Construction of Drought Stress Regulation Networks in Potato Based on SMRT and RNA Sequencing Data

Hongju Jian
Southwest University
Haonan Sun
Southwest University
Rongrong Liu
Southwest University
Wenzhe Zhang
Southwest University
Lina Shang
Southwest University
Jichun Wang
Southwest University
Vadim Khassanov
S.Seifullin Kazakh Agro Technical University
Dianqiu Lyu (smallpotatoes@126.com)
Southwest University

Research Article

Keywords: Solanum tuberosum, Drought stress, SMRT, Alternative splicing, LncRNAs, Alternative polyadenylation

Posted Date: March 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1456188/v1

Abstract

Background: Potato (*Solanum tuberosum*) is the fourth most important food crop in the world and plays an important role in food security. However, drought stress has a significantly negative impact on potato growth and production. Little information about the coordinated response in potato to drought stress has been obtained. The purpose of our research is to fully analyze the potato's response to drought stress.

Results: In this study, next-generation sequencing (NGS) and single-molecule real-time (SMRT) sequencing technology were used to study the transcription profiles in potato in response to drought stress. The leaves of the variety “Desiree” at the seedling stage after drought stress at six time points from 0 to 48 hours were used to perform NGS and SMRT sequencing. According to the sequencing data, a total of 12,798 differentially expressed genes (DEGs) were identified in six time points. The real-time (RT)-PCR results are significantly correlated with the sequencing data, confirming the accuracy of the sequencing data. Gene ontology and KEGG analysis show that these DEGs participate in response to drought stress through galactose metabolism, fatty acid metabolism, plant-pathogen interaction, glutathione metabolism and other pathways. Through the analysis of alternative splicing of 66,888 transcripts, the functional pathways of these transcripts were enriched, and 51,098 transcripts were newly discovered from alternative splicing events and 47,994 transcripts were functionally annotated. Moreover, 3,445 lncRNAs were predicted and enrichment analysis of corresponding target genes was also performed. Additionally, Alternative polyadenylation was analyzed by TADIS, and 26,153 poly (A) sites from 13,010 genes were detected in the Iso-Seq data.

Conclusion: Our research greatly enhanced potato drought-induced gene annotations and provides transcriptome-wide insights into the molecular basis of potato drought resistance.

Introduction

Drought is the most important environmental factor affecting crop yield. Studies have shown that drought can cause up to 50% of crop yield loss. As global warming intensifies the intensity, frequency and range of drought, resulting in more serious yield loss (Gong, Xiong et al. 2020). Therefore, breeding drought tolerant materials is an important goal of current breeding. Although lots of work on drought stress have been conducted in model plants, the molecular mechanism of drought response in potato, the fourth largest food crop, still needs to be further explored. The yield of potato is far from its physiological potential. Due to the drought-sensitivity of potato varieties, the yield is greatly reduced due to the lack of irrigation facilities or arid areas in the world. Drought stress leads to biochemical, molecular, physiological and morphological changes and significantly affects plant growth and yield (Salehi-Lisar and Bakhshayeshan-Agdam 2016). Materials with strong drought tolerance can reduce the impact of water shortage on plants. Therefore, screening and utilizing candidate drought stress response genes can stabilize potato yield in water-shortage area.

Plants have evolved very complex regulatory mechanisms to cope with the adverse environments (Blum 1996, Zhu 2016, Nejat and Mantri 2017). Many studies have been conducted to explore the impact of drought stress on physiological, molecular, and biochemical levels in various crops (Lamaoui, Jemo et al. 2018). Plants resist drought stress by reducing water loss (such as decreased leaf area, reduced stomata number and its conductance) and increasing water uptake (enhanced root formation, increased leaf thickness, and leaf rolling) in morphological aspect (Sicher, Timlin et al. 2012). There have been many reports on the rapid identification of the molecular mechanisms of drought response at the whole-genome level using next-generation sequencing (NGS) platforms or gene microarrays in different species. Based on these data, key genes, proteins or pathways were identified. Genes encoding proteins which are involved in water transport (aquaporin) (Alexandersson, Fraysse et al. 2005, Alexandersson, Danielson et al. 2010), scavenging of free oxygen radicals (superoxide dismutase, catalase, and peroxidase), maintaining cellular membrane integrity (proline, mannitol, glycine, and betaine), and protecting macromolecules (chaperones and late embryogenesis abundant proteins) (Shinozaki and Yamaguchi-Shinozaki 2007, Golldack, Li et al. 2014) directly protects plants against drought stress. The other genes involved in signal perception or transduction may also play critical roles in response to drought stress. Key transcription factors (TFs) such as MYB, WRKY, bZIPs, kinases such as SnRK2 and small molecules such as abscisic acid (ABA) have been functionally proved to play key roles in plant drought tolerance (Shinozaki, Yamaguchi-Shinozaki et al. 2003, Golldack, Li et al. 2014). Some key TFs such as ABA-dependent MYC/MYB, AREB/ABF (ABA-responsive element binding/ABA-binding factor), DREB (ABA-independent dehydration-responsive element-binding proteins), CRT/DRE (C-repeat/drought-responsive element) have been proved in many crops in response to drought stress (Yamaguchi-Shinozaki and Shinozaki 2006, Shin, Moon et al. 2011, Hu and Xiong 2014).

The *S. tuberosum* group Phureja clone DM1-3 516 R 44 (doubled monoploid) and the *S. tuberosum* group Tuberosum RH89-039-16 (heterozygous diploid) have been sequenced and released via the Potato Genome Sequence Consortium (PGSC) in 2011 (Xu, Pan et al. 2011). This provides a very good genomic reference information for transcriptome sequencing (Massa, Childs et al. 2011, Massa, Childs et al. 2013, Gong, Zhang et al. 2015, Galvez, Tai et al. 2016). For drought stress in potato, several studies have been conducted using RNA-Seq (Zhang, Liu et al. 2014, Moon, Ahn et al. 2018, Sprenger, Erban et al. 2018, Chen, Li et al. 2020). However, the completed genome information is still lack because most potato cultivars are the autotetraploid.

Although the NGS technology has many advantages, such as high throughput, high sensitivity, high accuracy and low cost, it has a major disadvantage, that is, the reading length is short, and it is difficult to ensure the accuracy of reconstructed transcripts in the assembly process, which brings severe challenges to eukaryotic transcriptome sequencing with many homologous genes (Steijger, Abril et al. 2013). In order to break the barrier, the third generation sequencing technology came into being. The third generation sequencing technology is a single molecule sequencing technology. In the process of sequencing, each RNA molecule is directly sequenced without PCR amplification. It has the characteristics of super long reading length and can cover the complete transcripts. The single molecular real time (SMRT) sequencing technology introduced by Pacific Bioscience (pacbio) was most widely used (Gupta and Verma 2019).

Through this sequencing technology, information such as variable splicing, fusion gene, gene family and lncRNA can be accurately identified, which is widely used in genome research, transcriptome sequencing and epigenetic analysis (Rhoads and Au 2015). However, the main problem in SMRT is the high error rate (mainly reflected in insertion or deletion errors) (Rhoads and Au 2015). In order to solve this problem, the NGS with high quality and high coverage sequencing data was used to correct SMRT data, so "SMRT + NGS" sequencing joint analysis is more and more widely used (Chao, Gao et al. 2019, Yan, Gao et al. 2019, Yao, Liang et al. 2020).

In previous studies on potato drought stress transcriptome sequencing, although many differentially expressed genes were identified and screened, clustered and functionally annotated, their analysis was not comprehensive and detailed (Evers, Lefevre et al. 2010, Zhang, Liu et al. 2014, Gong, Zhang et al. 2015). In our study, the leaves of the variety “Desiree” at the seedling stage after drought stress at six time points from 0 to 48 hours were used to perform NGS and SMRT sequencing. Our research will provide a valuable supplement to the potato genome and a broader regulatory mechanism in response to drought stress in potato.

Materials And Methods

Plant materials and drought stress treatments

A drought tolerant tetraploid potato “Desiree”, provided by Southwest University, was used in this study. After transplanting about 10 cm tube seedlings into hydroponic solution (1/4 Hoagland nutrient solution) for one week at 22/18℃ (16 h light / 8 h dark), the materials were divided into two parts, one for control and the other for drought stress treatment, which were treated with 20% polyethylene glycol (PEG)-6000. The aboveground tissues were taken at 0, 1, 3, 6, 12, 24 and 48 hours after drought stress treatment and immediately frozen at least 30 min in liquid nitrogen and kept at −80°C until use. The samples were collected for three biological replicates.

Library construction for Illumina and PacBio sequencing

Total RNA of all samples was extracted using EZ-10 DNA away RNA Mini-Preps Kit (Sangon Biotech, Shanghai, China) and the quality of total RNAs was evaluated using an Agilent 2100 Bio-analyzer (SA Pathology, Adelaide, SA, Australia). High-quality RNAs (RIN≥ 8.0) were used for the following cDNA synthesis and library construction. For Illumina RNA sequencing, 1.0 μg RNA per sample was used for cDNA synthesis with the NEBNext ® Ultra ™ II RNA Library Prep Kit for Illumina (NEB, USA) and sent to Biomarker Technologies Corporation (Beijing, China) for Illumina sequencing. For SMRT sequencing, equal amount of the total RNA of all samples were pooled together as the template for cDNA synthesis with the SMARTer PCR cDNA Synthesis Kit (Clontech, USA) in Biomarker Technologies Corporation (Beijing, China).

Analysis of the Illumina and SMRT sequencing data

The raw Illumina data was filtered to generate clean reads using FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects /fastqc/) (v0.11.5) and remove the adaptor contamination and low-quality reads using Trimmomatic (v0.35) (Bolger, Lohse et al. 2014). The high-quality clean reads were then mapped to the reference genome database (http://solanaceae.plantbiology.msu.edu/rh_potato_download.shtml) using HISAT2 version 2.1.0 using the default parameters (Kim, Landmead et al. 2015). For SMRT sequencing data, the circular consensus sequences (CCS) were generated from the PacBio Iso-Seq raw reads using smrtanalysis_2.3.0 (http://www.pacb.com/) and then processed into error corrected reads of insert (ROIs) with a minimum full pass of $\geq$ 3 and minimum accuracy of $\geq$ 0.9. These ROIs were classified as full-length, non-chemiric (FLNC) transcripts using smrtanalysis_2.3.0 based on the presence and location of the 5' and 3' cDNA primers and poly (A) tail. All full-length reads were aligned to the *S. tuberosum* group Tuberosum RH89-039-16 reference genome (http://solanaceae.plantbiology.msu.edu/rh_potato_download.shtml) using GMAP software (Wu and Watanabe 2005). Finally, all mapped reads were collapsed using a Python script in the ToFU package (https://github.com/PacificBiosciences/cDNA_primer/) with min-coverage=85% and min-identity=90%. All raw data was deposited in the National Center for Biotechnology Information (NCBI) sequence read archive with accession no. PRJNA728834.

Gene functional annotation, enrichment analysis, and weighted gene co-expression network (WGCNA) analysis

Gene functions was predicted using the following databases: NR (NCBI non-redundant protein sequences), Pfam (protein family), Swiss-Prot, KEGG (Kyoto Encyclopedia of Genes and Genomes), and GO (gene ontology). The GOseq package in R (version 3.4.3) and the KOBAS software were used to perform the GO and KEGG pathway enrichment, respectively. The gene expression levels were calculated using FPKM (fragments per kilobase of exon per million fragments mapped). Genes with FPKM values > 1 in each sample were used for WGCNA analysis using the R package WGCNA (v1.42).

Differential expression analysis and candidate gene determination

Differentially expressed genes (DEGs) were analyzed by DESeq and detected with the parameter FDR $\leq$ 0.01 & |Log2Ratio| $\geq$ 1. The plant transcription factors were predicted by iTAK software, and the differentially expressed transcription factors were screened by mapping to the DEGs database and the genes related to plant hormones in DEGs were screened based on Gene Ontology (GO) functional annotation. Twenty-nine DEGs were randomly selected and verified using the real-time (RT)-PCR to detect the accuracy of transcriptome data. Toolbox for Biologists v1.087 (Chen, Chen et al. 2020) was used to construct the heatmap.

Identification of lncRNAs and AS events

In order to analyze the coding potential in the transcripts, four most widely used computational approaches including coding potential calculator (CPC) (Li, Zhang et al. 2014), coding-non-coding Index (CNCI) (Altschul, Madden et al. 1997), coding potential assessment tool (CPAT) (Deng, Li et al. 2006), and

pfam protein structure domain analysis were combined. Transcripts were regarded as lncRNA candidates using the parameter: lengths > 200 nt & > 2 exons.

To identify the AS events, the AS talavista algorithm (Foissac and Sammeth 2007) based on the assembled GTF file from the PacBio Iso-Seq and Illumina data was conducted and five types of AS events including IR, ES, AD, AA and MEE were classified with the default settings.

Isoform identification and reverse transcription (RT)-PCR

To distinguish AS events, the reverse transcription (RT)-PCR was performed. Briefly, the RNA used for Iso-Seq was isolated using EZ-10 DNA away RNA Mini-Preps Kit (Sangon, https://www.sangon.com) and synthesized to complementary DNA using the Hifair® 1st Strand cDNA Synthesis SuperMix for qPCR gDNA digester plus (Yeasen). Isoform-specific primers were designed using Vector NTI software (**Supplementary Table S1**). PCR amplification was performed using 2×Taq PCR Master Mix (K1034, https://www.apexbt.com/) and the PCR products were showed in agarose gel stained with GoodView™ Nucleic Acid Stain (Apr-13-2021, HGV-, https://www.sbsbio.com).

BlazeTaq™ SYBR® Green qPCR Mix 2.0 (GeneCopoeia, USA) was used for RT-qPCR reactions on a CFX96 Real-time System (BIO-RAD, USA). The amplification program was briefly described in previous study (Wagner 2013). To ensure the precise and reproducible results, each sample was replicated three times and Ef1α in potato were used for each sample as an endogenous control (**Supplementary Table S1**). The formula $F = 2^{-\Delta\Delta Ct}$ was used to calculate the relative expression levels of selected genes.

Alternative polyadenylation analysis

Alternative splicing and alternative polyadenylation (APA) of pre-mRNAs greatly contribute to transcriptome diversity, coding capacity of a genome and gene regulatory mechanisms in eukaryotes. TADIS pipeline (Transcriptome Analysis Pipeline for Isoform Sequencing) was used to further analyze the full-length non-chimeric sequence (FLNC) to identify APA (Abdel-Ghany, Hamilton et al. 2016) and MEME analysis was also performed to identify poly (A) sequence signals in our data.

Results

Illumina and PacBio sequencing data analysis

To construct a comprehensive regulation network in response to drought stress, control and treated samples were performed SMRT sequencing on the PacBio platform and next-generation sequencing (NGS) on the Illumina Hi-Seq platform. High-quality RNA was extracted from above tissues of “Desiree” after PEG-6000 treatments (0 h, 1 h, 3 h, 6 h, 12 h, 24 h and 48 h) for Illumina Hi-Seq and pool together for construction of Iso-Seq librariy. After removing low-quality reads and adapter sequences, 38,347,790-46,349,352 reads were obtained and more than 80% sequences were mapped the reference genome (Table 1). The Pearson’s correlation coefficient showed that all correlation values between the three

replicates ranged from 0.734 to 0.925, indicating a perfect positive correlation and that the sequencing results could be used for further analysis (Fig. 1). For Iso-Seq, the PacBio platform generated 33.24 Gb clean data. Among these data, 464,332 circular consensus (CCS) read were obtained using Iso-seq pipeline with min full pass $\geq$3 and min predicted accuracy $\geq$ 0.9. Then, 375,387 full length non chimeric (FLNC) sequences were determined by searching for the polyA tail signal and the 5' and 3' cDNA primers and 121,434 high-quality consensus sequences were obtained by clustering the FLNC sequences. SMRT sequencing can offer better transcript integrity (up to 50 kb in length) than NGS and has greatly contributed in maize (Wang, Tseng et al. 2016), populous (Chao, Gao et al. 2019) oilseed rape (Yao, Liang et al. 2020) and sorghum (Mace, Tai et al. 2013, Abdel-Ghany, Hamilton et al. 2016), but no literature has been reported in potato. Significantly longer isoforms were detected by SMRT than by NGS assembly, demonstrating better isoform integrity (Fig. 2a). In addition, the overall expression levels of transcripts detected by SMRT were higher than those detected by NGS, indicating better quantification accuracy (Fig. 2b).

Analysis of gene expression and transcriptomic changes in response to drought treatment

DESeq was used for differential expression analysis between biologically duplicated sample groups. False Discovery Rate (FDR) was used as key indicators for differential expression gene screening. Fold Change$\geq$2 and FDR<0.01 were used as screening criteria. In total, 31,637 genes were detected in the library of six samples drought treated at different times, with common genes and unique genes between six libraries and the number of DEGs in D24h samples was the highest (Fig. 3a and Supplementary Table S4). Moreover, almost all of DEGs in the libraries showed moderate expression levels with only a small percentage of the genes expressed in high levels (Fig. 3b). As shown in Fig. 3c, more than 88% of the DEGs had a -3□fold change $\leq$3. Transcription factors (TFs) play important roles in plant abiotic stress responses (Jian, Ma et al. 2018). To further analyze the function of the DEGs, we used iTAK software to predict transcription factors in DEGs. In total, 664 DEGs belonging to 28 TF families were identified. The TF family members were unevenly distributed. Seven TF families, bHLH (69), MYB (60), ERF (53), bZIP (46), NAC (43) WRKY (35) and HD-ZIP (33) accounted for more than 51% (339/664) of all TFs (Fig. 3d). Plant hormones, such abscisic acid (ABA), auxin (IAA), and cytokinin (CK), play vital roles in abiotic stresses (Verma, Ravindran et al. 2016). In total, 919 DEGs belonging to eight classifications of plant hormones, and ABA, IAA and CK-related genes accounted for the majority, accounting for about 94% of all hormone-related genes (Fig. 3e). To combine the hormone profiling data with the expression profiles of genes involved in hormone metabolism and signaling, we analyzed 231 genes involved in plant hormone metabolism and visualized them as heat maps after functional classification (Fig. 4). Heat map results showed that DEGs were most involved in plant hormone signaling.

Gene profiling in response to drought stress in potato

To explore drought stress response mechanisms in potato, the expression patterns were calculated (Fig. 5a). Moreover, 98,624 genes were conducted using the K-means algorithm. Fourteen clusters were obtained and sixty-three pathways were enriched, such as carbon metabolism, carbon fixation in

photosynthetic organisms, porphyrin and chlorophyll metabolism, protein processing in endoplasmic reticulum and oxidative phosphorylation (Fig. 5b and Fig. 5c). Furthermore, WGCNA was also performed to screen genes response to drought stress by detecting molecular modules. Thirteen modules were generated and more than one module was significantly correlated with each sample. Blue, darkred, midnightblue, orange, cyan, lightgreen, and darkgreen were significantly correlated with DCK, D1h, D3h, D6h, D12h, D24h and D48h, respectively (Fig. 5d and Fig. 5e). Then, KEGG enrichment analysis were conducted to explore the enriched pathways in each module correlated to each sample (Fig. 5f). Two pathways, galactose metabolism (ko00052) and fatty acid metabolism (ko01212), were enriched in more than two treated samples. Plant-pathogen interaction (ko04626) and glutathione metabolism (ko00480) were enriched in early stages after drought stress. In addition, through the comparative analysis of the control group and the other six samples processed in different periods, a total of 12,798 differentially expressed genes after deduplication were found (Fig. S1, Supplementary Table S2 and S3). The functional enrichment of these DEGs during the six drought treatment period was conducted (Fig. 6a) and twenty-nine DEGs were selected to confirm the transcriptome sequencing data using the real-time (RT)-PCR (Fig. 6b).

Putative roles of AS for drought stress regulation in potato

AS plays critical roles in abiotic stress response by increasing genetic diversity in plants (Ling, Alshareef et al. 2017, Laloum, Martin et al. 2018). However, little information about AS involved in drought stress response has been obtained in potato. In this study, 21,756 AS events were categorized into five patterns, including 12,453 retained intron (RI), 2,493 skipping exon (SE), 236 mutually exclusive exons (MX), 4,481 alternative 3' splicing sites (A3), and 2,093 alternative 5' splicing sites (A5) using Isoform sequencing technology, which yields long reads without assembly (Fig. 7a). Among these AS events, IR events predominated, accounting for 57.2%, followed by A3, SE, A5 and MX (Fig. 7b), which was consistent with findings of previous reports in other plant species (Yan, Gao et al. 2019). Four random selected genes with AS events were successfully validated by reverse transcription (RT)-PCR (Fig. 7c).

To identify the co-expression network of genes involved in AS events, the WGCNA package was used and three significant modules were identified (brown, blue and turquoise). The CK sample correlated to the brown module (r=0.79, P=$2e^{-5}$), whereas the turquoise (r=0.71, P=$3e^{-4}$) and blue (r=0.77, P=$5e^{-5}$) modules were significantly correlated with samples after 6 and 24 h treatments, respectively (Fig. 7d). KEGG enrichment of the significant modules was also performed. Six, eleven and four pathways were significantly enriched in brown, turquoise and blue modules. The galactose metabolism (ko00052) was significantly enriched in turquoise and blue modules, which were treated with PEG-6000 after 6 and 24 h, respectively (Fig. 7e).

Putative roles of lncRNA involved in drought stress regulation in potato

LncRNA is a major component in transcriptome sequencing, and plays important roles in the regulation process after transcription (Rinn, Chang et al. 2012, Kwenda, Birch et al. 2016, Muret, Klopp et al. 2017, Li,

Liu et al. 2021). However, few information about lncRNA in potato has been obtained. In this study, 3,345 lncRNAs with various lengths were identified using composite computational algorithm based on the PacBio sequencing data (Fig. 8a, b). More than 54% of lncRNAs were located in the intergenic region while less than 7% were located in the antisense region according to reference annotations (Fig. 8c). To identify the co-expression network of target genes of lncRNAs, the WGCNA package was used and yellow and brown modules were significantly enriched (Fig. 8d). KEGG analysis indicated that the spliceosome (ko03040) and ubiquitin mediated proteolysis (ko04120) pathway were most enriched in CK and drought stress after 6 h samples, respectively.

Alternative polyadenylation analysis

In eukaryotes, polyadenylation is an important transcriptional modification in mRNAs (Elkon et al., 2013; Bentley, 2014). Alternative polyadenylation can easily increase the transcripts variety and regulate gene expression by producing different transcripts (Shen et al. 2011; Sherstnev et al., 2012; Fu et al., 2016). The iso-seq data allowed us to detect polyadenylation sites in potato using the program TAPIS (Abdel-Ghany, Hamilton et al. 2016). In this study, 26,153 poly (A) sites from 13,010 genes detected in the Iso-Seq data were identified (Supplementary Table S6). On average, 2.01 poly (A) sites per gene were found and 856 genes had at least five poly (A) sites (Fig. 9a). In addition, a clear nucleotide bias of the flanking region of all poly (A) sites was found. To confirm this phenomenon, we investigated the motifs of 50 nucleotides upstream of predominant sites using MEME analysis. Furthermore, three conserved polyadenylation signals, i.e. TTTTGT, GCCCCC and GGGGCG were overrepresented in upstream of poly (A) cleavage sites (Fig. 9b).

Discussion

Potato has important economic value, while they are vulnerable to drought stress. Therefore, it is important to understand the drought resistance mechanism of potatoes. Transcriptome research is one of the indispensable tools for understanding life processes. However, RNA-Seq technology based on the second-generation high-throughput sequencing platform often cannot accurately obtain or assemble complete transcripts, and cannot identify Isoform, homologous genes. Full-length transcriptome sequencing based on PacBio SMRT does not need to interrupt RNA fragments, and directly reverse transcribes the full-length cDNA, which enables the feasible determination of the whole transcriptome at the isoform level (Abdel-Ghany, Hamilton et al. 2016). Previous studies on potato drought stress only analyzed the DEGs based on the second-generation sequencing technology. In our research, NGS and SMRT sequencing technologies were used to study the transcription profile of potato in response to drought stress. The seedling leaves of the variety "Desiree" after drought stress were used for NGS and SMRT sequencing at six time points from 0 to 48 hours. Statistical analysis of sequencing data revealed the important roles of DEGs, alternative splicing, LncRNA and APA in potato under drought stress.

DEGs involved in drought stress response

In this study, SMRT and NGS technologies were combined to conduct a global survey of drought-treatment potato transcriptome. With the intensification of drought stress, the number of DEGs were increased. The number reached the highest at 24h after drought stress treatment (Fig. S1 and Supplementary Table S2). which was consistent with the findings previously reported in potato (Chen, Li et al. 2020). Transcription factors and plant hormone-related genes in differentially expressed genes were analyzed, and found that drought stress-related transcription factors such as bZIP and plant hormone-related genes such as ABA-related genes may played a dominant role. The functional annotation of differentially expressed genes confirmed their important role in drought stress, and 29 DEGs we verified using real-time (RT)-PCR (Supplementary Table S1 and Fig. 6b). Moreover, the WGCNA screening and KEGG enrichment analysis of drought stress response genes suggested that plant-pathogen interactions and glutathione metabolism pathways were enriched in the early stage of drought stress, and galactose metabolism and fatty acid metabolism pathways were enriched in at least two treatment samples, which indicated these pathways play an important role in the potato drought response process. Previous researches reported that galactose is involved in response to drought, high salt and low temperature tolerance (Taji, Ohsumi et al. 2002), and transgenic studies demonstrated that glutathione helped *Tortula ruralis* adapt to rehydration following rapid desiccation (Dhindsa 1991).

AS events involved in drought stress response

One of the advantages of PacBio sequencing is its ability to detect AS events. (Wang, Wang et al. 2018, Zhu, Li et al. 2018). For different biological responses, AS is an effective mechanism to increase the complexity and flexibility of the transcriptome. (Li, Dai et al. 2017). Previous studies have confirmed that AS events play critical roles in potato growth and abiotic stress response (Lu, Tian et al. 2008, Nicolas, Rodriguez-Buey et al. 2015). AS events identified during the potato drought stress response process were analyzed in this study. The proportion of different AS events was consistent with the findings previously reported in other plant species (Yan, Gao et al. 2019, Xie, Teng et al. 2020). Our experiment verified the AS events detected above, and the results showed that the size of the gel band fragments was consistent with the splice isoforms identified from the Iso-Seq data (Supplementary Table S1 and Fig. 7c) and the functional annotations of these genes also indicated that they were involved in drought stress response. In summary, we have verified that the four randomly selected genes undergo alternative splicing under drought stress, and their functional annotations are also related to plant stress tolerance. Our experiments results and sequencing data indicated that potatoes may increase their protein diversity under drought stress through alternative splicing, and thereby improve their environmental adaptability.

LncRNA plays an important role in drought stress response

As a type of non-coding RNA, lncRNA act as regulators in a series of biological processes (Heo, Lee et al. 2013). Previous researches suggested that lncRNA has potential functions in potato defense response and development (Kwenda, Birch et al. 2016, Ramirez Gonzales, Shi et al. 2021). In this study, a total of 3,445 lncRNAs of different lengths were identified through four methods: CPC analysis, CNCI analysis, pfam protein domain analysis, and CPAT analysis (Fig. 8a and Fig. 8b). The WGCNA and KEGG analysis

results of lncRNA target genes show the term ‘ubiquitin mediated proteolysis’ was enriched in the brown module, indicating that lncRNAs may be involved in the ubiquitination regulation in response to drought stress and it also provided additional effective candidate genes for future functional characterization (Fig. 8c).

APA is essential for drought resistance of potato

Polyadenylation refers to the covalent linkage of polyadenylic acid to messenger RNA (mRNA) molecules. In the process of protein biosynthesis, this is part of the way to produce mature mRNA ready for translation. In eukaryotes, polyadenylation is a mechanism that interrupts mRNA molecules at their 3'ends. The polyadenylic acid tail (or poly A tail) protects mRNA from exonuclease attack, and is very important for the termination of transcription, export of mRNA from the nucleus, and translation. A comprehensive genome-wide APA map draft consisting of 26,153 poly (A) sites from 13,030 genes was constructed in this study (Supplementary Table S6). The nucleotide bias was evaluated and obvious nucleotide bias in the flanking regions of all poly (A) sites were found (Fig. 9b). These findings are consistent with those of previous studies in *Brassica napus* (Yao, Liang et al. 2020). Two conserved polyadenylation motifs in *B. napus* were identified , while three conserved polyadenylation motifs in potato were identified, indicating that the polyadenylation motifs are species-specific (Li and Du 2014). Previous studies have shown that alternative polyadenylation of RNA is critical for gene function by increasing the complexity of the transcriptome and regulating gene expression (Wu, Liu et al. 2011). The APA locus found in our study may protect its transcription products through post-transcriptional modification of drought stress-related transcription factors and structural genes, and thereby regulate the drought resistance of potatoes.

Declarations

Ethics approval

The potato material used in this study has been approved by Southwest University, we comply with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora, All methods for this study are in compliance with the provisions of Chinese law and international guidelines and regulations.

Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and analyzed during the current study are available in the NCBI Sequence Read Archive repository under Bioproject PRJNA728834 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA728834).

Competing interests

The authors declared no conflict of interest in the authorship and publication of this document.

Funding

This study was supported by the National Key Research and Development Program of China (2018YFE0127900), the National Natural Science Foundation of China (32101659), the Science and Technology Partnership Program, Ministry of Science and Technology of China (KY201904016), the Talent Introduction Program of Southwest University Project (SWU019008) and the Germplasm Creation Research Program of Southwest University.

Authors' contribution

D.Q.L conceived and designed the experiments. H.J.J. and H.N.S. performed the experiments. H.J.J., H.N.S., R.R.L., W.Z.Z., L.N.S., J.C.W. and V.K. analyzed the data. H.J.J. and H.N.S. wrote the paper. All authors reviewed the manuscript.

Acknowledgements

The authors thank South Western University (SWU) for providing all the facilities for the research work.

References

1. Abdel-Ghany, S. E., M. Hamilton, J. L. Jacobi, P. Ngam, N. Devitt, F. Schilkey, A. Ben-Hur and A. S. Reddy. (2016). A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun. **7**: 11706. doi: 10.1038/ncomms11706
2. Alexandersson, E., J. A. H. Danielson, J. Rade, V. K. Moparthi, M. Fontes, P. Kjellbom and U. Johanson. (2010). Transcriptional regulation of aquaporins in accessions of *Arabidopsis* in response to drought stress. Plant J. **61**(4): 650-660. doi: 10.1111/j.1365-313X.2009.04087.x
3. Alexandersson, E., L. Fraysse, S. Sjovall-Larsen, S. Gustavsson, M. Fellert, M. Karlsson, U. Johanson and P. Kjellbom. (2005). Whole gene family expression and drought stress regulation of aquaporins. Plant Mol Biol. **59**(3): 469-484. doi: 10.1007/s11103-005-0352-1
4. Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**(17): 3389-3402. doi: 10.1093/nar/25.17.3389
5. Blum, A. (1996). Crop responses to drought and the interpretation of adaptation. J. Plant Growth Regul. **20**(2): 135-148.

6. Bolger, A. M., M. Lohse and B. Usadel. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. **30**(15): 2114-2120. doi: 10.1093/bioinformatics/btu170
7. Chao, Q., Z. F. Gao, D. Zhang, B. G. Zhao, F. Q. Dong, C. X. Fu, L. J. Liu and B. C. Wang. (2019). The developmental dynamics of the Populus stem transcriptome. Plant Biotechnol J. **17**(1): 206-219. doi: 10.1111/pbi.12958
8. Chen, C., H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He and R. Xia. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. Mol Plant. **13**(8): 1194-1202. doi: 10.1016/j.molp.2020.06.009
9. Chen, Y. K., C. H. Li, J. Yi, Y. Yang, C. X. Lei and M. Gong. (2020). Transcriptome response to drought, rehydration and Re-Dehydration in potato. Int J Mol Sci. **21**(1). doi: 10.3390/ijms21010159
10. Deng, Y., J. Li, S. Wu, Y. Zhu, Y. Chen and F. He. (2006). Integrated NR database in protein annotation system and its localization. CE. **32**(5): 71-74.
11. Dhindsa, R. S. (1991). Drought stress, enzymes of glutathione metabolism, oxidation injury, and protein synthesis in tortula ruralis. Plant Physiol. **95**(2): 648-651. doi: 10.1104/pp.95.2.648
12. Evers, D., I. Lefevre, S. Legay, D. Lamoureux, J. F. Hausman, R. O. G. Rosales, L. R. T. Marca, L. Hoffmann, M. Bonierbale and R. Schafleitner. (2010). Identification of drought-responsive compounds in potato through a combined transcriptomic and targeted metabolite approach. J Exp Bot. **61**(9): 2327-2343. doi: 10.1093/jxb/erq060
13. Foissac, S. and M. Sammeth. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res. **35**: W297-W299. doi: 10.1093/nar/gkm311
14. Galvez, J. H., H. H. Tai, M. Lague, B. J. Zebarth and M. V. Stromvik. (2016). The nitrogen responsive transcriptome in potato (*Solanum tuberosum* L.) reveals significant gene regulatory motifs. Sci Rep. **6**:26090. doi: 10.1038/srep26090
15. Golldack, D., C. Li, H. Mohan and N. Probst. (2014). Tolerance to drought and salt stress in plants: unraveling the signaling networks. Front Plant Sci. **5**:151. doi: 10.3389/fpls.2014.00151
16. Gong, L., H. X. Zhang, X. Y. Gan, L. Zhang, Y. C. Chen, F. J. Nie, L. Shi, M. Li, Z. Q. Guo, G. H. Zhang and Y. X. Song. (2015). Transcriptome profiling of the potato (*Solanum tuberosum* L.) plant under drought stress and water-stimulus conditions. PloS One **10**(5):e0128041. doi:10.1371/journal.pone.0128041
17. Gong, Z. Z., L. M. Xiong, H. Z. Shi, S. H. Yang, L. R. Herrera-Estrella, G. H. Xu, D. Y. Chao, J. R. Li, P. Y. Wang, F. Qin, J. J. Li, Y. L. Ding, Y. T. Shi, Y. Wang, Y. Q. Yang, Y. Guo and J. K. Zhu. (2020). Plant abiotic stress response and nutrient use efficiency. Sci China Life Sci. **63**(5): 635-674. doi: 10.1007/s11427-020-1683-x
18. Gupta, N. and V. K. Verma. (2019). Next-Generation sequencing and its application: empowering in public health beyond reality. In: Arora P. (eds) Microbial Technology for the Welfare of Society. Microorganisms for Sustainability, vol 17. Springer, Singapore. doi:10.1007/978-981-13-8844-6_15

19. Heo, J. B., Y. S. Lee and S. Sung. (2013). Epigenetic regulation by long noncoding RNAs in plants. Chromosome Res. **21**(6-7): 685-693. doi: 10.1007/s10577-013-9392-6
20. Hu, H. H. and L. Z. Xiong. (2014). Genetic engineering and breeding of drought-resistant crops. Annu Rev Plant Biol. **65**:715-41. doi: 10.1146/annurev-arplant-050213-040000
21. Jian, H. J., Ma, J. Q., Wei, L. J., Liu, P., Zhang, A. X., Yang, B., Li, J. N., Xu, X. F and liu, L. J. (2018). Integrated mRNA, sRNA, and degradome sequencing reveal oilseed rape complex responses to *Sclerotinia sclerotiorum* (Lib.) infection. Sci Rep **8**(1):10987. doi: 10.1038/s41598-018-29365-y
22. Kim, D., B. Landmead and S. L. Salzberg. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods. **12**(4): 357-60. doi: 10.1038/nmeth.3317
23. Kwenda, S., P. R. Birch and L. N. Moleleki. (2016). Genome-wide identification of potato long intergenic noncoding RNAs responsive to *Pectobacterium carotovorum* subspecies brasiliense infection. BMC Genomics **17**(1): 614. doi: 10.1186/s12864-016-2967-9
24. Laloum, T., G. Martin and P. Duque. (2018). Alternative splicing control of abiotic stress responses. Trends Plant Sci. **23**(2): 140-150. doi: 10.1016/j.tplants.2017.09.019
25. Lamaoui, M., M. Jemo, R. Datla and F. Bekkaoui. (2018). Heat and drought stresses in crops and approaches for their mitigation. Front Chem. **6**:26. doi: 10.3389/fchem.2018.00026
26. Li, A., J. Zhang and Z. Zhou. (2014). PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC bioinformatics. **15**(1): 311. doi: 10.1186/1471-2105-15-311
27. Li, X. Q. and D. Du. (2014). Motif types, motif locations and base composition patterns around the RNA polyadenylation site in microorganisms, plants and animals. BMC Evol Biol **14**: 162. doi: 10.1186/s12862-014-0162-7
28. Li, Y., C. Dai, C. Hu, Z. Liu and C. Kang. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. Plant J. **90**(1): 164-176. doi: 10.1111/tpj.13462
29. Li, M. X., Xie, S., Ma, L.P., Zhao, Z. C., Gong, H. B., Sun, Y. S and Huang, T. (2021). Transcriptome analysis reveals that long noncoding RNAs contribute to developmental differences between medium-sized ovarian follicles of Meishan and Duroc sows. Sci Rep **11**(1):22510.
30. Ling, Y., S. Alshareef, H. Butt, J. Lozano-Juste, L. Li, A. A. Galal, A. Moustafa, A. A. Momin, M. Tashkandi, D. N. Richardson, H. Fujii, S. Arold, P. L. Rodriguez, P. Duque and M. M. Mahfouz. (2017). Pre-mRNA splicing repression triggers abiotic stress signaling in plants. Plant J. **89**(2): 291-309. doi: 10.1111/tpj.13383
31. Lu, S. W., D. Tian, H. B. Borchardt-Wier and X. Wang. (2008). Alternative splicing: a novel mechanism of regulation identified in the chorismate mutase gene of the potato cyst nematode Globodera rostochiensis. Mol Biochem Parasitol. **162**(1): 1-15. doi: 10.1016/j.molbiopara.2008.06.002
32. Mace, E. S., S. S. Tai, E. K. Gilding, Y. H. Li, P. J. Prentis, L. L. Bian, B. C. Campbell, W. S. Hu, D. J. Innes, X. L. Han, A. Cruickshank, C. M. Dai, C. Frere, H. K. Zhang, C. H. Hunt, X. Y. Wang, T. Shatte, M. Wang, Z. Su, J. Li, X. Z. Lin, I. D. Godwin, D. R. Jordan and J. Wang. (2013). Whole-genome sequencing

reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun. **4**: 2320. doi: 10.1038/ncomms3320

33. Massa, A. N., K. L. Childs and C. R. Buell. (2013). Abiotic and biotic stress responses in *solanum tuberosum* group phureja DM1-3 516 R44 as measured through whole transcriptome sequencing. Plant Genome **6**: 3. doi:10.3835/plantgenome2013.05.0014
34. Massa, A. N., K. L. Childs, H. N. Lin, G. J. Bryan, G. Giuliano and C. R. Buell. (2011). The transcriptome of the reference potato genome *solanum tuberosum* group phureja clone DM1-3 516R44. PloS One **6**(10): e26801. doi: 10.1371/journal.pone.0026801
35. Moon, K. B., D. J. Ahn, J. S. Park, W. Y. Jung, H. S. Cho, H. R. Kim, J. H. Jeon, Y. I. Park and H. S. Kim. (2018). Transcriptome profiling and characterization of drought-tolerant potato plant (*Solanum tuberosum* L.). Mol Cells. **41**(11): 979-992. doi: 10.14348/molcells.2018.0312
36. Muret, K., Klopp, C., Wucher, V., Esquerré, D., Legeai, F., Lecerf, F., Désert, C., Boutin, M., Jehl, F., Acloque, H., Giuffra, E., Djebali, S., Foissac, S., Derrien, T and Lagarrigue, S. (2017). Long noncoding RNA repertoire in chicken liver and adipose tissue. Genet Sel Evol 49(1):6. doi: 10.1186/s12711-016-0275-0
37. Nejat, N. and N. Mantri. (2017). Plant immune system: crosstalk between responses to biotic and abiotic stresses the missing link in understanding plant defence. Curr Issues Mol Biol. **23**: 1-16. doi: 10.21775/cimb.023.001
38. Nicolas, M., M. L. Rodriguez-Buey, J. M. Franco-Zorrilla and P. Cubas. (2015). A recently evolved alternative splice site in the BRANCHED1a gene controls potato plant architecture. Curr Biol .**25**(14): 1799-809. doi: 10.1016/j.cub.2015.05.053
39. Ramirez Gonzales, L., L. Shi, S. B. Bergonzi, M. Oortwijn, J. M. Franco-Zorrilla, R. Solano-Tavira, R. G. F. Visser, J. A. Abelenda and C. W. B. Bachem. (2021). Potato *CYCLING DOF FACTOR 1* and its lncRNA counterpart *StFLORE* link tuber development and drought response. Plant J. **105**(4): 855-869. doi: 10.1111/tpj.15093
40. Rhoads, A. and K. F. Au. (2015). PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. **13**(5): 278-89. doi: 10.1016/j.gpb.2015.08.002
41. Rinn, J. L and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. Annu Rev Biochem. **81**:145-66. doi: 10.1146/annurev-biochem-051410-092902
42. Salehi-Lisar, S. Y. and Bakhshayeshan-Agdam, H. (2016). Drought stress in plants: causes, consequences, and tolerance. In: Hossain M., Wani S., Bhattacharjee S., Burritt D., Tran LS. (eds) Drought Stress Tolerance in Plants, Vol 1. Springer, Cham. doi: 10.1007/978-3-319-28899-4_1
43. Shin, D., S. J. Moon, S. Han, B. G. Kim, S. R. Park, S. K. Lee, H. J. Yoon, H. E. Lee, H. B. Kwon, D. Baek, B. Y. Yi and M. O. Byun. (2011). Expression of *StMYB1R-1*, a novel potato single MYB-Like domain transcription factor, increases drought tolerance. Plant Physiol. **155**(1): 421-32. doi: 10.1104/pp.110.163634
44. Shinozaki, K. and K. Yamaguchi-Shinozaki. (2007). Gene networks involved in drought stress response and tolerance. J Exp Bot. **58**(2): 221-7. doi: 10.1093/jxb/erl164

45. Shinozaki, K., K. Yamaguchi-Shinozaki and M. Seki. (2003). Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol. **6**(5): 410-17. doi: 10.1016/s1369-5266(03)00092-x

46. Sicher, R. C., D. Timlin and B. Bailey. (2012). Responses of growth and primary metabolism of water-stressed barley roots to rehydration. J Plant Physiol. **169**(7): 686-95. doi: 10.1016/j.jplph.2012.01.002

47. Sprenger, H., A. Erban, S. Seddig, K. Rudack, A. Thalhammer, M. Q. Le, D. Walther, E. Zuther, K. I. Kohl, J. Kopka and D. K. Hincha. (2018). Metabolite and transcript markers for the prediction of potato drought tolerance. Plant Biotechnol J. **16**(4): 939-950. doi: 10.1111/pbi.12840

48. Steijger, T., J. F. Abril, P. G. Engstrom, F. Kokocinski, T. J. Hubbard, R. Guigo, J. Harrow, P. Bertone and R. Consortium. (2013). Assessment of transcript reconstruction methods for RNA-seq. Nat Methods. **10**(12): 1177-84. doi: 10.1038/nmeth.2714

49. Taji, T., C. Ohsumi, S. Iuchi, M. Seki, M. Kasuga, M. Kobayashi, K. Yamaguchi-Shinozaki and K. Shinozaki. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. Plant J. **29**(4): 417-26. doi: 10.1046/j.0960-7412.2001.01227.x

50. Verma ,V., Ravindran, P and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. BMC Plant Biol 16:86. doi: 10.1186/s12870-016-0771-y

51. Wagner, E. M. (2013). Monitoring gene expression: quantitative real-time rt-PCR. Methods Mol Biol. **1027**: 19-45. doi: 10.1007/978-1-60327-369-5_2

52. Wang, B., E. Tseng, M. Regulski, T. A. Clark, T. Hon, Y. P. Jiao, Z. Y. Lu, A. Olson, J. C. Stein and D. Ware. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat Commun. **7**: 11708. doi: 10.1038/ncomms11708

53. Wang, M., P. Wang, F. Liang, Z. Ye, J. Li, C. Shen, L. Pei, F. Wang, J. Hu, L. Tu, K. Lindsey, D. He and X. Zhang. (2018). A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. New Phytol. **217**(1): 163-178. doi: 10.1111/nph.14762

54. Wu, T. D. and C. K. Watanabe. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. **21**(9): 1859-75. doi: 10.1093/bioinformatics/bti310

55. Wu, X., M. Liu, B. Downie, C. Liang, G. Ji, Q. Q. Li and A. G. Hunt. (2011). Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. Proc Natl Acad Sci U S A. **108**(30): 12533-8. doi: 10.1073/pnas.1019732108

56. Xie, L., K. Teng, P. Tan, Y. Chao, Y. Li, W. Guo and L. Han. (2020). PacBio single-molecule long-read sequencing shed new light on the transcripts and splice isoforms of the perennial ryegrass. Mol Genet Genomics. **295**(2): 475-489. doi: 10.1007/s00438-019-01635-y

57. Xu, X., S. K. Pan, S. F. Cheng, B. Zhang and P. G. S. Consortiu. et al. (2011). Genome sequence and analysis of the tuber crop potato. Nature **475**(7355): 189-95. doi: 10.1038/nature10158

58. Yamaguchi-Shinozaki, K. and K. Shinozaki. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol. **57**: 781-803. doi:

10.1146/annurev.arplant.57.032905.105444

59. Yan, H. L., Y. W. Gao, L. L. Wu, L. Y. Wang, T. Zhang, C. H. Dai, W. X. Xu, L. Feng, M. Ma, Y. G. Zhu and Z. Y. He. (2019). Potential use of the Pteris vittata arsenic hyperaccumulation-regulation network for phytoremediation. J Hazard Mater. **368**: 386-396. doi: 10.1016/j.jhazmat.2019.01.072
60. Yao, S., F. Liang, R. A. Gill, J. Huang, X. Cheng, Y. Liu, C. Tong and S. Liu. (2020). A global survey of the transcriptome of allopolyploid Brassica napus based on single-molecule long-read isoform sequencing and Illumina-based RNA sequencing data. Plant J **103**(2): 843-857. doi: 10.1111/tpj.14754
61. Zhang, N., B. Liu, C. Ma, G. Zhang, J. Chang, H. Si and D. Wang. (2014). Transcriptome characterization and sequencing-based identification of drought-responsive genes in potato. Mol Biol Rep. **41**(1): 505-17. doi: 10.1007/s11033-013-2886-7
62. Zhu, C. H., X. F. Li and J. Y. Zheng. (2018). Transcriptome profiling using Illumina- and SMRT-based RNA-seq of hot pepper for in-depth understanding of genes involved in CMV infection. Gene. **666**: 123-133. doi: 10.1016/j.gene.2018.05.004
63. Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. Cell **167**(2): 313-324. doi: 10.1016/j.cell.2016.08.029

Table

Table 1 is available in the Supplementary Files section.

Figures

Figure 1

Pearson correlation evaluation among twenty-one samples. DCK represents control groups; D1h represents drought treatment after 1 h; D3h represents drought treatment after 3 h; D6h represents Drought treatment 6h; D12h represents Drought treatment 12h; D24h represents Drought treatment 24h; D48h represents Drought treatment 48h. Each sample has three biological replicates.

Figure 2

Comparisons between NGS and SMRT. a. The difference between NGS and SMRT sequencing length. **b.** The difference in FPKM of NGS and SMRT sequencing.

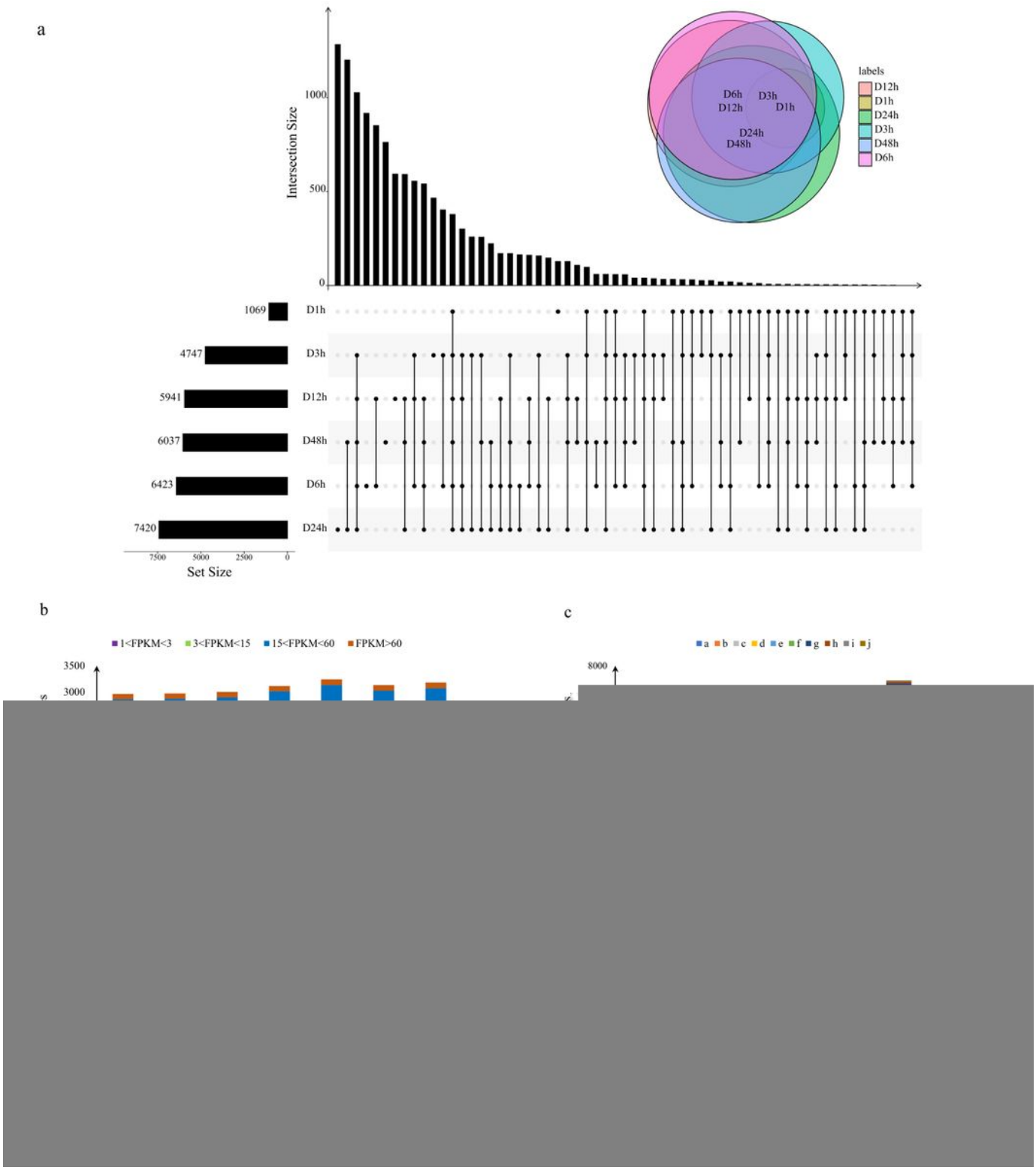

Figure 3

DEGs expression profiles of six samples treated at different periods after drought treatment. a. Venn Diagram of genes detected in the six samples. **b.** Statistical analysis of gene expressions identified in six samples. **c.** Fold changes in the DEGs identified after drought treatment. [(a) Number of DEGs with a log2 fold change ≤−9; (b) number of DEGs with −9 < log2 fold change ≤−7; (c) number of DEGs with −7 < log2 fold change ≤−5; (d) number of DEGs with −5 < log2 fold change ≤−3; (e) number of DEGs with −3 < log2

fold change ≤−1; (f) number of DEGs with 1 < log2 fold change ≤3; (g) number of DEGs with 3 <log2 fold change ≤5; (h) number of DEGs with 5 < log2 fold change ≤7; (i) number of DEGs with 7 < log2 fold change ≤9; (j) number of DEGs with log2 fold change 9]. **d.** Statistical analysis of genes encoding TFs identified in the DEGs. **e.** Statistical analysis of genes encoding hormones identified in the DEGs. [ABA: abscisic acid; IAA: auxin; BR: brassinosteroids; CK: cytokinin; GA: gibberellin; JA: jasmonic acid; SA: Salicylic acid].

Figure 4

Heat maps of DEGs associated with hormone metabolism and signal transduction in potato. Expression of genes involved in ABA: abscisic acid, IAA: auxin, CK: cytokinin, GA: gibberellin, JA: jasmonic acid, SA: Salicylic acid metabolism and signaling after drought treatment.

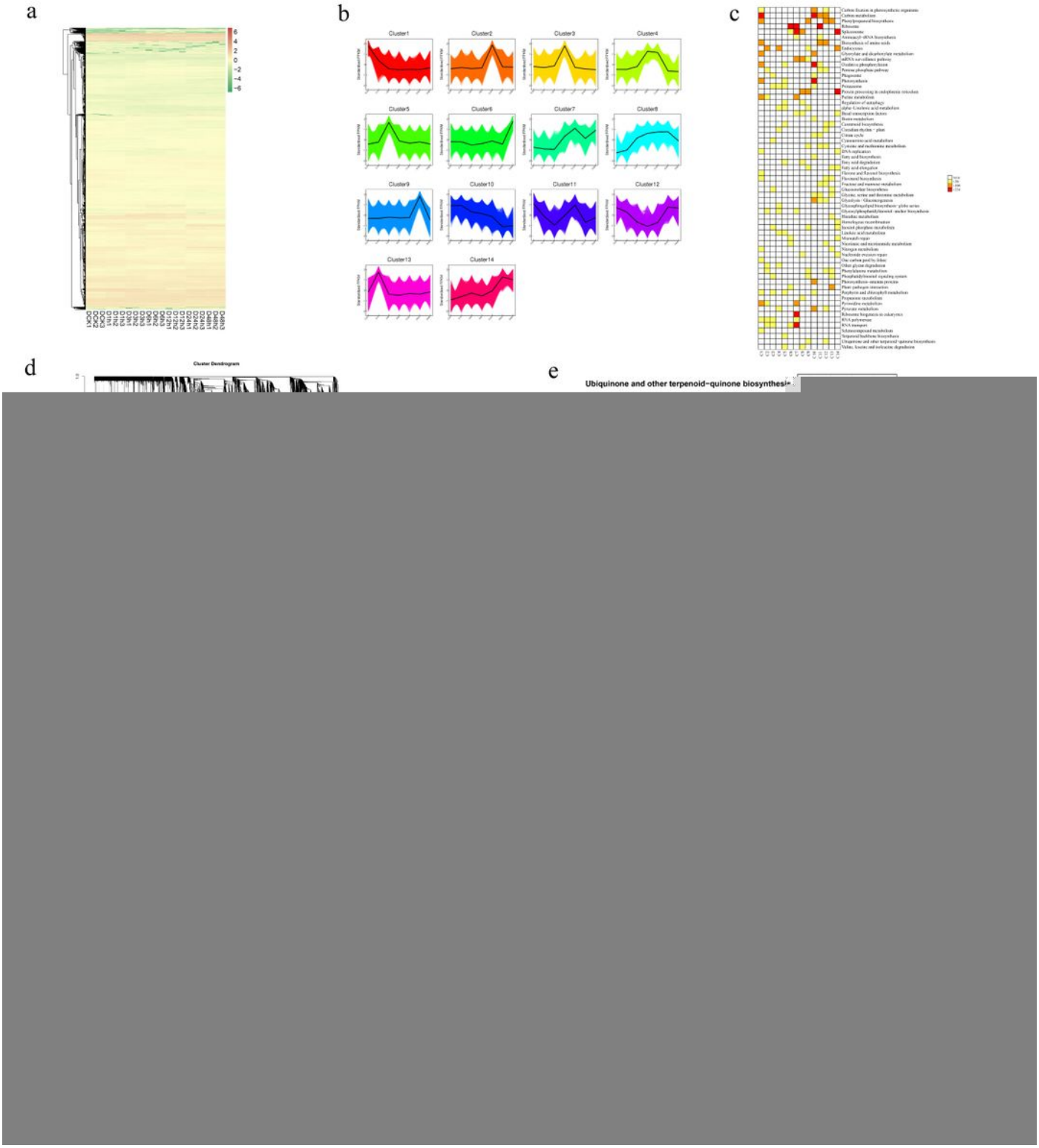

Figure 5

Global patterns of differentially expressed genes (DEGs) under drought stress. a. Hierarchical cluster analysis of DEGs. **b.** K-means clustering (C1−C14) showing transcriptome expression profiles. **c.** Gene Ontology enrichment among the fourteen clusters. Yellow to red, significant enrichment; white, none. **d.** Weighted gene co-expression network (WGCNA) analysis of DEGs. **e.** Expression levels of genes among the modules most significantly correlated with drought-treated samples. **f.** KEGG enrichment among the

modules most significantly correlated with drought-treated samples. The figure shows the heat map and KEGG enrichment of four drought-treated samples (D1h, D6h, D24h, D48h).

Figure 6

KEGG classification of differentially expressed genes in drought-treated samples and RT-PCR verification of several candidate genes. a. KEGG classification of differentially expressed genes in six drought-treated samples (D1h, D3h, D6h, D12h, D24h and D48h). **b.** Correlation evaluation between real-time (RT) PCR and RNA-seq.

Figure 7

Expression profiles and functional analysis of alternative splicing (AS) events. a. Classification of AS events. Diagrams show the following AS events: retained intron (RI), alternative 3' splicing sites (A3), alternative 5' splicing sites (A5), mutually exclusive exons (MX) and exon skipping (ES). The numbers of AS events. Filled boxes represent exons, introns are represented by black lines. **b.** The number of differentially expressed AS events in DCK, D1h, D3h, D6h, D12h, D24h and D48h. **c.** Reverse-transcription (RT)-PCR validation of AS events. The RT-PCR validation of AS events for four genes, all of them are intron retention. Gel bands in each figure show DNA markers and PCR results in five samples (DCK, D3h, D6h, D12h and D24h). The transcript structure of each isoform is shown in the right panel. Exons are represented by blue filled boxes, introns are represented by lines, and the dotted line indicates that there is no intron corresponding to the first isoform in the second isoform of the same gene. Primers are designed to span the splicing events. PCR primers (F, forward and R, reverse) are shown on the first isoform of each gene. The length of each expected PCR product is shown after the transcript structure. **d.** WGCNA analysis of genetic modules relating to each sample. **e.** Enriched KEGG pathway of AS events involved in each module.

Figure 8

Expression profiles and functional analysis of lncRNA. a. Venn diagram of long non-coding RNA (lncRNA) prediction using four databases. **b.** LncRNA length and targeted genes. **c.** Distribution of different lncRNAs. **d.** WGCNA analysis of genetic modules relating to each sample.

Figure 9

Alternative polyadenylation analysis. a. Distribution of the number of poly (A) sites per gene model. **b.** MEME analysis identified poly (A) signals in transcripts.

Supplementary Files

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

- Fig.S1.tif
- tableS1.xlsx
- tableS2.xlsx
- tableS3.xlsx
- tableS4.xlsx
- tableS5.xlsx
- tableS6.xlsx
- table1.xlsx