Formation and Comparative Analysis of Full-Length Transcriptome Sequencing and Next Generation Sequencing In *Medicago Sativa* L. Roots Under Abiotic Stress

Zhi-hong Fang
Sciences of ShanXi Agricultural Academy Institute of Animal Husbandry Veterinary, TaiYuan, 030032, ShanXi, China

Yonghong Shi (shiyh007@126.com)
Sciences of ShanXi Agriculture Academy Institute of Animal Husbandry Veterinary, TaiYuan, 030032, ShanXi, China

Xin-ming Wu
Sciences of ShanXi Agricultural Academy Institute of Animal Husbandry Veterinary, TaiYuan, 030032, ShanXi, China

Bin-lin Ren
Sciences of ShanXi Agricultural Academy Institute of Animal Husbandry Veterinary, TaiYuan, 030032, ShanXi, China

Yan Zhang
Sciences of ShanXi Agricultural Academy Institute of Animal Husbandry Veterinary, TaiYuan, 030032, China

Pu Guo
Sciences of ShanXi Agricultural Academy Institute of Animal Husbandry Veterinary, TaiYuan, 030032, China

Hui-wu Chi
Sciences of ShanXi agriculteral Academy Institute of Animal Husbandry Veterinary, TaiYuan, 030032, China

Jian-ning Liu
Sciences of ShanXi Agricultural Academy Institute of Animal Husbandry Veterinary

Yun-qi Wang
Sciences of ShanXi agricultural Academy Institute of Animal Husbandry Veterinary

Research article

Keywords: Medicago sativa L., full-length transcripts, single-molecule real-time sequencing, next generation sequencing, abiotic stress,

Posted Date: January 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-144129/v1

Abstract

Background: *Medicago sativa* L. (*M. sativa* L.) is a legume with high salt tolerance and a major forage crop with high biomass production. However, the large-scale full-length cDNA sequences of *M. sativa* L. in response to abiotic stress remain unclear.

Results: We provided the complete transcriptome for *M. sativa* L. roots under different abiotic stressors using a combination of single-molecule real-time sequencing and next generation sequencing. Our results indicated that there were 21.53 Gb clean reads, which consisted of 566,076 insert reads and 409,291 full-length non-chimeric reads. We obtained 194,286 consistent transcripts based on a cluster analysis of full-length reads, and 41,248 high quality transcript sequences based on non-full-length reads. After correction using second-generation data for third-generation low-quality data, we obtained 81,017 transcript sequences according to a cogent analysis. The sequence structural analysis acquired 33,058 simple sequence repeats and 42,725 complete coding sequence regions. In addition, 77,221 transcripts were annotated by eight functional databases; 3,043 lncRNAs were predicted and 4,971 alternative splicings were acquired. Moreover, we confirmed the levels of highly differentially expressed transcripts (ADH1, PEPC, MJG19.6, PCKA and GAPC1) in *M. sativa* L. roots under NaCl and polyethylene glycol stress.

Conclusions: Therefore, we fully and massively exposed the full-length transcripts related to abiotic stress in *M. sativa* L., which will lay the foundation for understanding gene regulation in *M. sativa* L. under abiotic stress.

Background

Medicago sativa L. (*M. sativa* L., Alfalfa) is a significant leguminous forage crop worldwide and a highly productive perennial forage species[1]. *M. sativa* L. is also an artificial forage species with the largest planting area and the widest distribution in China[2]. *M. sativa* L. is very adaptable, rich in nutrition, and has excellent quality[3]. Its stems and leaves are rich in proteins, minerals, vitamins, and carotene, and it is an important source of protein feed and forage for various livestock[4]. Moreover, *M. sativa* L. plays an important role improving soil, fertilizer, and soil and water conservation[5]. *M. sativa* L. is also known as the “king of herbage”[6]. Besides, the life span of *M. sativa* L. can be up to 30 years, and field cultivation is up to 7–10 years, with strong reproducibility[7]. However, growing researches have demonstrated that the production and quality of *M. sativa L*. were severely restricted by abiotic stress. Therefore, in-depth understanding of *M. sativa L.* resistance to abiotic stress is of great significance for the improvement of *M. sativa L.* yield.

Abiotic stresses, such as drought, high salt, darkness, low temperature, phytohormone, are common environmental factors in nature that severely restrict the growth and development of crops[8]. Currently, the annual yield loss of major crops due to abiotic stress is over 50% on average worldwide[9]. Studies have demonstrated that abiotic stress can dehydrate cells, destroy the cell membrane system, affect the

enzyme activity on the cell membrane, and disrupt metabolism[10, 11]. Besides, abiotic stress can reduce the photosynthetic rate, decrease the assimilation product, and change the respiration rate of crops[12, 13]. For example, salt stress (high concentration of NaCl) will cause a loss of water in plants, ion toxicity, nutrient imbalance, oxidative stress, and so on[14]. Drought stress is a vital stress factor affecting plant photosynthesis and growth[15]. Drought stress can indirectly lead to the occurrence of leaf photoinhibition, reduce photosynthetic efficiency, and cause serious damage to crops[16]. Dark stress can result in the complex characters including leaf senescence, the elongation of the hypocotyl and petiole, and early flowering[17]. Temperature is also a crucial ecological factor that restricts the growth and distribution of plant, and low temperature can markedly destroy the cell membrane system and reduce SOD activity[18]. Abscisic acid (ABA), as a kind of plant hormone, can alleviate the peroxidation of chlorophyll in plant leaves, contribute to the synthesis of chlorophyll[19]. At present, the effects of abiotic stress on crops have become a bottleneck in agricultural development in numerous areas[20]. Therefore, it is of significance to improve the tolerance of crops to abiotic stress by studying stress resistance in *M. sativa* L. However, the genome and transcriptome data of *M. sativa* L. roots under abiotic stressors have not been reported, which greatly hampers study of the underlying molecular mechanisms of abiotic stress during growth and development of *M. sativa* L.

A transcriptome is the complete set of transcripts for a certain type of cells or tissues in a specific developmental stage or physiological condition[21, 22]. A transcriptome analysis reveals the gene expression levels of organisms as well as structural variations and can be used to discover new genes[23]. The research methods and platforms for transcriptomes are undergoing rapid changes and bioinformatics analysis has also gradually improved. The single-molecule real-time (SMRT) sequencing technology from Pacific Biosciences (PacBio) quickly and accurately provides transcriptional information for organisms[24, 25]. Therefore, this technology has become a better alternative for full-length cDNA molecular sequencing and has been widely applied for whole-transcriptome profiling of humans and other species. The rapid development of next generation sequencing (NGS) over the past few years has also increased data throughput and read length, and simultaneously brought down sequencing costs[26]. This has allowed for new breakthroughs in the area of biology and ushered medical genetics into a new era.

In the present study, we utilized SMRT sequencing to generate and identify the full-length transcriptome of *M. sativa* L. roots under various abiotic stressors, such as darkness, low temperature (4 °C), NaCl (400 mM), polyethylene glycol (PEG) (25%) and abscisic acid (ABA, 100 μM). According to the transcriptome data, we analyzed the full-length transcriptome sequences by alternative splicing, simple sequence repeat (SSR), transcription factor, complete coding sequence (CDS), single nucleotide polymorphism (SNP), and transcript functional annotation. We also predicted the coding sequences and long non-coding RNAs (lncRNAs). In addition, we combined PacBio sequencing and NGS technology to further identify the differentially expressed transcripts and analyze the functional annotation in *M. sativa* L. roots under CK (control), 400 mM NaCl, and 25% PEG stress conditions. The results will be valuable for further investigations of *M. sativa* L. roots under abiotic stress.

Results

Overview of full-length transcriptome sequencing in abiotic stress-treated M. sativa **L. roots.** In this study, *M. sativa* L. roots were grown under different abiotic stressors (darkness, low temperature, 400 mM NaCl, 25% PEG, and 100 μM ABA) for 0, 3, 6, 12, 24, 48, and 72 h (Fig. 1A). Then, we extracted total RNA from the *M. sativa* L. roots, which were subjected to different abiotic stressors, equally mixed the RNAs together. After purification and repair, we constructed and examined the library using Agilent 2100 system (Agilent Technologies, Palo Alto, CA, USA). A full-length transcriptome sequencing experiment was carried out using PacBio Sequel (Fig. 1B). The obtained accurate full-length transcripts were analyzed using PacBio Single Molecule Real Time (SMRT™) DNA sequencing technology (Fig. 1C). A total of 21.53 Gb clean reads were obtained and utilized to correct the SMRT reads in *M. sativa* L. roots under abiotic stress. Based on full passes ≥ 0 and sequence accuracy > 0.80, we successfully obtained 566,076 reads-of-insert (ROI) (total bases: 1,058,177,144) with mean length of 1,869 bp, mean read quality of 0.96, and 19 passes.

Summary of Pacbio RS sequencing base on sequencing by synthesis (SMRT). Besides, the ROI read length distribution of each size bin and the 1–6 k size bin were shown, and the ROI read length was mainly distributed around 2000 (Fig. 2A and 2B). The number of full passes is related to the length of the cDNA, but generally decreases with the increase in cDNA length. The accuracy of the ROI sequence is affected by the number of full passes. The more the full passes contribute the higher the sequence accuracy. ROI quality values reflect the accuracy of the sequence. The distribution of full passes of ROI sequences was shown in Fig. 2C, and the distribution of ROI quality values was high (Fig. 2D). Besides, through the full-length transcriptome sequences, we obtained 34,871 filtered short reads, 116,022 non-full-length reads, 415,183 full-length reads, and 409,291 full-length non-chimeric reads. The average full-length non-chimeric read length (FLNC) was 1,916; the full-length percentage was 73.34% (**Table S2**). The FLNC length distributions of each size bin and the 1–6 k size bin were what was expected, and the FLNC length also mainly distributed around 2000 (Fig. 2E and 2F). According to the ROI classification for the 1–6 k size, the percentage of filtered short reads was 6.2%; the percentage of full-length (chimeric) was 1%; the percentage of full-length (non-chimeric) was 73.2%; the percentage of non-full-length (no poly-A) was 11.6%; and the percentage of the non-full-length (no primer) was 8.9% (Fig. 2G).

Alternative splicing and SSR analyses and the prediction of coding sequences. In our results, there were 194,286 consensus isoforms (average consensus isoform read length of 2,077), 41,248 polished high-quality isoforms, 152,636 polished low-quality isoforms; the percent of polished high-quality isoforms was 21.23%. The consensus isoform length distribution of each size bin was exhibited, and the consensus isoform length was also mainly distributed around 2000 (Fig. 3A). In addition, BUSCO was used to evaluate transcriptome integrity, and the results indicated that the number of complete and single-copies was 418, the number of complete and duplicated was 536, the number of fragmented transcriptomes was 29, and the number of missing transcriptomes was 457 (Fig. 3B). Meanwhile, BLAST software was applied to predict the candidate events of alternative splicing through pairwise comparison of the unreferenced transcriptomes in the three generations of after redundantion[27]. If the comparison

results satisfy the conditions in the previous research, they are considered as candidate alternative splicing events. And we also exhibited the alternative splicing events in **Table S3**. Furthermore, we analyzed SSRs using the MIcroSAtellite identification tool (MISA). Here, the detection of SSRs was part of the genetic analysis for *M. sativa* L. roots under abiotic stress. And the distribution of SSR types was displayed. There were 3944 compound SSR (c), 35 compound SSR with overlapping bases between two SSRs (c*), 13133 Mono-nucleotide (p1), 3535 Di nucleotide (p2), 6825 Tri nucleotide (p3), 349 Tetra nucleotide (p4), 117 Penta nucleotide (p5) and 113 Hexa nucleotide (p6) (**Table S4**). The coding region sequences and the corresponding amino acid sequences of the transcripts were predicted using TransDecoder software, and we also successfully obtained a total of 75,596 ORFs, including 42,725 complete ORFs. The complete length distribution of the predicted CDS coding proteins was shown in Fig. 3C.

LncRNA prediction and transcription factor analysis. Meanwhile, the numbers of lncRNA transcripts were presented in Fig. 4A, which was predicted by CPC, CNCI, CPAT and a pfam protein structure domain analysis. We discovered 3043 lncRNAs through the conjoint analysis of the CPC, CNCI, CPAT and pfam. Next, the target genes of the predicted 3,043 lncRNA sequences were also predicted using LncTar, and the predicted target genes were exhibited in **Table S5**. In addition, the transcription factors were predicted and analyzed using iTAK software; 8,336 transcription factors were predicted, and the distribution of the different transcription factors was shown in Fig. 4B. We discovered that the transcription factors mainly contained C3H, C2H2, bZIP, bHLH, GRAS, WRKY, SNF2, AP2/ERF-ERF, MYB-related, PHD, etc.

Functional annotation of the transcripts. In our results, 77,221 transcripts were annotated, including 32,297 transcripts in the COG database, 55,044 in the GO, 34,349 in the KEGG, 48,647 in the KOG, 62,168 in the Pfam, 54,521 in the Swiss-Prot, 72,017 in the eggNOG and 76,356 in the NR (**Table S6**). And the integrated_function annotation of all identified transcripts was also displayed in **Table S7.** According to the NR database, in the homologous species distribution, *Medicago truncatula* accounted for 80.03%, *Fusarium verticilliides* accounted for 3.92%, *Cicer arietinum* accounted for 2.66%, and *Medicago sativa* accounted for 2.66% (Fig. 5A). The GO annotation system was divided into biological processes (BP), molecular functions (MF), and cellular components (CC). We discovered that these differentially expressed transcripts under abiotic stresses were mainly enriched in CC terms (cell part, cell, organelle, membrane, organelle part, membrane part, macromolecular complex and cell junction, etc), MF terms (catalytic activity, binding, transporter activity, structural molecule activity, nucleic acid binding transcription factor activity, electron carrier activity, molecular transducer activity and enzyme regulator activity, etc) and BP terms (metabolic process, cellular process, single-organism process, response to stimulus, biological regulation, localization and cellular component organization or biogenesis, etc) (Fig. 5B). The COG database can be utilized to directly classify homologous gene products. Our results exhibited that general function prediction only accounted for 18.84%, replication, recombination and repair accounted for 10.4%, transcription accounted for 10.28%, and signal transduction mechanisms accounted for 9.84% (Fig. 5C). The eggNOG database serves as an annotation of functional descriptions and classifications for directly homologous groups. Our data revealed that the eggnog function classification under abiotic stresses mainly included posttranslational modification, protein turnover,

chaperones (7.89%), signal transduction mechanisms (6.87%), transcription (5.11%), and carbohydrate transport and metabolism (4.92%), etc (Fig. 5D). The KOG database was used to divide homologous genes from different species into different orthologous clusters. Our data displayed that the KOG function classification under abiotic stresses mainly contained general function prediction only, posttranslational modification, protein turnover, chaperones, signal transduction mechanisms, translation, ribosomal structure and biogenesis, and carbohydrate transport and metabolism, etc (Fig. 5E). The KEGG database is a collection of various pathways, representing molecular interactions and reaction networks. Our data from KEGG analysis disclosed that the enrichment pathways of the transcripts under abiotic stresses mainly included circadian rhythm-plant (ko04712), ubiquitin mediated proteolysis (ko04120), plant-pathogen interaction (ko04626), peroxisome (ko04146), endocytosis (ko04144), eyc. Our data exhibited the ubiquitin mediated proteolysis (Fig. 5F). We also exhibited all KEGG pathways and the related genes (**Table S8**).

Analysis of SNP and the expression of transcripts. The number of SNPs in HomoSNP, HeteSNP, and AllSNP was exhibited in **Table S9**. We also provide the SNP-related data in the **Table S10**. The distribution of SNP density was high in the 0–1 kb (Fig. 6A). In addition, the non-redundant transcripts obtained by third-generation sequencing (single-molecule real-time sequencing) were used as a reference for sequence alignment and subsequent analysis. STAR was used to conduct sequence alignment between Clean Reads and transcripts to obtain location information on transcripts. And the comparison of non-redundant transcripts between second-generation and third-generation sequences was shown in **Table S11**. The total distribution of expression for the transcriptomes was displayed in Fig. 6B. To further examine the degree of dispersion in transcript expression, a boxplot of fragments per kilobase of exon per million (FPKM) in each sample was prepared to intuitively compare the expression levels of the transcripts in different samples (Fig. 6C). The correlation of biological replicates can not only test the reproducibility of biological experiments, but also assess the reliability of differentially expressed genes. Pearson's correlation coefficient analysis was used as the evaluation index of biological repeatability and correlation. We discovered that the three biological replicates in each group were relatively good according to the heatmap of the pairwise comparisons (Fig. 6D).

Identification of differentially expressed transcripts in M. sativa **L. roots under CK, NaCl, and PEG stresses.** Furthermore, based on the full-length transcriptome sequencing, the differentially expressed transcripts were then identified by NGS technology in *M. sativa* L. roots under CK, 400 mM NaCl, and 25% PEG stress. The differentially expressed transcripts were screened by DESeq, and the correlation plots of the *M. sativa* L. transcripts under CK, 400 mM NaCl, and 25% PEG stress were exhibited in Fig. 7A. In addition, the volcano plot and hierarchical clustering analysis revealed the differentially expressed transcripts in *M. sativa* L. transcripts in response to the three stressors. There were 4,080 differentially expressed transcripts (2,609 downregulated and 1,471 upregulated) in the CK stress group compared to the 400 mM NaCl stress group; 5,854 transcripts were differentially expressed (3,241 downregulated and 2,613 upregulated) in the CK stress group compared to the 25% PEG stress group; 8463 transcripts were differentially expressed (3,896 downregulated and 4,567 upregulated) in the 400 mM NaCl stress group compared to the 400 mM NaCl stress group (Fig. 7B and 7C, Table 1).

Table 1
The statistics of the differentially expressed transcripts

DEG_Set	All_DEG	up-regulated	down-regulated
T01_T02_T03 vs. T04_T05_T06 CK stress vs. NaCl stress	4080	1471	2609
T01_T02_T03 vs. T07_T08_T09 CK stress vs. PEG stress	5854	2613	3241
T04_T05_T06 vs. T07_T08_T09 NaCl stress vs. PEG stress	8463	4567	3896

Annotation analysis of differentially expressed transcripts in M. sativa **L. roots under CK, NaCl, and PEG stresses.** According to the NGS results, the differentially expressed transcripts were also documented by GO, COG, eggNOG, KOG, and KEGG pathway analyses. The data from GO analysis showed that the differentially expressed transcripts in *M. sativa L*. roots under CK, NaCl, and PEG stresses were mainly enriched in CC terms (cell, cell part, organelle, membrane, membrane part, organelle part and macromolecular complex, etc), MF terms (catalytic activity, binding, transporter activity, nucleic acid binding transcription factor activity, structural molecule activity and electron carrier activity, etc) and BP terms (metabolic process, cellular process, single-organism process, response to stimulus, biological regulation, and localization, etc) (Fig. 8A). The COG function classification in M. sativa L. roots under CK, NaCl, and PEG stresses mainly included general function prediction only, signal transduction mechanisms, carbohydrate transport and metabolism, and transcription, etc (Fig. 8B). The eggNOG function classification in *M. sativa L*. roots under CK, NaCl, and PEG stresses mainly included carbohydrate transport and metabolism, osttranslational modification, protein turnover, chaperones, energy production and conversion, and amino acid transport and metabolism, etc (Fig. 8C). The KOG function classification in *M. sativa L.* roots under CK, NaCl, and PEG stresses mainly included General function prediction only, signal transduction mechanisms, posttranslational modification, protein turnover, chaperones, and translation, ribosomal structure and biogenesis, etc (Fig. 8D). In addition, the data from KEGG analysis presented that the enrichment pathways of differentially expressed transcripts between CK and NaCl stresses mainly included carbon metabolism, biosynthesis of amino acids, glycolysis/gluconeogenesis, starch and sucrose metabolism, ribosome and protein processing in endoplasmic reticulum, etc; the enrichment pathways of differentially expressed transcripts in PEG stress mainly included carbon metabolism, biosynthesis of amino acids, starch and sucrose metabolism, glycolysis/gluconeogenesis, phenylpropanoid biosynthesis, plant hormone signal transduction and plant-pathogen interaction, etc; the enrichment pathways of differentially expressed transcripts between NaCl and PEG stresses mainly included biosynthesis of amino acids, carbon metabolism, glycolysis/gluconeogenesis, ribosome, starch and sucrose metabolism, protein processing in endoplasmic reticulum and RNA transport, etc (Fig. 8E).Therefore, the differentially expressed transcripts

in *M. sativa L.* roots under CK, NaCl, and PEG stress were mainly enriched in carbon metabolism, biosynthesis of amino acids, glycolysis/gluconeogenesis pathways.

Screening and identification of differentially expressed transcripts. Based on the KEGG analysis, we selected the highly differentially expressed transcripts, which were related to carbon metabolism, biosynthesis of amino acids, glycolysis/gluconeogenesis pathways among the CK, NaCl, and PEG stress groups. As shown in Fig. 9A, hierarchical clustering exhibited the distributions of the Top10 up-regulated and Top10 down-regulated transcripts in Medicago sativa L. roots between NaCl and CK groups (Left), between PEG and CK groups (Middle), and between PEG and NaCl groups (Right). Subsequently, we also selected 5 upregulated and 5 downregulated transcripts between the three groups for verification. The results of the RT-qPCR assay disclosed that ADH1 was upregulated, PEPC and MJG19.6 were downregulated in the NaCl stress group relative to the CK stress group; PCKA and GAPC1 were upregulated, PEPC was downregulated in the PEG stress group compared to the CK stress group; BCAT2, PCKA and GAPC1 were upregulated, ADH1 and PEPC were downregulated in the PEG stress group compared to the NaCl stress group (Fig. 9B and 9C).

Discussion

Agricultural production is “open” large-scale production in a natural environment[28, 29]. Growth and development of crops are often limited due to the changeable natural environment [30]. Poor environmental conditions are often encountered during agricultural production, such as water loss or waterlogging damage; too low or too high temperature; and cold or heat damage[14, 31, 32]. These disasters caused by a poor environment are a tremendous threat to world crop production and a major problem that needs to be solved [33]. Abiotic stress is a commonly encountered environmental factor in nature that seriously affects plant growth and development[34]. Abiotic stress mainly includes low temperature, high salt, and drought conditions [31]. Low temperature stress decreases crop yield and quality[35]. Under low temperature stress, the photosynthetic capacity of plants decreases and excess light energy increases, leading to massive accumulation of reactive oxygen species (ROS)[36]. Excessive ROS attack proteins, nucleic acids, lipids, and other biomolecules, leading to cell death and tissue destruction[37]. Calcium ions are an important signaling molecule during plant stress [38]. Studies have shown that salt stress rapidly increases Ca^{2+} in plant cells[39–41]. PEG is an inert non-ionic long chain polymer that is soluble in water. PEG is the best material for studying drought stress in plants[42]. ABA is a small lipophilic plant hormone that plays an important role in the regulation of the plant stress response[43]. Some studies have suggested that ABA levels change significantly in plants under drought, salinity, and other stressors[44]. Light is an important environmental factor for plant growth[45], and darkness is often used to study light stress in plants[46, 47]. *M. sativa L.* has become a promising crop for use as a bioenergy feedstock; therefore, it is of significance to explore the mechanisms of abiotic stress and improve the stress resistance of *M. sativa* L. during agricultural production.

At present, several studies have reported the underlying relevant mechanisms of abiotic stress in *M. sativa L.* roots[48–50]. For instance, osmotic pressure and salt stress change the microtubule system of

interphase cells in *M. sativa L.* roots by changing salt ions and cations[48]; silicon priming significantly improves the tolerance of *M. sativa* L. to high alkaline stress[49]; the co-transformation of bar and CsLEA enhances the tolerance of *M. sativa* L. to drought and salt stress[50]. In addition, high-throughput sequencing platforms have also been applied to investigate the transcriptome changes in *M. sativa L.* under abiotic stress, including studying the potential mechanisms of the *M. sativa L.* response to cold stress[51]; transcriptome analysis of lead stress in *M. sativa L.* roots[52]; and metabolomics analysis of root-symbiotic rhizobia of *M. sativa L.* under alkali stress[53]. However, information from these studies has been insufficient.

Transcriptome research is a necessary tool to understand life processes. However, RNA-Seq2.0 technology based on the NGS high-throughput platform cannot accurately obtain or assemble complete transcript information, and cannot recognize the transcripts of isoforms, homologous genes, supergene families, or alleles, making it difficult to understand the meaning of life processes at a deeper level. Full-length transcriptome sequencing based on PacBio SMRT single-molecule real-time sequencing technology does not interrupt RNA fragments, and full-length cDNA obtained by reverse transcription using rapid amplification of cDNA ends technology can be used. The ultra-long reads (median 10 kb) of the platform contain a single complete transcript sequence. Furthermore, no assembly is required in the post-analysis, as what is measured is the result[54, 55]. The advantages of SMRT sequencing have been comprehensively studied [56]. SMRT sequencing produces full-length transcripts compared with short-read sequencing[57]. In addition, SMRT sequencing can be used to analyze alternative splicing, primary-precursor-mature RNA structures, and RNA processing[58]. In this study, *M. sativa* L. roots were treated with different stress conditions including dark, low temperature, 400 mM NaCl, 25% PEG, or 100 μM ABA for 0, 3, 6, 12, 24, 48, and 72 h. The full-length transcriptome was fully tested and analyzed by full-length transcriptome sequencing. We obtained a total of 21.53 Gb of clean data, including 566,076 ROI and 409,291 FLNC reads. A total of 194,286 consensus isoforms were identified by transcript clustering analysis of the FLNC reads. After removing the redundant sequences, we obtained 81,017 transcripts. We also predicted 4,971 alternative splicing events. In addition, we identified 33,058 SSR and 42,725 CDS regions. Furthermore, 77,221 transcripts were annotated and 3,043 lncRNAs were predicted. And all of these raw sequencing data has been successfully uploaded to the BioProject (Accession code: PRJNA531296; link: https://www.ncbi.nlm.nih.gov/sra/PRJNA531296)).

Based on the data and a combination of NGS and SMRT sequencing, we also screened the differentially expressed transcripts in *M. sativa* L. roots under CK (control), high salt (400 mM NaCl) and drought (25% PEG) conditions by NGS. In addition, the differentially expressed transcripts were documented by GO, COG, eggNOG, KOG, and KEGG pathway analyses. GO analysis, as a gene functional classification system, can comprehensively describe the genetic characteristics and gene products of organisms[59]. In our study, GO analysis uncovered that differentially expressed transcripts in *M. sativa* L. roots under NaCl and PEG stresses may be involved in the metabolic process, cellular process, single-organism process, response to stimulus, biological regulation. Meanwhile, we discovered that catalytic activity, binding, transporter activity, nucleic acid binding transcription factor activity might be associated with the NaCl and PEG stresses in *M. sativa* L. roots. Besides, our data exhibited that the differentially expressed

transcripts were also mainly enriched in cell, organelle, and membrane. Therefore, we suggested that these differentially expressed transcripts in cells have made significant contributions in the cellular metabolic process of *M. sativa* L. after NaCl and PEG stress through regulating the catalysis, binding, transport, or transcriptional activity.

Moreover, based on the KEGG analysis, we discovered that these differentially expressed transcripts in *M. sativa* L. roots under NaCl and PEG stresses were mainly enriched in carbon metabolism, biosynthesis of amino acids, glycolysis/gluconeogenesis pathways. Carbon metabolism is the uptake, transport, storage and utilization of carbon by plants[60]. Carbon metabolism can affect the root growth and water absorption of plants by changing physiological and hydraulic functions, thus affecting the stress resistance of plants[61]. The biosynthesis of amino acids has also been demonstrated to enhance plant adaptability to abiotic stress by participating in the alteration of certain physiological metabolism, the regulation of related gene expression and the activity of key enzymes in plants[62, 63]. Besides, it has been confirmed that the glycolysis/gluconeogenesis pathway mainly responds to the abiotic stress by regulating the ATP supply of plants[64]. Therefore, these three pathways might be in connection with the NaCl and PEG stresses of *M. sativa* L.

In our study, we further screened the Top10 upregulated and Top10 downregulated transcripts in M. sativa L. roots under NaCl stress vs CK, PEG stress vs CK, and NaCl stress vs. PEG stress based on the carbon metabolism, biosynthesis of amino acids, glycolysis/gluconeogenesis pathways. We demonstrated that ADH1 was upregulated, PEPC and MJG19.6 were downregulated in *M. sativa* L. roots under NaCl stress compared to that under CK; PCKA and GAPC1 were upregulated, PEPC was downregulated in *M. sativa* L. roots under PEG stress compared to that under CK; meanwhile, BCAT2, PCKA and GAPC1 were upregulated, ADH1 and PEPC were downregulated in *M. sativa* L. roots under PEG stress compared to that under NaCl stress. Therefore, we proved that ADH1, PEPC and MJG19.6 were associated with NaCl stress in *M. sativa* L. roots; PCKA, GAPC1 and PEPC were associated with PEG stress in *M. sativa* L. roots; PCKA, GAPC1, ADH1 and PEPC were also differently expressed between NaCl stress and PEG stress. Among them, ADH1 (MF: alcohol dehydrogenase activity, zinc ion binding; CC: cytoplasm; BP: oxidation-reduction process) was connected with glycolysis/gluconeogenesis, fatty acid degradation, tyrosine metabolism, alpha-Linolenic acid metabolism pathways. PEPC (MF: phosphoenolpyruvate carboxylase activity; CC: cytosol; BP: carbon fixation, tricarboxylic acid cycle and photosynthesis) was connected with pyruvate metabolism, carbon fixation in photosynthetic organisms and carbon metabolism pathways. MJG19.6 (MF: histidine dehydrogenase activity, zinc ion binding, NAD binding; BP: histidine biosynthetic process, spermidine biosynthetic process, response to UV, pollen development, oxidation-reduction process; CC: chloroplast stroma) was related to histidine metabolism and biosynthesis of amino acids pathways. PCKA (MF: phosphoenolpyruvate carboxykinase (ATP) activity, ATP binding, kinase activity; BP: gluconeogenesis, phosphorylation) was related to glycolysis/gluconeogenesis, citrate cycle (TCA cycle), pyruvate metabolism, carbon fixation in photosynthetic organisms, carbon metabolism. GAPC1 (MF: NAD + activity, NADP binding, NAD binding; CC: cytoplasm; BP: glycolytic process, oxidation-reduction process) was in connection with glycolysis/gluconeogenesis, carbon fixation in photosynthetic organisms, carbon metabolism and

biosynthesis of amino acids. Therefore, we suggested that these transcripts might provide novel insight into stress-response mechanisms in *M. sativa* L. roots. In future research, we will also dig into a large amount of experimental data, and focus on the verification of relevant transcripts.

Conclusions

In this study, we comprehensively identified the full-length transcriptome sequences in *M. sativa* L. roots under abiotic stress (low temperature, darkness, high salt, drought, and ABA) for the first time. We also identified the differentially expressed transcripts and analyzed the annotations of the differentially expressed transcripts in *M. sativa* L. roots under high salt and drought conditions by combining PacBio sequencing and NGS technologies. Besides, we also screened and identified the aberrantly expressed transcripts (ADH1, PEPC, MJG19.6, PCKA and GAPC1), which might be potentially involved in the influence process of *M. sativa* L. roots under NaCl and PEPC stresses. Therefore, ADH1, PEPC, MJG19.6, PCKA and GAPC1 might serve as potential biomarkers for the vigorous growth of *M. sativa* L. On the whole, our study systematically investigated the full-length transcriptome sequences in *M. sativa* L. roots under abiotic stress. This study provides useful information on the abiotic stress-mediated response in *M. sativa* L. as well as other plants. However, studies are needed to further validate the enormous amount of biological information in our study. Furthermore, it will be necessary to further explore functions and mechanisms of transcripts which we have identified.

Materials And Methods

Plant cultivation and stress treatment. *M. sativa* L. was provided by Institute of Animal Husbandry and Veterinary Science, Beijing Academy of Agricultural Sciences, Beijing, China. The plants were grown in soil in a greenhouse under a light/ (16 h)-dark (8 h) regime, temperature of 25 ± 1 °C, and relative humidity of 80 ± 5% and were watered twice weekly. For NaCl treatment, *M. sativa* L. seedlings were cultured by increasing the concentration of NaCl (1/4 (100 mM) NaCl, 1/2 (200 mM) NaCl and 400 mM NaCl) in medium to avoid shock effect; For PEG treatment, *M. sativa* L. seedlings were also cultured by increasing the concentration of PEG (Final concentration was 25% PEG) in medium; For darkness treatment, *M. sativa* L. seedlings were cultured in a completely dark incubator; For low temperature (4 °C) treatment, we adopted light incubator and temperature-controlled refrigerator to simulate the low-temperature environment, M. sativa L. seedlings were moved into the controlled temperature refrigerator (the cooling speed was 2 °C/h, and the final constant temperature was 4 °C). For ABA treatment, M. sativa L. seedlings were moved into the medium including100 μM ABA. After collection, the root samples were carefully washed with the corresponding temperature of sterile water and dried, and stored in liquid nitrogen to reduce damage.

RNA sample preparation, Library preparation and SMRT sequencing. The *M. sativa* L. roots were exposed to different stress conditions, including darkness, low temperature (4 °C), 400 mM NaCl, 25% PEG and 100 μM ABA for 0, 3, 6, 12, 24, 48, and 72 h. Total RNA was isolated using TRIzol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Electrophoresis on a 1% agarose gel was used to

detect RNA degradation and contamination. The NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Rockland, DE, USA) was used to analyze the purity and concentration of RNA (OD260/280). RNA integrity was evaluated by the RNA Nano 6000 Assay Kit on the Agilent Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA). All total RNA samples were equally mixed together for the following experiments. Then the poly (T) oligo-attached magnetic beads were utilized to purify the mRNA from the mixed RNAs (5 μg RNAs with concentration > 50 ng/μL). Divalent cations were applied to manage fragmentation under the high temperatures. The full-length cDNA of mRNA was synthesized using the SMARTer™ PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). The remaining overhangs were used to generate blunt ends according to exonuclease/polymerase activities. BluePippin (Sage Science, Beverly, MA, USA) was used to screen full-length cDNA fragments and construct the cDNA libraries. The filtered cDNAs were re-amplified by polymerase chain reaction (PCR) assay, and the fragment size distribution was assessed with the Qubit fluorometer (Life Technologies, Carlsbad, CA, USA). Full-length cDNA ends were repaired, and SMRT joints were connected. The cDNAs were re-screened to obtain the sequencing library using BluePippin. The libraries were evaluated quantitatively with the Qubit2.0 DNA kit (Life Technologies, Dalian, China). The size of the libraries was detected by the Agilent 2100 instrument. Finally, cull-length transcriptome sequencing was performed using PacBio Sequel.

Next-generation sequencing (NGS). *M. sativa* L. roots were exposed to CK, NaCl, and PEG stress for 72 h, respectively. The samples in each group contained three biological replicates. Total RNA was isolated from the samples in each group using TRIzol (Invitrogen, Carlsbad, CA, USA). Total RNA (5 μg RNAs with concentration > 50 ng/μL) was digested with DNase I (New England Biomedical, Frankfurt, Germany). The sample was purified using Agencourt RNAClean XP Beads and fragmented. The First Strand Master Mix and Super Script II reverse transcription (Invitrogen) were used to synthesize the first-strand cDNA, and the Second Strand Master Mix was used to generate the second-strand cDNA. After end repairing, the cDNA fragments were amplified using the PCR Master Mix. Finally, the library was quantitatively determined with the Agilent 2100 bioanalyzer instrument and RT-qPCR. The Illumina HiSeq xten platform was used to qualify the libraries.

Isoform sequence clustering. Iterative sequence clustering was analyzed by SMRT analysis software with the iterative clustering for error correction algorithm. Similar sequences were clustered, each of which yielded a consensus isoform. The consistent sequences in each cluster were corrected by applying quiver. Finally, we obtained high-quality transcripts (HQ, high-quality isoforms) with > 99% accuracy.

Transcriptome integrity assessment. BUSCO was used to evaluate the integrity of the transcriptome after removing the redundancies [65].

Alternative splicing analysis. Pre-mRNA can be spliced in a variety of ways. Different exons are selected to produce different mature mRNAs, which are translated into different proteins and constitute the diversity of biological characters. This kind of post-transcriptional mRNA processing is called alternative

splicing. Alternative splicing can be predicted based on the transcripts of three generations of no-reference transcriptome after de-redundancy.

Simple sequence repeat (SSR) analysis. Transcripts > 500 bp were screened, and the SSRs were analyzed using the MIcroSAtellite identification tool.

Prediction of coding sequences. Coding sequence (CDS) sequences were predicted using TransDecoder software based on the open reading frame (ORF) length, amino acid sequence, Pfam database protein structure domain, and log-likelihood score[66].

SNP analysis. STAR software was adopted to compare the Reads of each sample with the Unigene sequence[67], and SNP sites were identified using GATK based on the SNP Calling process of RNA-sequence[68]. These SNP sites can be applied to analyze whether they can affect gene expression levels or the types of protein products. The identification criteria of SNP include: (1) No more than 3 consecutive single base mismatches occurred in the range of 35 bp; (2) The standardized SNP quality value is greater than 2.0.

LncRNA prediction from PacBio sequences. It is necessary to screen the coding potential of the transcripts to determine whether they have coding potential and to filter out the transcripts with coding potential to obtain the predicted lncRNAs. Based on the previous studies[69, 70], the lncRNAs were then predicted using the coding-non-coding index (CNCI)[27], pfam protein structure domain analysis, coding potential assessment tool (CPAT), and the coding potential calculator (CPC)[71].

Prediction of lncRNA targeted transcripts. Target genes were predicted for the 3,043 predicted lncRNA sequences. LncRNA and mRNA acted on each other due to the complementary pairing of the bases. LncRNA target genes were mainly predicted using the LncTar target gene prediction tool.

Transcription factor analysis. A transcription factor refers to the sequence of nucleotides in upstream gene-specific proteins. These proteins regulate the combination of RNA polymerase and the DNA template, and then regulate gene transcription. ITAK software was used to predict the plant transcription factors[72].**Functional annotation of transcripts.** BLAST software (version 2.2.26)[27] was applied to compare the non-redundant transcript sequences with the 8 databases to obtain the annotation information of the transcripts. And the 8 databases included NR[73], Swissprot[74], gene ontology (GO; http://www.geneontology.org) [75], clusters of orthologous groups of proteins (COG; http://www.ncbi.nlm.nih.gov/COG) [76], euKaryotic ortholog groups (KOG)[77], Pfam (http://pfam.janelia.org/) [78], non-supervised orthologous groups (eggnog; http://eggnog.embl.de), and Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.ad.jp/kegg/)[79] databases.

RNA extraction and quantitative real-time PCR (RT-qPCR) assay. The extracted total RNAs from different groups of *M. sativa* L. roots were utilized to synthesize cDNA using SMARTer™ PCR cDNA Synthesis Kit (Clontech). The expression levels of the genes were determined using Bestar™ qPCR MasterMix (DBI Bioscience, cat. no. #2043) on the ABI 7500 Real-time PCR system (Applied Biosystems, Foster City, CA,

USA) according to the manufacturer's instructions. The relative levels were calculated using the $2^{-\triangle\triangle Ct}$ method[80]. The sequences of the gene primers are displayed in **Table S1**.

Statistical analysis. Experimental data are presented as mean ± standard error, and analyzed with Graphpad Ver. Prism 7 software (GraphPad Prism Software, La Jolla, CA, USA). The results were analyzed using Student's *t*-test and one-way analysis of variance analysis, as appropriate. A *P*-value < 0.05 was considered significant.

Abbreviations

ROI: reads of insert; FLNC: full-length non-chimeric; FL: full-length; nFL: non-full-length; SSR: simple sequence repeats; CDS: complete coding sequence; lncRNAs: long non-coding RNA; SMRT: single-molecule real-time; NGS: Next-generation sequencing; PEG: polyethylene glycol; SNP: single nucleotide polymorphism; DEGs: differentially expressed genes; CK: control; MISA: MIcroSAtellite identification tool; CPC: calculator; CNCIL: coding-non-coding index; CPAT: coding potential assessment tool; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; COG: Orthologous Groups of protein; KOG: euKaryotic Ortholog Groups; SNP: single nucleotide polymorphism; FPKM: fragments per kilobase of transcript per million fragments mapped.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All raw sequencing data for the *Medicago sativa* Raw sequence reads have been deposited at the BioProject under accession code PRJNA531296. And the SRA records are accessible with the link (https://www.ncbi.nlm.nih.gov/sra/PRJNA531296).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Construction of Modern Agricultural Industrial Technology System (CARS-34), the Key Research and Development Program of Shanxi Province (201703D211002-9-2, 201903D211012 and 201703D221012-4), Shanxi Academy of Agricultural Sciences Director Youth

Guidance Special Project (yydzx04), Breeding Project of Shanxi Academy of Agricultural Sciences (17yzgc113) and Doctoral Foundation of Shanxi Academy of Agricultural Sciences (YBSJJ1507). Funding bodies granted the funds based on the research proposal. The bodies had no impact on the experimental design, data analysis or writing of the manuscript.

Authors' contributions

ZF and YS designed the research. ZF and YS performed the research. ZF, XW, BR, YZ, PG analyzed the data. ZF, HC, JL and YW wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Fu C, et al. Alfalfa (Medicago sativa L.). Methods Mol Biol. 2015;1223:213–21.
2. Wang D, et al. Genetic Engineering of Alfalfa (Medicago sativa L.). Protein Pept Lett. 2016;23(5):495–502.
3. Panchenko L, Muratova A, Turkovskaya O. Comparison of the phytoremediation potentials of Medicago falcata L. And Medicago sativa L. in aged oil-sludge-contaminated soil. Environ Sci Pollut Res Int. 2017;24(3):3117–30.
4. Jiang J, et al. Overexpression of Medicago sativa TMT elevates the alpha-tocopherol content in Arabidopsis seeds, alfalfa leaves, and delays dark-induced leaf senescence. Plant Sci. 2016;249:93–104.
5. Zhang Y, Sledge MK, Bouton JH. Genome mapping of white clover (Trifolium repens L.) and comparative analysis within the Trifolieae using cross-species SSR markers. Theor Appl Genet. 2007;114(8):1367–78.
6. Ma QG, et al. Characterization of Chalcones from Medicago sativa L. and Their Hypolipidemic and Antiangiogenic Activities. J Agric Food Chem. 2016;64(43):8138–45.
7. Shu Y, et al., *Genome-Wide Investigation of MicroRNAs and Their Targets in Response to Freezing Stress in Medicago sativa L., Based on High-Throughput Sequencing*. G3 (Bethesda), 2016. **6**(3): p. 755–65.
8. Sah SK, Reddy KR, Li J. Abscisic Acid and Abiotic Stress Tolerance in Crop Plants. Front Plant Sci. 2016;7:571.
9. Al-Khayri JM, Naik PM, Alwael HA. *In* Vitro Assessment of Abiotic Stress in Date Palm: Salinity and Drought. Methods Mol Biol. 2017;1637:333–46.
10. Zandalinas SI, Mittler R. ROS-induced ROS release in plant and animal cells. Free Radic Biol Med. 2018;122:21–7.

11. Wu H, Tito N, Giraldo JP. Anionic Cerium Oxide Nanoparticles Protect Plant Photosynthesis from Abiotic Stress by Scavenging Reactive Oxygen Species. ACS Nano. 2017;11(11):11283–97.
12. Schwalm CR, et al., *Assimilation exceeds respiration sensitivity to drought: A FLUXNET synthesis*. Global Change Biology. 16(2): p. 657–670.
13. Vega JM, et al., *Effect of abiotic stress on photosynthesis and respiration in Chlamydomonas reinhardtii: Induction of oxidative stress*. Enzyme & Microbial Technology. 40(1): p. 163–167.
14. Liang W, et al. Plant salt-tolerance mechanism: A review. Biochem Biophys Res Commun. 2018;495(1):286–91.
15. Vurukonda SS, et al. Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. Microbiol Res. 2016;184:13–24.
16. Lawas LMF, et al. Molecular mechanisms of combined heat and drought stress resilience in cereals. Curr Opin Plant Biol. 2018;45(Pt B):212–7.
17. Santos SGD, et al. Dark septate endophyte decreases stress on rice plants. Braz J Microbiol. 2017;48(2):333–41.
18. Hua W, et al., *Effects of Low Temperature Stress on SOD Activity and Membrane Deroxidization of Apricot Flowers*. Journal of Fruit Science, 2000: p. 197–201.
19. Pompeu GB, et al. Abscisic acid-deficient sit tomato mutant responses to cadmium-induced stress. Protoplasma. 2017;254(2):771–83.
20. Choudhury FK, et al. Reactive oxygen species, abiotic stress and stress combination. Plant J. 2017;90(5):856–67.
21. Foley SW, et al. Transcriptome-wide measurement of plant RNA secondary structure. Curr Opin Plant Biol. 2015;27:36–43.
22. McKenzie P, et al. A Common Variant in MTHFD1L is Associated with Increased Risk for Spina Bifida. Journal of Molecular Clinical Medicine. 2018;1(1):19–22.
23. Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.
24. Nakano K, et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Hum Cell. 2017;30(3):149–61.
25. Yanhu L, Lu W, Li Y. [The principle and application of the single-molecule real-time sequencing technology]. Yi Chuan. 2015;37(3):259–68.
26. Zhang SH, et al. [Review of Second Generation Sequencing and Its Application in Forensic Genetics]. Fa Yi Xue Za Zhi. 2016;32(4):282–9.
27. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402.
28. Batary P, et al. The role of agri-environment schemes in conservation and environmental management. Conserv Biol. 2015;29(4):1006–16.

29. Lesk C, Rowhani P, Ramankutty N. Influence of extreme weather disasters on global crop production. Nature. 2016;529(7584):84–7.
30. Nakamichi N. Adaptation to the local environment by modifications of the photoperiod response in crops. Plant Cell Physiol. 2015;56(4):594–604.
31. Zandalinas SI, et al. Plant adaptations to the combination of drought and high temperatures. Physiol Plant. 2018;162(1):2–12.
32. Ghosh D, Xu J. Abiotic stress responses in plant roots: a proteomics perspective. Front Plant Sci. 2014;5:6.
33. Kosova K, et al. Biological Networks Underlying Abiotic Stress Tolerance in Temperate Crops–A Proteomic Perspective. Int J Mol Sci. 2015;16(9):20913–42.
34. Zhu JK. Abiotic Stress Signaling and Responses in Plants. Cell. 2016;167(2):313–24.
35. Jha UC, Bohra A, Jha R. Breeding approaches and genomics technologies to increase crop yield under low-temperature stress. Plant Cell Rep. 2017;36(1):1–35.
36. Sewelam N, Kazan K, Schenk PM. Global Plant Stress Signaling: Reactive Oxygen Species at the Cross-Road. Front Plant Sci. 2016;7:187.
37. Redza-Dutordoir M, Averill-Bates DA. Activation of apoptosis signalling pathways by reactive oxygen species. Biochim Biophys Acta. 2016;1863(12):2977–92.
38. Ranty B, et al. Calcium Sensors as Key Hubs in Plant Responses to Biotic and Abiotic Stresses. Front Plant Sci. 2016;7:327.
39. Choi WG, et al. Salt stress-induced Ca2 + waves are associated with rapid, long-distance root-to-shoot signaling in plants. Proc Natl Acad Sci U S A. 2014;111(17):6497–502.
40. Li L, et al. A phosphoinositide-specific phospholipase C pathway elicits stress-induced Ca(2+) signals and confers salt tolerance to rice. New Phytol. 2017;214(3):1172–87.
41. Zhao R, et al. The Arabidopsis Ca(2)(+)-dependent protein kinase CPK27 is required for plant response to salt-stress. Gene. 2015;563(2):203–14.
42. Ferreira BG, Teixeira CT, Isaias RM. Efficiency of the Polyethylene-Glycol (PEG) Embedding Medium for Plant Histochemistry. J Histochem Cytochem. 2014;62(8):577–83.
43. Shanker AK, et al. Drought stress responses in crops. Funct Integr Genomics. 2014;14(1):11–22.
44. Osakabe Y, et al. ABA control of plant macroelement membrane transport systems in response to water deficit and high salinity. New Phytol. 2014;202(1):35–49.
45. Lima VA, et al. Growth, photosynthetic pigments and production of essential oil of long-pepper under different light conditions. An Acad Bras Cienc. 2017;89(2):1167–74.
46. Silva-Navas J, et al. D-Root: a system for cultivating plants with the roots in darkness or under different light conditions. Plant J. 2015;84(1):244–55.
47. Ye XQ, et al. Submergence Causes Similar Carbohydrate Starvation but Faster Post-Stress Recovery than Darkness in Alternanthera philoxeroides Plants. PLoS One. 2016;11(10):e0165193.

48. Lazareva EM, Baranova EN, Smirnova EA. *REORGANIZATION OF INTERPHASE MICROTUBULES IN ROOT CELLS OF MEDICAGO SATIVA L. DURING ACCLIMATION TO OSMOTIC AND SALT STRESS CONDITION*. Tsitologiia, 2017. 59(1): p. 34–44.
49. Liu D, et al. Silicon Priming Created an Enhanced Tolerance in Alfalfa (Medicago sativa L.) Seedlings in Response to High Alkaline Stress. Front Plant Sci. 2018;9:716.
50. Zhang J, et al. Co-transforming bar and CsLEA enhanced tolerance to drought and salt stress in transgenic alfalfa (Medicago sativa L.). Biochem Biophys Res Commun. 2016;472(1):75–82.
51. Zhou Q, et al., *Multiple Regulatory Networks Are Activated during Cold Stress in Medicago sativa L*. Int J Mol Sci, 2018. 19(10).
52. Xu B, et al. Transcriptomic and physiological analyses of Medicago sativa L. roots in response to lead stress. PLoS One. 2017;12(4):e0175307.
53. Song T, et al. Metabolomic Analysis of Alfalfa (Medicago sativa L.) Root-Symbiotic Rhizobia Responses under Alkali Stress. Front Plant Sci. 2017;8:1208.
54. Gordon SP, et al. Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA. Sequencing. 2015;10(7):e0132628.
55. Thomas S, et al., *Long-Read Sequencing of Chicken Transcripts and Identification of New Transcript Isoforms*. 2014. 9(4): p. e94650.
56. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13(5):278–89.
57. Guo X, et al., *SMRT Sequencing for Parallel Analysis of Multiple Targets and Accurate SNP Phasing*. 2015. 5(12): p. 2801–2808.
58. Li Y, et al., *Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry*. 2017. 90(1): p. 164.
59. Consortium GO. The Gene Ontology (GO) database and informatics resource. Nucleic acids research. 2004;32(suppl_1):D258–61.
60. Dusenge ME, Duarte AG, Way DA. Plant carbon metabolism and climate change: elevated CO(2) and temperature impacts on photosynthesis, photorespiration and respiration. New Phytol. 2019;221(1):32–49.
61. Quick WP, et al., *The effect of water stress on photosynthetic carbon metabolism in four species grown under field conditions*. Plant Cell & Environment. 15(1): p. 25–35.
62. Cao XC, et al. Advances in studies of absorption and utilization of amino acids by plants: A review. Chin J Appl Ecol. 2015;26(3):919–29.
63. Qi Y, et al., *Preparation comprising amino acids and plants and its activity in the alcohol detoxification*. Modutech, 2015.
64. Qaddoori AG. *Antimicrobial evaluation of selected medicinal plants using molecular approach*. 2016.
65. Simao FA, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.

66. Haas BJ, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512.
67. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
68. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
69. Yang L, et al. Genome-wide identification of long non-coding RNA and mRNA profiling using RNA sequencing in subjects with sensitive skin. Oncotarget. 2017;8(70):114894–910.
70. Peng R, et al. Characterization and Analysis of Whole Transcriptome of Giant Panda Spleens: Implying Critical Roles of Long Non-Coding RNAs in Immunity. Cell Physiol Biochem. 2018;46(3):1065–77.
71. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics. 2014;15:311.
72. Zheng Y, et al. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. Mol Plant. 2016;9(12):1667–70.
73. Deng Y, et al., *Integrated nr Database in Protein Annotation System and Its Localization*. 2006. 32(5): p. 71–72.
74. Apweiler R, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004;32(Database issue):D115-9.
75. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
76. Tatusov RL, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33–6.
77. Koonin EV, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 2004;5(2):R7.
78. Finn RD, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222-30.
79. Kanehisa M, et al. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004;32(Database issue):D277-80.
80. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402–8.

Figures

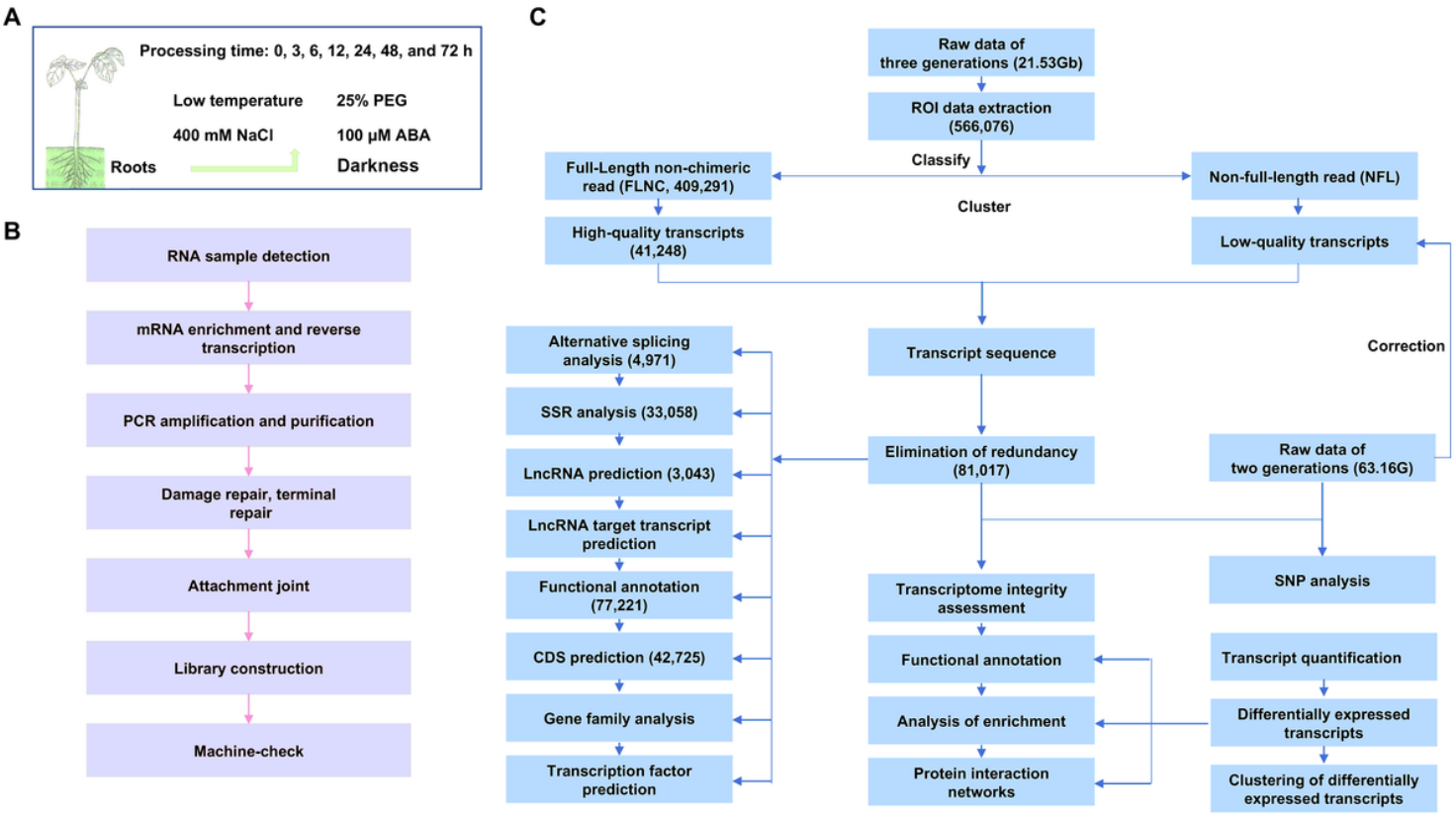

Figure 1

The experimental and analytical processes for full-length transcriptome sequencing. (A) Schematic graph of Medicago sativa L. treatment. (B) Flow diagram of full-length transcriptome sequencing experiment. (C) Flow chart for the bioinformatics analysis of the full-length transcriptome.

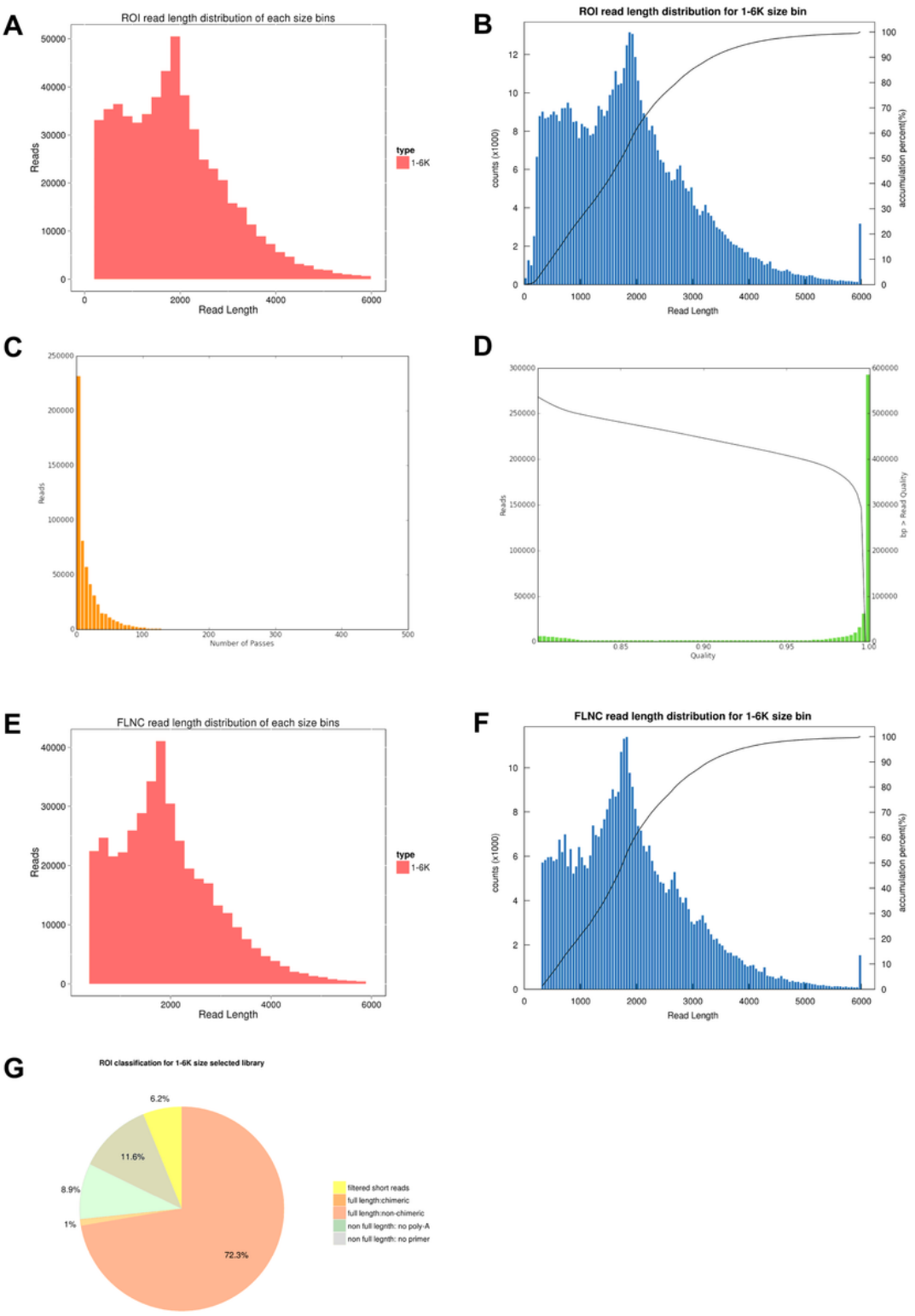

Figure 2

Summary of Pacbio RS sequencing based on sequencing by synthesis (SMRT). (A, B) Reads-of-insert (ROI) length distribution of the cDNA library. The X-axis represents the ROI length distribution, the Y-axis (left) represents the frequency distribution for ROI sequence length, and the Y-axis (right) represents the cumulative frequency curve for ROI sequence length. (D) A full passes distribution of the ROI sequences. The X-axis indicates the number of full passes in the ROI sequences, the Y-axis indicates the number of

ROI sequences of the corresponding full passes. (D) The distribution of ROI quality values. The X-axis represents the distribution of ROI quality values, the Y-axis (left) represents the frequency distribution of ROI sequence quality values, and the Y-axis (right) represents the number of bases greater than the corresponding quality values. (E, F) FLNC length distribution. The X-axis represents the length distribution of the FLNC sequences, the Y-axis (left) represents the frequency distribution of the FLNC sequence lengths, and the Y-axis (right) represents the cumulative frequency curve of the FLNC sequence lengths. (G) ROI classification for 1–6 k size selected library. FLNC, full-length non-chimeric read.

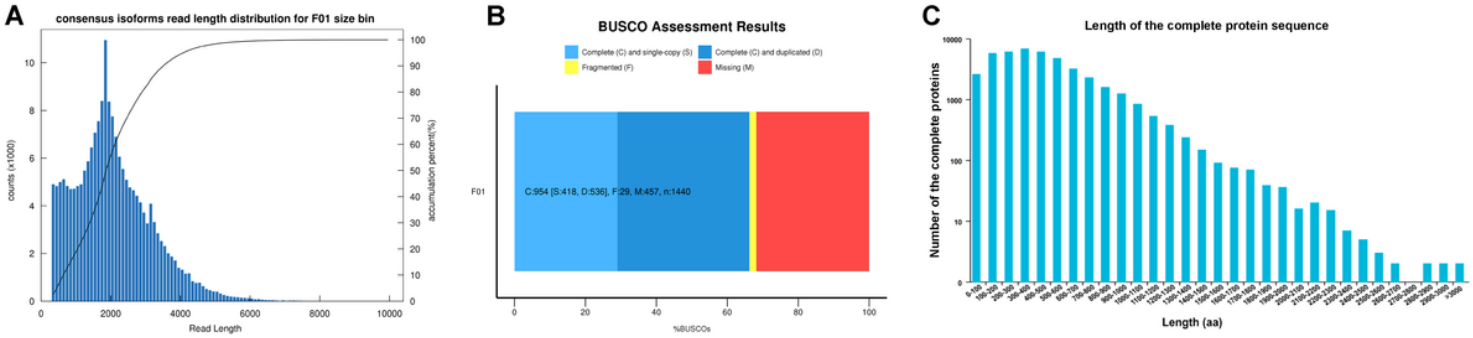

Figure 3

Alternative splicing and SSR analyses and coding sequences prediction. (A) The length distribution of the consensus isoforms was analyzed using the iterative clustering for error correction (ICE) method. The X-axis represents the length distribution of the consensus isoform sequences. The Y-axis (left) represents the frequency distribution of the consensus isoform sequence lengths; the Y-axis (right) represents the cumulative frequency curve of consensus isoform sequence lengths. (B) Transcriptome integrity was assessed using BUSCO. (C) The distribution of SSR type was evaluated with the MIcroSAtellite identification tool (MISA). c, Compound SSR; c*, compound SSR with overlapping bases between two SSRs; p1, Mono-nucleotide; p2, Di nucleotide; p3, Tri nucleotide; p4, Tetra nucleotide; p5, Penta nucleotide; p6, Hexa nucleotide. (D) The complete length distribution of the predicted CDS coding proteins. SSR, simple sequence repeats; CDS, complete coding sequence.

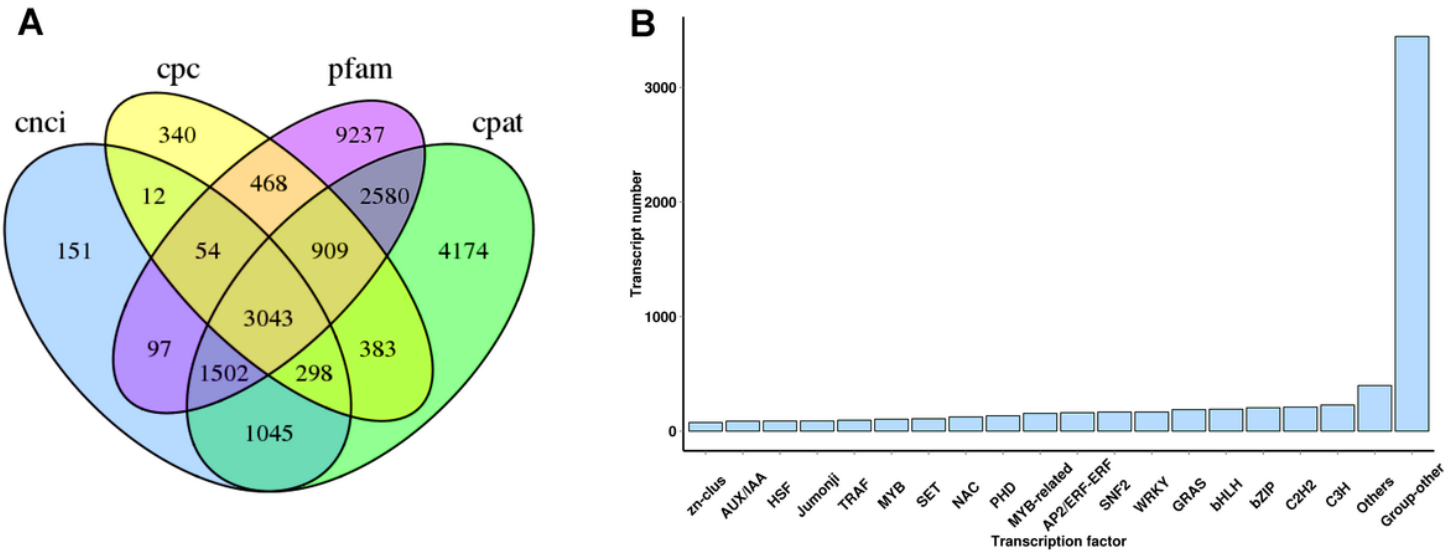

Figure 4

LncRNA prediction and transcription factor analysis. (A) Venn diagram of lncRNAs number predicted by CPC, CPAT, CNCI, and pfam protein structure domain analyses. CPC, calculator; CNCI, coding-non-coding index; CPAT, coding potential assessment tool. (B) Distribution of different transcription factors.

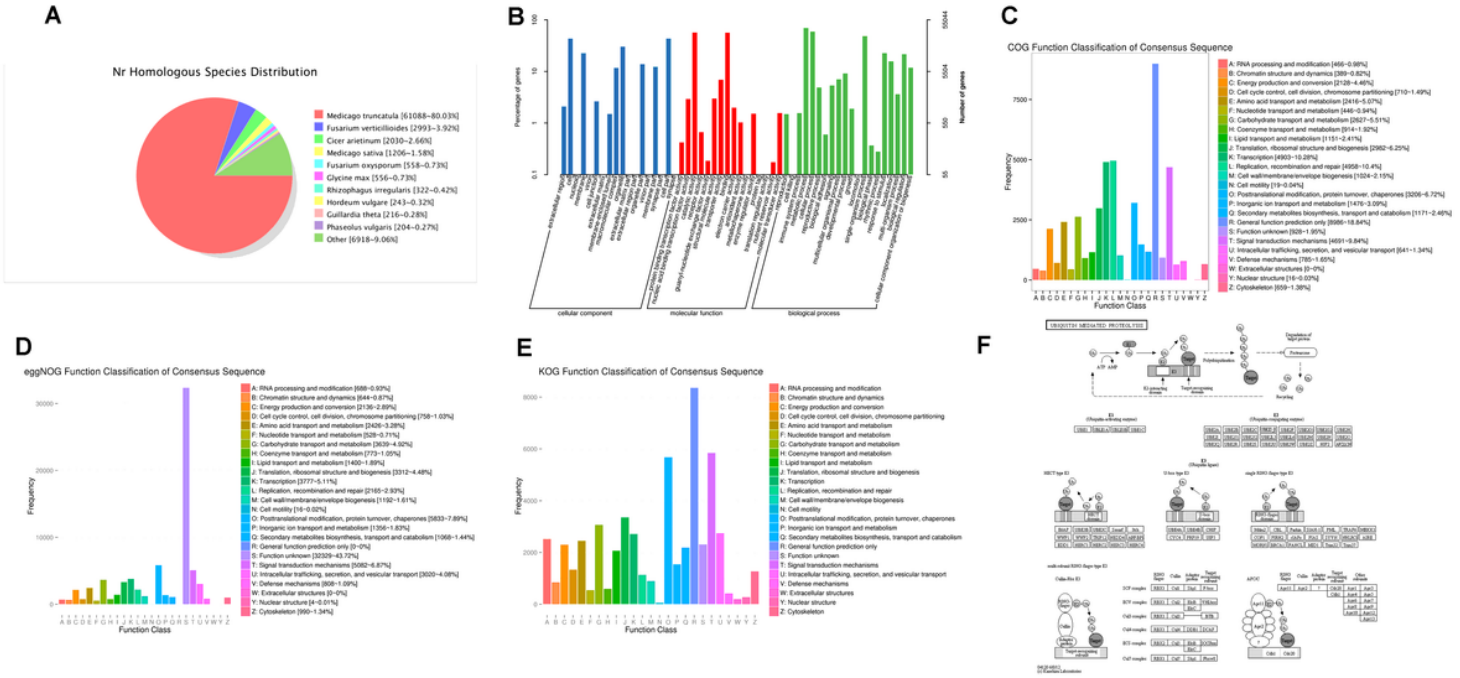

Figure 5

Functional annotation of the transcripts. (A) Homologous species distribution of Medicago sativa L. roots under abiotic stress annotated in the NR database. (B) Gene Ontology Consortium (GO) annotation of the M. sativa L. transcripts under abiotic stress. (C) Clusters of Orthologous Groups of protein (COG) annotation of M. sativa L. transcripts under abiotic stress. (D) eggNOG annotation of M. sativa L. transcripts under abiotic stress. (E) Clusters of euKaryotic Ortholog Groups (KOG) annotation of M. sativa L. transcripts under abiotic stress. (F) The Kyoto Encyclopedia of Genes and Genomes (KEGG) was applied to analyze the related pathway (Ubiquitin mediated proteolysis; ko04120) of M. sativa L. transcripts under abiotic stress.

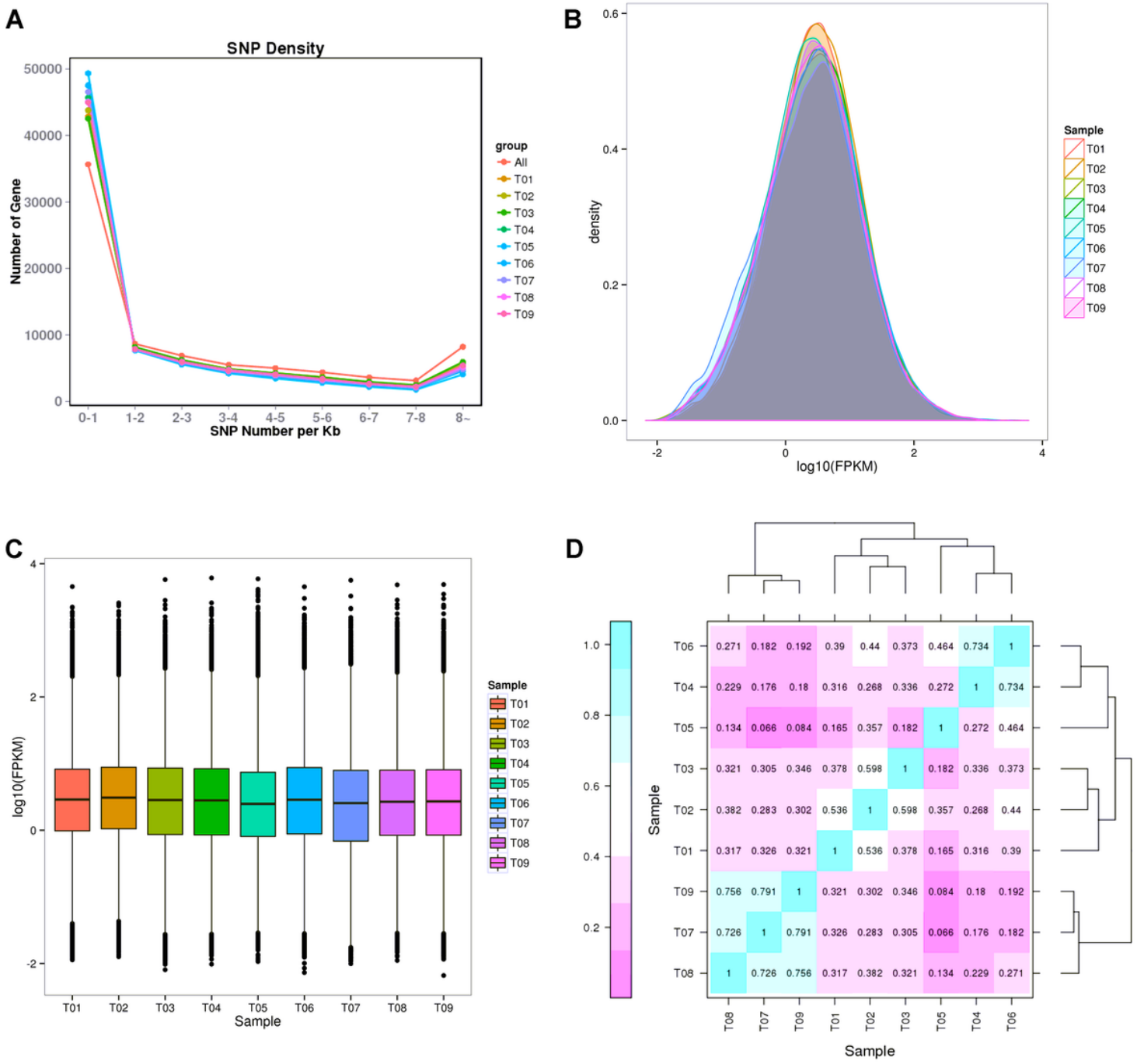

Figure 6

Analysis of SNP and expressions of the transcripts. (A) The distribution of the SNP density was analyzed using STAR software. (B) The density distribution of FPKM in each sample. (C) The boxplot of FPKM in each sample. (D) The heatmap of the pairwise comparison of Medicago sativa L. transcripts under abiotic stress. The colored bar represents the expression levels of transcripts (log2FPKM+1). SNP, single nucleotide polymorphism; FPKM, fragments per kilobase of transcript per million fragments mapped.

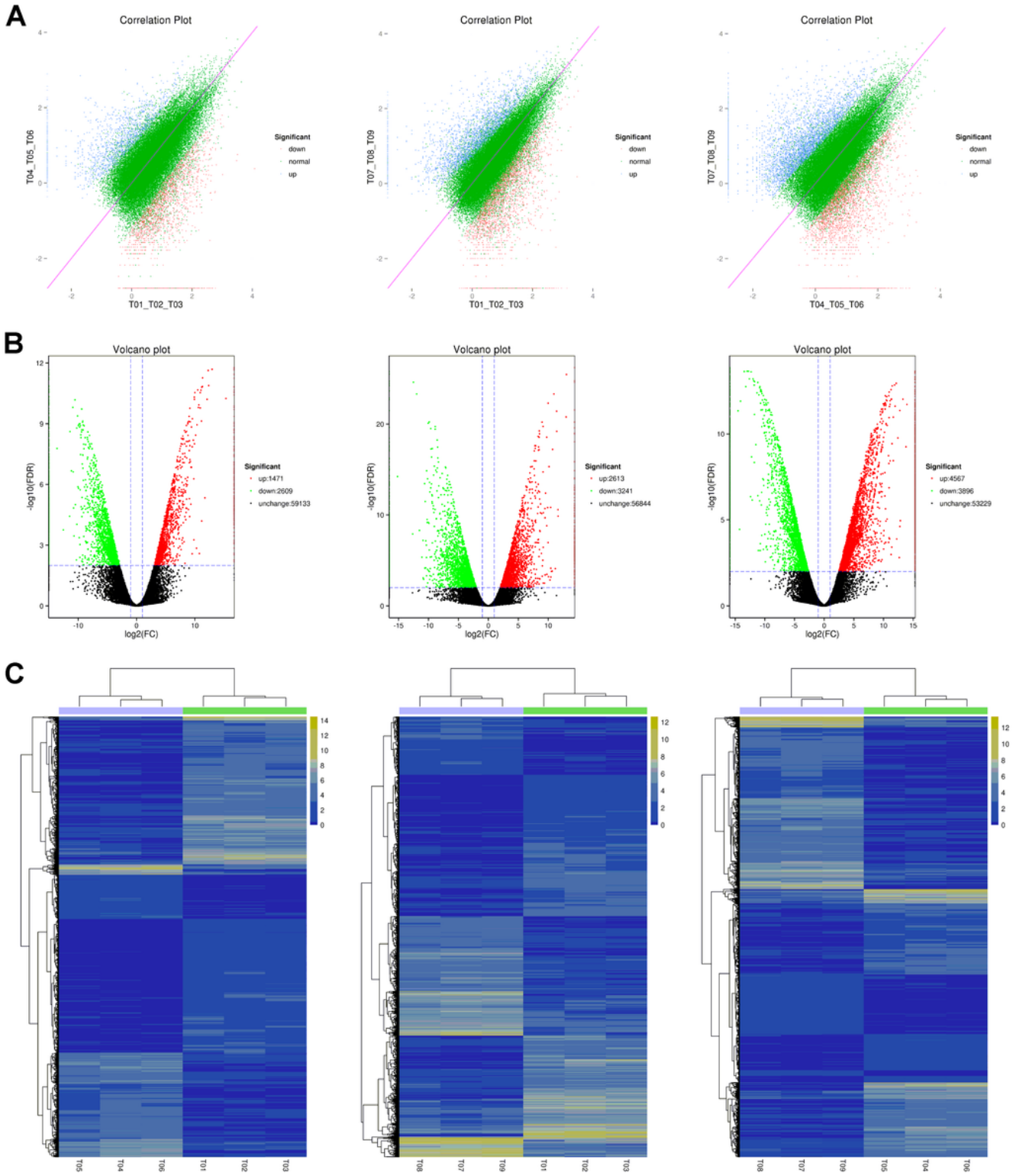

Figure 7

Identification of differentially expressed transcripts in Medicago sativa L. roots under abiotic stress. (A) Correlation plots of M. sativa L. transcripts under CK, 400 mM NaCl, and 25% PEG stress. (B) Volcano plot indicates the differentially expressed M. sativa L. transcripts under control (CK), 400 mM NaCl, and 25% PEG stress. The red points indicate significantly upregulated transcripts; the green points indicate significantly downregulated transcripts, and the black points indicate the transcripts that did not change.

(C) A hierarchical clustering analysis of differentially expressed transcripts in M sativa L. roots under CK, 400 mM NaCl, and 25% PEG stress.

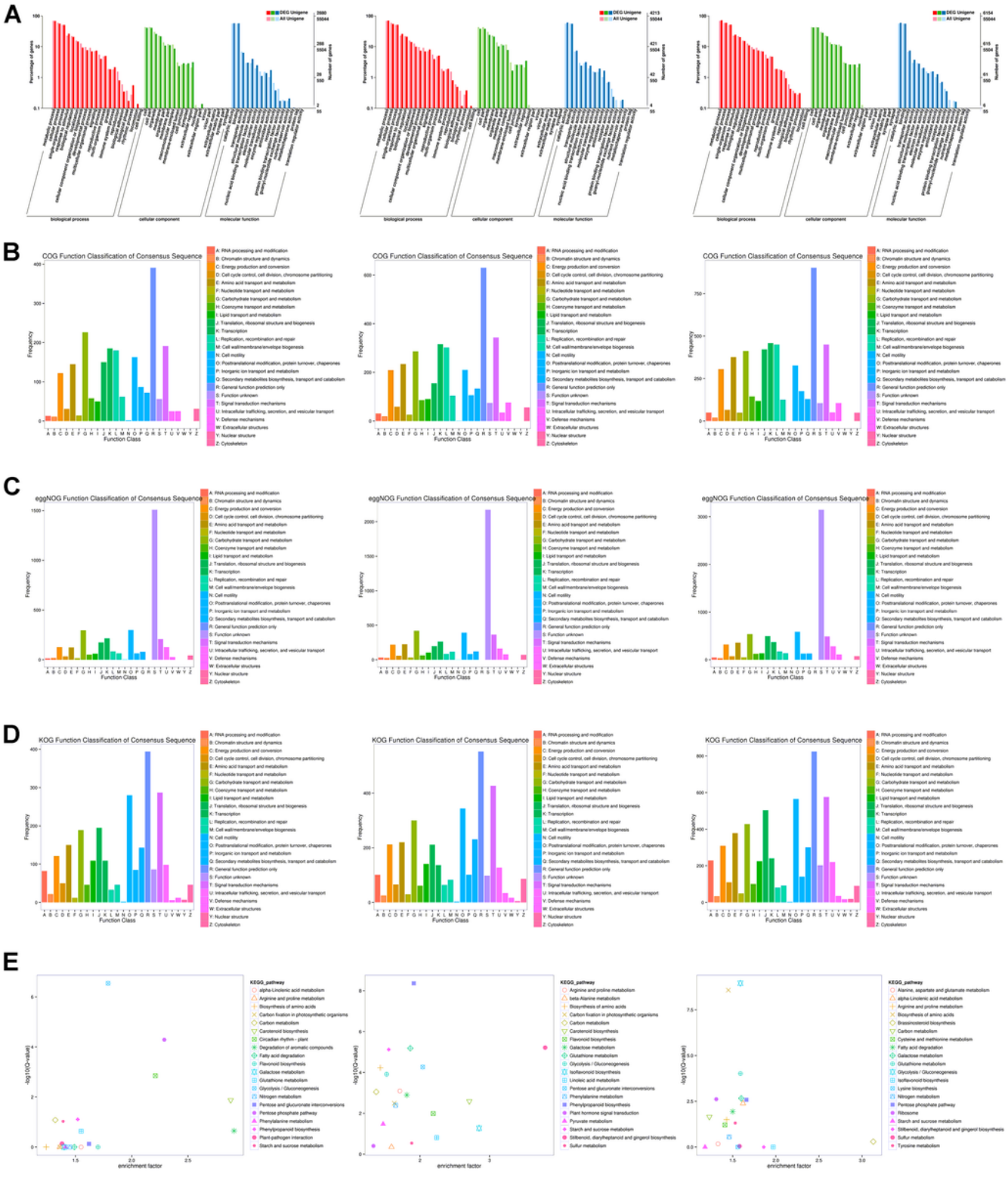

Figure 8

Annotation analysis of differentially expressed transcripts in Medicago sativa L. roots under abiotic stress. (A) GO annotation of M. sativa L. transcripts under CK, 400 mM NaCl, and 25% PEG stress, respectively. (B) COG annotation of M. sativa L. transcripts under CK, 400 mM NaCl, and 25% PEG stress,

respectively. (C) eggNOG annotation of M. sativa L. transcripts under CK, 400 mM NaCl, and 25% PEG stress, respectively. (D) KOG annotation of M. sativa L. transcripts under CK, 400 mM NaCl, and 25% PEG stress, respectively. (E) Scatterplot of enriched KEGG pathway shows the top 20 pathways enriched in M. sativa L. transcripts under CK, 400 mM NaCl, and 25% PEG stress, respectively.

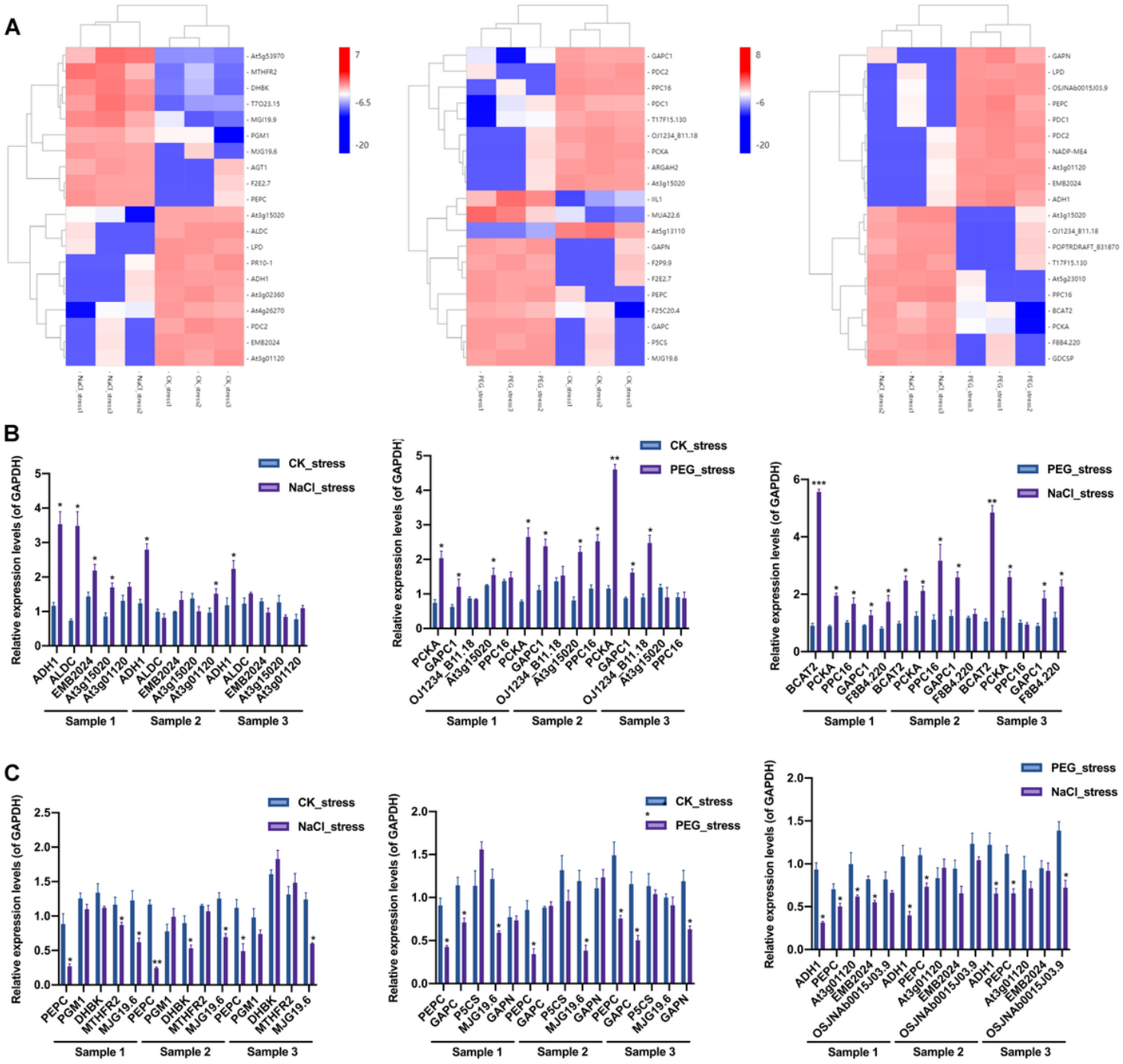

Figure 9

Screening and verification of abnormally expressed transcripts. (A) Hierarchical clustering of the highest differentially expressed transcripts in Medicago sativa L. roots under CK (control), NaCl, and PEG stress. Red represents high gene expression and blue represents low gene expression. (B) The expressions of 5 upregulated transcripts were identified by RT-qPCR assay in Medicago sativa L. roots among CK (control),

NaCl, and PEG stress. (C) RT-qPCR analysis was applied to confirm the expressions of 5 downregulated transcripts in Medicago sativa L. roots among CK (control), NaCl, and PEG stress. *P<0.05, **P<0.01, ***P<0.001.

Supplementary Files

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

- AdditionalTableS1.docx
- AdditionalTableS2.docx
- AdditionalTableS3.xls
- AdditionalTableS4.docx
- AdditionalTableS5.xls
- AdditionalTableS6.docx
- AdditionalTableS7.xls
- AdditionalTableS8.xlsx
- AdditionalTableS9.docx
- AdditionalTableS10.xls
- AdditionalTableS11.docx