

RESEARCH ARTICLE

Open Access



Characterization of full-length transcriptome in *Saccharum officinarum* and molecular insights into tiller development

Haifeng Yan^{1†}, Huiwen Zhou^{1†}, Hanmin Luo¹, Yegeng Fan¹, Zhongfeng Zhou¹, Rongfa Chen¹, Ting Luo¹, Xujuan Li², Xinlong Liu², Yangrui Li¹, Lihang Qiu^{1*} and Jianming Wu^{1*}

Abstract

Background: Although extensive breeding efforts are ongoing in sugarcane (*Saccharum officinarum* L.), the average yield is far below the theoretical potential. Tillering is an important component of sugarcane yield, however, the molecular mechanism underlying tiller development is still elusive. The limited genomic data in sugarcane, particularly due to its complex and large genome, has hindered in-depth molecular studies.

Results: Herein, we generated full-length (FL) transcriptome from developing leaf and tiller bud samples based on PacBio Iso-Seq. In addition, we performed RNA-seq from tiller bud samples at three developmental stages (T0, T1 and T2) to uncover key genes and biological pathways involved in sugarcane tiller development. In total, 30,360 and 20,088 high-quality non-redundant isoforms were identified in leaf and tiller bud samples, respectively, representing 41,109 unique isoforms in sugarcane. Likewise, we identified 1063 and 1037 alternative splicing events identified in leaf and tiller bud samples, respectively. We predicted the presence of coding sequence for 40,343 isoforms, 98% of which was successfully annotated. Comparison with previous FL transcriptomes in sugarcane revealed 2963 unreported isoforms. In addition, we characterized 14,946 SSRs from 11,700 transcripts and 310 lncRNAs. By integrating RNA-seq with the FL transcriptome, 468 and 57 differentially expressed genes (DEG) were identified in T1vsT0 and T2vsT0, respectively. Strong up-regulation of several pyruvate phosphate dikinase and phosphoenolpyruvate carboxylase genes suggests enhanced carbon fixation and protein synthesis to facilitate tiller growth. Similarly, up-regulation of linoleate 9S-lipoxygenase and lipoxygenase genes in the linoleic acid metabolism pathway suggests high synthesis of key oxylipins involved in tiller growth and development.

Conclusions: Collectively, we have enriched the genomic data available in sugarcane and provided candidate genes for manipulating tiller formation and development, towards productivity enhancement in sugarcane.

Keywords: Crop productivity, Genomic data, C₄ plant, Carbon fixation, Linoleic acid, Gene expression

* Correspondence: qiulihang2017@126.com; wujianming2004@126.com

† Haifeng Yan and Huiwen Zhou contributed equally to this work.

¹Sugarcane Research Institute of Guangxi Academy of Agricultural Sciences, Guangxi Key Laboratory of Sugarcane Genetic Improvement, and Key Laboratory of Sugarcane Biotechnology and Genetic Improvement (Guangxi), Ministry of Agriculture, East Daxue Road 172, Nanning 530004, Guangxi, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Sugarcane (*Saccharum officinarum* L.) is an important economic crop of the grass family. It is cultivated in the tropic and subtropic regions and represents the main source of world's sucrose [1]. Biofuel, fiber, fertilizer and several other byproducts are also derived from sugarcane production [2]. It is estimated that more than 45 million farmers are involved in the sugarcane production worldwide and nearly 2 billion tonne of sugarcane were produced in 2019 [1]. Despite an extensive breeding, worldwide sugarcane average yield (84 t/ha) is far below the theoretical potential (384 t/ha), therefore, considerable efforts are still needed to increase the crop productivity [3].

Sugarcane can be harvested for many years and the plant is mainly composed of stalks, which are derived from tillers. Similar to other grasses, a single sugarcane plant can produce multiple stalks [4]. Tillering is the sprouting of lateral buds, which can subsequently develop into mature stalks, therefore it is an important component of sugarcane yield. Elucidating the molecular mechanism of tiller development is critical for sugarcane productivity.

Tillering has early catalyzed attention of researchers. In rice, the cloning and functional identification of the gene *MOC1* has marked a breakthrough in the tillering regulation mechanism [5]. Later on, several genes such as *MOC3/TAB1/SRT1*, *LAX1*, *LAX2*, *FON1*, *SLR1* and *TAD1* have been reported to coordinately interact with *MOC1* for tiller formation and development in rice [5–13]. It has been reported that *TaD27-B* gene controls tiller number in hexaploid wheat by regulating strigolactone content [14]. In maize, a complex gene regulatory network involving *tb1*, *gt1*, *tru1*, *sugary1* and *tin1* controls tiller development [15–19]. Collectively, it has become evident that at the genome level, tillering is a multigenic trait underlined by fine coordination of the expression of many genes involved in various biological pathways such as cell cycle, growth and development and phytohormone signaling [20, 21]. In sugarcane, no specific gene controlling tiller development has been documented so far. More importantly, the molecular mechanism of tiller development in sugarcane is still elusive.

Genomics-assisted breeding has become a revolutionary strategy for crop improvement and has high potentials for sugarcane improvement [22]. However, it requires high-quality genomic resources such as complete genome sequences, re-sequencing and gene expression data. Sugarcane modern cultivars (*Saccharum spp.*, $2n = 100\text{--}120$) are heteropolyploid with a large (~10 Gb) and highly complex genome [23]. Several initiatives have been launched to generate genomics resources in sugarcane, particularly, for developing a high-quality reference genome. This resulted in the release of several genome sequences, most of them being of low-quality,

fragmented and incomplete [24–29]. In addition, several transcriptome data based on next generation RNA-seq technologies were generated in sugarcane, providing important gene reservoirs for functional studies [3]. Recently, third generation sequencing technologies such as single molecule real-time (SMRT) sequencing developed by PacBio (Pacific Biosciences of California, Menlo Park, CA, USA) are used for transcriptome studies [30]. SMRT generates high quality full-length (FL) transcripts and facilitates the identification of isoforms and repeat regions in the genome. For species without a good reference genome sequence, SMRT could further enrich the available genomic resources through the reconstruction of the coding genome [31, 32].

The present study aimed at reconstructing the FL transcriptome of sugarcane and investigating the molecular basis of tiller development. To achieve this objective, we prepared two separate PacBio Iso-Seq libraries from developing leaf and axillary tiller bud tissues of sugarcane seedlings to reconstruct and characterize the FL transcriptome. In addition, we performed RNA-seq from axillary tiller bud tissues at three developmental stages to uncover differentially expressed genes and biological pathways underlying tiller development in sugarcane.

Results

Construction and annotation of *S. officinarum* full-length transcriptome

We constructed two SMRT libraries for leaf samples (F01) and tiller bud samples (F02). Libraries were sequenced each with 3 cells, yielding 19,7 and 23,63 Gb clean data for F01 and F02, respectively. A total of 570,055 and 357,140 CCS were identified in F01 and F02, respectively, and classified as FL based on the presence of 5' primers, 3' primers and poly(A) tail (Table 1). The distribution of transcript lengths ranged from 150 to 8000 and 200 to 10,000 bp in F01 and F02, respectively (Fig. 1a, b). After, polishing using RNA-Seq reads, clustering and demultiplexing of FL transcripts, 30,360 and 20,088 high-quality non-redundant FLNC were identified in F01 and F02, respectively (Table 2). We identified 1063 and 1037 AS events in F01 and F02, respectively (Table S1; S2). By merging FLNC transcript lists from the two libraries, we identified 41,109 unique FLNC transcripts in *S. officinarum* spanning 91,227,518 bp. A total of 40,343 CDSs were predicted in the FL transcriptome with length ranging from 100 to 2500 bp (Fig. 1c). Functional annotation of the FLNC transcripts was conducted using eight different public databases and 39,581 transcripts were successfully annotated in at least one database (Table 3; Table S3). These annotated transcripts were grouped into 3640 gene families (Table S4),

Table 1 PacBio sequencing data statistics

Samples	Library	cDNA size (kb)	CCS Number	Read Bases of CCS (bp)	Mean Read Length of CCS (bp)	Mean Number of Passes
Leaf	F01	1–6	570,055	1,407,018,638	2468	19
Tiller bud	F02	1–6	357,140	934,768,566	2617	25

including 1166 PK, 1134 TF and 324 TR genes (Table S5).

We compared the unique FLNC obtained in this study with previous FL transcriptome sequencing in sugarcane [31, 32]. We observed that 85% of the FLNC identified in this study was conserved among other sugarcane cultivars while 2963 FLNC did not match previous Iso-Seq study of sugarcane (Fig. 1d).

Characterizations of SSRs and lncRNAs

We also examined the presence of SSRs and lncRNAs in the FL transcriptome of *S. officinarum*. We identified 14,

946 SSRs in 11,700 transcripts, dominated by tri- and mono-nucleotide SSR types (Fig. 2a; Table S6). A total of 2535 genes contained more than 1 SSR and 879 compound SSRs were detected. With regard to lncRNAs, we detected 310 lncRNAs conserved among the four tools used for lncRNA prediction (CPC, CNCI, Pfam, and CPAT) (Fig. 2b; Table S7).

Short-reads transcriptome

RNAs extracted from tiller bud samples collected at T0, T1 and T2 in three biological replicates, were used for RNA-seq analysis based on the Illumina HiSeq Ten X

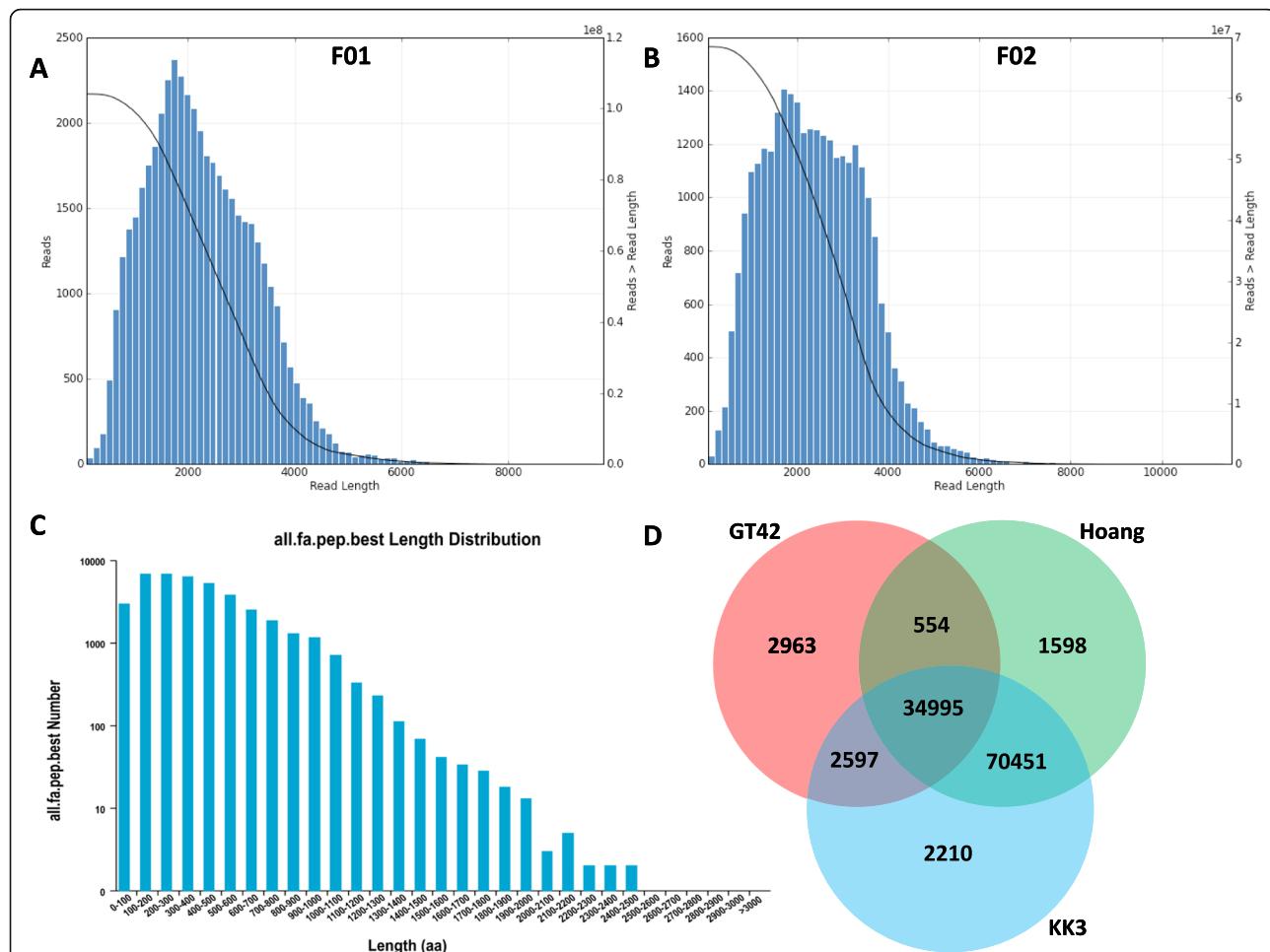


Fig. 1 Overview of full-length transcriptome sequencing using PacBio Sequel platform in *S. officinarum*. **a** The length distribution of reads in F01 representing the leaf library. The horizontal axis represents the length, the vertical axis represents the number of reads within the length range; **b** The length distribution of reads in F02 representing the tiller bud library; **c** CDS length distribution in the two libraries; **d** Venn diagram depicting the shared and specific number of FLNC in three Iso-Seq transcriptome datasets. Hoang, KK3 and GT42 refer to Hoang et al. [31], Piriyapongsa et al. [32] and the current study, respectively

Table 2 Processing of PacBio data and identification of FLNC

Samples	Library	Number of undesired primer reads	Number of undesired poly-A reads	Number of filtered short reads	Number of full-length nonchimeric Reads (FLNC)	Non redundant FLNC
Leaf	F01	61,250	413,783	309	475,033	30,360
Tiller bud	F02	37,436	266,227	190	303,663	20,088

platform. In total, nine RNAs were sequenced, yielding more than 500 million raw short-reads (Table S8). After filtering out low-quality reads, we obtained ~79 Gb clean data representing ~99% of the raw data. On average, the clean reads had a Q30 score of ~93% and a GC content of ~54%, showing a high quality of the sequencing data. The clean reads were mapped to *S. officinarum* FL transcriptome and gene expression was estimated using the FPKM method. We performed a principal component analysis (PCA) based on FPKM data to assess the pattern of clustering of biological replicates and samples. As shown in Fig. 2c, all biological replicates were clustered together, indicating a high correlation between them. In addition, we observed a clear separation of the tiller bud samples collected from the three growth stages. These results suggest significant effects of growth stages on *S. officinarum* tiller bud transcriptome.

Differentially expressed genes (DEG) in tiller bud tissues over growth stages

Gene expression profiles of tiller bud samples at T0 were compared with T1 and T2 in order to identify DEGs. We identified 468 DEGs (369 up- and 99 down-regulated) in T1vsT0 while only 57 DEGs (40 up- and 17 down-regulated) were obtained in T2vsT0. KEGG enrichment analysis of the DEGs in T1vsT0 showed carbon fixation pathways and pyruvate metabolism as the most enriched biological pathways (Fig. 3a). Concerning the

comparison T2vsT0, we observed that carbon fixation pathways, linoleic acid metabolism and pyruvate metabolism were the most enriched KEGG pathways (Fig. 3b).

We randomly selected ten DEGs and quantified their transcript levels using qRT-PCR method (Table S9). The gene *Actin* was used as endogenous control for gene expression normalization. The results showed that transcript levels of all selected genes were significantly altered at T1 and T2 as compared to T0 similarly as observed in the RNA-seq data (Fig. 3c). These observations indicate the reliability of the DEG analysis conducted in this study.

Carbon fixation pathways

Carbon fixation in plants is the process by which inorganic carbon is converted to organic compounds. In this study, 81 DEGs mapped to the carbon fixation pathways were screened out. Interestingly, all these DEGs were annotated as pyruvate phosphate dikinase (*ppdk*) and phosphoenolpyruvate carboxylase (*ppc*). The *ppdks* are involved in the conversion of pyruvate into phosphoenol pyruvate, which is subsequently converted to oxaloacetate by *ppcs* (Fig. 4). Oxaloacetate is then reduced to malate, which represents the input of the Calvin cycle. All the 31 *ppdks* were strongly upregulated at T1 and T2 as compared to T0. Similarly, the 50 *ppcs* identified were all up-regulated at T1 as compared to T0. Altogether, these results suggest that *S. officinarum* tiller development is underlined by a strong synthesis of oxaloacetate.

Table 3 Annotation statistics of the FLNC

Annotation database	Annotated number	300 < =length (bp) < 1000	Length (bp) > =1000
COG	16,596	1370	15,221
GO	33,953	3495	30,448
KEGG	17,198	1897	15,287
KOG	24,191	2071	22,113
Pfam	33,970	3173	30,793
Swissprot	29,225	2695	26,519
eggNOG	38,644	4069	34,557
nr	39,491	4263	35,200
All annotated	39,581	4311	35,242

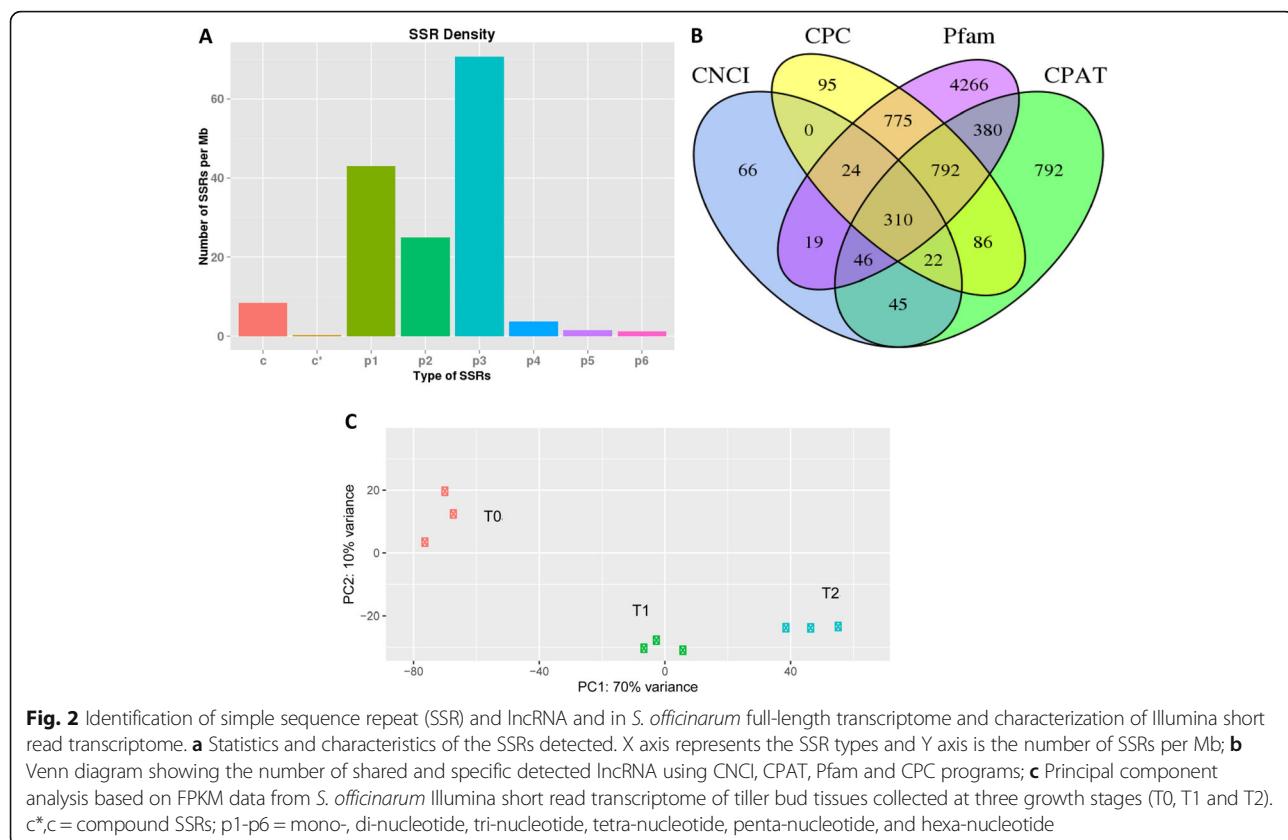


Fig. 2 Identification of simple sequence repeat (SSR) and lncRNA and in *S. officinarum* full-length transcriptome and characterization of Illumina short read transcriptome. **a** Statistics and characteristics of the SSRs detected. X axis represents the SSR types and Y axis is the number of SSRs per Mb; **b** Venn diagram showing the number of shared and specific detected lncRNA using CNCI, CPAT, Pfam and CPC programs; **c** Principal component analysis based on FPKM data from *S. officinarum* Illumina short read transcriptome of tiller bud tissues collected at three growth stages (T0, T1 and T2). $c^*, c = \text{compound SSRs}; p1-p6 = \text{mono-, di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide}$

Linoleic acid metabolism

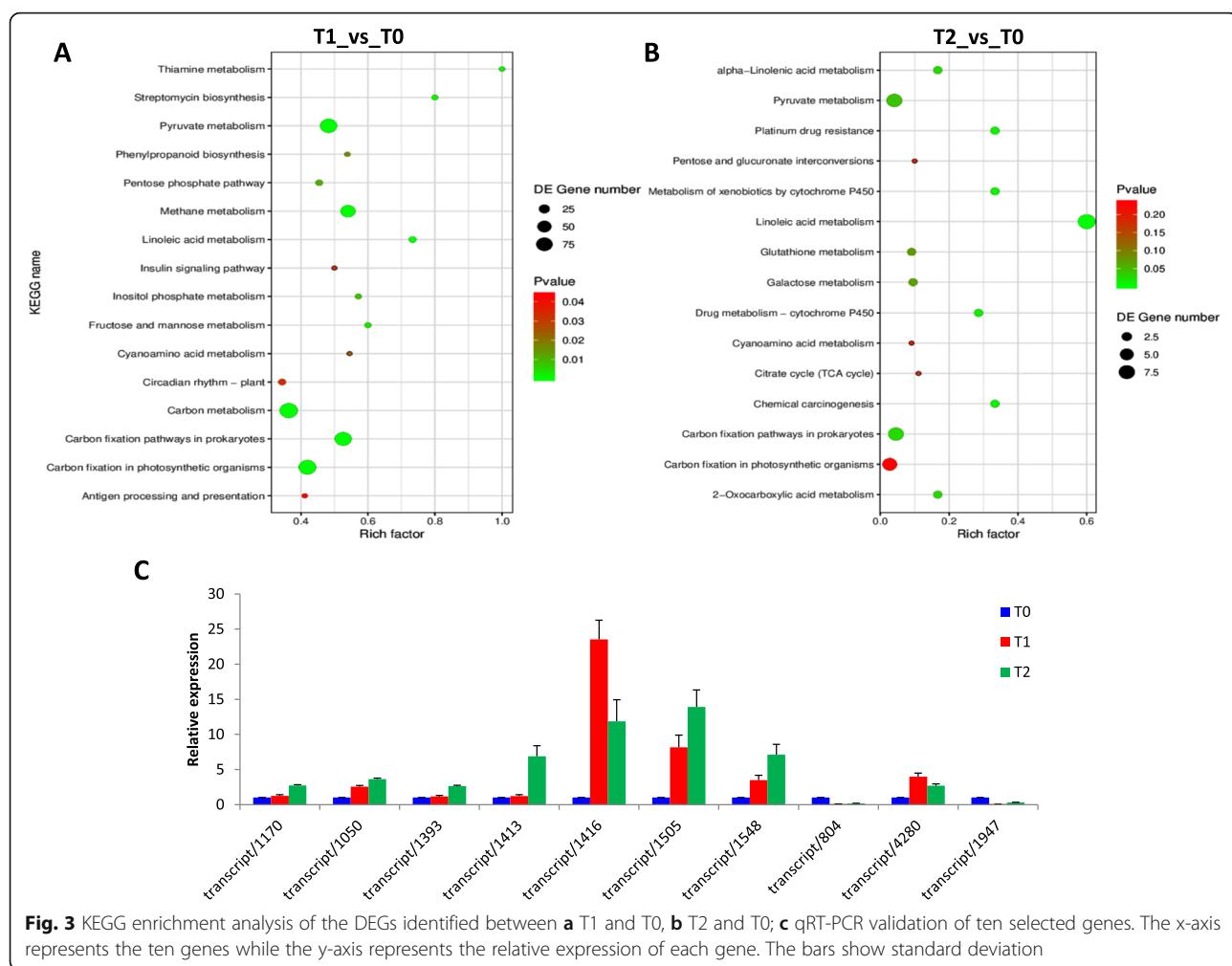
A total of 11 DEGs were mapped to the linoleic acid metabolism pathway, including 10 linoleate 9S-lipoxygenase (*LOX1_5*) and one lipoxygenase (*LOX2_S*) (Fig. 5). *LOX1_5* and *LOX2_S* convert linoleate to (9S)-HPODE and (13S)-HPODE, respectively, which are converted into a large class of oxygenated polyenoic fatty acids called oxylipins [33]. Up-regulation of these genes in T2 and T1 as compared to T0 indicates a mechanism towards high synthesis of oxylipins, which may be crucial for *S. officinarum* tiller growth.

Transcription factors

Plant transcription factors (TFs) regulate numerous physiological programs important for plant growth and development. In this study, we detected nine DEGs encoding TFs belonging to five different families (Table 4). Most of the TFs were differentially expressed between T1 and T0 and data showed that they were up-regulated. Only the gene *transcript/1947* was down-regulated at T2vsT0. We predict that these TFs play vital roles in regulating structural genes involved in *S. officinarum* tiller development.

Discussion

The largest genomic data in sugarcane has been recently released by Souza et al. [28]. It spanned 4.26 Gb long, representing ~ 30% of the estimated genome size and 373,869 gene models were predicted. Hence, it is obvious that a high proportion of gene models are still unreported in sugarcane. Besides, it is well known that more than 95% of genes experience alternative splicing events, leading to multiple isoforms of each gene [34]. Deciphering the landscape of isoforms in sugarcane transcriptome will provide a useful resource for improving gene model prediction and annotation as adopted in various plant species such as in wheat, maize, panax and sorghum [35–38]. In this study, we employed Pacbio Iso-Seq to generate full-length transcriptome from two sugarcane tissues. In total, 30,360 and 20,088 FLNC transcripts were obtained in leaf and tiller tissues, respectively. The variation of the number of FLNC transcripts in both tissues highlights the tissue specific expression pattern of several genes in sugarcane, which could associate with important agronomic traits and thus provide new marker tools for breeding programs [32]. Compared with previous FL transcriptome sequencing in sugarcane [31, 32], we obtained the lowest number of FLNC transcripts. This can be explained by

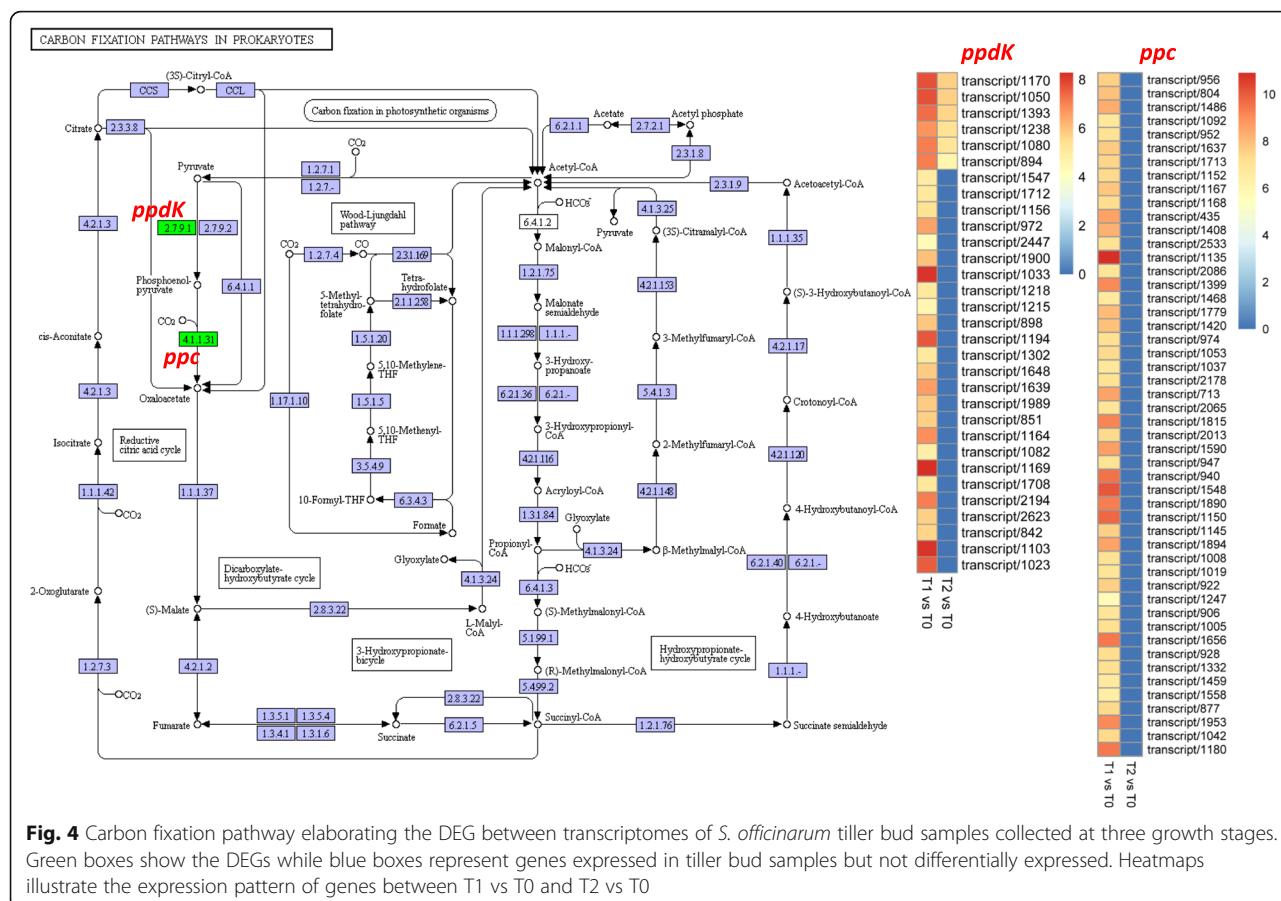


several factors such as the tissues sampled, the genotype, the growth stage and even the PacBio transcriptome sequencing and data processing procedures. For example, Hoang et al. [31], pooled RNA from internode, leaf and root collected at different developmental stages in 22 sugarcane varieties. They reported 107,598 unique transcripts which is more than the double of the number of FLNC transcripts identified in this study. Nonetheless, we obtained the highest number of specific FLNC transcripts in this study. Notably, there is no previous report of axillary tiller bud tissue transcriptome in sugarcane, hence, our results provide new catalog of useful transcripts for improving genome annotation in sugarcane.

Beyond being advantageous to recover full-length transcripts, PacBio Iso-Seq allows identification of lncRNAs and repetitive sequences such as SSRs [39]. Growing evidences demonstrate that lncRNAs are key regulators of gene expression and genome stability in plants and are involved in functions such as vernalization, fertility, photomorphogenesis, phosphate homeostasis, protein re-localization, modulation of chromatin loop dynamics

[40, 41]. We reported here 310 lncRNAs with high confidence, which will be instrumental for further illuminating the complex biology of sugarcane. SSRs are the most widely used molecular markers in plant thanks to their relatively abundance, co-dominance, high polymorphism, easy and low-cost procedure [42]. Importantly, it has been demonstrated that transcriptome based SSRs are linked to functional genes, hence they can be used to study their association with phenotypic variation and the flanking sequences are more likely to be conserved among close or distant species, making their use as markers for comparative mapping easier [43]. Currently, a limited number of sugarcane-specific SSR markers are available [44], hence, the 14,946 SSRs detected in our study, pending validation and screening of the most polymorphic ones, will be very useful for genotyping and genetic diversity studies.

By integrating RNA-seq data from axillary tiller bud at three developmental stages with our FL transcriptome, we investigated the DEGs and enriched pathways associated with tiller development in sugarcane seedlings. We



obtained higher number of DEGs at T1 vs T0 as compared to T2 vs T0, indicating that major transcriptome readjustment occurs in order to initiate tiller bud germination in *S. officinarum* seedlings. Two key biological pathways were found enriched during tiller development in *S. officinarum* seedlings: carbon fixation and linoleic acid metabolism pathways.

Carbon fixation is the process by which CO_2 is incorporated into organic compounds [45], which are used to store energy and as building blocks for other important plant biomolecules. It is well known that enhancing carbon fixation is a viable approach for improving plant growth and biomass accumulation [46–48]. Two key gene families were strongly up-regulated in the carbon fixation pathways during tiller development in sugarcane seedlings: pyruvate phosphate dikinase (*ppdk*) and phosphoenolpyruvate carboxylase (*ppc*). In species such as sugarcane that use C_4 photosynthesis, the actions of these two enzymes lead to synthesis of oxaloacetate, which is reduced to malate, as input of the Calvin cycle [49]. Transgenic plants overexpressing *ppdk* or *ppc* displayed improved carbon assimilation and growth [50–54]. Collectively, we deduced that up-regulation of *ppc* and *ppdk* increases the supply of 4-carbon carboxylic

acids, providing high carbon skeletons to sustain high amino acid and protein synthesis responsible for increased metabolic processes and tiller growth in sugarcane seedlings.

Plant lipoxygenases (*LOX*) oxidize polyunsaturated fatty acids such as linolenic and linoleic acids into fatty acid hydroperoxides, which are converted into oxylipins [33]. *LOX*-derived oxylipins are involved in various physiological processes of plants, including growth and development. For example, Kolomiets et al. [55] demonstrated that *POTLX-1* controls tuber growth and development in potato, probably by initiating the synthesis of oxylipins that regulate cell growth during tuber formation. In *Arabidopsis*, it has been shown that oxylipins produced by the 9-*LOX* pathway regulate lateral root development [56]. Similarly, the *lox3–4* knockout maize mutants displayed reduced root length and plant height compared with the wild type, indicating that *ZmLOX3* is required for normal plant development Gao et al. [57]. The strong upregulation of *LOX* genes in this study indicates a mechanism towards high synthesis of key oxylipins involved in tiller growth and development in sugarcane. Jasmonic acid (JA) is a type of oxylipins and several studies have demonstrated that JA contents affect

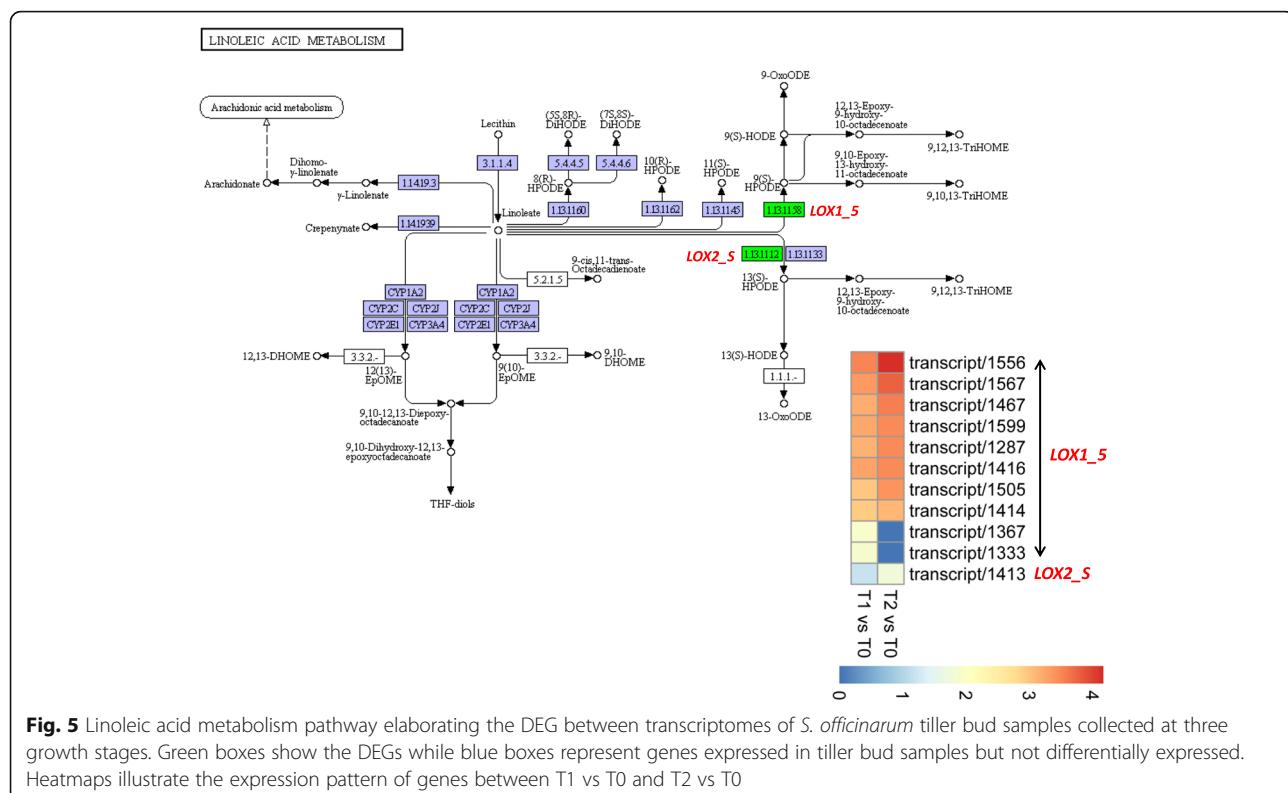


Fig. 5 Linoleic acid metabolism pathway elaborating the DEG between transcriptomes of *S. officinarum* tiller bud samples collected at three growth stages. Green boxes show the DEGs while blue boxes represent genes expressed in tiller bud samples but not differentially expressed. Heatmaps illustrate the expression pattern of genes between T1 vs T0 and T2 vs T0

tillering in grass [58, 59]. Future studies should investigate developing sugarcane tiller bud samples to clarify whether JA or other types of oxylipins promote tiller development.

Conclusions

In this study, we enriched the genomic data available in sugarcane by generating and characterizing the full-length transcriptome from developing leaf and tiller bud tissues. The novel transcripts identified will be useful for ongoing efforts of genome annotation and gene model prediction in sugarcane. By integrating RNA-seq data from developing axillary tiller bud tissues, we identified important genes involved in the carbon fixation and

linoleic acid pathways differentially expressed during tiller development. Further in-depth investigations of these candidate genes will provide prospects for controlling tiller outgrowth and productivity in sugarcane.

Methods

Plant material

In this study, *S. officinarum* L. cultivar Guitang 42 (GT42) was employed as plant material. It is an excellent cultivar with high-yield, high-sugar, lodging-resistance and suitable for mechanized production, developed by Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences, China [60]. Plant material was obtained from Sugarcane Research Institute, Guangxi

Table 4 Transcription factors differentially expressed over *S. officinarum* tiller growth stages

Gene ID	T1vsT0 (Log 2 fold change)	T2vsT0 (Log 2 fold change)	TF family
transcript/1548	10.14	0.00	B3-ARF
transcript/3906	3.85	0.00	B3-ARF
transcript/4280	1.97	0.00	B3-ARF
transcript/313	1.07	0.00	C3H
transcript/2433	1.53	0.00	MYB
transcript/1947	0.00	-1.54	MYB
transcript/3790	2.40	0.00	RWP-RK
transcript/2444	3.27	0.00	RWP-RK
transcript/2533	7.00	0.00	WRKY

Academy of Agricultural Sciences, China. The formal identification of the material was undertaken by the corresponding author of this study Prof Lihang Qiu. No voucher of the plant material has been deposited in a genebank. Healthy stems were selected, cut into single bud segments, then washed with water and soaked for 30 min. Thirty stems with only one bud were used as propagules and planted in pots ($24 \times 19.5 \times 26.5$ cm), with 3 propagules per pot. Each pot was filled with 11 kg of peat nutrient soil, covered with plastic film and placed in a greenhouse located at the experimental field of Sugarcane Research Institute of Guangxi Academy of Agricultural Sciences (latitude: 22.85, longitude: 108.24, altitude 50.17 m) and the microclimate conditions for sugarcane growth in the greenhouse were as follows: average temperature = 36.52 ± 0.76 °C, average relative humidity = $75.66 \pm 2.4\%$ and average light intensity = $1884.1 \pm 39.25 \mu\text{mol m}^{-2} \text{s}^{-1}$. When the buds sprouted out of the soil, the plastic film was removed, and water was sprayed in each pot to ensure a normal development. Subsequently, samples of sugarcane leaf tissues and tiller buds were taken at different growth stages, including T0 (2–3 leaves stage; establishment stage), T1 (4–5 leaves stage; beginning of tillering stage), and T2 (6–7 leaves stage; full tillering stage). For leaf samples, the first true leaf was collected from the three plants of each pot and mixed to form a biological replicate. For the tiller bud samples, at the T0 the axillary tiller buds start to germinate belowground. At T1, the axillary tiller buds reach about 0.5 cm–1 cm length belowground. At T2, the axillary tiller buds emerge from the soil and the length reaches about 2 cm. At each stage, axillary tiller buds were collected from the three plants of each pot and mixed to form a biological replicate. Samples were collected in three biological replicates from plants grown in different pots and quickly frozen with liquid nitrogen and stored at –80 °C.

Library construction and single-molecule real-time (SMRT) sequencing

Total RNA was extracted by grinding mixed leaf samples collected at the three growth stages and mixed tiller bud samples from the three growth stages, separately, in TRIzol reagent (Life technologies) following the manufacturer's protocol and two libraries were constructed. The integrity of the RNAs was determined with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and agarose gel (1%) electrophoresis. The purity and concentration of the RNAs were determined with the Nanodrop micro-spectrophotometer (Thermo fisher Scientific, Waltham, MA, USA). 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 (Clontech Laboratories, USA). 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. Moreover, 1–6 kb size selection was performed using the Blue Pippin TM Size-Selection System. Then, large-scale PCR was performed for the next SMRTbell library construction. cDNAs were repaired for DNA damage and ligated to sequencing adapters. The SMRTbell template was annealed to sequencing primer, bound to polymerase, and sequenced on the PacBio Sequel platform using P6-C4 chemistry with 10 h movies by Biomarker Technology Co. (Beijing, China).

PacBio long-read processing

PacBio data processing followed standard procedures established by Biomarker Technology Co. (Beijing, China). The raw sequencing reads of cDNA libraries were classified and clustered into transcript consensus using the SMRT Link v5.0.1 pipeline [61]. First, circular consensus sequence (CCS) reads were extracted out of subreads BAM file. Next, 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 from the analysis. Subsequently, the full-length non-chimeric reads were clustered by Iterative Clustering for Error Correction software to generate the cluster consensus isoforms. To improve accuracy of PacBio reads, two strategies were employed. First, the nFL reads were used to polish the above obtained cluster consensus isoforms by Quiver software [62] to obtain the FL polished high quality consensus sequences (accuracy ≥90%). Next, the low quality isoforms were further corrected using Illumina short reads by using the LoRDEC tool (version 0.8) [63]. Then, the final transcriptome isoform sequences were filtered by removing the redundant sequences with software CD-HIT-v4.6.7 [64] using an identity threshold of 0.99.

Annotation of genes and identification of transcription factor (TF), transcriptional regulators (TR) and protein kinases (PK)

We used TransDecoder software (<https://github.com/TransDecoder/TransDecoder/wiki>) for the coding sequence (CDS) prediction. Next, DNA or protein sequences of the FLNC transcripts were submitted to various public databases for functional annotation: clusters of orthologous groups/eukaryotic orthologous groups (COG/KOG) database (<http://www.ncbi.nlm.nih.gov/COG>), Gene Ontology (GO), NCBI non-redundant protein (Nr), Swiss-Prot protein (Swissprot) database (<http://www.expasy.ch/sprot>), Kyoto Encyclopedia of

Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg>), evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) database (<http://eggnog5.embl.de/>) and Pfam protein families (Pfam) database (<https://pfam.xfam.org/>). To identify plant transcription factors (TFs), transcriptional regulators (TRs) and protein kinases (PKs), transcripts were submitted to Plant Transcription Factor database (PlantTFdb: <http://planttfdb.gao-lab.org/>, [65]) and iTAK (v.1.5, [66]) with the best match result [67].

Simple sequence repeats (SSR) prediction

MIcroSATellite (MISA, <http://pgrc.ipk-gatersleben.de/misa/>) is a package that identifies seven types of simple sequence repeat (SSR): mono- nucleotide, di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide, hexa-nucleotide and compound SSR (hybrid microsatellite). We screened transcripts with length > 500 bp using the following parameters: unit_size_min_repeats: 2–6, 3–5, 4–4, 5–4, 6–4; compounds (max_difference_between_2_SSRS): 100.

Characterization of long non-coding RNAs (lncRNAs)

Four tools were used to predict the lncRNAs: CNCI (v.2) [68], CPC (v.0.9-r2) [69]), Pfam protein families (Pfam) database (<https://pfam.xfam.org/>) and CPAT (v.3.0.0) [70]. In this study, we kept the commonly detected lncRNAs between the four tools as the most probable lncRNAs.

Alternative splicing (AS) detection

AS transcripts were predicted following methods described by Pan *et al.* [71]. All sequences were run all-vs-all BLAST with high identity settings [72] against the assembled FL transcriptome. BLAST alignments that met all criteria were considered products of candidate AS events: (1) both sequence (query and subject) lengths exceeded 1000 bp and the alignment contained 2 high-scoring segment pairs (HSPs); (2) the alternative splicing gap exceeded 100 bp and was located ≥ 100 bp from the 3'/5' end; and (3) a 5-bp overlap was allowed for all alternative transcripts.

RNA-Seq library construction and sequencing

High-quality RNAs were extracted from tiller buds collected at the three growth stages in triplicate using TRI-ZOL® reagent (TIANGEN, Beijing, China) according to the manufacturer's protocol. Extracted RNA from different samples was purified by using RNase-free DNase I (TaKaRa, Kyoto, Japan) to remove the genomic DNA contamination. cDNA library construction and sequencing on Illumina HiSeq Ten X platform were conducted following standard procedures established by Biomarker

Technology Co. (Beijing, China) and fully described by Li *et al.* [73].

Gene expression quantification and differentially expressed genes (DEG) analysis

Raw RNA-Seq data were processed by the fastQC v0.11.2. Paired-end clean reads were aligned to the reference FL transcriptome using Bowtie2 [74] with default parameters. Gene expression was estimated by the number of fragments per kilobase of the transcript sequence per million base pairs sequenced (FPKM) method. The differential expression analysis was performed using the DESeq2 package. Genes with fold change ≥ 2 and a false discovery rate (FDR) < 0.05 were considered as significantly differentially expressed genes (DEGs) in comparative analysis. DEGs were subjected to enrichment analysis of KEGG [75] and COG [76] pathways. To validate DEG analysis, ten genes were selected and their expression levels were validated based on qRT-PCR approach in three biological and three technical replicates following descriptions of Dossa *et al.* [77].

Abbreviations

DEG: Differentially expressed genes; FL: Full length; FLNC: Full length non coding; PCA: Principal component analysis; FPKM: Fragments Per Kilobase of Transcript per Million Fragments Mapped; TF: Transcription factor; lncRNA: Long non coding RNA; SSR: Simple sequence repeat

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-021-02989-5>.

Additional file 1: Table S1. Alternative splicing events and isoforms identified in *S. officinarum* full-length transcriptome based on leaf sample library. **Table S2.** Alternative splicing events and isoforms identified in *S. officinarum* full-length transcriptome based on tiller bud sample library. **Table S3.** Gene annotation statistics. **Table S4.** Gene family statistics. **Table S5.** List of genes annotated as transcription factors (TF), transcriptional regulators (TRs) and protein kinases (PK). **Table S6.** Characterization of the simple sequence repeats in *S. officinarum* full-length transcriptome. **Table S7.** Characterization of lncRNAs in *S. officinarum* full-length transcriptome. LncRNAs highlighted in red are those commonly identified by the four software. **Table S8.** Statistics of Illumina RNA-seq in *S. officinarum* tiller bud over three growth stages. **Table S9.** Primer sequences of genes used for qRT-PCR validation of differentially expressed genes.

Acknowledgements

Not applicable.

Authors' contributions

H. Y., H. Z., H. L. and Y. F. collected plant materials and performed the experiments; Z. Z., R. C., T. L., Xu. L., Xi. L. and Y. L. carried out the bioinformatic analyses; H. Y. and H. Z. wrote the paper; L. Q. and J. W. designed the study, guided the research and provided funding. All authors have read and approved the final version of this manuscript.

Funding

This work was funded by National Key Research and Development Program of China (2019YFD1000503), National Natural Science Foundation of China (31701363, 31960416 and 31360312), Guangxi Natural Science Foundation Project (2018GXNSFAA138149, 2018GXNSFDA294004, 2016GXNSFBA380034 and 2015GXNSFDA139011), Guangxi Science and Technology Project (Guike AD19245080), Guangxi Academy of Agricultural Sciences Fund Project

(2020YM22, 31701363, 2018YM02 and 2018YT01). The Funder has no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The RNA-seq data has been submitted to NCBI SRA: PRJNA723212 and accessible at www.ncbi.nlm.nih.gov/sra/PRJNA723212.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Author details

¹Sugarcane Research Institute of Guangxi Academy of Agricultural Sciences, Guangxi Key Laboratory of Sugarcane Genetic Improvement, and Key Laboratory of Sugarcane Biotechnology and Genetic Improvement (Guangxi), Ministry of Agriculture, East Daxue Road 172, Nanning 530004, Guangxi, China. ²Sugarcane Research Institute of Yunnan Academy of Agricultural Sciences, East Lingquan Road 172, Kaiyun 661600, Yunnan, China.

Received: 28 October 2020 Accepted: 27 April 2021

Published online: 22 May 2021

References

1. Faostat 2018. Accessible at : <http://www.fao.org/faostat/en/#home>. Consulted on 25/08/2020.
2. Wei Y-A, Li Y-R. Status and trends of sugar industry in China. Sugar Tech. 2006;8(1):10–5. <https://doi.org/10.1007/BF02943735>.
3. Diniz AL, Ferreira SS, Caten FT, Margarido GRA, Dos Santos JM, Barbosa GV, et al. Genomic resources for energy cane breeding in the post genomics era. Comput Struct Biotechnol J. 2019;10:1404–14.
4. King NJ, Mungomery RW, Hughes SG. Manual of cane growing. New York: American Elsevier Publishing Company, Inc; 1965. p. 375.
5. Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, et al. Control of tillering in rice. Nature. 2003;422(6932):618–21. <https://doi.org/10.1038/nature01518>.
6. Liao Z, Yu H, Duan J, Yuan K, Yu C, Meng X, et al. SLR1 inhibits MOC1 degradation to coordinate tiller number and plant height in rice. Nat Commun. 2019;10(1):2738. <https://doi.org/10.1038/s41467-019-10667-2>.
7. Lu Z, Shao G, Xiong J, Jiao Y, Wang J, Liu G, et al. MONOCULM 3, an ortholog of *WUSCHEL* in rice, is required for tiller bud formation. J. Genet Genomics. 2015;42:71–8.
8. Mjomba F, Zheng Y, Liu H, Tang W, Hong Z, Wang F, et al. Homeobox is pivotal for *OsWUS* controlling tiller development and female fertility in rice. G3. 2016;6:2013–21.
9. Oikawa T, Kyozuka J. Two-step regulation of *LAX PANICLE1* protein accumulation in axillary meristem formation in rice. Plant Cell. 2009;21(4): 1095–108. <https://doi.org/10.1105/tpc.108.065425>.
10. Shao G, Lu Z, Xiong J, Wang B, Jing Y, Meng X, et al. Tiller bud formation regulators MOC1 and MOC3 cooperatively promote tiller bud outgrowth by activating FON1 expression in Rice. Mol Plant. 2019;12(8):1090–102. <https://doi.org/10.1016/j.molp.2019.04.008>.
11. Tabuchi H, Zhang Y, Hattori S, Omae M, Shimizu-Sato S, Oikawa T, et al. *LAX PANICLE2* of rice encodes a novel nuclear protein and regulates the formation of axillary meristems. Plant Cell. 2011;23(9):3276–87. <https://doi.org/10.1105/tpc.111.088765>.
12. Tanaka W, Ohmori Y, Ushijima T, Matsusaka H, Matsushita T, Kumamaru T, et al. Axillary meristem formation in rice requires the *WUSCHEL* ortholog *TILLERS ABSENT1*. Plant Cell. 2015;27(4):1173–84. <https://doi.org/10.1105/tpc.5.00074>.
13. Xu C, Wang Y, Yu Y, Duan J, Liao Z, Xiong G, et al. Degradation of MONOCULM 1 by APC/C (TAD1) regulates rice tillering. Nat Commun. 2012; 3(1):750. <https://doi.org/10.1038/ncomms1743>.
14. Zhao B, Wu TT, Ma SS, Jiang DJ, Bie XM, Sui N, et al. *TaD27-B* gene controls the tiller number in hexaploid wheat. Plant Biotechnol J. 2020;18(2):513–25. <https://doi.org/10.1111/pbi.13220>.
15. Dong ZB, et al. Ideal crop plant architecture is mediated by *tassels replace upper ears1*, a BTB/POZ ankyrin repeat gene directly targeted by TEOSINTE BRANCHED1. Proc Natl Acad Sci U S A. 2017;114(41):E8656–64. <https://doi.org/10.1073/pnas.1714960114>.
16. Kebröm TH, Brutnell TP. Tillering in the *sugary1* sweet corn is maintained by overriding the *teosinte branched1* repressive signal. Plant Signal Behav. 2015; 10(12):e1078954. <https://doi.org/10.1080/15592324.2015.1078954>.
17. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat Genet. 2011; 43(11):1160–4. <https://doi.org/10.1038/ng.942>.
18. Whipple CJ, Kebröm TH, Weber AL, Yang F, Hall D, Meeley R, et al. *grassy tillers1* promotes apical dominance in maize and responds to shade signals in the grasses. Proc Natl Acad Sci U S A. 2011;108(33):E506–12. <https://doi.org/10.1073/pnas.1102819108>.
19. Zhang X, Lin Z, Wang J, Liu H, Zhou L, Zhong S, et al. The *tin1* gene retains the function of promoting tillering in maize. Nat Commun. 2019;10(1):5608. <https://doi.org/10.1038/s41467-019-13425-6>.
20. Domagalska MA, Leyser O. Signal integration in the control of shoot branching. Nat Rev Mol Cell Biol. 2011;12(4):211–21. <https://doi.org/10.1038/nrm3088>.
21. Wang YH, Li JY. Branching in rice. Curr Opin Plant Biol. 2011;14(1):94–9. <https://doi.org/10.1016/j.pbi.2010.11.002>.
22. Kandel R, Yang X, Song J, Wang J. Potentials, challenges, and genetic and genomic resources for sugarcane biomass improvement. Front Plant Sci. 2018;9:151. <https://doi.org/10.3389/fpls.2018.00151>.
23. Piperidis N, D'Hont A. Sugarcane genome architecture decrypted with chromosome-specific oligo probes. Plant J. 2020;103(6):2039–51. <https://doi.org/10.1111/tpj.14881>.
24. de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GMQ, Del Bem LE, Vicentini R, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics. 2014;15:540. <https://doi.org/10.1186/1471-2164-15-540>.
25. Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun. 2018;9. <https://doi.org/10.1038/s41467-018-05051-5>.
26. Nascimento LC, Yanagui K, Jose J, Camargo ELO, Grassi MCB, Cunha CP, et al. Unraveling the complex genome of *Saccharum spontaneum* using Polyploid Gene Assembler. DNA Res. 2019;26:205–16. <https://doi.org/10.1093/dnare/dsz001>.
27. Riaño-Pachón DM, Mattiello L. Draft genome sequencing of the sugarcane hybrid SP80-3280. F1000Res. 2017;6:861. <https://doi.org/10.12688/f1000research.11859.2>.
28. Souza GM, Van Sluys MA, Lemke CG, Lee H, Margarido GRA, Hotta CT, et al. Assembly of the 373K gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. Gigascience. 2019;8(12):giz129. <https://doi.org/10.1093/gigascience/giz129>.
29. Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat Genet. 2018;50:1565–73. <https://doi.org/10.1038/s41588-018-0237-2>.
30. Wang B, Kumar V, Olson A, Ware D. Reviving the Transcriptome studies: an insight into the emergence of single-molecule Transcriptome sequencing. Front Genet. 2019;10:384. <https://doi.org/10.3389/fgene.2019.00384>.
31. Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, et al. 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. 2017;18(1):395. <https://doi.org/10.1186/s12864-017-3757-8>.
32. Piriapongsa J, Kaewprommal P, Vaiwiri S, Anuntakrun S, Wirojirasarak W, Punpee P, et al. Uncovering full-length transcript isoforms of sugarcane cultivar Khon Kaen 3 using single-molecule long-read sequencing. PeerJ. 2018;6:e5818. <https://doi.org/10.7717/peerj.5818>.
33. Porta H, Rocha-Sosa M. Plant lipoxygenases. Physiological and molecular features. Plant Physiol. 2002;130(1):15–21. <https://doi.org/10.1104/pp.010787>.
34. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40(12):1413–5. <https://doi.org/10.1038/ng.259>.
35. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, et al. A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun. 2016;7(1):11706. <https://doi.org/10.1038/ncomms11706>.

36. Dong L, Liu H, Zhang J, Yang S, Kong G, Chu JS, et al. Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genomics*. 2015;16(1):1039. <https://doi.org/10.1186/s12864-015-2257-y>.
37. Jo IH, Lee J, Hong CE, Lee DJ, Bae W, Park SG, et al. Isoform sequencing provides a more comprehensive view of the Panax ginseng Transcriptome. *Genes*. 2017;8(9):228. <https://doi.org/10.3390/genes8090228>.
38. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun*. 2016;7(1):11708. <https://doi.org/10.1038/ncomms11708>.
39. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. 2015;13(5):278–89. <https://doi.org/10.1016/j.gpb.2015.08.002>.
40. Chekanova JA. Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol*. 2015;27:207–16. <https://doi.org/10.1016/j.pbi.2015.08.003>.
41. Liu X, Hao L, Li D, Zhu L, Hu S. Long non-coding RNAs and their biological roles in plants. *Genomics, Proteomics Bioinformatics*. 2015;13(3):137–47. <https://doi.org/10.1016/j.gpb.2015.02.003>.
42. Dossa K, Yu J, Liao B, Cisse N, Zhang X. Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: SisatBase. *Front Plant Sci*. 2017; 8:1470. <https://doi.org/10.3389/fpls.2017.01470>.
43. Tulsiani NJ, Hamid R, Jacob F, Umretiya NG, Nandha AK, Tomar RS, et al. Transcriptome landscaping for gene mining and SSR marker development in coriander (*Coriandrum sativum* L.). *Genomics*. 2020;112(2):1545–53. <https://doi.org/10.1016/j.ygeno.2019.09.004>.
44. Pan Y. Development and integration of an SSR-based molecular identity database into sugarcane breeding program. *Agronomy*. 2016;6(2):28. <https://doi.org/10.3390/agronomy6020028>.
45. Bar-Even A, Noor E, Lewis NE, Milo R. Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci*. 2010;107(19):8889–94. <https://doi.org/10.1073/pnas.0907176107>.
46. Bar-Even A. Daring metabolic designs for enhanced plant carbon fixation. *Plant Sci*. 2018;273:71–83. <https://doi.org/10.1016/j.plantsci.2017.12.007>.
47. Betti M, Bauwe H, Busch FA, Fernie AR, Keech O, Levey M, et al. Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement. *J Exp Bot*. 2016;67(10): 2977–88. <https://doi.org/10.1093/jxb/erw076>.
48. Ducat DC, Silver PA. Improving carbon fixation pathways. *Curr Opin Chem Biol*. 2012;16(3–4):337–44. <https://doi.org/10.1016/j.cbpa.2012.05.002>.
49. Brown RH, Byrd GT, Bouton JH, Bassett CL. Photosynthetic characteristics of segregates from hybrids between *Flaveria brownii* (C4-like) and *Flaveria linearis* (C3-C4). *Plant Physiol*. 1993;101:825–31.
50. Giuliani R, Karki S, Covshoff S, Lin HC, Coe RA, Koteyeva NK, et al. Transgenic maize phosphoenolpyruvate carboxylase alters leaf-atmosphere CO₂ and ¹³CO₂ exchanges in *Oryza sativa*. *Photosynth Res*. 2019;142(2):153–67. <https://doi.org/10.1007/s11120-019-00655-4>.
51. Sen P, Ghosh S, Sarkar SN, Chanda P, Mukherjee AS, Datta K, et al. Pyramiding of three C₄ specific genes towards yield enhancement in rice. *Plant Cell Tissue Organ Cult*. 2017;128(1):145–60. <https://doi.org/10.1007/s11240-016-1094-2>.
52. Yadav S, Rathore MS, Mishra A. The pyruvate-phosphate Dikinase (C₄-SmPPDK) gene from *Suaeda monoica* enhances photosynthesis, carbon assimilation, and abiotic stress tolerance in a C₃ plant under elevated CO₂ conditions. *Front Plant Sci*. 2020;11:345. <https://doi.org/10.3389/fpls.2020.00345>.
53. Yadav S, Mishra A. Ectopic expression of C₄ photosynthetic pathway genes improves carbon assimilation and alleviate stress tolerance for future climate change. *Physiol Mol Biol Plants*. 2020;26(2):195–209. <https://doi.org/10.1007/s12298-019-00751-8>.
54. Zhang J, Bandopadhyay A, Sellappan K, Wang G, Xie H, Datta K, et al. Characterization of a C₄ maize pyruvate orthophosphate dikinase expressed in C₃ transgenic rice plants. *Afr J Biotechnol*. 2010;9(2):234–42.
55. Kolomiets MV, Hannapel DJ, Chen H, Tymeson M, Gladon RJ. Lipoxygenase is involved in the control of potato tuber development. *Plant Cell*. 2001; 13(3):613–26. <https://doi.org/10.1105/tpc.13.3.613>.
56. Vellosoillo T, Martínez M, López MA, Vicente J, Cascón T, Dolan L, et al. Oxylipins produced by the 9-lipoxygenase pathway in *Arabidopsis* regulate lateral root development and defense responses through a specific signaling cascade. *Plant Cell*. 2007;19(3):831–46. <https://doi.org/10.1105/tpc.106.046052>.
57. Gao X, Starr J, Göbel C, Engelberth J, Feussner I, Tumlinson J, et al. Maize 9-lipoxygenase ZmLOX3 controls development, root-specific expression of defense genes, and resistance to root-knot nematodes. *Mol Plant-Microbe Interact*. 2008;21(1):98–109. <https://doi.org/10.1094/MPMI-21-1-0098>.
58. Liu X, Li F, Tang J, Wang W, Zhang F, Wang G, et al. Activation of the jasmonic acid pathway by depletion of the hydroperoxide lyase OsHPL3 reveals crosstalk between the HPL and AOS branches of the oxylipin pathway in rice. *PLoS One*. 2012;7(11):e50089. <https://doi.org/10.1371/journal.pone.0050089>.
59. Liu R, Finlayson SA. Sorghum tiller bud growth is repressed by contact with the overlying leaf. *Plant Cell Environ*. 2019;42(7):2120–32. <https://doi.org/10.1111/pce.13548>.
60. Wang L, Liao J, Tan F, Tang S, Huang J, Li X, et al. Breeding of new high-yield, high-sugar and lodging-resistant sugarcane variety Guitang 42 and its high-yield cultivation technique. *J Southern Agric*. 2015;46:1361–6.
61. Gordon SP, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, et al. Widespread Polycistronic transcripts in Fungi revealed by single-molecule mRNA sequencing. *PLoS One*. 2015;10(7):e0132628. <https://doi.org/10.1371/journal.pone.0132628>.
62. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, et al. Non hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10:563–9.
63. Salmela L, Rivals E. LORDEC: accurate and efficient long read error correction. *Bioinformatics*. 2014;30(24):3506–14. <https://doi.org/10.1093/bioinformatics/btu538>.
64. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
65. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*. 2017;45(D1):D1040–5. <https://doi.org/10.1093/nar/gkw982>.
66. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, et al. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant*. 2016;9(12):1667–70. <https://doi.org/10.1016/j.molp.2016.09.014>.
67. Yang X, Xia X, Zhang Z, Nong B, Zeng Y, Wu Y, et al. Identification of anthocyanin biosynthesis genes in rice pericarp using PCAMP. *Plant Biotechnol J*. 2019;17(9):1700–2. <https://doi.org/10.1111/pbi.13133>.
68. Sun L, Luo H, Bu D, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013. <https://doi.org/10.1093/nar/gkt646>.
69. Kong L, Zhang Y, Ye ZQ, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(Web Server):W345–9.
70. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41(6):e74. <https://doi.org/10.1093/nar/gkt006>.
71. Pan L, Guo M, Jin X, Sun Z, Jiang H, Han J, et al. Full-length Transcriptome survey and expression analysis of parasitoid wasp *Chouioia cunea* upon exposure to 1-Dodecene. *Sci Rep*. 2019;9(1):18167. <https://doi.org/10.1038/s41598-019-54710-0>.
72. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402. <https://doi.org/10.1093/nar/25.17.3389>.
73. Li L, Wu H, Ma X, Xu W, Liang Q, Zhan R, et al. Transcriptional mechanism of differential sugar accumulation in pulp of two contrasting mango (*Mangifera indica* L.) cultivars. *Genomics*. 2020;112(6):4505–15. <https://doi.org/10.1016/j.ygeno.2020.07.038>.
74. Langmead B, Salzberg S. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
75. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*. 2008;36:D480–4. <https://doi.org/10.1093/nar/gkm882>.
76. Tatusov RL, Galperin MY, Natale DA. The COG database: a tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000; 28(1):33–6. <https://doi.org/10.1093/nar/28.1.33>.
77. Dossa K, Mmadi MA, Zhou R, Zhang T, Su R, Zhang Y, et al. Depicting the core transcriptome modulating multiple abiotic stresses responses in sesame (*Sesamum indicum* L.). *Int J Mol Sci*. 2019;20(16):3930.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.