Plant Biotechnology Reports (2020) 14:713–723
https://doi.org/10.1007/s11816-020-00647-0

Online ISSN 1863-5474
Print ISSN 1863-5466

ORIGINAL ARTICLE

Comparative transcriptome analysis of Sweetpotato (*Ipomoea batatas* L.) and discovery of genes involved in starch biosynthesis

Meng Kou[1,2] · Zai-Xing Su[1] · Yun-Gang Zhang[1] · Qiang Li[1,2]

Received: 11 October 2020 / Revised: 2 November 2020 / Accepted: 4 November 2020 / Published online: 12 November 2020

Abstract

Sweetpotato, *Ipomoea batatas* L., *Convolvulaceae*, rich in starch, is a major staple economically important root crop worldwide. Limited functional genomics resources and whole genome studies could be improved through the development of next-generation sequencing technologies. Illumina sequencing of three sweetpotato varieties, Xushu 18, Xu 781 and Xushu 25, including de novo transcriptome assembly, functional annotation, and differential expression of potential starch biosynthesis-related genes are described. A total of 2508, 1807 and 3169 Mb clean reads were generated for Xushu 18, Xu 781 and Xushu 25, and were assembled into 32,112, 35,788 and 36,955 unigenes with an average length of 787, 802 and 824 bp, respectively. BLAST against the NCBI non-redundant protein, the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Gene Ontology (GO) databases have classified these unigenes into functional categories for understanding gene functions and regulation pathways. The proportion of up-regulated DEGs was lower than that of the down-regulated DEGs among three sweetpotato cultivars. The similar proportion of the up/down-regulated gene number was presented between Xushu 25 and one of its parents (Xushu18 and Xu 781). Moreover, 14 unigenes related to starch biosynthesis in Xushu 18 and Xu 781were searched out from the transcriptome sequencing data. Furthermore, gene expression of these genes in storage root during different developing stages was also investigated, 6 types of gene expression patterns were summarized. The novel sequence dataset developed in this study will be helpful for the further genetic characterization and studies of starch biosynthesis in sweetpotato.

Keywords Sweetpotato · Transcriptome · RNA-seq · Starch biosynthesis · Gene expression

Introduction

Sweetpotato [*Ipomoea batatas* (L.) Lam.] is the world's seventh most important food crop after maize, rice, wheat, potatoes, cassava and barley (FAO 2018). It is adaptive to drought conditions and various climates, and can produce high yielding in marginal lands and under different farming systems (Antonio et al. 2011). Being the largest producer, China plants approximate four million hectares annually with fresh output of one hundred million tons, accounting for 80% of total worldwide supply (Ma et al. 2010). Therefore, sweetpotato plays an important role in food security and in industrial development in China. One of the most distinctive characters of sweetpotato is its root, which is a specialized storage organ that can accumulate large amounts of starch, constituting 15–20% of the fresh weight or 50–80% of the dry weight of the sweetpotato storage root (Antonio et al. 2011; Bovell-Benjamin 2007). A range of starch-related genes in many

Meng Kou and Zai-Xing Su contribute equally to this manuscript.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11816-020-00647-0) contains supplementary material, which is available to authorized users.

✉ Qiang Li
instrong@163.com

[1] Xuzhou Institute of Agricultural Sciences in Jiangsu Xuhuai District, Sweetpotato Research Institute, Chinese Academy of Agricultural Sciences/Key Laboratory of Biology and Genetic Breeding of Sweetpotato, Ministry of Agriculture and Rural Affairs, Xuzhou 221131, China

[2] School of Life Science, Jiangsu Normal University, Xuzhou 221116, China

higher plants such as wheat and rice has been mainly studied, and the pathway of starch biosynthesis is fairly well understood. However, little about starch biosynthesis pathway was known in sweetpotato, especially those genes associated with precursor substances metabolism of starch.

The next-generation sequencing (NGS) technology is a newly developed way for genome analysis, which can deliver a very large amount of inexpensive and accurate genome sequence information. Transcriptome sequencing (RNA-Seq) is a highly effective method for gene discovery based on NGS technology (Wang et al. 2009), which is available not only to model species with reference genome sequences but also especially to non-model organisms without corresponding genome sequences as a reference (Grabherr et al. 2011; Gongora-Castillo and Buell 2013). Meanwhile, transcriptome analyses have been used for identifying differentially expressed genes at different developmental stages and in different tissues. In previous studies, researchers mainly concentrated on development of EST-derived microsatellite markers using the Roche-454 pyrosequencing technology (Schafleitner et al. 2010) and Illumina paired-end sequencing technology (Wang et al. 2010). Tao et al. specifically expressed transcripts in different tissues and at all sorts of developmental stages of sweetpotato were identified using Illumina NGS (Tao et al. 2012). Later, the sweet potato polyadenylated RNA viral sequences and expression levels were detected in three sweet potato cultivars according to the Digital Gene Expression (DGE) profiling results (Gu et al. 2014). The purple-flesh sweetpotato attracts attention because of its being rich in anthocyanins, which possess physiological functions and biological activity. Xie et al. comprehensively analyzed the transcriptome sequencing on storage roots of the purple sweetpotato, and found that UDP-glucose-flavonoid 3-*O*-glucosyltransferase (*UFGT*) was one of the key structural gene for anthocyanin biosynthesis (Xie et al. 2012). Most recently, a mutant sweet potato (white skin and flesh) transcriptome from the purple-fleshed sweetpotato cultivar (Ningzishu 1), and its wild-type, were characterized by Illumina paired-end sequencing technology (Ma et al. 2016).

In this study, we used Illumina sequencing technology to carry out comparative transcriptome analysis for three sweetpotato varieties. We also analyzed DEGs related to starch biosynthesis among the cultivars and discovered new genes of the pathway. The present transcriptome analysis provides useful information for further functional genomic research on this root crop and a reference for studying starch metabolic processes.

Materials and methods

Plant materials

The seedlings of three virus-free sweetpotato varieties, Xushu 18 (medium and high yield, medium starch content, better adaptability, high resistance to root rot, now is still one of main cultivars in China, Sequence Number: Batata_137-T1), Xu 781 (high starch content, high resistance to stem nematode, precious parental material, selected from the seedlings of International Potato Center, Sequence Number: Batata_137-T2), and Xushu 25 (high starch content, high resistance to root rot and stem nematode, is a hybrid of Xushu 18 (♀) and Xu 781 (♂), Sequence Number: Batata_137-T3) were planted in the field of Xuzhou institute of Agricultural Sciences, XuHuai region, Jiangsu Province. Tissue samples of young leaves, leaf petioles and stems were collected and mixed together at 40 days after transplanting for transcriptome sequencing and qRT-PCR validation. The storage roots of these samples were picked 40, 70, 100 and 130 days after transplanting (DAT), which means initial period of expansion (IPE), medium period of expansion (MPE), peak period of expansion (PPE) and harvest period (HP), then frozen immediately in liquid nitrogen, then stored at − 80 ℃ for further use. Each sample had three replicates.

cDNA library construction and transcriptome sequencing

The samples of three sweetpotato varieties were delivered to Biomarker (BMK) Technologies CO., Ltd (Beijing, China) for the construction of transcript library by Illumina Gene Expression Sample Prep Kit (Illumina Inc., San Diego, CA) following the manufacturer's protocols. The high-throughput sequencing was conducted using Illumina Cluster Station and Illumina HiSeq™ 2000 System (Illumina Inc., San Diego, CA).

Transcriptome assembly and gene function annotation

To get high-quality transcriptome sequence data, all reads with adaptor contamination and more than 5% unknown nucleotides were discarded. Then de novo assembly of the resting clean reads was performed by Trinity (Grabherr et al. 2011). First, we combined reads with selected overlap lengths to form longer fragments (Contigs). Second, the contigs were assembled to scaffold construction (Transcripts). Third, paired-end reads were mapped for filling the gaps in the scaffolds so as to obtain sequences with fewest Ns. These sequences were defined as Unigenes. Subsequently, multiple transcripts by alternative splicing in assembly results of the

three sweetpotato samples were clustered to a gene, finally a unigene library of the three samples was constructed. Assembled unigenes were carried out to align with the non-redundant protein (Nr) database, the SwissProt database, the GO database, the Cluster of Orthologous Groups (COG) database and KEGG pathway database by running BLASTX at an E-value threshold of 1.0E-5. Compared sequencing reads with unigene sequences in the library, unigene expression abundance information in the corresponding sample would be obtained by the RPKM (reads per kb per million reads) method (Mortazavi et al. 2008). Differentially expressed genes (DEGs) were defined using false discovery rate (FDR) < 0.01 and absolute log2-fold changes ≥ 1. HemI software was used to produce heat maps of gene expression (Deng et al. 2014). OriginPro 2018C (OriginLab Co., Ltd Northampton, USA) was used to make Venn diagrams.

RNA isolation and qRT-PCR validation

Total RNAs were extracted using the Quick RNA Isolation Kit (Huayueyang Biotechnology Co., Ltd, Beijing, China), and treated with DNase I (Fermentas, USA) according to the manufacturer's instructions. RNA quality and quantity were determined by a Nanodrop spectrophotometer 1000 (Thermo Fisher Scientific Inc., Waltham, MA, USA). RNA integrity was further measured by 1.2% agrose gel electrophoresis. First-strand cDNA was synthesized from 2 μg of total RNA by a reverse transcriptase M-MLV kit (Promega, China) following the manufacturer's protocol. cDNA was diluted 10-fold for qRT-PCR analysis.

Primers for qRT-PCR were designed using Primer Premier 5.0 (Supplementary Table S1). The qRT-PCR was performed with the ABI StepOnePlus Real-Time PCR Detection System (Applied Biosystems, USA). 20 μl of total reaction volume contained 10 μl SYBR Green Real time PCR Master Mix (TOYOBO, Osaka, Japan), 1 μl diluted cDNA, 0.8 μl of each primer (10 μM), and 7.4 μl of dd H_2O. The qRT-PCR programs were as follows: 95 ℃ for 1 min, followed by 40 cycles of 95 ℃ for 15 s, 60 ℃ for 15 s, and 72 ℃ for 20 s. Each sample was run in three biological replicates. The *IbActin* gene was used as an internal standard (Park et al. 2012). The comparative CT method ($2^{-\Delta\Delta}$CT method) was used to analyze the relative expression levels of genes during different developing stages (Schmittgen and Livak 2008).

Results

Illumina sequencing and de novo sequence assembly

The contigs were used to assemble transcripts (the sequence whose length was less than 200 bp was removed) through Trinity software. Then the transcripts of three sweetpotato varieties were clustered to unigene data, separately (Supplementary Table S2). For Xushu 18, 58.26% of the screened reads (55,120) were incorporated into assembled sequences (unigene), the N50 length was 1209 bp. For Xu 781, 53.75% of the screened reads (66,582) were incorporated into assembled sequences (unigene), the N50 length was 1282 bp. Similarly, in case of Xushu 25, 47.98% of the screened reads (77,029) were incorporated into assembled sequences (unigene), the N50 length was 1346 bp. For total unigene of the three sweetpotato varieties, the length of 16,884 (30.33%) unigenes was between 200 and 300 bp, the length of 15,242 (27.38%) unigenes was between 300 and 500 bp, the length of 10,658 (19.15%) unigenes was between 500 and 1000 bp, the length of 8694 (15.62%) unigenes was between 1000 and 2000 bp, the length of 4189 (7.53%) unigenes was more than 2000 bp. In a word, a total of 55,667 unigenes with an average length of 748 bp were generated by de novo software Trinity. Average lengths of the unigenes were larger than that obtained previously from "Guangshu 87" (Wang et al. 2010) and "Jingshu 6" (Xie et al. 2012), but were a little smaller than "Ningzishu 1 and its mutant" (Ma et al. 2016), probably owing to using different assembly softwares. Here, de novo transcriptome analysis of three sweetpotato varieties Xushu 18, Xu 781 and Xushu 25 using Illumina Genome Analyzer IIx System. A lots of high-quality clean reads with the total length of 2508, 1807 and 3169 Mb were generated for Xushu 18, Xu 781 and Xushu 25, respectively. The GC content of the three samples data was relatively consistent and an average GC content approximated to 49%. The average Phred score value of more than 99% cycle ≥ 20. These findings indicated that better sequencing and assembly results were useful for gene subsequent function analysis.

Transcriptome annotation and differentially expressed genes (DEGs) analysis

To annotate the transcriptome of sweetpotato, 55,667 assembled unigenes were queried against various protein databases in NCBI using BLASTX, a total of more than 38,000 unigenes had been functionally annotated. As shown in Fig. 1 and Supplementary Table S3, we found out 35,883 sequences (64.46%) similar to proteins in the Nr database, 47.16% of the sequences between 300–1000 bp in length were successfully annotated, whereas 35.14% of the longer transcripts (> 1000 bp) were hit. A search for homology against the UniProtKB-Swissprot database produced 26,600 hits (47.78%) and 28,329 hits (50.89%) annotated in Nt (NCBI non-redundant nucleotide sequences). The GO protein database comparison displayed a slightly lower number of hits (25,994, 46.70%) than the Nr search, 11,482 (44.17%) sequences between 300 and 1000 bp in length were aligned, nevertheless 10,410 (40.05%%) unigenes greater

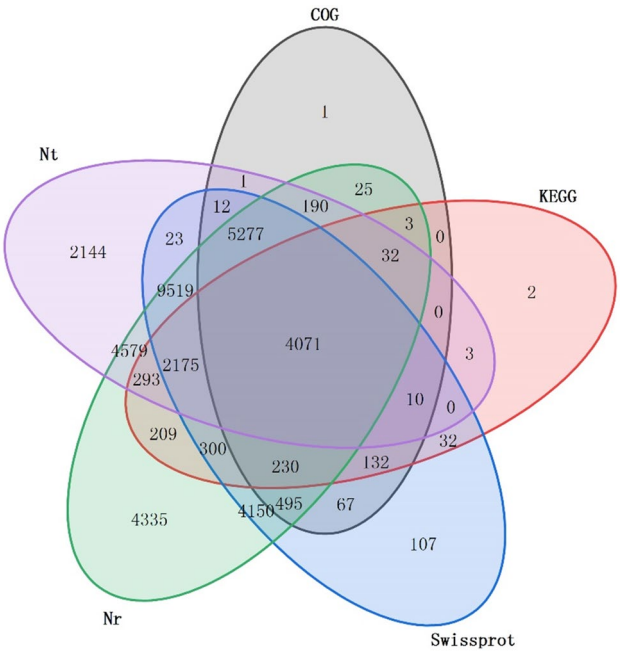

Fig. 1 Venn diagrams showing transcriptome annotation in different database

than 1000 bp were annotated. According to the BLASTX results, 10,546 (18.94%) and 7492 (13.46%) of 55,667 unigenes could be hit in databases of COG and KEGG.

The GO database is suitable for different species, which can define and describe gene and protein. In terms of the BLASTX results against Nr protein database, unigenes could be associated with GO classification from Nr database. The sweetpotato unigenes were categorized into three main GO categories: biological process, cellular component and molecular function. These GO terms were further distributed into 19 (Fig. 2a and c) and 18 (Fig. 2b) main sub-categories respectively. Among these categories, the "cell", "binding activity", "enzyme activity", "structural molecule activity", "transporter activity", "cellular process", "development" and "physiological processes" terms were discovered to have correlation with relatively more number of unigenes than other GO terms. COG database is built on genome encoding protein and phylogenetic relationships of bacteria, algae, and eukaryotes, which can be used for orthologous classification of gene products. Among the sweetpotato unigene sets, a total of 10,546 hits were categorized (E-value ≤ 1.0E-5) in 25 functional COG clusters (Fig. 3). Thus, only a small proportion of the unigenes (18.94%) carried protein domains with annotation for COG categories. The top categories based on number of orthologies were "translation, ribosomal structure, and biogenesis", "general function prediction only", "posttranslational modification, protein turnover, and chaperones", "carbohydrate transport and metabolism", "inorganic ion transport and metabolism", "amino acid transport and metabolism", "chromatin structure and dynamics", "energy production and conversion", and "secondary metabolites biosynthesis, transport and catabolism". The most interesting thing was that whether in COG database of Xushu 18 vs Xu 781, or Xushu 18 vs Xushu 25, or Xu 781 vs Xushu 25, "translation, ribosomal structure, and biogenesis" was always the largest category. It was implied that the sweetpotato tissues used in the study processed extensive translation activities. The two categories comprising "nuclear structure" and "defense mechanisms" respectively, representing the two small COG. KEGG is a major public database of pathway (Kanehisa et al. 2007). Unigenes of sweetpotato were mapped to KEGG pathways using the translated amino acids to run BLASTX against the KEGG database. A total of 425 unigenes of Xushu 18 vs Xu 781 were mapped to 112 pathways, and ribosome had the largest number of unigenes (101 members, 23.8%, ko03011), followed by chaperones and folding catalysts (17 members, 4.0%, ko03110), protein processing in endoplasmic reticulum (14 members, 3.3%, ko04141), photosynthesis proteins (12 members, 2.8%, ko00194). 154 unigenes of Xushu 18 vs Xushu 25 were mapped to 62 pathways, including chromosome (14 members, 9.1%, ko03036), ribosome (13 members, 8.4%, ko03011), photosynthesis proteins (11 members, 7.1%, ko00194), systemic lupus erythematosus (11 members, 7.1%, ko05322), and so on. 212 unigenes of Xu 781 vs Xushu 25 were mapped to 76 pathways, and ribosome had the largest number of unigenes (49 members, 23.1%, ko03011). As we all known, sweetpotato is one of the most important root food crops, which can supply people with carbohydrates and feed on billions of humans in many tropical and sub-tropical regions of the world, especially in developing countries. Therefore, we mainly focused on these pathways concerning starch biosynthesis. For Xushu 18 vs Xu 781, 5 unigenes, 3 unigenes and 1 unigene were involved in those promising pathways including fructose and mannose metabolism (81 genes, ko00051), starch and sucrose metabolism (184 genes, ko00500), and carbohydrate digestion and absorption (22 genes, ko04973) separately. For Xushu 18 vs Xushu 25, only 2 unigenes were discovered to be concerned with fructose and mannose metabolism. For Xu 781 vs Xushu 25, we found that 2 unigenes and 1 unigene were engaged in fructose and mannose metabolism and starch and sucrose metabolism respectively.

To elucidate the different phenotypes of three genotypes sweetpotato, differentially expressed genes analysis was used to determine DEGs. As shown in Fig. 4, the proportion of up-regulated DEGs was lower than that of the down-regulated DEGs among three sweetpotato cultivars. In the Xushu18 vs Xu781 comparison, we observed a total of 110 up-regulated DEGs and 426 down-regulated DEGs; in the Xushu18 vs Xushu25 comparison, we selected a total of 78 up-regulated DEGs and 152 down-regulated DEGs;

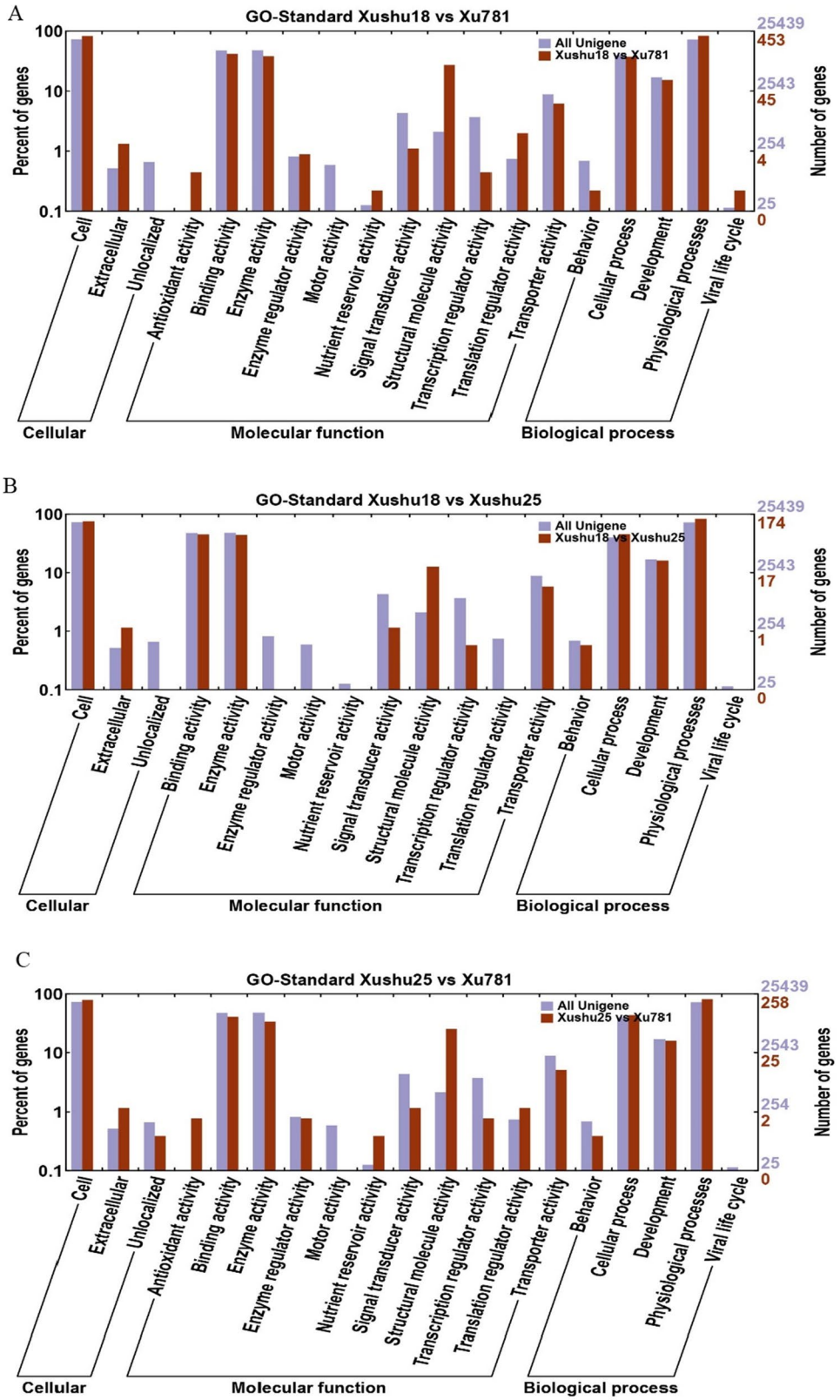

Fig. 2 Histogram presentation of GO classification of the sweetpotato transcriptome. Based on significant hits against the Nr database, GO terms associated with sweetpotato unigenes were summarized into 3 main GO categories (biological process, cellular component, and molecular function) and 19 (**a** and **c**) and 18 (**b**) sub-categories respectively. **a**: Xushu 18 vs Xu 781; **b**: Xushu 18 vs Xushu 25; **c**: Xushu 25 vs Xu 781

in the Xushu25 vs Xu781 comparison, we found 112 up-regulated DEGs and 223 down-regulated DEGs. Interestingly, the similar proportion of the up/down-regulated gene number was presented between Xushu 25 vs one of its parents (Xushu18 and Xu 781), in which up-regulated genes accounted for about one-third of all DEGs and up-regulated

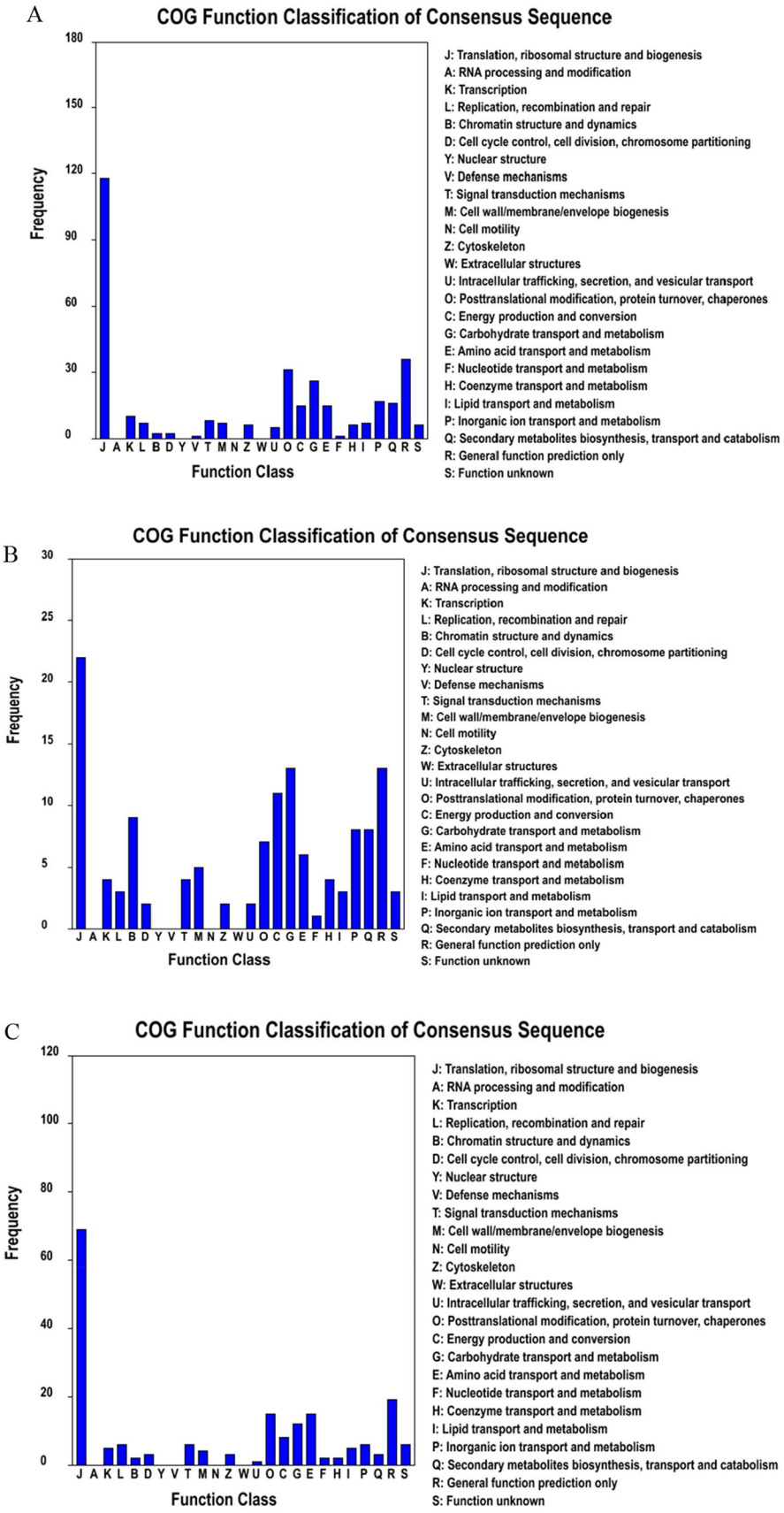

Fig. 3 Histogram presentation of COG functional classification for the sweetpotato transcriptome. Of 35,883 unigenes in the Nr database, 10,546 unigenes show significant homologies to the COGs database, which were classified into 25 COG categories. A: Xushu 18 vs Xu 781; B: Xushu 18 vs Xushu 25; C: Xushu 25 vs Xu 781

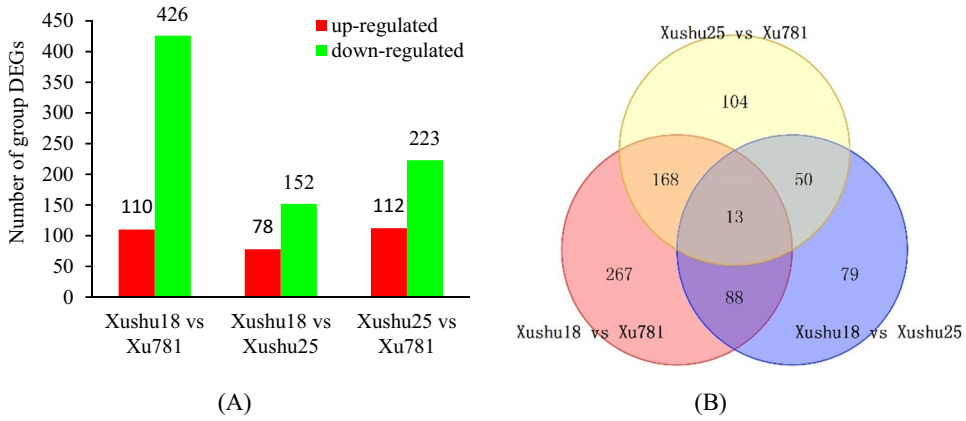

Fig. 4 Overview of differentially expressed genes (DEGs) between three sweetpotato cultivars. **a** Histogram of group DEGs; **b** Venn diagrams for analysis of total DEGs

genes occupied two-thirds (Fig. 4a). Furthermore, 769 DEGs were found and 13 DEGs overlapped among the three comparisons were identified by Venn diagram analysis (Fig. 4b).

Candidate genes involved in starch biosynthesis between Xushu 18 and Xu 781 were selected from RNA-Seq results

Because of being rich in starch, sweetpotato is widely applied to industrial and agricultural production as raw material. To date, many genes related to storage starch biosynthesis in storage roots of sweetpotato had been isolated (Wang et al. 1999; Lee et al. 2000; Kimura et al. 2000; Kim et al. 2005; Hamada et al. 2006). However, those genes associated with starch biosynthesis in the process of photosynthesis were rarely focused on. According to the annotation results, 14 unigenes related to starch biosynthetic process were obtained from 536 DEGs using "starch" as search word (Supplementary Table S1). There into, 6 of them annotated by KEGG, 13 of them annotated by Nt, and all of them annotated by SwissProt, TrEMBL, and Nr. The homologous sequences of these genes in NCBI were captured by BLAST, and named comparing with that of genes in the Tair (https://www.arabidopsis.org/) databases. The names of 11 unigenes could be confirmed as follows: beta-amylase 3 (*BAM3*), glycine dehydrogenase P-protein 1 (*GLDP1*), glucose-1-phosphate adenylyltransferase large subunit 1 (*AGPb1A*), fructose-bisphosphate aldolase 2 (*FBA2*), fructose-bisphosphate aldolase 5 (*FBA5*), microtubule-associated protein 1 (*MAP65-1*), NDR1/HIN1-LIKE 1 (*NHL1*), phosphoethanolamine N-methyltransferase 2 (*PEAM2*), phosphoethanolamine N-methyltransferase 3 (*PEAM3*), sedoheptulose-1,7-bisphosphatase (*SBPase*), and Alpha-xylosidase 1 (*XYL1*). The other three unigenes were unable to ascertain their names owing to too little comparative information. Therefore, their names were represented by codes as follows: *T1-23698*, *T1-30106*, and *T2-33205*.

Differential gene expression analysis in starch biosynthetic process between Xushu 18 and Xu 781 by qRT-PCR verification

Total 18 DEGs involved in starch metabolic pathway have been ascertained from the sequencing results (Fig. 5a, b). For the 40 DAT, among the 14 starch-related DEGs selected in our experiment, the qRT-PCR expression patterns of BAM3, AGPb1A, FBA5 and other 6 genes were consistent to that of RNA-seq, showing a tendency of up-regulation or down-regulation, although the regulatory multiples are different. However, the qRT-PCR expression patterns presented the opposite results with RNA-seq in GLDP1, FBA2 and other 3 genes (Fig. 5c).

To explore the expression pattern of those candidate genes involved in starch biosynthesis during different growth stages, 14 genes were selected and qRT-PCR were carried out in storage root of Xushu 18 and Xu 781. According to the changes in the relative expressions of 14 candidate genes, the expression patterns of these genes can be summarized into 6 types as followings: (I) always increase; (I) always decrease; (III) firstly increase and then decrease; (IV) firstly decrease and then increase; (V) firstly increase then decrease and increase; (VI) firstly decrease then increase and decrease (Fig. 6). The expression changes of *BAM3* in Xu 781, *GLDP1* in Xushu 18 and Xu 781, *IbAGPb1A* in Xushu 18, *SBPase* in Xushu 18, *T1-30106* in Xushu 18 and Xu 781, and *T2-33205* in Xushu 18 belonged to type I. The expression trends of *BAM3* in Xushu 18, *IbAGPb1A* in Xu 781, *IbFBA5* in Xushu 781, *MAP65-1* in Xushu 18 and Xu 781, *NMT2* in Xushu 18 and Xu 781, *NMT3* in Xushu 18, *T1-23698* in Xu 781 and *T2-33205* in Xu 781 were assigned to type III. While changes in gene expression levels of *IbFBA2* in Xushu 18 and Xu 781, *IbFBA5* in Xushu 18, *SBPase* in Xu 781, and *T1-23698* in Xushu 18 were geared to type V. Furthermore, the expression patterns of *XYL1* in Xushu 18 and Xu 781 belonged to type II, and that of *NHL1* in Xushu 18 and Xu 781 was assigned to type IV, and then that of

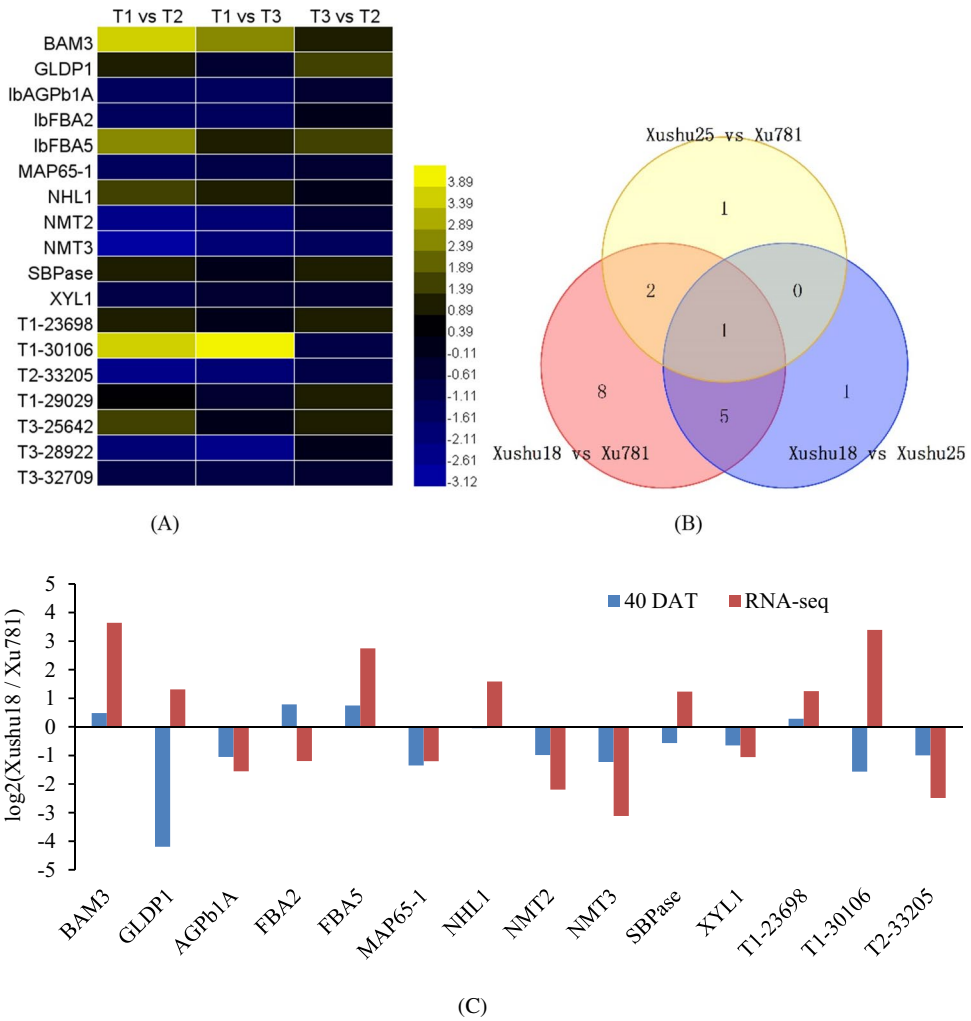

Fig. 5 The differentially expressed genes related to starch metabolism. **a** Yellow and blue indicate up- and down-regulated transcripts, respectively (log2-fold change), "T1" represents Xushu 18, "T2" represents Xu 781, and "T3" represents Xushu 25; **b** Venn diagrams for analysis of DEGs related to starch metabolic; **c** Validation of sequencing data by qRT-PCR

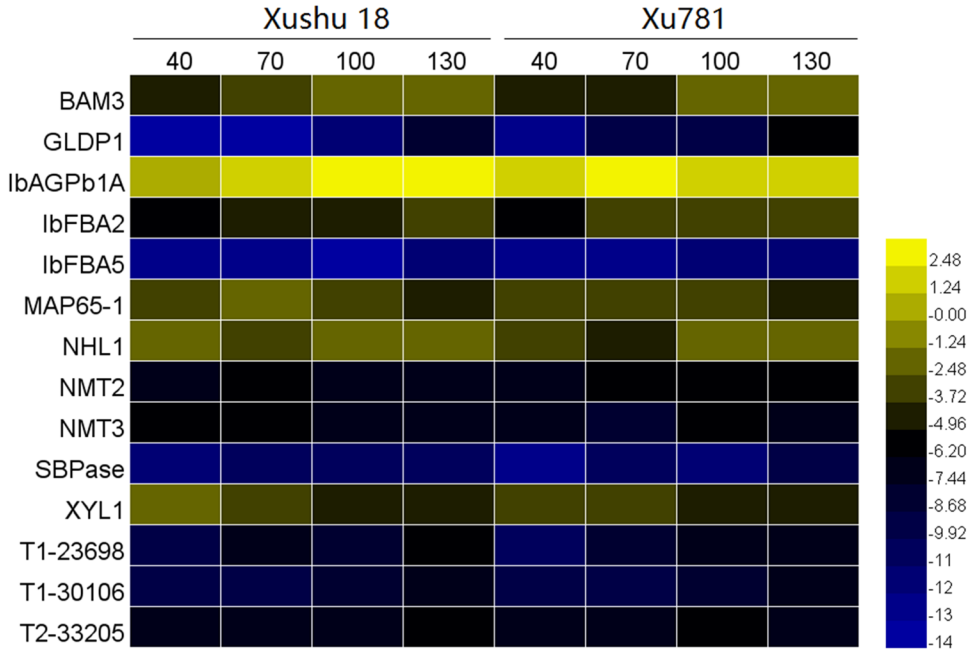

Fig. 6 The expression patterns of 14 candidate genes related to starch biosynthetic. The actual expression levels of each gene were compared by Actin standardization. Yellow and blue indicate up- and down-regulated transcripts, respectively, from the three comparisons (log2-fold change). "T1" represents Xushu 18, "T2" represents Xu 781.

NMT3 in Xu 781 fitted into type VI. Our expression results were consistent with those studies reported by Wang et al. (Wang et al. 2015). In the previous report, the expression patterns of 22 selected genes were validated by RT-PCR based on the microarray data. These 22 genes presented various expression profilings from 10 to 120 days after planting, such as early increase and late decrease, early decrease and late increase, and other expression patterns (Wang et al. 2015). Furthermore, we also found that the expression levels of 14 tested genes varied greatly between Xushu 18 and Xu 781 (Fig. 6), which showed consistence with the conclusions reached between TNG31 (low-amylose content) and TNN18 (high-amylose content) (Lai et al. 2016). Previous qRT-PCR results demonstrated that the expression levels of *AGPase*, *GBSS*, *SS* (starch synthase) and *SBE* related to starch synthesis in TNN18 storage roots were significantly higher than those in TNG31 storage roots, while the RNA levels of *SBEII*, *DBE* and *ISA* (isoamylase) in TNN18 were lower.

In the study, based on the transcriptome sequencing data (Fig. 5c, Supplementary Figure S1), the expression level of 7 genes in Xushu 18 was up-regulated comparing to that in Xu 781, while the expression level of other 7 genes was down-regulated. The expression of *MAP65-1*, *GLDP1* and *FBA5* in different expanding stages of sweetpotato storage root was contrary to the transcriptome sequencing results, and only the expression of *NHL1* perfectly matched with the RNA-seq data. However, the expression of the rest of the genes was partly identical to the sequencing data.

Discussion

Sweetpotato is not only a staple food and feedstuff, but also a promising source of bio-energy for fuel production (Santa-Maria et al. 2009). Thus, it is urgent for crop breeders to improve the starch content and quality of sweetpotato by gene engineering, because conventional hybridization is limited owing to its high heterozygosity and self-incompatible nature (Dhir et al. 1998). In recent decades, more and more researches on the sweetpotato were also focused on increasing the dry matter by regulating starch biosynthesis-related genes in the storage root. Suppression of granule-bound starch synthase I (GBSSI) (Otani et al. 2007) and the starch branching enzyme (SBE) (Shimada et al. 2006) by RNA interference (RNAi) resulted in amylose-free starch content and high-amylose content of sweetpotato separately (Kitahara et al. 2007). However, whether these studies can enhance the starch content of storage roots by genetic engineering remains unknown. Tanaka et al. found that the transgenic sweetpotato plants overexpressing the *SRF1* gene (encodes a Dof zinc finger transcription factor) increased the dry matter and starch content in the storage roots (Tanaka et al. 2009). Transgenic tobacco plants overexpressing the sweetpotato *IbSnRK1* gene significantly increased soluble sugar and starch levels comparing with that of the wild-type (Jiang et al. 2013). Bae et al. overexpressed the sweetpotato *IbEXP1* in *Arabidopsis* and accumulated more starch in each seed (Bae et al. 2014). The *IbAATP* (a plastidic ATP/ADP transporter gene)-overexpressing sweetpotato plants significantly increased the starch and amylose content and altered fine structure of amylopectin (Wang et al. 2016).

With the development of NGS technology, de novo assembly has been conducted for transcriptome annotation, marker identification and gene discovery of model species and non-model species (Grabherr et al. 2011). Owning to a large amount of chromosomes, complex genetic background and uncompleted genome sequencing of sweetpotato, many functional genes had been acquired by transcriptome sequencing method (Schafleitner et al. 2010; Wang et al. 2010; Tao et al. 2012; Xie et al. 2012; Ma et al. 2016). Compared with Roche/454 sequencing technology, the Illumina/Solexa platform is more accurate and gains deeper coverage (Su et al. 2011). Using previous research results uploaded to bioinformatics database, researchers can effectively annotate sequencing data, and preliminarily understand the functions and subcellular location of the genes. In this study, two starch biosynthesis-related genes in sweetpotato storage root, *BAM3* (Yoshida et al. 1992) and *AGPb1A* (Lee et al. 2000), had been cloned a few years earlier. Four genes (*GLDP1*, *FBA2*, *FBA5*, and *SBPase*), coding those important enzymes involved in photosynthetic carbon fixation, were sought out. The glucose produced via photosynthesis is the precursor substance of starch biosynthesis. An increase in sucrose and starch accumulation was found in *SBPase* sense transgenic tomato plants (Ding et al. 2016). Photosynthetic and growth rates of *GDCP* antisense potato lines were reduced compared to wild-type potato (Winzer et al. 2001). The interesting thing is that we also explored out five genes (*MAP65-1*, *NHL1*, *NMT2*, *NMT3* and *XYL1*), which played a crucial role in plant growth and environmental stress response in the previous study (Keogh et al. 2009; Günl and Pauly 2011; Krtková et al. 2016). In addition, three unigenes coding uncharacterized protein were searched out, which implied that these genes maybe participate in starch metabolism.

Acknowledgements This work was supported by the National Key R&D Program of China (2019YFD1001300, 2019YFD1001302), the earmarked fund for China Agriculture Research System(CARS-10,Sweetpotato), Xuzhou academy of agricultural sciences research fund project (XM2019001), and the Key R&D Plan of Xuzhou City (KC18127).

Author contributions QL designed the study, ZXS conducted the experiments, MK and ZXS analyzed the results and performed the figures. MK drafted the manuscript. QL and YGZ revised the manuscript. All authors read and approved the manuscript.

References

Antonio GC, Takeiti CY, de Oliveira RA, Park KJ (2011) Sweetpotato: production, morphological and physicochemical characteristics, and technological process. Fruit Veg Cereal Sci 5:1–18

Bae JM, Kwak MS, Noh SA, Oh MJ, Kim YS, Shin JS (2014) Overexpression of sweetpotato expansin cDNA (IbEXP1) increases seed yield in Arabidopsis. Transgenic Res 23(4):657–667

Bovell-Benjamin AC (2007) Sweet potato: a review of its past, present, and future role in human nutrition. Adv Food Nutr Res 52:1–59

Deng WK, Wang YB, Liu ZX, Cheng H, Xue Y (2014) HemI: a toolkit for illustrating heatmaps. PLoS ONE 9(11):e111988

Ding F, Wang ML, Zhang SX, Ai XZ (2016) Changes in SBPase activity influence photosynthetic capacity, growth, and tolerance to chilling stress in transgenic tomato plants. Sci Rep 6:32741

Dhir SK, Oglesby J, Bhagsari AS (1998) Plant regeneration via somatic embryogenesis, and transient gene expression in sweet potato protoplasts. Plant Cell Rep 17:665–669

Food and Agriculture Organization of the United Nations (2018). FAO Statistical Databases, FAOSTAT. https://www.fao.org/faostat/zh/#data/QC

Gongora-Castillo E, Buell CR (2013) Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. Nat Prod Rep 30(4):490–500

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnol 29(7):644–652

Gu YH, Tao X, Lai XJ, Wang HY, Zhang YZ (2014) Exploring the polyadenylated RNA virome of sweet potato through high-throughput sequencing. PLoS ONE 9(6):e98884

Günl M, Pauly M (2011) *AXY3* encodes a α-xylosidase that impacts the structure and accessibility of the hemicellulose xyloglucan in Arabidopsis plant cell walls. Planta 233(4):707–719

Hamada T, Kim SH, Shimada T (2006) Starch-branching enzyme I gene (*IbSBEI*) from sweet potato (*Ipomoea batatas*); molecular cloning and expression analysis. Biotechnol Lett 28(16):1255–1261

Jiang T, Zhai H, Wang FB, Yang N, Wang B, He SZ, Liu QC (2013) Cloning and characterization of a carbohydrate metabolism- associated gene IbSnRK1 from sweetpotato. Sci Hortic 158(4):22–32

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2007) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480–D484

Keogh MR, Courtney PD, Kinney AJ, Dewey RE (2009) Functional characterization of phospholipid *N*-Methyltransferases from *Arabidopsis* and soybean. J Biol Chem 284(23):15439–15447

Kim SH, Hamada T, Otani M, Shimada T (2005) Cloning and Characterization of sweetpotato isoamylase gene (*IbIsa1*) Isolated from tuberous root. Breed Sci 55(4):453–458

Kimura T, Ideta O, Saito A (2000) Identification of the gene encoding granule-bound starch synthase I in sweet potato (Ipomoea batatas (L.) Lam.). Plant Biotechnol 17(3):247–252

Kitahara K, Hamasuna K, Nozuma K, Otani M, Hamada T, Shimada T, Fujita K, Suganuma T (2007) Physicochemical properties of amylose-free and high-amylose starches from transgenic sweetpotatoes modified by RNA interference. Carbohyd Polym 69(2):233–240

Krtková J, Benáková M, Schwarzerová K (2016) Multifunctional microtubule-associated proteins in plants. Front Plant Sci 7:474

Lai YC, Wang SY, Gao HY, Nguyen KM, Nguyen CH, Shih MC, Lin KH (2016) Physicochemical properties of starches and expression and activity of starch biosynthesis-related genes in sweet potatoes. Food Chem 199:556–564

Lee SS, Bae JM, Oh MS, Liu JR, Harn CH (2000) Isolation and characterization of polymorphic cDNAs partially encoding ADP-glucose pyrophosphorylase (AGPase) large subunit from sweet potato. Mol Cells 10(1):108–112

Ma PY, Bian XF, Jia ZD, Guo XD, Xie YZ (2016) De novo sequencing and comprehensive analysis of the mutant transcriptome from purple sweet potato (Ipomoea batatas L.). Gene 575(2):641–649

Ma D, Li H, Tang J, Xie Y, Li Q, Cao Q, Zhang Y, Zhang A (2010) Current status and future prospects of development of sweet potato industry in China. Sweet potato in food and energy security: In: Proceedings of China Xuzhou 4th international sweet potato symposium and 4th China–Japan–Korea sweet potato workshop. Xuzhou, China, pp 3–10

Mortazavi A, Wold B, Williams BA, Schaeffer L, Mccue K (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628

Otani M, Hamada T, Katayama K, Kitahara K, Kim SH, Takahata Y, Suganuma T, Shimada T (2007) Inhibition of the gene expression for granule-bound starch synthase I by RNA interference in sweet potato plants. Plant Cell Rep 26:1801–1807

Park SC, Kim YH, Ji CY, Park S, Jeong JC (2012) Stable internal reference genes for the normalization of real-time PCR in different sweetpotato cultivars subjected to abiotic stress conditions. PLoS ONE 7(12):e51502

Santa-Maria M, Pecota KV, Yencho CG, Allen G, Sosinski B (2009) Rapid shoot regeneration in industrial "high starch" sweetpotato (Ipomoea batatas L.) genotypes. Plant Cell Tiss Organ Cult 97(1):109–117

Schafleitner R, Tincopa LR, Palomino O, Rossel G, Robles RF, Alagon R, Rivera C, Quispe C, Rojas L, Pacheco JA, Solis J, Cerna D, Kim JY, Hou J, Simon R (2010) A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. BMC Genomics 11(1):604

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3(6):1101–1108

Shimada T, Otani M, Hamada T, Kim SH (2006) Increase of amylose content of sweet potato starch by RNA interference of the starch branching enzyme II gene (IbSBEII). Plant Biotechnol 23:85–90

Su C, Chao Y, Alex Chang Y, Chen W, Chen C, Lee A, Hwa K, Shin M (2011) De Novo assembly of expressed transcripts and global analysis of the Phalaenopsis aphrodite Transcriptome. Plant Cell Physiol 52(9):1501–1514

Tanaka M, Takahata Y, Nakayama H, Nakatani M, Tahara M (2009) Altered carbohydrate metabolism in the storage roots of sweet potato plants overexpressing the SRF1 gene, which encodes a Dof zinc finger transcription factor. Planta 230(4):737–746

Tao X, Gu YH, Wang HY, Zheng W, Li X, Zhao CW, Zhang YZ (2012) Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [Ipomoea batatas (L.) Lam]. PLoS ONE 7(4):e36234

Yoshida N, Hayashi K, Nakamura K (1992) A nuclear gene encoding *β*-amylase of sweet potato. Gene 120(2):255–259

Wang SJ, Yeh KW, Tsai CY (1999) Molecular characterization and expression of starch granule-bound starch synthase in the sink and source tissues of sweet potato. Physiol Plantarum 106(3):253–261

Wang YN, Li Y, Zhang H, Zhai H, Liu QC, He SZ (2016) A plastidic ATP/ADP transporter gene, IbAATP, increases starch and amylose contents and alters starch structure in transgenic sweetpotato. J Integr Agr 15(9):1968–1982

Wang ZY, Fang BP, Chen JY, Zhang XJ, Luo ZX, Huang LF, Chen XL, Li YJ (2010) De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development

of cSSR markers in sweetpotato (Ipomoea batatas). BMC Genomics 11(1):726

Wang Z, Fang B, Chen X, Liao M, Chen J, Zhang X, Huang L, Luo Z, Yao Z, Li Y (2015) Temporal patterns of gene expression associated with tuberous root formation and development in sweetpotato (*Ipomoea batatas*). BMC Plant Biol 15:180

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1):57–63

Winzer T, Heineke D, Bauwe H (2001) Growth and phenotype of potato plants expressing an antisense gene of P-protein of glycine decarboxylase under control of a promoter with preference for the mesophyll. Ann Appl Biol 138(1):9–15

Xie FL, Burklew CE, Yang YF, Liu M, Xiao P, Zhang BH, Qiu DY (2012) De novo sequencing and a comprehensive analysis of purple sweet potato (Impomoea batatas L.) transcriptome. Planta 236(1):101–113