance have been found to be harboured in the Dt sub-genome as opposed to At sub-genome (Wang et al.

2014; Zhou et al. 2014). The following five genes were found to be the putative key genes with a positive net effect on enhancing drought tolerance in cotton: *Gh\_D01G0182*, *Gh\_D01G0218*, *Gh\_D01G0183*, *Gh\_D01G0205* and *Gh\_A01G1774*. These five genes could be further exploited for the development of more drought and salt resilient cotton genotypes.

### Discussion

Drought stress poses a serious threat to the normal growth and development of crops and in many cases leads to plant death, resulting in a total loss of yield in agricultural crops (Nakashima et al. 2014). Cotton is an important crop and indispensable source of raw material for the textile industries; however, its production over the years has been in steady decline, due to various environmental stress factors (Dabbert and Gore 2014).



Cotton is generally partially tolerant to various environmental stresses, though it is highly susceptible at the seedling, flowering and boll formation stages, which affects stand establishment and overall production if boll abortion at the boll formation stages (Wang et al. 2016). Improvement of cotton cultivar performance under abiotic stress conditions has been a challenge owing to its narrow genetic base, which is the result of intensive selection, inbreeding and incompatibility between various genotypes (Kottapalli et al. 2016). To broaden the narrow genetic base of elite cotton cultivars such as the commonly grown upland cotton, *G. hirsutum*, the use of the wild progenitors has been explored and has led to generating new genotypes with improved performance under various environmental stresses (Pushpam and Raveendran 2006). In this research, we explored backcross inbred lines; BC<sub>2</sub>F<sub>2</sub> generations were developed from *G. tomentosum* and *G. hirsutum* to map QTLs related to drought tolerance traits and explore any genes within the QTL regions possibly related to drought stress.

In the evaluation of the phenotypic traits under the drought stress condition, all the physiological and morphological traits showed significant reduction compared with the traits measured under a controlled environment (a well watered condition). From visual observation, the most notable morphological feature among the BC<sub>2</sub>F<sub>2</sub> populations under the water stress condition was a decrease in plant height. The reduction in plant height could be attributed to shortened stem length. The results obtained are in agreement with a previous finding in which water deficit was reported to have a negative effect on plant growth and development. For example, in maize, growth is greatly affected by declining soil moisture content resulting in either a decreased growth rate or the plant becoming stunted in growth (Hsiao et al. 1970). Moreover, in soybean, stem length elongation is affected under drought stress; soybean plants exposed to declining moisture conditions have a reduced plant height compared with non stressed plants (Specht et al. 2001). The effect of water deficit on plant growth has also been

observed in okra (*Abelmoschus esculentus* (L.) Moench), in which its height was significantly reduced due to increased leaf senescence and cessation in cell elongation after exposure to a drought stress condition (Bhatt and Srinivasa 2005). Plant growth inhibition during drought exposure is primarily due to a loss of turgor arising from a lack of water availability (Farooq et al. 2011). Plant growth rates are reduced more rapidly than photosynthetic activity under drought conditions, implying that plants actively reduce growth in response to drought stress (Todaka et al. 2015). The reduction in plant height is primarily attributed to a reduction in two main cell cycle processes, cell expansion and elongation (Mantovani and Iglesias 2008).

Correlation analysis aids in understanding of overall contribution of various plant traits on each other (Gibert et al. 2016). Excised leaf water loss (ELWL) had a negative correlation with cell membrane stability (CMS). The results obtained in the correlation analysis are consistent with previous findings. For example, shoot fresh weight (SFW) and shoot dry weight (SDW) were found to be highly correlated under the abiotic stress condition. Similarly, root fresh weight (RFW) has been reported to be highly correlated to root dry weight (RDW) (Li et al. 2005). Relative leaf water content (RLWC) correlated positively with cell membrane stability (CMS). RLWC is a measure of plant water status in a given environment and is correlated with drought stress tolerance and yield in crop plants (Almeselmani et al. 2011; Lugojan and Ciulca 2011). Relative leaf water content (RLWC) has a direct effect on cellular membrane integrity. Loss of leaf turgor causes dehydration in cells and eventually cell membrane damage. In this study, a positive correlation between RLWC and CMS indicates that the plant with a higher water content may maintain cellular membrane integrity under drought stress. CMS is linked to drought stress tolerance and yield in plants (Almeselmani et al. 2011). And therefore, it is an important trait in the evaluation of plants in relation to drought tolerance (Rahman et al. 2008). Furthermore, the negative correlation between excised leaf water loss (ELWL) with CMS indicates that lower water loss from leaves help maintain relative water content and hence cell membrane stability. The dynamics of water balance in plant tissues regulates turgor pressure and directly affects the extensibility of the cell wall (Marshall and Dumbroff 1999).

The stomata plays a critical role in plant water relationships (Buckley 2005). The rate of water loss is highly correlated with the number, location and size of the stomatal pore (Drake et al. 2013). Through microscopic examination of the abaxial (lower leaf surface) and the adaxial (upper leaf surface) regions of the leaf surface of the two parental lines, together with their BC<sub>2</sub>F<sub>1</sub>

generation, significant variation was detected in both number and size of the stomatal pores. The tolerant cultivar, *G. tomentosum*, had fewer stomata on either side of the leaf with a relatively reduced stomatal pore compared with the drought susceptible cultivar, *G. hirsutum*. The reduction in stomatal size and number on exposure to drought is an adaptive feature to enhance plant survival under drought stress. Several studies have reported a significant reduction in stomatal number in plants under a drought stress condition. For example, a study conducted on perennial grass species in relation to a varying soil moisture condition showed that the stomatal number correlated positively with soil moisture content levels (Xu and Zhou 2008). A similar finding has also been observed in rice (Karaba et al. 2007).

Genetic map is a vital tool in the exploration of the plant genome, and it provides vital information on the level of allele introgression during breeding periods (De Sousa et al. 2015). The most recent linkage map developed from F<sub>2:3</sub> generation derived from *G. hirsutum* and *G. tomentosum* was done by Zheng et al. (2016). They used simple sequence repeat (SSR) markers in which they generated a map size of 3 328.2 cM, with 1 295 markers which amplified 1 342 loci. The map had an average marker distance of 2.57 cM; the average distance was relatively high and therefore not precise for providing valid results on gene action or yield predicted with drought related QTLs. In this study, we employed genotyping by sequence (GBS) to generate the SNPs. A total of 10 888 SNPs were used in the development of the genetic map with a map size of 4 191.3 cM and an average marker distance of 0.1047 cM. This was 25% reduction between two flanking markers compared with the 2.57 cM previously obtained by Zheng et al. (2016). The map we developed allowed identification of QTLs with higher resolution than what was obtained in earlier reports. Thus, the detected QTLs are reliable and true to type for future application in breeding for drought tolerance in cotton.

Trait introgression from parental lines to their offspring's is governed by the level of heritability. When the heritability percentage of a trait is high, manipulation become easy. Various traits exhibit high heritability percentages, ranging from 62.5%, as detected for cell membrane stability (CMS), to a maximum of 95.9%, as observed for plant height (PH). Higher heritability percentages show that the traits are easy to manipulate and are inheritable. Similar results have been observed in a number of studies in cotton genotypes under abiotic stress conditions (Oluoch et al. 2016). Low heritability could be due to environmental influence; thus, high heritability is highly recommended for trait-based selection in relation to abiotic stress tolerance (Würschum 2012). In this research, 30 QTLs were consistent for 11 traits:

DSB, SLW, SPAD, TDB, FSB/FRB, DLW, CMS, FSB, LFW, TFB and DSB/DRB, with a range of broad sense heritability between 62.5% to a maximum of 84.4%, which explained the phenotypic variation of 0 to 75.8%. The QTLs were mapped asymmetrically within the two sub-genomes of the tetraploid cotton (AD) genome, 17 and 13 QTLs located in At and Dt sub-genomes, respectively. The results were in agreement with a previous study, which showed that stable QTLs were detected in both At and Dt sub-genomes (Zheng et al. 2016). The contribution of Dt sub-genome towards abiotic tolerance has been widely investigated. A high number of QTLs related to salt stress has been mapped in the Dt sub-genome as opposed to the At sub-genome. For example, Oluoch et al. (2016) found 11 significant QTLs located in the Dt sub-genome while only a single QTL was located in the At sub-genome.

Genes have a greater influence on various phenotypic traits of the plants under abiotic stress exposure (Omholt et al. 2000). There are four types of gene actions: additive effect (Ae), dominant effect (De), partial dominance (PD) and over dominance (OD) (Omholt et al. 2000). In this study, all four gene actions were observed. A majority of the QTLs exhibited over dominance gene action, regulating 10 QTLs. This was closely followed by the dominance effect with 8 QTLs, then partial dominance with 7 QTLs. The additive effect regulated only with 5 QTLs. The results obtained were not in agreement with the previous finding by Oluoch et al. (2016), in which the partial dominant effect was found to be higher than the other gene actions. QTL determination alone is not sufficient without deeper insight into the various genes deemed to be controlling the trait mapped. The 30 stable QTLs were grouped into 16 clusters. Cluster 1 had 4 QTLs; cluster 2, 3, 5, 7, 8, 9, 11, 12, 13, 15 and 16 each had a single QTL; cluster 4 had 5 QTLs; cluster 6 had 2 QTLs; cluster 10 had 6 QTLs and cluster 14 had 2 QTLs. Based on the number of QTLs per cluster, clusters 1, 4 and 10 were the major QTL clusters, with more than 2 QTLs in each. Because the two parental lines were phenotypically varied, *G. hirsutum* is superior in most phenotypic traits compared with the donor parent, *G. tomentosum*. We therefore, undertook to identify the candidate genes for physiological traits from the QTL regions contributed by the donor parent. Three traits were considered, cell membrane stability (CMS), saturated leaf weight (SLW) and chlorophyll content as measured by SPAD values. The CMS is a main cellular target common to different stresses, and the CMS has been extensively used as a selection criterion for different abiotic stresses, including drought and high temperature in wheat (Ciulca et al. 2017). The obtained results in the present study indicated that *G. tomentosum* had higher cell membrane stability compared

with the recurrent parent, *G. hirsutum*, as was evident through the ion leakage concentration. Several investigators have reported that differences in the CMS might result from differences in leaf structure (Kocheva et al. 2014), cell wall composition (Marcia 2009) and the degree of membrane lipid saturation (Kumar 2012). Thus, determining of candidate genes within these QTL regions was important for determining vital genes responsible for drought stress tolerance in the wild cotton progenitor, *G. tomentosum*, which was used as the donor parent.

The identified candidate genes within the QTL regions regulating CMS, ELWL and SLW were further analyzed to elucidate their roles in enhancing drought stress tolerance in cotton. Based on phylogenetic analysis, all the candidate genes were allocated to four groups, in which some orthologous gene pairs were obtained from QTL regions regulating different traits, such as *Gh\_D01G0223\_TSJT1* (stem-specific protein TSJT1) and *Gh\_A01G1774* (GDSL esterase/lipase) obtained from QTL regions controlling SLW and CMS, respectively. The stem-specific protein TSJT1 have a profound role in enhancing drought stress tolerance in rice. TSJT1 was up-regulated four-fold across all tissues and stages under drought stress conditions (Sircar and Parekh 2015). The detection of the *TSJT1* gene among the identified candidate genes showed that the ortholog pair could perform a similar function in enhancing drought stress tolerance in cotton. Other ortholog pairs obtained from different QTL regions controlling different traits were, for example, *Gh\_D01G0201* (pyruvate dehydrogenase E1 component subunit beta-1, mitochondrial) and *Gh\_A05G3285* (probable receptor-like protein kinase At2g42960); *Gh\_D01G0179* (non-specific lipid-transfer protein 13) and *Gh\_A05G3286* (protein NLP5); *Gh\_A01G1948* (laccase-4) and *Gh\_A05G3284*. A majority of the orthologous genes were found to belong to the same functional domain. For example, *Gh\_D01G0228 (EC13)* and *Gh\_D01G0229 (EC13)* were associated with a functional description of Enoyl-CoA delta isomerase 3, an enzyme that functions in fatty acid degradation (Voldina and Steinbüchel 2014). Fatty acid is integral in the process of fiber formation in cotton (Qin et al. 2007). The detection of this protein, especially with its present within the QTL region controlling saturated leaf weight, possibly means that the gene has a multifunctional role within the plant because unsaturated fatty acids such as oleic acid have a regulatory role in water uptake in barley (Cozzolino et al. 2014).

Analysis of physiochemical properties of the mined genes was critical to determine various aspects such as the molecular weight (aa), GRAVY values, PI values and charge because these properties are important in determining the possible roles of the various genes. A majority of the genes had negative GRAVY values with high

charge, indicating that most of the genes were hydrophobic in nature, a property shared among most of the stress related genes, such as *LEA* genes (Hand et al. 2011). Hydrophobicity enables drought related proteins to be tolerant to desiccation. Gene structural analysis revealed that most of the genes were disrupted by introns. Introns place a great burden on genes because they require a spliceosome, which is among the largest molecular complexes in the cell (Wahl et al. 2009). However, the intron:exon ratios were relatively low. Some genes had no intron disruption, such as *Gh\_D01G0209*, *Gh\_A05G2521*, *Gh\_A01G1946*, *Gh\_D01G0189*, *Gh\_D01G0228*, *Gh\_D01G0229*, *Gh\_D01G0219*, *Gh\_D01G0177*, *Gh\_A01G1944*, *Gh\_D01G0180* and *Gh\_D01G0181*. Some of these genes were highly up-regulated in the analysis of RNA sequence data under the salt and drought stress condition.

Large numbers of proteins encoding the candidate genes identified within the QTL regions were found to be embedded in the chloroplast, nucleus and cytoplasm. Chloroplastic membranes and their membrane bound structures are very vulnerable to oxidative stress because large quantities of reactive oxygen species (ROS) can be released from these membranes when plants are exposed to a stress condition. ROS cause an extensive de-esterification and peroxidation of membrane lipids, as well as protein denaturation and/or DNA mutations (Bowler et al. 1992). The delicate balance of ROS release and detoxification is always affected when plants are exposed to drought stress conditions. The continuous elimination of ROS inhibits oxidative damage, thus enabling plants to maintain various physiological and biochemical pathways uninterrupted. The presence of these proteins encoding the candidate genes could be linked to the induction of various antioxidant enzymes such as peroxidase (POD) and superoxide dismutase (SOD). High concentrations of antioxidants have a regulatory role in maintaining the ROS levels within a threshold tolerable by plants. A number of genes have a regulatory role in mobilizing the antioxidant enzymes, such as *LEA2* genes (Magwanga et al. 2018c). Intense drought stress leads to massive water loss, resulting in intense plasmolysis of the tonoplast. This causes an increased concentration of cellular solutes, which possibly reach a toxic threshold for certain proteins and or enzymes (Cruz de Carvalho 2008), thereby intensifying the detrimental effects on the photosynthetic machinery, the cytosol and other organelles, thus affecting the membrane stability and its integrity. In this study, we found critical genes with profound roles in cell structural integrity as revealed through gene ontology (GO) analysis. For example, *Gh\_A01G1940* was found to be involved in the cellular component, specifically on the integral component of membranes (GO: 0016021 and GO: 0016020); *Gh\_A01G1943* was involved in the nucleus

(GO: 0005634); and *Gh\_A01G1948* had a functional role in the apoplast (GO: 0048046). These genes were mined within the QTL region controlling saturated leaf weight (SLW). The flow of water in and out of the cell is governed by a passive process but the membrane is significant; any damage offsets the osmolytes, and thus causing an excessive leakage of ions out of the cell (Cooper 2000).

Furthermore, when plants are under stress, the release of ROS is accelerated; thus, the faster elimination is a survival strategy of the plant. Within the three major QTL cluster regions, we found vital genes involved in the process of ubiquitination (*Gh\_D01G0188*). Ubiquitination is a biological process that has been found to aid plant tolerance to various abiotic stresses. Ubiquitination and phosphorylation sites regulate ROS (Liu and Min 2002). The detection of genes linked to the ubiquitination mechanism provides an indication of the introgression of drought tolerance traits from the donor drought resistant parent to the segregating backcross in bred lines.

The roots are the main organ in the uptake of water from the soil or other water reservoirs. When drought occurs, the root becomes the first organ to be affected (Robbins and Dinneny 2015). More genes are expected to be highly up-regulated in the roots compared with other plant organs. However, in these groups of genes, more were up-regulated at the stem regions, with 16 genes, and in the leaf and root, 15 and 11 genes were observed, respectively. This indicated that, the plant organs work in a synchronized manner when drought occurs to increase the plant's ability to tolerate the drought effect for a relatively long period. The high number of up-regulated genes in the leaf could be responsible for maintaining stomatal conductance and rapid elimination of the reactive oxygen species being released from the cells (Hardy et al. 1995).

In relation to the RNA sequence expression profile, 15 genes were highly up regulated under drought and salt stress conditions: *Gh\_D01G0234*, *Gh\_D01G0231*, *Gh\_D01G0201*, *Gh\_A01G1947*, *Gh\_D01G0215*, *Gh\_A01G1774*, *Gh\_D01G0205*, *Gh\_D01G0229*, *Gh\_A01G1939*, *Gh\_D01G0186*, *Gh\_D01G0182*, *Gh\_D01G0218*, *Gh\_D01G0206*, *Gh\_D01G0210* and *Gh\_D01G0183*. It is interesting that among the highly up regulated genes, some were also targeted by various miRNAs. For example, *Gh\_D01G0234* was targeted by ghr-miR7497, *Gh\_D01G0205* was targeted by ghr-miR164 and ghr-miR2948-5p, *Gh\_D01G0229* was targeted by ghr-miR394a and ghr-miR394b, *Gh\_D01G0186* was targeted by ghr-miR399e, and *Gh\_D01G0182* was targeted by ghr-miR7499. A number of miRNAs have a functional role under stress in various plants, including drought stress. For example, miR394a/b, which targets *Gh\_D01G0229*, is a conserved and versatile miRNA with

multiple functional roles under various abiotic stresses (Huang et al. 2010). The presence of miR394a/b has been reported in a number of plants, such as *Arabidopsis thaliana* (Jones-Rhoades and Bartel 2004), *Oryza sativa* (Zhang et al. 2007) and *Brassica napus* (Zhao et al. 2012). Therefore, genes targeted by miR394a/b could have a direct functional role in enhancing drought and salt stress in upland cotton.

The expression pattern of genes in various tissues at varying stress exposure provides important information on the functional correlation of the genes to the stress factor under investigation (Shinozaki and Yamaguchi-Shinozaki 2007). We analyzed 15 genes through RT-qPCR analysis on root, leaf and stem tissue samples obtained from the two cotton genotypes grown under the drought stress condition. More genes were highly up-regulated on the various tissues of *G. tomentosum* than *G. hirsutum*. The higher up-regulation of genes in the tolerant parental line, *G. tomentosum*, showed that the tolerant genotype had the ability to induce more stress related genes when exposed to the drought condition, thereby increasing its tolerance level. Similar findings have been observed between two maize genotypes in which more genes were up-regulated in the tolerant genotypes under a drought condition compared with less tolerant genotypes (Hayano-Kanashiro et al. 2009). Out of 15 genes, 5 were found to be putative key genes. This conclusion was informed by the results obtained from RNA seq expression analysis, and miRNA target and RT-qPCR validation. These genes can be explored and be used in breeding of cotton genotypes with improved drought tolerance.

## Conclusions

We developed a semi-wild segregating backcross inbred line (BC<sub>2</sub>F<sub>2</sub>) from two tetraploid cotton species, an elite cultivated *G. hirsutum* and its wild progenitor *G. tomentosum*. The population was successfully genotyped through the GBS approach and the map generated is the finest genetic map developed from an interspecific cross to date. The map size was 4 191.3 cM, with an average marker distance of 0.1047 cM. The maps developed allowed us to identify 30 consistent QTLs with higher precision than what was possible in earlier studies. Thus, the QTLs detected are reliable and true to type for future application in breeding for drought tolerance in cotton. Within the major QTL clusters, we mined 89 genes belonging to different gene families. The genes were analyzed and their physiochemical properties showed that they were involved in diverse cellular, molecular and biological processes, as evident through gene ontology results. Expression profiling in various tissues suggested that the mined genes were highly active in modulating cotton growth and development under

drought and salt stress conditions. Moreover, by integrating RNA-seq data and RT-qPCR analysis, we were able to determine five putative candidate genes, which could be of significance in the regulatory response to drought and salt stress tolerance in cotton. The findings of this research provide fundamental steps for future exploration of the identified candidate genes within the QTL regions to understand their specific roles in enhancing abiotic stress tolerance in cotton. In addition, the cotton breeders could use of the key QTLs identified in this study to the development of much more drought tolerant cotton genotypes with improved performance under drought stress conditions.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s42397-020-0043-0>.

**Additional file 1: Figure S1.** Flow chart of the development of the mapping population of the BC<sub>2</sub>F<sub>2</sub> progenies from *G. hirsutum* the recurrent parent and *G. tomentosum* the donor parent.

**Additional file 2: Figure S2.** Frequency distribution of the 16 traits for physio-morphological traits of BC<sub>2</sub>F<sub>2</sub> PH-plant height, SPAD-chlorophyll content; SLW: saturated leaf weight; FSB: fresh shoot biomass; FSB/FRB: fresh shoot biomass/fresh root biomass; DRB: dry root biomass; TDB: total dry biomass; ELW: excised leaf weight; CMS: cell membrane stability; ELWL: excised leaf water loss; FLW: fresh leaf weight; DSB/DRB: dry shoot biomass/dry root biomass; TFB: total fresh biomass; RLWC: relative leaf water content; FRB: fresh root biomass and DLW: dry leaf weight.

**Additional file 3: Figure S3.** Consistent QTLs mapped onto their specific chromosomes.

**Additional file 4: Table S1.** List of primers used for the qRT-PCR validation for the mined genes within the major QTL cluster regions.

**Additional file 5: Table S2.** The BIL seedlings performance under drought and controlled conditions in two different environments.

**Additional file 6: Table S3.** Genotypic correlation coefficient among different component traits under drought stress.

**Additional file 7: Table S4.** Characteristics of the high-density genetic map.

**Additional file 8: Table S5.** Genome co-ordinates of all the SNPs with physical positions (bp), genetic position (cM) and the alleles.

**Additional file 9: Table S6.** Consistent QTLs for physio-morphological drought related traits detected in the semi wild BIL population.

**Additional file 10: Table S7.** Mined genes within 3 major QTLs cluster regions for physiological traits related to drought stress in cotton.

**Additional file 11: Table S8.** Detailed information of the sub cellular location prediction of the mined genes by the use of TargetP, Prowler and Wolfpsort online tools.

**Additional file 12: Table S9.** GO functional annotation of genes mined within the three major QTL cluster regions.

**Additional file 13.** Newick structure and protein sequences used in phylogenetic tree analysis.

**Additional file 14: Table S10.** miRNA targeting the mined genes within the three major QTL clusters.

## Abbreviations

CDS: Coding sequence; GBS: Genotyping by sequence; GO: Gene ontology; GRAVY: Grand average of hydropathy; PCV: Phenotypic coefficient of variation; QTL: Quantitative trait loci

### Acknowledgements

The authors are grateful for the help of the teachers and students of the Wild Cotton Project of the Institute of Cotton Research of Chinese Academy of Agricultural Sciences, for their support in the course of this research work.

### Authors' contributions

Magwanga RO, Wang KB and Liu F designed the experiment. Magwanga RO, Lu P and Kirungu JN implemented and collected the data. Magwanga RO analyzed the results and prepared the manuscript. Kirungu JN, Lu P, Liu F, Cai XY, Zhou ZL, Agong SG and Wang KB revised the manuscript. All authors reviewed and approved the final manuscript.

### Funding

This research program was financially sponsored by the National Natural Science Foundation of China (31671745, 31530053) and the National key research and development plan (2016YFD0100306).

### Availability of data and materials

Not applicable.

### Ethics approval and consent to participate

No ethical nor consent to participate in this research was sought. The research work was conducted as per the broad mandate of the cotton research institute (CRI), which is the state owned research institute charged with the responsibility to develop, carry out research and approve all the cotton breeding work within China.

### Consent for publication

No consent to publish the work was sort; publication is highly encouraged by Institute of Cotton Research Institute, Chinese Academy of Agricultural Sciences (ICRCAAS) as a way to disseminate significant funding's in cotton breeding science.

### Competing interests

The authors declare no any form of competing interest.

### Author details

<sup>1</sup>State Key Laboratory of Cotton Biology/Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China.  
<sup>2</sup>Jaramogi Oginga Odinga University of Science and Technology, P.O Box 210-40601, Bondo, Kenya.

Received: 14 October 2019 Accepted: 1 January 2020

Published online: 02 February 2020

### References

- Almeselmani M, Abdullah F, Hareri F, et al. Effect of drought on different physiological characters and yield component in different varieties of Syrian durum wheat. *J Agric Sci*. 2011;3:3. <https://doi.org/10.5539/jas.v3n3p127>.
- Argyrokastritis IG, Papastilianou PT, Alexandris S. Leaf water potential and crop water stress index variation for full and deficit irrigated cotton in Mediterranean conditions. *Agric Agric Sci Procedia*. 2015;4:463–70. <https://doi.org/10.1016/j.jaaspro.2015.03.054>.
- Ariani A, Berny Mier y Teran JC, Gepts P. Genome-wide identification of SNPs and copy number variation in common bean (*Phaseolus vulgaris* L.) using genotyping-by-sequencing (GBS). *Mol Breed*. 2016;36:87. <https://doi.org/10.1007/s11032-016-0512-9>.
- Ashraf M. Inducing drought tolerance in plants: recent advances. *Biotechnol Adv*. 2010;28:169–83. <https://doi.org/10.1016/j.biotechadv.2009.11.005>.
- Ali Q, Ahsan M, Qaisrani SA, et al. Genetic variability and correlation analysis for various morpho-physiological traits in maize (*Zea mays* L.) For green fodder yield. *American-Eurasian J Agric & Environ Sci*. 2015;15(7):1298–303. <https://doi.org/10.5829/idosi.aejaes.2015.15.7.12704>.
- Azhar FM, McNeilly T. The genetic basis of variation for salt tolerance in *Sorghum bicolor* (L.) Moench seedlings. *Plant Breed*. 1988;101:114–21. <https://doi.org/10.1111/j.1439-0523.1988.tb00275.x>.
- Barnabás B, Jäger K, Fehér A. The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ*. 2008;31:11–38. <https://doi.org/10.1111/j.1365-3040.2007.01727.x>.
- Barrs HD, Weatherley PE. A re-examination of the relative turgidity technique for estimating water deficits in leaves. *Aust J Biol Sci*. 1962;15:413–28. <https://doi.org/10.1071/BI9620413>.
- Bernatzky R, Tanksley SD. Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics*. 1986;112:887–98.
- Bhatt RM, Srinivasa RNK. Influence of pod load on response of okra to water stress. *Indian J Plant Physiol*. 2005;10:54–9.
- Blum A. Drought resistance, water-use efficiency, and yield potential - are they compatible, dissonant, or mutually exclusive? *Aust J Agric Res*. 2005;56:1159–68.
- Bowler C, Montagu MV, Inze D. Superoxide dismutase and stress tolerance. *Annu Rev Plant Physiol Plant Mol Biol*. 1992;43:83–116. <https://doi.org/10.1146/annurev.pp.43.060192.000503>.
- Buckley TN. The control of stomata by water balance. *New Phytol*. 2005;168:275–92.
- Chakravarthy VSK, Reddy TP, Reddy VD, Rao KV. Current status of genetic engineering in cotton (*Gossypium hirsutum* L.): an assessment. *Crit Rev Biotechnol*. 2012;8551:1–18. <https://doi.org/10.3109/07388551.2012.743502>.
- Chapagain AK, Hoekstra AY, Savenije HHG, Gautam R. The water footprint of cotton consumption: an assessment of the impact of worldwide consumption of cotton products on the water resources in the cotton producing countries. *Ecol Econ*. 2006;60:186–203. <https://doi.org/10.1016/j.ecolecon.2005.11.027>.
- Chen Y, Liu Z-H, Feng L, et al. Genome-wide functional analysis of cotton (*Gossypium hirsutum*) in response to drought. *PLoS One*. 2013;8:e80879. <https://doi.org/10.1371/journal.pone.0080879>.
- Ciulca S, Madoşa E, Ciulca A, et al. The assessment of cell membrane stability as an indicator of drought tolerance in wheat. In: International multidisciplinary scientific GeoConference surveying geology and mining ecology management, SGEM; 2017.
- Civelek M, Lusić AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*. 2014;15:34–48.
- Clarke JM, McCaig TN. Excised-leaf water retention capability as an indicator of drought resistance of Triticum genotypes. *Can J Plant Sci*. 1982;62:571–8.
- Cooper GM. Membrane structure and function. *Cell A Mol Approach*. 2000:67–82. <https://doi.org/10.1080/03050068.2013.803783>.
- Cozzolino D, Roumeliotis S, Eglinton JK. The role of total lipids and fatty acids profile on the water uptake of barley grain during steeping. *Food Chem*. 2014;151:231–5. <https://doi.org/10.1016/j.foodchem.2013.11.073>.
- Cruz de Carvalho MH. Drought stress and reactive oxygen species: production, scavenging and signaling. *Plant Signal Behav*. 2008;3:156–65. <https://doi.org/10.4161/psb.3.3.5536>.
- da Silva MJR, Tecchio MA, Domiciano S, et al. Phenology, yield and fruit quality of 'Paluma' guava tree at different pruning times. *Ciênc Agrotec*. 2016;40:317–25. <https://doi.org/10.1590/1413-70542016403033315>.
- Dabbert TA, Gore MA. Challenges and perspectives on improving heat and drought stress resilience in cotton. *The Journal of Cotton Science*. 2014;18:393–409.
- Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
- Davey J, Hohenlohe P, Etter P, et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 2011;12:499–510. <https://doi.org/10.1038/nrg3012>.
- De Sousa CC, De Figueirêdo LC, Arrais MDGM, Valente SEDS. Genetic map of cotton with molecular markers. *Am J Plant Sci*. 2015;6:289–97. <https://doi.org/10.4236/ajps.2015.62033>.
- Deschamps S, Llaca V, May GD. Genotyping-by-sequencing in plants. *Biology (Basel)*. 2012;1:460–83. <https://doi.org/10.3390/biology1030460>.
- Desmarais E, Lanneluc I, Lagnel J. Direct amplification of length polymorphisms (DALP), or how to get and characterize new genetic markers in many species. *Nucleic Acids Res*. 1998;26:1458–65. <https://doi.org/10.1093/nar/26.6.1458>.
- Drake PL, Froend RH, Franks PJ. Smaller, faster stomata: scaling of stomatal size, rate of response, and stomatal conductance. *J Exp Bot*. 2013;64:495–505. <https://doi.org/10.1093/jxb/ers347>.
- Elshire RJ, Glaubitz JC, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379>.
- Fan Y, Shabala S, Ma Y, et al. Using QTL mapping to investigate the relationships between abiotic stress tolerance (drought and salinity) and agronomic and physiological traits. *BMC Genomics*. 2015;16:43. <https://doi.org/10.1186/s12864-015-1243-8>.
- Fang DD, Jenkins JN, Deng DD, et al. Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in

- upland cotton (*Gossypium hirsutum* L.). BMC Genomics. 2014;15:397. <https://doi.org/10.1186/1471-2164-15-397>.
- Farooq M, Bramley H, Palta JA, Siddique KHM. Heat stress in wheat during reproductive and grain-filling phases. CRC Crit Rev Plant Sci. 2011;30:491–507.
- Fernández AP, Strand Å. Retrograde signaling and plant stress: plastid signals initiate cellular stress responses. Curr Opin Plant Biol. 2008;11:509–13.
- Fleury D, Jefferies S, Kuchel H, Langridge P. Genetic and genomic tools to improve drought tolerance in wheat. J Exp Bot. 2010;61:3211–22.
- Fokar M, Nguyen HT, Blum A. Heat tolerance in spring wheat. I estimating cellular thermotolerance and its heritability. Euphytica. 1998;104:1–8. <https://doi.org/10.1023/a:1018346901363>.
- Foulongne-Oriol M, Rodier A, Rousseau T, Savoie JM. Quantitative trait locus mapping of yield-related components and oligogenic control of the cap color of the button mushroom, *Agaricus bisporus*. Appl Environ Microbiol. 2012;78:2422–34. <https://doi.org/10.1128/AEM.07516-11>.
- Fryxell PA. The natural history of the cotton tribe (Malvaceae tribe Gossypieae). Econ Bot. 1982;36:126–7. <https://doi.org/10.2307/4254361>.
- Furuta T, Ashikari M, Jena KK, et al. Adapting genotyping-by-sequencing for Rice F<sub>2</sub> populations. G3: Genes/Genomes/Genetics. 2017;7:881–93. <https://doi.org/10.1534/g3.116.038190>.
- Gibert A, Gray EF, Westoby M, et al. On the link between functional traits and growth rate: meta-analysis shows effects change with plant size, as predicted. J Ecol. 2016;104:1488–503. <https://doi.org/10.1111/1365-2745.12594>.
- Dessimoz C, Škunca N. The gene ontology handbook. In: Methods in molecular biology, vol. 1446. (Dordrecht: Springer. 1984–2020. <https://doi.org/10.1007/978-1-4939-3743-1>. (Open Access)
- Hand SC, Menze MA, Toner M, et al. LEA proteins during water stress: not just for plants anymore. Annu Rev Physiol. 2011;73:115–34. <https://doi.org/10.1146/annurev-physiol-012110-142203>.
- Hardy JP, Anderson VJ, Gardner JS. Stomatal characteristics, conductance ratios, and drought-induced leaf modifications of semiarid grassland species. Am J Bot. 1995;82:1–7. <https://doi.org/10.2307/2445779>.
- Haworth M, Killi D, Materassi A, et al. Impaired stomatal control is associated with reduced photosynthetic physiology in crop species grown at elevated [CO<sub>2</sub>]. Front Plant Sci. 2016;7. <https://doi.org/10.3389/fpls.2016.01568>.
- Hayano-Kanashiro C, Calderón-Vásquez C, Ibarra-Laclette E, et al. Analysis of gene expression and physiological responses in three Mexican maize landraces under drought stress and recovery irrigation. PLoS One. 2009;4(10):e7531. <https://doi.org/10.1371/journal.pone.0007531>.
- Henley S. Principles and procedure of statistics: a biometrical approach. New York: McGraw-Hill International; 1983.
- Hoagland DR, Arnon DI. The water-culture method for growing plants without soil. Circular. California Agricultural Experiment Station; 1950: p.32.
- Hsiao TC, Acevedo E, Henderson DW. Maize leaf elongation: continuous measurements and close dependence on plant water status. Science. 1970; 168:590–1. <https://doi.org/10.1126/PP.438.1169>.
- Ma LL, Zhao YP, Wang YM, et al. QTLs analysis and validation for fiber quality traits using maternal backcross population in upland. 2017;8:2168. <https://doi.org/10.3389/fpls.2017.02168>.
- Huang SQ, Xiang AL, Che LL, et al. A set of miRNAs from *Brassica napus* in response to sulphate deficiency and cadmium stress. Plant Biotechnol J. 2010;8:887–99. <https://doi.org/10.1111/j.1467-7652.2010.00517.x>.
- Illumina. Understanding Illumina quality scores. Tech Note Informatics. San Diego: Illumina, Inc. 2014. p. 1–2.
- Jones-Rhoades MW, Bartel DP. Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. Mol Cell. 2004;14:787–99. <https://doi.org/10.1016/j.molcel.2004.05.027>.
- Kalra YP. Handbook of reference methods for plant analysis. Boca Raton: CRC Press; 1998.
- Karaba A, Dixit S, Greco R, et al. Improvement of water use efficiency in rice by expression of HARDY, an Arabidopsis drought and salt tolerance gene. Proc Natl Acad Sci U S A. 2007;104:15270–5. <https://doi.org/10.1073/pnas.0707294104>.
- Khan NU, Marwat KB, Hassan G, et al. Genetic variation and heritability for cotton seed, fiber and oil traits in *Gossypium hirsutum* L. Pakistan J Bot. 2010;42: 615–25.
- Kimber G. Basis of the diploid-like meiotic behaviour of polyploid cotton. Nature. 1961;191:98–100. <https://doi.org/10.1038/191098a0>.
- Kocheva KV, Landjeva SP, Georgiev GI. Variation in ion leakage parameters of two wheat genotypes with different Rht-B1 alleles in response to drought. J Biosci. 2014;39:753–9. <https://doi.org/10.1007/s12038-014-9471-7>.
- Konieczny A, Ausubel FM. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 1993;4:403–10.
- Kosambi DD. The estimation of map distances from recombination values. Ann Eugenics. 1943;12:172–5. <https://doi.org/10.1111/j.1469-1809.1943.tb02321.x>.
- Kottapalli P, Ulloa M, Kottapalli KR, et al. SNP marker discovery in pima cotton (*Gossypium barbadense* L.) leaf transcriptomes. Genomics Insights. 2016;9:51–60. <https://doi.org/10.4137/GELS40377>.
- Krizman M, Jakse J, Baricevic D, et al. Robust CTAB-activated charcoal protocol for plant DNA extraction. Acta Agric Slov. 2006;87:427–33.
- Kumar RR. Protection against heat stress in wheat involves change in cell membrane stability, antioxidant enzymes, osmolyte, H<sub>2</sub>O<sub>2</sub> and transcript of heat shock protein. Int J Plant Physiol Biochem. 2012;4. <https://doi.org/10.5897/IJPPB12.008>.
- Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11:241–7. <https://doi.org/10.1038/ng1195-241>.
- Li F, Fan G, Lu C, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol. 2015a;33:524–30. <https://doi.org/10.1038/nbt.3208>.
- Li FL, Bao WK. New insights into leaf and fine-root trait relationships: implications of resource acquisition among 23 xerophytic woody species. Ecol Evol. 2015; 5:5344–51. <https://doi.org/10.1002/ece3.1794>.
- Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. Bioinformatics. 2010;26:589–95. <https://doi.org/10.1093/bioinformatics/btp698>.
- Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAM tools. Bioinformatics. 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li J, Lv J, Liu P, et al. Genome survey and high-resolution backcross genetic linkage map construction of the ridgetail white prawn *Exopalaemon carinicauda* applications to QTL mapping of growth traits. BMC Genomics. 2019;20:598. <https://doi.org/10.1186/s12864-019-5981-x>.
- Li XJ, Li M, Zhou Y, et al. Overexpression of cotton *RAV1* gene in *Arabidopsis* confers transgenic plants high salinity and drought sensitivity. PLoS One. 2015b;10:e118056. <https://doi.org/10.1371/journal.pone.0118056>.
- Li Z, Mu P, Li C, et al. QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland japonica rice in three environments. Theor Appl Genet. 2005;110:1244–52. <https://doi.org/10.1007/s00122-005-1958-z>.
- Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet. 1989;44:397–401.
- Liu J, Pattey E, Miller JR, et al. Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. Remote Sens Environ. 2010;114:1167–77. <https://doi.org/10.1016/j.rse.2010.01.004>.
- Liu Y, Min W. Thioredoxin promotes ASK1 ubiquitination and degradation to inhibit ASK1-mediated apoptosis in a redox activity-independent manner. Circ Res. 2002;90:1259–66. <https://doi.org/10.1161/01.RES.0000022160.64355.62>.
- Lopes MS, Bastiaansen JWM, Harlizius B, et al. A genome-wide association study reveals dominance effects on number of teats in pigs. PLoS One. 2014;9: e105867. <https://doi.org/10.1371/journal.pone.0105867>.
- Lu Y, Zhang S, Shah T, et al. Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc Natl Acad Sci. 2010;107:19585–90. <https://doi.org/10.1073/pnas.1006105107>.
- Lugoian C, Ciulca S. Evaluation of relative water content in winter wheat. J Hortic For Biotechnol. 2011;15:173–7.
- Magadum S, Banerjee U, Murugan P, et al. Gene duplication as a major force in evolution. J Genet. 2013;92:155–61. <https://doi.org/10.1007/s12041-013-0212-8>.
- Magwanga RO, Lu P, Kirungu JN, et al. GBS mapping and analysis of genes conserved between *Gossypium tomentosum* and *Gossypium hirsutum* cotton cultivars that respond to drought stress at the seedling stage of the BC<sub>2</sub>F<sub>2</sub> generation. Int J Mol Sci. 2018a;19. <https://doi.org/10.3390/ijms19061614>.
- Magwanga RO, Lu P, Kirungu JN, et al. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. BMC Genet. 2018b;19:6. <https://doi.org/10.1186/s12863-017-0596-1>.
- Magwanga RO, Lu P, Kirungu JN, et al. Cotton late embryogenesis abundant (LEA2) genes promote root growth and confers drought stress tolerance in

- transgenic *Arabidopsis thaliana*. G3: Genes|Genomes|Genetics. 2018c;200423. <https://doi.org/10.1534/g3.118.200423>.
- Mantovani A, Iglesias RR. Factors limiting seed germination of terrestrial bromeliads in the sandy coastal plains (Restinga) of Maricá, Rio de Janeiro, Brazil. *Rodriguésia*. 2008;59:135–50.
- Marcia MM. Feruloylation in grasses: current and future perspectives. *Mol Plant*. 2009;2:861–72. <https://doi.org/10.1093/mp/ssp067>.
- Mardani Z, Rabei B, Sabouri H, Sabouri A. Mapping of QTLs for germination characteristics under non-stress and drought stress in rice. *Rice Sci*. 2013;20:391–9. [https://doi.org/10.1016/S1672-6308\(13\)60150-X](https://doi.org/10.1016/S1672-6308(13)60150-X).
- Marshall JG, Dumbroff EB. Turgor regulation via cell wall adjustment in white spruce. *Plant Physiol*. 1999;119:313–20. <https://doi.org/10.1104/pp.119.1.313>.
- Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front Plant Sci*. 2014;5. <https://doi.org/10.3389/fpls.2014.00170>.
- Nguyen HT, Babu RC, Blum A. Breeding for drought resistance in rice: physiology and molecular genetics considerations. *Crop Sci*. 1997;37:1426–34. <https://doi.org/10.2135/cropsci1997.0011183X003700050002x>.
- Oluoch G, Zheng J, Wang X, et al. QTL mapping for salt tolerance at seedling stage in the interspecific cross of *Gossypium tomentosum* with *Gossypium hirsutum*. *Euphytica*. 2016;209:223–35. <https://doi.org/10.1007/s10681-016-1674-6>.
- Omholt SW, Plahte E, Øyehaug L, Xiang K. Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics*. 2000;155:969–80.
- Ontology G. Gene Ontology. *Nat Rev Genet*. 2009;2009:1–13.
- Paran I, Michelmore RW. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor Appl Genet*. 1993;85:985–93. <https://doi.org/10.1007/BF00215038>.
- Parent B, Suard B, Serraj R, Tardieu F. Rice leaf growth and water potential are resilient to evaporative demand and soil water deficit once the effects of root system are neutralized. *Plant Cell Environ*. 2010;33:1256–67. <https://doi.org/10.1111/j.1365-3040.2010.02145.x>.
- Pleasant JM, Wendel JF. Reproductive and pollination biology of the endemic Hawaiian cotton, *Gossypium tomentosum* (Malvaceae) 1. *Pacific Sci*. 2010;64:45–55. <https://doi.org/10.2984/64.1.045>.
- Poland J, Endelman J, Dawson J, et al. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome J*. 2012;5:103. <https://doi.org/10.3835/plantgenome2012.06.0006>.
- Prado K, Maurel C. Regulation of leaf hydraulics: from molecular to whole plant levels. *Front Plant Sci*. 2013;4. <https://doi.org/10.3389/fpls.2013.00225>.
- Pushpam R, Raveendran TS. Production of interspecific hybrids between *Gossypium hirsutum* and jassid resistant wild species *G. raimondii* and *G. armourianum*. *Cytologia*. 2006;71(4):407–18.
- Qi H, Wang N, Qiao W, et al. Construction of a high-density genetic map using genotyping by sequencing (GBS) for quantitative trait loci (QTL) analysis of three plant morphological traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica*. 2017;213. <https://doi.org/10.1007/s10681-017-1867-7>.
- Qin Y-M, Hu C-Y, Pang Y, et al. Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. *Plant Cell Online*. 2007;19:3692–704. <https://doi.org/10.1105/tpc.107.054437>.
- R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013.
- Rahman M, Ullah I, Ahsraf M, et al. Genotypic variation for drought tolerance in cotton. *Agronomy for Sustainable Development*. 2008;28(3):439–47.
- Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. *J Magnetic Resonance Imaging*. 2012;36(5):1011–4.
- Rhoades MW, Reinhart BJ, Lim LP, et al. Prediction of plant microRNA targets. *Cell*. 2002;110:513–20. [https://doi.org/10.1016/S0092-8674\(02\)00863-2](https://doi.org/10.1016/S0092-8674(02)00863-2).
- Robbins NE, Dinneny JR. The divining root: moisture-driven responses of roots at the micro- and macro-scale. *J Exp Bot*. 2015;66:2145–54.
- Salimath SS, de Oliveira AC, Godwin ID, Bennetzen JL. Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome*. 1995;38:757–63. <https://doi.org/10.1139/g95-096>.
- Salmela L. Correction of sequencing errors in a mixed set of reads. *Bioinformatics*. 2010;26:1284–90. <https://doi.org/10.1093/bioinformatics/btq151>.
- Shinozaki K, Yamaguchi-Shinozaki K. Gene networks involved in drought stress response and tolerance. *J Exp Bot*. 2007;58:221–7.
- Singh UM, Yadav S, Dixit S, et al. QTL hotspots for early vigor and related traits under dry direct-seeded system in rice (*Oryza sativa* L.). *Front Plant Sci*. 2017;8. <https://doi.org/10.3389/fpls.2017.00286>.
- Sircar S, Parekh N. Functional characterization of drought-responsive modules and genes in *Oryza sativa*: a network-based approach. *Front Genet*. 2015;6. <https://doi.org/10.3389/fgene.2015.00256>.
- Specht JE, Chase K, Macrauder M, et al. Soybean response to water: a QTL analysis of drought tolerance. *Crop Sci*. 2001;41:493–509.
- Stubber CW, Edwards MD, Wendel JF. Molecular-marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop Sci*. 1987;27:639–48. <https://doi.org/10.2135/cropsci1987.0011183X002700040006x>.
- Sunkar R, Chinnusamy V, Zhu J, Zhu JK. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci*. 2007;12:301–9.
- Swamy BPM, Kaladhar K, Anuradha K, et al. QTL analysis for grain iron and zinc concentrations in two *O. nivara* derived backcross populations. *Rice Sci*. 2018;25:197–207. <https://doi.org/10.1016/j.rsci.2018.06.003>.
- Swinnen S, Thevelein JM, Nevoigt E. Genetic mapping of quantitative phenotypic traits in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. 2012;12:215–27.
- Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. *Science*. 2010;327:818–22. <https://doi.org/10.1126/science.1183700>.
- Todaka D, Shinozaki K, Yamaguchi-Shinozaki K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Front Plant Sci*. 2015;6:84. <https://doi.org/10.3389/fpls.2015.00084>.
- Van Ooijen JW, Voorrips RE. JoinMap 3.0 - a software for the calculation of genetic linkage maps completely. Wageningen: Plant Res Int; 2001. p. 1–51.
- Volodina E, Steinbüchel A. (S)-3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase (FadB) from fatty acid degradation operon of *Ralstonia eutropha* H16. *AMB Express*. 2014;4:1–9. <https://doi.org/10.1186/s13568-014-0069-0>.
- Vos P, Hogers R, Bleeker M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*. 1995;23:4407–14. <https://doi.org/10.1093/nar/23.21.4407>.
- Wahl MC, Will CL, Lüthmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009;136:701–18.
- Wang J, Sun N, Deng T, et al. Genome-wide cloning, identification, classification and functional analysis of cotton heat shock transcription factors in cotton (*Gossypium hirsutum*). *BMC Genomics*. 2014;15:961. <https://doi.org/10.1186/1471-2164-15-961>.
- Wang R, Ji S, Zhang P, et al. Drought effects on cotton yield and fiber quality on different fruiting branches. *Crop Sci*. 2016;56:1265–76. <https://doi.org/10.2135/cropsci2015.08.0477>.
- Wang S, Basten CJ, Zeng ZB. Windows QTL cartographer 2.5. *J Infect Dis*. 2011;204(Suppl):198–9.
- Watson SJ, Sowden RG, Jarvis P. Abiotic stress-induced chloroplast proteome remodelling: a mechanistic overview. *J Exp Bot*. 2018. <https://doi.org/10.1093/jxb/ery053>.
- Wendel JF, Cronn RC, Spencer Johnston J, James Price H. Feast and famine in plant genomes. *Genetica*. 2002;115(1):37–47. <https://doi.org/10.1023/A:1016020030189>.
- Wilfinger WW, Mackey K, Chomczynski P. 260/280 and 260/230 ratios NanoDrop® ND-1000 and ND-8000 8-sample spectrophotometers. *BioTechniques*. 1997;22:474–81. <https://doi.org/10.7860/JCDR/2015/11821.5896>.
- Williams JGK, Kubelik AR, Livak KJ, et al. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res*. 1990;18:6531–5. <https://doi.org/10.1093/nar/18.22.6531>.
- Wood V. Use and misuse of the Gene Ontology annotations. *Nature Reviews Genetics*. 2008;9:509–515. <https://doi.org/10.1038/nrg2363>.
- Würschum T. Mapping QTL for agronomic traits in breeding populations. *Theor Appl Genet*. 2012;125:201–10.
- Xie F, Wang Q, Sun R, Zhang B. Deep sequencing reveals important roles of microRNAs in response to drought and salinity stress in cotton. 2015;66:789–804. <https://doi.org/10.1093/jxb/eru437>.
- Xu Z, Zhou G. Responses of leaf stomatal density to water status and its relationship with photosynthesis in a grass. *J Exp Bot*. 2008;59:3317–25. <https://doi.org/10.1093/jxb/ern185>.
- Zhang B, Wang Q, Wang K, et al. Identification of cotton microRNAs and their targets. *Gene*. 2007;397:26–37. <https://doi.org/10.1016/j.gene.2007.03.020>.
- Zhang J, Stewart J, Mac. Economical and rapid method for extracting cotton genomic DNA. *J Cott Sci*. 2000;4:193–201.

- Zhang J, Yu J, Pei W, et al. Genetic analysis of Verticillium wilt resistance in a backcross inbred line population and a meta-analysis of quantitative trait loci for disease resistance in cotton. *BMC Genomics*. 2015a;16:577.
- Zhang T, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015b;33:531–7. <https://doi.org/10.1038/nbt.3207>.
- Zhang X, Zhai C, He L, et al. Morphological, cytological and molecular analyses of a synthetic hexaploid derived from an interspecific hybrid between *Gossypium hirsutum* and *Gossypium anomalum*. *Crop J*. 2014;2:272–7. <https://doi.org/10.1016/j.cj.2014.06.009>.
- Zhang ZS, Hu MC, Zhang J, et al. Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Mol Breed*. 2009;24:49–61. <https://doi.org/10.1007/s11032-009-9271-1>.
- Zhao Y-T, Wang M, Fu S-X, et al. Small RNA profiling in two *Brassica napus* cultivars identifies MicroRNAs with oil production- and development-correlated expression and new small RNA classes. *Plant Physiol*. 2012;158: 813–23. <https://doi.org/10.1104/pp.111.187666>.
- Zheng JY, Oluoch G, Riaz Khan MK, et al. Mapping QTLs for drought tolerance in an F<sub>2,3</sub> population from an inter-specific cross between *Gossypium tomentosum* and *Gossypium hirsutum*. *Genet Mol Res*. 2016;15. <https://doi.org/10.4238/gmr.15038477>.
- Zhou M, Sun G, Sun Z, et al. Cotton proteomics for deciphering the mechanism of environment stress response and fiber development. *J Proteome*. 2014; 105:74–84.

**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)

