Genotypic variations in GST genes reveal a regulatory role in the accumulation of caffeoylquinic acids in leafy sweet potato

Deka Reine Judesse SOVIGUIDI

Yangtze University

Yi Liu

Yangtze University

Rui Pan

Yangtze University

Salah Fatouh Abou-Elwafa

Assiut University

Wenying Zhang (✉ wyzhang@yangtzeu.edu.cn)

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

Xinshun Yang

Hubei Academy of Agricultural Sciences

Research Article

Keywords: Leafy sweet potato, Caffeoylquinic acids, Glutathione S-transferase, Gene expression, Abiotic stress

Posted Date: February 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1213676/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Leafy sweet potato is rich in caffeoylquinic acids (CQAs), including monoCQAs and diCQAs, which are important for plant resistance to environmental stresses. Knowledge about genes responsible for the accumulation of these CQAs is limited. However, the implication of *glutathione S-transferase* (*GST*) genes in the metabolism of plant secondary compounds has been reported. Yet, the mechanism of CQA-related *GST* proteins is not well understood. In this study, two sweet potato *GST* genes i.e., *itb01g35330* (*lbGSTTCHQD*) and *itb09g30700* (*lbGSTT*) were selected from our transcriptome database as they exhibited secondary metabolism functions based on gene ontology classification, and investigated for their association with the accumulation of CQAs. Sequence comparison of the *lbGST* coding regions amplified from EC16 and FS7-6 leafy sweetpotato varieties revealed some mutations at different sites in the amino acid sequences between the two varieties. Besides, the genes were highly expressed in EC16 compared to FS7-6, which corresponded to the amounts of CQAs in both varieties. The amplified sequences from EC16 were further transformed into *Agrobacterium tumefaciens* and subsequently introduced into *Nicotiana benthamiana* and *Nicotiana tabacum* for transient expression and transgenic transformation, respectively. Gene expression profiling in both experimental methods revealed significant increases of *lbGST* transcripts in transformed leaves, resulting in enhanced amount of monoCQAs compared to the wild types. However, no significant change was observed in the level of diCQAs. Moreover, the *lbGSTs* responded positively to salinity and oxidative stresses. Our findings suggested that *lbGSTTCHQD* and *lbGSTT* genes might be involved in the accumulation of monoCQAs in leafy sweet potato.

Key Message

Two sweet potato glutathione S-transferase genes (*lbGSTs*) were involved in the accumulation of mono-caffeoylquinic acids (monoCQAs) and tobacco transgenic plants resistance to salinity and oxidative stresses.

Introduction

Sweet potato (*Ipomoea batatas*) is an important crop with nutritional properties beneficial for animal and human consumption (Alam 2021; Martino et al. 2021). In China, its tuberous roots have long been considered the only edible parts of the plant, however, more recently the leaves and tender stems of leafy sweet potato are increasingly consumed as fresh vegetables, making the production of these new cultivars with improved foliar biofortification profitable for farmers (Tang et al. 2021). Leafy sweet potato is rich in vitamins, proteins, carotenoids, flavonoids, anthocyanins and caffeoylquinic acids (CQAs) (Suárez et al. 2020; Liu et al. 2021). The CQAs, including monoCQAs (1-CQA, 3-CQA, 4-CQA, 5-CQA) and diCQAs (1,3-diCQA, 1,4-diCQA, 1,5-diCQA, 3,4-diCQA, 3,5-diCQA, 4,5-diCQA) possess antioxidant properties, which are important for the treatment of chronic human diseases and the resistance of plants to environmental adversaries (Naveed et al. 2018; Cheevarungrapakul et al. 2019). In view of these different properties, increasing the CQA contents in leafy sweet potato would be of great benefit.

Among the CQAs, 5-CQA (chlorogenic acid) is the predominant phenolic acid in plants such as *Nicotiana tabacum* (tobacco), *Solanum lycopersicum* (tomato), *Coffea arabica* (Coffee), *Malus domestica* (apple), etc. (Soviguidi et al. 2022). Its biosynthesis begins in the cytoplasm via the pathway of phenylpropanoids (PPs), and the acid is further transported and stored in the vacuoles (Li et al. 2019). However, little is known about the genes encoding proteins that are implicated in the vacuolar accumulation of CQAs. It has been reported that *glutathione S-transferase* (*GST*), a group of proteins divided into 8 subfamilies namely, eukaryotic translation elongation factor 1 gamma (EF1By), Tau (U), Zeta (Z), dehydroascorbate reductase (DHAR), Lambda (L), tetrachlorohydroquinone dehalogenase (TCHQD), Phi (F) and Theta (T), is involved in the sequestration of several secondary metabolites and plant detoxification under harmful stress conditions (Czerniawski and Bednarek 2018; Estévez and Hernández 2020; Hao et al. 2021). For example, *OsGSTU30* a member of Tau subfamily was shown to enhance *Arabidopsis thaliana* tolerance to drought stress (Srivastava et al. 2019); whereas *CaGSTL1* belonging to Lambda subfamily improved the resistance of chickpea to salinity stress (Ghangal et al. 2020). *GSTs* also act as carriers of flavonoids and anthocyanins. They bind to anthocyanins forming the complexes GST-anthocyanins, which transport the proteins to vacuoles for storage (Pérez-Díaz et al. 2016; Biała and Jasiński 2018; Lin et al. 2020). The positive correlations between the expressions of *CmGST1*, *PpGST1* and *IbGSTF4* genes with the accumulation of anthocyanins has been reported in *Chrysanthemum morifolium* (Li et al. 2021), *Prunus persica* (Zhao et al. 2020) and *Arabidopsis thaliana* (Kou et al. 2019), respectively. In addition, *CsGSTa* and *CsGSTb* were found to be related to proanthocyanins and flavonols contents in *Camellia sinensis* (Liu et al. 2019). However, the CQA-related GST protein mechanism has not been extensively studied. Moglia *et al.*, demonstrated that the silencing of a *GST* gene increased the level of diCQAs in tomato (Moglia et al. 2011). Nevertheless, the role of *GSTs* in the accumulation of CQAs remains to be addressed. We hypothesize that, similar to anthocyanins, CQAs also bind to GST enzymes that would mediate their vacuolar sequestration through GST-CQA complexes.

Previously, we identified 51 *GST* genes in the leafy sweet potato varieties EC16 and FS7-6 using transcriptomic datasets. These genes were classified into TCHQD, Theta, Lambda, Tau and EF1By subfamilies based on comparison with *Arabidopsis* *GST* proteins. Different members of each subfamily, exhibiting secondary metabolism functions according to the classification of gene ontology were preselected to investigate their roles in the accumulation of CQAs. Two members i.e., *itb01g35330* and *itb09g30700* belonging to TCHQD and Theta subfamilies were finally selected for further analysis as their sequences were confirmed in both varieties. The full-length coding sequences of the genes were identified. Their levels of expression were analyzed in different plant tissues of EC16 and FS7-6 varieties, which possess contrasting phenotypes in terms of CQA contents and tolerance to abiotic stresses. The genes were further characterized by transient and transgenic transformations in *N. benthamiana* and *N. tabacum*, respectively. Moreover, the transcript abundance of the two genes in transgenic and wild type *N. tabacum* plants was analyzed under salinity and oxidative stress conditions.

Materials And Methods

Plant materials and growing conditions

Leafy sweet potato varieties EC16 and FS7-6 were grown in Food Crops Institute, Hubei Academy of Agricultural Sciences, Wuhan, China. Twenty days after planting, tissues from stem tips, stems, leaves and roots were collected and immediately frozen in liquid nitrogen. The samples were stored in refrigerator (-80°C) for further CQAs content estimation and gene expression profiling. *Nicotiana benthamiana* and *Nicotiana tabacum* L. cv. Wisconsin-38 were used for transient and transgenic transformations, respectively. For evaluation under abiotic stress, the seedlings of wild type *N. tabacum* and transgenic lines overexpressing *lbGSTTCHQD* and *lbGSTT* were grown on 1/2 Murashige and Skoog (MS) medium containing 5% H₂O₂ and 200 mM NaCl for oxidative and salinity stress treatments, respectively. Three replications were set for each treatment. The plants growing conditions were 26 ± 2 °C of temperature, 8 hours of dark and 16 hours of light. After 40 days, leaves were collected and relative expression levels of the *lbGST* genes were analyzed.

Bioinformatic analyses and identification of *GST* genes

A number of 51 *GST* genes divided into TCHQD, Theta, Lambda, Tau and EF1By subfamilies were previously identified in the EC16 and FS7-6 varieties based on transcriptome database (NCBI SRA, accession number PRJNA592001), and comparison with Arabidopsis *GST* proteins. Among these genes, different members from each subfamily possessing secondary metabolism functions were first preselected to dissect their roles in the accumulation of CQAs. Two genes i.e., *itb01g35330* and *itb09g30700* belonging to the TCHQD and Theta subfamilies were finally selected as their sequences were confirmed in EC16 and FS7-6. The genes were designated as *lbGSTTCHQD* and *lbGSTT*, and their amino acid sequences retrieved from the Sweetpotato Genomic Resources (<http://sweetpotato.uga.edu/>) were used to perform multiple alignment sequences with GST proteins from other plant species using ClustalX2. MEGA7 software was employed to generate the phylogenetic tree. The online tool PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) was used for identification of *cis*-elements in the promoter regions of the *lbGST* genes. The sequence information of the *lbGST* genes is presented in Additional file 1 Table S1.

Extraction of genomic DNA, total RNA and gene expression analysis

The Cetyltrimethylammonium bromide method was used for the isolation of genomic DNA. TransZol Up and TransScript® All-in-One kits were employed for the extraction of total RNA and first-strand cDNA following the instructions of the manufacturer (TransGen Biotech, China). The cDNA was 10-times diluted for gene expression profiling using the RT-qPCR. The expression analysis was performed using TransStart® Tip Green qPCR SuperMix, with thermal cycling conditions of 94 °C for 30s, 40 cycles of 94 °C for 5s, 56 °C for 15s and 72 °C for 10s. The relative expression levels of the *lbGST* genes were determined in triplicate using 2^{-ΔΔCT} method. β-actin was used as a reference gene. The primer sequences of the *lbGSTs* are presented in Additional file 2: Table S2.

Determination of CQA contents

The CQAs were quantified using the method described by Liu et al. (Liu et al. 2021) with few modifications. Briefly, 50 mg of each sample was dissolved in 70% ethanol (v/v) followed by water bath (60°C, 1h) and then centrifugation (12000 rpm, 10 min). The supernatant was further filtered, and 20 µL of the reaction was analyzed using the High-performance liquid chromatography (HPLC) Agilent 1260 system (Agilent Technologies Inc., USA) as described by the authors. The 3-CQA, 3,4-diCQA, 3,5-diCQA, 4-CQA, 4,5-diCQA and 5-CQA were detected at 326 nm and their levels were determined based on comparison with the peak area and the retention time of standard CQAs obtained from Sigma-Aldrich (USA).

Sequence amplification of *lbGSTs* and vectors construction

First, the complete coding sequences of *lbGSTTCHQD* (981 bp) and *lbGSTT* (723 bp) were amplified using cDNAs from the leaves of EC16 and FS7-6 as templates. PCR amplification was performed with specific primers designed for each gene. The products were recovered and cloned into the TaKaRa pMD19-T vector, and the resulting plasmids pMD19-*lbGSTs* were sequenced. Each gene was confirmed in EC16 and FS7-6, and the alignment of their sequences exhibited slight difference of amino acids between the two varieties. The amplified coding sequences from EC16 were ultimately selected for further experiments as this variety accumulated high amounts of CQAs compared to FS7-6. Based on pCambia1300-GFP overexpression vector, specific restriction sites corresponding to each gene were determined and primers were designed to amplify the *lbGSTs* from the pMD19-*lbGSTs*. The PCR products were subsequently digested with Kpn1/BamHI and BamHI/XbaI for *lbGSTTCHQD* and *lbGSTT*, respectively, and ligated to the corresponding restriction enzyme sites of the pCambia1300-GFP vector. Each clone was grown overnight under 200 rpm at 37°C in 30 ml of Luria-Bertani (LB) liquid medium containing 50 mg/L kanamycin (Kan). The constructed pCambia1300-*lbGST* overexpression vectors were further introduced into the *Agrobacterium tumefaciens* EHA105 competent cells using the freeze-thaw method. *A. tumefaciens* containing each construct was grown at 28°C in LB medium containing 100 mg/L rifampicin and 50 mg/L Kan under 200 rpm. The integrity of the inserted genes and vectors was confirmed by sequencing.

Transient expression of *lbGSTs*

The *A. tumefaciens* suspensions harboring each gene of interest were centrifuged for 10 min at 5000 rpm. Cells were collected, washed twice, and resuspended in a pH = 5.6 buffer containing 10 mM MES and 10 mM MgCl₂. Optical densities (OD) were further measured at 0.6 to OD600 and the bacterial solutions were incubated at room temperature for 2 h. Acetosyringone (100 mg/L) was added, and the culture solutions were injected into the leaves of 4-weeks old *N. benthamiana* plants. Five days later, the leaves were harvested for HPLC and gene expression profiling.

Plant transformation

The *N. tabacum* transgenic plants were generated using *Agrobacterium*-mediated gene transfer system. *Agrobacterium* cells containing *lbGSTTCHQD* and *lbGSTT* constructs were harvested and washed twice

with MS liquid medium containing 2 mg/L 2,4-D to remove residual LB medium and toxicants. The pellets were then resuspended in the same medium supplemented with 100 mg/L acetosyringone, and OD were measured at 0.6 to OD600. *N. tabacum* seeds were sterilized for 10 min in 2% sodium hypochlorite and cultivated on solid MS medium after rinsing with distilled water. Two-week old leaves were further collected for genetic transformation. The leaves were cut into 1×1 cm pieces and pre-cultured in dark for 3 days on M1 medium containing MS solid medium, 2.0 mg/L 6-benzyladenine (6-BA) and 0.5 mg/L naphthalene acetic acid (NAA). After pre-culture, the leaves were soaked in *A. tumefaciens* suspensions containing the constructs for 10 min with constant shaking. The infected leaves were co-cultured in dark on M1 medium for 3 days and the explants were rinsed three times with MS liquid mediums containing 600, 300 and 150 mg/L cefotaxime (Cef), respectively. Afterwards, they were transferred to M2 selective shooting medium containing solid MS, 2 mg/L 6-BA, 0.5 mg/L NAA, 600 mg/L Cef and 100 mg/L Kan. Four weeks later, kanamycin-resistant shoots were cultivated in M3 selective rooting medium containing 1/2 solid MS supplemented with 2 mg/L 6-BA, 0.1 mg/L indole-3-acetic acid, 600 mg/L Cef and 100 mg/L Kan. Ten independent well-rooted explants were randomly selected for the PCR detection of positive transgenic lines. Relative expressions of *lbGSTs* were determined and the lines with higher transcriptional levels were selected for further analysis.

Statistical analyses

Data were analyzed with SPSS software version 25 (IBM, USA) using analysis of variance (one-way). Duncan's multiple-range test ($p < 0.05$) and Student's t-test ($p < 0.05$) were employed for means comparisons. OriginPro 9.0 software was used to generate the graphs.

Results

CQAs content in sweet potato plant tissues

HPLC was performed to investigate the content of CQAs in the leafy sweet potato varieties EC16 and FS7-6. The analysis revealed the presence of six CQAs, including 3-CQA, 3,5-diCQA, 3,4-diCQA, 4,5-diCQA, 4-CQA and 5-CQA in different tissues of stem tips, stems, leaves and roots. These CQAs are highly accumulated in stems and leaves compared to stems tips and roots in both varieties (Fig. 1A). The 3,5-diCQA is the predominant acid in stems followed by 3-, 5- and 4-CQA. Meanwhile, in stem tips and leaves, 3,5-diCQA and 5-CQA are the main CQAs, whereas the roots accumulated more 4- and 5-CQA. Conversely, 3,4- and 4,5-diCQA are the lowest CQAs in all tissues of both varieties. Notably, 5-CQA is one of the major CQAs that is present in large quantity simultaneously in all tissues of EC16 and FS7-6. As shown in Fig. 1B, the total amount of CQAs is significantly higher in EC16 than in FS7-6. It could be speculated that the content of CQAs varied among cultivars and plant tissues.

Phylogenetic analysis of *lbGST* genes

For investigation of the evolutionary relationships, orthologs of *lbGSTTCHQD* and *lbGSTT* in other plant species were identified using NCBI BLASTP. Phylogenetic tree analysis indicated that *lbGSTT* protein is

closely related to the *GST* from *Nicotiana attenuata* (XP_019264208.1:1-231), whereas *lbGSTTCHQD* protein is a homologous to the GST protein of *Theobroma cacao* (EOY15086.1:1-322) (Fig. 2). This result showed close homology and domain similarity between the *lbGSTs* and GST proteins from other plant species.

Analysis of *lbGST* promoter regions

To identify relevant factors that regulate the expression of the *GST* genes, the upstream 2 kb of transcription start site of *lbGSTTCHQD* and *lbGST* was analyzed using the PlantCARE online tool. Various cis-regulatory elements, including core promoter element of light-responsive elements (G-box), transcription start (TATA-box), abscisic acid responsive elements (ABRE) and other environmental stress responsive-elements such as DRE, MYB and MYC were identified in the promoters of the *lbGST* genes (Table 1). The ABRE and G-box elements were discovered in the promoter of *lbGSTTCHQD* gene, while the plant metabolism cis-acting regulatory element CAAT-box, TATA-box, and DRE and were found in the promoter of *lbGSTT* gene. Moreover, the transcription factors MYC and MYB, which play crucial roles in PPs metabolism, plant growth and defense mechanisms were found in both genes, indicating that the transcription of the *lbGST* genes might be regulated by diverse physiological and environmental factors, which are essential for the accumulation of CQAs in plant tissues.

Expression profiles of *lbGST* genes

Different transcripts of *lbGST* genes were found in stem tips, stems, leaves and roots of untransformed EC16 and FS7-6 plants. Compared to other tissues, leaves and stems accumulated more CQAs (Fig. 1B). The expression of *lbGSTTCHQD* was highly elevated in the leaves of EC16 compared to the leaves of FS7-6 variety but was most concentrated in stem tips, stems and roots of FS7-6 (Fig. 3A). *lbGSTT* was highly expressed in leaves and roots of both varieties (Fig. 3B). The relative expression levels of these genes varied considerably among tissues and varieties. Our results suggest that the accumulation of CQAs in each of the leafy sweetpotato tissues might not be directly related to the expression pattern of the *lbGST* genes. However, we observed higher expressions of *lbGSTTCHQD* and *lbGSTT* in EC16 compared to FS7-6, which corresponded to the maximum concentration of CQAs in EC16 variety.

Comparison of *lbGST* gene sequences between EC16 and FS7-6

The comparison of protein sequences obtained after amplification of *lbGST* genes from EC16 and FS7-6, showed similarities of more than 95% between the two varieties for each gene. However, slight variations in amino acid sequences were also observed. For examples, the amino acids leucine (L) and tyrosine (Y) in *lbGSTTCHQD* synthesized from the EC16 variety were substituted with phenylalanine (F) and cysteine (C), respectively, in FS7-6 variety (Fig. 4A). Additionally, the arginine (R) and glycine (G) amino acids in *lbGSTT* gene from EC16 were replaced by histidine (H) and valine (V), respectively in FS7-6 (Fig. 4B). These molecular mutations could alter the functions of the target protein in both varieties, and probably the content of CQAs. Since the EC16 variety exhibited higher amount of CQAs, the amplified sequences

from this variety were then used to determine whether the accumulation of CQAs would be correlated with the overexpression of *lbGSTTCHQD* and *lbGSTT*.

***lbGSTTCHQD* and *lbGSTT* enhance CQAs accumulation in *Nicotiana* transformed plants**

Transient expression and genetic transformation of *lbGSTTCHQD* and *lbGSTT* genes in *Nicotiana* plants were employed to evaluate the implication of those two genes in the accumulation of CQAs. Suspensions of the *Agrobacterium tumefaciens* containing the constructed pCambia1300-*lbGST* and pCambia1300-*lbGSTT* overexpression vectors from the EC16 variety were transferred into the *N. benthamiana* leaves. Gene expression profiling revealed significant increases more than 4-fold in the expression levels of *lbGSTTCHQD* and *lbGSTT* genes in the transformed leaves compared to the untransformed leaves (Figs. 5A, B). Furthermore, the overexpression of *lbGSTTCHQD* and *lbGSTT* enhanced the levels of monoCQAs, while the diCQAs were less accumulated in comparison with the untransformed leaves (Figs. 5C, D). For genetic transformation, the coding sequences of *lbGSTTCHQD* and *lbGSTT* driven by the CaMV35S promoter were transformed into *N. tabacum* via *Agrobacterium*-mediated transformation system and kanamycin-resistant explants were generated within 3 months (Figs. 6A, B, C, D). Ten independent transgenic lines were randomly selected, and integration of the transgene was verified by PCR. All the ten selected lines produced the expected products of 981 bp and 723 bp for the integration of *lbGSTTCHQD* and *lbGSTT*, respectively (Figs. 6E, F), indicating that the *lbGSTs* were incorporated into the genome of *N. tabacum*. RT-qPCR analysis exhibited significant increases of *lbGSTTCHQD* and *lbGSTT* transcript levels (3-5-folds) in the transgenic lines compared to the wild type plants (Figs. 6G, H). The three lines that exhibited higher expression levels of each gene were further selected for the quantification of CQAs. Plants expressing the *35S::lbGSTTCHQD* or *35S::lbGSTT* transgenes showed significant accumulation of monoCQAs compared to the wild types. Among those highly accumulated monoCQAs, 5-CQA was the highest (Figs. 6I, J). However, no significant increases were observed in the content of diCQAs except for lines 2 expressing the *35S::lbGSTTCHQD* transgene and line 8 expressing the *35S::lbGSTT* transgene. In general, the *lbGST* genes enhance the accumulation of monoCQAs up to 20% for *lbGSTT* transgene and 55% for *lbGSTTCHQD* transgene, but no association with the diCQAs accumulation was observed (Fig. 6K). These results indicated that the sweet potato *GST* gene expression profiles correlate with the accumulation of monoCQAs in transformed *N. benthamiana* leaves and transgenic *N. tabacum* plants overexpressing either *lbGSTTCHQD* or *lbGSTT*.

Expression profile of *lbGST* genes under abiotic stresses

Seedlings from the three tobacco transgenic lines selected previously based on the observed higher expression levels of *lbGSTTCHQD* and *lbGSTT* genes and wild types were subjected to salinity and oxidative stresses. Forty days after stress induction, total RNA was extracted from the leaves, cDNA was synthesized, and the gene expression profiles were analyzed. The *lbGSTTCHQD* was significantly upregulated in all lines except for line 4 under oxidative stress conditions (Figs. 7A, B). High expression levels of *lbGSTT* were detected in the transgenic lines 8 and 10 under both stress conditions, whereas no

significant difference was detected in the transgenic line 2 (Figs. 7C, D). Our results showed that *lbGSTTCHQD* and *lbGSTT* responded positively to salinity and oxidative stress conditions.

Discussion

Leafy sweet potato has been shown to contain high amount of several CQAs that are important for plant development and human health (Tang et al. 2021). In this study, two leafy sweet potato varieties, i.e., EC16 and FS7-6 were investigated for their content in mono and diCQAs. It was found that these two varieties contain 3-CQA, 3,5-diCQA, 3,4-diCQA, 4,5-diCQA, 4-CQA and 5-CQA, which are highly concentrated in stems and leaves compared to stems tips and roots (Fig. 1A). Our result is similar to that of Li *et al.*, who demonstrated that both sweet potato stems and leaves represent a rich source of various CQAs (Li et al. 2017). In the latter study, 5-CQA, 4,5- and 3,5-diCQA were the predominant phenolic acids in the leaves and stems of fourteen different cultivars of sweet potato with higher CQAs accumulated in the leaves compared to the stems. However, in our study the 5-CQA, 3-CQA and 3,5-diCQA were the most accumulated acids in the leaves and stems of EC16 and FS7-6 and their amounts were significantly higher in stems compared to the leaves. This could be explained by the differences in genotypes used in the respective study. Moreover, plants from the EC16 variety accumulated more CQAs compared to the FS7-6 variety (Fig. 1B), which is in agreement with the findings of Liu et al., (Liu et al. 2021). It could be assumed that the differences in CQAs contents between varieties are due to the genetic heritage of each variety, which probably influences their ability to resist various environmental stresses that stimulate the enzymatic genes involved in the PPs pathway responsible for CQAs biosynthesis. The level of CQAs could also vary depending on developmental stages and plant tissues (Chen et al. 2017).

Sweet potato *GST* genes *itb01g35330* (*lbGSTTCHQD*) and *itb09g30700* (*lbGSTT*) were further investigated for their association with the accumulation of CQAs. These two genes shared close homology with the *GST* from other plant species (Fig. 2). Their expression analysis showed different patterns in stem tips, stems, leaves and roots of EC16 and FS7-6 cultivars (Figs. 3A, B). The stems, which contained higher concentration of CQAs compared to the leaves, exhibited low transcripts of *lbGSTTCHQD* and *lbGSTT*, suggesting that the tissue-specific expression level of the genes does not correlate with the amount of CQAs in the tissues. However, the transcripts abundance of the two genes is associated with the amount of the total CQAs in both varieties, where the EC16 variety that accumulated more CQAs showed more transcript abundance of the *lbGSTTCHQD* and *lbGSTT* genes compared to the FS7-6 variety. In addition, different levels of expression were observed between *lbGSTTCHQD* and *lbGSTT* genes. This could be due to the variability in the types, numbers and positions of each regulatory element present in the promoter regions, which would potentially influence the mechanism of *GST* related to the accumulation of CQAs.

Sequence comparison of the coding regions of the *lbGST* genes amplified using leaves from EC16 and FS7-6 varieties as templates revealed some variations between the two varieties at the amino acid sequences level. We observed mutations at different sites in the amino acid sequences (Figs. 4A, B), which might affect the synthesis and the functions of the target GST proteins and consequently the

accumulation of CQAs in both varieties. The coding regions of *lbGSTTCHQD* and *lbGSTT* from the EC16 variety were further transferred into *A. tumefaciens* and also transformed into *N. benthamiana* and *N. tabacum* for transient expression (Figure 5A-D) and transgenic transformation analyses, respectively (Figure 6A-K). It was found that the overexpression of the two genes promotes the accumulation of monoCQAs but has no effect on diCQAs accumulation. Among the monoCQAs, 5-CQA was predominately accumulated compared to the other CQAs. This result confirmed the abundant level of chlorogenic acid in plants, which is consistent with previous studies (Thi Lan Khanh et al. 2018; Frond et al. 2019). As it was reported that after synthesis in cytoplasm, CQAs are transported to vacuoles for storage, we speculate that the monoCQAs bind to the active sites of the *lbGSTs*, inducing modifications in the structures of these proteins, and thereby induce the formation of the *lbGST*-monoCQAs complexes which facilitate their accumulation. The mechanism of CQAs related *GST* genes is probably the same as that of anthocyanins in which *GST* mediates the accumulation of anthocyanins (Lai et al. 2021). Besides, another study reported that transportation of secondary metabolites occurs through membranes or vesicles (Gani et al. 2021). The *lbGSTTCHQD* and *lbGSTT* genes could also be associated with membrane transporters to ensure the vacuolar accumulation of monoCQAs. The non-correlation between the overexpression of *lbGST* genes and the content of diCQAs could be explained by an insufficient synthesis of proteins that is essential for diCQAs accumulation in plants. Lower synthesis of proteins during translation of mRNA potentially led to the inhibition of the enzymatic activity of *GSTs*, the reduction of metabolic flux and consequently to a low level of diCQAs. This storage process might not be similar with that of the monoCQAs, which probably accumulated more rapidly and easily even with a low amount of *GST* proteins. Moreover, it has been reported that diCQAs are produced from monoCQAs (Cheevarungrapakul et al. 2019), which could explain the lower amount of diCQAs compared to monoCQAs. Consistently, diCQAs are shown to be less stable than monoCQAs and rapidly degrade into methylated or cis-isomer forms (Suryadi and Mun'im 2021). So far, little is known about the biosynthesis of diCQAs (Magana et al. 2021). It is possible that the accumulation mechanism of these compounds requires other transferase genes that function together with *GST* genes in the process of transporting the acids for their vacuolar sequestration. The identification of novel candidate transferase genes might be important in increasing the level of diCQAs.

The role of *lbGSTTCHQD* and *lbGSTT* in plant response to salinity and oxidative stresses was further determined, and we observed significant upregulation of the genes under abiotic stress conditions (Figs. 7A-D). *lbGSTTCHQD* and *lbGSTT* were induced by both salinity and oxidative stresses, indicating that the stress regulatory-elements DRE, MYB and MYC found in the promoter regions of the *lbGSTTCHQD* and *lbGSTT* (Table 1) play crucial roles in enhancing stress tolerance of transgenic plant. Overall, the *lbGST* genes could mediate the transport and accumulation of monoCQAs into vacuoles, from where they participate in plant detoxification under diverse environmental stresses.

Conclusion

In the present study, the involvement of *itb01g35330* (*lbGSTTCHQD*) and *itb09g30700* (*lbGSTT*) in the accumulation of CQAs was investigated using both transient expression and transgenic transformation.

The results revealed positive correlations between the overexpression of *lbGSTTCHQD* and *lbGSTT* genes and the accumulation of monoCQAs in tobacco plants. However, the genes had no effect on the level of diCQAs. These two genes could play important roles in the metabolism of monoCQAs in leafy sweet potato. The transcript levels of the genes were also induced by salinity and oxidative stresses, suggesting the implication of those two genes in plant tolerance to abiotic stresses. The metabolism of CQAs is not only related to their biosynthesis, but also to their transportation and accumulation. Since *GSTs* serve as protein transporters in secondary metabolisms, the manipulation of these genes would help to improve the sequestration of CQAs during the growth and development of plants. Further research is therefore needed to better understand the regulatory mechanisms of *GST* genes leading to increased accumulation of CQAs.

Declarations

Funding This work was supported by National Key R&D Program of China (2018YFD1000700, 2018YFD1000705, 2019YFD1001300, 2019YFD1001305), Science and Technology Development Plan Project of Jingzhou City, Hubei Province, China (2018-37), Characteristic Discipline of Hubei Academy of Agricultural Sciences (2015TSXK06) and Science and Technology Innovation Center of Hubei Academy of Agricultural Sciences (2007-620-001-03).

Author contributions DRJS performed the experiments, analyzed and interpreted the data, and drafted the original manuscript. YL and RP participated in transient expression experiments. SFA contributed to review and editing the manuscript. WYZ and XSY participated in supervision, project administration, funding acquisition, review and editing. All authors have read and approved the final manuscript.

Conflict of interest The authors declare that there are no conflicts of interests.

Data availability The data generated and analyzed during the current study are available from the corresponding author upon reasonable request.

References

1. Alam MK (2021) A comprehensive review of sweet potato (*Ipomoea batatas* [L.] Lam): Revisiting the associated health benefits. *Trends in Food Science & Technology* 115: 512-529. <https://doi.org/10.1016/j.tifs.2021.07.001>
2. Biała W, Jasiński M (2018) The Phenylpropanoid Case – It Is Transport That Matters. *Frontiers in Plant Science* 9: 1-8. <https://doi.org/10.3389/fpls.2018.01610>
3. Cheeverungnapakul K, Khaksar G, Panpetch P, Boonjing P, Sirikantaramas S (2019) Identification and functional characterization of genes involved in the biosynthesis of caffeoylquinic acids in sunflower (*Helianthus annuus* L.). *Frontiers in plant science* 10: 968. <https://doi.org/10.3389/fpls.2019.00968>
4. Chen Z, Liu G, Liu Y, Xian Z, Tang N (2017) Overexpression of the *LmHQT1* gene increases chlorogenic acid production in *Lonicera macranthoides* Hand-Mazz. *Acta Physiologiae Plantarum*

- 39: 1-10. <https://doi.org/10.1007/s11738-016-2310-8>
5. Czerniawski P, Bednarek P (2018) Glutathione S-Transferases in the Biosynthesis of Sulfur-Containing Secondary Metabolites in Brassicaceae Plants. *Frontiers in Plant Science* 9: 1-8. <https://doi.org/10.3389/fpls.2018.01639>
 6. Estévez IH, Hernández MR (2020) Plant glutathione S-transferases: an overview. *Plant Gene* 23: 100233. <https://doi.org/10.1016/j.plgene.2020.100233>
 7. Frond AD, Iuhas CI, Stirbu I, Leopold L, Socaci S, Andreea S, Ayvaz H, Andreea S, Mihai S, Diaconeasa Z (2019) Phytochemical characterization of five edible purple-reddish vegetables: Anthocyanins, flavonoids, and phenolic acid derivatives. *Molecules* 24: 1536. <https://doi.org/10.3390/molecules24081536>
 8. Gani U, Vishwakarma RA, Misra P (2021) Membrane transporters: the key drivers of transport of secondary metabolites in plants. *Plant Cell Reports* 40: 1-18. <https://doi.org/10.1007/s00299-020-02599-9>
 9. Ghangal R, Rajkumar MS, Garg R, Jain M (2020) Genome-wide analysis of glutathione S-transferase gene family in chickpea suggests its role during seed development and abiotic stress. *Molecular biology reports* 47: 2749–2761. <https://doi.org/10.1007/s11033-020-05377-8>
 10. Hao Y, Xu S, Wang H, Kong L, Sun S (2021) Comparative Analysis of the Glutathione S-Transferase Gene Family of Four Triticeae Species and Transcriptome Analysis of GST Genes in Common Wheat Responding to Salt Stress. *International journal of genomics* 2021: 6289174. <https://doi.org/10.1155/2021/6289174>
 11. Kou M, Liu Y-j, Li Z-y, Zhang Y-g, Tang W, Yan H, Wang X, Chen X-g, Su Z-x, Arisha MH (2019) A novel glutathione S-transferase gene from sweetpotato, IbGSTF4, is involved in anthocyanin sequestration. *Plant Physiology and Biochemistry* 135: 395-403. <https://doi.org/10.1016/j.plaphy.2018.12.028>
 12. Lai B, You Y, Zhang L, Wang Q, Chen F, Luo G, Du L, Wang H (2021) Identification and functional characterization of RsGST1, an anthocyanin-related glutathione S-transferase gene in radish. *Journal of Plant Physiology* 263: 153468. <https://doi.org/10.1016/j.jplph.2021.153468>
 13. Li M, Jang GY, Lee SH, Kim MY, Hwang SG, Sin HM, Kim HS, Lee J, Jeong HS (2017) Comparison of functional components in various sweet potato leaves and stalks. *Food science and biotechnology* 26: 97-103. <https://doi.org/10.1007/s10068-017-0013-6>
 14. Li Y, Kong D, Bai M, He H, Wang H, Wu H (2019) Correlation of the temporal and spatial expression patterns of HQT with the biosynthesis and accumulation of chlorogenic acid in *Lonicera japonica* flowers. *Horticulture research* 6: 1-14. <https://doi.org/10.1038/s41438-019-0154-2>
 15. Li Y, Liu X, Li F, Xiang L, Chen K (2021) The Isolation and Identification of Anthocyanin-Related GSTs in *Chrysanthemum*. *Horticulturae* 7: 231. <https://doi.org/10.3390/horticulturae7080231>
 16. Lin Y, Zhang L, Zhang J, Zhang Y, Wang Y, Chen Q, Luo Y, Zhang Y, Li M, Wang X (2020) Identification of Anthocyanins-Related Glutathione S-Transferase (GST) Genes in the Genome of Cultivated Strawberry (*Fragaria × ananassa*). *International journal of molecular sciences* 21: 8708. <https://doi.org/10.3390/ijms21228708>

17. Liu Y, Jiang H, Zhao Y, Li X, Dai X, Zhuang J, Zhu M, Jiang X, Wang P, Gao L (2019) Three *Camellia sinensis* glutathione S-transferases are involved in the storage of anthocyanins, flavonols, and proanthocyanidins. *Planta* 250: 1163-1175. <https://doi.org/10.1007/s00425-019-03206-2>
18. Liu Y, Su W, Wang L, Lei J, Chai S, Zhang W, Yang X (2021) Integrated transcriptome, small RNA and degradome sequencing approaches proffer insights into chlorogenic acid biosynthesis in leafy sweet potato. *Plos one* 16: e0245266. <https://doi.org/10.1371/journal.pone.0245266>
19. Magana AA, Kamimura N, Soumyanath A, Stevens JF, Maier CS (2021) Caffeoylquinic acids: Chemistry, biosynthesis, occurrence, analytical challenges, and bioactivity. *The plant journal* 107: 1-21. <https://doi.org/10.1111/tpj.15390>
20. Martino JA, Di Feo LdV, Paccioretti M, Contardi CA, Sanchez MA, Ortega LI, Pastor SE (2021) First report of *Berkeleyomyces basicola* (synonymous: *Thielaviopsis basicola*) on roots of sweet potato (*Ipomoea batatas* (L.) Lam) in Argentina. *Revista de la Facultad de Ciencias Agrarias UNCuyo* 53: 283-287. <https://doi.org/10.48162/rev.39.027>
21. Moglia A, Lanteri S, Comino C, Butelli E, Martin C (2011) The silencing of a GST gene increases the content of health-promoting dicaffeoylquinic acids in tomato. Joint meeting AGI-SIBV-SIGA, Società Italiana Genetica Agraria
22. Naveed M, Hejazi V, Abbas M, Kamboh AA, Khan GJ, Shumzaid M, Ahmad F, Babazadeh D, FangFang X, Modarresi-Ghazani F, WenHua L, XiaoHui Z (2018) Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomed Pharmacother* 97: 67-74. <https://doi.org/10.1016/j.biopha.2017.10.064>
23. Pérez-Díaz R, Madrid-Espinoza J, Salinas-Cornejo J, González-Villanueva E, Ruiz-Lara S (2016) Differential roles for VviGST1, VviGST3, and VviGST4 in proanthocyanidin and anthocyanin transport in *Vitis vinifera*. *Frontiers in plant science* 7: 1166. <https://doi.org/10.3389/fpls.2016.01166>
24. Soviguidi DRJ, Pan R, Liu Y, Rao L, Zhang W, Yang X (2022) Chlorogenic Acid Metabolism: The Evolution and Roles in Plant Response to Abiotic Stress. *Phyton-International Journal of Experimental Botany* 91: 239-255. <https://doi.org/10.32604/phyton.2022.018284>
25. Srivastava D, Verma G, Chauhan AS, Pande V, Chakrabarty D (2019) Rice (*Oryza sativa* L.) tau class glutathione S-transferase (OsGSTU30) overexpression in *Arabidopsis thaliana* modulates a regulatory network leading to heavy metal and drought stress tolerance. *Metallomics* 11: 375-389. <https://doi.org/10.1039/c8mt00204e>
26. Suárez S, Mu T, Sun H, Añón MC (2020) Antioxidant activity, nutritional, and phenolic composition of sweet potato leaves as affected by harvesting period. *International Journal of Food Properties* 23: 178-188. <https://doi.org/10.1080/10942912.2020.1716796>
27. Suryadi H, Mun'im A (2021) Effect of gamma irradiation on the caffeoylquinic acid derivatives content, antioxidant activity, and microbial contamination of *Pluchea indica* leaves. *Heliyon* 7: e07825. <https://doi.org/10.1016/j.heliyon.2021.e07825>

28. Tang C-C, Ameen A, Fang B-P, Liao M-H, Chen J-Y, Huang L-F, Zou H-D, Wang Z-Y (2021) Nutritional composition and health benefits of leaf-vegetable sweet potato in South China. *Journal of Food Composition and Analysis* 96: 103714. <https://doi.org/10.1016/j.jfca.2020.103714>
29. Thi Lan Khanh P, Chittrakorn S, Rutnakornpituk B, Phan Tai H, Ruttarattanamongkol K (2018) Processing effects on anthocyanins, phenolic acids, antioxidant activity, and physical characteristics of Vietnamese purple-fleshed sweet potato flours. *Journal of Food Processing and Preservation* 42: e13722. <https://doi.org/10.1111/jfpp.13722>
30. Zhao Y, Dong W, Zhu Y, Allan AC, Lin-Wang K, Xu C (2020) PpGST1, an anthocyanin-related glutathione S-transferase gene, is essential for fruit coloration in peach. *Plant biotechnology journal* 18: 1284-1295. <https://doi.org/10.1111/pbi.13291>

Tables

Table 1 cis-elements in the promoter of IbGST genes

Site name	IbGSTTCHQD	IbGSTT	Sequence	Function
ABRE	2	-	ACGTGGC	Abscisic acid responsive element binding factor
DRE	-	2	TACCGACAT	Dehydration and cold responsive element
MYB	-	1	AACTAAA/C	Growth, PPs metabolism, drought, salinity, and cold responsive element
MYC	2	1	CACATG	PPs metabolism, growth, drought responsive element
G-box	1	-	CACGTG	Light responsive factor
CAAT-box	-	1	CCAAT	Cis-acting regulatory element related to plant metabolism and development
TATA-box	-	15	TATAAA	Core promoter element of transcription start

Figures

Figure 1

Caffeoylquinic acids content in EC16 and FS7-6 sweet potato varieties. **A** Level of 3-CQA, 4-CQA, 5-CQA, 3,4-diCQA, 3,5-diCQA and 4,5-diCQA in stem tips, stems, leaves and roots of EC16 and FS7-6. The CQAs were highly accumulated in stems and leaves compared to stems tips and roots. Alphabetic letters indicate significant differences based on Duncan's multiple-range test ($p < 0.05$). **B** Comparison of total

caffeoylquinic acids concentration between EC16 and FS7-6 in different tissues. The total amount of CQAs was significantly higher in EC16 than in FS7-6. ** indicate significant differences according to Student's t-test ($p < 0.01$).

Figure 2

Phylogenetic analysis of GST proteins from sweet potato and other plant species. The phylogenetic tree showed close homology and domain similarity between the IbGSTs and other plant GST proteins. Ib: Ipomoea batatas, Na: Nicotiana attenuata (XP_019264208.1:1-231), Sl: Solanum lycopersicum (XP_004245615.1:1-250), Cs: Camellia sinensis (XP_028065230.1:1-235), Jc: Jatropha curcas (XP_037493609.1), Pv: Pistacia vera (XP_031255948.1), Hb: Hevea brasiliensis (XP_021674985.1:2-240), Me: Manihot esculenta (XP_021629763.1:1-231), Hi: Handroanthus impetiginosus (PIN12481.1:1-334), Sa: Striga asiatica (GER57004.1:1-322), Ar: Actinidia rufa (GFZ10448.1:1-319), Ac: Actinidia chinensis (PSR99570.1:1-330), Cf: Cephalotus follicularis (GAV56714.1:1-317), Tc: Theobroma cacao (EOY15086.1:1-322).

Figure 3

Expression profile of IbGST genes in stem tips, stems, leaves and roots of EC16 and FS7-6 sweet potato varieties. **A** Relative expression of IbGSTTCHQD in different tissues of EC16 and FS7-6. IbGSTTCHQD was most expressed in the leaves of EC16, while its expression level was lower in the leaves of FS7-6 but high in stem tips, stems and roots. **B** Relative expression of IbGSTT in different tissues of EC16 and FS7-6. IbGSTT was highly expressed in leaves and roots of both varieties. Alphabetic letters indicate significant differences (Duncan's multiple-range test, $p < 0.05$).

Figure 4

Comparison of IbGSTTCHQD and IbGSTT sequences in leafy sweet potato varieties EC16 and FS7-6. Slight differences of amino acids were observed in the protein sequences of both varieties for each gene.

Figure 5

Relative expression of *IbGST* genes and CQAs content in infiltrated *Nicotiana benthamiana* leaves. **A, B** qRT-PCR analysis of *IbGSTTCHQD* and *IbGSTT* in *N. benthamiana*. Expression analysis revealed significant increases of more than 4-fold in treated leaves compared to the wild-types. **C, D** CQAs concentration in *N. benthamiana*. *IbGSTTCHQD* and *IbGSTT* enhanced the level of monoCQA contents but had no effects on diCQAs in comparison with the control, * and ** indicate significant differences according to Student's t-test $p < 0.05$ and $p < 0.01$, respectively.

Figure 6

Stable expression of IbGSTTCHQD and IbGSTT in *Nicotiana tabacum*. **A** *N. tabacum* leaves transformed with *Agrobacterium tumefaciens* harboring IbGST genes, **B** Shoot induction on selective shooting medium, **C** Transgenic plantlet root induction on selective rooting medium, **D** Transgenic plant after 4 weeks on selective rooting medium. **E, F** PCR-detection of overexpressing IbGSTTCHQD and IbGSTT in transgenic lines. All the lines produced the expected products of 981 bp and 723 bp for IbGSTTCHQD and

IbGSTT, respectively. **G, H** Relative expression of IbGSTTCHQD and IbGSTT in transgenic lines. Expression analysis exhibited significant increases of transcript levels of about 3 to 5-fold higher than that in wild-type plants. **I, J** Caffeoylquinic acids content in overexpressing-IbGSTTCHQD and IbGSTT transgenic lines. Significant accumulation of monoCQAs and no changes of diCQA contents were observed for both genes compared to the wild types. **K** Average CQAs contents comparison between wild-types and overexpressing-IbGSTTCHQD and IbGSTT transgenic lines. IbGSTT and IbGSTTCHQD promote the accumulation of monoCQAs up to 20 and 55%, where the genes were not correlated with the accumulation of diCQAs. M: 2000 bp DNA marker; W: water as negative control; P: pCAMBIA1300-IbGSTs as positive control; WT: wild-type plant control; L1-L10: transgenic lines. * and ** indicate significant differences according to Student's t-test $p < 0.05$ and $p < 0.01$, respectively.

Figure 7

Response of overexpressing-IbGSTTCHQD and IbGSTT transgenic plants to salinity and oxidative stresses. The IbGSTs were induced by both stress treatments. * and ** indicate significant differences

according to Student's t-test $p < 0.05$ and $p < 0.01$, respectively.

Supplementary Files

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

- [Additionalfile1TableS1.SequenceinformationofIbGSTgenes.xlsx](#)
- [Additionalfile2TableS2.PrimersequencesofIbGSTgenes.xlsx](#)