

Genome-wide association study reveals genetic variation and candidate genes of lint yield components under salt field conditions in cotton (*Gossypium hirsutum* L.)

Guozhong Zhu

Nanjing Agricultural University

Wenwei Gao

Xinjiang Agricultural University

Xiaohui Song

Nanjing Agricultural University

Fenglei Sun

Xinjiang Agricultural University

Sen Hou

Nanjing Agricultural University

Na Liu

Xinjiang Agricultural University

Yajie Huang

Xinjiang Agricultural University

Dayong Zhang

Nanjing Agricultural University

Zhiyong Ni

Xinjiang Agricultural University

Quanjia Chen

Xinjiang Agricultural University

Wangzhen Guo (✉ moelab@njau.edu.cn)

<https://orcid.org/0000-0003-3333-7147>

Research article

Keywords: Genome-wide association study, Lint percentage, Boll number per plant, Boll weight per plant, Salt stress, Transcriptome analysis, Haplotype, *Gossypium hirsutum*

Posted Date: October 10th, 2019

DOI: <https://doi.org/10.21203/rs.2.15945/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 14th, 2020. See the published version at <https://doi.org/10.1186/s12870-019-2187-y>.

Abstract

Background Salinity is one of the most decisive environmental factors limiting the productivity of cotton. However, the key genetic components leading to the reduction of cotton yield in saline-alkali soils are still unclear. Results Here, we evaluated three main components of lint yield across 316 *G. hirsutum* accessions, including single boll weight (SBW), lint percentage (LP) and boll number per plant (BNPP), under four salt conditions for two years. Phenotypic analysis indicated that LP showed no change under different salt conditions, however BNPP decreased significantly while SBW increased slightly under high salt condition. Based on 57,413 high-quality single nucleotide polymorphisms (SNPs) and genome-wide association studies (GWAS) analysis, a total of 42, 91 and 25 stable quantitative trait loci (QTLs) were identified for SBW, LP and BNPP, respectively. Few overlapped QTLs and no significant phenotypic correlation among the three traits was observed. Gene Ontology (GO) analysis indicated that their regulatory mechanisms were also quite different. There were 8 overlapped QTLs for LP while fewer for SBW and BNPP identified by comparing different salt conditions. We detected that 10 genes from the 8 stable LP QTLs were predominantly expressed during fiber development. Further, haplotype analyses found that a MYB gene (GhMYB103) with the two SNP variations in cis-regulatory and coding regions, was significantly correlated with lint percentage, implying the crucial role in lint yield. With transcriptome analysis, we identified that 40 candidate genes from BNPP QTLs were salt-inducible. However, these genes exhibited different regulation pattern. Genes related to carbohydrate metabolism and cell structure maintenance were rich in high salt condition, while genes related to ion transport were active in low salt condition. Conclusions This study provides a foundation for elucidating cotton salt tolerance mechanism and contributes gene resources for developing varieties with high yield and salt stress tolerance in upland cotton.

Background

Salinity is one of the most decisive environmental factors limiting the productivity of crop plants [1]. Salinity stress affected about one billion hectares of arid and semi-arid areas globally [2] and is becoming progressively more severe due to climatic changes, unscientific irrigation and excessive fertilization [3]. Scientists and agronomists have done lots of efforts to make improvements and utilization of saline soil. In addition to using physical and chemical methods to reduce salt content, screening or breeding high-salt-tolerant crops by modern molecular means is economic and effective way to solve the present situation.

In the past few decades, several methods, such as molecular marker [4], transgene technology [5], transcriptome sequencing [6], and genome-wide association study (GWAS) based on single nucleotide polymorphism (SNP) [7], have been used for investigating mechanism of salt tolerance and mining elite alleles in plants. GWAS is used widely in plants because of effectively associating genotypes with phenotypes and detecting many natural allelic variations simultaneously by using natural populations. Liu et al. (2017) identified 13 novel loci affecting maize kernel oil concentration by combining SNPs (1.25Mb) from reduced genome sequencing and high-density arrays (600K) with RNA-sequencing data

[8]. Shi et al. (2017) reported 11 loci associated with stress-susceptibility indices (SSIs) of vigor index (VI) and mean germination time (MGT) by screening 6,361,920 SNPs on 478 rice accessions [9]. These loci/genes after prior to functional analysis may be used in molecular assisted breeding.

Cotton (*Gossypium spp.*) could be used for soil reclamation as a pioneer crop of saline-alkali land due to its high salt tolerance [7]. However, the growth and development can still be affected by adverse salt conditions. Soil salinity ranging from 8 to 18 dS/m resulted in yield losses of 15 to 55% in cotton [10]. Therefore, discovering the limiting factors, further screening and identifying the candidate genes in the salt-tolerance response pathway is an effective way to increase cotton yield and provide reference for other plants. Many candidate genes that underlie traits such as fiber yield and quality [11-13], seed oil composition and protein content [14], have been revealed. Above these, several studies related stress tolerance have been reported in cotton. Using 145 simple-sequence repeat (SSR) markers, 60 quantitative trait loci (QTLs) associated with ten salt-tolerance related traits were detected [4]. Via genome-wide SNP analysis by genotyping by sequencing (GBS), a total of 66 QTLs for 10 traits related to salinity were identified and 12 candidate genes in QTLs might play crucial roles in salt tolerance in cotton [15]. Using CottonSNP63K array, a total of 23 SNPs that represented seven genomic regions were significantly associated with the two salt-tolerance-related traits, relative survival rate and salt tolerance level at the seedling stage in Upland cotton [7]. Using genome re-sequencing data, nine SNP rich regions associated with relative fresh weight, relative stem length, relative water content and comprehensive index of salt tolerance under salt condition were reported in Asiatic cotton [16]. So far, several studies focused on salt-tolerance related traits, few reports was to investigate fiber yield under stress conditions including salt stress, which had more practical value for saline-alkali soils utilization.

In the current study, we investigated three main components of lint yield, single boll weight (SBW), lint percentage (LP) and boll number per plant (BNPP), across 316 *G. hirsutum* accessions with diverse origins under four different salt conditions for two-years. Phenotypic variation of the three lint yield components under different salt environments showed that BNPP is a major limiting factor causing the reduction of lint yield per plant (LYPP). Further, GWAS analysis was conducted by applying 57,413 SNPs to identify QTLs and candidate genes associated with LP, SBW and BNPP. Current study systematically clarified the genetic characterization that LP was the most stable and BNPP was most easily affected under salt field conditions. The results provide new insight into understanding the mechanisms of salt stress and breeding improvement of boll number with enhanced salt tolerance in cotton.

Results

Phenotypic variation of lint yield components under salt condition

At maturity stage, three traits including SBW, LP and BNPP were measured in 316 upland cotton accessions (Additional file 1: Table S1) with four salt conditions (Additional file 2: Fig. S1A and Additional file 3: Table S2). The average values of SBW, LP and BNPP traits ranged from 5.28g to 6.84g (Additional file 4: Table S3), 37.89% to 40.39% (Additional file 5: Table S4) and 5.01 to 7.74 (Additional file 6: Table

S5) under four different salt conditions, respectively. The BNPP had the largest coefficients of variation (CV) ranging from 22.53% to 48.08%, and the LP had the smallest (8.15%~12.68%). For decreasing the environmental errors, we further evaluated the BLUP value of the three traits in the same salt condition. The BLUP value showed that the distribution of SBW was at 5~7 g (Additional file 2: Fig. S1B) and LP at 35~45% (Additional file 2: Fig. S1C). However, the BNPP showed larger difference with 3~7 in salt condition A and B, and 5~9 in salt condition C and D (Additional file 2: Fig. S1D).

Paired-samples *t*-test was conducted to further investigate the phenotypic variation. The SBW was slightly higher in condition A than in B, C and D conditions (Fig. 1A). However, no significant difference of LP was detected under all four salt conditions (Fig. 1B). Most significantly, the BNPP decreased with the increase of salt concentrations (Fig. 1C). In addition, the LYPP also showed the decrease with the increase of salt concentrations mainly due to change of BNPP (Fig. 1D). We compared respectively the differences of SBW, LP, BNPP and LYPP, at the highest with the lowest salt conditions, and the results showed that SBW increased by 5.29%, LP remained unchanged, BNPP decreased by 17.75%, and LYPP decreased by 16.79%. Taken together, Lint yield of cotton was mainly affected by boll number decreased under high salt concentration, although SBW slightly increase relevant to severe boll numbers decrease.

Correlation analysis was conducted among SBW, LP and BNPP under four salt conditions (Additional file 7: Fig. S2). The results showed that there was low correlation among the three traits. As for the same trait, under different salt environments, LP showed high correlation with R value from 0.75 to 0.83, followed by SBW with R value from 0.43 to 0.63. BNPP showed no or weak correlation in four salt conditions, further indicating that BNPP was largely affected by salt concentration.

GWAS of the lint yield related traits

With the genotypic data of 57,413 high-quality SNPs [17], the GWAS was conducted for the three traits with different environments and the BLUP values using six methods (“mrMLM”, “FASTmrMLM”, “FASTmrEMMA”, “pKWmEB”, “ISIS EM-BLASSO” and “pLARmEB”) of multi-loci MLM model in R package “mrMLM” [18]. In total, 854 QTNs on 26 chromosomes were identified as significantly associated with the three traits (Additional file 8: Table S6). Following the linkage disequilibrium (LD) value [17] and referring the method reported by Song et al. (2019) [19], the 200-kb upstream and downstream of QTNs were defined as one QTL. In total, 600 QTLs including 151 of SBW, 417 of LP and 112 of BNPP, were detected with 1446 times under different environments and BLUPs by six multi-loci MLM models (Table 1). For each trait, most QTLs (85 of SBW, 235 of LP and 65 of BNPP) were detected only once, implying these QTLs be apt to be affected by environmental condition (Additional file 9: Fig. S3). To improve the reliability and stability of associated QTLs, we selected QTLs detected three or more times across different methods or environments as stable QTLs for further analysis. As a result, 42, 91 and 25 QTLs were identified in SBW, LP and BNPP, respectively (Table 1).

The chromosomal distribution showed that these stable QTLs were widely distributed on 26 chromosomes, with more QTLs of SBW and BNPP located on At sub-genome than on Dt sub-genome, while QTLs of LP showed opposite (Fig. 2A). Most of SBW located on chromosomes A11 and A12, LP on A08, D06 and D13, and BNPP on A05, A12 and D07 (Fig. 2B). The vein diagram of these stable QTLs showed that no overlapped QTL was detected within three traits and most of QTLs were specific for individual trait (Fig. 3A), implying great differences in regulating pathway for the three traits. We further analyzed the QTLs of each trait under different salt conditions, only one QTL of SBW was detected under all four salt conditions, and no overlapped QTL detected in BNPP (Fig. 3B, C). However, most QTLs of LP were overlapped under different salt conditions, there were 8 QTLs of LP detected under all four salt conditions (Fig. 3D). The results suggest complex and various regulatory mechanism involved in SBW and BNPP under different salt conditions but stable and high heritability of LP against salt stress.

Identification of candidate genes in QTLs

Potential candidate genes in these stable QTL regions were extracted based on the released *G. hirsutum* TM-1 genome [20]. In total, 1166, 2748, and 711 candidate genes were identified in the QTL regions for SBW, LP and BNPP, respectively (Fig. 2C), with most genes distributed on chromosome A12, D13 and A12 for SBW, LP and BNPP, respectively (Fig. 2D). With GO analysis, the genes in QTL regions for SBW enriched in “embryo development” and “regulation of cell shape” (Additional file 10: Fig. S4 and Additional file 11: Table S7). The genes from LP QTLs mainly enriched in several pathways, including “regulation of organ growth”, hormone and ROS regulation such as “regulation of gibberellic acid mediated signaling pathway”, “positive regulation of reactive oxygen species metabolic process”, “brassinosteroid biosynthetic process” and “defense response by cell wall thickening”, and carbohydrate metabolism such as “glycosylation”, “glucose metabolic process”, “monosaccharide biosynthetic process” and “hexose biosynthetic process”, which was consistent with the previous reports that these Go items played the crucial roles in fiber development [11, 21]. In addition, we identified 14, 21 and 10 genes related to “Golgi vesicle transport”, “plant-type secondary cell wall biogenesis” and “glucose metabolic process”, respectively, however, none of these process-related genes were found in the QTL regions of BNPP and SBW (Additional file 12: Fig. S5 and Additional file 11: Table S7). The function of genes associated with BNPP mainly enriched in “mitotic cell cycle”, “ion transmembrane transport” and “polysaccharide catabolic process” (Additional file 13: Fig. S6 and Additional file 11: Table S7), implying that BNPP was closely related to stress response.

Genes relevant to LP

Via tissue and organ transcriptome profile [22], we identified 182 genes from LP QTLs with predominant expression during fiber development. Of them, 29, 35, 73 and 45 genes highly expressed at 10 DPA, 15 DPA, 20 DPA and 25 DPA, respectively (Additional file 14: Fig. S7 and Additional file 15: Table S8). Further,

we focused on the genes in the regions of 8 overlapped QTLs under all four salt conditions and identified 10 genes predominantly expressed during fiber development (Table S9). Of them, *Gh_A05G2488* encoding auxin transport facilitator family called PIN-FORMED LIKE proteins (PILS), located in a high frequency associated QTL (A05: 32377816-33100112) which was detected 21 times with multi-methods and environments. Auxin is essential for plant growth and development. Prominent auxin carriers with fundamental importance during plant development are PIN-FORMED (PIN) proteins [23, 24]. In *Arabidopsis*, overexpressing *PILS* genes could reduce hypocotyl and root growth [25]. However, auxin accumulation can promote cell initiation (-2 to 2 DPA) in the fiber cell. Transgenic assay showed that ovule-specific suppression of multiple *GhPIN* genes inhibited both fiber initiation and elongation in cotton [26], indicating that *Gh_A05G2488* might play an important role in the fiber development. Two genes, *Gh_D13G0342* encoding RAB GTPase homolog G3F (RABG3F) and *Gh_A10G2138* encoding prenylated RAB acceptor protein 1 (PRA1), located on QTL (D13: 3297533-3912736) and QTL (A10: 99896949-100396471), respectively. Previous reports suggested that *RAB* genes played crucial roles in cell polarity growth, including elongation of pollen tubes [27] and root hairs [28]. In addition, *RAB* genes also involved in cotton fiber development [29]. Prenylated Rab acceptor 1 (PRA1) domain proteins are small transmembrane proteins that regulate vesicle trafficking as receptors of Rab GTPases. AtPRA1 proteins were localized in the endoplasmic reticulum, Golgi apparatus, and endosomes/prevacuolar compartments, indicating a function in both secretory and endocytic intracellular trafficking pathways [30]. These results indicate that the RAB and RAB acceptor protein have potential important functions in cotton fiber development.

Cotton seed transfer cells are enriched in callose, which regulated fiber elongation and secondary wall thickening [31]. Zhang et al. (2007) showed that the transcription factor MYB103 affects callose dissolution during the anther development in *Arabidopsis* [32]. We found *Gh_D03G1419* encoding MYB103 transcription factor (named as *GhMYB103*), located in a high frequency associated QTL (D03: 42299450-43529568). Sequence analysis showed that two QTNs associated with LP located in the 3739 bp upstream (TM55216, D03: 42969311) and exon (TM55217, D03: 42973276) regions of the gene, respectively (Fig. 4A). The single nucleotide mutation (from C to G) at TM55217 locus led to the change of amino acid from leucine (L) to valine (V) (Fig. 4A). Through Student's *t* test, we found that LP with A genotype in TM55216 was significantly higher than with G genotype (Fig. 4B), and with G genotype in TM55217 significantly higher than with C genotype (Fig. 4C). The two QTNs could generate 3 haplotypes including H1: AG, H2: AC and H3: GC. LP with AG and AC haplotypes were significantly higher than that with GC haplotypes. However, there was no significant difference between AG and AC haplotypes, implying that QTN TM55216 might play more important roles in LP (Fig. 4D).

In addition, we integrated the LP QTLs with GWAS signals published in previous reports [11, 12], and found three QTLs (A12: 602614-743324; D13: 58792627-59289811; A09: 4676815-5076815) overlapped with GWAS signals. The three QTLs were detected in different salt conditions. Further, 10 genes from the QTL regions were identified to be predominant expression during fiber development (Additional file 16: Table S9). Of them, *ABP1* [33] and *CPK17* [34] was related to hypocotyl growth or fiber development.

Taken together, the genes from these stable LP QTLs and expressed predominantly in fiber developmental stages play an important role for the LP improvement in breeding practice.

Genes relevant to BNPP

The QTLs of BNPP were easily affected in different environments. No overlapped QTL under all four salt conditions were detected and most QTLs were identified only in one or two salt environments. For example, QTL TM57617_TM57620 (D05: 24.3-24.8 Mb, detected with 16 times) and TM74225 (D10: 24.1-24.5 Mb, detected with 7 times) were identified in salt conditions C and D. However, QTL TM52041_TM52044 (D02: 49.2-49.7 Mb, detected with 6 times) was identified only in salt condition A, and QTL TM29006 (A08: 81.7-82.1 Mb, detected with 5 times) was identified only in salt condition B. These results indicated that the genes regulating the number of bolls were various in different salt conditions.

A large number of items related to ion transport were enriched with GO analysis (Additional file 13: Fig. S6 and Additional file 11: Table S7). In order to further explore the mechanism of stress tolerance on the increase of boll number, we performed RNA-seq analysis using salt stress transcriptome in *Gossypium* acc. TM-1. With the filter of FPKM ≥ 3 , 500 genes (446 in roots and 395 in leaves) in QTL regions were obtained. With GO annotation and differential expression analysis, we further focused on 40 salt-inducible stress response genes, of them, 6 commonly identified in roots and leaves (Additional file 17: Fig. S8, Additional file 18: Table S10 and Additional file 19: Table S11). *Gh_A04G1216* encoded a high-affinity K⁺ transporter 1 (HKT1). *AtHKT1* limits the root-to-shoot sodium transportation and is believed to be essential for salt tolerance in *Arabidopsis thaliana* [35]. *Gh_A05G3239* encoded a peroxidase superfamily protein (POD), which had been proved playing an important role in anti-oxidation under salt stress in cotton [36]. *Gh_A08G1183* encoded a mitogen-activated protein kinase (MAPK), which was widely reported to be associated with salt tolerance in cotton [37]. Besides ion transport Go term, carbohydrate metabolism was active under salt stress, for example, 10 genes were identified in the enriched Go term “polysaccharide catabolic process” and 3 of them differentially expressed under salt stress. Cell cycle regulation is of pivotal importance for plant growth and development [38]. The 33 genes were identified in the enriched Go term “cell cycle” with 8 of them differentially expressed under salt stress. These candidate genes could contribute to increasing boll number under salt environment in cotton.

In order to explore the key genes and regulation mechanism of boll number under high and low salt conditions, we compared the genes located in QTL regions of BNPP between condition A and D. Totally, 204 and 265 genes were identified under salt condition A and D, respectively. GO enrichment analyses showed the significantly different function classification of genes between high and low salt conditions. Go terms related to “polysaccharide metabolic process”, “carbohydrate catabolic process” and “cell wall organization” were enriched under high salt condition A (Fig. 5A, Additional file 20: Table S12) and “cell

cycle”, “ion transmembrane transport” and “regulation of signal transduction” under low salt condition D (Fig. 5B), which indicate that carbohydrate metabolism and cell structure maintenance play an crucial role under high salt condition and ion transport is more basal under low salt condition. Under high salt condition, several candidate genes associated with BNPP were detected. *Gh_A11G1551* encodes a proline dehydrogenase 1 (ProDH1), also called early responsive to dehydration 5 (ERD5), which have been studied extensively, especially under abiotic stress [39]. To counteract osmotic stress caused by salt stress, some plants accumulate several kinds of compatible osmolytes, such as proline, glycine betaine, and sugar alcohols, to protect macromolecules and remains osmotic pressure equilibrium inside and outside cell membrane [40]. The expression of ProDH2, a highly homologous gene of *ProDH1*, can promote proline accumulation under stress conditions [41]. For energy metabolism, *Gh_A05G1912* encodes an isoamylase 3 (ISA3) which contributes to starch breakdown. *Atisa3* mutants have more leaf starch and a slower rate of starch breakdown than wild-type plants [42]. Under low salt condition, three candidate genes were identified to play important roles in the balance between sodium and potassium. In detail, *Gh_A10G0441* encodes a potassium transporter 1 (KUP1) [43], *Gh_A12G0074* encodes a high affinity K⁺ transporter 5 (HAK5) [44] and *Gh_A12G0061* encodes a sodium hydrogen exchanger 2 (NHX2) [45].

Discussion

With the decrease in arable land area and the deterioration of soil environments throughout the world, there is an urgent need to improve for stress tolerance in crop plants. Xinjiang is the main cotton production area in China, but the soil salinization is serious. Excavating elite alleles that can increase cotton yield under saline-alkali conditions is of great significance for new varieties breeding with high yield and stress resistance. With the development of high-throughput sequencing technology and new statistical methods, GWAS provides a fast and effective method for functional gene mining in plants. Several GWAS signals for vegetative growth index related to salt tolerance have been detected at germination and seeding stages in previous studies [4, 16]. However, there are few GWAS studies on cotton yield traits under stress condition. In present study, we focused on three lint yield traits (SBW, LP and BNPP) and identified favorable associated QTLs and elite alleles under four salt conditions. The results provides new insights into the genetic basis of salt tolerance and the identification of the novel alleles underlying the variation in the salt-tolerant traits and candidate genes, allowing for accelerating the progress of cotton tolerance breeding.

Cotton yield decreases under salt stress. The previous studies reported that SBW, LP and BNPP remarkably decreased under salt stress [46-48]. In present study, we selected four different field soil salt environments with two year replicates to investigate the lint yield components of cotton accessions. Through phenotypic analysis, LP did not change significantly under different salt conditions and SBW was slightly higher in salt condition A than in other three conditions, might be relevant to the decreased BNPP. It's worth noting that BNPP decreased significantly with the increase of total salt content, which results in the reduction of cotton lint yield. Totally, BNPP is the most susceptible factor to salt stress compared to the other two traits. Compared with the lowest salt condition, the highest salt condition

could cause 16.79% loss of LYPP, which was mainly caused by the reduction of BNPP. Hence, it is important to maintain the boll number of cotton under salt stress for cotton lint yield. In addition, there was less correlation among SBW, LP and BNPP traits, implying the different biological processes in the regulation of the three traits under salt stress.

Increasing yield is a major goal in cotton breeding program. LP is an important yield component and a critical economic index for cotton cultivars. Although phenotypic analysis showed no significant change of LP under different salt conditions in present study, GWAS results indicated that there were fewer overlapped QTLs compared with the previous reports in non-salt conditions [11, 12], implying that some specific genes contribute to the improvement of LP under salt condition. We also found a large number of QTLs associated with LP on chromosomes A08 and D08, which were reported to contain many QTLs or key genes related to fiber development [11, 12]. However, the distribution of QTLs associated with LP was quite different from report by Su et al (2016) [49]. Taken together, LP is a complex quantitative trait, and the majority of loci detected in our study were novel and might be related to salt stress. Especially, 8 QTLs were identified commonly in four salt conditions, which could contribute to the increase of LP under salt stress. Further, we identified 10 genes which were closely related to fiber development in these overlapped LP QTLs, such as *PILS*, *RAB* and *MYB*. In Arabidopsis, RabA4d is necessary for the proper regulation of pollen tube growth. Loss of RabA4d leads to the destruction of pollen tube growth and changes in the structure of the cell wall [50]. MYB transcription factor is also well known to play crucial roles in fiber development. *GhMYB212* RNAi plants (*GhMYB212i*) accumulated less sucrose and glucose in developing fibers, and had shorter fibers and a lower lint index [51]. Zhang et al. (2007) showed that the transcription factor MYB103 affects callose dissolution during the anther development in Arabidopsis [32]. Although many genes related to LP have been identified by GWAS analysis, however, candidate genes in this study may play a more important role in improving LP under salt stress. In addition, we also identified three QTLs overlapping with the LP loci reported previously from GWAS analysis and further identified 10 candidate genes in the QTL regions. These studies could provide genes resource for improving LP in both salt and normal environments.

GWAS analysis on BNPP are relatively rare, especially under salt stress. Our studies showed that salt stress can lead to a significant decrease in boll number per plant which was consistent with previous reports [46, 52], indicating that boll number is the first limiting factor for increasing cotton lint yield under stress environments. We also found the overlapped QTLs associated with BNPP was few under different salt conditions, implying a complex regulatory mechanism for BNPP production. GO analysis showed that genes associated with BNPP mainly involved in “mitotic cell cycle”, “ion transmembrane transport” and “polysaccharide catabolic process”. Of them, a large number of ion transport related processes are enriched, which suggests that the excellent ion transport capacity plays a key role in salt tolerance of cotton. Na⁺ accumulation can lead to ion poisoning, which induces decline of biomass and yield losses in crop plants [1]. Maintaining ion homeostasis by ion uptake and compartmentalization is crucial for plant growth during salt stress. With RNA-seq analysis, we found that *HKT1*, which is known to play a role in the removal of Na⁺ from the xylem and bring it back to the root, down-regulated under salt stress.

Overexpression of *HKT1* in roots can decrease Na^+ accumulation in the shoot and significantly improve salt tolerance in *Arabidopsis thaliana* [53]. Interestingly, *HKT1* was also found down-regulated under salt stress in *G. davidsonii*, a cotton D-genome diploid species with important properties of salinity stress resistance [54]. It suggests that the function of *HKT1* could be improved in cotton for increasing the stress tolerance. We also found that no overlapped QTLs associated with BNPP was detected by comparing under high and low salt conditions, implying a complex regulatory mechanism under different salt conditions. The enriched genes under high salt condition are mainly related to energy metabolism and maintenance of cell morphology. High salt stress can lead to a decrease in photosynthetic efficiency of plants [55]. Under non-stressed conditions, plants use the majority of the energy to maintenance vegetative and reproductive growth. However, plants need to allocate more energy to resist the stress with the increase of salt concentration [1]. In addition, high salt concentration can also increase osmotic stress, and plants need to synthesize more osmolytes to maintain cell morphology. In this study, we found that *ISA3* played crucial roles in energy metabolism and *ProDH1* contributed to maintenance of cell morphology. In addition, the enriched genes under low salt condition are mainly related to ion transport. As a salt-tolerant crop, cotton suffers less salt damage under low salt conditions, which may be due to efficient ion transport capacity. In this study, *KUP1*, *HAK5* and *NHX2* were identified that contributed to ion homeostasis. It suggests that the active sodium and potassium ion exchange capacity at low salt concentration is the basis of salt tolerance in cotton. Meanwhile, it also reflects the different demand for stress resistance under different salt stress conditions in cotton.

Several reports have suggested that the lint yield can be improved by altering the expression of salt-tolerant genes in cotton. For example, overexpressing *AvDH1* can decrease membrane ion leakage, along with increased activity of superoxide dismutase and lead to salinity tolerance and increasing yield in cotton [52]. Overexpression *SNAC1*, which belongs to the stress-related NAC superfamily of transcription factors, could improve drought and salt tolerance by enhancing root development and reducing transpiration rate in transgenic cotton [56]. In this study, we first report phenotypic and GWAS analysis of three lint yield components in cotton under salt stress and found that BNPP was the most important factor for cotton lint yield in saline-alkali soil environment. Further, we identified a large number of elite alleles which contributed to improvement of lint yield under salt conditions. These findings will help us understand the salt tolerance mechanism in cotton and provide improvement for breeding cultivars in saline-alkali soil environment.

Methods

Plant materials and field experiments

A total of 316 upland cotton accessions, with 303 cultivars/lines collected from different regions of China and 13 landraces introduced from the United States, were used in this study (Additional file 1: Table S1).

In 2016 and 2017, the 316 Upland cotton accessions were planted in 4 different salt field concentration in Xinjiang Agricultural University Experimental Station (43°20'–45°20'E, 84°45'–86°40'N). The soil total salt

contents were measured with five point sampling for each environment. In detail, the total salt contents in the four environments were identified as 19 g/Kg (condition A), 10 g/Kg (condition B), 7 g/Kg (condition C), and 5 g/Kg (condition D), respectively, with two replicate plots for each salt condition (Additional file 2: Fig. S1A and Additional file 3: Table S2). With wide/narrow row alternation plantation mode (10 cm for narrow row and 66 cm for wide row), each accession was grown two rows with 2 m row length and 0.10 m between plants for each plot. Drip fertilization beneath mulched film was used for plant growth. Other agronomic practices were same for all the treatments.

Phenotype investigation and data analysis

Ten plants for each accession in each plot were selected randomly from middle part of each row and tagged for identification to record the data for SBW, LP and BNPP. At plant maturity (approximately 70% boll open), BNPP was counted with ten biological replicates. A total of 20 well developed open boll samples (2 bolls per plant) from the middle branches of tagged plants were harvested and weighed for SBW and calculation of LP. To reduce environmental influences, the best linear unbiased predictors (BLUPs) based on mixed linear model for the three traits under each salt condition were estimated using the function of 'lmer' in the lme4 package [57]. In order to explore the effect of salt concentration on cotton lint yield, LYPP was calculated by multiplying SBW, LP and BNPP.

Paired-samples *t*-test and correlation analysis among different salt conditions and traits were performed using SPSS software, respectively. Visualization of correlation analysis was performed by R package "Performance Analytics".

GWAS analysis

Genomic DNA of the 316 cotton accessions was extracted according to the method described by Paterson et al. (1993) [58]. CottonSNP80K array was applied to genotype the 316 cotton accessions. The SNP genotyping and population structure were reported in our previous study [17]. A total of 57,413 SNPs (calling rate ≥ 0.9 and minor allele frequency (MAF) ≥ 0.05) were used for GWAS analysis. To explore the SNP-trait association, multi-locus random-SNP-effect mixed linear model (mrMLM) [18] was employed using the R package "mrMLM" with the following parameters: Critical P-value in rMLM: 0.001; Search radius of candidate gene (Kb): 100; Critical LOD score in mrMLM: 3. And the Q+K model was used. Population structure (Q) matrix was calculated using admixture 1.3 with $k = 3$, and kinship (K) matrix was calculated by the R package "mrMLM". The BLUP values and single environments of three traits under different salt conditions were individually used for the GWAS.

QTLs and candidate genes identification

Following the LD value [17] and referring the method reported by Song et al. (2019) [19], the 200 kb upstream and downstream of QTNs were defined as QTLs. The distance of two QTNs less than 200 kb will be merged into a single QTL. The QTLs of each trait discovered three or more times using different methods or in different environments were selected as stable QTLs. Putative candidate genes located in the stable QTL regions were extracted by self-written shell scripts from the reference genome TM-1 [20]. Gene ontology (GO) analysis was implemented using AgriGO V2.0 with SEA method [59].

RNA-seq analysis

For investigating the genes related to lint yield, the transcriptome profiles of TM-1 tissues were download from NCBI Sequence Read Archive collection PRJNA490626 [22]. Twenty three tissues or development stages including root, stem, leaf, petal, torus, sepal, bract, anther, filament, pistil, ovule and fiber tissues at -3, 0, 1, 3, 5 days post anthesis (DPA), ovules at 10, 15, 20, and 25 DPA, and fiber tissues at 10, 15, 20, and 25 DPA, were selected to identify the expression pattern by calculating Z-score. For investigating the genes related to stress tolerance, the transcriptome profiles of TM-1 roots and leaves treated by salt and control were downloaded from NCBI Sequence Read Archive collection PRJNA532694 and PRJNA490626, respectively. Both of the RNA-seq reads were mapped to the *G. hirsutum* acc. TM-1 genome using a Tophat spliced aligner with default parameters [60]. The genome-matched reads from each library were assembled with Cufflinks [61]. Cuffmerge was then used to merge the individual transcript assemblies into a single transcript set. Lastly, Cuffdiff was used to detect differentially expressed genes (DEGs) with a cutoff of 0.05 q-value. Three biological replicates from each sample were used for RNA-seq experiments. The heatmap of predominant expressed genes and salt stress response genes were performed with Mev software (<http://mev.tm4.org>).

Abbreviations

BLUPs: The best linear unbiased predictors; BNPP: Boll number per plant; CV: Coefficients of variation; GBS: Genotyping by sequencing; GO: Gene ontology; GWAS: Genome-wide association studies; LD: Linkage disequilibrium; LP: Lint percentage; LYPP: Lint yield per plant; MAF: Minor allele frequency; QTLs: Quantitative trait loci; SBW: Single boll weight; SNP: Single nucleotide polymorphism; SSR: Simple-sequence repeat.

Declarations

Acknowledgements

We thank the Dr. Waqas Amjid from College of Agriculture in Nanjing Agricultural University for helpful editing and comments.

Authors' contributions

Experiments were designed by WZG and QJC. Experiments were performed by GZZ, WWG, XHS, FLS, SH, NL, YJH, DYZ and ZYN. GZZ and WZG drafted the manuscript, WZG revised the manuscript. All authors read and approved the final manuscript.

Funding

This program was financially supported in part by National Key R & D Program for Crop Breeding (2018YFD0100400), National Natural Science Foundation in China (31771885), the Fundamental Research Funds for the Central Universities (KYYJ201701; KYYJ201801), Jiangsu Collaborative Innovation Center for Modern Crop Production project (No.10), and Tianshan Youth Program in Xinjiang Uygur Autonomous Region (2018Q016).

Availability of data and materials

RNA-Seq data in this study have been deposited at the National Center of Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) under the accessions PRJNA490626, PRJNA532694 and PRJNA490626.

Ethical approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Competing interests

The authors declared that they had no competing interests.

Author details

¹State Key Laboratory of Crop Genetics and Germplasm Enhancement, Engineering Research Center of Hybrid Cotton Development (the Ministry of Education), Nanjing Agricultural University, Nanjing 210095, China. ²Engineering Research Center for Cotton (the Ministry of Education), Xinjiang Agricultural University, Nanjing 210095, China.

References

1. Zorb C, Geilfus CM, Dietz KJ. Salinity and crop yield. *Plant Biol.* 2019;21(S1):31-38.
2. Dagar JC, Minhas PS. Global perspectives on agroforestry for the management of salt-affected soils. In: Dagar J, Minhas P, editors. *Agroforestry for the management of waterlogged saline soils and poor-quality waters.* New Delhi: Springer India; 2016. p. 5-32.

3. Han J, Shi J, Zeng L, Xu J, Wu L. Effects of nitrogen fertilization on the acidity and salinity of greenhouse soils. *Environ Sci Pollut Res Int*. 2015;22(4):2976-2986.
4. Du L, Cai C, Wu S, Zhang F, Hou S, Guo W. Evaluation and exploration of favorable QTL alleles for salt stress related traits in cotton cultivars (*hirsutum* L.). *PloS One*. 2016;11(3):e0151076.
5. Li Z, Li L, Zhou K, Zhang Y, Han X, Din Y, Ge X, Qin W, Wang P, Li F et al. *GhWRKY6* acts as a negative regulator in both transgenic *Arabidopsis* and cotton during drought and salt stress. *Front Genet*. 2019;10:392.
6. Zhang B, Chen X, Lu X, Shu N, Wang X, Yang X, Wang S, Wang J, Guo L, Wang D et al. Transcriptome analysis of *Gossypium hirsutum* reveals different mechanisms among NaCl, NaOH and Na₂CO₃ stress tolerance. *Sci Rep*. 2018;8:13527.
7. Sun ZW, Li HL, Zhang Y, Li ZK, Ke HF, Wu LQ, Zhang GY, Wang XF, Ma ZY. Identification of SNPs and candidate genes associated with salt tolerance at the seedling stage in cotton (*Gossypium hirsutum*). *Front Plant Sci*. 2018;9:1011.
8. Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, Wang X, Jin M, Li W, Zhang Q et al. Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. *Mol Plant*. 2017;10(3):414-426.
9. Shi YY, Gao LL, Wu ZC, Zhang XJ, Wang MM, Zhang CS, Zhang F, Zhou YL, Li ZK. Genome-wide association study of salt tolerance at the seed germination stage in rice. *BMC Plant Biol*. 2017;17(1):92.
10. Satir O, Berberoglu S. Crop yield prediction under soil salinity using satellite derived vegetation indices. *Field Crop Res*. 2016;192:134-143.
11. Fang L, Wang Q, Hu Y, Jia Y, Chen J, Liu B, Zhang Z, Guan X, Chen S, Zhou B et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet*. 2017;49(7):1089-1098.
12. Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat Genet*. 2018;50(6):803-813.
13. Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat Genet*. 2017;49(4):579-587.
14. Yuan Y, Wang X, Wang L, Xing H, Wang Q, Saeed M, Tao J, Feng W, Zhang G, Song XL et al. Genome-wide association study identifies candidate genes related to seed oil composition and protein content in *Gossypium hirsutum* *Front Plant Sci*. 2018;9:1359.
15. Diouf L, Pan ZE, He SP, Gong WF, Jia YH, Magwanga RO, Romy KRE, Rashid HO, Kirungu JN, Du XM. High-density linkage map construction and mapping of salt-tolerant QTLs at seedling stage in upland cotton using genotyping by sequencing (GBS). *Int J Mol Sci*. 2017;18(12):2622.
16. Dilnur T, Peng Z, Pan Z, Palanga KK, Jia Y, Gong W, Du X. Association analysis of salt tolerance in asiatic cotton (*Gossypium arboreum*) with SNP markers. *Int J Mol Sci*. 2019;20(9):2168.

17. Hou S, Zhu G, Li Y, Li W, Fu J, Niu E, Li L, Zhang D, Guo W. Genome-wide association studies reveal genetic variation and candidate genes of drought stress related traits in cotton (*Gossypium hirsutum*). *Front Plant Sci.* 2018;9:1276.
18. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, Zhang J, Dunwell JM, Xu S, Zhang YM. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci Rep.* 2016;6:19444.
19. Song C, Li W, Pei X, Liu Y, Ren Z, He K, Zhang F, Sun K, Zhou X, Ma X et al. Dissection of the genetic variation and candidate genes of lint percentage by a genome-wide association study in upland cotton. *Theor Appl Genet.* 2019;132(7):1991-2002.
20. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* acc. TM-1) provides a resource for fiber improvement. *Nat Biotech.* 2015;33(5):531-537.
21. Wang F, Zhang J, Chen Y, Zhang C, Gong J, Song Z, Zhou J, Wang J, Zhao C, Jiao M et al. Identification of candidate genes for key fibre-related QTLs and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *barbadense* introgressions. *Plant Biotechnol J.* 2019.
22. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet.* 2019;51(4):739-748.
23. Benkova E, Michniewicz M, Sauer M, Teichmann T, Seifertova D, Jurgens G, Friml J. Local, efflux-dependent auxin gradients as a common module for plant organ formation. 2003;115(5):591-602.
24. Dubrovsky JG, Sauer M, Napsucialy-Mendivil S, Ivanchenko MG, Friml J, Shishkova S, Celenza J, Benkova E. Auxin acts as a local morphogenetic trigger to specify lateral root founder cells. *Proc Natl Acad Sci USA.* 2008;105(25):8790-8794.
25. Barbez E, Kubes M, Rolcik J, Beziat C, Pencik A, Wang B, Rosquete MR, Zhu J, Dobrev PI, Lee Y et al. A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. 2012;485(7396):119-122.
26. Zhang M, Zeng JY, Long H, Xiao YH, Yan XY, Pei Y. Auxin regulates cotton fiber initiation via GhPIN-mediated auxin transport. *Plant Cell Physiol.* 2017;58(2):385-397.
27. Peng J, Ilarslan H, Wurtele ES, Bassham DC. AtRabD2b and AtRabD2c have overlapping functions in pollen development and pollen tube growth. *BMC Plant Biol.* 2011; 11:25.
28. Blanco FA, Meschini EP, Zanetti ME, Aguilar OM. A small GTPase of the Rab family is required for root hair formation and preinfection stages of the common bean-Rhizobium symbiotic association. *Plant Cell.* 2009;21(19):2797-2810.
29. Li P, Guo W. Genome-wide characterization of the Rab gene family in *Gossypium* by comparative analysis. *Bot Stud.* 2017;58(1):26.
30. Alvim Kamei CL, Boruc J, Vandepoele K, Van den Daele H, Maes S, Russinova E, Inze D, De Veylder L. The *PRA1* gene family in *Plant Physiol.* 2008; 147(4):1735-1749.

31. Guerriero G, Fugelstad J, Bulone V. What do we really know about cellulose biosynthesis in higher plants? *J Integr Plant Biol.* 2010;52(2):161-175.
32. Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Li H, Zhang HQ, Zhang S, Wang DM, Wang QX et al. Transcription factor *AtMYB103* is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *Plant J.* 2007;52(3):528-538.
33. Dahlke RI, Fraas S, Ullrich KK, Heinemann K, Romeiks M, Rickmeyer T, Klebe G, Palme K, Luthen H, Steffens B. Protoplast swelling and hypocotyl growth depend on different auxin signaling pathways. *Plant Physiol.* 2017;175(2):982-994.
34. Xiao G, Zhao P, Zhang Y. A pivotal role of hormones in regulating cotton fiber development. *Front Plant Sci.* 2019;10:87.
35. An D, Chen JG, Gao YQ, Li X, Chao ZF, Chen ZR, Li QQ, Han ML, Wang YL, Wang YF et al. *AtHKT1* drives adaptation of *Arabidopsis thaliana* to salinity by reducing floral sodium content. *PLoS Genet.* 2017;13(10):e1007086.
36. Ibrahim W, Qiu CW, Zhang C, Cao F, Shuijin Z, Wu F. Comparative physiological analysis in the tolerance to salinity and drought individual and combination in two cotton genotypes with contrasting salt tolerance. *Physiol Plantarum.* 2019;165(2):155-168.
37. Zhang J, Zou D, Li Y, Sun X, Wang NN, Gong SY, Zheng Y, Li XB. GhMPK17, a cotton mitogen-activated protein kinase, is involved in plant response to high salinity and osmotic stresses and ABA signaling. *PloS One.* 2014;9(4):e95642.
38. Inze D, De Veylder L. Cell cycle regulation in plant development. *Annu Rev Genet.* 2006;40:77-105.
39. Rizzi YS, Cecchini NM, Fabro G, Alvarez ME. Differential control and function of *Arabidopsis ProDH1* and *ProDH2* genes on infection with biotrophic and necrotrophic pathogens. *Mol Plant Pathol.* 2017;18(8):1164-1174.
40. Delauney AJ, Verma DPS. Proline biosynthesis and osmoregulation in plants. *Plant J.* 1993;4(2):215-223.
41. Funck D, Eckard S, Muller G. Non-redundant functions of two proline dehydrogenase isoforms in *Arabidopsis*. *BMC Plant Biol.* 2010;10:70.
42. Delatte T, Umhang M, Trevisan M, Eicke S, Thorneycroft D, Smith SM, Zeeman SC. Evidence for distinct mechanisms of starch granule breakdown in plants. *J Biol Chem.* 2006;281(17):12050-12059.
43. Wang Y, Lu J, Chen D, Zhang J, Qi K, Cheng R, Zhang H, Zhang S. Genome-wide identification, evolution, and expression analysis of the KT/HAK/KUP family in pear. 2018;61(10):755-765.
44. Brauer EK, Ahsan N, Dale R, Kato N, Coluccio AE, Pineros MA, Kochian LV, Thelen JJ, Popescu SC. The Raf-like kinase *ILK1* and the high affinity K⁺ Transporter *HAK5* are required for innate immunity and abiotic stress response. *Plant Physiol.* 2016;171(2):1470-1484.
45. Barragan V, Leidi EO, Andres Z, Rubio L, De Luca A, Fernandez JA, Cubero B, Pardo JM. Ion exchangers NHX1 and NHX2 mediate active potassium uptake into vacuoles to regulate cell turgor

- and stomatal function in *Arabidopsis*. *Plant Cell*. 2012;24(3):1127-1142.
46. Ibrahim W, Zhu YM, Chen Y, Qiu CW, Zhu S, Wu F. Genotypic differences in leaf secondary metabolism, plant hormones and yield under alone and combined stress of drought and salinity in cotton genotypes. *Physiol Plantarum*. 2019;165(2):343-355.
 47. Hussein MM, Abou-Baker NH. The contribution of nano-zinc to alleviate salinity stress on cotton plants. *Roy Soc Open Sci*. 2018;5(8):171809.
 48. Peng J, Liu J, Zhang L, Luo J, Dong H, Ma Y, Zhao X, Chen B, Sui N, Zhou Z et al. Effects of soil salinity on sucrose metabolism in cotton leaves. *PloS One*. 2016;11(5):e0156241.
 49. Su J, Fan S, Li L, Wei H, Wang C, Wang H, Song M, Zhang C, Gu L, Zhao S et al. Detection of favorable QTL alleles and candidate genes for lint percentage by GWAS in Chinese upland cotton. *Front Plant Sci*. 2016;7:1576.
 50. Szumlanski AL, Nielsen E. The Rab GTPase RabA4d regulates pollen tube tip growth in *Arabidopsis thaliana*. *Plant Cell*. 2009;21(2):526-544.
 51. Sun WJ, Gao ZY, Wang J, Huang YQ, Chen Y, Li JF, Lv ML, Wang J, Luo M, Zuo KJ. Cotton fiber elongation requires the transcription factor *GhMYB212* to regulate sucrose transportation into expanding fibers. *New Phytol*. 2019;222(2):864-881.
 52. Chen J, Wan S, Liu H, Fan S, Zhang Y, Wang W, Xia M, Yuan R, Deng F, Shen F. Overexpression of an apocynum venetum DEAD-Box helicase gene (*AvDHT1*) in Cotton confers salinity tolerance and increases yield in a saline field. *Front Plant Sci*. 2015;6:1227.
 53. Moller IS, Gilliham M, Jha D, Mayo GM, Roy SJ, Coates JC, Haseloff J, Tester M. Shoot Na⁺ exclusion and increased salinity tolerance engineered by cell type-specific alteration of Na⁺ transport in *Arabidopsis*. *Plant Cell*. 2009;21(7):2163-2178.
 54. Zhu G, Li W, Zhang F, Guo W. RNA-seq analysis reveals alternative splicing under salt stress in cotton, *Gossypium davidsonii*. *BMC Genomics*. 2018;19(1):73.
 55. Munns R, Gilliham M. Salinity tolerance of crops - what is the cost? *New Phytol*. 2015;208(3):668-673.
 56. Liu G, Li X, Jin S, Liu X, Zhu L, Nie Y, Zhang X. Overexpression of rice NAC gene *SNAC1* improves drought and salt tolerance by enhancing root development and reducing transpiration rate in transgenic cotton. *PloS One*. 2014;9(1):e86895.
 57. Bates D, Machler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1-48.
 58. Paterson AH, Brubaker CL, Wendel JF. A rapid method for extraction of cotton (*Gossypium*) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol Biol Rep*. 1993;11(12):122-127.
 59. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. AgriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res*. 2017;45:W122-W129.
 60. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. 2009;25(9):1105-1111.

61. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 2012;7(3):562-578.

Table

Table 1 GWAS analysis of three traits in four salt conditions by multi-loci MLM model.

Trait	Salt condition	Number of stable/total QTL	Count of stable/total QTL*
SBW	A	21/49	69/103
	B	12/39	56/90
	C	13/40	48/76
	D	18/47	50/87
	Total	42/151	223/356
LP	A	49/136	113/209
	B	56/188	131/276
	C	57/165	110/224
	D	50/106	97/159
	Total	91/417	451/868
BNPP	A	7/30	25/53
	B	9/31	32/60
	C	8/24	29/50
	D	8/34	27/59
	Total	25/112	113/222
Total		150/600	787/1446

* indicated the counts of QTLs detected in different methods, years or replications. SBW: single boll weight; LP: lint percentage; BNPP: boll number per plant. A, B, C, and D represented the total salt contents in the four soil environments, with 19 g/Kg (condition A), 10 g/Kg (condition B), 7 g/Kg (condition C), and 5 g/Kg (condition D), respectively.

Additional Files

Additional file 1: Table S1. Information on 316 cotton accessions used in this study. (XLSX 20 kb)

Additional file 2: Figure S1. Sketch map of soil salt concentration and distribution of the phenotypic data for three lint yield components. A: Sketch map of soil salt distribution with different total salt content. The experimental field was divided into eight parts including four different total salt content with two

replications for each salt condition. B-D: Distribution of phenotypic data of single boll weight (B), lint percentage (C) and boll number per plant (D) under four salt conditions. (TIFF 469 kb)

Additional file 3: Table S2. Measurement of total soil salt content in four different environments. (XLSX 11 kb)

Additional file 4: Table S3. Phenotypic data statistics of single boll weight under four salt conditions. (XLSX 10 kb)

Additional file 5: Table S4. Phenotypic data statistics of lint percentage under four salt conditions. (XLSX 10 kb)

Additional file 6: Table S5. Phenotypic data statistics of boll number per plant under four salt conditions. (XLSX 10 kb)

Additional file 7: Figure S2. Correlation analysis among three lint yield components and in different salt conditions for each trait. The red boxes indicated the correlation among three lint yield components. The green boxes indicated the correlation among different salt conditions for each trait. The number in these boxes indicated correlation coefficient (R value). *, **, and *** indicated P value at the 0.05, 0.01 and 0.001 levels, respectively. (TIFF 2,605 kb)

Additional file 8: Table S6. QTNs and QTLs of single boll weight, lint percentage and boll number per plant detected by multi-loci MLM model. (XLSX 151 kb)

Additional file 9: Figure S3. Distribution on detected times for 600 associated QTLs from three lint yield components, respectively. The x-axis represents the detected times; y-axis represents the number of QTLs. (TIFF 216 kb)

Additional file 10: Figure S4. The enriched biological processes of candidate genes associated with single boll weight. (TIFF 230 kb)

Additional file 11: Table S7. GO enrichment of genes associated with the three traits. (XLSX 21 kb)

Additional file 12: Figure S5. The enriched biological processes of candidate genes associated with lint percentage. (TIFF 552 kb)

Additional file 13: Figure S6. The enriched biological processes of candidate genes associated with boll number per plant. (TIFF 430 kb)

Additional file 14: Figure S7. Heatmap of predominantly expressed genes associated with lint percentage in cotton fiber development. The number indicated different tissues or development stages, 1: root; 2: stem; 3: leaf; 4: petal; 5: torus; 6: sepal; 7: bract; 8: anther; 9: filament; 10: pistil; 11: -3DPA ovule and fiber; 12: 0DPA ovule and fiber; 13: 1DPA ovule and fiber; 14: 3DPA ovule and fiber; 15: 5DPA ovule and fiber; 16:

10DPA ovule; 17: 15DPA ovule; 18: 20DPA ovule; 19: 25DPA ovule; 20: 10DPA fiber; 21: 15DPA fiber; 22: 20DPA fiber; 23: 25DPA fiber. (TIFF 586 kb)

Additional file 15: Table S8. Expression pattern of 182 predominant expressed genes during fiber development in different tissues and organs. (XLSX 64 kb)

Additional file 16: Table S9. Candidate genes related to lint percentage identified in stable QTLs. (XLSX 11 kb)

Additional file 17: Figure S8. Heatmap of candidate genes related to boll number per plant under salt stress. The stress response genes located in QTLs were salt-inducible in roots (A) and leaves (B) under salt stress. The gene names marked in red represent differential expression in both roots and leaves. (TIFF 639 kb)

Additional file 18: Table S10. Expression pattern of 19 candidate genes related to boll number per plant under salt stress in roots. (XLSX 13 kb)

Additional file 19: Table S11. Expression pattern of 27 candidate genes related to boll number per plant under salt stress in leaves. (XLSX 13 kb)

Additional file 20: Table S12. GO enrichment of genes associated with boll number per plant under salt condition A and D. (XLSX 12 kb)

Figures

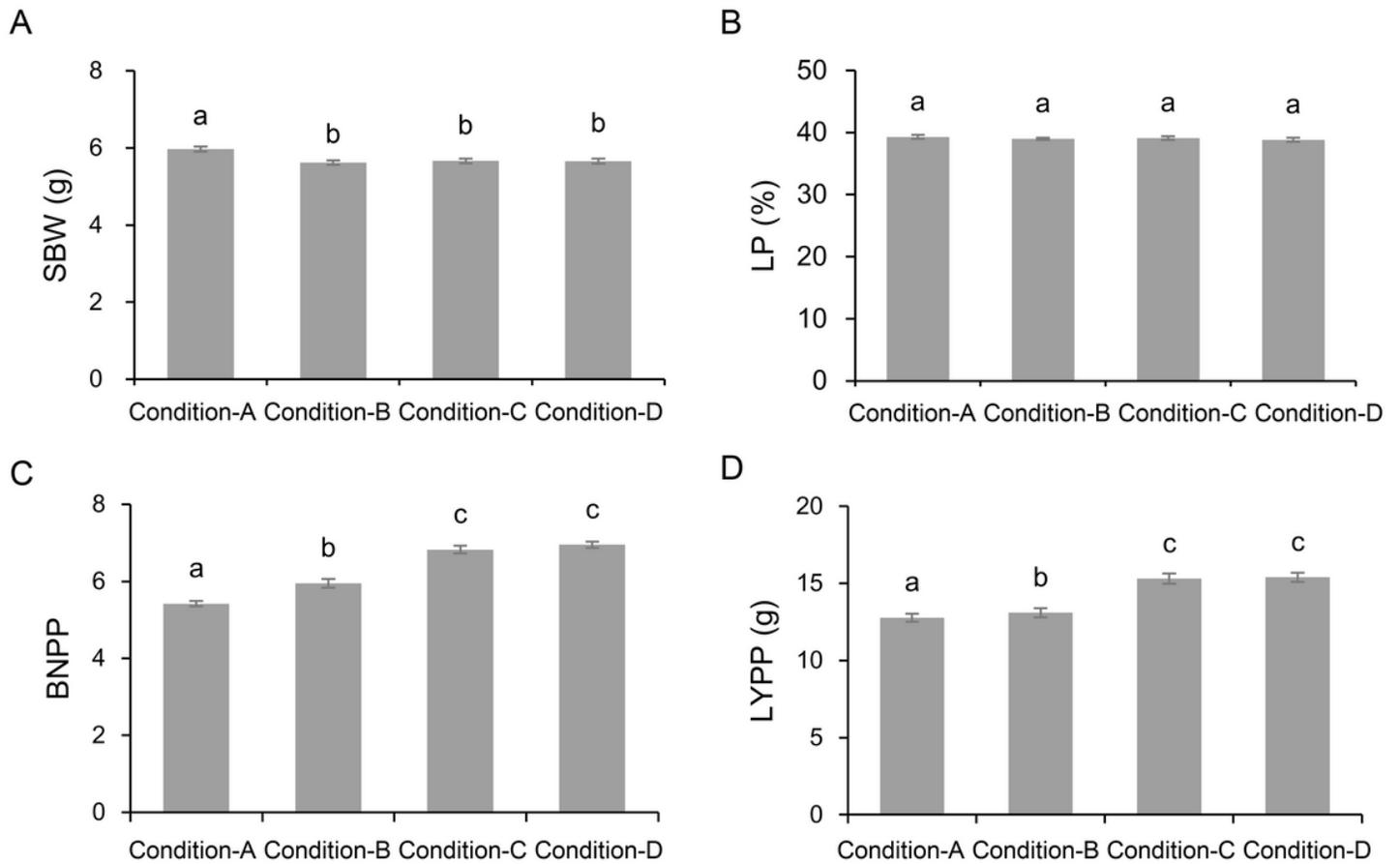


Figure 1

Comparison of different phenotypic data under four salt conditions. A. single boll weight (SBW). B. lint percentage (LP). C. boll number per plant (BNPP) D. lint yield per plant (LYPP). Significant difference of single trait was calculated with paired-samples t-tests.

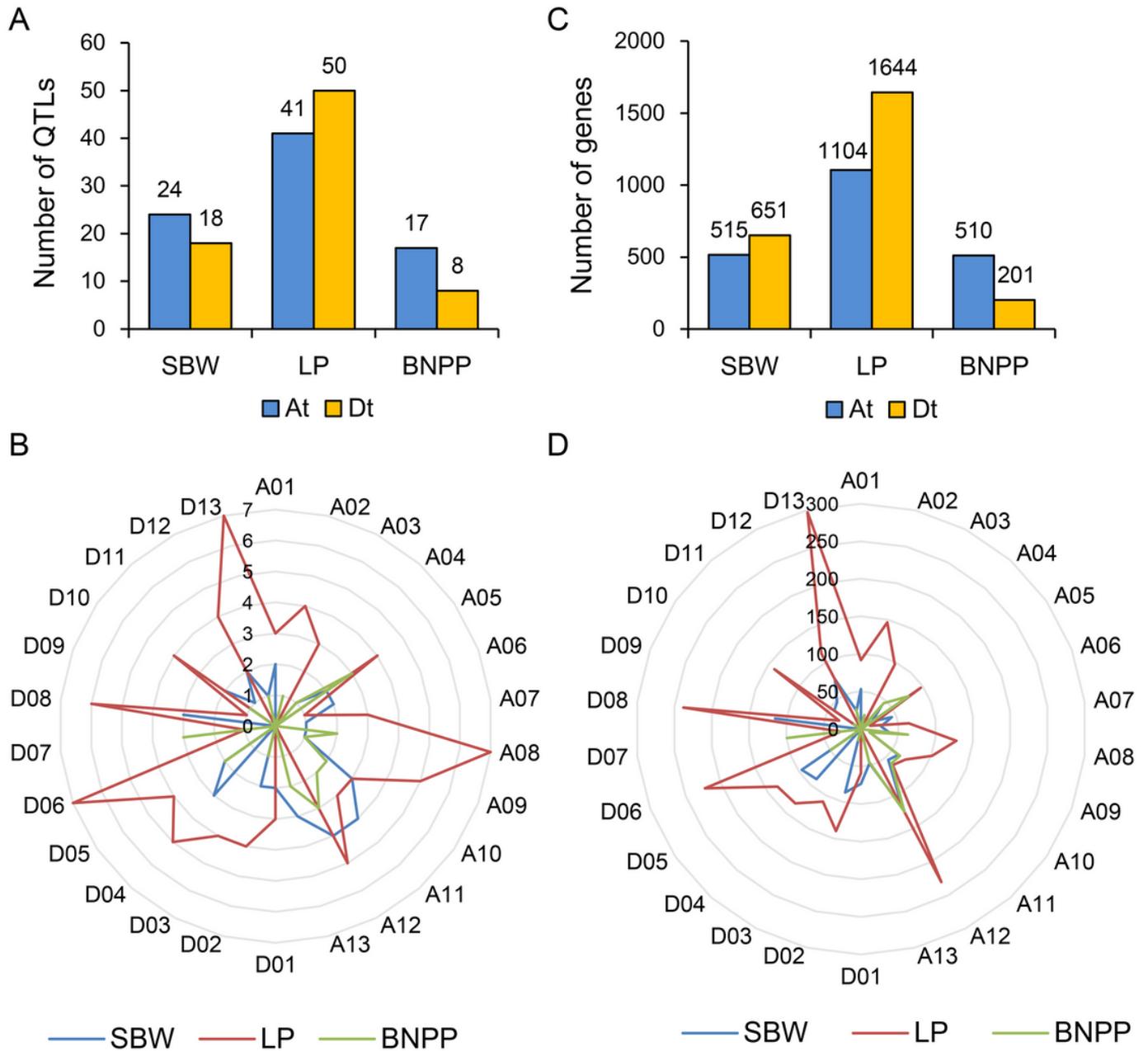


Figure 2

Genomic distribution of QTLs and candidate genes associated with the three traits. A. Numbers of QTLs on At and Dt sub-genome. B. Numbers of QTLs on 26 chromosomes. C. Numbers of candidate genes on At and Dt sub-genome. D. Numbers of candidate genes on 26 chromosomes.

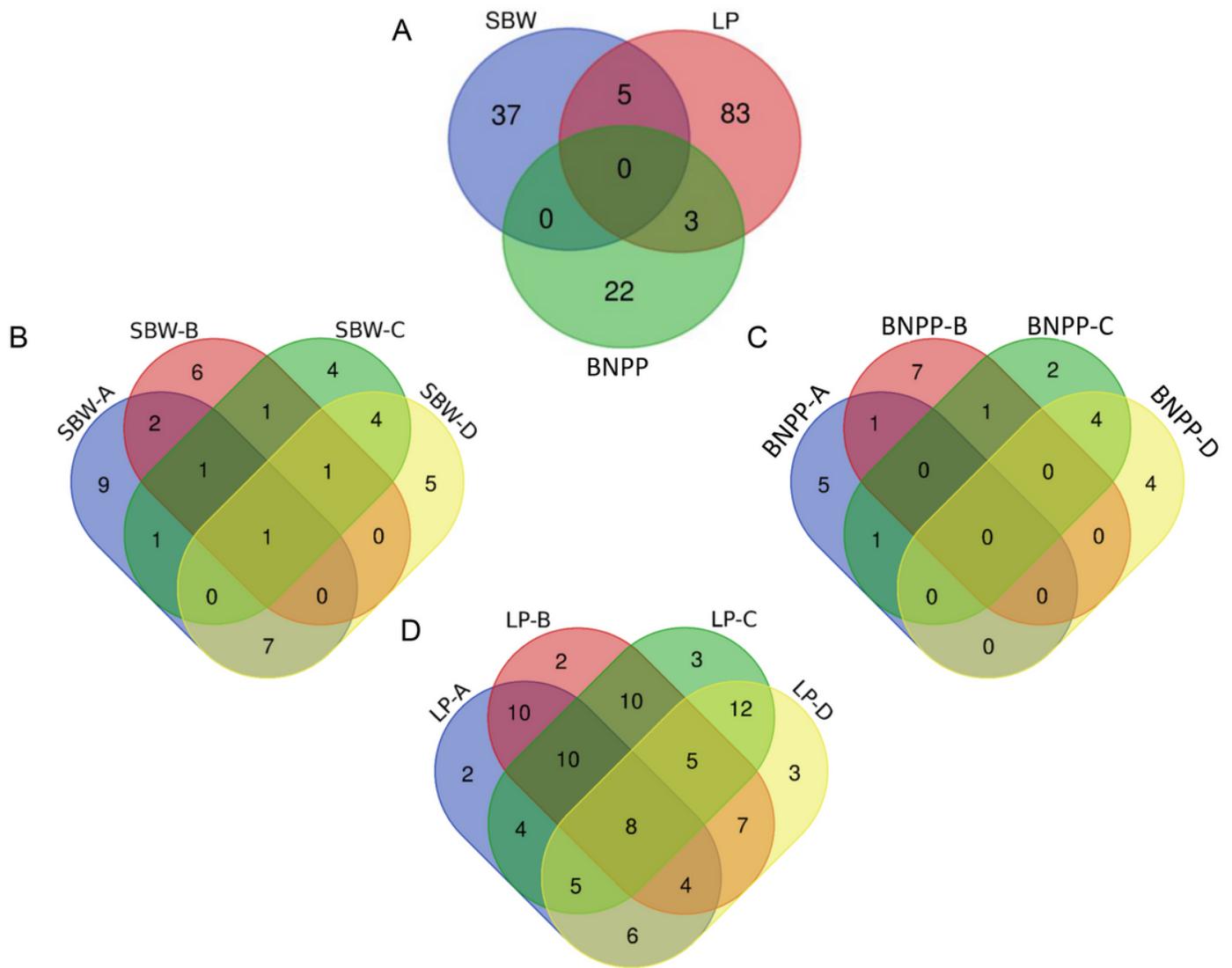


Figure 3

Venn diagram of QTLs associated with three traits under four different salt conditions. A. Venn diagram of QTLs among three traits. B. Venn diagram of QTLs associated with SBW in four salt conditions. C. Venn diagram of QTLs associated with LP in four salt conditions. D. Venn diagram of QTLs associated with BNPP in four salt conditions.

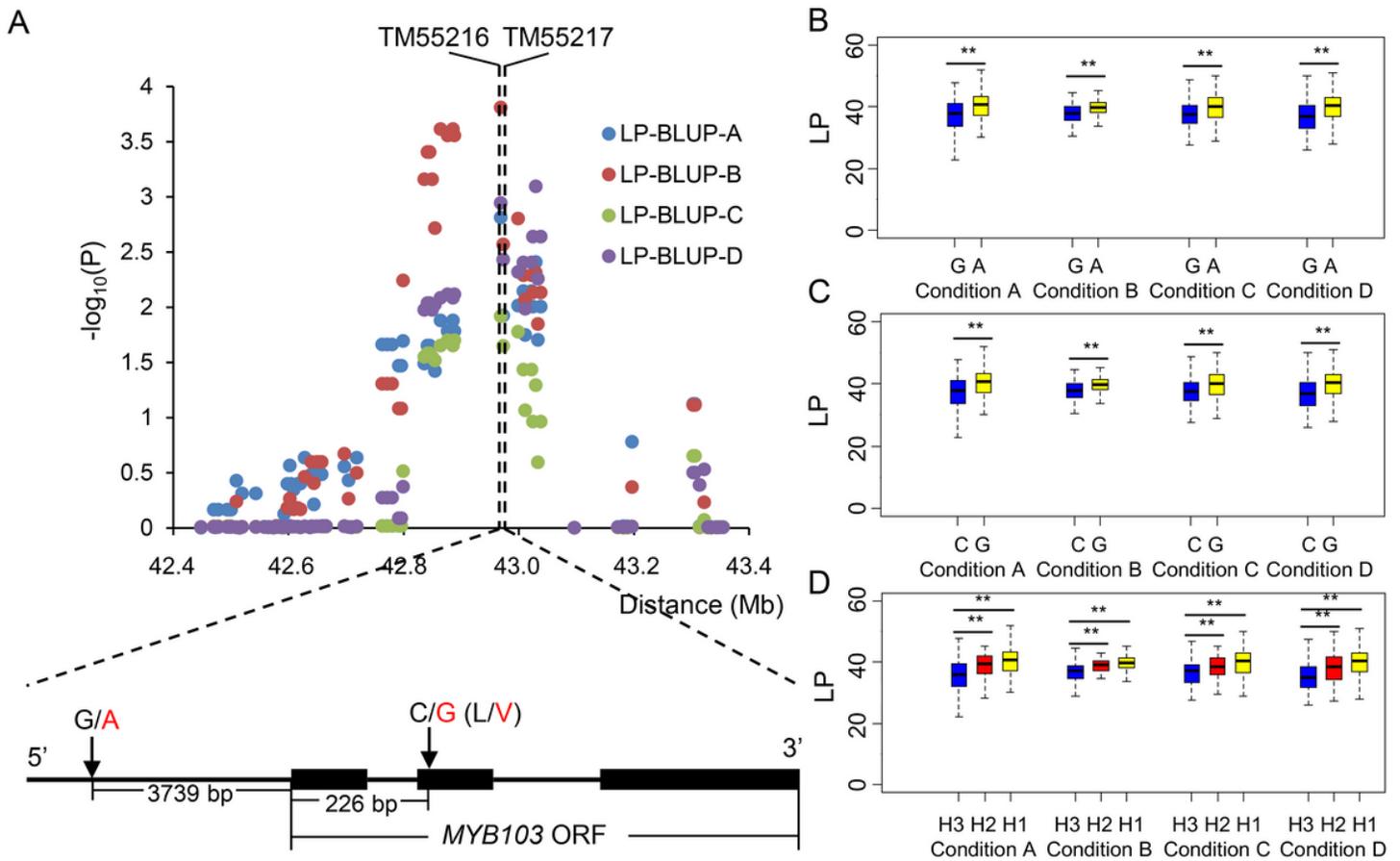


Figure 4

Haplotype analysis of the candidate gene MYB103 on chromosome D03. A. Manhattan plots of SNPs around MYB103 for the best linear unbiased prediction (BLUP) of LP across the four different salt conditions and the location of two QTNs related to MYB103. L indicated leucine and V indicated valine. B. Box plots for the phenotypic values of QTN TM55216. C. Box plots for the phenotypic values of QTN TM55217. D. Box plots for the phenotypic values of haplotype from the two QTN combinations. H1 indicated A-G genotype, H2 indicated A-C genotype, and H3 indicated G-C genotype. ** indicated P value at the 0.01 level with student's t test, respectively.

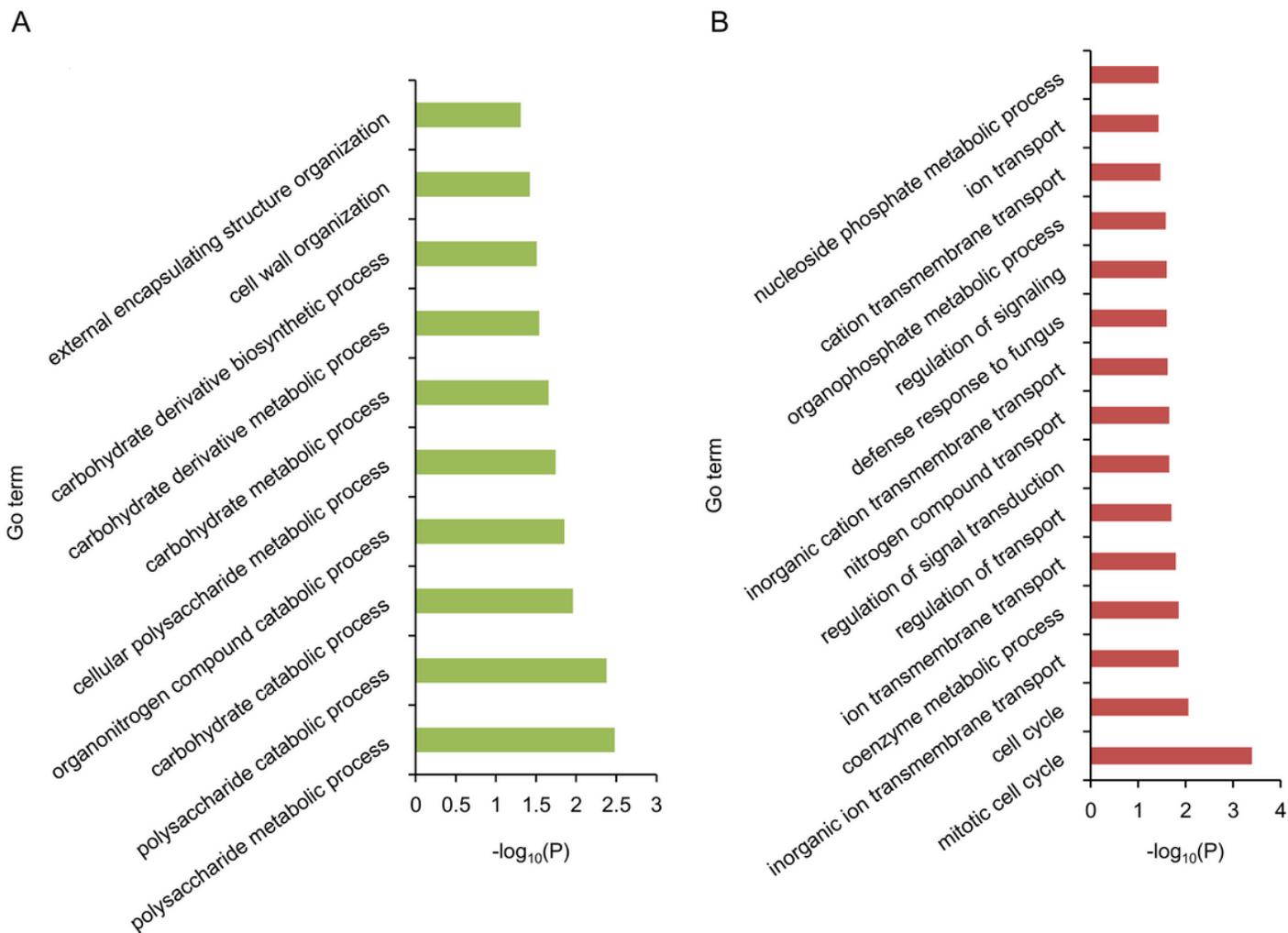


Figure 5

enrichment of genes associated with BNPP under salt condition A and D. A. salt condition A. B. salt condition D.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile17Fig.S8.tif](#)
- [Additionalfile9Fig.S3.tif](#)
- [Additionalfile7Fig.S2.tif](#)
- [Additionalfile19TableS11.xlsx](#)
- [Additionalfile20TableS12.xlsx](#)
- [Additionalfile15TableS8.xlsx](#)
- [Additionalfile1TableS1.xlsx](#)

- [Additionalfile3TableS2.xlsx](#)
- [Additionalfile5TableS4.xlsx](#)
- [Additionalfile4TableS3.xlsx](#)
- [Additionalfile14Fig.S7.tif](#)
- [Additionalfile13Fig.S6.tif](#)
- [Additionalfile8TableS6.xlsx](#)
- [Additionalfile11TableS7.xlsx](#)
- [Additionalfile2Fig.S1.tif](#)
- [Additionalfile18TableS10.xlsx](#)
- [Additionalfile16TableS9.xlsx](#)
- [Additionalfile10Fig.S4.tif](#)
- [Additionalfile12Fig.S5.tif](#)
- [Additionalfile6TableS5.xlsx](#)