

Comparative Transcriptome Analysis of ‘GR2’ and ‘GL05136’ in Mature-stage of Sugarcane

Yefei Lu, Kewei Ou, Qing Luo, Yu Zhang, Qin Cheng, Pengjin Zhu, Quanguang Zhou, Xinhua Pang, Ping Lv✉
Guangxi Subtropical Crops Research Institute, Nanning, 530001, China P.R., China

✉ Corresponding author email: plvgx@126.com

Molecular Plant Breeding, 2021, Vol.12, No.13 doi: [10.5376/mpb.2021.12.0013](https://doi.org/10.5376/mpb.2021.12.0013)

Received: 20 Apr., 2021

Accepted: 07 May, 2021

Published: 20 May, 2021

Copyright © 2021 Lu et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Lu Y.F., Ou K.W., Luo Q., Zhang Y., Cheng Q., Zhu P.J., Zhou Q.G., Pang X.H., and Lv P., 2021, Comparative transcriptome analysis of ‘GR2’ and ‘GL05136’ in mature-stage of sugarcane, Molecular Plant Breeding, 12(13): 1-9 (doi: [10.5376/mpb.2021.12.0013](https://doi.org/10.5376/mpb.2021.12.0013))

Abstract Sugarcane ‘GR2’ and ‘GL05136’ belong to late and early maturing high sugar varieties. Transcriptome sequencing was performed on the mixed samples of the leaves and stems of the two varieties to compare the differences in gene expression and to screen and clone genes that may be involved in sugar synthesis. Two comparison methods of interspecific at the same period and intraspecific at different periods were adopted, and four comparison groups were set up in the two varieties. KEGG and GO databases were used to analyze the overall transcriptome and the gene differential expression in the related fields of sugar synthesis. The results showed that there were 97 gene sequences of 5 enzymes related to sugar synthesis and several high-expression genes whose functions were unknown. Subsequently, the function of the high-expression genes could be verified by building a transgenic sugarcane validation system, improving the molecular mechanism theory of sugar metabolism, developing high-sugar molecular markers and even building high-expression transgenic varieties.

Keywords Sugarcane; Sugar synthesis; Transcriptome; ‘GR2’; ‘GL05136’

The sugar content of sugarcane (*Saccharum officinarum*) is high when it is pressed in factory, which has a positive influence on the economic benefit of sugar factory. For example, in the crushing season from 2017 to 2018, the total sugarcane yield in Guangxi is 5.08×10^{10} kg, and the sugar yield is 11.89%. Every 0.01% of the cane sugar content means 5.08×10^6 kg sugar, and its economic value is about 25.4 million yuan (5 yuan/kg sugar), which shows the importance of the sugar content of sugarcane (Guangxi Sugar Industry Development Office, 2018, 2017/2018 Guangxi Sugar Industry Annual Report, pp.1). However, sugar content of sugarcane is affected by multiple factors, such as temperature (Luo et al., 2018), humidity (Yao et al., 2002) and mineral elements (Liang, 2012). These factors have great variability, low controllable range and high control cost. Therefore, as an important factor affecting sugar content, sugarcane varieties can be artificially interfered to increase sugar content and reduce production costs.

Sugar metabolism in sugarcane is complex, controlled by multiple genes and interacts with environment. However, sugarcane is a heteroploid heterozygote. So far, the whole genome sequencing of cultivated varieties has not been completed (Zhang et al., 2018), which leads to a dilemma in genetic research on sugar metabolism mechanism of sugarcane. Based on the present situation, the mechanism of sugar metabolism in sugarcane can be studied from comparative transcriptomics. On the one hand, we can clone tissue-specific promoter or high expression genes related to sugar synthesis, which can be used to construct transgenic plants with high sugar content (Lu et al., 2013). On the other hand, genes related to sugar metabolism can be cloned and developed as molecular markers of high sugar content (Ni et al., 2016) to speed up the breeding process (Dai et al., 2013).

In the early stage, our research team collected and preserved more than 200 sugarcane germplasm resources in China, and successfully bred a new sugarcane variety ‘GR2’ with parent combinations ‘Roc20’ and ‘Yuetang 91-976’, which are characterized by late maturity, high yield, high sugar content and strong perennial root (Cheng et al., 2018). According to the results of our experiment in 2016-2017, the average sucrose content of ‘GR2’ from November to December, December to January and January to February were 12.01%, 13.4% and 13.74%,

respectively. 'GL05136' is one of the varieties with the largest planting area in Guangxi. Its parent combination is 'CP81-1254'×'ROC22'. In the past five years, the average annual planting area of 'GL05136' has exceeded 13.3×10^4 hm² (Guangxi Sugar Industry Development Office, 2018, 2017/2018 Guangxi Sugar Industry Annual Report, pp.3), which is characterized by early maturity, high yield, high sugar content and strong adaptability (Lu and Lu, 2015). The average sucrose content of 'GL05136' from November to December, December to January and January to February were 14.19%, 14.1% and 14%, respectively. The parents and maturity stages of the two varieties were different, and the high sugar content of 'GL05136' lasted for a long time. When 'GR2' reached the best maturity stage, 'GL05136' could still maintain high sugar content. Based on two varieties of high sugar cane at different maturity stages, the related genes of high differential expression of sugar synthesis were found by comparative transcriptomics, and many unknown genes with high expression were also found. The next step is to analyze the relationship between transcriptomics and proteomics, to study the molecular mechanism of sugar metabolism, meanwhile, to develop high sugar molecular markers, and even to construct high expression transgenic varieties.

1 Results and Analysis

1.1 Annotation analysis of transcriptome assembly

A total of 51 900 Unigenes were obtained after transcriptome sequencing and assembling. The proportion of Unigenes with the length of 300~500 bp, 500~1 000 bp, 1 000~2 000 bp and 2 000 bp were 27.60%, 27.26%, 24.78% and 20.36%, respectively. And the total length was 67 098 304 bp, N50 was 1 966 bp, mean length was 1 292.84 bp. The Clean Data of each sample was compared with the assembled Unigene library, showed that the Mapped Ratio was more than 70%, which proved that the integrity of the transcriptome assembly was high. Among 51 900 Unigenes, 36159 Unigenes have annotation information, and the annotation rate is 69.67% (Table 1).

Table 1 Annotation result statistics of unigenes

Anno Database	Annotated Number	300<=length	length>=1 000
COG_Annotation	11 235	3 818	7 417
GO_Annotation	25 992	10 552	15 440
KEGG_Annotation	12 063	4 933	7 130
KOG_Annotation	18 242	6 897	11 345
Pfam_Annotation	23 524	8 229	15 295
Swissprot_Annotation	20 265	6 815	13 450
eggNOG_Annotation	32 700	13 618	19 082
NR_Annotation	35 767	15 602	20 165
All Annotated	36 159	15 913	20 246

1.2 Analysis of differentially expressed genes

1.2.1 Statistical analysis differentially expressed genes

When screening and detecting differentially expressed genes, independent statistics of the abundant gene expression values will cause false positive effects on the detection results. In order to reduce this effect, certain screening criteria will be set in the statistical analysis. In this study, $FDR < 0.01$, and the $FC \geq 2$. The results of differentially expressed genes in this study are shown in Table 2. In DEG set, former vs latter, former is control, and latter one is treated sample. So that the up regulated gene indicates that the expression level of latter gene is higher than that of former gene, otherwise it is down regulated gene.

Table 2 Number of differentially expressed genes

Group	DEG set	All DEG	Up regulated	Down regulated
A	05136-1 vs 05136-2	953	780	173
B	05136-2 vs GR2	2 532	1 186	1 346
C	GR2-1 vs GR2-2	811	655	156
D	05136-1 vs GR2-1	2 309	591	1 718

Group A and C, as intraspecific comparison groups of two different varieties, total number of differentially expressed genes was close, and significantly lower than that in group B and D. In intraspecific control group, the number of up regulated genes was significantly higher than that of down regulated genes. In the interspecific

comparison group, the number of down regulated genes was higher than that of up regulated genes, which was consistent with the characteristics of 'GL05136' with higher sugar content than 'GR2', and the number of up regulated genes in group B was significantly higher than that in group D, which was also consistent with the trend of 'GR2' with higher sugar content in two periods. In order to directly reflect the difference degree and statistical significance of gene expression level in each differential expression gene set, volcano map is used to display the relevant information (Figure 1). The ordinate $-\log_{10}(\text{FDR})$ in the map represents the negative logarithm of the error detection rate of genes. The larger the value is, the more reliable the differentially expressed genes are. The abscissa $\log_2(\text{FC})$ represents the sum of the multiple of the difference in expression of the gene in two samples. The larger the value, the greater the difference of the expression multiple between the two samples. The volcano map shows that the points with $\log_2(\text{FC})$ values greater than 5 and -5 in group B and D are significantly more than those in groups A and C, which proved that B and D groups had more differentially expressed genes and higher differentially expressed multiples. And a large proportion of these differentially expressed genes had a $\log_{10}(\text{FDR})$ value greater than 10, which proved that the differentially expressed genes were reliable.

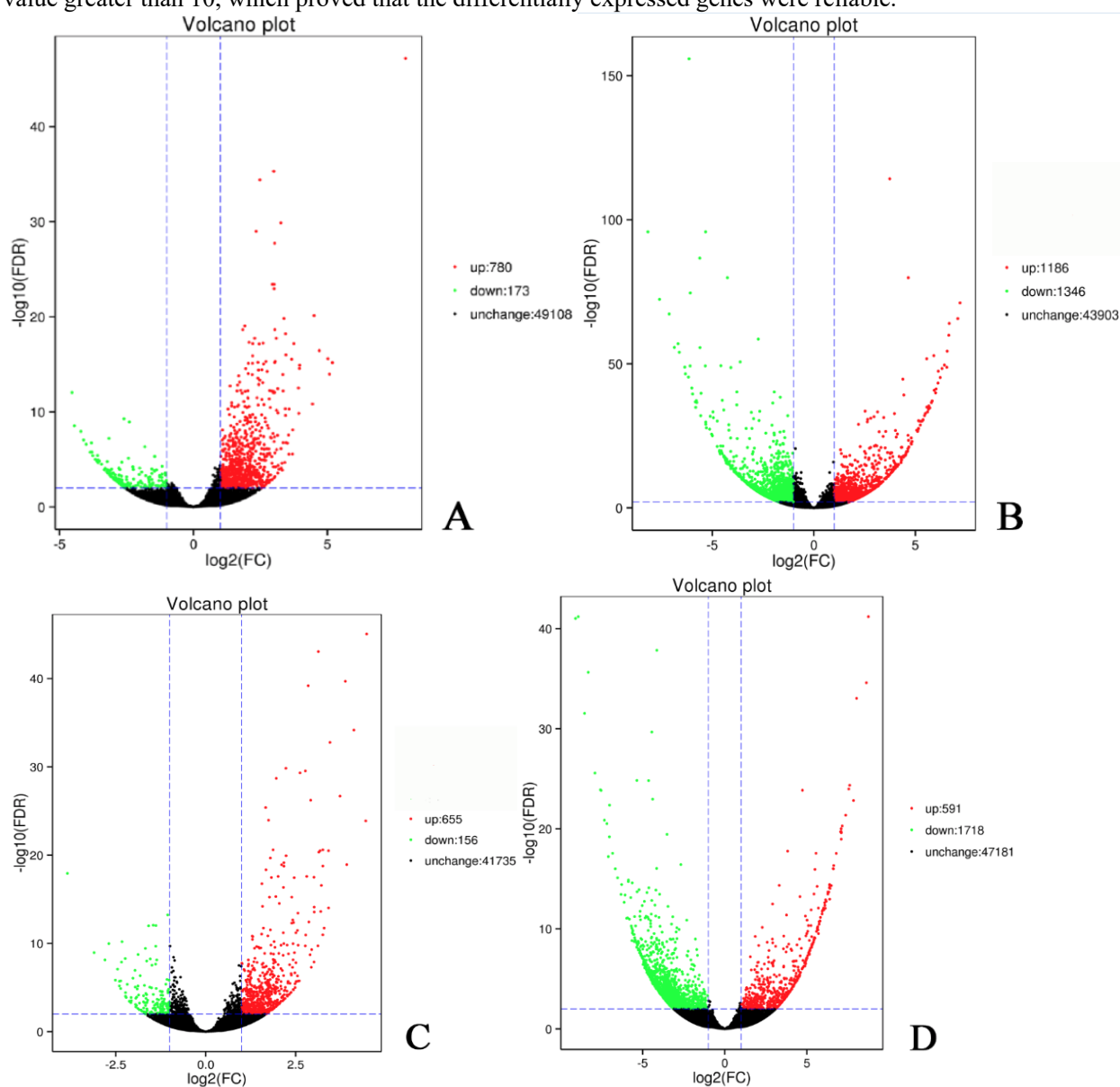


Figure 1 Volcano map of differentially expressed genes

Note: A: Volcano map of differentially expressed genes of 'GL05136' in January and February; B: Volcano map of differentially expressed genes of 'GR2' and 'GL05136' in February; C: Volcano map of differentially expressed genes of 'GR2' in January and February; D: Volcano map of differentially expressed genes of 'GR2' and 'GL05136' in January

1.2.2 GO enrichment analysis of differentially expressed genes

GO databases is the international standard classification system of gene function, which describes the functional properties of genes and gene products. GO annotation results (Figure 2) showed that among the differentially expressed gene sets, the up regulated gene have high similarity in the common analysis results of the four control groups. And the expression in metabolic process, cell process, cell, cell component, binding site, catalytic activity and other fields showed an obvious upward trend. In the intraspecific contrast (A, C), the obvious characteristic was that the gene expression of 05136-2 vs 05136-1 in group A showed an upward trend in translocation and nucleoid, while GR2-2 vs GR2-1 in group C showed an upward trend in supramolecular complexes. In interspecific comparison (B, D), the expression trend was inconsistent with intraspecific comparison. For example, in the analysis of interspecific samples collected in January, there were almost no differentially expressed genes in terms of multicellular biological processes, growth, extracellular regions, transcription factor activity, protein binding, etc. While, in the analysis of interspecific samples collected in February, the gene expression levels of these four species were significantly increased.

1.2.3 KEGG pathway analysis of differentially expressed genes

In the annotation results of KEGG, most of the differentially expressed genes were concentrated in the metabolic pathway, in which intraspecific contrast accounted for about 70% of the corresponding differentially expressed genes, and interspecific contrast accounted for about 50%.

Further analysis of the metabolic pathways of intraspecific up regulated genes showed that most of the differentially expressed genes were concentrated in carbohydrate metabolism and energy metabolism pathways. Among them, there were 78 up regulated genes in carbohydrate metabolism and energy metabolism in group A, accounting for 61% of the metabolic pathways. And 55 up regulated genes were found in carbohydrate metabolism and energy metabolism in group C, accounting for 56% of the metabolic pathway.

In interspecific comparison (B, D) analysis, due to the less up regulated genes and the more down regulated genes, the analysis focuses on the down regulated genes. Group D, as a sample group collected in January, had a total of 187 down regulated genes in carbohydrate metabolism, energy metabolism and amino acid metabolism, accounting for 97% of the metabolic process. There are 73 down regulated genes in the translation pathway of genetic information processing, accounting for 66.4% of the genetic information processing process. Group B, as a sample group collected in February, had a total of 107 down regulated genes, and there were 30 down regulated genes in carbohydrate metabolism, energy metabolism and amino acid metabolism, accounting for 50% of the metabolic process. And there were 5 down regulated genes in the translation pathway of genetic information processing, accounting for 25% of the genetic information processing (Table 3). Among the metabolic pathways, starch and sucrose metabolism are closely related to sugar synthesis (Table 4).

1.3 Analysis of genes related to sugar synthesis

Searching gene bank for known enzyme genes directly related to sugar synthesis in sugarcane, such as sucrose synthase, sucrose phosphate synthase, invertase, fructokinase and phosphofructokinase, there were 14 sucrose synthase genes, 3 sucrose phosphate synthase genes, 23 invertase genes, 37 fructokinase genes and 20 phosphofructokinase genes (Table 5, only 3 genes with high FPKM were selected for each enzyme). Analysis of the differentially expressed genes in Table 5 showed no significant increase or decrease in the expression of genes in two different stages of the two different varieties. Sucrose phosphate synthase c185455, invertase c171030, c181545 were significantly increased, while sucrose synthase c179828 was significantly decreased. These differentially expressed genes can be used as candidate genes for further functional expression verification.

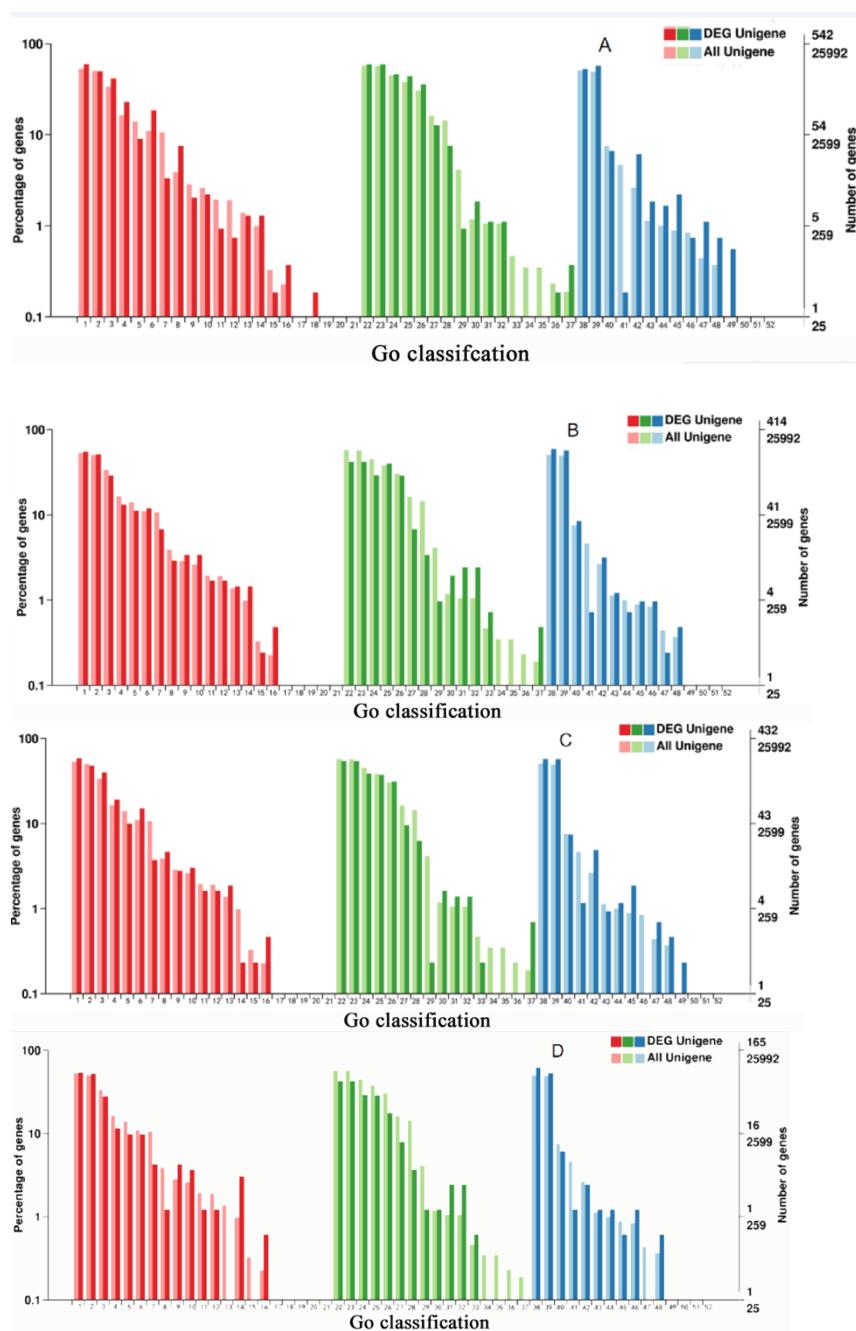


Figure 2 GO classification of the DEGs between Up-regulated genes

Note: A: GO classification of the DEGs between Up-regulated genes of 'GL05136' in January and February; B: GO classification of the DEGs between Up-regulated genes of 'GR2' and 'GL05136' in February; C: GO classification of the DEGs between Up-regulated genes of 'GR2' in January and February; D: GO classification of the DEGs between Up-regulated genes of 'GR2' and 'GL05136' in January; 1~21 are biological processes: 1: Metabolic process; 2: Cellular process; 3: Single organic process; 4: Biological regulation; 5: Positioning; 6: Stimulus response; 7: Tissue or biogenesis of cell components; 8: Signal; 9: Development process; 10: Multicellular biological processes; 11: Breeding; 12: Regeneration process; 13: Multi-biological process; 14: Detoxification; 15: They grow. 16: Immune system program; 17: Rhythmic process; 18: Transposition; 19: Cell killing; 20: Biological clock; 21: Biofacies; 22~37 are the cell components: 22: Cells; 23: Cell components; 24: Organelles; 25: Cell membrane; 26: Membrane element; 27: Organelle part; 28: Macromolecular complex; 29: Cell membrane closed the lumen; 30: Extracellular region; 31: Cell connection; 32: Syncytium; 33: Supramolecular compounds; 34: Virions; 35: Virion portion; 36: Quasi nuclear; 37: Part of the extracellular region; 38~52 are molecular functions: 38: Binding site; 39: Catalytic activity; 40: Transport activity; 41: Structural molecular activity; 42: Nucleic acid binding transcription factor activity; 43: Molecular function regulator; 44: Electronic carrier activity; 45: Signal sensor activity; 46: Antioxidant activity; 47: Transcription factor activity, protein binding; 48: Molecular sensor activity; 49: Nutrient reservoir activity; 50: Protein labeling; 51: Metal chaperone activity; 52: Translation regulator activity

Table 3 Analysis of KEGG pathway

Group	Type	Total	Cellular processes	Environmental information processing	Genetic information processing	Metabolism	Organismal systems
A	All DEG	202	8 (3.96%)	23 (11.39%)	5 (2.48%)	146 (72.28%)	20 (9.9%)
	Up Regulated	173	5 (2.89%)	18 (10.4%)	4 (2.31%)	128 (73.99%)	18 (10.40%)
	Down Regulated	29	3 (10.34%)	5 (17.24%)	1 (3.45%)	18 (62.07%)	2 (6.90%)
B	All DEG	163	5 (3.07%)	21 (12.88%)	32 (19.63%)	94 (57.67%)	11 (6.75%)
	Up Regulated	56	2 (3.57%)	5 (8.93%)	12 (21.43%)	34 (60.71%)	3 (5.36%)
	Down Regulated	107	3 (2.8%)	16 (14.95%)	20 (18.69%)	60 (56.07%)	8 (7.48%)
C	All DEG	158	4 (2.53%)	18 (11.39%)	8 (5.06%)	115 (72.78%)	13 (8.23%)
	Up Regulated	135	2 (1.48%)	17 (12.59%)	5 (3.70%)	98 (72.59%)	13 (9.63%)
	Down Regulated	23	2 (8.7%)	1 (4.35%)	3 (13.04%)	17 (73.91%)	0 (0.00%)
D	All DEG	385	31 (8.05%)	9 (2.34%)	118 (30.65%)	213 (55.32%)	14 (3.64%)
	Up Regulated	35	2 (5.71%)	2 (5.71%)	8 (22.86%)	20 (57.14%)	3 (8.57%)
	Down Regulated	350	29 (8.29%)	7 (2.00%)	110 (31.43%)	193 (55.14%)	11 (3.14%)

Table 4 Analysis of Starch and sucrose metabolism pathway

Type	A	B	C	D
All DEG	12	12	16	16
Up Regulated	12	3	16	2
Down Regulated	0	9	0	14

Table 5 Genes of sugar synthesis

Enzyme	Gene ID	GR2-1 FPKM mean value	GR2-2 FPKM mean value	05136-1 FPKM mean value	05136-2 FPKM mean value
Sucrose synthase	c181561	72.8	61.5	94.5	78.6
	c179828	143.6	97.9	160.9	92.3
	c189045	86.4	81.5	76.9	77.5
Sucrose phosphate Synthase	c166170	56.2	53.4	45.9	49.8
	c185455	87.3	150.3	59.5	155.2
Phosphofructokinase	c168364	36.6	33.1	44.2	33.1
	c181750	64.6	76.5	57.9	74.3
	c179565	70.6	67.0	90.8	79.3
Fructokinase	c184634	92.5	110.1	108.4	144.8
	c187537	55.8	46.3	41.5	43.8
	c173434	54.5	61.1	42.8	30.0
Invertase	c182175	41.8	44.4	40.9	52.6
	c171030	22.9	48.1	36.2	154.5
	c186842	40.6	49.3	28.1	88.2
	c181545	20.3	78.7	10.5	86.6

1.4 Differentially expressed genes with high expression

In view of the lack of systematic studies on the complete mechanism of sugar synthesis in sugarcane, therefore, in addition to the enzyme genes directly related to sugar synthesis mentioned above, those differentially expressed genes with expression, especially those genes with unknown function or related to the three major metabolisms (sugar, protein and lipid metabolism) should be included in the study. Among the different expression gene sets, the $\log_2(\text{FC})$ value was used as the screening standard, and the 5 genes with the highest $\log_2(\text{FC})$ value in each group were selected for functional analysis. The results showed that (Table 6) Group A, as the intraspecific comparison of 'GL05136' in different periods, had several unknown functional genes with different expression levels, and the gene with the highest expression level was a monooxygenase gene. In group C, as intraspecific comparison of 'GR2' in different periods, the differentially expressed genes with higher expression were all related to amino acid metabolism. Group B and group D, as interspecific comparison groups of 'GL05136' and

'GR2' in the same period, should be analyzed together. The results showed that the expression of c173306 and c155181 genes in 'GR2' was higher than that in 'GL05136' in two different periods.

Table 6 Functional analysis of high expression genes

Group	Gene ID		GO annotation	KEGG pathway	NR database
A	c169663	7.91	Monooxygenase activity	Unknown	Sorghum bicolor
	c168932	5.18	Unknown	Unknown	Unknown
	c179427	5.07	Unknown	Unknown	Hypothetical protein PAHAL
	c165757	5.02	Serine-type endopeptidase inhibitor activity	Unknown	Trypsin inhibitor
	c180938	4.70	Hydrolase activity	IAA-amino acid hydrolase	Uncharacterized protein
B	c168487	7.20	Unknown	Unknown	Unknown
	c155181	7.09	Unknown	Unknown	ERBB-3 BINDING PROTEIN
	c173306	6.67	Binding	Meiotic recombination protein	Sorghum uncharacterized protein
	c175732	6.64	Unknown	Unknown	Hypothetical protein
	c168831	6.57	Kinase activity	Unknown	Aegilops uncharacterized protein
C	c171132	4.47	Unknown	Unknown	Setaria italica hypothetical protein
	c150544	4.45	Phenylpropanoid biosynthetic process	Unknown	Sorghum bicolor dirigent protein
	c164549	4.12	Amino acid transmembrane transport	Solute carrier family	Sorghum bicolor amino acid transporter
	c163749	3.92	Unknown	Homeobox-leucine zipper protein	Homeobox-leucine zipper protein
	c163233	3.88	Plastid	Unknown	Sorghum bicolor chaperone protein
D	c173306	8.75	Binding	Meiotic recombination protein	Sorghum uncharacterized protein
	c155181	8.62	Unknown	Unknown	ERBB-3 Binding Protein
	c158922	8.03	Arginine catabolic process	Arginine decarboxylase	Arginine decarboxylase
	c97530	7.84	ADP binding	Unknown	Panicum hypothetical protein
	c134319	7.61	Unknown	Unknown	Unknown

2 Discussion

The characteristic of this study is that there are many ways of comparative analysis, including the interspecific comparison of different sugarcane varieties in the same period and the intraspecific comparison of the same sugarcane varieties in different periods. It is expected that the differentially expressed genes related to sucrose synthesis can be found accurately by various comparative analysis methods during sucrose synthesis.

In terms of transcriptome sequencing results, we completed the transcriptome sequencing of 12 sugarcane samples from two varieties in two periods and obtained 101 Gb Clean Data. Clean Data of each sample was 8.00 Gb. Tang et al. (2018) completed 8 sugarcane samples comparative transcriptome and obtained 57 Gb Clean Data. Qiu et al. (2018) completed 2 sugarcane samples comparative transcriptome and obtained 11.6 Gb Clean Data. Compared with the results of them, we further enriched the transcriptome information of sugarcane.

There are more intuitive analysis results in the analysis of KEGG pathway. Group A and Group C were used as intraspecific comparative analysis in different periods, the genes of starch and sucrose metabolism pathway were all up regulated. While Group B and Group D were used as interspecific comparative analysis in the same period, the genes of the starch and sucrose metabolic pathways mainly were down regulated. The results were consistent with the characteristics of 'GL05136' with higher sugar content than 'GR2' and 'GL05136' maturing earlier than 'GR2'.

Among the known enzymes regulating sugar synthesis of sugarcane, sucrose synthase can catalyze and decompose sucrose at the same time. Sucrose phosphate synthase catalyzed reversible reaction of UDPG and fructose-6-phosphate to form UDP and sucrose-6-phosphate. Invertase, also known as sucrase, catalyzes a reversible reaction between sucrose and water to produce fructose and glucose in the process of sucrose metabolism. Fructokinase is responsible for the phosphorylation of fructose produced by invertase. Phosphofructokinase is the rate limiting enzyme of glycolysis, which can act on fructose-6-phosphate (Li, 2010, China Agricultural Press, pp.14-20). In this study, a total of 97 enzyme genes were found, which can be verified from these genes in an attempt to find molecular markers for high-sugar traits.

In the field of sugarcane, no related research results have been found on oxygenase. In the study of monooxygenase in other fields (Liu and Liu, 2004), it mainly exists in microorganisms and is a mixed functional oxidase, which can catalyze the direct oxygenation of organic molecules. 'GL05136' and 'GR2' are new sugarcane varieties, there are few studies on their genetic resources, so there are many genes which are not marked in GO and KEGG database in the analysis of high expression genes. As for which genes lead to the difference between the two sugarcane varieties in maturity and sugar content, and which genes can be developed as analysis markers for sugarcane breeding with high sugar content, it is necessary to construct the verification system of transgenic sugarcane at the later stage and carry out the gene verification in the process of selection and breeding.

3 Materials and Methods

3.1 Experimental materials

Using 'GR2' and 'GL05136' planted in the field of the second experimental area in Guangxi Subtropical Crops Research Institute as materials. Samples were taken at the end of January and the end of February 2019, and the mixed samples of +1 leaf and stem tissues were collected and immediately frozen in liquid nitrogen. Illumina Novaseq sequencing was carried out by Biomarker Technologies. Three replicates were set for each group, and a total of 12 samples were collected (Table 7). GR2-1 represents the samples collected in January of 'GR2', GR2-2 represents the samples collected in February of 'GR2', 05136-1 represents the samples collected in January of 'GL05136', and 05136-2 represents the samples collected in February of 'GL05136'.

Table 7 Sample information

No.	Sample ID	OD _{260/280}	OD _{260/230}	Sucrose content (%)	Average sucrose content (%)	Remark
1	05136-1-1	2.08	1.09	17.3	16.9	3 duplicate samples collected of 'GL05136' on January 28
2	05136-1-2	1.98	0.5	16.9		
3	05136-1-3	2.1	1.86	16.4		
4	GR2-1-1	2.04	0.5	14.2	14.0	3 duplicate samples collected of 'GR2' on January 28
5	GR2-1-2	2.06	0.08	13.5		
6	GR2-1-3	2.05	1.72	14.2		
7	05136-2-1	2.00	1.54	15.7	16.1	3 duplicate samples collected of 'GL05136' on February 28
8	05136-2-2	1.94	1.3	17.1		
9	05136-2-3	2.07	1.61	15.5		
10	GR2-2-1	2.12	1.43	13.7	14.5	3 duplicate samples collected of 'GR2' on February 28
11	GR2-2-2	2.11	1.93	14.6		
12	GR2-2-3	2.1	1.86	15.1		

3.2 Difference analysis scheme

In this experiment, different varieties in the same period and the same varieties in different periods were used for comparative analysis, and four comparative groups were set up. The specific comparative analysis scheme is as follows (Table 8).

Table 8 Analysis scheme design

Group	DEG Set		Annotation
A	05136-1	vs	05136-2
B	05136-2	vs	GR2-2
C	GR2-1	vs	GR2-2
D	05136-1	vs	GR2-1

3.3 Analysis of differentially expressed genes

Among GO, KEGG, NR, COG, KOG, Swiss-Prot, eggNOG and many other databases, GO is an international standardized gene functional classification system that provides dynamic updating of standard vocabulary. KEGG is a database used to analyze the metabolic pathways of gene expression products in cells. The rest of the databases are not very useful for the analysis of functional genes in sugarcane, so this experiment focuses on the analysis of differentially expressed genes in sugarcane with GO and KEGG databases.

Author's contributions

LYF designed and carried out the study. OKW and LQ completed the sample collection; ZY and CQ completed the data analysis and drafted the manuscript. ZPJ, ZQG and PXH participated in the experimental design, analysis of experimental results. LP conceived of the project, directed the design of the study, data analysis, draft and revision. All authors read and approved the final manuscript.

Acknowledgments

The study was supported by the Science and Technology Development Fund of Guangxi Academy of Agricultural Sciences (2020YM137), the Special Fund of Innovation-Driven Development of Guangxi (AA17202042-5), and the Scientific Research and Technology Development Program of Guangxi (1598006-1-1B).

Reference

- Cheng Q., Pang X.H., Lv P., Zhou Q.G., Tan Q.L., Jin G., Zhu P.J., and Ou K.W., 2018, Identification of two guire sugarcane varieties (lines) using inter simple sequence repeat (ISSR) marker, *Kexue Jishu Yu Gongcheng (Science Technology and Engineering)*, 18(8): 191-195
- Dai P.H., La P., Guo C.K., Gao W.W., Zhang J.X., Luo S.P., 2013, Study on optimization for SSR reaction system and molecular marker of important agronomy traits in *Beta vulgaris*, *Xinjiang Nongye Kexue (Xinjiang Agricultural Sciences)*, 50(7): 1199-1205
- Liang F.X., 2012, Physiological and molecular basis of drought resistance enhanced by Si application in sugarcane, Dissertation for Ph.D., College of Agriculture, Guangxi University, Supervisor: Li Y.R., pp.20-40
- Liu Z.P., and Liu S.J., 2004, Advances in the molecular biology of nitrifying microorganisms, *Yingyong Yu Huanjing Shengwu Xuebao (Chinese Journal of Applied & Environmental Biology)*, 10(4): 521-525
- Lu W.X., and Lu L.W., 2015, Breeding and characteristics of new sugarcane varieties Guilui 05136, *Ganzhe Tangye (Sugarcane and Cane Sugar)*, (4): 1-6
- Lu Y.F., Wu L., Qin X.L., Feng J.X., and Liu J.L., 2013, Cloning of the gene encoding cellobiohydrolase I from *Trichoderma harzianum* A25-2 and expression of the sequence encoding the enzyme's catalytic domain in *Trichoderma viride* HP35-3, *Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 32(5): 569-574
- Luo S.S., He H.L., Tang L.Q., Liu L.J., and Lu Z.J., 2018, Effects of temperature and rainfall on sugarcane growth, *Zhongguo Tangliao (Sugar Crops of China)*, 40(1): 13-15
- Ni H.T., Ni H.B., Zhang F.S., 2016, Application of molecular marker technology in sugarbeet breeding, *Zhongguo Nongye Tongbao (Chinese Agricultural Science Bulletin)*, 32(16): 132-137
- Qiu L.H., Luo H.M., Chen R.F., Huang X., Chen Z.L., Fan Y.G., Chen D., Li Y.R., and Wu J.M., 2018, Establishment and preliminary analysis on transcriptome of sugarcane between stalk and tiller based on RNA-Seq technology, *Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 37(3): 1271-1279
- Tang S.Y., Yang L.T., and Li Y.R., 2018, Comparative analysis on transcriptome among different sugarcane cultivars under low temperature stress, *Shengwu Jishu Tongbao (Biotechnology Bulletin)*, 34(12): 26-34
- Yao R.L., Li Y.R., Zhang G.R., and Yang L.T., 2002, Endogenous hormone levels at technical maturity stage in sugarcane, *Sugar Tech*, 4: 14-18
<https://doi.org/10.1007/BF02956874>
- Zhang J.S., Zhang X.T., Tang H.B., Zhang Q., Hua X.T., Ma X.K., Zhu F., Jones T., Zhu X.H., Bowers J.E., Wai C.M., Zheng C.F., Shi Y., Chen S., Xu X.M., Yue J.J., Nelson D.R., Huang L.X., Zhen L., Xu H.M., Zhou D., Wang Y.J., Hu W.C., Lin J.S., Deng Y.J., Pandey N., Mancini M., Zerpa D., Nguyen J.K., Wang L.M., Yu L., Xin Y.H., Ge L.F., Arro J., Han J.O., Chakrabarty S., Pushko M., Zhang W.P., Ma Y.H., Ma P.P., Lv M.J., Chen F.M., Zheng G.Y., Xu J.S., Yang Z.H., Deng F., Chen X.Q., Liao Z.Y., Zhang X.X., Lin Z.C., Lin H., Yan H.S., Kuang Z., Zhong W.M., Liang P.P., Wang G.F., Yuan Y., Shi J.X., Hou J.X., Lin J.X., Jin J.J., Cao P.J., Shen Q.C., Jiang Q., Zhou P., Ma Y.Y., Zhang X.D., Xu R.R., Liu J., Zhou Y.M., Jia H.F., Ma Q., Qi R., Zhang Z.L., Fang J.P., Fang H.K., Song J.J., Wang M.J., Dong G.R., Wang G., Chen Z., Ma T., Liu H., Dhungana S.R., Huss S.H., Yang X.P., Sharma A., Trujillo J.H., Martinez M.C., Hudson M., Riascos J.J., Schuler M., Chen L.Q., Braum M.D., Li L., Yu Q.Y., Wang J.P., Wang K., Schatz M.C., Heckerman D., Van Sluys M., Souza G.M., Moore P.H., Sankoff D., VanBuren R., Paterson A.H., Nagai C., and Ming R., 2018, *Nature Genetics*, DOI: 10.1038/s41588-018-0293-7
<https://doi.org/10.1038/s41588-018-0293-7>