Transcriptome sequencing and whole genome expression profiling of purple-fleshed sweetpotato under salt stress

Mohamed Hamed Arisha

Zagazig University Faculty of Agriculture <https://orcid.org/0000-0002-3547-2505>

Hesham Aboelnasr

national research center

Liu Yaju

xuzhou research institute

Tang Wei

xuzhou sweet research institute

Gao Run Fei

xuzhou research institute

Yan Hui

xuzhou research institute

Kou Meng

Xuzhou research institute

Wang Xin

xuzhou research institute

Zhang Yungang

xuzhou research institute

Li Qiang (✉ jsxzgysz@163.com)

Research article

Keywords: Purple-fleshed sweetpotato, Salt stress, Expression profile, RNA-sequencing, Transcriptome

Posted Date: July 1st, 2019

DOI: <https://doi.org/10.21203/rs.2.10828/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Abstract Background: Purple-fleshed sweetpotato (PFSP) is one of the most important crops in the world which helps to bridge the food gap and contribute to solve the malnutrition problem especially in developing countries. Salt stress is seriously limiting its production and distribution. Due to lacking of reference genome, transcriptome sequencing is offering a rapid approach for crop improvement with promising agronomic traits and stress adaptability. **Results:** Five cDNA libraries were prepared from the third true leaf of PFSP at seedlings stage (Xuzi-8 cultivar) treated with 200 mM NaCl for 0, 1, 6, 12, 48 hours. Using second and third generation technology, Illumina sequencing generated 170,344,392 clean high-quality long reads that were assembled into 15,998 unigenes with an average length 2178 base pair and 96.55% of these unigenes were functionally annotated in the NR protein database. A number of 537 unigenes failed to hit any homologs which may be considered as novel genes. The current results indicated that sweetpotato plants behavior during the first hour of salt stress was different than the other three time points. Furthermore, expression profiling analysis identified 4, 479, 281, 508 significantly expressed unigenes in salt stress treated samples at the different time points including 1, 6, 12, 48 hours, respectively as compared to control. COG analysis revealed that the highest number of genes was categorized as “post-translational modification, protein turnover, chaperones” followed by “signal transduction mechanisms”. Gene ontology (GO) enrichment analysis showed that salt stress caused changes in metabolic process (Secondary and amino acids), hormone response, cellular processes and single organism processes. These findings suggest that salt stress tolerance in PFSP plants may be mainly depended on three factors including the balance between protein biosynthesis and degradation, preventing or correcting damage caused by miss-folding due to salt stress and transforming the certain stimulus induced by salt stress into a biochemical signal which activates more genes specifically involved in salt stress tolerance.

1. Introduction

Sweetpotato (*Ipomoea batatas* (L.) Lam.), the only crop plant belongs to *Convolvulaceae* family with starchy storage roots. Purple-fleshed sweetpotato (PFSP) considered to be an important source for anthocyanin which displays strong antioxidant properties [1]. It is also considered as an important staple source of calories and proteins which consumed by all age groups. In terms of agricultural production sweetpotato considered as the seventh most important food crop in the world [2].

Salinity is a global problem caused vast area of lands remaining uncultivated. Exposure of sweetpotato plants to salt stress resulting in problems such as ion imbalance, mineral deficiency, osmotic stress, ion toxicity and oxidative stress [3]. Ultimately, these conditions interact with several cellular components including DNA, protein, lipids and pigments. That's in rule impeding plant development and affect sweetpotato production [4]. Therefore, introducing of salt tolerant sweetpotato cultivar became necessary.

With the fact of environmental stress and climate change there is an urgent need to accelerate crops breeding with higher production and stress tolerance traits [5]. In sweetpotato transcriptome sequencing offers a rapid approach for crop improvement with promising agronomic traits and stress adaptability. Several transcriptome sequencing studies have been conducted on hexaploidy sweetpotato genome [6-8]. However,

having a complex genome structures ($2n=6x=90$), sweetpotato still didn't achieve a reference genome which covered a few percent of genome, so still a long way from the reference genome [9].

In the present study, second and third generation sequencing technology were used to establish a useful database of transcriptomes sequencing as well as differentially expressed genes in PFSP leaves under salt stress conditions. In total 102,845,433 high quality reads were assembled into 16,856 transcripts giving 15,998 unigenes. Our results provide novel insights into PFSP response to salt stress and identified numerous specific genes involved in salt stress defense mechanisms. That's in role can be used to guide future efforts towards breeding of sweetpotato salt resistant cultivars.

2. Materials And Methods

2.1. Plant materials

Xuzi-8, a high quality and early mature cultivar with purple flesh storage roots, was used in this experiment. Its storage roots contain 6% soluble sugar and more than 80 mg anthocyanin /100 g fresh weight. This cultivar was obtained from Xuzhou Institute of Agricultural Sciences in Jiangsu Xuhuai District, China.

2.2. Treatments and experimental design

Apical stem cuttings (15-20 cm) of Xuzi-8 cultivar were growing in hydroponics culture using Hoagland nutrient medium for four weeks. Seedlings were exposed to salt stress (200 mM NaCl were added to Hoagland solution). 0.1 g sample was taken from the upper third leaf after 0, 1, 6, 12, 48 hour from salt treated and untreated plants for the next generation sequencing. While, for the third-generation sequencing, one sample was collected by mixing the upper third leaf of sweetpotato seedlings after 1, 6, 12, 48 hour of salt stress and one control samples. All collected samples were immediately frozen in liquid nitrogen and stored at -80°C until processing for RNA extraction and all treatments were done in triplicate.

2.3. RNA extraction

Total RNA was extracted from collected samples with the TRIZOL method (Life technologies, Carlsbad, CA) according to the manufacturer's protocol. RNA degradation and contamination were monitored on 1% agarose gels. RNA purity was checked using the Nanophotometer® spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

2.4. Library Preparation for Transcriptome Analysis

A total amount of 3 µg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's protocol and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer $\times 5X$. First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase

RNAse H. Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNAse H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

2.5. Clustering and sequencing

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina platform and paired-end reads were generated.

2.6. Data analysis Quality control

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing poly-N and low-quality reads from raw data. At the same time, Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data with high quality.

2.7. Transcriptome assembly

The data from all libraries/samples were pooled into one big left fq file (read1 files) and right fq files (read2 files) into one big right [10]. Transcriptome assembly was accomplished based on the left and right fq using Trinity [11] with min_kmer_cov set to 2 by default and all other parameters set default.

2.8. Gene functional annotation

Gene function was annotated based on the following databases: Nr (NCBI non-redundant protein sequences, <https://www.ncbi.nlm.nih.gov/>) with E-value cut-off of $1e-5$ Nt (NCBI non-redundant nucleotide sequences <https://www.ncbi.nlm.nih.gov/>) with E-value cut-off of $1e-5$ Pfam (Protein family, <http://pfam.sanger.ac.uk/>) with E-value cut-off of $1e-2$ KOG/COG (Clusters of Orthologous Groups of proteins, <https://www.ncbi.nlm.nih.gov/cog/>) with E-value cut-off of $1e-3$ and Swiss-Prot (A manually annotated and reviewed protein sequence database, <http://www.ebi.ac.uk/uniprot/>) with E-value cut-off of $1e-5$.

Based on the NR and Pfam annotations, Blast2GO (v2.5) was used to obtain GO (Gene Ontology) annotations (<http://www.geneontology.org>) according to the molecular functions, biological processes and cellular component ontologies[12]. GO enrichment analysis of the differentially expressed genes (DEGs) was implemented by the GO-seq R packages based Wallenius non-central hyper-geometric distribution [13] which can adjust for gene length bias in DEGs.

KEGG [14] is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (<http://www.genome.jp/kegg/>). KOBAS software were used [15] to test the statistical enrichment of DEGs in KEGG.

2.9. SNP calling and SSR detection

Picard - tools v1.41 and samtools v0.1.18 were used to sort, remove duplicated reads and merge the bam alignment results of each sample. GATK software was used to perform SNP calling. Variants were kept for quality using the following parameters (1) mapping quality filter equal to PASS; (2) Quality Depth (QD) >2; (3) Mapping Quality (MQ) > 40; (5) QUAL >30; Moreover, variants were further filtered if coverage < 10, if cluster SNPs more than 2 in 5bp window, if SNP around Indel within 5bp.

SSR of the transcriptome were identified using MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>), and primer for each SSR was designed using Primer3 (<http://primer3.sourceforge.net/releases.php>).

2.10 Quantification of gene expression levels and differential expression analysis

Gene expression levels were estimated by RSEM [16] for each sample. Clean data were mapped back onto the assembled transcriptome. Read count for each gene was obtained from the mapping results.

Differential expression analysis of two conditions/groups was performed using the DESeq2R package. DESeq2 provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value ≤ 0.001 found by DESeq2 were assigned as differentially expressed [17]. $Q\text{-value} \leq 0.001 \& |\log_2 FC| \geq 1$ was set as the threshold for significantly differential expression.

3. Results

3.1. Sequencing and *de novo* assembly of sweetpotato transcriptome under salt stress conditions

Five cDNA libraries were prepared from the third true leaf of PFSP seedlings (Xuzi-8 cultivar) treated with 200 mM NaCl for 0, 1, 6, 12, 48 hours. These libraries were sequenced using Illumina high-throughput sequencing platform. After removing the low-quality reads and all possible contamination, a total of 170,344,392 clean reads with Q20 >96.73% and a GC percentage between 45.07% and 46.50% (Table 1). Each library was represented by over than 30 million high-quality reads, with number ranging from 32,830,183 to 35,663,873. Due to the lack of a reference genome, the clean reads resulted from the transcriptome sequences were aligned and assembled using Trinity software. After further clustering and assembly, a total of 21,497,466; 20,272,643; 21,954,725; 19,121,890 and 19,998,709 mapped reads were obtained with percentage 60.26%, 61.79%, 61.62%, 59.33% and 58.87% of total reads at different time points (0, 1, 6, 12, 48 hours), respectively. The quality of these assemblies and unigenes length distribution are shown in Table 1. Statistics on unigenes

and transcripts length resulted from mixed second and third generations sequencing were performed using PacBio's officially recommended cogent software (Table1 and 2, Figure 1). In addition, the total number of CDS was 30,615 of which 23,245 CDS mapped to the protein database.

Table 1: Next generation sequencing statistical summary of sequenced and assembled results

Table 2: Third generation sequencing statistical summary of sequenced and assembled results

Figure 1: Assembly result sequence length distribution map of transcripts and unigenes in XuZi-8 sweetpotato cultivar. The horizontal axis represents the length intervals of the transcripts and unigenes, and the vertical axis represents the number of transcripts and unigenes.

3.2. Functional annotation

To annotate the obtained unigenes, a BlastX search against the NR NCBI protein database with cut-off E-value of 10^{-5} based on sequence similarity. In total, 15,461 unigenes were detected (Table 3) that comparability with known gene sequence in all databases corresponding to approximately 96.64% of total unigenes including Clusters of orthologous groups (COG), Gene ontology (GO), Kyoto encyclopaedia of genes and genomes (KEGG), eukaryotic orthologous group (KOG), protein family (PFAM), Swiss-Prot., NCBI non-redundant protein sequences (Nr). According to Figure 2, the species that gave the best BlastX matches was *Nicotiana sylvestris* (21.30%) followed by *Nicotiana tomentosiformis* (20.69%), *Solanum tuberosum* (9.16%), *Sesamum indicum* (7.46%), *Coffea canephora* (6.00%), *Solanum lycopersicum* (4.74%), *Solanum penellii* (4.58%), *Ipomoea batatas* (3.07%), *Vitis vinifera* (2%), *Ipomoea nil* (1.57%) and others (19.22%).

Table 3: Statistics of unigenes annotated in public database

Figure 2: Species distribution of the top BlastX matches of the transcriptome unigenes in the non-redundant protein database (Nr) data base.

3.3. Gene ontology (GO) and KOG classifications

For functional categories of 15,998 successfully annotated unigenes, a total of 12,481 genes (78.01%) (Table 3) were assigned to at least one GO term. These GO terms were categorized into 48 functional groups divided into three categories including biological process, cellular component and molecular function (Figure 3). For biological process, the highest categories were metabolic (8291 unigenes, 53.55%) followed by cellular process (7774, 50.21%) then single organism process (6181, 39.92%). In the category of molecular function, the most abundant groups included catalytic activity (6369, 41.14%) and binding (6513, 42.07%). Furthermore, the most abundant group for cellular components was cell parts (7198, 46.49%). According to GO annotation results there were 199 unigenes recorded to contribute in biological processes involved in response to salt stress. Moreover, rigorous algorithm ($FDR \leq 0.001$, $\log_2 \text{FC-ratio} \geq 1$) was used to measure the significance level of the 199 obtained genes. Out of these 199 unigenes there were no genes significantly expressed during the first hour involved in response to salt stress. Furthermore, there were three unigenes significantly expressed at 6, 12, 48 hour of salt stress (unigene243, unigene_3644, unigene_7343). In addition, at 6 hours of salt stress there were two unigenes including unigene_7826 and unigene_9718 significantly expressed. At 12 hours of salt stress number of 13368 and 13566 unigenes were significantly expressed.

Another two unigenes (15315 and 9718) were significantly expressed at 48 hours of salt stress. Furthermore, depend on COG annotation, there were 6 genes out of the 199 involved in the defense mechanisms and 7 unigenes have uncharacterized protein sequence.

Genetic orthologous relationships, combines evolutionary relationships were used to classify the potential functions into different orthologous clusters (COG). In total 10,020 genes were subdivided into 25 functional classes as shown in Figure 4. Among the 25 groups, “general function prediction only” represented the largest group (1,871 unigenes, 16.89%) followed by “post translational modification, protein turnover, chaperons” (1,271 unigenes, 11.47%) then “signal transduction mechanisms” (1,037 unigenes, 9.36%). In addition, it was interesting to note that 87 genes were aligned to the “defense mechanisms” cluster (Figure 4). Rigorous algorithm ($FDR \leq 0.001$, $\log_2 FC\text{-ratio} \geq 1$) were applied to measure the significance level of the 87 obtained genes. Out of these defense mechanisms genes there were no significant expressed genes during the first hour of salt stress. Furthermore, there was one unigene (unigene802) significantly expressed during 6, 12, 48 hours of salt stress, while there are another gene (unigene_1120) significantly expressed only at 6 and 48 hours of salt stress. On the other hand, the remained genes were normally expressed without any significant difference as compared to control. In addition, there were 8 unigenes with uncharacterized protein sequence (according to nr annotation results).

Figure 3: Gene ontology (GO) classifications in sweetpotato (Xuzi-8 cultivar), the percentage indicate the proportion of unigenes with the GO annotations.

Figure 4: Clusters of orthologous groups (COG) classification in Xuzi-8 sweetpotato cultivar. Genes from the same Orthologous have the same function, so that direct functional annotations to other members of the same KOG cluster.

3.4 KEGG annotation

KEGG pathway annotation for 15,461 unigenes was obtained. A total of 5,965 sequences were assigned to 125 pathways. The largest groups in the KEGG pathways were “Metabolic pathways (ko01100)” (1,508 unigenes, 25.28%) and “Biosynthesis of secondary metabolites (ko01110)” (733 unigenes, 12.28%), which came in the first rank. Followed by “Carbon metabolism (ko01200)” (336 unigenes, 5.63%), “Ribosome (ko03010)” (303 unigenes, 5.08%), “Plant hormone signal transduction (ko04075)” (249 unigenes, 4.17%), “Biosynthesis of amino acids (ko01230)” (249 unigenes, 4.17%) and “Photosynthesis (ko00195)” (199 unigenes, 3.34%). These specific enrichments KEGG pathways and mechanisms are involved in response to salt stress in sweetpotato (Xuzi-8 cultivar).

Figure 5: The most enriched KEGG clusters in Xuzi-8 sweetpotato cultivar. The most enriched 22 clusters out of 123 clusters were presented in this figure.

3.5. Expression patterns of PFSP unigenes in response to salt stress

The results in Figure 6 showed that the highest number of DEGs was induced at 48 hours of salt stress followed by 6 h then 12 h, while 1 h gave the lowest number of DEGs. Transcriptional level at 1, 6, 12, 48 h as compared to control induced expression values 4, 529, 341 and 663 as up-regulated unigenes, and 0, 672,

422 and 1531 as down-regulated, respectively. Furthermore, there was 15,534; 14,450; 14,703 and 13,330 normally expressed unigenes during 1, 6, 12 and 48 hours of salt stress. In addition, there were 119 up-regulated genes, 87 down-regulated genes, 12,384 genes normal and 211 unknown genes common under all durations of salt stress (Figure 6).

Figure 6: Statistical chart of DEGs transcriptome in response to salt stress. Transcriptional level of five libraries including 1, 6, 12 and 48 hours of salt stress treatment as compared to control.

3.6. Detection of salt-induced genes related to salt tolerance

RPKM read counts were used to identify DEGs significance level between control and salt-stressed samples using the rigorous algorithm ($FDR \leq 0.001$, $\log_2$ FC-ratio ≥ 1). As shown in Figure 6 there were 119 up-regulated genes common in various durations of salt stress (6, 12, 48 hours). Out of these 119 unigenes there were 92 genes significantly increased at 6, 12, 48 hours of salt stress. Pfam results showed that of these 92-salt induced genes, 77 unigenes (83.69%) were found to have a known function. A number of 5 unigenes (5.43%) out of the significantly expressed genes showed no homology with known sequence. Furthermore, 4, 479, 281, 508 unigenes were significantly expressed in salt stress treated samples at the different time points of salt stress including 1, 6, 12, 48 hours, respectively. During the first hour of salt stress there were four significantly expressed unigenes including (unigene11357, unigene15171, unigene16391 and unigene7424). After 6 hours a number of 479 unigenes were significantly expressed, these genes belong to 45 different protein families and most of these families are involved in stress tolerance or defense mechanisms, metabolism, *etc.* At 12 hours of salt stress, a number of 281 unigenes were significantly expressed contained in 32 different gene families. There were 508 unigenes significantly expressed in leaf tissues after 48 hours of salt stress with 63 different gene families (Figure 7). Based on the abovementioned results, all of these significantly unigenes during the three time points of salt stress (6, 12, 48 hours) definitely have a role under salt stress conditions or may be contribute in regulating the salt response in Xuzi- 8 sweetpotato cultivar.

Figure 7: Comparison of four transcriptomes for classification of DEGs and statistics of sequence annotation of DEGs. (A) and (D); Venn diagram analysis of up-regulated unigenes and all induced unigenes, respectively. (B) and (E) Hierarchical cluster analysis of up-regulated unigenes and all induced unigenes, respectively. (C) and (F); Statistical chart of sequence annotation of up-regulated unigenes and all induced unigenes, respectively. Sequence showed homology with known sequences were classified as “hits” and sequence showed no homology with known sequences were classified as “no hits”.

3.7. SSR and SNP identification

For further application of sweetpotato SSRs and SNPs were discovered using assembled transcriptomes. A total of 24,559 SSRs was identified in transcriptomes present in 15,976 sequences. Furthermore, the number of sequence containing more than one SSR was 5,762 and the number of SSRs present in compound formation was 2,215. In addition, the major types of the identified SSRs were mono-nucleotide (19,380), di-nucleotide (2,791), tri-nucleotide (2,199), tetra-nucleotide (123), penta-nucleotide (33) and hexa-nucleotide (25). The most SSR motif was A/T (19282) followed by AG/CT (1,964), AT/AT (669), AAG/CTT (596), CCG/CGG (330), AAT/ATT (309), ATC/ATG (247) and AGC/CTG (217). A total of 562174 SNPs between transcriptomes was identified, among which 348,495 were transitions, and 213,679 were transversions. These

SSRs and SNPs identified in this study provided a valuable resource for future studies on genetic linkage mapping and the analysis of interesting traits in sweetpotato.

4. Discussion

Sweetpotato is a hexaploid non-model crop with a complex genome lacking high quality reference genome [18]. To date the current understanding of the complex physiological and molecular mechanisms of salt tolerance in sweetpotato remains limited [19, 20]. High throughput RNA sequencing is required for identification of candidate genes involved in salt stress tolerance. In addition, it will be helpful for better understanding of salts stress tolerance mechanisms in sweetpotato [5].

In the present study, deep sequencing analyses of PFSP under different durations of salt stress were characterized. Using second and third generation technology, Illumina sequencing generated 170,344,392 clean high-quality long reads that were assembled into 15,998 unigenes with an average length 2178 base pair and 96.55% of these unigenes were functionally annotated in the NR protein database. A number of 537 unigenes failed to hit any homologs which may be considered as novel genes.

The current results revealed that the obtained sequences highly matched with top seven species with percent ranging from 4.58 to 21.3% including *Nicotiana sylvestris*, *Nicotiana tomentosiformis*, *Solanum tuberosum*, *Sesamum indicum*, *Coffea canephora*, *Solanum lycopersicum* and *Solanum pennelli*. In addition, the online available sequence of sweetpotato genome gave only 3.07% matches coming in the 8th rank with the current experiment obtained sequence. That's mean that 96.93% of our data are not present in the sweetpotato online data. Therefore, the current annotation results will improve sweetpotato genome annotation and facilitate the discovery of genetic resources that are responsible for salt stress in hexaploid sweetpotato.

The results in Figure 3 showed that the most dominant Go-terms which were identified during the different durations of salt stress included metabolic, cellular, single organism and response to stimulants. These results are in agree with [21] in diploid halophytic sweetpotato, [22] in *Ipomoea imperati* [23]. The present COG results indicated that there were 16.89% belongs to “general function prediction only” which mean that these genes are poorly characterized, in addition, 4.34% genes with unknown function. The highest number of genes was categorized as “post-translational modification, protein turnover, chaperones” in our study. Which mean that, sweetpotato plants under salt stress conditions, starts the key mechanisms of chemical modifications including regulating of enzymes activity, localization and interaction with other cellular molecules such as proteins, nucleic acid, lipids and co-factors [24]. Furthermore, protein turnover related genes control the balance between protein biosynthesis and degradation, which is very important for determining the resistant or sensitivity of the plant to salt stress. That's may be due to if the protein biosynthesis is more than breakdown indicates an anabolic state, in this case the plant tolerance to stress will be enriched [25-28]. Moreover, Protein chaperones' major function is to prevent or correct damage caused by miss-folding due to salt stress [29-32]. The second COG group of genes involved in “signal transduction mechanisms” which is mainly transforming the certain stimulus induced by salt stress into a biochemical signal which activates more genes specifically involved in salt stress tolerance [33].

In the present study, to identify significant gene expression changes under salt stress, differentially expressed unigenes were analyzed using the rigorous algorithm equation ($FDR \leq 0.001$, $\text{Log}_2\text{FC-ratio} \geq 1$) depending on RPKM reads.

The current results indicated sweetpotato plants behavior during the first hour of salt stress was different than the other three time points. During the first hour of salt stress a number of four unigenes only were expressed with significant difference (unigene11357, unigene15171, unigene16391 and unigene7424). The first unigene (11357) belongs to SBP gene family which have important role in leaf development, vegetative phase change and may be have a role in stress response [34]. The second (unigene-15171) belongs to HSP70 which is essential regulator to maintain internal cell stability and prevent aggregation under physical or chemical pressure [35]. The third unigene (16391) is a plant invertase/ pectin methyl-esterase inhibitor. The invertase affects growth and development consisting with its activities such as stress response. In addition, pectin methyl-esterase determines the solidity of cell wall including root development and permeability. The fourth unigene (7424) belongs to uncharacterized protein family and unknown function [36]. According to the current results presenting of only four significantly expressed unigenes in leaf tissues during the first hour of salt stress may be due to salt stress signals slightly received or didn't received by leaves during the first hour [37] in XuZi-8 sweetpotato cultivar.

At six hours of salt stress, there were 529 up-regulate unigenes giving 479 significantly expressed unigenes. A number of ten genes were the highest (four folds higher than control) among all significantly expressed genes including (unigene128, 16087, 16131, 2278, 4002, 55, 5655, 9231, and 9447). Furthermore, it was noted that these superior unigenes are three different functions including protein detoxification, energy production and conservation, and dehydrin. In addition, out of these 479 unigenes there were six unigenes involved in the defense mechanisms against salt stress including (Unigene128, 13856, 13920, 14004, 396, and 6144). Except for the unigene128 and 396 all the other unigenes belong to ABC- transporter G-family which is involved in growth, development and response to abiotic stress. In addition, the unigene396 belongs to pentatricopeptide-repeat proteins (PPR- family) which play important roles in organelle RNA metabolism, plant development and involved in abiotic stress response which have been previously reported in rice [38, 39]. There were 5 genes specifically involved in salt stress response including (unigene243, 3644, 7343, 7826 and 9718). The unigene243 and 3644 belong to ABA insensitive-like protein which involved in ABA and salt stress responses and act as a positive component in glucose signal transduction [40]. The unigene7343 belongs to EID-like F-box protein family which is a known drought response regulation. Unigene7826 were found to be belonged to WCOR 413-like protein which is responsible for cryoprotection of plasma membrane against dehydration stress which has been discovered earlier in sweetpotato genome, these findings are in agree with [41, 42] in wheat. The unigene 9718 was detected as a putative salt stress responsive protein isoform-1 which has been discovered before in sweetpotato genome.

At twelve hours of salt stress a number of 341 unigenes were up-regulated with 281 significantly expressed unigenes. Out of the 281 genes there were three unigenes (unigene3153, 3480 and 3682) expressed with four folds more than control, all these four genes belonging to Ahpc-TSA family with intracellular signal transduction biological function which constitutes an enzymatic defense against salt stress [43]. There were 10 unigenes among the 281 unigenes, are specifically involved in defense mechanisms against salt stress

(unigene10470, 11483, 12478, 12588, 12916, 13856, 14004, 2507, 4295 and 6144) which were included under ABC-transporter family [44]. In addition, there were 5 unigenes with a biological function as response to salt stress (unigene13368, 13566, 234, 3644 and 7343) which belongs to three different families including BEL1-like homeodomain protein-1 (BLH-1), ABA insensitive like protein (bZIP transcription factor) and EID-like F-box protein-3 [45].

At 48 hours of salt stress, there were 663 up-regulated unigenes with 508 significantly expresses. A number of 9 unigenes were the highest significantly expressed among the 508 unigenes. Out of these 9 unigenes, four unigenes (unigene16087, 16288, 2278 and 9017) are responsible for energy production and conservation; two unigenes (unigene2618, 2863) were responsible for signal transduction mechanism, one unigene (unigene657) as a reverse transcriptase, one unigene (unigene10431) as protein import into nucleus, and one (unigene7343) was involved in posttranslational modification, protein turnover, chaperones. In addition, there were 6 unigenes are specifically involved in defense mechanisms to salt stress including unigene10470, 12916, 2507, 4295, 5014 and 6144. All these six unigenes belong to ABC-transporter family which is involved in growth, development and stress response [44]. Moreover, a number of 5 unigenes (unigene15315, 243, 3644, 7343 and 9718) were having the same biological function as response to salt stress. Out of the five unigene four have been expressed significantly in leaf tissues during 6, 12 hours of salt stress, while the unigene15315 (cysteine-rich receptor-like protein kinase-2) start to expressed significantly after 48 hours of salt stress [46].

Declarations

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Li Qiang, and Arisha designed the study. Arisha conducted the experiments, analyzed the results, performed the figures and wrote the manuscript. Li Qiang Liu Ya-ju, Hesham Aboelnasr revised the manuscript. Tang Wei, Kou Meng, Yan Hui, Gao Run-fei, Wang Xin and Zhang Yun-gang assisted in the experimental procedures.

Acknowledgements

This work was supported by the earmarked fund for China Agriculture Research System (CARS-10, Sweetpotato), the “333 talent project” fund of Jiangsu Province (BRA2016265), and China, Talented Young Scientist Program of Ministry of Science and Technology.

References

1. Truong V, Avula R, Pecota K, Yencho CJHov, processing v: Sweetpotatoes. 2011:717-737.

2. Bouwkamp JC: Sweet potato products: A natural resource for the tropics: CRC Press; 2018.
3. Zörb C, Geilfus CM, Dietz KJJPB: Salinity and crop yield. 2019, 21:31-38.
4. Majeed A, Muhammad Z: Salinity: A Major Agricultural Problem—Causes, Impacts on Crop Productivity and Management Strategies. In: *Plant Abiotic Stress Tolerance*. Springer; 2019: 83-99.
5. Bhadauria V: Next-generation Sequencing and Bioinformatics for Plant Science: Caister Academic Press; 2017.
6. Tao X, Gu Y-H, Wang H-Y, Zheng W, Li X, Zhao C-W, Zhang Y-ZJPo: Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [*Ipomoea batatas* (L.) Lam.]. 2012, 7(4):e36234.
7. Lin Y, Zou W, Lin S, Onofua D, Yang Z, Chen H, Wang S, Chen XJPo: Transcriptome profiling and digital gene expression analysis of sweet potato for the identification of putative genes involved in the defense response against *Fusarium oxysporum* f. sp. *batatas*. 2017, 12(11):e0187838.
8. Li R, Zhai H, Kang C, Liu D, He S, Liu QJljog: De novo transcriptome sequencing of the orange-fleshed sweet potato and analysis of differentially expressed genes related to carotenoid biosynthesis. 2015, 2015.
9. Loebenstein G, Thottappilly G: The Sweetpotato: Springer Netherlands; 2009.
10. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QJNb: Full-length transcriptome assembly from RNA-Seq data without a reference genome. 2011, 29(7):644.
11. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber MJNp: De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. 2013, 8(8):1494.
12. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa AJNar: High-throughput functional annotation and data mining with the Blast2GO suite. 2008, 36(10):3420-3435.
13. Young MD, Wakefield MJ, Smyth GK, Oshlack AJGb: Gene ontology analysis for RNA-seq: accounting for selection bias. 2010, 11(2):R14.
14. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu TJNar: KEGG for linking genomes to life and the environment. 2007, 36(suppl_1):D480-D484.
15. Mao X, Cai T, Olyarchuk JG, Wei LJB: Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. 2005, 21(19):3787-3793.
16. Li B, Dewey CNJBb: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. 2011, 12(1):323.

17. Storey JD, Tibshirani RJPotNAoS: Statistical significance for genomewide studies. 2003, 100(16):9440-9445.
18. Kumpatla SP, Buyyarapu R, Abdurakhmonov IY, Mammadov JA: Genomics-assisted plant breeding in the 21st century: technological advances and progress. In: *Plant breeding*. Intechopen; 2012.
19. Gupta B, Huang B: Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. *Int J Genomics* 2014, 2014:701596-701596.
20. Sunkar RJMMB: Plant stress tolerance. 2010, 639:401.
21. Luo Y, Reid R, Freese D, Li C, Watkins J, Shi H, Zhang H, Loraine A, Song B-H: Salt tolerance response revealed by RNA-Seq in a diploid halophytic wild relative of sweet potato. *Sci Rep* 2017, 7(1):9624-9624.
22. Solis J, Baisakh N, Brandt SR, Villordon A, La Bonte D: Transcriptome Profiling of Beach Morning Glory (*Ipomoea imperati*) under Salinity and Its Comparative Analysis with Sweetpotato. *PLOS ONE* 2016, 11(2):e0147398.
23. Luo Q, Teng W, Fang S, Li H, Li B, Chu J, Li Z, Zheng Q: Transcriptome analysis of salt-stress response in three seedling tissues of common wheat. *The Crop Journal* 2019.
24. Alberts B, Bray D, Hopkin K, Johnson AD, Lewis J, Raff M, Roberts K, Walter P: Essential cell biology: Garland Science; 2013.
25. Mahajan S, Tuteja N: Cold, salinity and drought stresses: An overview. *Archives of Biochemistry and Biophysics* 2005, 444(2):139-158.
26. Murata N, Takahashi S, Nishiyama Y, Allakhverdiev SI: Photoinhibition of photosystem II under environmental stress. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 2007, 1767(6):414-421.
27. Mittler R, Vanderauwera S, Gollery M, Van Breusegem F: Reactive oxygen gene network of plants. *Trends in Plant Science* 2004, 9(10):490-498.
28. Parida AK, Das AB: Salt tolerance and salinity effects on plants: a review. *Ecotoxicology and Environmental Safety* 2005, 60(3):324-349.
29. Botella MA, Rosado A, Bressan RA, Hasegawa PMJPas: Plant adaptive responses to salinity stress. 2005:37-70.
30. Wang W, Vinocur B, Shoseyov O, Altman AJTips: Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. 2004, 9(5):244-252.
31. Wang W, Vinocur B, Altman AJP: Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. 2003, 218(1):1-14.
32. Zhu J-KJC: Abiotic stress signaling and responses in plants. 2016, 167(2):313-324.

33. Zhu J-KJArope: Salt and drought stress signal transduction in plants. 2002, 53(1):247-273.
34. Yang Z, Wang X, Gu S, Hu Z, Xu H, Xu C: Comparative study of SBP-box gene family in Arabidopsis and rice. *Gene* 2008, 407(1-2):1-11.
35. Mayer MP, Bukau B: Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* 2005, 62(6):670-684.
36. Hothorn M, Wolf S, Aloy P, Greiner S, Scheffzek K: Structural Insights into the Target Specificity of Plant Invertase and Pectin Methylesterase Inhibitory Proteins. 2004, 16(12):3437-3447.
37. Negrao S, Schmöckel S, Tester M: Evaluating physiological responses of plants to salinity stress, vol. 119; 2016.
38. Chen G, Zou Y, Hu J, Ding Y: Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC Genomics* 2018, 19(1):720.
39. Xing H, Fu X, Yang C, Tang X, Guo L, Li C, Xu C, Luo K: Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. *Sci Rep* 2018, 8(1):2817.
40. Roscoe TT, Guilleminot J, Bessoule JJ, Berger F, Devic M: Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in Arabidopsis. *Plant & cell physiology* 2015, 56(6):1215-1228.
41. Danyluk J, Carpentier E, Sarhan F: Identification and characterization of a low temperature regulated gene encoding an actin-binding protein from wheat. *FEBS letters* 1996, 389(3):324-327.
42. Kovalchuk N, Smith J, Bazanova N, Pyvovarenko T, Singh R, Shirley N, Ismagul A, Johnson A, Milligan AS, Hrmova M *et al*: Characterization of the wheat gene encoding a grain-specific lipid transfer protein TdPR61, and promoter activity in wheat, barley and rice. *Journal of Experimental Botany* 2012, 63(5):2025-2040.
43. Cha MK, Hong SK, Lee DS, Kim IH: Vibrio cholerae thiol peroxidase-glutaredoxin fusion is a 2-Cys TSA/AhpC subfamily acting as a lipid hydroperoxide reductase. *The Journal of biological chemistry* 2004, 279(12):11035-11041.
44. Nguyen VN, Moon S, Jung KH: Genome-wide expression analysis of rice ABC transporter family across spatio-temporal samples and in response to abiotic stresses. *Journal of plant physiology* 2014, 171(14):1276-1288.
45. Marrocco K, Zhou Y, Bury E, Dieterle M, Funk M, Genschik P, Krenz M, Stolpe T, Kretsch T: Functional analysis of EID1, an F-box protein involved in phytochrome A-dependent light signal transduction, vol. 45; 2006.
46. G Kazanietz M, Wang S, Milne B, E Lewin N, L Liu H, M Blumberg P: Residues in the second cysteine-rich region of protein kinase C delta relevant to phorbol ester binding as revealed by site-directed mutagenesis,

Tables

Table 1: Next generation sequencing statistical summary of sequenced and assembled results

	0 h	1 h	6 h	12 h	48 h
Total Reads	35,663,873	32,830,183	35,632,937	32,241,116	33,976,283
Mapped Reads	21,497,466	20,272,643	21,954,725	19,121,890	19,998,709
Mapped Ratio	60.26%	61.79%	61.62%	59.33%	58.87%
Nt	10,699,162,000	9,849,054,900	10,689,881,200	9,672,334,900	10,192,884,900
GC (%)	45.53	46.50	46.17	45.29	45.07
Q20 (%)	96.72	96.63	96.77	96.75	96.77
Q30 (%)	92.03	91.66	92.08	91.91	91.96
N Percentage	00.00	00.00	00.00	00.00	00.00

3: Nt, total number of clean nucleotides; The GC percentage is the proportion of guanidine and cyto eotides among total nucleotides; The Q20 and Q30 percentage is the proportion of nucleotides with a quality v and 30, respectively; The N percentage is the proportion of unknown nucleotides in clean reads.

Table 2: Third generation sequencing statistical summary of sequenced and assembled results

	Unigenes	Transcripts
Total number of sequences	15,998	16,856
Total sequences length	34,848,832	36,928,928
Maximum length	9,135	9,135
Minimum length	208	208
Average length	2,178	2,190
Percent GC	43.10%	43.12%
N40	3,698	3,733
N50	2,652	2,701
N60	2,131	2,155
N70	1,785	1,795
N80	1,503	1,509
N90	1,162	1,168

N50, represents sorting the assembled transcripts from long to short by length, accumulating the length of the transcript to 50% of the total length, corresponding to the length of the transcript, and so on.

Table 3: Statistics of unigenes annotated in public database

Annotated Database	Annotated Number		300<=length<1000		length>=1000	
	Value (%)		Value (%)		Value (%)	
COG	6982	(43.64%)	853	(12.22%)	6129	(87.78%)
GO	12480	(78.01%)	1750	(14.02%)	10728	(85.96%)
KEGG	5965	(37.29%)	806	(13.51%)	5159	(86.49%)
KOG	10020	(62.63%)	1164	(11.62%)	8856	(88.38%)
Pfam	13569	(84.82%)	1725	(12.71%)	11844	(87.29%)
Swiss-Prot.	13428	(83.94%)	1675	(12.47%)	11751	(87.51%)
Nr	15446	(96.55%)	2115	(13.69%)	13329	(86.29%)
All	15461	(96.64%)	2125	(13.74%)	13334	(86.24%)

Figures

Figure 1

Assembly result sequence length distribution map of transcripts and unigenes in Xuzi-8 sweetpotato cultivar. The horizontal axis represents the length intervals of the transcripts and unigenes, and the vertical axis represents the number of transcripts and unigenes.

Nr Homologous Species Distribution

Figure 2

Species distribution of the top BlastX matches of the transcriptome unigenes in the non-redundant protein database (Nr) data base.

Figure 3

Gene ontology (GO) classifications in sweetpotato (Xuzi-8 cultivar), the percentage indicate the proportion of unigenes with the GO annotations.

Figure 4

Clusters of orthologous groups (COG) classification in XuZi-8 sweetpotato cultivar. Genes from the same Orthologous have the same function, so that direct functional annotations to other members of the same KOG cluster.

Figure 5

The most enriched KEGG clusters in XuZi-8 sweetpotato cultivar. The most enriched 22 clusters out of 123 clusters were presented in this figure.

Figure 6

Statistical chart of DEGs transcriptome in response to salt stress. Transcriptional level of five libraries including 1, 6, 12 and 48 hours of salt stress treatment as compared to control.

Figure 7

Comparison of four transcriptomes for classification of DEGs and statistics of sequence annotation of DEGs. (A) and (D); Venn diagram analysis of up-regulated unigenes and all induced unigenes, respectively. (B) and (E) Hierarchical cluster analysis of up-regulated unigenes and all induced unigenes, respectively. (C) and (F); Statistical chart of sequence annotation of up-regulated unigenes and all induced unigenes, respectively. Sequence showed homology with known sequences were classified as “hits” and sequence showed no homology with known sequences were classified as “no hits”.

Supplementary Files

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

- [2Additionalfile2.docx](#)
- [1Additionalfile1.xlsx](#)
- [4Additionalfile4.xlsx](#)
- [3AdditionalFile3.docx](#)