

# Characterization of full-length transcriptome and mechanisms of sugar accumulation in *Annona squamosa* fruit

REN FANG<sup>#</sup>; WEIXIONG HUANG<sup>#</sup>; JINYAN YAO; XING LONG; JI ZHANG; SHUANGYUN ZHOU; BIAO DENG; WENZHONG TANG; ZHENYU AN<sup>\*</sup>

Horticultural Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, 530007, China

**Key words:** PacBio sequencing, Sugar metabolism, Gene expression, Alternative splicing, Simple sequence repeat

**Abstract:** *Annona squamosa* is a multipurpose fruit tree employed in nutritional, medicinal, and industrial fields. Its fruit is significantly enriched in sugars, making it an excellent species to study sugar accumulation in fruit. However, the scarcity of genomic resources hinders genetic studies in this species. This study aimed at generating large-scale genomic resources in *A. squamosa* and deciphering the molecular basis of its high sugar content. Herein, we sequenced and characterized the full-length transcriptome of *A. squamosa* fruit using PacBio Iso-seq. In addition, we analyzed the changes in sugar content over five fruit growth and ripening stages, and we applied RNA-sequencing technology to investigate the changes in gene expression related to sugar accumulation. A total of 783,647 circular consensus sequences were generated, from which we obtained 48,209 high-quality, full-length transcripts. Additionally, 1,838 transcription factors and 1,768 long non-coding RNAs were detected. Furthermore, we identified 10,400 alternative splicing events from 2,541 unigenes having on average 2–4 isoforms. A total of 15,061 simple sequence repeat (SSR) motifs were discovered and up to three primer pairs were designed for each SSR locus. Sugars mainly accumulate during the ripening stage in *A. squamosa*. Most of the genes involved in sugar transport and metabolism in the fruit were progressively repressed overgrowth and ripening stages. However, sucrose phosphate synthase involved in sucrose synthesis and more importantly, isoamylase, alpha-amylase, beta-amylase, 4-alpha-glucanotransferase genes involved in starch degradation displayed positive correlations with sugar accumulation in fruit. Overall, we provide here a high-quality, full-length transcriptome assembly which will facilitate gene discovery and molecular breeding of *A. squamosa*. We found that starch degradation during fruit ripening was the main channel for sugar accumulation in *A. squamosa* fruit, and the key genes positively linked to sugar accumulation could be further studied to identify targets for controlling sugar content in *A. squamosa* fruit.

## Introduction

*Annona squamosa* Linn. also known as sugar apple is a popular fruit tree widely cultivated in tropical and subtropical areas. The soft granular juicy and sugary pulp of the fruit is enriched in dietary fibers, carbohydrates, and various vitamins, minerals, and fatty acids (Brandão and Santos, 2016). The seeds contain a significant amount of oil with various potential industrial applications (Mariod *et al.*, 2010; Omkaresh *et al.*, 2017; Zahid *et al.*, 2018a; Abdulrahman *et al.*, 2019). Besides, *A. squamosa* has been shown to possess important medicinal attributes. Different tissues, including leaf, root, fruit, stem, seed, flower, and

bark of *A. squamosa* are sources of medicinally important compounds such as cyclic peptides, reomerine, annonine, dehydroreomerine, bullatacinone, squamone, annoreticuin, isoannoreticuin, dopamine, salsolinol, and coclaurine with antioxidative, anti-inflammatory, insecticidal, antitumor, anti-diabetic, analgesic, hepatoprotective, antimicrobial properties (Pandey and Barve, 2011; Perez *et al.*, 2004; Yu *et al.*, 2005; Patel and Kumar, 2008; Saleem *et al.*, 2008; Chavan *et al.*, 2010; Salman and Senthilkumar, 2015; Zahid *et al.*, 2018b).

*A. squamosa* fruit is referred to as 'one of the most delicious fruits known to man' due to the high sugar content (Levetin and McMahon, 2008). In fact, it contains up to 28% total sugars with variation among genotypes (Brandão and Santos, 2016). This value is by far higher than other sweet fruits such as mango (~15%, Saleem-Dar *et al.*, 2016) and banana (~20%, Forster *et al.*, 2003). High sugar content is a key fruit quality criterion and driver of

\*Address correspondence to: Zhenyu An, anderlu0396@163.com

<sup>#</sup>These authors contributed equally to this work

Received: 18 July 2020; Accepted: 06 September 2020



consumer choice (Vimolmangkang *et al.*, 2016); therefore, the development of cultivars with high sugar content is a priority in fruit breeding programs.

Sugar transport, metabolism, and accumulation have been well studied in plants (Nguyen-Quoc and Foyer, 2001). Sugar transporters are the main carriers of sucrose from source tissues to sink tissues such as fruit. They play a preponderant role in sugar accumulation in fruit. Besides, several genes, such as sucrose synthase and invertases, are involved in the conversion of sucrose into glucose, fructose, and other sugars. Activities of sucrose phosphate synthase or sucrose phosphate phosphatase lead to the re-synthesis of sucrose from fructose 6-phosphate and uridine diphosphate-glucose (Rolland *et al.*, 2006; Li *et al.*, 2012). In addition to sucrose transport and metabolism in fruit, starch degradation is another important mechanism for sugar accumulation. Starch is mainly accumulated in developing fruit, while during fruit ripening, starch is hydrolyzed into soluble sugars by the activity of several genes such as amylases (Cordenunsi-Lysenko *et al.*, 2019). It has been shown that starch content increased up to 35% in the pulp of developing banana while at late ripening, starch content was less than 1%. Meanwhile, soluble sugars increased from 1% in developing banana to reach up to 20% of the fresh weight at the ripening stage (Soares *et al.*, 2011).

Using the RNA-sequencing approach, the global change in gene expression levels and mechanisms of sugar metabolism have been reported in various fruit crops such as banana, melon, and watermelon (Asif *et al.*, 2014; Gao *et al.*, 2018; Schemberger *et al.*, 2020). However, the molecular mechanisms of the significantly high sugar content in *A. squamosa* fruit have not been elucidated. In non-model species without reference genome sequence, full-length transcriptome generated by third-generation sequencing platforms, such as PacBio RS and Nanopore, provides high quality basic genomic resources (Chen *et al.*, 2019; Qian *et al.*, 2019; Yue *et al.*, 2020).

The scarcity of genomic resources in *A. squamosa* limits gene discovery, functional analysis, and molecular breeding (Gupta *et al.*, 2015; Liu *et al.*, 2016). This study aimed to

generate a high-quality, full-length transcriptome in *A. squamosa* and clarify the molecular basis of its high sugar content. We successfully sequenced and characterized for the first time the full-length transcriptome of *A. squamosa* using PacBio Sequel platform. With the relatively high error rate of PacBio Sequel, we integrated data from Illumina short-read sequencing at five fruit growth and ripening stages to improve sequencing accuracy (Koren *et al.*, 2012). We further examined changes in gene expression in sugar transport and metabolism, and we also investigated mechanisms of high sugar content in *A. squamosa* fruit.

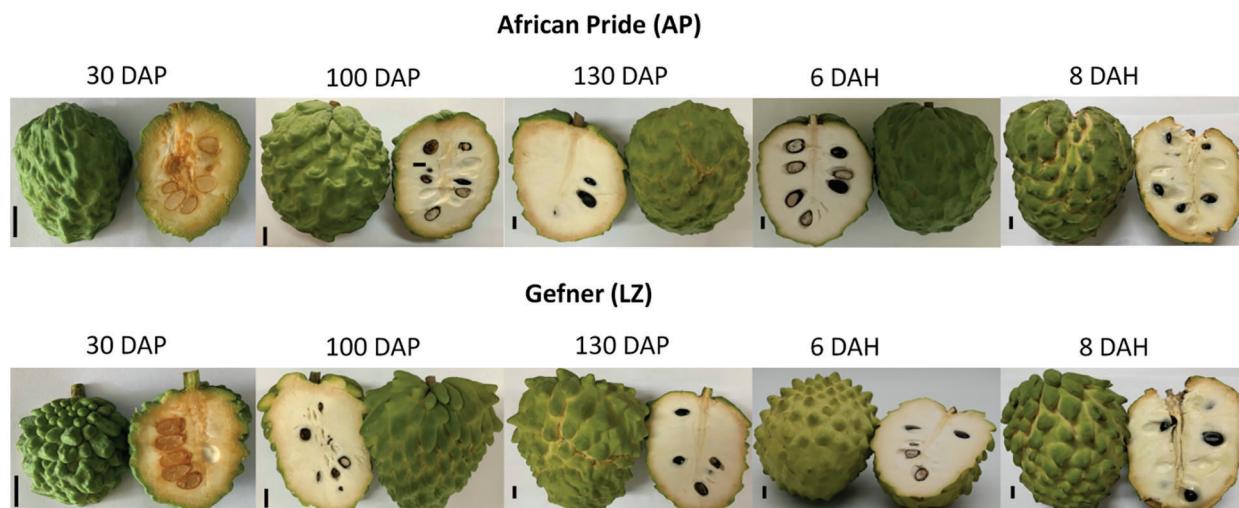
## Materials and Methods

### Plant materials

Two cultivars of *Annona squamosa* Linn. ('African Pride' and 'Gefner') were used as plant materials in this study. The cultivars 'African Pride' and 'Gefner' originated in South Africa and Israel, respectively (Paull and Duarte, 2011). These two cultivars were used in this study because they are widely grown in China and very popular among customers. The materials were obtained from the Horticultural Research Institute, Guangxi Academy of Agricultural Sciences, China, and formal identification of the materials was conducted by the corresponding author of this paper. Thirty fruit samples, for each cultivar, were collected from 5-year-old trees at five growth and ripening stages from April to July 2019 (Fig. 1). Fruit from each cultivar and stage were separated into three replicates (ten fruits/replicate) and pulp samples were collected, frozen in liquid nitrogen, and used for RNA-seq. In addition, samples from both cultivars at all stages were mixed and used for PacBio sequel sequencing.

### Analysis of fruit sugar content

The soluble sugars such as sucrose, glucose, and fructose of *A. squamosa* fruit collected at the different growth and ripening stages were analyzed using high-performance liquid chromatography (HPLC) as described by Folgado *et al.* (2017).



**FIGURE 1.** Pictures of fruits from two *A. squamosa* cultivars at five growth and ripening stages. DAP: days after pollination; DAH: days after harvest; AP: 'African Pride' cultivar; LZ: 'Gefner' cultivar. Bar = 1 cm.

Approximately 200 mg of each sample was accurately weighed in three replicates and inserted in a 2 mL EP tube. Then, 0.6 mL 2-chlorophenylalanine (4 ppm) methanol ( $-20^{\circ}\text{C}$ ) was added and vortexed for 30 s; 100 mg glass beads were added, and the samples were put into a TissueLysis II tissue grinding machine. Samples were ground at 25 Hz for 60 s. The tubes were placed in an ultrasound bath at room temperature for 15 min; Then, centrifuged at  $25^{\circ}\text{C}$  for 10 min at  $1750\times g$ , and the supernatant was filtered through a  $0.22\text{-}\mu\text{m}$  membrane to obtain the prepared samples for liquid chromatography-mass spectrometry. Chromatographic separation was performed in a Thermo Ultimate 3000 system equipped with an ACQUITY UPLC<sup>®</sup> HSS T3 ( $150 \times 2.1$  mm,  $1.8 \mu\text{m}$ , Waters) column maintained at  $40^{\circ}\text{C}$ . The temperature of the autosampler was  $8^{\circ}\text{C}$ . Gradient elution of analytes was carried out with 0.1% formic acid in water (D) and 0.1% formic acid in acetonitrile (C) or 5 mM ammonium formate in water (B) and acetonitrile (A) at a flow rate of 0.25 mL/min. The concentration of sugars in the tissues was calculated based on the molar concentration obtained from the measurements, and the initially fresh weight used.

#### Library construction and single molecule real-time sequencing

Total RNA was extracted by grinding mixed *A. squamosa* fruit tissues from the five growth and ripening stages and two cultivars in TRIzol reagent (Life Technologies) on dry ice and processed following the protocol provided by the manufacturer. The integrity of the RNA was determined with the Agilent 2100 Bioanalyzer and agarose gel electrophoresis. The purity and concentration of the RNA were determined with the Nanodrop microspectrophotometer (ThermoFisher). The mRNA was enriched by Oligo(dT) magnetic beads. Then the enriched mRNA was reverse transcribed into cDNA using Clontech SMARTer PCR cDNA Synthesis Kit. PCR cycle optimization was used to determine the optimal amplification cycle number for the downstream large-scale PCR reactions. Then, the optimized cycle number was used to generate double-stranded cDNA. In addition,  $>4$  kb size selection was performed using the Blue Pippin<sup>™</sup> Size-Selection System and mixed equally with the no-size-selection cDNA. Then, large-scale PCR was performed for the next SMRTbell library construction. cDNAs were DNA damage repaired, end repaired, and ligated to sequencing adapters. The SMRTbell template was annealed to the sequencing primer and bound to polymerase and sequenced on the PacBio Sequel platform using P6-C4 chemistry with 10 h movies by Gene Denovo Biotechnology Co. (Guangzhou, China).

#### PacBio long-read processing

The raw sequencing reads of cDNA libraries were classified and clustered into transcript consensus using the SMRT Link v5.0.1 pipeline (Gordon *et al.*, 2015) supported by Pacific Biosciences. Briefly, circular consensus sequence (CCS) reads were extracted out of the subreads BAM file. Then, CCS reads were classified into full-length non-chimeric (FL), non-full-length (nFL), chimeras, and short reads based on cDNA primers and poly-A tail signal. Short

reads ( $<50$  bp) were discarded. Subsequently, the full-length non-chimeric (FLNC) reads were clustered by Iterative Clustering for Error Correction (ICE) software to generate the cluster consensus isoforms. To improve the accuracy of PacBio reads, two strategies were employed. First, the non-full-length reads were used to polish the above-obtained cluster consensus isoforms by Quiver software (Chin *et al.*, 2013) to obtain the FL polished, high-quality consensus sequences (accuracy  $\geq 99\%$ ). Next, the low-quality isoforms were further corrected using Illumina short reads obtained from the same samples by the LoRDEC tool (version 0.8) (Salmela and Rivals, 2014). Then, the final transcriptome isoform sequences were filtered by removing the redundant sequences with software CD-HIT-v4.6.7 (Fu *et al.*, 2012) using an identity threshold of 0.99.

#### Annotation of isoforms and transcription factor analysis

Basic annotation of isoforms includes protein functional annotation, pathway annotation, COG/KOG functional annotation, and Gene Ontology (GO) annotation. To annotate the isoforms, they were BLAST analyzed against the NCBI non-redundant protein (Nr) database (<http://www.ncbi.nlm.nih.gov>), the Swiss-Prot protein database (<http://www.expasy.ch/sprot>), the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg>), and the COG/KOG database (<http://www.ncbi.nlm.nih.gov/COG>) with BLASTx program (<http://www.ncbi.nlm.nih.gov/BLAST/>) at an *E*-value threshold of  $1e-5$  to evaluate sequence similarity with genes of other species. GO annotation was analyzed by Blast2GO software (Conesa *et al.*, 2005) with Nr annotation results of isoforms. Then, the functional classification of isoforms was performed using the WEGO software (Ye *et al.*, 2018). Protein coding sequences of isoforms were aligned by hmmscan to PlantTFdb (<http://planttfdb.gao-lab.org/>; Jin *et al.*, 2017) to predict transcription factor (TF) families.

#### Simple sequence repeats prediction and primer design

The MicroSatellite (MISA, <http://pgrc.ipk-gatersleben.de/misa/>) was employed for microsatellite mining in the whole transcriptome. Parameters were set as follows: definition (unit\_size- min\_repeats): 2-6, 3-5, 4-4, 5-4, 6-4; interruptions (max\_difference\_between\_2\_SSRs): 100. Based on the MISA results, Primer3 (version 1.1.4) was used to design up to three primer pairs in the flanking regions of simple sequence repeats (SSR).

#### Characterization of long non-coding RNAs

The software CNCI (version 2) (Sun *et al.*, 2013) and CPC (version 0.9-r2, Kong *et al.*, 2007) were used to assess the protein-coding potential of transcripts without annotations by default parameters for potential long non-coding RNAs (lncRNA).

#### Alternative splicing detection and validation

To analyze alternative splicing (AS) events of transcript isoforms, Coding GENome reconstruction Tool (Cogent) (Li *et al.*, 2017) was firstly used to partition transcripts into gene families based on k-mer similarity and reconstruct each family into a coding reference genome based on De

Brujin graph methods. Then, the software SUPPA (Alamancos *et al.*, 2015) was used to analyze alternative splicing events of transcript isoforms.

#### RNA-Seq library construction and sequencing

After high-quality RNA was extracted, mRNA was enriched by Oligo(dT) beads. Then, the enriched mRNA was fragmented into short fragments using fragmentation buffer and reverse transcribed into cDNA with random primers. Second-strand cDNA was synthesized by DNA polymerase I, RNase H, dNTP, and buffer. Then, the cDNA fragments were purified with QiaQuick PCR extraction kit, end repaired, poly(A) added, and ligated to Illumina sequencing adapters. The ligation products were size selected by agarose gel electrophoresis, PCR-amplified, and sequenced using Illumina HiSeq™4000 by Gene Denovo Biotechnology Co. (Guangzhou, China).

#### Gene expression quantification and differentially expression analysis

In order to get high-quality clean reads, raw reads were filtered according to the following rules: (a) Removing reads containing adapters; (b) Removing reads containing more than 10% of unknown nucleotides (N); (c) Removing low-quality reads containing more than 40% of low quality (Q-value  $\leq 20$  bases). The high-quality clean reads were mapped to the reference transcriptome using short reads alignment tool Bowtie2 (Langmead and Salzberg, 2012) by default parameters. The gene abundances were calculated and normalized to Reads Per kb per Million reads (RPKM) (Mortazavi *et al.*, 2008).

To identify differentially expressed genes (DEG) across groups, the edgeR package (<http://www.r-project.org/>) was used. We identified genes with a fold change  $\geq 2$  and a false discovery rate (FDR)  $< 0.05$  in a comparison as significant DEGs. DEGs were then subjected to enrichment analysis of KEGG pathways (Kanehisa *et al.*, 2008).

#### Gene expression analysis using quantitative real time PCR

The quantitative real-time PCR (qRT-PCR) was performed to validate the RNA-seq analysis on RNA extracted from pulp samples of *A. squamosa* fruits at 30 days after pollination and 6 days after harvest. Technical procedures were as fully described by Dossa *et al.* (2019). Specific primer pairs were designed using PrimerPremier 3 tool (primer length:  $18 \pm 2$  bp; search node: automatic; amplicon size: 70–200 bp; melting temperature is  $60 \pm 3^\circ\text{C}$ ) for the selected genes (Tab. S1). RNA was first transcribed into cDNA using a cDNA synthesis kit (Invitrogen, Carlsbad, CA, USA). The qRT-PCR was conducted on a Roche Lightcycler<sup>®</sup> 480 instrument using the SYBR Green Master Mix (Vazyme), according to the manufacturer's protocol. The *GAPDH* gene was used as an endogenous control.

## Results

#### Portray of the full-length transcriptome sequencing and annotation

High-quality RNAs were extracted from the fruits of two *A. squamosa* cultivars ('African Pride' [AP] and 'Gefner' [LZ]) at three growth (30, 100, and 130 days after pollination

[DAP]) and two ripening stages (6 and 8 days after harvest [DAH]) (Fig. 1) and used for single-molecule real-time (SMRT) Bell libraries on the PacBio Sequel platform using the latest P6–C4 chemistry. The specific bioinformatics analysis pipeline for data is outlined in Fig. 2A. In total, 57 Gb clean reads with 32,757,220 subreads were generated (Tab. S1). A total of 783,647 circular consensus sequences (CCS) with a mean length of 2,398 bp were classified as FL transcripts based on the presence of 5'-primers, 3'-primers, and poly(A) tails. After polishing, clustering, and demultiplexing of FL transcripts, we obtained 48,209 non-redundant high-quality FL non-chimeric (FLNC) transcripts with an average length of 2,076.6 bp, N50 of 2,500 bp, and GC content of 44.58% (Tab. S2). The length distribution of the obtained FLNC sequences ranged from 61 to 11,414 bp (Fig. 2B), which is similar to previous SMRT reports in plants (Feng *et al.*, 2019; Yue *et al.*, 2020).

Functional annotation of the FLNC transcripts was conducted by using four different public databases, including Nr, Swissprot, KOG, and KEGG (Apweiler, 2001; Koonin *et al.*, 2004; Kanehisa *et al.*, 2008; Wilke *et al.*, 2012). We identified homologs for 45,989, 39,569, 32,035 and 23,647 FLNC transcripts in Nr, Swissprot, KOG and KEGG, respectively. This represents in total 46,008 successfully annotated FLNC transcripts (95% of the total FLNC transcripts) with 19,993 transcripts (41% of the total FLNC transcripts) simultaneously annotated in all databases (Fig. 2C). The comparison of the FLNC transcript sequences with available genome sequenced plant species from the Nr database showed that *A. squamosa* shared the highest transcript number with *Nelumbo nucifera* (Fig. 2D).

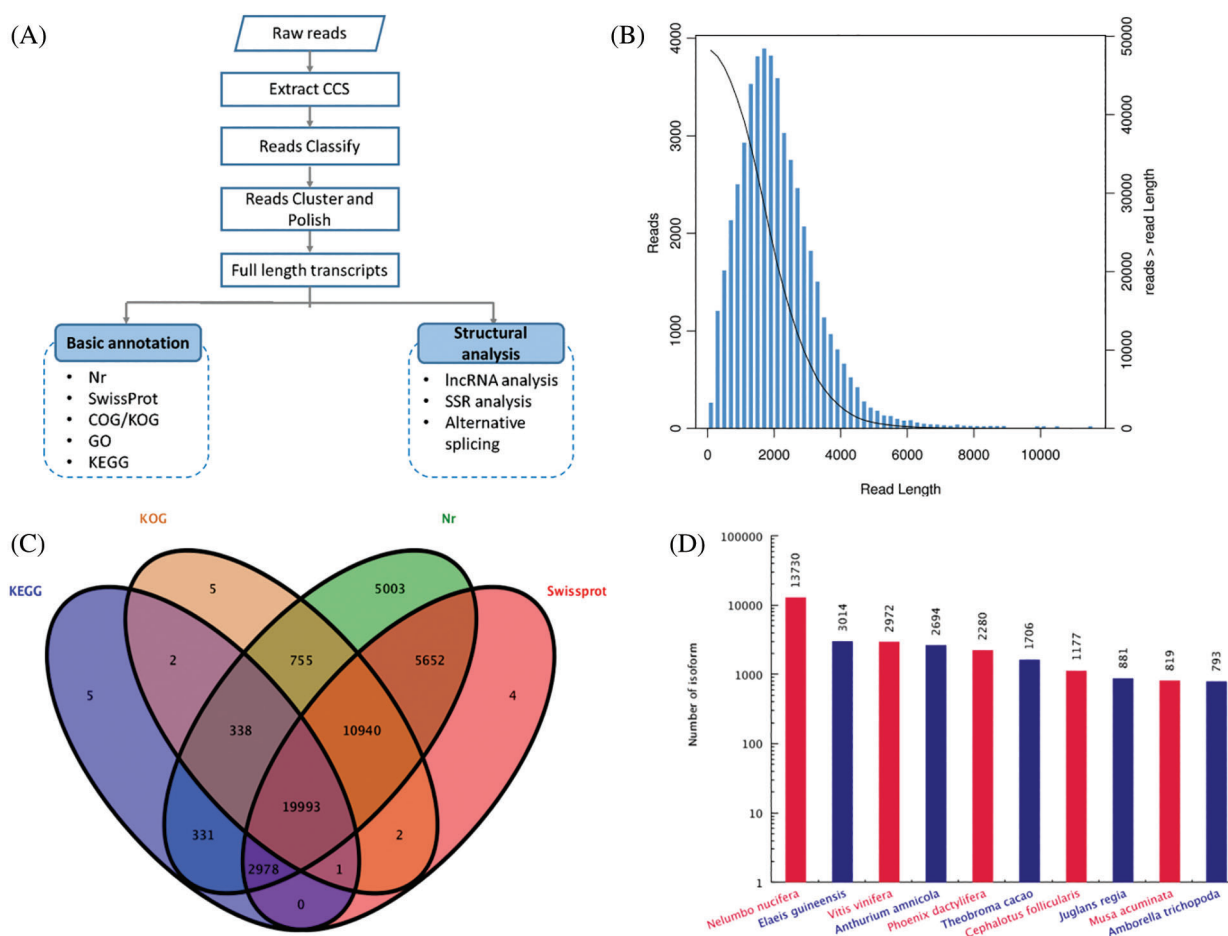
We further compared the FLNC sequences to PlantTFdb (Jin *et al.*, 2017) to predict TF families. In total, 1,838 TF encoding genes were detected and classified into 53 families (Tab. S3). C3H and bHLH family members were the most abundant, while STAT and S1Fa-like families had each a single member.

#### Alternative splicing analysis

Since PacBio sequencing generates full-length transcripts, it offers the possibility of identifying alternative splicing (AS) by directly comparing isoforms of the same gene without *de novo* assembly (Qian *et al.*, 2019). In this study, we identified 10,400 AS events from 2,541 unigenes (Tab. S3). Most of the unigenes had 2–4 isoforms, while *COGENT005277* featured 48 isoforms (Fig. 3A, Tab. S4). To obtain detailed information, the number of unigenes falling in each of the seven primary AS types, including exon skipping (SE), alternative 5' and 3' splice sites (A5/A3), mutually exclusive exons (MX), intron retention (RI), and alternative first and last exons (AF/AL), was investigated (Fig. 3B). RI was by far the most abundant AS event (47% of the total unigenes regulated by AS) in *A. squamosa*, while only one unigene (*COGENT004817*) showed MX AS event (Fig. 3C, Tab. S4).

#### Long non-coding RNAs and Simple Sequence Repeat characterizations

The FL transcript sequences that were not annotated to the four databases were submitted to the CNCI (Sun *et al.*, 2013) and CPC (Kong *et al.*, 2007) programs, and only



**FIGURE 2.** Overview of full-length transcriptome sequencing and annotation in *A. squamosa*. (A) The pipeline of bioinformatics analysis of the PacBio Sequel data. (B) The length distribution of isoforms. The horizontal axis represents the length; the vertical axis represents the number of isoforms within the length range. (C) Functional annotation against various databases. Protein databases, including Swiss-Prot, KOG, KEGG, and NR, were used. (D) Comparison of the full-length transcriptome of *A. squamosa* with genome sequenced plant species in the NR database. The numbers above bars show the shared isoforms between *A. squamosa* and corresponding species.

lncRNAs predicted by both programs were kept. From the 48,209 FLNC transcripts, 1,768 (3.67%) lncRNAs were commonly identified by both programs in *A. squamosa* (Fig. 4A).

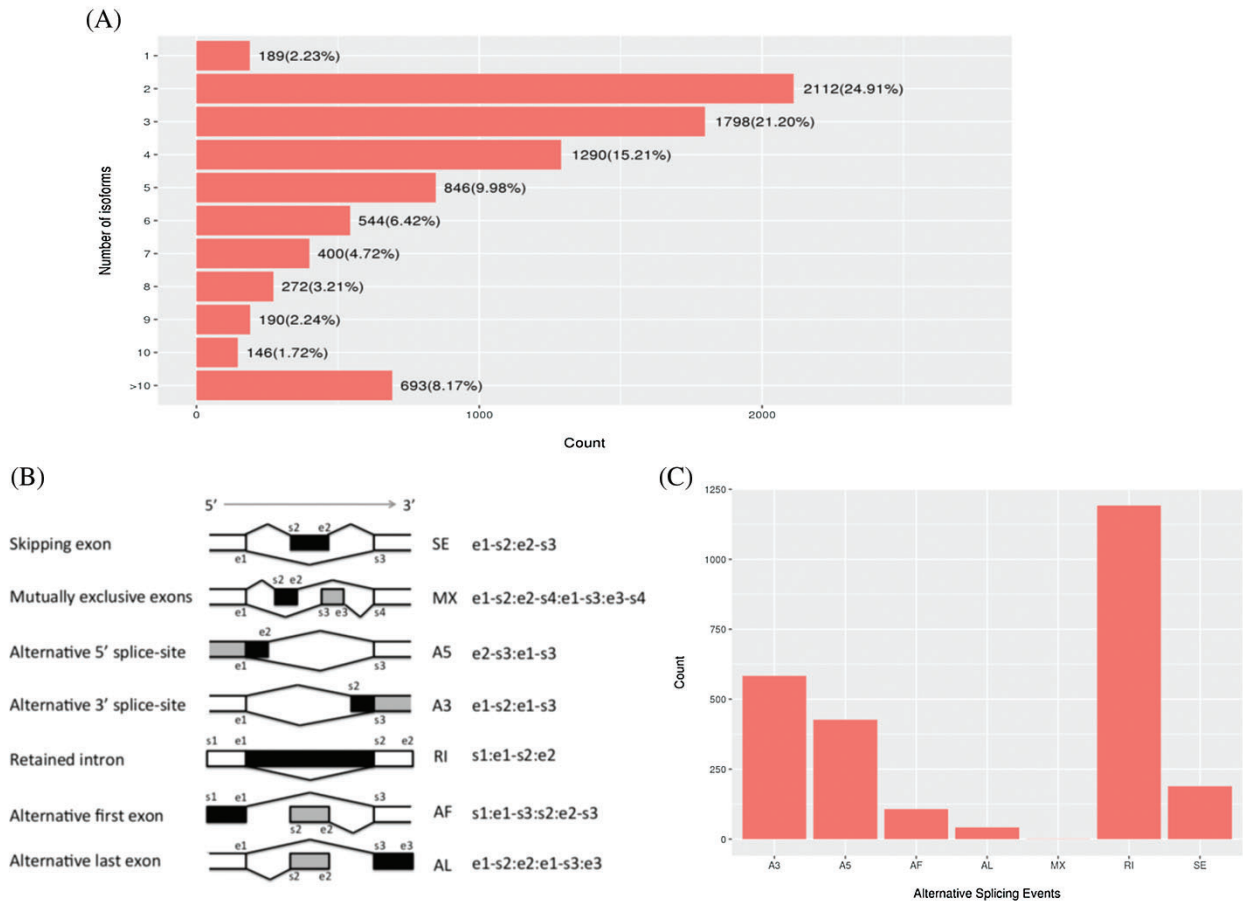
SSRs are easily detected by the standard PCR technology, representing one of the most used molecular markers (Dossa et al., 2017). Here, we characterized the SSRs in *A. squamosa* transcriptome. In total, we annotated 15,061 SSRs contained in 11,574 sequences, representing 24% of the total FLNC transcripts (Tab. S5). Di- and tri-nucleotide SSRs were the most abundant (6,534 and 5,881, respectively), while penta-nucleotide SSRs were the least represented (466) (Fig. 4B). Moreover, 1403 SSRs present in compound formation were detected (Tab. S5). Using Primer3 (version 1.1.4), we designed up for three primer pairs to each SSR locus.

#### RNA-seq and differentially expressed transcripts overgrowth/ripening stages

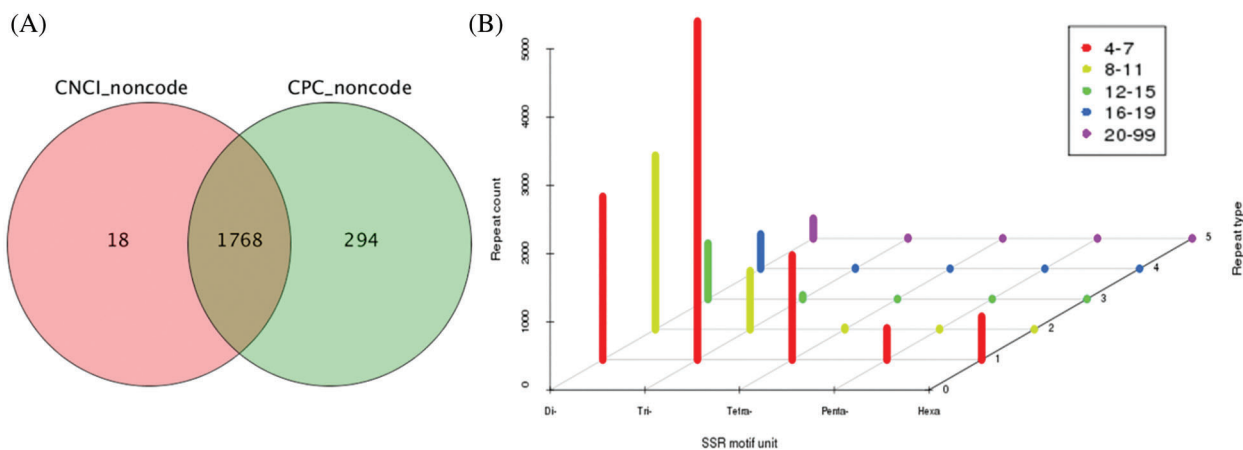
We further sequenced short reads transcriptome from fruit samples of the two cultivars at the five growth and ripening stages using the Illumina platform. The short reads were used in the polishing process of FL transcripts (Fig. 2A). Globally, 30 samples were sequenced (three biological replicates for each sample), generating over 276 Gb raw data. After cleaning, 273 Gb data were kept for subsequent

analysis, and quality assessment showed a high Q30 score (>93%) (Tab. S6). Statistics of the short reads mapping to the FL transcriptome are presented in Tab. S7. Overall, high mapping ratios, ranging from 86.71 to 90.98 %, were obtained. Gene expression was estimated based on the Reads Per kb per Million reads (RPKM) method, and overall, 47,967 genes were found expressed in all samples. Based on the gene expression profiles, we performed a principal component analysis (PCA) of the 30 samples to assess the quality of the biological replicates and clustering patterns of samples. As shown in Fig. 5A, all biological replicates clustered together, showing high correlations between them. In addition, we observed that samples from the two cultivars grouped together at the different stages except at stage 4 (6 DAH), indicative of profound transcriptome changes between the two cultivars at this specific stage. The PC1 separated the first two stages (30 and 100 DAP) from the latter stages, while the PC2 clearly separated stages 4 and 5 from earlier stages.

In order to detect differentially expressed genes (DEG), pairs of samples were compared by using stage 1 (30 DAP) as a reference for each cultivar. Genes with  $|\log_2 \text{Fold Change}| > 1$  and false discovery rate (FDR)  $< 0.05$  were selected as DEGs (Fig. 5B). The numbers of DEGs ranged



**FIGURE 3.** Characteristics of alternative splicing (AS) events and isoform numbers in *A. squamosa*. (A) The graph counts the number of isoforms that the gene contains. The Y-axis indicates the number of isoforms per gene, and the X-axis represents the number of genes and percentages that contain the corresponding number of isoforms. (B) Schematic diagram of the primary seven types of AS events. (C) AS event types and number of genes corresponding to each AS event. The X-axis represents the types of AS events, while the Y-axis represents the number of genes.

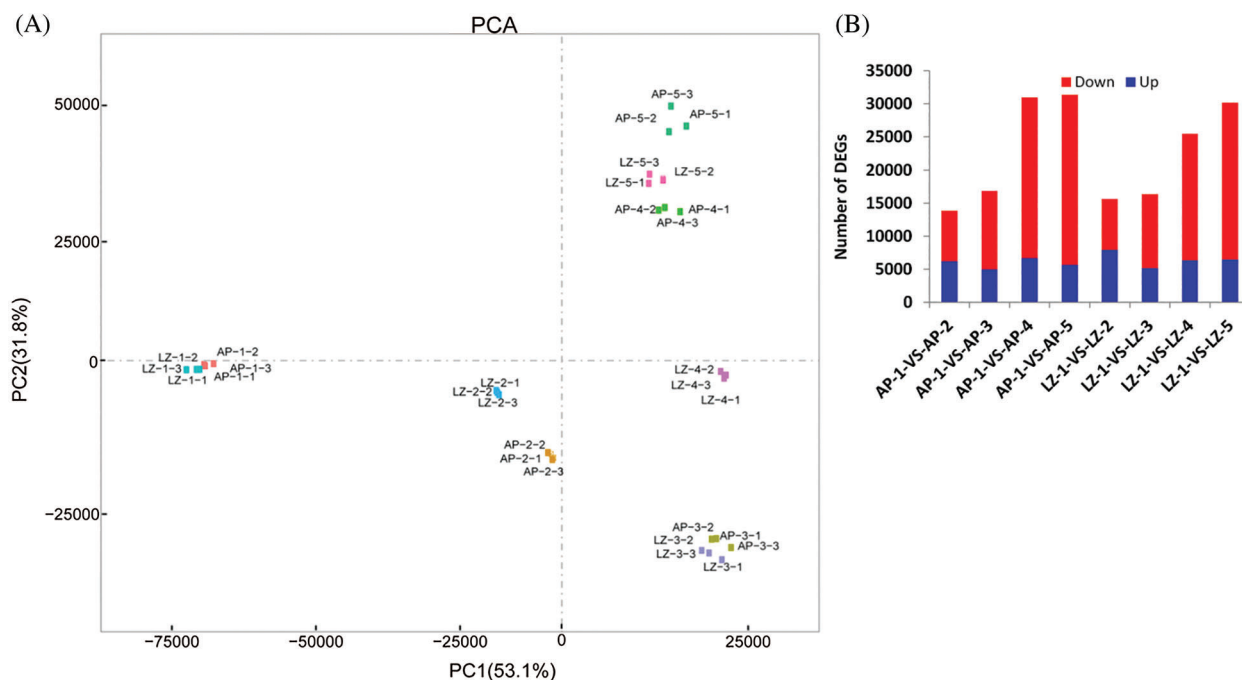


**FIGURE 4.** Identification of long non-coding RNAs (lncRNA) and simple sequence repeat (SSR) in *A. squamosa* full-length transcriptome. (A) Venn diagram showing the number of shared and specific detected long non-coding RNAs using CNCI and CPC programs. (B) Statistics and characteristics of the SSRs detected. The X-axis represents the SSR types, while the Y-axis values represent the number of repetitions to be matched by color legend, and the Z-axis is the number of SSRs.

from 13,842 (AP-1-VS-AP-2) to 31,399 (AP-1-VS-AP-5), reflecting that as the fruit maturation process goes on, more gene expression levels are altered. In particular, the majority of DEGs were downregulated, and as suspected from the PCA result, conspicuous transcriptome changes occurred at stages 4 and 5 in both cultivars.

KEGG annotation of the DEGs revealed that most of them are involved in metabolism, particularly in carbohydrate metabolism, energy metabolism, and amino acid metabolism pathways (Fig. S1).

To validate the DEG analysis, we selected ten genes involved in carbohydrate metabolism and performed qRT-PCR



**FIGURE 5.** Principal component analysis (PCA) based on RPKM data (A) and differential expressed genes between groups of samples (B). AP: African pride, LZ: Gefner; Stages 1–5 represent 30 DAP, 100 DAP, 130 DAP (harvest day), 6 DAH and 8 DAH. DAP: days after pollination and DAH: days after harvest.

analysis at 30 DAP and 6 DAH (Fig. S2). The *GAPDH* gene was used as an endogenous control. The regulation patterns of the expression levels of the selected genes were similar to those in the RNA-seq report with a strong Pearson correlation score ( $R^2 = 0.83$ ). This result confirms well the reliability of the DEG analysis and subsequent interpretation.

#### Differentially expressed transcripts involved in sugar and starch metabolism

Sugar content in fruit is an important quality criterion; therefore, we studied the genes involved in sugar accumulation in *A. squamosa*. First, we analyzed the pattern of sugar accumulation in *A. squamosa* fruit overgrowth and ripening stages in the two cultivars. As shown in Tab. 1, the trend of sugar accumulation was quite similar between the two cultivars. Globally, sugar content increased from the early growth stages to the ripening stages. Glucose and fructose were the dominant sugars, and a striking sugar accumulation occurred at 6 DAH. That indicates

that sugar accumulates mainly post-harvest (during ripening) in *A. squamosa*. To uncover the molecular mechanism of sugar accumulation in *A. squamosa*, we investigated the changes in gene expression related to sugar metabolism. In total, we detected 257 DEGs belonging to 27 families playing various functions in sugar accumulation. We mapped these DEGs to the sugar metabolism pathway, and the general patterns of regulation of these gene families in late stages (100 DAP, 130 DAP, 6 DAH and 8 DAH) compared to the first stage (30 DAP) are presented in Fig. 6. Genes highlighted in red and blue are those constantly up-regulated and down-regulated, respectively, after 30 DAP. Genes highlighted in gray display both up-regulation and down-regulation patterns (mixed regulation) at specific developmental stages after 30 DAP. Genes in white boxes are those not differentially expressed between developmental stages (Fig. 6).

In sink cells, sucrose is converted to fructose and glucose by invertase (INV, EC 3.2.1.26) and alpha-glucosidase ( $\alpha$ GL,

**TABLE 1**

**Determination of sugar content in fruit of two *Annona squamosa* cultivars over five growth and ripening stages. AP = African pride, DAP = Day after pollination, DAH = Day after harvest, ND = not detected.**

| Growth/ripening stages | Soluble sugar (%) |              | Glucose (g/100g) |             | Fructose (g/100g) |             | Sucrose (g/100g) |             |
|------------------------|-------------------|--------------|------------------|-------------|-------------------|-------------|------------------|-------------|
|                        | AP                | Gefner       | AP               | Gefner      | AP                | Gefner      | AP               | Gefner      |
| 30 DAP                 | 1.5 ± 0.49        | 1.81 ± 0.11  | 0.63 ± 0.03      | 0.74 ± 0.02 | 0.62 ± 0.01       | 0.74 ± 0.08 | ND               | ND          |
| 100 DAP                | 1.88 ± 0.27       | 2.4 ± 0.31   | 0.65 ± 0.02      | 1 ± 0.08    | 0.78 ± 0.04       | 1 ± 0.11    | ND               | ND          |
| 130 DAP (Harvest day)  | 2.2 ± 0.41        | 2.17 ± 0.35  | 0.91 ± 0.05      | 0.85 ± 0.04 | 1 ± 0.09          | 0.94 ± 0.14 | ND               | ND          |
| 6 DAH                  | 20.1 ± 1.60       | 14.52 ± 0.96 | 9.8 ± 0.64       | 7 ± 0.77    | 9.1 ± 0.78        | 7 ± 0.96    | 0.7 ± 0.09       | ND          |
| 8 DAH                  | 19.03 ± 2.33      | 18.18 ± 1.58 | 9.6 ± 1.04       | 9.1 ± 1.05  | 8.4 ± 0.94        | 7.9 ± 1.02  | 0.52 ± 0.06      | 0.73 ± 0.02 |

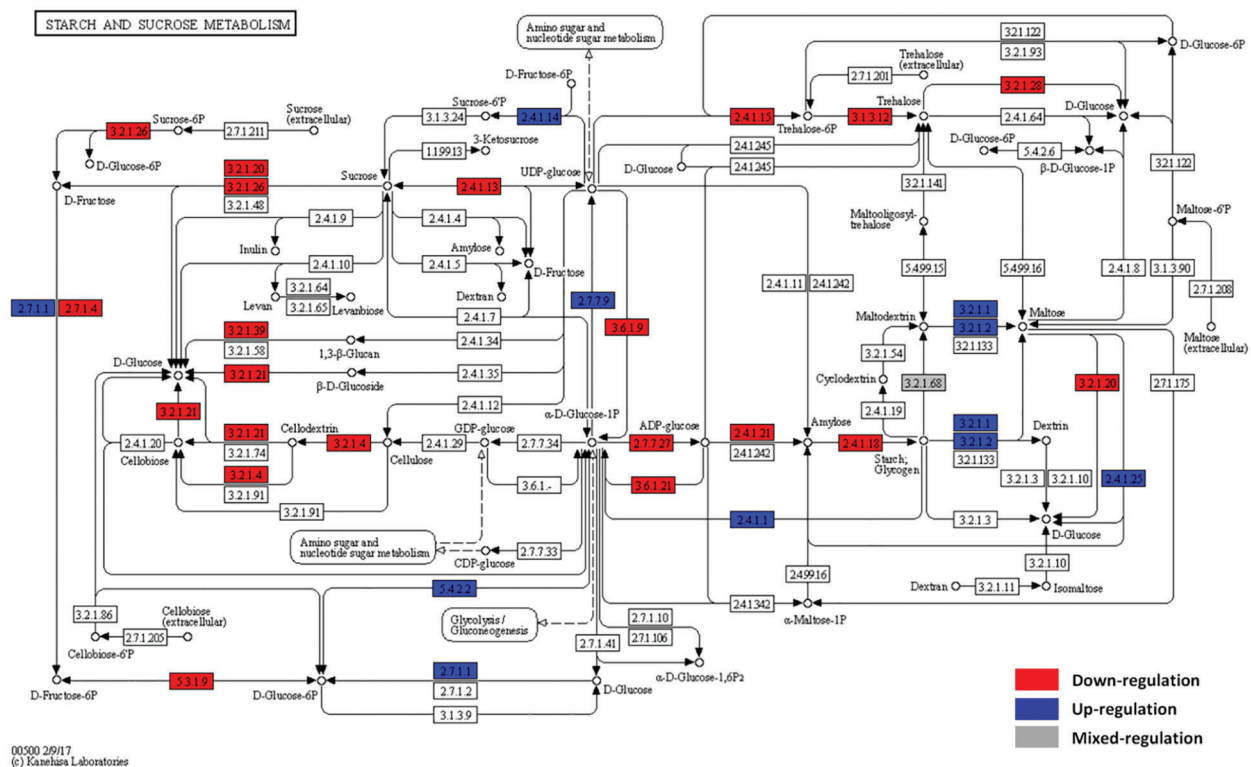
EC 3.2.1.20). We identified four INV and ten  $\alpha$ GL genes, which were mostly down-regulated mainly at 6 and 8 DAH in both cultivars (Fig. 7), suggesting that these genes do not positively contribute to sugar accumulation in *A. squamosa*. We also identified 39 sucrose synthase (SuSy, EC 2.4.1.13) genes, known to convert sucrose into fructose and UDP-glucose (UDPG). Most SuSy DEGs were downregulated in both cultivars, with more profound gene repression at 6 and 8 DAH (Fig. 7). We deduce that SuSy is negatively correlated with sugar accumulation in *A. squamosa*. Hexokinase (HK, EC 2.7.1.1) and fructokinase (FK, EC 2.7.1.4) phosphorylate glucose and fructose to glucose 6-phosphate (G6P) and fructose 6-phosphate (F6P). We identified only one HK gene steadily upregulated, while five FK genes were strongly downregulated, mainly at 6 and 8 DAH in both cultivars (Fig. 7). This suggests a mechanism toward limiting the conversion of glucose and fructose. We also identified 11 sucrose phosphate synthase (SPS, EC 2.4.1.14) genes that re-synthesize sucrose by combining F6P and UDPG. Most of the SPS was upregulated, indicative of their positive contribution to sugar accumulation. UDPG is also formed through actions of phosphoglucomutase (PGM, EC 5.4.2.2) and UTP-glucose-1-phosphate uridylyltransferase (UGP, EC 2.7.7.9). Here, we identified one PGM (up-regulated) and three UGP, two of which were upregulated, which may have led to high UDPG synthesis. This coupled with the strong induction of SPS may promote high sucrose synthesis. Globally, most of the gene families involved in sucrose conversion into glucose and fructose, or the subsequent conversion of glucose and fructose into other metabolites, were downregulated over fruit growth

and ripening stages. Only genes involved in the synthesis of sucrose were found primarily up-regulated (Figs. 6 and 7).

During fruit postharvest ripening, starch is degraded into soluble sugars (Xiao et al., 2018). In this study, we found that the genes involved in starch synthesis (glucose-1-phosphate adenylyltransferase [EC 2.7.7.27], 1,4- $\alpha$ -glucan branching enzyme [EC 2.4.1.18] and starch synthase [EC 2.4.1.21]) were highly repressed during late growth and ripening stages as compared to the early growth stage (Figs. 6 and 7). However, genes contributing to starch degradation were strongly up-regulated, including two  $\alpha$ -amylase (AMY, EC 3.2.1.1), 36  $\beta$ -amylase (EC 3.2.1.2), and 13 4- $\alpha$ -glucanotransferase (EC 2.4.1.25) genes (Figs. 6 and 7). Likewise, six isoamylases (ISA, EC 3.2.1.68) genes involved in starch degradation were down-regulated prior harvest period but were strongly up-regulated after harvest, indicating that they positively contribute to sugar accumulation in *A. squamosa* fruit during ripening.

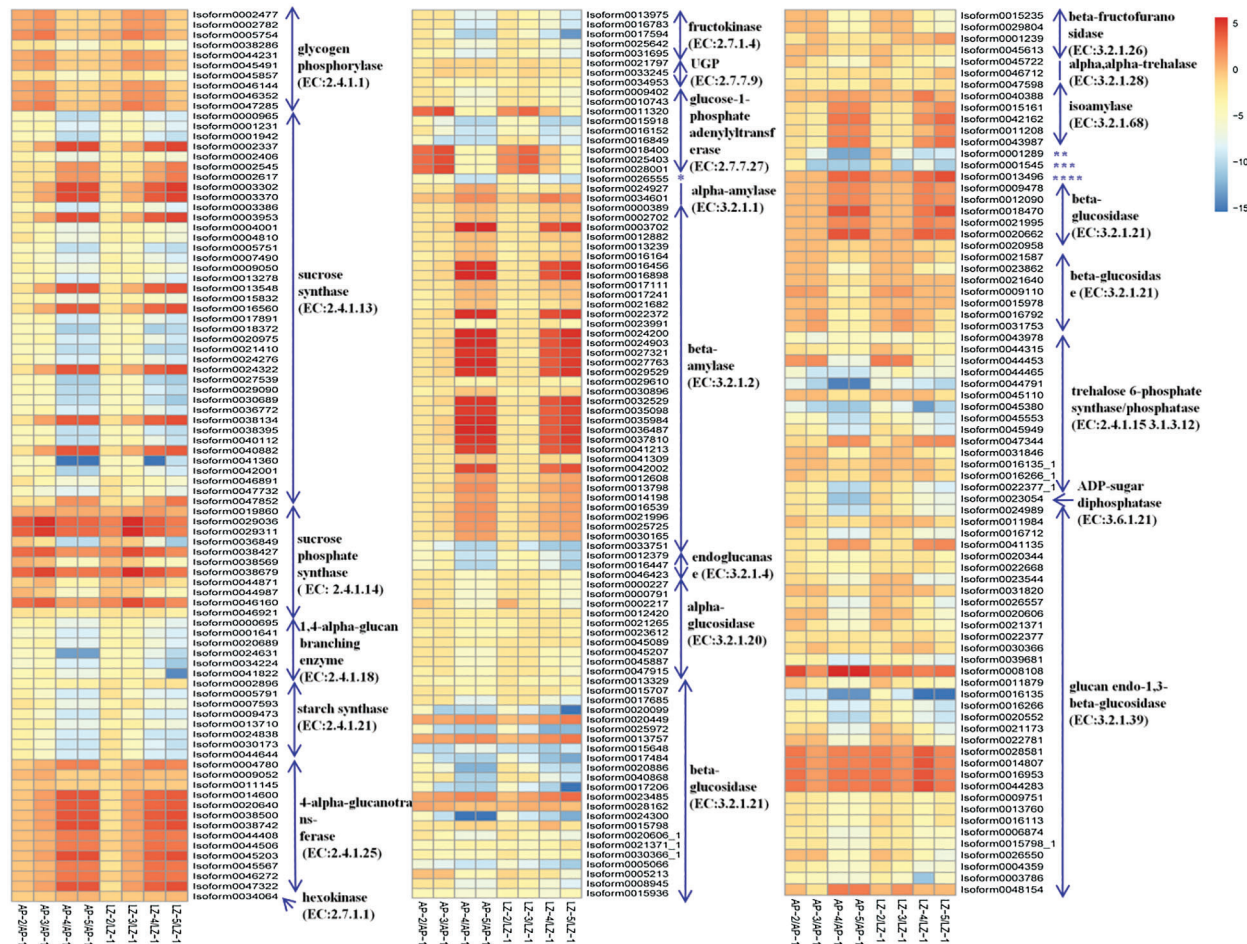
Collectively, analysis of sugar metabolism-related DEGs revealed that starch degradation, which is triggered during ripening, is the most active channel for soluble sugar accumulation in *A. squamosa* fruit.

*Differentially expressed transcripts involved in sugar transport*  
Sucrose transport from source tissues to sink tissues is mediated by special genes belonging to the families of sucrose transporters (SUC), sugar will eventually be exported transporters (SWEET), sugar transport protein (STP), etc. In this study, we searched for all DEGs annotated as SUC, SWEET and STP. In total, four, six, and six SUC, SWEET, and STP DEGs, respectively, were



**FIGURE 6.** The differentially expressed genes (DEG) between late stages (100 DAP, 130 DAP, 6 DAH and 8 DAH) compared to the first stage (30 DAP) were mapped to the starch and sucrose metabolism pathway. In total, 257 DEGs belonging to 27 families were identified and the general patterns of regulation of these gene families were highlighted with colors. Genes highlighted in red, blue, white, and gray are those constantly down-, up-, no- and mixed regulated (down- and up-regulated), respectively.





**FIGURE 7.** Heatmap displaying the expression fold change ( $\log_2$  fold change) of differentially expressed genes involved in starch and sugar metabolism. AP: African pride, LZ: Gefner; Stages 1–5 represent 30 days after pollination (DAP), 100 DAP, 130 DAP (Harvest day), 6 days after harvest (DAH) and 8 DAH. \*trehalose 6-phosphate phosphatase (EC:3.1.3.12), \*\*ectonucleotide pyrophosphatase/phosphodiesterase family member 1/3 (EC:3.1.4.1 3.6.1.9), \*\*\*glucose-6-phosphate isomerase (EC:5.3.1.9), \*\*\*\*phosphoglucomutase (EC:5.4.2.2).

identified across the different growth and ripening stages (Tab. 2). SUCs were globally upregulated in late stages as compared to the 30 DAP, indicating that sucrose transport to *A. squamosa* fruit is continuous. Nonetheless, the increase in SUC gene expression was not very high, meaning this process is not very decisive for sugar accumulation. On the opposite, SWEET and STP genes were more active at 30 DAP and subsequently strongly repressed at later stages. Collectively, our data suggest that sugar transporters play a preponderant role at the early growth stages, but, during ripening, their contribution to sugar accumulation in *A. squamosa* fruit is very limited.

### Discussion

In this study, we generated, for the first time, full-length transcriptome data for *A. squamosa*. Although the numerous nutritional, medicinal, and industrial applications of this species, the lack of genome sequence and limited transcriptome data (Gupta *et al.*, 2015; Liu *et al.*, 2016) impair genomic research in *A. squamosa*. Hence, our FL transcriptome provides crucial genomic data to instigate gene expression, gene discovery, cloning, and evolutionary studies. We identified 48,209 high-quality FL non-chimeric (FLNC) transcripts with an average length of 2,076.6 bp, N50 of 2,500 bp (Tab. S2). The longest unigene has a length

of 11,414 bp, which could not be identified by short-read sequencer such as Illumina (Hoang *et al.*, 2017). However, the number of unigenes detected in this study was lower than Illumina based RNA-seq (71,948 unigenes), which is understandable since RNA-seq can recover more transcripts than PacBio sequencing due to relatively higher sequencing depth (Rhoads and Au, 2015).

The advantage of full-length transcriptome sequencing is the possibility to uncover the alternative splicing (AS) events and isoforms (Qian *et al.*, 2019), which are major mechanisms for the enhancement of transcriptome diversity (Keren *et al.*, 2010). We observed that only ~5% of the total unigenes yielded different isoforms in *A. squamosa* fruit transcriptome (Tab. S4), which is quite lower than reports in *Crocus sativus* (33.7%, Qian *et al.*, 2019), *Lateolabrax maculatus* (28%, Tian *et al.*, 2019) and *Gossypium australe* (50.83%, Feng *et al.*, 2019). This low percent of genes with isoforms could be explained by the fact that only fruit tissue was used in this study, while other studies combined various tissues. Therefore, we propose that future works should sample various tissues such as flower, leaf, stem, seed, and root in order to have a better insight into the complexity of AS and isoforms in *A. squamosa* transcriptome and to discover more genes. Intron retention (RI) was found to be the most abundant AS in *A. squamosa* (Fig. 3), which is

TABLE 2

Log<sub>2</sub> fold change of the sugar transporter genes in *Annona squamosa* at late growth and ripening stages as compared to early stage (30 DAP).

| Gene ID               | Symbol  | AP-2/AP-1 | AP-3/AP-1 | AP-4/AP-1 | AP-5/AP-1 | LZ-2/LZ-1 | LZ-3/LZ-1 | LZ-4/LZ-1 | LZ-5/LZ-1 |
|-----------------------|---------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| <i>Isoform0003410</i> | SUC     | 1.36      | 0.00      | 1.47      | 1.56      | 1.43      | 0.00      | 2.17      | 1.47      |
| <i>Isoform0006940</i> | SUC     | 0.00      | 0.00      | 1.05      | 0.00      | 0.00      | 0.00      | 1.54      | 0.00      |
| <i>Isoform0007555</i> | SUC     | 0.00      | 0.00      | 1.35      | 1.06      | 0.00      | 0.00      | 1.35      | 1.18      |
| <i>Isoform0008947</i> | SUC     | 1.31      | 0.00      | 1.95      | 1.65      | 1.68      | 0.00      | 2.59      | 1.97      |
| <i>Isoform0038501</i> | STP9    | -2.37     | -6.69     | -4.95     | -6.69     | 0.00      | -2.17     | -5.91     | -5.91     |
| <i>Isoform0021219</i> | STP7    | 1.76      | 0.00      | -1.76     | -2.58     | 1.26      | 0.00      | -1.73     | -2.09     |
| <i>Isoform0019302</i> | STP5    | 1.58      | 0.00      | -1.57     | -2.40     | 1.99      | 1.58      | 0.00      | -1.66     |
| <i>Isoform0020821</i> | STP14   | 2.32      | -1.75     | -5.69     | -5.27     | 1.06      | 0.00      | -3.27     | -7.80     |
| <i>Isoform0015882</i> | STP13   | -1.63     | -2.55     | 4.25      | 2.83      | -2.52     | -3.68     | 0.00      | 3.20      |
| <i>Isoform0047494</i> | STP13   | 0.00      | -4.25     | 2.58      | 0.00      | 0.00      | -3.81     | 0.00      | 1.68      |
| <i>Isoform0008390</i> | SWEET2A | -1.54     | -2.95     | -2.39     | -1.87     | 0.00      | -2.73     | -2.00     | -2.48     |
| <i>Isoform0032697</i> | SWEET17 | -5.31     | -7.63     | -8.56     | -10.48    | -3.33     | -7.85     | -6.87     | -9.17     |
| <i>Isoform0021186</i> | SWEET16 | -5.01     | -4.43     | -9.34     | -9.34     | -2.39     | -8.45     | -8.45     | -5.13     |
| <i>Isoform0030431</i> | SWEET16 | -12.01    | -12.01    | -12.01    | -12.01    | -4.24     | -12.26    | -7.53     | -12.26    |
| <i>Isoform0032807</i> | SWEET1  | 0.00      | -1.41     | -8.21     | -8.47     | -1.36     | -1.34     | -7.45     | -9.10     |
| <i>Isoform0033589</i> | SWEET1  | 0.00      | -3.47     | -7.65     | -8.81     | 0.00      | -3.55     | -8.48     | -8.32     |

common in plants (Marquez *et al.*, 2012; Shen *et al.*, 2014; Feng *et al.*, 2019). We also reported lncRNAs and, more importantly, simple sequence repeat (SSR) that represents a popular marker type for plant genotyping (Dossa *et al.*, 2017). Only a few SSR markers developed in related species have been used for diversity analysis in *A. squamosa* (Nagori *et al.*, 2018); therefore, the SSR markers along with their primer sequences provided in this study (Tab. S5) will be a great resource for future diversity studies and molecular breeding.

*A. squamosa* fruit is one of the sweetest fruits containing up to 28% total sugars at the ripening stage (Brandão and Santos, 2016). Sugar content analysis in developing fruit compared to ripening fruit demonstrated that sugars mainly accumulate during the ripening stage with a continuous increase of sugars over the fruit growth and ripening stages (Tab. 1). Sugar accumulation is controlled by both the translocation of sugars and their metabolism in developing and ripening fruits (Yativ *et al.*, 2010). By analyzing genes involved in sugar metabolism, we found that most of the genes were highly active at the early growth stage (30 DAP) but were repressed at late growth and ripening stages. Similar observations were found for sugar transporter related genes and starch synthesis genes (Tab. 2, Figs. 6 and 7). On the opposite, genes involved in the synthesis of sucrose were found up-regulated during ripening, and this correlates with the detection of sucrose in ripe fruit (Tab. 1). Likewise, genes involved in starch degradation were markedly induced during ripening. For example, we observed the up-regulation of sucrose phosphate synthase genes after 30 DAP. It was reported that over-expression of sucrose phosphate synthase genes enhances sucrose content in sugarcane and tomato fruit (Nguyen-Quoc *et al.*, 1999;

Anur *et al.*, 2020). We identified several invertase genes downregulated after 30 DAP in both cultivars, indicative of their negative correlation with sugar accumulation. In accordance with our report, Wang *et al.* (2015) demonstrated that ectopic over-expression of *EjVIN* from loquat in tobacco significantly decreased sucrose content. In this study, we noticed the high activity of sucrose synthase genes at early developmental stages, while at late stages, the expression levels of sucrose synthase genes were strongly downregulated. The action of sucrose synthase genes on sucrose cleavage was reported to regulate the import and compartmentation of sucrose in the early stage of tomato fruit development (Demnitz-King *et al.*, 1997; D'Aoust *et al.*, 1999). In addition, it has been documented through transgenic experiments that suppression of sucrose synthase gene expression leads to reduced starch accumulation in potato tubers (Zrenner *et al.*, 1995) and carrot taproots (Tang and Sturm, 1999). These results showed the key roles of sucrose synthase genes in sucrose import and starch accumulation in sink tissues. Therefore, we conclude that the high activity of sucrose synthase genes promotes sucrose import and starch accumulation in developing *A. squamosa* fruits. And during fruit ripening stages, the accumulated starch is degraded, contributing to the fruit sweetness. In support of this conclusion, we observed several alpha-amylase, beta-amylase, 4-alpha-glucanotransferase, and isoamylase genes highly induced during *A. squamosa* fruit ripening, which are known to be involved in sugar degradation in fruit (Purgatto *et al.*, 2001; Radchuk *et al.*, 2009). Concerning sugar transporters, we observed that they play a preponderant role at the early growth stages of *A. squamosa* fruit. Meteier *et al.* (2019) found that over-expression of the *VvSWEET4* in grapevine promotes sugar

transport. Similarly, ectopic expression of *PbSUT2* enhances sucrose content in tomato fruit (Wang *et al.*, 2016). Collectively, our data showed that during early fruit development stages in *A. squamosa*, sucrose is transported and loaded into fruit, and starch is highly accumulated. During late development and ripening stages, starch is degraded to promote sugar accumulation in the fruit. Extensive studies in banana, tomato, and apple also showed a shift from starch synthesis to starch breakdown in ripening fruit leading to the accumulation of soluble sugars, with a significant impact on fruit taste and flavor (Beck and Ziegler, 1989; Nascimento *et al.*, 2000; Mota *et al.*, 1997; Thammawong and Arakawa, 2010; Maria *et al.*, 2016; Cordenunsi-Lysenko *et al.*, 2019). Interestingly, we identified 56 genes involved in starch degradation during ripening, which should be further studied to uncover the transcription factors that regulate their expression levels (Xiao *et al.*, 2018). This information will be cardinal for controlling the level of starch degradation during *A. squamosa* fruit ripening, since it has been shown that genotypic variation in sugar content is mainly associated with the ability to breakdown accumulated starch (Cordenunsi-Lysenko *et al.*, 2019).

## Conclusions

We generated a high-quality, full-length transcriptome assembly using PacBio Iso-seq, representing the most complete genomic data available in *A. squamosa*. We further provided extensive short reads data and elucidated the molecular mechanisms of sugar accumulation in developing and ripening fruits. Our work will facilitate future functional and evolutionary studies as well as molecular breeding mainly aiming at controlling sugar content in *A. squamosa* fruit.

## Supplementary Figures

**Figure S1.** KEGG annotation of the differentially expressed genes identified between pairs of samples. AP = African pride, LZ = Gefner; Stages 1-5 represent 30 DAP, 100 DAP, 130 DAP (Harvest day), 6 DAH and 8 DAH. DAP = day after pollination and DAH = day after harvest.

**Figure S2.** qRT-PCR validation of ten selected genes. (A-J) Gene expression analysis based on qRT-PCR. The x-axis represents the ten genes while the y-axis represents the relative expression of each gene. The bars show standard deviation. AP = African pride, LZ = Gefner; Stages 1 and 4 represent 30 DAP and 6 DAH. (K) Pearson correlation analysis between RNA-seq and qRT-PCR expression profiles. DAP = day after pollination and DAH = day after harvest.

## Supplementary Tables

**Table S1.** Primer sequences of genes used for qRT-PCR validation of differentially expressed genes

**Table S2.** PacBio sequencing data statistics

**Table S3.** List of genes annotated as transcription factors

**Table S4.** Alternative events and isoforms identified in *Annona squamosa* full-length transcriptome

**Table S5.** Characterization of the simple sequence repeat and development of primer pairs in *Annona squamosa* full-length transcriptome

**Table S6.** Sequencing data statistics of Illumina RNA-seq

**Table S7.** Mapping statistics of Illumina RNA-seq data

**Author Contributions:** RF, WH and ZA conceived and designed the study. RF, WH, JY, XL, JZ, SZ, BD, WT executed the plant infection experiments, RNA extraction and sequencing data analysis, and result interpretation. WH performed the qRT-PCR validation experiment and revised the drafts of the manuscript; RF drafted the manuscript; ZA supervised the study, provided funding, and revised drafts of the manuscript. All the authors read and approved the manuscript.

**Availability of Data and Materials:** The RNA-seq datasets generated in the current study were submitted to NCBI SRA: PRJNA639613. All the supplementary tables and figures are available from the corresponding author on request.

**Funding Statement:** This work was funded by the key research and development program of Guangxi, China (Grant No. AB19245004), the Science and Technology Major Project of Guangxi, China (Grant No. AA17204026, Grant No. AA17204097-7), the Department of Agriculture and Rural Affairs of Guangxi, China (Grant No. nycytgxgxcxtd-17-15), the Science and technology project of Guangxi Academy of Agricultural Sciences, China (Grant No. 2015YT48). The funder has no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- Abdualrahman MAY, Ma H, Zhou C, Yagoub AEA, Ali AO, Tahir HE, Wali A (2019). Postharvest physicochemical properties of the pulp and seed oil from *Annona squamosa* L. (Gishta) fruit grown in Darfur region. *Sudan Arabian Journal of Chemistry* **12**: 4514–4521. DOI 10.1016/j.arabjc.2016.07.008.
- Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyra E (2015). Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* **21**: 1521–1531. DOI 10.1261/rna.051557.115.
- Anur RM, Mufithah N, Sawitri WD, Sakakibara H, Sugiharto B (2020). Overexpression of sucrose phosphate synthase enhanced sucrose content and biomass production in transgenic sugarcane. *Plants* **9**: 200. DOI 10.3390/plants9020200.
- Apweiler R (2001). Functional information in SWISS-PROT: the basis for large-scale characterisation of protein sequences. *Briefings in Bioinformatics* **2**: 9–18. DOI 10.1093/bib/2.1.9.
- Asif MH, Lakhwani D, Pathak S, Gupta P, Bag SK, Nath P, Trivedi PK (2014). Transcriptome analysis of ripe and unripe fruit tissue of banana identifies major metabolic networks involved in fruit ripening process. *BMC Plant Biology* **14**: 94. DOI 10.1186/s12870-014-0316-1.

- Beck E, Ziegler P (1989). Biosynthesis and degradation of starch in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**: 95–117. DOI 10.1146/annurev.pp.40.060189.000523.
- Brandão APME, Santos DYAC (2016). Nutritional value of the pulp of different sugar apple cultivars (*Annona squamosa* L.). In: Simmonds MSJ, Preedy VR, eds., *Nutritional Composition of Fruit Cultivars*, pp. 195–214. Cambridge, MA: Academic Press.
- Chavan MJ, Wakte PS, Shinde DB (2010). Analgesic and anti-inflammatory activities of 18-acetoxyent-kaur-16-ene from *Annona squamosa* L. *Bark. Inflammopharmacology* **18**: 135–141.
- Chen S, Qiu G, Yang M (2019). SMRT sequencing of full-length transcriptome of seagrasses *Zostera japonica*. *Scientific Reports* **9**: 376. DOI 10.1038/s41598-019-51176-y.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**: 563–569. DOI 10.1038/nmeth.2474.
- Conesa A, Gotz S, García-Gómez JM, Terol J, Talón M, Robles M (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676. DOI 10.1093/bioinformatics/bti610.
- Cordenunsi-Lysenko BR, Nascimento J, Castro-Alves VC, Purgatto E, Fabi JP, Peroni-Okyta F (2019). The starch is (not) just another brick in the wall: The primary metabolism of sugars during banana ripening. *Frontiers in Plant Science* **10**: e0166308. DOI 10.3389/fpls.2019.00391.
- D'Aoust MA, Yelle S, Nguyen-Quoc B (1999). Antisense inhibition of tomato fruit sucrose synthase decreases fruit setting and the sucrose unloading capacity of young fruit. *Plant Cell* **11**: 2407–2418. DOI 10.2307/3870964.
- Demnitz-King A, Ho LC, Baker DA (1997). Activity of sucrose hydrolyzing enzymes and sugar accumulation during tomato fruit development. *Plant Growth Regulation* **22**: 193–201. DOI 10.1023/A:1005985213777.
- Dossa K, Mmadi MA, Zhou R, Zhang T, Su R, Zhang Y, Wang L, You J, Zhang X (2019). Depicting the core transcriptome modulating multiple abiotic stresses responses in sesame (*Sesamum indicum* L.). *International Journal of Molecular Sciences* **20**: 3930. DOI 10.3390/ijms20163930.
- Dossa K, Yu J, Liao B, Cisse N, Zhang X (2017). Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: SisatBase. *Frontiers in Plant Science* **8**: 3389. DOI 10.3389/fpls.2017.01470.
- Feng S, Xu M, Liu F, Cui C, Zhou B (2019). Reconstruction of the full-length transcriptome atlas using PacBio Iso-Seq provides insight into the alternative splicing in *Gossypium australe*. *BMC Plant Biology* **19**: 1303. DOI 10.1186/s12870-019-1968-7.
- Folgado R, Panis B, Sergeant K, Renaut J, Swennen R (2015). Unravelling the effect of sucrose and cold pretreatment on cryopreservation of potato through sugar analysis and proteomics. *Cryobiology* **71**: 432–441. DOI 10.1016/j.cryobiol.2015.09.006.
- Forster M, Rodríguez ER, Martín JD, Romero C (2003). Distribution of nutrients in edible banana pulp. *Food Technology and Biotechnology* **41**: 167–171.
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152. DOI 10.1093/bioinformatics/bts565.
- Gao L, Zhao S, Lu X, He N, Zhu H, Dou J, Liu W (2018). Comparative transcriptome analysis reveals key genes potentially related to soluble sugar and organic acid accumulation in watermelon. *PLoS One* **13**: e0190096. DOI 10.1371/journal.pone.0190096.
- Gordon S P, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, Kang D, Underwood J, Grigoriev I V, Figueroa M, Schilling J S, Chen F, Wang Z, Zheng D (2015). Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One* **10**: e0132628. DOI 10.1371/journal.pone.0132628.
- Gupta Y, Pathak AK, Singh K, Mantri SS, Singh SP, Tuli R (2015). *De novo* assembly and characterization of transcriptomes of early-stage fruit from two genotypes of *Annona squamosa* L. with contrast in seed number. *BMC Genomics* **16**: 86. DOI 10.1186/s12864-015-1248-3.
- Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ (2017). A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and *de novo* assembly from short read sequencing. *BMC Genomics* **18**: 122. DOI 10.1186/s12864-017-3757-8.
- Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **45**: D1040–D1045. DOI 10.1093/nar/gkw982.
- 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 Research* **36**: D480–D484. DOI 10.1093/nar/gkm882.
- Keren H, Lev-Maor G, Ast G (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics* **11**: 345–355. DOI 10.1038/nrg2776.
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* **35**: W345–W349. DOI 10.1093/nar/gkm391.
- Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Rogozin IB, Smirnov S, Sorokin AV, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology* **5**: R7. DOI 10.1186/gb-2004-5-2-r7.
- Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM (2012). Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology* **30**: 693–700. DOI 10.1038/nbt.2280.
- Langmead B, Salzberg S (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359. DOI 10.1038/nmeth.1923.
- Levetin E, McMahon K (2008). *Plants and Society: The Botanical Connections to Our Lives*. 5<sup>th</sup> edition, pp. 88–102. McGraw-Hill, Iowa: Dubuque.
- Li J, Haratalee Y, Denton MD, Feng Q, Rathjen JR, Qu Z, Adelson DL (2017). Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis. *Cell Discovery* **3**: 338. DOI 10.1038/celldisc.2017.31.

- Li M, Feng F, Cheng L (2012). Expression patterns of genes involved in sugar metabolism and accumulation during apple fruit development. *PLoS One* 7: e33055. DOI 10.1371/journal.pone.0033055.
- Liu K, Feng S, Pan Y, Zhong J, Chen Y, Yuan C, Li H (2016). Transcriptome analysis and identification of genes associated with floral transition and flower development in sugar apple (*Annona squamosa* L.). *Frontiers in Plant Science* 7: 1695.
- Maria T, saniklidis G T, Delis C, Nikolopoulou A, Nikoloudakis N, Karapanos I, Aivalakis G (2016). Gene transcript accumulation and enzyme activity of  $\beta$ -amylases suggest involvement in the starch depletion during the ripening of cherry tomatoes. *Plant Gene* 5: 8–12. DOI 10.1016/j.plgene.2015.10.004.
- Mariod AA, Elkheir S, Ahmed YM, Matthaus B (2010). A. squamosa and C. nilotica seeds, the effect of the extraction method on the oil composition. *Journal of the American Oil Chemists' Society* 87: 763–769. DOI 10.1007/s11746-010-1548-3.
- Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Research* 22: 1184–1195. DOI 10.1101/gr.134106.111.
- Meteier E, La Camera S, Goddard ML, Laloue H, Mestre P, Chong J (2019). Overexpression of the VvSWEET4 transporter in grapevine hairy roots increases sugar transport and contents and enhances resistance to *Pythium irregulare*, a soilborne pathogen. *Frontiers in Plant Science* 10: 766. DOI 10.3389/fpls.2019.00884.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods* 5: 621–628. DOI 10.1038/nmeth.1226.
- Mota R, Lajolo F, Cordenunsi B (1997). Composition of carbohydrates from banana cultivars (*Musa* spp.) during ripening. *Food Science and Technology* 17: 94–97. DOI 10.1590/S0101-20611997000200005.
- Nagori R, Sharma P, Purohit SD (2018). Molecular characterization of diversity in *Annona squamosa* L.: A high value tropical medicinal plant. *International Journal of Phytocosmetics and Natural Ingredients* 5: 4. DOI 10.15171/ijpni.2018.04.
- Nascimento JRO, Cordenunsi BR, Lajolo FM (2000). Sucrose synthase activity and expression during development and ripening in bananas. *Journal of Plant Physiology* 156: 605–611. DOI 10.1016/S0176-1617(00)80219-9.
- Nguyen-Quoc B, Foyer CH (2001). A role for 'futile cycles involving invertase and sucrose synthase in sucrose metabolism of tomato fruit. *Journal of Experimental Botany* 52: 881–889. DOI 10.1093/jexbot/52.358.881.
- Nguyen-Quoc B, N'Tchobo H, Foyer CH, Yelle S (1999). Overexpression of sucrose phosphate synthase increases sucrose unloading in transformed tomato fruit. *Journal of Experimental Botany* 50: 785–791. DOI 10.1093/jxb/50.335.785.
- Omkaresh BR, Suresh R, Yatish KV (2017). Optimization of *Annona squamosa* oil biodiesel production by using response surface methodology. *Biofuels* 8: 377–382. DOI 10.1080/17597269.2016.1231957.
- Pandey N, Barve D (2011). Phytochemical and pharmacological review on *Annona squamosa*. *International Journal of Research in Pharmaceutical and Biomedical Sciences* 2: 2229–3701.
- Patel JD, Kumar V (2008). *Annona squamosa* phytochemical analysis and antimicrobial screening. *Journal of Pharmacy Research* 1: 34–38.
- Paull RR, Duarte O (2011). Tropical fruits. In: Atherton J, eds. *Crop Production Science in Horticulture Series*, vol. 1, 2nd edition, pp. 123–152. CAB International, Oxfordshire.
- Perez AMC, González EA, Morales MC, Toriz F (2004). Oxoaporphine alkaloids in guatteradiospyroides baill and *Annona squamosa* (Annonaceae). *International Journal of Experimental Botany* 53: 53–55.
- Purgatto E, Lajolo FM, Oliveira do Nascimento Jão R, Cordenunsi BR (2001). Inhibition of  $\beta$ -amylase activity, starch degradation and sucrose formation by indole-3-acetic acid during banana ripening. *Planta* 212: 823–828. DOI 10.1007/s004250000441.
- Qian X, Sun Y, Zhou G, Yuan Y, Li J, Huang H, Xu L, Li L (2019). Single-molecule real-time transcript sequencing identified flowering regulatory genes in *Crocus sativus*. *BMC Genomics* 20: 21. DOI 10.1186/s12864-019-6200-5.
- Radchuk VV, Borisjuk L, Sreenivasulu N, Merx K, Mock HP, Rolletschek H, Wobus U, Weschke W (2009). Spatiotemporal profiling of starch biosynthesis and degradation in the developing barley grain. *Plant Physiology* 150: 190–204. DOI 10.1104/pp.108.133520.
- Rhoads A, Au KF (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* 13: 278–289. DOI 10.1016/j.gpb.2015.08.002.
- Rolland F, Baena-Gonzalez E, Sheen J (2006). Sugar sensing and signaling in plants: Conserved and novel mechanisms. *Annual Review of Plant Biology* 57: 675–709. DOI 10.1146/annurev.arplant.57.032905.105441.
- Saleem MTS, Christina AJM, Chidambaranathan N, Ravi V, Gauthanam K (2008). Hepatoprotective activity of *Annona squamosa* (Linn) on experimental animal model. *International Journal of Applied Research in Natural Products* 1: 1–7.
- Salman HA, Senthilkumar R (2015). Antimicrobial activity of *Annona squamosa* L. and *Annona reticulata* L. against clinical isolates of mutants streptococci the causative agents of dental caries. *Asian Journal of Pharmaceutical and Clinical Research* 8: 152–155.
- Salmela L, Rivals E (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30: 3506–3514. DOI 10.1093/bioinformatics/btu538.
- Saleem-Dar M, Oak P, Chidley H, Deshpande A, Giri A, Gupta V (2016). Nutrient and flavor content of mango (*Mangifera indica* L.) cultivars: An appurtenance to the list of staple foods. In: Simmonds MSJ, Preedy VR (eds.), *Nutritional composition of fruit cultivars*, Chap. 19. (Switzerland), 445–468. DOI 10.1016/B978-0-12-408117-8.00019-2.
- Schemberger MO, Stroka MA, Reis L, de Souza Los KK, de Araujo GAT, Sfeir MZT, Galvão CW, Etto RM, Baptistão ARG, Ayub RA (2020). Transcriptome profiling of non-climacteric 'yellow' melon during ripening: Insights on sugar metabolism. *BMC Genomics* 21: 1. DOI 10.1186/s12864-020-6667-0.
- Shen Y, Zhou Z, Wang Z, Li W, Fang C, Wu M, Ma Y, Liu T, Kong LA, Peng DL, Tian Z (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26: 996–1008. DOI 10.1105/tpc.114.122739.
- Soares CA, Peroni-Okita FHG, Cardoso MB, Shitakubo R, Lajolo FM, Cordenunsi BR (2011). Plantain and banana starches: Granule structural characteristics explain the differences in their starch degradation patterns. *Journal of Agricultural and Food Chemistry* 59: 6672–6681. DOI 10.1021/jf201590h.
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y (2013). Utilizing sequence intrinsic composition to classify

- protein-coding and long non-coding transcripts. *Nucleic Acids Research* **41**: e166. DOI 10.1093/nar/gkt646.
- Tang GQ, Sturm A (1999). Antisense repression of sucrose synthase in carrot (*Daucus carota* L.) affects growth rather than sucrose partitioning. *Plant Molecular Biology* **41**: 465–479. DOI 10.1023/A:1006327606696.
- Thammawong M, Arakawa O (2010). Starch to sugar conversion in "Tsgaru" apples under ethylene and 1-methylcyclopropene treatments. *Journal of Agricultural Science and Technology* **12**: 617–626.
- Tian Y, Wen H, Qi X, Zhang X, Liu S, Li B, Sun Y, Li J, He F, Yang W, Li Y (2019). Characterization of full-length transcriptome sequences and splice variants of *Lateolabrax maculatus* by single-molecule long-read sequencing and their involvement in salinity regulation. *Frontiers in Genetics* **10**: 11706. DOI 10.3389/fgene.2019.01126.
- Vimolmangkang S, Zheng H, Peng Q, Jiang Q, Wang H, Fang T, Liao L, Wang L, He H, Han Y (2016). Assessment of sugar components and genes involved in the regulation of sucrose accumulation in peach fruit. *Journal of agricultural and food chemistry* **64**: 6723–6729. DOI 10.1021/acs.jafc.6b02159.
- Wang Y, Chen J, Feng J, Qin Q, Huang J (2015). Overexpression of a loquat (*Eriobotrya japonica* Lindl.) vacuolar invertase affects sucrose levels and growth. *Plant Cell, Tissue and Organ Culture* **123**: 99–108. DOI 10.1007/s11240-015-0817-0.
- Wang LF, Qi XX, Huang XS, Xu LL, Jin C, Wu J, Zhang SL (2016). Overexpression of sucrose transporter gene PbSUT2 from *Pyrus bretschneideri*, enhances sucrose content in *Solanum lycopersicum* fruit. *Plant Physiology and Biochemistry* **105**: 150–161. DOI 10.1016/j.plaphy.2016.04.019.
- Wilke A, Harrison T, Wilkening J, Field D, Glass EM, Kyrpides N, Mavrommatis K, Meyer F (2012). The M5nr: A novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* **13**: 141. DOI 10.1186/1471-2105-13-141.
- Xiao YY, Kuang JF, Qi XN, Ye YJ, Wu ZX, Chen JY, Lu WJ (2018). A comprehensive investigation of starch degradation process and identification of a transcriptional activator MabHLH6 during banana fruit ripening. *Plant Biotechnology Journal* **16**: 151–164. DOI 10.1111/pbi.12756.
- Yatav M, Harary I, Wolf S (2010). Sucrose accumulation in watermelon fruits: Genetic variation and biochemical analysis. *Journal of Plant Physiology* **167**: 589–596. DOI 10.1016/j.jplph.2009.11.009.
- Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, Xu H, Huang X, Li S, Zhou A, Zhang X, Bolund L, Chen Q, Wang J, Yang H, Fang L, Shi C (2018). WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Research* **46**: W71–W75. DOI 10.1093/nar/gky400.
- Yu JG, Luo XZ, Sun L, Li DY, Huang WH, Liu CY (2005). Chemical constituents from the seeds of *Annona squamosa*. *Acta Pharmaceutica Sinica* **40**: 153–158.
- Yue J, Wang R, Ma X, Liu J, Lu X, Thakar SB, An N, Liu J, Xia E, Liu Y (2020). Full-length transcriptome sequencing provides insights into the evolution of apocarotenoid biosynthesis in *Crocus sativus*. *Computational and Structural Biotechnology Journal* **18**: 774–783. DOI 10.1016/j.csbj.2020.03.022.
- Zahid M, Arif M, Rahman A, Singh K, Mujahid M (2018b). Solvent extraction and gas chromatography-mass spectrometry analysis of *Annona squamosa* L. seeds for determination of bioactives, fatty acid/fatty oil composition, and antioxidant activity. *Journal of Dietary Supplements* **15**: 613–623. DOI 10.1080/19390211.2017.1366388.
- Zahid M, Mujahid M, Singh PK, Farooqui S, Singh K, Parveen S, Arif M (2018a). *Annona squamosa* Linn. (custard apple): An aromatic medicinal plant fruit with immense nutraceutical and therapeutic potentials. *International Journal of Pharmaceutical Sciences and Research* **3**: 1745–1759.
- Zrenner R, Salanoubat M, Willmitzer L, Sonnewald U (1995). Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum* L.). *Plant Journal* **7**: 97–107. DOI 10.1046/j.1365-313X.1995.07010097.x.