

# Integrated transcriptome, small RNA and degradome sequencing approaches proffer insights into chlorogenic acid (CGA) biosynthesis in leafy sweet potato

**Yi Liu**

Food Crops Institute, Hubei Academy of Agricultural Sciences

**Wenjin Su**

Food Crops Institute, Hubei Academy of Agricultural Sciences

**Lianjun Wang**

Food Crops Institute, Hubei Academy of Agricultural Sciences

**Jian Lei**

Food Crops Institute, Hubei Academy of Agricultural Sciences

**Shasha Chai**

Food Crops Institute, Hubei Academy of Agricultural Sciences

**Chunhai Jiao**

Hubei Academy of Agricultural Sciences

**Wenyang Zhang** (✉ [wyzhang@yangtzeu.edu.cn](mailto:wyzhang@yangtzeu.edu.cn))

Yangtze University <https://orcid.org/0000-0002-9926-4115>

**Xinsun Yang**

Food Crops Institute, Hubei Academy of Agricultural Sciences

---

## Research article

**Keywords:** Leafy sweet potato, TP, CGAs, Transcriptome, Small RNA (sRNA), microRNAs (miRNAs)

**Posted Date:** March 24th, 2020

**DOI:** <https://doi.org/10.21203/rs.2.14343/v3>

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

---

## Abstract

Background: Phenolic compounds play key roles in health protection and leafy sweet potato is an excellent source of total phenolics (TP). The chlorogenic acid (CGA) family, which includes caffeoylquinic acid (CQA), 3,4-O-dicaffeoylquinic acid (3,4-diCQA), 3,5-O-dicaffeoylquinic acid (3,5-diCQA) and 4,5-O-dicaffeoylquinic acid (4,5-diCQA), constitutes the major components of phenolic compounds in leafy sweet potato. However, the mechanism of CGA biosynthesis in leafy sweet potato is unclear. The objective of present study is to dissect the mechanisms of CGA biosynthesis by using transcriptome, small RNA (sRNA) and degradome sequencing. Results: Transcriptome sequencing of twelve samples (triple replicates) from one low-CGA content genotype and one high-CGA content genotype at two stages (65 and 85 days after planting) identified a total of 2333 common differentially expressed genes (DEGs). The enriched DEGs were related to photosynthesis, starch and sucrose metabolism and phenylpropanoid biosynthesis. In this study, functional genes CCR, CCoAOMT and HCT in the CGA biosynthetic pathway were uniformly downregulated, indicating the way to lignin was altered, and two possible CGA biosynthetic routes were hypothesized. The sRNA sequencing identified a total of 38 DE miRNAs. Using in silico approaches, 1799 targets were predicated for 38 DE miRNAs. The target genes were enriched in lignin and phenylpropanoid catabolic processes. Transcription factors (TFs) such as *ap2/ERF* and Squamosa promoter binding protein-like (SPL) predicated in silico were validated by degradome sequencing. Association analysis of the DE miRNAs and transcriptome datasets identified that miR156 family targeted DHQ / SDH (3-dehydroquinate dehydratase/shikimate dehydrogenase), the key gene in the phenylpropanoid pathway. Conclusions: This study established comprehensive functional genomic resources for the CGA biosynthesis and provided insights into the molecular mechanisms involving in this process. The results also enabled the first perceptions of the regulatory roles of mRNAs and miRNAs and offered candidate genes for leafy sweet potato improvement s.

## Background

Sweet potato (*Ipomoea batatas* (L.) Lam.) is the seventh most important food crop in the world due to its wide adaptability, high nutrition and high productivity [1]. In the past, the tuberous roots of sweet potato were the main organs harvested. However, in recent years, the tender stems and leaves of certain sweet potato varieties consuming as fresh vegetables have become increasingly popular in many regions [2]. In central and southern China, leafy sweet potato contributes enormous economic values to the farmers. Its yield exceeds 75,000 kg/ha each year with the price about 0.59 USD/kg, making the total output value reach as much as 44,117 USD/ha [3]. Thus, planting leafy sweet potato is a commercially viable venture.

The nutritional attributes of leafy sweet potato are increasingly being recognized. It is rich in vitamins, minerals, dietary fibres, phenolics and proteins. These characteristics make it a candidate vegetable for reducing malnutrition [4, 5]. Among its nutritional components, phenolics have attracted particular attention because they can reduce the risks of serious human afflictions, such as cancer and cardiovascular diseases, and protect human body from oxidative stress which causes fatigue and aging [6-10]. Phenolics are the most abundant type of secondary metabolites produced in leafy sweet potato [2, 9, 10]. The major phenolic components in leafy sweet potato are caffeoylquinic acid derivatives, including CQA, 3,4-diCQA, 3,5-diCQA and 4,5-diCQA which belong to the CGA family [11-13] (Fig. 1a).

Currently, considerable information has been extensively reviewed regarding the steps in the phenylpropanoid metabolic pathway that lead to CGA biosynthesis in plants [14-18]. An overview of this pathway was summarized in Fig. 1b: In route 1, hydroxycinnamoyl D-glucose: quinate hydroxycinnamoyl transferase (HCGQT) catalysed the formation of CGAs from caffeoyl D-glucoside and quinic acid in the roots of sweet potato [19, 20]. In route 2, enzyme hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase (HQT) catalysed the formation of CGAs from caffeoyl-CoA and quinic acid in tomato [21, 22]. In route 3, p-coumaroyl quinate was synthesized by the catalysis of hydroxycinnamoyl-CoA (HCT) and subsequently hydroxylated by p-coumarate 3'-hydroxylase (C3'H) to form CGAs. This route was reported in *Arabidopsis* [23, 24].

Gene studies showed that functional genes and TFs in the pathway were closely related to the biosynthesis of CGAs. A transesterification reaction between caffeoyl D-glucose and D-quinic acid was discovered in the CGA biosynthetic pathway of the roots of sweet potato via the isotope tracer method [19]. In addition, HCGQT extracted from sweet potato roots was found to catalyse the formation of CGAs in *in vitro* experiments [20]. A high level of *HCT* expression could increase CGA accumulation in *Solanaceous* species [25]. In *Lonicera japonica*, the *HQT* gene was found to positively regulate CGA biosynthesis [26]. Overexpression of *HQT* gene isolated from *Cynara cardunculus* var. *scolymus* in *Nicotiana tabacum* led to rechannel of the phenylpropanoid pathway [27]. Some TFs were also reported to regulate the biosynthesis of CGA. *MYB1* was an important transcriptional activator of *C4H* (*PAL1*), while *MYB3* and *5* were found to act on the promotor of *PAL3* in carrots [28]. *ZIP8* could be specifically bound to the G-box element of *PAL2* 5'-UTR in *Lonicera japonica*, with its overexpressing leading to a decrease in CGA content [26]. The biosynthesis of many phenolic compounds was also recorded to be regulated by the *WRKY* family, for example, *WRKYs* 38, 45, 60, 89 and 93 acted as activators for *HCT2* in poplar [29]. In *Salvia miltiorrhiza*, *AP2/ERF1* was reported to be able to increase the phenolic acid level [30]. However, despite the abundance of CGA compounds in sweet potato leaves, which is far more than that of the roots [11] and growing recognition of their importance to human health, there were few data in the literatures concerning genes involved in the CGA biosynthetic pathway in leafy sweet potato.

In addition, small non-coding RNAs had been extensively studied to be participated in epigenetic regulations by altering gene expression. Small RNAs, such as miRNAs, composed a class of endogenous small non-coding RNAs that ranged from approximately 20 to 24 nt in length. They negatively regulated expression of their target genes at the post-transcriptional and translational levels and played crucial roles in diverse biological processes, including plant growth and development, viral defence, metabolism and apoptosis [31]. Although much progresses had been made in miRNA research in plants, including a few studies in sweet potato [32-35], the mechanisms of miRNA regulating CGA biosynthesis in leafy sweet potato remained unclear.

To better understand the basis of the high phenolic levels of leafy sweet potato and elucidate the global expression patterns of genes and miRNAs involved in the CGA biosynthetic pathway, the present study employed transcriptome, small RNA and degradome sequencing approaches using two leafy sweet potato genotypes. These genotypes comprised one high-CGA content genotype and one low-CGA content genotype. The comprehensive and integrated analysis of different datasets identified DE mRNAs, DE miRNAs, and DE miRNA targets in CGA biosynthetic pathway and proposed the possible CGA biosynthetic routes in leafy sweet potato.

## Results

### *Analyses of TP and CGA compositions*

The search for CGA-related candidate genes and miRNAs was initiated by profiling two CGA-producing leafy sweet potato genotypes (E: EC16; F: Fushu No.7-6) at two stages (S1: 65 days after planting; S2: 85 days after planting) using Folin-Ciocalteu and HPLC methods following Xu et al. [12]. As illustrated in Fig. 2a, the TPCs of E were ~1.6 and ~1.7 fold those of F at S1 and S2, respectively. In Fig. 2b and c, at S1, the contents of CQA, 3,4-diCQA, 3,5-diCQA, and 4,5-diCQA in E were ~2.4, ~1.7, ~3.0, and ~2.4 fold those in F; at S2, the contents of CQA, 3,4-diCQA, 3,5-diCQA, and 4,5-diCQA in E were ~11.7, ~7.7, ~32.5 and ~24.7 fold those in F. Overall, the TP and CGA contents differed significantly between genotypes at each stage and between stages of each genotype (T-TEST,  $P < 0.01$ ). Obviously, within the same management condition, the TP and CGA contents of E were significantly higher than F; S1 notably higher than S2.

### *Transcriptome sequencing and analyses*

The RNA-seq reads for two genotypes at two stages (three replicates) included 1675.7 million reads, with individual libraries containing 128.4 to 185.7 million reads (Table 1). Reads from each sample were mapped to the reference genome [36] using HISAT [37]. BLAST mapping [38] revealed 29834 (91.84%) genes with homology to protein sequences in the Nr database. The expression level distributions of expressed genes were shown in Additional file 12: Fig. S1a. Correlation analysis showed that L10 revealed low correlation to the other double replicated samples, and therefore were removed from further DE mRNA and miRNA analysis (Additional file 12: Fig. S1b).

### *Differentially expressed gene analyses and annotation*

The TP and CGA compounds increased across the four pairwise comparisons (FS2 vs. FS1; ES2 vs. ES1; FS2 vs. ES2; FS1 vs. ES1), thus co-regulated DEGs across the comparisons indicated the pivotal steps in the pathway of CGA biosynthesis. In total, 6961 DEGs were found across the four comparisons (Additional file 1: Table S1). The number of DEGs ranged from 1315 (690 upregulated; 625 downregulated) for FS1 vs. ES1 to 4482 (2196 upregulated; 2286 downregulated) for FS2 vs. FS1 (Fig. 3a). A total of 1685 and 711 DEGs exhibited common expression patterns between FS2 vs. FS1 and ES2 vs. ES1, between FS1 vs. ES1 and FS2 vs. ES2; an overlap of 63 common DEGs were found across the four comparisons (Fig. 3b). In comparisons of stage-specific and genotype-specific groups, 2333 common DEGs that were identified at least in two comparisons were then considered for further analysis.

To functionally characterize expression genes, firstly, we used the BLAST algorithm to annotate 6961 DEGs based on the eggNOG, KEGG, Pfam, GO, Nr and Swiss-Prot databases. As a result, 4426, 2289, 2678, 1273 DEGs were annotated for FS2 vs. FS1, ES2 vs. ES1, FS2 vs. ES2, FS1 vs. ES1; detailed annotation information was provided in Additional file 2: Table S2. Out of these DEGs, 2333 common DEGs were assigned to 47 GO terms in Additional file 3: Table S3, including the biological process (20), molecular function (14) and cellular component (13) categories. GO enrichment analysis of common DEGs revealed that catalytic activity (GO:0003824), oxidation-reduction process (GO:0055114), oxidoreductase activity (GO:0016491) ranked in the top 20 most significant enriched terms in Fig 4a. Furthermore, the pathway analysis of common DEGs was carried out to understand the molecular mechanism using KEGG database. The DEGs were found to represent 288 pathways (Additional file 4: Table S4). The enrichment analysis suggested that photosynthesis-antenna proteins (map00196), starch and sucrose metabolism (map00500), drug metabolism-cytochrome P450 (map00982) and phenylpropanoid biosynthesis (map00940) were among the most enriched pathways (Fig. 4b). A total of 134 transcription factors (TFs) belonging to 26 families were identified differentially expressed. Among them, *C2C2* (18), *AP2/ERF* (16), *MYB-related* (11), *bHLH* (11) were the most overrepresented TF families (Fig. 4c).

### ***Metabolic pathway and gene analyses during CGA accumulation***

To provide a global view of leafy sweet potato secondary metabolism, common DEGs with different map ids were further submitted for analysis via the online Interactive Pathway (ipath) explorer v2 (Fig. 5a) [39]. The metabolic pathways such as pentose phosphate pathway (Fig. 5b), phenylalanine biosynthetic pathway (Fig. 5c), CGA biosynthetic pathway (Fig. 5d) and flavonoid biosynthesis showed enhanced, which were in accordance with the results of GO analysis. As pentose phosphate metabolism, phenylalanine biosynthetic pathway and CGA biosynthetic pathway were vital steps for CGA biosynthesis, genes involving in these pathways were fully illustrated.

In this work, genes participating in CGA biosynthesis showed special expression patterns. For the common expression pattern, two synonymous DEGs of *G6PDH* (itb02g05910; itb03g00300) which encoded glucose-6-phosphate 1-dehydrogenase were upregulated, providing sufficient NADPH and phosphoenolpyruvic acid (PEP) for shikimic acid pathway. *DHQ/SDH* (itb14g20920) which catalyzed the biosynthesis of phenylalanine was upregulated as well. *HCT* that had three synonymous DEGs (itb03g29460; itb01g04710; itb01g04740) expressed differentially across the four comparisons. Yet itb03g29460 was the most expressed one, which was downregulated in ES2 vs. ES1 and FS2 vs. FS1. *CCR*, which encoded cinnamoyl-CoA reductase, had 4 synonymous DEGs (itb09g17150; itb09g17200; itb02g23900; itb07g23820). They were downregulated in different comparisons. *CCoAOMT* that encoded caffeoyl-CoA O-methyltransferase had three synonymous DEGs (itb12g05230; itb01g21750; itb12g20360), all of which were downregulated. In this study, *PAL* (itb09g15750) was found to be only upregulated in the comparison FS2 vs. ES2. *C3'H* (p-coumarate 3'-hydroxylase), another name *CYP 98A3* (itb01g24570) was upregulated in the comparison FS2 vs. ES2. The schematic of metabolic data related to leafy sweet potato CGA accumulation was briefly illustrated in Fig. 6.

### ***High-throughput small RNA sequencing***

The small RNA sequencing resulted in 248.6 million clean reads, with 14.4 to 30.0 million reads per library. Reads with length > 17 nt and < 33 nt were kept, following by the removal of ribosomal RNA (rRNA), transfer RNA (tRNA), small nucleolar RNA (snoRNA), and repetitive sequences (Table 1). The length distribution patterns of the sRNAs were similar in the eleven libraries. They ranged from 18 to 30 nt, of which 24 nt were the most abundant size (Fig. 7a). In order to identify known miRNAs, the filtered reads were searched against the miRNAs in miRBase. A total of 149 known miRNAs were obtained. As some of the known miRNAs were aligned with more than one pre-miRNAs, the detailed information of all aligned miRNAs was listed in Table S5. The length distribution of known miRNAs exhibited a peak at 21 nt (Fig. 7b), similar to the results reported in previous research in sweet potato and other species [32-35]. Reads that could not be mapped to miRBase were subjected to novel miRNA predication by miRDeep2 and the most length distribution of novel miRNA was 24 nt following by 21 nt (Fig. 7c) [40]. A total of 22 novel miRNAs were identified and listed in Table S5. The negative folding free energies of the hairpin structures of novel miRNAs ranged from -68.37 to -26.52 kcal mol<sup>-1</sup> with an average of -43.47 kcal mol<sup>-1</sup>. The minimal folding free energy index (MFEI) of novel miRNAs ranged from 0.86 to 1.69 with an average of 1.13.

### ***DE miRNA expression during CGA accumulation***

miRNAs were considered as DE miRNAs if they had absolute values of log<sub>2</sub> (Fold Change) ≥ 1 and FDR (False discovery rate) ≤ 0.05. A total of 9, 7, 18 and 9 miRNAs were identified as DE miRNAs across the four comparisons and 5 miRNAs were common differently expressed ones (Additional file 6: Table S6). The majority of DE miRNAs showed a trend of downregulation during CGA accumulation (Fig. 7d). miR156, miR166, miR167 and miR858 were found in different comparisons, which had been reported to be involved in phenylpropanoid pathway [41].

### ***Target predication via in silico and degradome approaches***

To explore the function of miRNAs, computational program was performed to predict their target genes. All identified 171 miRNAs were predicated to have 1799 targets via TargetFinder software with the score value < 4 [42]. The annotations for the 1799 miRNA targets were based on the GO, KEGG, eggNOG, Nr, Swiss-Prot and Pfam databases (Additional file 7: Table S7). The targets were uniformly assigned to 20 biological processes, 14 cellular components and 11 molecular functions. The most abundant 20 GO terms were demonstrated in Fig. 8a. The significant enriched GO terms like lignin catabolic process (GO:0046274), phenylpropanoid catabolic process (GO:0046271), lignin metabolic process (GO:0009808) and phenylpropanoid metabolic process (GO:0009698) were listed in Additional file 8: Table S8 and they were all involved in CGA accumulation pathway. Furthermore, KEGG annotation was carried out to explore the pathways in which the identified targets were involved. A total of 220 pathways were identified indicating the highly diverse functions of the targets. Phenylpropanoid biosynthesis (map00940) which was CGA accumulation related pathway were among the most 20 abundant pathways (Fig. 8b).

Using degradome sequencing, a total of 21.94 Mb clean tags and 7,892,630 cluster tags were obtained. The cluster tags were aligned to the transcriptome and Rfam database for cleavage site analysis. After processing and analysis with the software CleaveLand [43], 158 miRNA-mRNA pairs were totally identified (Additional file 9: Table S9). Target analysis showed that many cleaved-target genes by miRNAs were TF genes, including *AP2/ERF*, *bZIP*, *TCP*, *MYB*, *SPL*, etc. Some miRNAs had more than one target genes, like miR530a targeted *microtubule-associated protein 70-1-like* and *bHLH130-like* genes. On the contrary, same gene can be targeted by more than one miRNA, for instance, miR394c and miR384-5p shared the same target *F-box*. TFs such as *AP2/ERF* (itb14g16290) and *SPL* (itb01g24030) predicated *in silico* were validated in degradome sequencing.

### **Correlation analysis of DE miRNA expression profiles and their target genes**

The expression of both DE miRNAs (from small RNA-seq) and their target genes (from RNA-seq) were integrated to infer the mediatory role of miRNAs during CGA biosynthesis. Coherent interactions were the ones in which the expression of miRNA was upregulated when the expression of target mRNA was downregulated, and vice versa. In this study, the spearman mathematical method with the criterions of index  $\leq -0.8$  and P value  $\leq 0.05$  was employed. As a result, the correlation analysis of DE miRNA and their target mRNA expression profiles identified a total of 1462, 66, 319 and 19 miRNA-mRNA interaction pairs across four comparisons (FS2 vs. FS1; ES2 vs. ES1; FS2 vs. ES2; FS1 vs. ES1) (Additional file 10: Table S10). From these coherent pairs, miR156g-5p, miR156k-5p, miR156i-5p, miR156h-5p, miR156j-5p, miR156e in FS2 vs. FS1 which belonged to miR156 family were downregulated and their target *DHQ/SDH* (itb14g20920) was upregulated with the correlation rate of -0.9. These results were in accordance with quantitative reverse transcription-polymerase chain reaction (qRT-PCR) result. In addition, through associate analysis miR156g-5p, miR156k-5p, miR156i-5p, miR156h-5p, miR156j-5p, miR156 were found downregulated and their target gene *SPL* (itb02g07930) was upregulated in FS2 vs. FS1, the negative correlation of which was -0.9.

### **Validation of differential gene and miRNA expression**

qRT-PCR analysis was carried out to validate the expression patterns of genes and miRNAs obtained from the RNA and small RNA sequencing. The expression of enzyme encoding-genes (*HCT*, *CCoAOMT*, *CCR*) in the CGA biosynthetic pathway, two synonymous *G6PD* genes (itb03g00300, itb02g05910) in the pentose pathway and one phenylalanine biosynthesis-related gene *DHQ/SDH* were validated via qRT-PCR (Fig. 9; Additional file 11: Table S11), and the results were consistent with that of the mRNA sequencing, except the deviations of *CCR* in FS2 vs. FS1 and ES2 vs. ES1. As thirty genes were investigated in qRT-PCR experiments (data not shown), *CCR* was the only one with deviations and also the results turned out the same after the qRT-PCRs of *CCR* were repeated three times, so it was considered to be downregulated in this study. In addition, six miRNAs, namely, Nov-m2294-5p, Nov-m3917-3p, Nov-m4613-3p, sly-miR168a-5p, stu-miR156e and tcc-miR530a were validated by qRT-PCR as well (Fig. 9; Additional file 11: Table S11). Similar expression trends (upregulated or downregulated) were observed between the qRT-PCR analysis and the sRNA sequencing results.

## **Discussion**

Leafy sweet potato is extremely popular among consumers in China, because it is beneficial to the health. CGAs in leafy sweet potato are not only the key attributes for health care by fresh consuming, but also have potential applications in food and pharmaceutical industries. As a result, the major objective of this research is to comprehensively study the CGA metabolism and investigate the molecular basis of this pathway in leafy sweet potato. The availability of diverse germplasm resources and high-throughput approaches, namely, transcriptome, small RNA and degradome sequencing, provide an opportunity to dissect the mechanism.

We investigated two well-characterized genotypes at two stages for their different CGA accumulation (Fig. 2a-c). Although a few genes involved in CGA biosynthesis had been reported in other plant species [14, 22-30], the molecular mechanisms underlying in leafy sweet potato remains largely unknown. In the present study, a total of 29834 genes were identified based on the reference guided transcriptome analysis of two genotypes at two stages. The sequences for all genes in route 2 and 3 were assembled and the differentially expressed functional genes were concentrated in these two routes. Therefore, it was reasonable to conclude that route 2 and 3 were the main pathways for the biosynthesis of CGA in leafy sweet potato. This result was not in accordance with that of the report by Kojima and Uritani [20], in which the biosynthesis of CGA was assumed by route 1. We speculated that the mechanism of CGA biosynthesis in the leaves of sweet potato was different from that of the root. There were 4426, 2289, 2678, 1273 DEGs annotated for FS2 vs. FS1, ES2 vs. ES1, FS2 vs. ES2 and FS1 vs. ES1. As overlapping DEGs between FS2 vs. FS1 and ES2 vs. ES1, between FS2 vs. ES2 and FS1 vs. ES1 were potential DEGs explaining CGA differences based on stage- and genotype- specific, 2333 common DEGs identified at least in two comparisons were therefore considered for further analysis.

It has been acknowledged that the accumulation of CGAs is a multifaceted process that can be traced to the pentose phosphate pathway, where the precursor phosphoenolpyruvic acid (PEP) is produced. Following this process is the shikimic acid pathway, which is the main pathway

leading to the production of phenylalanine, and then CGAs are produced by the catalysis of enzyme cascades. The formations of PEP, biosynthesis of phenylalanine and CGAs were supported by the KEGG enrichment analysis presented in the study. The expression of the important enzymes of *G6PD* and *DHQ/SDH*, which involved in the biosynthesis of PEP and phenylalanine, were integrally upregulated in high-CGA content samples compared to low-CGA content samples. In this study, ipath that integrated 123 KEGG maps from 183 species was employed as an open access online tool to indicate the enhanced catalytic activities. From the ipath map, pathways of pentose phosphate, phenylalanine biosynthesis, phenylpropanoid biosynthesis and flavonoid biosynthesis were all indicated in red, meaning related genes were active in these pathways.

As two possible routes (route 2 and 3) for CGA biosynthesis were indicated in the research, the initiation study for both routes started from DEG *PAL*. Phenylalanine was catalyzed by *PAL* to form cinnamic acid. DEG *PAL* in this study was found only in FS2 vs. ES2 instead of the common DEG aggregation. In addition, the following downstream two enzyme genes, *ACL* and *Cinnamate 4-hydroxylase (C4H)*, which catalyzed the formation of p-Coumaroyl-CoA, did not express significantly differently among comparisons in this research. These results were quite different from other higher plants, for them *PAL*, *ACL* and *C4H* were key enzymes in CGA accumulation [44]. However, the intriguing phenomenon occurred on three common DEGs *HCT*, *CCR* and *CCoMAT* in the phenylalanine pathway. The expression of *CCR* and *CCoMAT* were downregulated indicating the lignin biosynthesis was altered to CGA accumulation. For *HCT*, all homologous genes were upregulated except the most highly expressed one (itb03g29460). Though having been shown to synthesize caffeoylquinic acid *in vitro* [18], *HCT* was involved both upstream and downstream of the 3-hydroxylation step (Fig. 1b). Its inhibition could affect predominant caffeoylquinic acid catabolized into caffeoyl CoA which led to the lignin biosynthesis, and thus the CGA accumulation occurred. The same phenomenon had been reported by Hoffmann et al. [17]. The CGA mechanism also involved a number of TFs like *C2C2*, *AP2/ERF*, *MYB-related*, *bHLH*, etc.

miRNAs have emerged as master modulators of gene expression and are promising tools for crop improvement [45]. A few studies in sweet potato had reported the genome-wide discoveries of miRNAs [32-35], but no study had yet characterized the roles of miRNAs in CGA biosynthesis. In the present study, a total of 149 known and 22 novel miRNAs were identified. The expression pattern of the isolated miRNAs were analyzed, more miRNAs were downregulated than upregulated across the four comparisons (Fig. 7d) and thirty-eight miRNAs were recognized as DE miRNAs. Most of the DE miRNAs were known miRNAs. miR156, miR166, miR167 and miR858 family members were confirmed, which were reported to be either positively or negatively involved in regulating anthocyanin and catechin biosynthesis [46-48]. As anthocyanin and catechin belonged to flavonoid pathway which had close correlations with CGAs, miR166d-3p, miR166c-3p, miR166a-3p, miR166b, miR167b-3p, miR156a, miR156b, miR156c, miR156d-5p, miR156j-5p, miR156k-5p, miR156g-5p, miR156h-5p, miR156i-5p, miR156e, miR858a, miR858b identified in this study were potential targets for manipulation CGA contents for further study.

By target predication, 1799 miRNA targets were annotated based on the GO, KEGG, eggNOG, Nr, Swiss-Prot and Pfam databases (Additional file 7: Table S7). GO and KEGG analysis showed that genes were enriched in lignin and phenylpropanoid catabolic process. Degradome sequencing that had been successfully applied to identify miRNA targets in many plant species [49, 50] were employed to verify the predication results. In degradome sequencing analysis, the majority target genes were transcription factors, containing *SPL*, *HD-ZIP* and *MYB* genes. These TFs were all reported to be related to the phenylpropanoid pathway. For instance, *SPLs* played important roles in plant growth and development. The miR156/*SPL* module was reported to participate in the biosynthesis of phenylpropanoids by destabilizing the MYB-bHLH-WD (MBW) complex and directly preventing the expression of anthocyanin biosynthetic genes in *Litchi chinensis* [51]; as *HD-ZIP* TFs played crucial roles in shoot apical meristem and organ polarity, the blockage of miRNA165/166 caused the upregulation of *HD-ZIP* TFs and increased IAA content accompanied by enhanced anthocyanin [41]. In this research, the analyses of degradome sequencing demonstrated that *SPLs* were targeted by miR156, *HD-ZIP* by miR166, *MYB* by miR159. These results suggested that miR156, miR166 and miR159 might be involved in CGA biosynthetic pathway.

The miRNA and mRNA association analysis revealed that miR156 and miR858 families were strongly negatively correlated with the mRNAs. miR156g-5p, miR156k-5p, miR156i-5p, miR156h-5p, miR156j-5p, miR156e in FS2 vs. FS1 commonly negatively targeted DEG *DHQ/SDH* (itb14g20920) and *SPL* (itb02g07930); while miR858b was upregulated in FS2 vs. ES2 and its target *MYB* (itb12g01510) was downregulated. All these target genes were noted to be associated with CGA biosynthesis.

## Conclusions

In summary, the present study integrated mRNA and miRNA expression data along with degradome analysis to identify key factors in CGA biosynthesis in leafy sweet potato. The study revealed complex mechanism, in which pentose metabolism and lignin biosynthesis were all related to CGA biosynthesis and routes 2 and 3 were the possible CGA biosynthetic pathway in leafy sweet potato. A set of genes and miRNAs were identified as crucial roles for the CGA biosynthesis. They could serve as targets for further research of gene functions. This study provided a foundation for uncovering the CGA biosynthetic system in leafy sweet potato, and the results could be used to improve leafy sweet potato varieties for both consumer health benefits and pharmaceutical use in the future.

## Methods

### *Plant materials*

From seven investigated varieties (lines), the contents of TP and CGAs of which varied between the genotypes and stages, two brilliant leafy sweet potato variety and line Fushu No. 7-6 (F) and EC16 (E), which exhibited significant differences between each other from the points of genotype and stage, were chosen for further study (data not shown). Variety Fushu No. 7-6 was bred by Fujian Academy of Agricultural Sciences and introduced into Hubei Academy of Agricultural Sciences as a resource. EC16 was one of the progenies of Fushu No. 7-6. Both of them were kept in the plant nursery of Food Crops Institute of Hubei Academy of Agricultural Sciences and cultivated in potting soil on May 2<sup>nd</sup>, 2017 and grown under standard production practices. The leaves of the two genotypes were sampled at two stages: 65 days (S1) and 85 days (S2) after planting. Each sample was pooled with leaves from three individual plants and three biological replicates were collected. Part of the samples were immediately frozen in liquid nitrogen and stored at -80°C in a freezer for transcriptomic analysis. The remaining samples were rinsed gently and dried in a blast drier (Shangce, Wuhan, China) at 70°C. After powdered by a blender, the dehydrated samples were filtered through a 60-mesh sieve and then were placed in the sealed plastic bags, maintaining in a freezer at -20°C for further TP and CGA measurement analyses.

### *Determination of TPC*

TP content was determined by following Xu et al.[12] with a few modifications. The powders of the samples were extracted 25 times (w/v) with 70% ethanol for 40 min in a 80°C water bath. After the solution was centrifuged at 5000 × g for 10 min, the residue was re-extracted with 70% ethanol as described above. The supernatants were combined, concentrated in a rotary evaporator and filtered. The crude solution was diluted with distilled water to 100 ml. One ml of the prepared solution was mixed with 1 ml of Folin-Ciocalteu reagent (Guoyao, Shanghai, China), 3 ml of 7.5% Na<sub>2</sub>CO<sub>3</sub>, and 5.0 ml of distilled water in a test tube and allowed to react at 45°C for 1.5 h in a water bath. The absorbance was measured at 756 nm using a UV-2880 spectrophotometer (UNICO, Shanghai, China). A calibration curve of gallic acid (ranging from 0 to 0.05 mg/ml) (Guoyao, Shanghai, China) was prepared, and the TPC was expressed as mg GAE (Gallic acid equivalent) per gram of DW.

### *Determination of CGA contents by HPLC*

The powder of the samples was extracted 50 times (w/v) with 70% ethanol for 40 min in a 80°C water bath. The solution was centrifuged at 5000 g for 10 min, and the residue was re-extracted twice with 70% ethanol as described above. The supernatant was filtered through a cellulose acetate membrane filter (0.2 µm, Advantec, Japan) and used for analysis. A 20 µl portion of the filtrate was injected into the HPLC Agilent 1260 system (Agilent Technologies Inc., USA) with the ZORBAX Eclipse Plus C18 column. The procedures of eluting were described as below: firstly, the column oven temperature was set at 40°C and the mobile phase consisting of 0.1% (v/v) formic acid (A) and 100% (v/v) acetonitrile (B) was prepared. Then, elution procedure was performed with a linear gradient as follows: 10% to 40% B from 0 to 30 min; 40% to 10% B from 30 to 30.1 min; holding at 10% B from 30.1 to 35 min. The flow rate was 0.5 ml/min and CGAs were detected at 326 nm. The retention times of the CGA compounds were compared with standard reagents, including CQA, 3,4-CQA, 3,5-CQA and 4,5-CQA (Sigma, St. Louis, MO, USA).

### *Library construction and sequencing*

Total RNA was extracted from the samples using the RNeasy Plus Mini kit (Qiagen, Valencia, CA, USA). The purity, concentration and integrity of RNA were confirmed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). For RNA-seq, 1.5 µg RNA per sample was used as input material for rRNA removal using the Ribo-Zero rRNA Removal Kit (Epicentre, Madison, WI, USA). Sequencing libraries were generated using the NEBNext<sup>®</sup> Ultra<sup>™</sup> Directional RNA Library Prep Kit following the manufacturer's instructions. The library fragments of ~280 bp in length were purified on 2% low range ultra agrose followed by PCR amplified using phusion DNA polymerase (NEB). The products were prepared for paired-end sequencing on the Illumina High-Seq 2500 sequencing platform (Illumina, Inc.; San Diego, CA, US). For sRNA-seq, 1 µg RNA per sample was used as input material. Sequencing libraries were generated using the NEBNext<sup>®</sup> Ultra<sup>™</sup> small RNA Library Prep Kit following the manufacturer's instructions. Shortly, cDNAs were synthesized by reverse transcription and amplified with 12 PCR cycles. After the purification of the PCR products, deep sequencing was performed on the Illumina High-Seq 2500 sequencing platform.

### *Expression analysis and annotation*

Raw data were first processed with SeqPrep and Sickle software to remove reads containing poly-N and adaptor sequences. The clean RNA-seq reads were filtered and mapped to the *Ipomoea triloba* (NSP323.v3) genome using HISAT software [37]. The mapped reads were assembled by StringTie. Gene expression levels were estimated using FPKM (fragments per kilobases of transcript per million fragments mapped) values calculated by RSEM software [52]. An mRNA was considered as a DE mRNA via the DESeq2 R package when it exhibited a two-fold or higher expression change and its FDR was below 0.01 in the comparisons [53]. To functionally characterize the pathway and expression clusters, the BLAST algorithm was used to annotate DE mRNAs based on the eggNOG, KEGG, Swiss-Prot, and Nr databases.

For sRNAs, the clean reads were aligned to the *Ipomoea triloba* genome [36] via Bowtie2 [54], and then the reads aligned with the reference genomes were searched against miRbase and Rfam to detect known miRNAs. The prediction of precursors for the novel miRNAs was performed by using miRDeep2 [40], and ones with the MFEIs of precursors (pre-miRNAs) above 0.85 were considered to be novel. Moreover, the normalized copy number of the novel miRNAs were required to be  $\geq 10$  in at least one small RNA library to avoid potential false positive. The expression levels of miRNAs in each sample were calculated and normalized by the transcript per million (TPM) algorithm. Differential expression analyses were carried out using the DESeq R package (1.10.1). miRNAs with absolute values of  $\log_2$  (Fold Change)  $\geq 1$  and  $FDR \leq 0.05$  were considered DE miRNAs.

### ***Degradome sequencing***

Degradome library construction was conducted with the method previously described by German et al. [55], with some modifications. mRNAs were isolated by Dynabeads and then endonucleased. The fragments containing 5'-monophosphates were ligated with 5' adaptors and used to generate first-strand cDNA. Single-end sequencing was performed on the Illumina High-Seq 2500 sequencing platform.

### ***qRT-PCR***

qRT-PCR analyses were carried out to determine the reliability of the RNA-seq results for expression profile analysis. All primers were designed according to the mRNA sequences and miRNA mature sequences and were synthesized commercially in the company Tianyi Huiyuan, Wuhan. The primer sequence information was presented in Additional file 11: Table S11. qRT-PCRs for mRNAs and miRNAs were carried out in a 20  $\mu$ l system: 2.0  $\mu$ l cDNA product, 10  $\mu$ l 2  $\times$  qPCR Mix, and 2.5  $\mu$ M for each of the forward and reverse primers. The reactions were incubated in a Real Time System Thermocycler for 10 min at 95°C, followed by 40 cycles of 15 s at 95°C, 60 s at 60°C. All reactions were run in three replicates and  $\beta$ -actin served as the endogenous reference gene. The  $2^{-\Delta\Delta CT}$  method was employed to analyse the relative changes of genes and miRNAs [56]. T-TEST was employed to analyse the data generated from qRT-PCRs.

## **List Of Abbreviations**

CGA: Chlorogenic acid; TP: Total phenolics; CQA: Caffeoylquinic acid; 3,4-diCQA: 3,4-O-dicaffeoylquinic acid; 3,5-diCQA: 3,5-O-dicaffeoylquinic acid; 4,5-diCQA: 4,5-O-dicaffeoylquinic acid; sRNA: small RNA; DEGs: Differentially expressed genes; miRNAs: microRNAs; TFs: Transcription factors; AP2/ERF: apetala2/ethylene response factor; SPL: Squamosa promoter binding protein-like; DHQ/SDH: 3-dehydroquininate dehydratase/shikimate dehydrogenase; HCGQT: Hydroxycinnamoyl D-glucose: quinate hydroxycinnamoyl transferase; HQT: Hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase; HCT: Hydroxycinnamoyl-CoA transferase; C3'H: p-coumarate 3'-hydroxylase; PAL: Phenylalanine ammonia-lyase; Nr: NCBI non-redundant protein sequences; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; eggNOG: Clusters of Orthologous Groups; Swiss-Prot: A manually annotated and reviewed protein sequence database; ipath: Interactive Pathway; G6PDH: Glucose-6-phosphate 1-dehydrogenase; C4H: Cinnamate 4-hydroxylase; CCR: Cinnamoyl-CoA reductase; CCoAOMT: Caffeoyl-CoA O-methyltransferase; C3'H/ CYP 98A3: p-coumarate 3'-hydroxylase; FPKM: Fragments per kilobase of transcript per million fragments mapped; rRNA: Ribosomal RNA; tRNA: Transfer RNA; snoRNA: Small nucleolar RNA; MFEI: Minimal folding free energy index; FDR: False discovery rate; qRT-PCR: Quantitative reverse transcription-polymerase chain reaction; PEP: Phosphoenolpyruvic acid; C4H: Cinnamate 4-hydroxylase; MYB-bHLH-WD; GAE: Gallic acid equivalent; DW: dry weight; pre-miRNA: miRNA precursors; TPM: Transcript per million;

## **Declarations**

### ***Ethics approval and consent to participate***

Not applicable.

### ***Consent for publication***

Not applicable.

## Availability of data

The clean data in this study is available in NCBI Sequence Read Archive (SRA) database under the submission number SUB6617451 for mRNAs, SUB6567093 for miRNAs and SUB6626042 for degradome sequencing.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the National Key R&D Program of China (2018YFD1000700, 2018YFD1000705); the International Scientific Cooperation Project of Hubei Province (2018AHB012); the Open Project Program of Biology and Biotechnology, Key Laboratory of Ministry of Agriculture and Rural Affairs; the Science and Technology Innovation Center of Hubei Academy of Agricultural Sciences (2007-620-001-03); China Agriculture Research System (CARS-11-C-15); the Characteristic Discipline of Hubei Academy of Agricultural Sciences (2015TSXK06); the Hubei Province Technological Innovation Major Project (2017ABA149); the Project of Excel Youth of Hubei Academy of Agricultural Sciences. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Authors' contributions

WZ and XY designed the study. YL and WS conducted the pre-experiments of TP and CGA measurements from various samples, and analyzed the data and prepared the manuscript under the supervision of WZ and XY. LW, JL, SC, and CJ provided manuscript revision advices. We also thank all the fellows in Dr. Yang's lab. All authors have read and approved the manuscript.

## Acknowledgements

We thank the anonymous reviewers for their comments on the manuscript.

## List Of The Softwares And Databases

### List of the softwares and databases

Software/	Version	Sources
<b>database</b>		
Sickle		<a href="https://github.com/najoshi/sickle">https://github.com/najoshi/sickle</a>
SeqPrep		<a href="https://github.com/jstjohn/SeqPrep">https://github.com/jstjohn/SeqPrep</a>
HISTA2	Version 2.1.0 2.1.0	<a href="http://ccb.jhu.edu/software/hisat2/index.shtml">http://ccb.jhu.edu/software/hisat2/index.shtml</a>
Stringtie	Version 1.3.3b	<a href="https://ccb.jhu.edu/software/stringtie/">https://ccb.jhu.edu/software/stringtie/</a>
RSEM	Version 1.3.1	<a href="http://deweylab.biostat.wisc.edu/rsem/">http://deweylab.biostat.wisc.edu/rsem/</a>
DESeq2	Version 1.24.0 11.24.01.24.01.24.0	<a href="http://bioconductor.org/packages/stats/bioc/DESeq2/">http://bioconductor.org/packages/stats/bioc/DESeq2/</a>
Bowtie2	1.24.0 Version 2.2.9	<a href="http://downloads.sourceforge.net/project/bowtie-bio/bowtie2/2.2.9/bowtie2-2.2.9-linux-x86_64.zip">http://downloads.sourceforge.net/project/bowtie-bio/bowtie2/2.2.9/bowtie2-2.2.9-linux-x86_64.zip</a>
miRDeep2	Version 2.0.0.5	<a href="https://www.mdc-berlin.de/content/mirdeep2-documentation">https://www.mdc-berlin.de/content/mirdeep2-documentation</a>
Blast2go	Version 2.5	<a href="https://www.blast2go.com/">https://www.blast2go.com/</a>
eggNOG	Version 5.0	<a href="http://eggnogdb.embl.de/#/app/home">http://eggnogdb.embl.de/#/app/home</a>
KEGG	Version 2017.08	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>
GO	—	<a href="http://www.geneontology.org/">http://www.geneontology.org/</a>
Swiss-prot	Version 2019.7.1	<a href="ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz">ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz</a>
Nr	Version 2019.6.26	<a href="ftp://ftp.ncbi.nlm.nih.gov/blast/db/">ftp://ftp.ncbi.nlm.nih.gov/blast/db/</a>
miRbase	Version Release 22	<a href="http://www.mirbase.org/">http://www.mirbase.org/</a>
Rfam	Version 14.1	<a href="http://rfam.janelia.org/">http://rfam.janelia.org/</a>
Pfam	Version v32.0	<a href="http://pfam.xfam.org/">http://pfam.xfam.org/</a>
ipath	Version 3	<a href="https://pathways.embl.de/">https://pathways.embl.de/</a>

## References

- Rodriguez-Bonilla L, Cuevas HE, Montero-Rojas M, Bird-Pico F, Luciano-Rosario D, Siritunga D. Assessment of genetic diversity of sweet potato in Puerto Rico. *PLoS One*. 2014;9:e116184.
- Islam MS, Yoshimoto M, Yahara S, Okuno S, Ishiguro K, Yamakawa O. Identification and characterization of foliar polyphenolic composition in sweetpotato (*Ipomoea batatas*) genotypes. *J Agric Food Chem*. 2002;50:3718-22.
- Mei X, Yang X, He JJ, Shi JB, Liu Y, Cai S, et al. Factor analysis and synthetic evaluation of the main qualities of vegetable sweetpotato advanced selections. *Plant Sci J*. 2016;34:614-21. *In Chinese*

4. Ishida H, Suzuno H, Sugiyama N, Innami S, Tadokoro T, Maekawa A. Nutritive evaluation on chemical components of leaves, stalks and stems of sweet potatoes (*Ipomoea batatas* *poir*). Food Chem. 2000;68:359-67.
5. Sun H, Mu T, Xi L, Zhang M, Chen J. Sweet potato (*Ipomoea batatas*) leaves as nutritional and functional foods. Food Chem. 2014;156:380-9.
6. Hagerman AE, Riedl KM, Jones GA, Sovik KN, Ritchard NT, Hartzfeld PW, et al. High molecular weight plant polyphenolics (*Tannins*) as biological antioxidants. J Agric Food Chem. 1998;46:1887-92.
7. Robards K, Prenzler PD, Tucker G, Swatsitang P, Glover W. Phenolic compounds and their role in oxidative processes in fruits. Food Chem. 1999;66:401-36.
8. Manach C, Scalbert A, Morand C, Rémésy C, Jiménez L. Polyphenols: food sources and bioavailability. Am J Clin Nutr. 2004;79:727-47.
9. Johnson M, Pace RD. Sweet potato leaves: properties and synergistic interactions that promote health and prevent disease. Nutr Rev. 2010;68:604-15.
10. Ojong PB, Njiti V, Guo Z, Gao M, Besong S, Barnes SL. Variation of flavonoid content among sweetpotato accessions. J Am Soc Hortic Sci. 2008;133:819-24.
11. Zheng W, Clifford MN. Profiling the chlorogenic acids of sweet potato (*Ipomoea batatas*) from China. Food Chem. 2008;106:147-52.
12. Xu W, Liu L, Hu B, Sun Y, Ye H, Ma D, et al. TPC in the leaves of 116 sweet potato (*Ipomoea batatas*) varieties and Pushu 53 leaf extracts. J Food Comp Anal. 2010;23:599-604.
13. Jung J-K, Lee S-U, Kozukue N, Levin CE, Friedman M. Distribution of phenolic compounds and antioxidative activities in parts of sweet potato (*Ipomoea batata*) plants and in home processed roots. J Food Comp Anal. 2011;24:29-37.
14. Niggeweg R, Michael AJ, Martin C. Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat Biotechnol. 2004;22:746-54.
15. Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. Plant Cell. 1995;7:1085-97.
16. Douglas CJ. Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. Trends Plant Sci. 1996;1:171-8.
17. Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, et al. Silencing of hydroxycinnamoyl-coenzyme a shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. Plant Cell. 2004;16:1446-65.
18. Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M. Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. J Biol Chem. 2003;278:95-103.
19. Villegas RJ, Kojima M. Purification and characterization of hydroxycinnamoyl D-glucose. Quinate hydroxycinnamoyl transferase in the root of sweet potato, *Ipomoea batatas* J Biol Chem. 1986;261:8729-33.
20. Kojima M, Uritani I. Studies on chlorogenic acid biosynthesis in sweet potato root tissue using trans-cinnamic acid-2-<sup>14</sup>C and quinic acid-G-<sup>3</sup>H. Plant Cell 1972;13:311-9.
21. Ulbrich B, Zenk MH. Partial purification and properties of hydroxycinnamoyl-CoA: quinate hydroxycinnamoyl transferase from higher plants. Phytochemistry. 1979;18:929-33.
22. Rhodes MJC, Woollorton LSC. The enzymic conversion of hydroxycinnamic acids to p-coumarylquinic and chlorogenic acids in tomato fruits. Phytochemistry. 1976;15:947-51.
23. Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, et al. CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. J Biol Chem. 2001;276:36566-74.
24. Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, et al. The *Arabidopsis* REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J. 2002;30:33-45.
25. Strack D, Gross W. Properties and activity changes of chlorogenic acid:glucaric acid caffeoyltransferase from tomato (*Lycopersicon esculentum*). Plant Physiol. 1990;92:41-7.
26. Zhang J, Wu M, Li W, Bai G. Regulation of chlorogenic acid biosynthesis by hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase in *Lonicera japonica*. Plant Physiol Biochem. 2017;121:74-9.
27. Mudau S, Steenkamp P, Piater L, De Palma M, Tucci M, Madala N, et al. Metabolomics-guided investigations of unintended effects of the expression of the *hydroxycinnamoyl quinate hydroxycinnamoyltransferase (hqf1)* gene from *Cynara cardunculus* var. *scolymus* in *Nicotiana tabacum* cell cultures. Plant Physiol Biochem. 2018;127:287-98.
28. Bartley GE, Avena-Bustillos RJ, Du W-X, Hidalgo M, Cain B, Breksa AP. Transcriptional regulation of chlorogenic acid biosynthesis in carrot root slices exposed to UV-B light. Plant Gene. 2016;7:1-10.
29. Zhang J, Yang Y, Zheng K, Xie M, Feng K, Jawdy SS, et al. Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of *HCT2* in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in *Populus*. New Phytol. 2018;220:502-16.

30. Sun M, Shi M, Wang Y, Huang Q, Yuan T, Wang Q, et al. The AP2/ERF transcription factor SmERF115 positively regulates the biosynthesis of phenolic acids in *Salvia miltiorrhiza*. *J Exp Bot*. 2019;70:243-54.
31. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009;136:215-33.
32. Lin JS, Lin CC, Lin HH, Chen YC, Jeng ST. MicroR828 regulates lignin and H<sub>2</sub>O<sub>2</sub> accumulation in sweet potato on wounding. *New Phytol*. 2012;196:427-40.
33. Sun R, Guo T, Cobb J, Wang Q, Zhang B. Role of microRNAs during flower and storage root development in sweet potato. *Plant Mol Biol Rep*. 2015;33:1731-9.
34. Bian X, Ma P, Jia Z, Guo X, Xie Y. Identification of miRNAs in sweet potato by Solexa sequencing. *Russ J Plant Physiol*. 2016;63:283-92.
35. Xie Z, Wang A, Li H, Yu J, Jiang J, Tang Z, et al. High throughput deep sequencing reveals the important roles of microRNAs during sweetpotato storage at chilling temperature. *Sci Rep*. 2017;7:16578.
36. Wu S, Lau KH, Cao Q, Hamilton JP, Sun H, Zhou C, et al. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nat Commun*. 2018;1-12.
37. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12:357-60.
38. Altschul SF, Madden TL, Schaffer 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;25:3389-402.
39. Letunic I, Yamada T, Kanehisa M, Bork P. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci*. 2008;33:101-3.
40. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012;40:37-52.
41. He L, Tang R, Shi X, Wang W, Cao Q, Liu X, et al. Uncovering anthocyanin biosynthesis related microRNAs and their target genes by small RNA and degradome sequencing in tuberous roots of sweetpotato. *BMC Plant Biol*. 2019;19:232.
42. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr Biol*. 2008;18:758-62.
43. Addo-Quaye C, Miller W, Axtell MJ. CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*. 2009;25:130-1.
44. Chen Z, Tang N, You Y, Lan J, Liu Y, Li Z. Transcriptome analysis reveals the mechanism underlying the production of a high quantity of chlorogenic acid in young leaves of *Lonicera macranthoides* Hand.- PLoS One. 2015;10(9):e0137212.
45. Tang J, Chu C. MicroRNAs in crop improvement: fine-tuners for complex traits. *Nat Plants*. 2017;3:17077.
46. Sun P, Cheng C, Lin Y, Zhu Q, Lin J, Lai Z. Combined small RNA and degradome sequencing reveals complex microRNA regulation of catechin biosynthesis in tea (*Camellia sinensis*). *PLoS One*. 2017;12:e0171173.
47. Gou JY, Felippes FF, Liu CJ, Weigel D, Wang JW. Negative regulation of anthocyanin biosynthesis in *Arabidopsis* by a miR156-targeted SPL transcription factor. *Plant Cell*. 2011;23:1512-22.
48. Jia X, Shen J, Liu H, Li F, Ding N, Gao C, et al. Small tandem target mimic-mediated blockage of microRNA858 induces anthocyanin accumulation in tomato. *Planta*. 2015;242:283-93.
49. Ku YS, Wong JW, Mui Z, Liu X, Hui JH, Chan TF, et al. Small RNAs in plant responses to abiotic stresses: regulatory roles and study methods. *Int J Mol Sci*. 2015;16:24532-54.
50. Fang YN, Zheng BB, Wang L, Yang W, Wu XM, Xu Q, et al. High-throughput sequencing and degradome analysis reveal altered expression of miRNAs and their targets in a male-sterile cybrid pummelo (*Citrus grandis*). *BMC Genomics*. 2016;17:591.
51. Liu R, Lai B, Hu B, Qin Y, Hu G, Zhao J. Identification of microRNAs and their target genes related to the accumulation of anthocyanins in *Litchi chinensis* by high-throughput sequencing and degradome analysis. *Front Plant Sci*. 2016;7:2059.
52. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
53. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
54. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
55. German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, et al. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotechnol*. 2008;26:941-6.
56. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔCT</sup> method. *Methods*. 2001;25:402-8.

## Table

**Table 1.** Summary of transcriptome and small RNA sequencing data.

Sample	Sample replicate	Transcriptome sequencing		Small RNA sequencing							
		No. of clean reads	Percentage of mapped reads	raw reads	No. of clean reads	17 nt < Reads <33 nt	Repeat reads	rRNA reads	tRNA reads	snRNA reads	miRNA reads
ES1	L01	174683830	76.32	25927605	24095270	24094597	133724	208381	17766	446	106044
	L02	166992924	79.98	23865315	22321800	22320871	141021	395043	14084	684	148547
	L03	149408286	79.82	32115499	29998859	29997913	187042	344301	19841	648	211084
FS2	L04	138978770	80.06	21378200	20333242	20330797	92533	905305	97290	2787	286047
	L05	141222622	84.71	23317399	22270590	22266892	102186	486652	50227	1061	499324
	L06	139989810	81.51	15117444	14387309	14385108	58805	292689	49024	1352	285295
ES2	L07	153472418	80.64	26286171	24518825	24518144	159900	1034571	50200	1601	171786
	L08	142547416	76.26	29131977	27250685	27248979	148601	296936	36413	631	365179
	L09	128383618	78.67	18473381	17240894	17240236	102292	209377	32750	541	201224
FS1	L11	154290016	80.96	24627110	23088334	23087243	113446	296534	17196	470	227489
	L12	185704656	79.65	24673614	23078309	23077146	113972	316806	12650	565	211848
Total		1675674366		264913715	248584117	248567926	1353522	4786595	397441	10786	2713867

Notes: E denoted EC16; F denoted Fushu No. 7-6; S1 denoted 65 days after planting; S2 denoted 85 days after planting. L01-L12 denoted different samples.

## Supplementary Files Section

Additional file 1. Table S1 DEGs identified across the four comparisons.

Additional file 2. Table S2 Detailed annotation information of the four comparisons Additional file 3. Table S3 GO term analysis of DEGs

Additional file 4. Table S4 KEGG analysis of DEGs

Additional file 5. Table S5 Known and novel miRNAs identified across the four comparisons

Additional file 6. Table S6 DE miRNA expression patterns

Additional file 7. Table S7 The annotation of identified miRNAs targets

Additional file 8. Table S8 GO enrichment analysis of the identified miRNAs targets

Additional file 9. Table S9 miRNA targets identified from degradome sequencing Additional

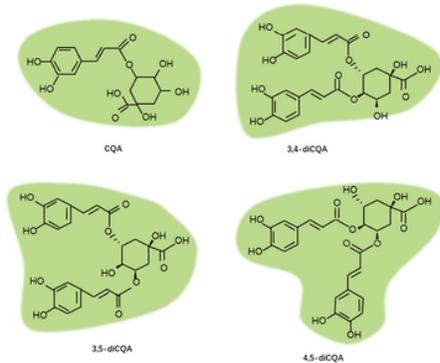
Additional file 10. Table S10 Association analysis between DE miRNAs and mRNAs

Additional file 11. Table S11 Sequence information of mRNAs and miRNAs for qRT-PCR analyses

Additional file 12. Figure S1. Results of RNA-seq transcriptome assembly. **a** Numbers and levels of expressed genes from different samples. **b** Correlation-based clustering analyses of RNA-seq expression across all replicates.

## Figures

a



b

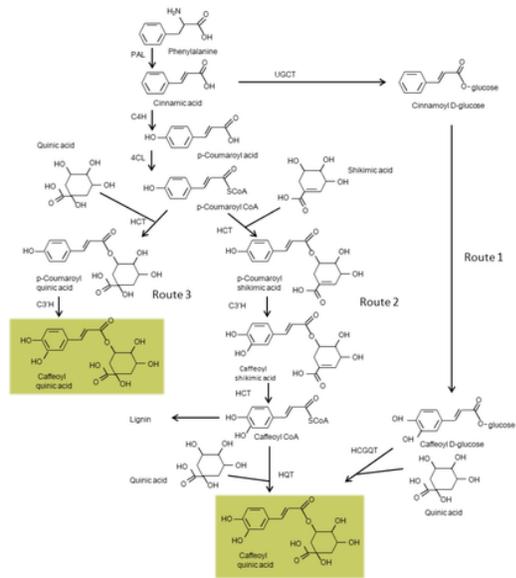
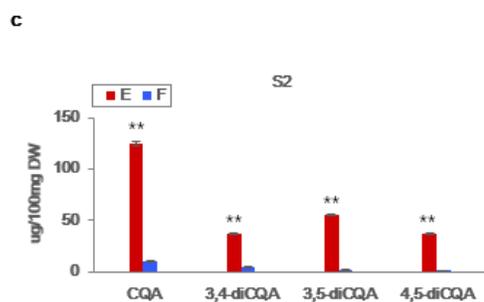
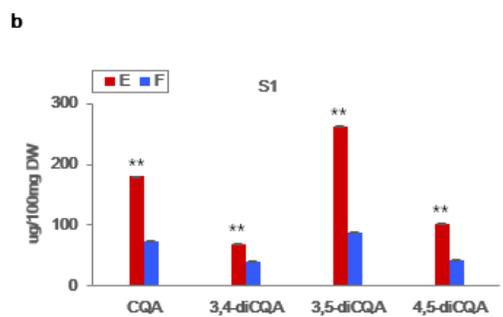
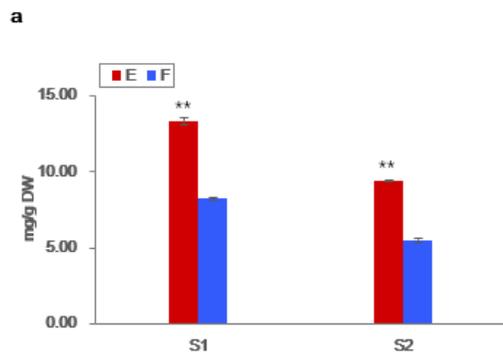


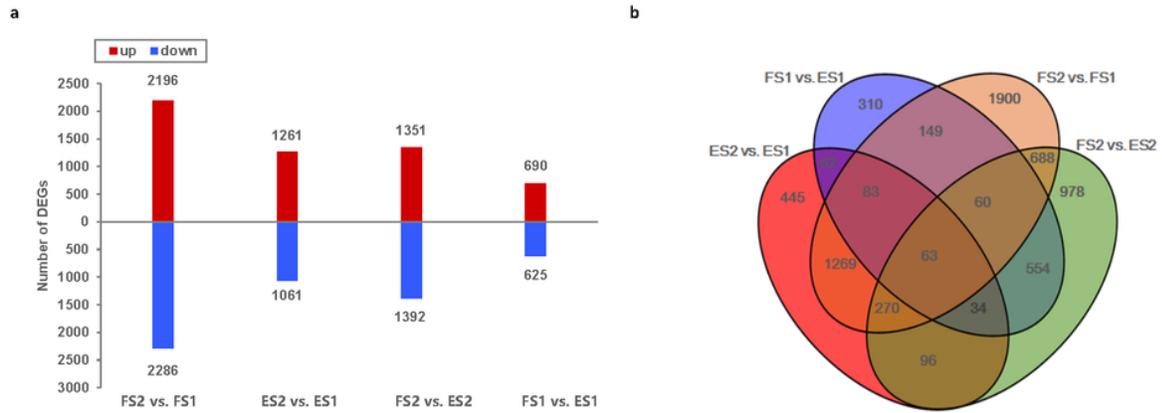
Figure 1

Specialized CGAs found in leafy sweet potato. a Structures of major CGAs. b. The proposed pathway responsible for the CGA biosynthesis [12].



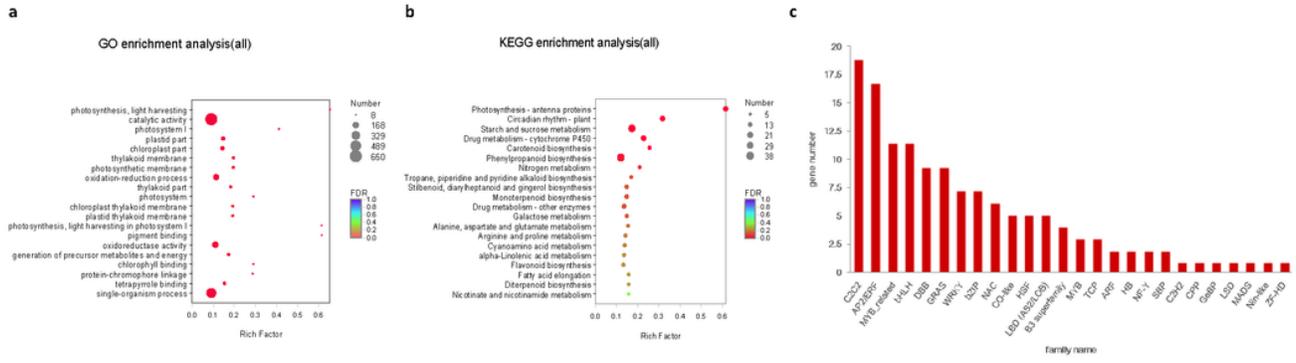
**Figure 2**

Measurements of TP and CGA contents between genotypes E and F at stages S1 and S2. a Measurements of TPC in genotypes E and F at stages S1 and S2. b Measurements of CGA monomers in genotypes E and F at S1. c Measurements of CGA monomers in genotypes E and F at S2. Error bars indicated  $\pm$ SD (N=3). \*\* T-TEST, P<0.01.



**Figure 3**

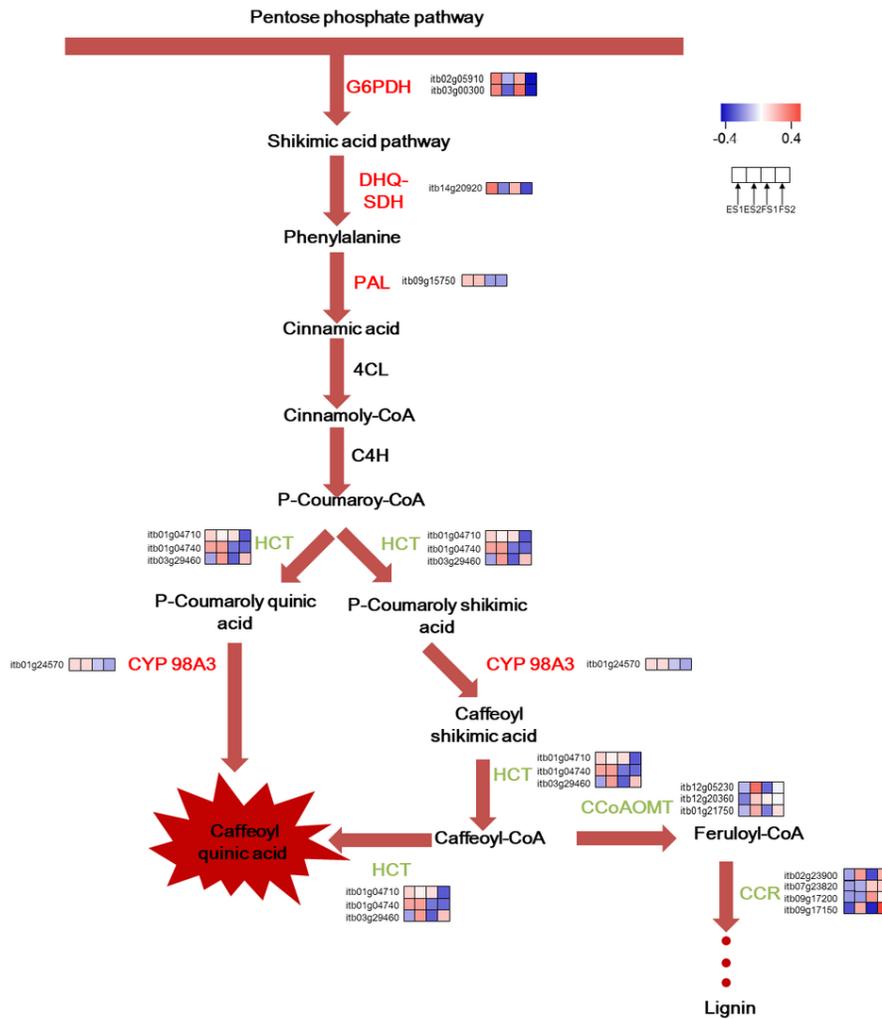
DE mRNA analyses. a Numbers and levels of DEGs in the four comparisons. b Venn diagram of DE mRNAs among the four comparisons.



**Figure 4**

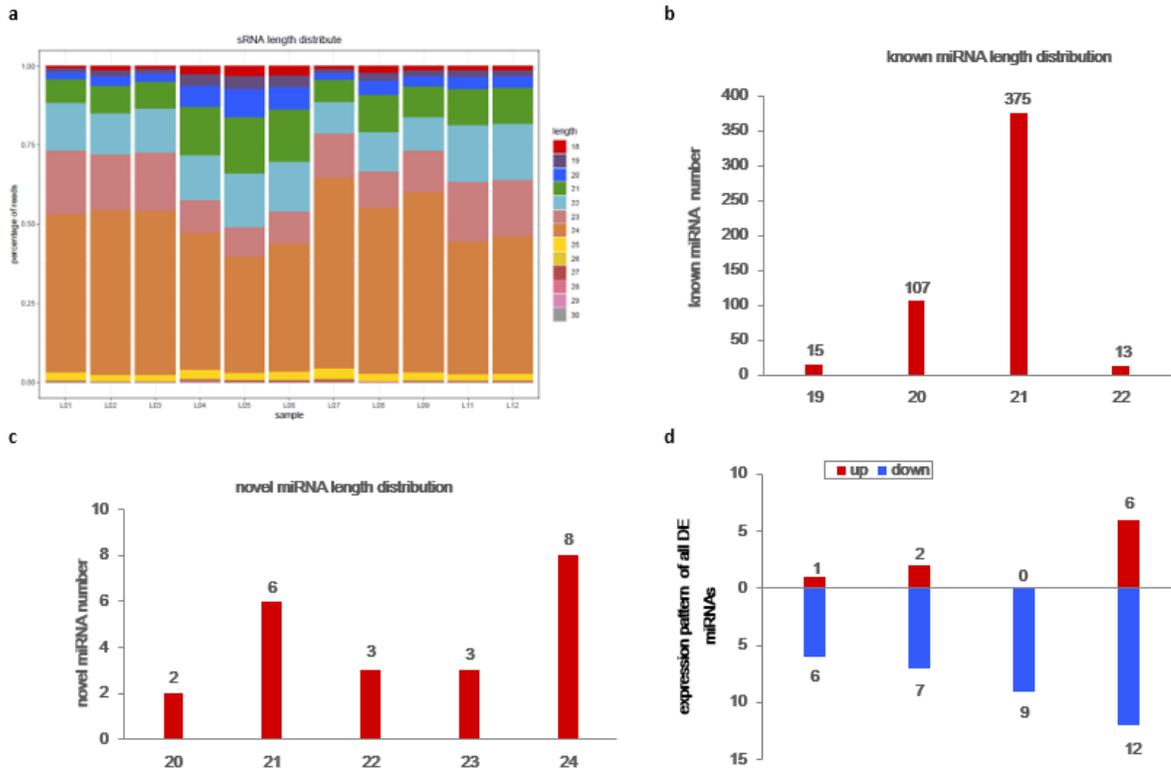
Annotation and pathway analyses of the identified DEGs. a GO enrichment analysis of common DEGs. b The scatter plot of enriched KEGG pathways of DEGs. c Number of genes from TF families in the common DEGs.



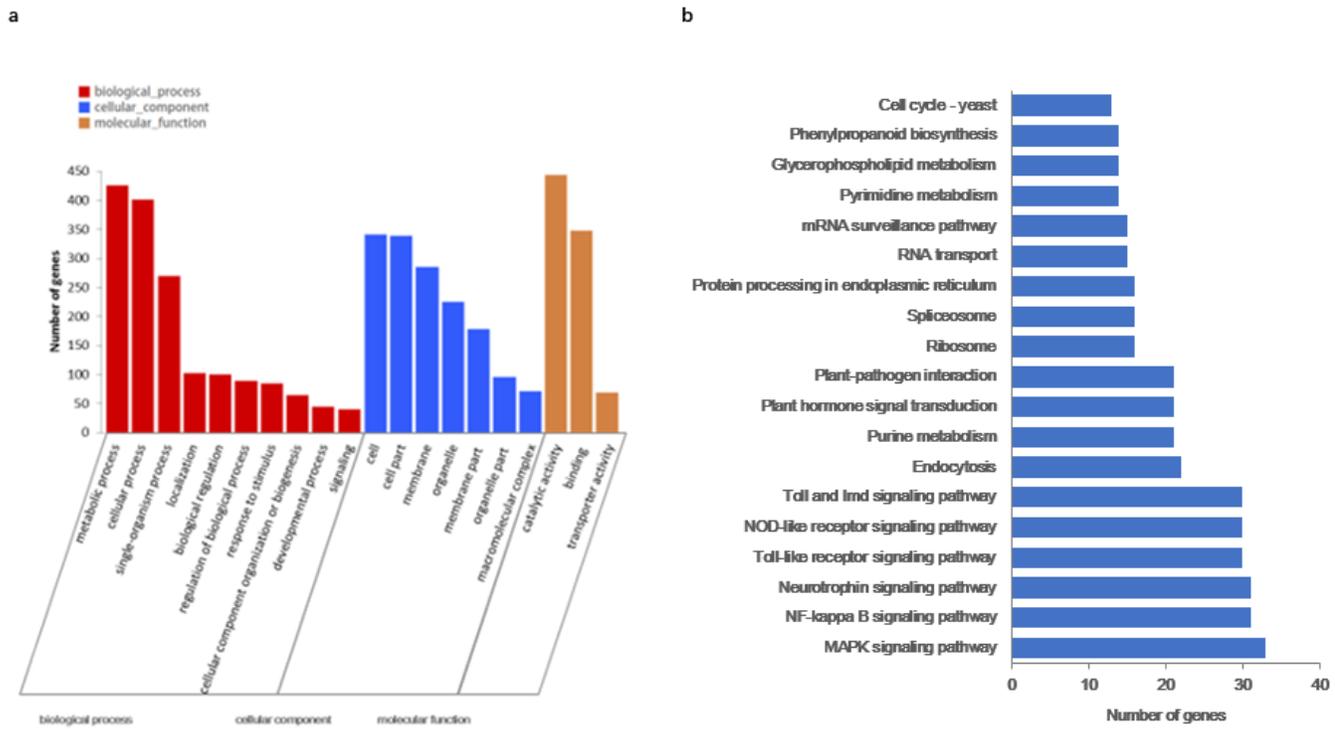


**Figure 6**

A hypothetical representation of the expression patterns of genes involved in the CGA biosynthetic pathway in leafy sweet potato. Upregulated Genes were shown in red, downregulated genes in green and non-DEGs in black.

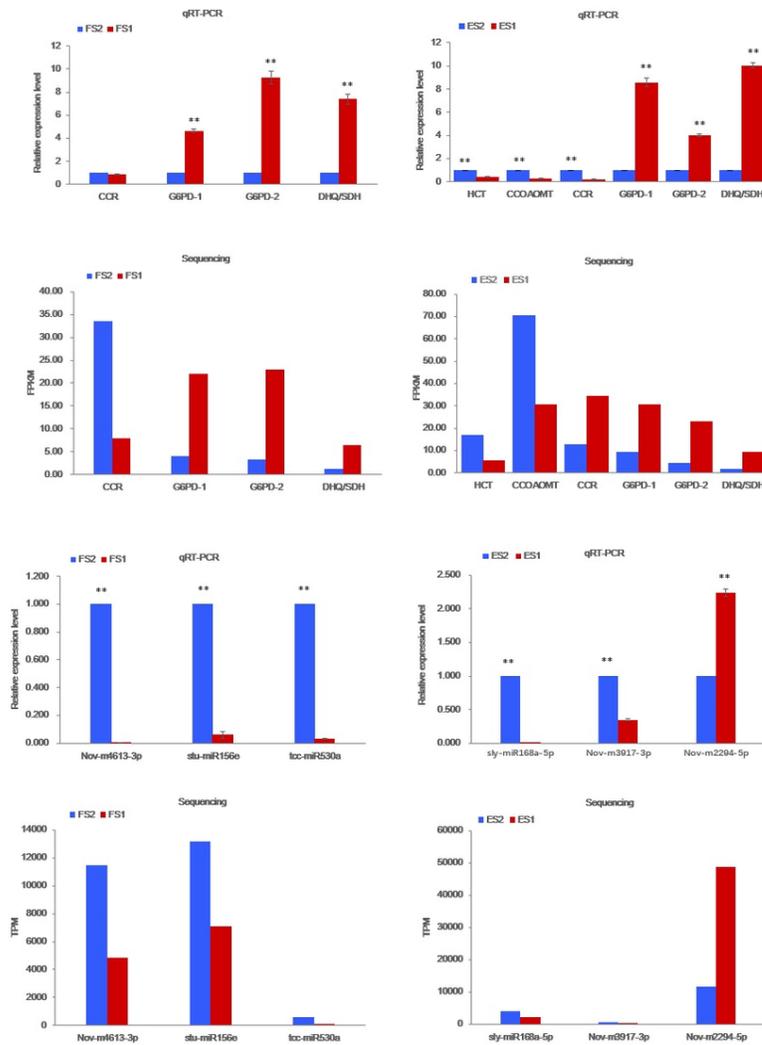


**Figure 7**  
 Size distribution of miRNAs by sRNA-seq. a Length distribution of sRNAs. b Numbers of length distribution of known miRNAs. c Numbers of length distribution of novel miRNAs. d Expression pattern of DE miRNAs.



**Figure 8**

DE miRNA target genes' annotations. a GO analysis of annotated DE miRNA target genes. b Major pathways of all annotated DE miRNA target genes.



**Figure 9**

Expression patterns of mRNAs and miRNAs determined by qRT-PCR. Expression levels of mRNAs and miRNAs were normalized according to the level of  $\beta$ -actin in qRT-PCRs. Three biological replicates were performed for each mRNA and miRNA. T-TEST was employed to analyse the data. The data were presented as the mean  $\pm$  SD. \*\* $P < 0.01$

## Supplementary Files

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

- [TableS11.xlsx](#)
- [FigS1.docx](#)
- [TableS10.xlsx](#)
- [TableS5.xlsx](#)
- [TableS8.xlsx](#)
- [TableS2.xlsx](#)
- [TableS1.xlsx](#)
- [TableS9.xls](#)
- [TableS7.xlsx](#)

- [TableS4.xlsx](#)
- [TableS6.xlsx](#)
- [TableS3.xlsx](#)