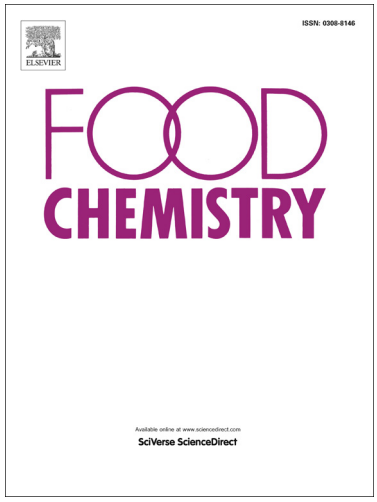
Journal Pre-proofs



Genome-wide analyses reveals a glucosyltransferase involved in rutin and emodin glucoside biosynthesis in tartary buckwheat

Qinggang Yin, Xiaoyan Han, Zongxian Han, Qingfu Chen, Yuhua Shi, Han Gao, Tianyuan Zhang, Gangqiang Dong, Chao Xiong, Chi Song, Wei Sun, Shilin Chen

PII: S0308-8146(20)30340-X
DOI: https://doi.org/10.1016/j.foodchem.2020.126478
Reference: FOCH 126478

To appear in: *Food Chemistry*

Received Date: 3 October 2019
Revised Date: 23 February 2020
Accepted Date: 23 February 2020

Please cite this article as: Yin, Q., Han, X., Han, Z., Chen, Q., Shi, Y., Gao, H., Zhang, T., Dong, G., Xiong, C., Song, C., Sun, W., Chen, S., Genome-wide analyses reveals a glucosyltransferase involved in rutin and emodin glucoside biosynthesis in tartary buckwheat, *Food Chemistry* (2020), doi: https://doi.org/10.1016/j.foodchem.2020.126478

Genome-wide analyses reveals a glucosyltransferase involved in rutin and emodin glucoside biosynthesis in tartary buckwheat

Qinggang Yin[1*] Xiaoyan Han[2*] Zongxian Han[3] Qingfu Chen[4] Yuhua Shi[1] Han Gao[3] Tianyuan Zhang[1] Gangqiang Dong[5] Chao Xiong[1] Chi Song[1] Wei Sun[1#] Shilin Chen[1#]

1. Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

2. Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

3. School of Chemistry Chemical Engineering and Life Sciences, Wuhan University of Technology, No. 122, Lo lion Road, Wuhan, Hubei, 430070, China

4. Research Center of Buckwheat Industry Technology, Guizhou Normal University, Baoshan Beilu 116, Guiyang 550001, China

5. Amway（China）Botanical R&D centre, Wuxi 214115, China

*These authors contributed equally to this work.

#Corresponding authors: Wei Sun, E-mail: wsun@icmm.ac.cn; Shilin Chen, E-mail:. slchen@icmm.ac.cn, phone: 0086-01064014411.

Highlights

Temporal and spatial transcripts of 106 FtUGTs were analyzed in genome level.

FtUGT73BE5 was identified dominating the first step of rutin glycosylation.

FtUGT73BE5 was demonstrated to catalyze emodin to form emodin 6-*O*-glucoside.

MeJA up-regulates transcript of *FtUGT73BE5*; it was related to anti-adversity.

Abstract

With people's increasing needs for health concern, rutin and emodin in tartary buckwheat have attracted much attention for their antioxidant, anti-diabetic and reducing weight function. However, the biosynthesis of rutin and emodin in tartary buckwheat is still unclear; especially their later glycosylation contributing to make them more stable and soluble is uncovered. Based on tartary buckwheat' genome, the gene structures of 106 UGTs were analyzed; 21 candidate FtUGTs were selected to enzymatic test by comparing their transcript patterns. Among them, FtUGT73BE5 and other 4 FtUGTs were identified to glucosylate flavonol or emodin *in vitro*; especially rFtUGT73BE5 could catalyze the glucosylaton of all tested flavonoids and emodin. Furthermore, the identical *in vivo* functions of FtUGT73BE5 were demonstrated in tartary buckwheat hairy roots. The transcript profile of *FtUGT73BE5* was consistent with the accumulation trend of rutin in plant; This gene may relate to anti-adversity for its transcripts were up-regulated by MeJA, and repressed by ABA.

Key words

Tartary buckwheat; Rutin; Emodin; Glucoside; Biosynthesis; Glucosyltransferase; Genome-wide

1 Introduction

Tartary buckwheat (*Fagopyrum tataricum*) belongs to the genus *Fagopyrum*, a member of the eudicot family Polygonaceae which comprises of about 23 species (Chen, 2018). Distributed in subtropical and temperate regions, tartary buckwheat is planted especially in mountainous areas where other crops such as rice and wheat are not adapted (Tsuji & Ohnishi, 2000; Liu, Chen, Yang, &Chiang, 2008). Due to the beneficial values of flavonoids and emodin, tartary buckwheat is more popular as a healthy food than a crop. Tartary buckwheat is an important natural source of flavonoids for human, such as rutin, which are minor in main crops like rice (*Oryza sativa*), wheat (*Triticum aestivum*), sorghum (*Sorghum bicolor*), highland barley (*Hordeum vulgare*) and soybean (*Glycine max*) (Fabjan, Rode, Kosir, Wang, Zhang, & Kreft, 2003). Previous studies indicated that these flavonoids have antioxidant, anti-diabetic, anti-inflammatory, anti-neuroprotective and cytotoxic effects (Yasui *et al.,* 2016; Zhang *et al.,* 2017; Al-Snafi, 2017). Another functional compound—emodin exists in tartary buckwheat, which plays a major role in preventing type II diabetes and reducing weight according to recent reports (Lee, Ku, & Bae*,* 2013; Peng *et al.,* 2013).

Though the structures of rutin and emodin are different, they possess a common precursor malonyl Co-A, catalyzed by polyketone synthase (PKS), and then stably exists in plant through the same glycosylation (Fig.S1). Unmodified rutin is quercetin, which are biosynthesized by chalcone synthase (CHS) which is a kind of PKS, chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavanone 3'-hydroxylase (F3'H) and flavonol synthase (FLS) according to the study of model plants, such as *Glycine max* (Rodas *et al.,* 2014) and *Citrus* (Frydman *et al.,* 2004). However, there are few reports regarding to these genes in tartary buckwheat. The synthesis of emodin sorted to anthraquinone is mainly through polyketone pathway (Bringmann & Irmer, 2008). The polyketone pathway can be divided into 3 stages: octaketide chain were formed using acetyl Co-A and malonyl Co-A as the starting substrates by PKS. After reduction, decarboxylation and oxidation, chrysophanol, aloe emodin and rhein were formed from octaketide chain. Alternatively, emodin

could be produced by hydrolyzed, decarboxylation and dehydration of octaketide chain. Generally, quercetin and emodin are finally modified by UDP-glycosyltransferase (UGT) to produce rutin and emodin glycosides respectively.

Glycosides are the final form of rutin and emodin in plant, which determine their pharmaceutical functions (Jing, Li, Hu, Jiang, Qin, & Zheng, 2014; Lee *et al.,* 2013). Glycosylation of rutin is divided into two steps. First, quercetin is glucosylated by 3-hydroxyl, followed by rhamosylation of the 6-hydroxyl of glucose group in quercetin 3-*O*-glucoside (An, Yang, Kim, & Ahn, 2016). FaGT6 (*Fragaria ananassa*), Cp3OGT (*Citrus paradisi*), GmUGT78K1 (*G. max*) and VvGT5 (*Vitis Vinifera*) were *in vitro* proved to involve in the first step (Griesser *et al.,* 2008; Owens, & McIntosh, 2009; Kovinich, Saleem, Arnason, & Miki, 2010; Mizohata *et al.,* 2013), but their functions *in vivo* are still blurry. GmUGT79A6, Cm1, 6RhaT and FeUGT79A8 (*F. esculentum*) were demonstrated responsible for the second step of rutin glycosylation (Frydman *et al.,* 2004; Rodas *et al.,* 2014; Koja, Ohata, Maruyama, Suzuki, Shimosaka, & Taguchi, 2018). A previous report identified three UGTs involved in the glucosylation of cyanidin 3-*O*-glucoside in *F. tataricum* by enzymatic tests *in vitro* (Zhou *et al.,* 2016), meanwhile FeCGTa (UGT708C1) and FeCGTb (UGT708C2) were identified to have *C*-glucosylation activity toward 2-hydroxyflavanones, dihydrochalcone, trihydroxyacetophenones and other related compounds with chemical structures similar to 2′,4′,6′-trihydroxyacetophenone (Nagatomo, Usui, Ito, Kato, Shimosaka, & Taguchi, 2014). Since emodin possess very low solubility, the practical potency as oral medicine will be discounted compared with its theoretical biological activity (Lee *et al.,* 2013). That was in accordance with later report that the anthraquinone glucosides had higher bioavailability and better pharmacological activity (Ghimire, Koirala, Pandey, Jung, & Sohng, 2015). Emodin exists 3 hydroxyls distributed 1-, 6- and 8- position, which would be the target of glycosylation. There is only one report about the glucosylation of emodin in plant: CtUGT73AE1 catalyzed the formation of emodin 8-*O*-glucoside *in vitro* (Xie *et al.,* 2014). However, no reports about glycosylation other positions of emodin has

been published to date. Usually the structure of protein determines its function in plants; it's reasonable to speculate the homolog gene of *CtUGT73AE1* in tartary buckwheat may encode similar functional protein.

The entire sequences from genome contribute to the feasibility for illuminating the superfamily of protein, such as UGTs, which possess functional redundancy and diversity (Zhang *et al.,* 2017). Based on the published genomes, integrated UGTs for their key secondary metabolites were identified systematically, such as flavonoids in Arabidopsis and *L. japonicus*, 2-phenylethanol in peach, andrographolide in *Andrographis paniculata* and *ginsenoside* in ginseng (Caputi, Malnoy, Goremykin, Nikiforova, & Martens, 2012; Yin, Shen, Chang, Tang, Gao, & Pang, 2017; Wu et al. 2017; Xu et al. 2017; Sun et al. 2018). Equally predictable, the available genome of *F. tataricum* would provide a scheme for the systematic analysis of the whole pathway of rutin synthesis, especially for the analysis of UGTs.

Herein, using the available genome and transcriptome data of *F. tataricum*, we identified 106 UGT genes. Among 14 cloned FtUGTs, not only FtUGT73BE5 was able to glycosylate all tested flavonoids, but also could catalyze emodin to form emodin 6-*O*-glucoside *in vitro*. Subsequently, the functions of FtUGT73BE5 were verified using *F. tataricum* hairy root system. Furthermore, methyl-jasmonate (MeJA) were demonstrated to up-regulate the transcription of *FtUGT73BE5* and slightly increase rutin content in seedlings. Our study sheds light on the functional diversity and biochemical mechanism of UGTs, involving in the biosynthesis of rutin and emodin glucosides in *F. tataricum*; these understandings would also benefit to developing healthy tartary buckwheat food.

2 Materials and Methods

2.1 Materials and chemicals

F. tataricum (BT18) plants were grown in an illumination incubator under control (16 h/8 h day/night cycle at 25℃/22℃, respectively, with a relative humidity of 40%). The roots, stems, leaves, flowers and seeds at different developing stages were collected (Sd1-Sd3 represent prophase, mid-term, later period of ovary bulge, Sd4-Sd6 represent prophase, mid-term, later period of grout, and Sd7-Sd9 represent prophase, mid-term, later period of mature); these samples were immediately frozen in liquid nitrogen, and stored at -80℃ for further use. Sterile seeds of *F. tataricum* were grown on Murashige and Skoog medium (MS) after soaking in warm water (37℃) for 20 min. The substrates tested in the present study were purchased from Xili Limited Co. (Shanghai, China) and Indofine (Hillsborough, NJ, USA). UDP-glucose and UDP-glucuronic acid were purchased from Sigma-Aldrich (Oakville, CA, USA). UDP-rhamnose was enzymatic synthesized using methods mentioned in Rautengarten et al. (2014). Chemicals used in this study were all of analytical or HPLC grade.

2.2 RNA extraction, sequencing and reads filtering

In order to examine the expression patterns of FtUGT genes associated with rutin and emodin glycosides biosynthetic pathway, RNAs from leaves, roots, stems, flowers and seeds of different developing stages were sequenced by Illumina HiSeq2000 platform. High-quality RNA was separately extracted from the fresh samples of *F. tataricum* using TaKaRa MiniBEST Plant RNA Extraction Kit. Libraries were constructed following the manufacturer's instructions, and sequence was carried out by using the Illumina HiSeq2000 platform. The raw reads were filtered following the same parameters, which used the Custom Perl script to generate contigs' database (Ravi & Mukash, 2012), and finally resulted in a total of ~37.4 Gb clean data. RNA-Seq reads from each tissue were individually mapped to the annotated genes from published genome of *F. tataricum*

(Trapnell *et al.* 2017) with default parameters. The expression level of each gene was evaluated by Fragments Per Kilobase per Million mapped fragments (FPKM) and calculated using Tophat and Cufflinks (version 2.1.1) with default parameters (Trapnell *et al.* 2017).

2.3 Isolation and sequence analysis of FtUGTs

Primers for the 14 FtUGT genes were designed according to the published genome of *F. tartaricum*. All the forward and reverse primers (see Supplementary Data S3) for gene cloning contained corresponding restriction sites. Mixed cDNAs from root, stem, leaves and seeds of *F. tataricum* were used for gene amplification. The PCR products were purified and digested using the corresponding restriction enzymes, and then ligated to the pMAL-c2x vector (New England BioLabs, Ipswich, MA, USA) digested with the same restriction enzymes for expression of recombinant protein in *Escherichia coli*.

Multiple sequence alignment of the deduced amino acid sequences was carried out by using DNASTAR. Predicted amino acid sequences of UGTs were aligned using Clustal X2 and further used for phylogenetic analysis. The neighbor-joining phylogenetic tree was constructed with 1000 boot-strap replicates using MEGA4.0 software (Tamura, Dudley, Nei, & Kumar, 2007).

2.4 Enzyme assay and product identification

Gaining recombinant UGT proteins in *E. coli* was followed as previously described (Yin *et al.,* 2017); the crude proteins were used to enzymatic activity test. Optimal temperature for FtUGT73BE5 was measured in 50 mM Tris-HCl (pH 7.0) with 2.5 μg purified protein, 2 mM UDP-glucose, 10 mM DTT and 100 μM quercetin or emodin, under temperature from 25 to 50℃.

To assess the optimal pH, the recombinant protein was put in the 50 mM Tris-HCl (pH 6, pH 7 or pH 8) on 37 ℃. The substrate specificity of FtUGT73BE5 was analyzed with 100 μM candidate flavonoids and emodin for 30 minutes on 37 ℃. For kinetic analysis of the recombinant FtUGT73BE5 protein, purified enzymes (10 μg) were incubated in reaction mixtures comprising 10 mM DTT, 50 mM Tris-HCl (pH 7.0), and 2 mM UDP-glucose or UDP-glucuronic acid or UDP-rhamnose crude (20 μl enzymatic solution with UDP-rhamnose), in a final volume of 100 μl. The concentration of the tested acceptor substrates ranged from 0 to 400 μM. Reactions were stopped by addition of methanol after 30 min incubation at 37℃. Samples were centrifuged at 14, 000 rpm for 10 min at 4℃, and further analyzed by HPLC or UPLC. The kinetic parameters *K*m and *K*cat were calculated by using the Hyper 32 program (http://hyper32.software.informer.com/).

Enzymatic reaction solution was filtered through 0.22 μm membranes, 20 μL aliquot was used to analyse the new products by HPLC (Yin *et al.,* 2017), then a 1μL aliquot was injected into UPLC-MS/MS for identifying the products. A 1290 series UPLC was coupled with a 6470 triple quadrupole mass spectrometer via an AJS-ESI interface (Agilent Technologies, Waldbronn, Germany). Samples were eluted on an Agilent Eclipse Plus C18 column (RRHD 1.8 μm, 2.1×50 mm). Acetonitrile was used as mobile phase A, water solution containing 0.1% formic acid and 0.1% ammonium formate was used as mobile phase B. The analytes were eluted using a linear gradient program: 0~12 min, 95%→5% B and washed for 1 min. The flow rate was 0.20 ml min^{-1} and column temperature was 45℃. Mass spectrometer was operated in negative ion mode, with sheath gas temperature at 300℃, gas flow at 5.0 L min^{-1}, and nebulizer gas at 45 psi. Capillary voltage was set at 3500 V, nozzle voltage 500 V, and delta EMV 200 V. Product Ion mode was used, and two precursor-product ions were selected for the qualification of glycosylated products, as shown in

Supplementary Table S3. Quantify data for all flavonol and emodin metabolites was analyzed by MassHunter (version B.07.00).

2.5 Homology modeling and docking statistic

Homology models of the FtUGT73BE5 were built, using the 3D structure of UGT84A1 (The UniProt Knowledgebase: Q5XF20) as a template, with the SWISS-MODEL server at http://swissmodel.expasy.org (Biasini *et al.*, 2014). The amino acid sequence identity between FtUGT73BE5 protein and UGT84A1 was 40%. UDP-Glucose and quercetin 3-*O*-glucoside were respectively docking with the model structure of FtUGT73BE5 using the Igemdock 2.1 program. The docking result with UDP-Glucose and quercetin 3-*O*-glucoside was visualized with the Pymol molecular graphics system at http://www.pymol.org.

2.6 Expression analysis by quantitative real-time PCR

Total RNA was isolated from roots, stems, leaves, flowers, and developing seeds of *F. tataricum* by using an RNAprep Pure Plant Kit (Tiangen Biotech Co., Beijing, China). The *histone H3* (*His3 gene*) was used as a housekeeping gene in qRT-PCRs (Zhou *et al.,* 2016).

Total RNA from UGT-overexpression or -knockdown hairy roots were extracted. PCR system was the same as previously described (Yin *et al,* 2017). Primer sequences used for qRT-PCR were listed in Supplementary Data S2.

2.7 Treatment of *F. tataricum* s with hormones

The cotyledons from one week seedlings of *F. tataricum* were collected, and then soaked in water with DMSO (100 μM), Me-jasmonic acid (MeJA; 100 μM) and abscisic acid (ABA; 100 μM) for 4, 14, and 24 h in dark, respectively. The samples were harvested and immediately frozen in liquid nitrogen, and stored at -80℃ until use.

2.8 Overexpression and knockdown *FtUGT73BE5* in *F. tataricum* hairy root

To investigate whether FtUGT73BE5 has the same catalytic activity toward quercetin and emodin *in vivo* as observed *in vitro*, the plasmids for overexpression and RNA interference (RNAi) of *FtUGT73BE5* were constructed by Gateway system (Invitrogen Corporation). Briefly speaking, the ORF region of the *FtUGT73BE5* driven by 35S CaMV promoter was subcloned to the binary vector pK7WG2D for gene overexpressing in plant, while the specific fragment about 300bp length of the *FtUGT73BE5* was cloned to the binary vector pK7GWIWG2II2D for RNAi. The resulting overexpression (pK7WG2D-*FtUGT73BE5*) and RNAi (pK7GWIWG2II2D-*FtUGT73BE5*) constructs were transformed into *Agrobacterium ATCC 10060* to generate transgenic hairy roots of *F. tataricum*.

The generation of tartary buckwheat hairy root was operated as before (Zhang *et al.,* 2019). After selecting for one week, GFP signal was detected under a fluorescent display instrument. The transgenic hairy roots were further cultured in B5 liquid medium with the same antibiotics for another three weeks in flask before harvest. Finally, the hairy root samples were used for qRT-PCR, flavonol and emodin metabolites analysis. In order to investigate whether FtUGT73BE5 has the same catalytic activity toward emodin *in vivo*, emodin (final concentration was 100 μM) was added to the B5 liquid medium, co-cultured with transgenic hairy roots for 24 h as similar to before due to the reason of emodin can hardly be detected in hairy roots of *F. tataricum* (Karimi *et al.,* 2002).

2.9 Analysis of flavonol and emodin metabolites by UPLC/MS/MS

The *FtUGT73BE5* overexpression and knockdown hairy root lines were selected and extracted for flavonol and emodin analysis by UPLC/MS/MS. 10 mg dry weight sample was extracted with 1 ml methanol in an ultrasonic bath at 25℃ for 30 min. Then the supernatant was filtered through a membrane (pore diameter is 0.22 μm) after centrifuging 10 min (12, 000 rpm) under 4℃. Finally, a 1 μl aliquot was injected into UPLC/MS/MS for subsequently analysis. The UPLC/MS/MS method was the same as mentioned above in 2.6: metabolites were detected in multiple reaction monitoring (MRM) mode, the MRM transitions of two precursor-product ion were selected as references for each compound, as shown in Table S3. Data was analyzed by MassHunter (version B.07.00).

2.10 Statistical analysis

Statistical analyses were performed using Excel (Microsoft Office, Microsoft). P-values were calculated using an unpaired, two-legged Student's *t*test (**p < 0.01; *p < 0.05; ns, not significant). Data represent means ± standard deviation (n ≥ 3).

3 Results and discussions

3.1 Characteristics of FtUGT genes in *F. tataricum* genome

With the help of the UGT nomenclature committee, the submitted 106 UGT genes were named; proteins encoded by these genes are comprise of 300 to 600 amino acids, and have significant PSPG motifs near their C-terminals (Data S1). Our analysis revealed that 38 of the 106 FtUGT genes

276 contained no introns, while 48 FtUGTs contained only a single intron (Fig.S2). The number of
277 FtUGTs without intron (38 out of 106) is apparent small than that of other plants, such as peach (72
278 out of 168) (Wu *et al.,* 2017). Their physical locations mapped on 8 chromosomes with a relatively
279 high density on the chromosome 1 (14 FtUGTs), 6 (16 FtUGTs), 7 (18 FtUGTs) and 8 (15 FtUGTs)
280 (Fig.S3).
281
282 We constructed a phylogenetic tree with these 106 FtUGTs and 119 AtUGTs (Fig. 1). According to
283 the phylogenetic tree, FtUGTs clustered into 23 families, 14 evolutional groups (A-H, J-N, and P)
284 and an uncharacterized group. Even though the number of groups become less, a new and an
285 unclassified one (UGT93 family and UGT95 family) were emerged comparing to AtUGTs. Seven
286 groups cover 77% members of FtUGTs: E (15), B (14), G (13), H (12), D (11), L (9) and A (8) (Fig.
287 1). Several UGTs in other plant, belonged to A, B, D, E, F, G and L evolutional groups, had been
288 reported to involve in flavonoid or emodin glycosylation. LjUGT72AD1 and LjUGT72Z2 from
289 group E were identified to glycosylate flavonol in *L. japonicas* hairy root (Yin *et al.,* 2017).
290 GmUGT79A6 (*G. max*) and FeUGT79A8 belonged to group A were reported to involve in rutin
291 biosynthesis (Rodas *et al.,* 2014; Koja et al., 2018). Therefore, in this study, FtUGTs sorted to
292 group A, B, D, E, F, G and L were preferentially chosen to be further investigated.
293
294 **3.2 Expression profiles of UGT genes in *F. tataricum***
295
296 We retrieved the expression profiles of 106 full-length UGT genes from transcriptome of *F.*
297 *tataricum* (Fig. 2 and Data S2). Hierarchical clustering analysis of the transcript data indicated that
298 57 UGT genes expressed in developing seeds. Among them, 20 UGT genes possessed high
299 expression levels in late development stage of seeds, other 31 UGT genes were highly expressed at
300 the stage of ovary bulge, while six UGT genes were highly expressed at the stage of ovary bulge

and seed maturation. Furthermore, in total of 11, 23 and 11 UGT genes were highly expressed in roots, stems, and leaves, respectively.

Due to the expression profiles, the evolutionary analysis and UGT family identification related to flavonoid or emodin in other species, total 19 FtUGTs were selected for their high expression in seed and close relationship with flavonoids and emodin glycosylations. These candidates include family UGT79s in group A, UGT89s in group B, UGT73s in group D, UGT71s, UGT72s and UGT88s in group E, UGT714s in group F, UGT85s in group G and UGT75s and UGT84s in group L (Table S1) (Dhaubhadel, Farhangkhoee, & Chapman, 2008, Xie *et al.,* 2014, Funaki *et al.,* 2015, Yin *et al.,* 2017). Two FtUGTs from unclassified group were also screened for theirs expression pattern; FtUGT93R1 was highly expressed in roots and seeds, while FtUGT95D1 was top one in expression level among all UGTs in roots, leaves and seeds. Finally, 21 FtUGTs were selected as candidates according to phylogenetic tree and transcriptomics analysis.

3.3 Sequence analyses and enzymatic activity of FtUGTs

Comparing to the sequences of reported functional UGTs toward flavonoids and emodin, 21 candidate FtUGTs were distributed into five clusters in an unrooted phylogenetic tree, that is in 3GT (1 FtUGT), 5GT (8), 7GT (4), CGT (6) and Branch forming clusters (2); while FtUGT95D1 located in an independent cluster (Fig. S4A). Fourteen FtUGTs out of 21 candidates were cloned and expressed in *E. coli* to produce recombinant proteins to investigate their enzymatic activities eventually. These 14 putative FtUGTs shared a relatively high identity within the PSPG box with 'W-2x-Q-3x-L-8x-H-x-G-2x-S-2x-E-17x-Q' motif near their C-terminal ends; FtUGT714C1 composed of 45 amino acids in its PSPG box, differed from other FtUGTs, which possessed 44 amino acids in PSPG box (Fig. S4B & S4C). Their open reading frames ranged from 1257 bp to

1548 bp in length encoding proteins from 419 to 515 amino acids. The last amino acids in their PSPG box were all glutamine (Q), suggesting that they may use UDP-glucose as sugar donor like previously reported UGT proteins (Kubo, Arai, Nagashima, & Yoshikawa, 2004).

Enzymatic tests were carried out by using 11 flavonoids like aglycones and emodin as sugar acceptors, and using UDP-glucose, UDP-glucuronic acid and UDP-rhamnose as sugar donors (Fig. S4C). Our results revealed that recombinant UGT73BE5 (rUGT73BE5), rUGT75R2, rUGT75R3, rUGT85AD3, and rUGT89G2 could use quercetin as substrate, while rUGT73BE5, rUGT75R2 and rUGT75R3 could use emodin as substrate; these five UGTs could use UDP-glucose but not UDP-glucuronic acid or UDP-rhamnose as sugar donor (Fig. 3A; Table S2). Among them, rUGT73BE5 and rUGT75R2 also showed glucosylate activity toward quercetin 3-*O*-glucoside. Further test of rUGT73BE5 and rUGT75R2 to seven other substrates showed that rUGT73BE5 possessed more complex substrate diversity than rUGT75R2 (Table S2).

Our data revealed that all of these enzymatic products were characterized as addition of one or two glucoses (molecular weight increased 162 or 324) to yield their corresponding aglycones (Fig.3B). The enzymatic products of rFtUGT73BE5 with quercetin substrate included quercetin 3-*O*-glucoside and quercetin di-glucoside comparing with authentic standards (Fig. 3B). Furthermore, when quercetin 3-*O*-glucoside was used as substrate, the enzymatic products were identified as two new quercetin di-glucosides by MS (Fig.3A and 3C). Similarly, FtUGT73BE5 could catalyze luteolin and kaempferol to form their di-glucoside (Fig. S5). Enzymatic products of FtUGT73BE5 with other tested substrates were all mono-glycoside confirmed by MS (Fig. S5). FtUGT75R2 could also catalyze quercetin to produce quercetin monoglucoside and di-glucoside, while merely catalyze other substrates to produce mono-glucoside (Fig. S6). Since we mainly focused on the procedure from quercetin to quercetin 3-*O*-glucoside, the glycosylated position of these di-glucosides was not identified further.

The enzymatic products of rUGT73BE5, rUGT75R2 and rUGT75R3 with emodin substrate were predicted as emodin 6-*O*-glucoside, as compared with authentic standards of emodin 1-*O*-glucoside and emodin 8-*O*-glucoside (Fig. 3D & 3E, Fig. S7). A previous work reported that some kind of *E. coli* could produce emodin glycoside mixtures including emodin 6-*O*-glucoside and emodin 8-*O*-glucoside, but which protein contributed to these procedures was not stated (Zhang, Ye, Zhan, Chen, &Guo, 2004). The majority enzymatic product of FtUGT73BE5 was emodin 6-*O*-glucoside, with extremely minor emodin 8-*O*-glucoside detected by UPLC-QQQ, which differing from CtUGT73AE1 mainly catalyzing emodin to form 8-*O*-glucoside (Xie *et al.,* 2014). Moreover, CtUGT73AE1 was observed to catalyze the reversible glycosyltrasferase reactions, which could be employed for deglycosylation to form UDP-glucose in certain cases. Since FtUGT73BE5 shares 43% amino acids identity with CtUGT73AE1, whether it has similar reversing glycosylation ability might to be investigated in the future.

We further studied the characteristics of rFtUGT73BE5, including optimum pH value, optimum enzymatic temperature, substrate specificity and enzymatic kinetic properties (Fig S8 and Fig. 4A & 4B). Among 6.0 to 8.0, the optimum pH value for producing quercetin and emodin were both 7.0. The enzymatic ability were enhanced with the increasing of temperature from 25℃ to 50℃. The optimum enzymatic ability for quercetin was 190.5 p*k*at/mg protein at 50℃, while that was 170 p*k*at/mg protein for emodin (Fig S8 A&B). The reaction rate for quercetin was 99.8 p*k*at/mg protein and for emodin was 88.8 p*k*at/mg protein, which both were the highest in the 10 tested substrates except for naringenin (127.6 p*k*at/mg protein) and luteolin (268.8 p*k*at/mg protein) (Fig S8 C).

Enzymatic kinetic test showed that the affinity of rUGT73BE5 for kaempferol (40.26 μM) and quercetin 3-*O*-glucoside (22.42 μM) was stronger than that for quercetin and emodin. The high

K_{cat}/K_m values for kaempferol and quercetin 3-*O*-glucoside were 0.53×10^{-3} s^{-1} μM^{-1} and 0.47×10^{-3} s^{-1} μM^{-1}, respectively. These results indicated that rFtUGT73BE5 was more efficient toward quercetin 3-*O*-glucoside than other substrates. Phylogenetic analysis demonstrated that FtUGT73BE5 was clustered into 7-*O*-group; however our enzymatic study showed that it possessed multiple functions toward different substrates with different positions. These results suggested that the correlation between amino acid similarity of glycosyltransferases and their functions was not strong (Chang, Singh, Phillips, & Thorson, 2011). Therefore, predicting function by protein structure would be more practical than amino acids sequence alignment (Sharma, Panigrahi, & Suresh, 2014). To explore the potential molecular basis for the highly enzymatic activity of FtUGT73BE5 towards quercetin 3-*O*-glucoside, we analyzed the interactions between amino acid residues and the substrates, including quercetin 3-*O*-glucoside and UDP-glucose. This analysis predicted the interaction between ten amino acids (His17, Gly18, Cys287, Gln355, His370, Trp373, Asn374, Ser375, Glu394 and Gln395) and the sugar donor UDP-glucose in FtUGT73BE5 by H-bond. Moreover, other ten amino acids (Gly18, His19, Gly140, Ser141, Cys287, Ser290, Gln355, Asn374, Ser375 and Glu394) could potentially form H-bond with quercetin 3-*O*-glucoside (Fig. 4C). The docking results suggested that the ligand-bindings on FtUGT73BE5 in a central cleft were formed by the N- and C-terminal domains (Fig. 4C); it was big enough to hold both quercetin 3-*O*-glucoside and UDP-glucose, which could explain why rFtUGT73BE5 possesses high catalyzed activity for quercetin 3-*O*-glucoside.

3.4 *In vivo* function of *FtUGT* in *F. tataricum* hairy roots

In order to evaluate the *in vivo* functions of FtUGT73BE5, we carried out overexpression driven by 35S promoter and knockdown through RNA interference (RNAi) in *F. tataricum* hairy root via *Agrobacterium rhizogenes*-mediated transformation. At least 10 transgenic hairy root lines were

obtained; qRT-PCR was used to quantify expression levels in over-expression and RNAi lines (Fig. 5A & 5B and Fig. S9A).

Compared to control, rutin content increased from 0.77 to 1.29-fold in three *FtUGT73BE5* over-expression transgenic lines (Fig. 5C). The intermediate quercetin 3-*O*-glucoside increased from 0.32- to 1.15-fold compare to control (Fig. 5C). The results appeared in *FtUGT73BE5* knockdown lines were coincided. Compared to control, quercetin 3-*O*-glucoside decreased from 0.6- to 0.65-fold while rutin content similarly decreased from 0.54- to 0.64-fold in three *FtUGT73BE5* RNA interference lines (Fig. S9B & C). These results demonstrated that FtUGT73BE5 could catalyze the formation of rutin and quercetin 3-*O*-glucoside *in vivo*.

Since emodin and emodin glucosides were hardly detected in *F. tataricum* hairy root, the function of FtUGT73BE5 was verified by feeding emodin substrate to transgenic hairy roots (Fig. 5D & 5E). The result indicated that emodin 6-*O*-glucoside was mainly detected in *FtUGT73BE5* over-expressing hairy roots; the content of emodin 6-*O*-glucoside in these overexpression lines was significantly higher than that in the control line, ranging from 17.36- to 22.73-fold. These results suggested that FtUGT73BE5 could glucosylate emodin to produce emodin 6-*O*-glucoside in *F. tataricum* hairy root .

As a superfamily, the UGT functions possess quite diversity and redundancy, which makes them difficult to identify in native species. In model plants, such as citrus, strawberry, Medicago, soybean and *L. japonicus*, their UGT functions were commonly found verifing in native plants (Frydman et al. 2004; Griesser et al. 2008; Pang, Peel, Sharma, Tang, & Dixon, 2008; Kojas et al. 2014; Yin et al. 2017). A few researches had focus on buckwheat UGTs: FeCGTa and FeCGTb were confirmed to have C-glucosylation activity toward flavonoids *in vitro* (Nagatomo *et al*. 2014); FeUGT79A8 was identified to rhamnosylate quercetin 3-*O*-glucoside in tobacco(Kojas et al. 2018); three FtUGTs

were identified to involve in anthocyanidin glucosylation *in vitro* (Zhou et al. 2016). Differently, in our study FtUGT73BE5 was illustrated could glucosylate quercetin and emodin not only *in vitro* through enzymatic tests, but also in native plants.

3.5 The expression pattern of *FtUGT73BE5* and its relationship with hormone

In order to characterize the function of *FtUGT73BE5* in plant, we further analyzed their expression levels in various tissues and different developing stages of seeds using qRT-PCR, *FtUGT75R2* was selected as control (Fig. 6A). The expression patterns of the two FtUGTs were similar with the data from transcriptome; *FtUGT73BE5* was highly expressed in leaves and seeds of *F. tataricum*, which consisted with the accumulation trends of rutin in these tissues. This result further confirmed that FtUGT73BE5 was truly responsible for the biosynthesis of rutin in leaves and seeds of *F. tataricum* (Fig. 6A). The transcript level of FtUGT73BE5 was down-regulated in the procedure of seed maturation; these may be caused by the accumulation of ABA at the mature stage of seeds.

Rutin was produced when plants suffer from adversity (Zhou et al., 2016). In our study, the rutin content increased about 5% after treated with ABA or MeJA, despite it reached as high as 34.9 mg/g dry weight in the cotyledon (Fig. 6B & C).

A previous study indicated that ABA and MeJA induced the expression of *LjUGT72AD2* and *LjUGT72Z1* from *L. japonicas*, which encoding two flavonol *O*-glucosyltransferases (Yin *et al.,* 2017). Similarly, the transcription of *FtUGT73BE5* was induced by MeJA for more than 4.26-fold after 14 h comparing to control (Fig. 6B). Conversely, after treated with ABA, the transcription of *FtUGT73BE5* was depressed by 0.17-fold after 5 h compared with untreated controls (Fig. 6B).

These experimental results demonstrated that ABA could inhibit the expression of FtUGT73BE5, which was coincided with the expression trend of FtUGT73BE5 in mature seeds.

4 Conclusion

In this study, 106 FtUGTs were systematically and comprehensively analyzed in genome level, from gene structure, chromosome location to phylogenetic category of encoding proteins. Even though less total number, FtUGTs formed a new evolutional group and an unclassified group (UGT93 familiy and UGT95 family) comparing with AtUGTs. The transcript profiles of FtUGTs in different temporal and spatial stages were analyzed. Almost half of total FtUGT genes (57 UGTs) were highly expressed in developing seeds.

A glucosyltransferase FtUGT73BE5 was identified involving to glucosylate rutin and emodin *in vitro* and *in vivo*. Among 21 candidate FtUGTs, screened based on published tartary buckwheat genome and their spatial and temporal transcripts, FtUGT73BE5 and others 4 FtUGTs were identified to glucosylate flavonol or emodin *in vitro*. rFtUGT73BE5 especially possessing substrate diversity could catalyze glucosylaton of all tested flavonoids and emodin. Furthermore, the functions of FtUGT73BE5 *in vivo* were demonstrated in hairy roots of tartary buckwheat. Rutin content was increased at most by 1.29-fold in *FtUGT73BE5* over-expressing lines, while the emodin 6-*O*-glucoside content was increased at most by 22.73-fold in transgenic lines when supplying emodin substrate.

The transcript profile of *FtUGT73BE5* was consistent with the accumulation trend of rutin in plant. It may relate to anti-adversity for its transcript level was up-regulated by MeJA, but repressed by ABA. Our work about glycosylation for rutin and emodin glucosides in tartary buckwheat not only

illuminates their biosynthetic pathway and biological interests, but also provides a theory basis for developing healthy food products with tartary buckwheat.

Acknowledgements

Illumina sequencing was done at Novogene co. ltd. This research was supported by Beijing Natural Science Foundation (7192138), the National Natural Science Foundation of China (81703647, 31471562), the Fundamental Research Funds for the Central public welfare research institutes (ZZ13-YQ-097), National Key R&D Program of China (2019YFC1711100), and National Science and Technology Major Project "Creation of Major New Drugs" (2019ZX09201005-006-004, No. 2019ZX09731-002).

Accession numbers

The GenBank accession numbers of sequences for this research as follow, FtUGT714C1, MH197430; FtUGT85A68, MH197422, FtUGT85A66, MH197421; FtUGT75R3, MH197425; FtUGT75R2, MH197426; FtUGT89G2, MH197419; FtUGT95D1, MH197417; FtUGT73E11, MH197413; FtUGT85AC1, MH197420; FtUGT79J1, MH197416; FtUGT73BE5, MH197414; FtUGT85AD3, MH197423; FtUGT85AD1, MH197424; FtUGT93R1, MH197429. SRA accession of transcripts data from different tissues of *F. tataricum*: SRP141278.

References

AL-Snafi, A.E. (2017). Therapeutic importance of *Ephedra alata* and *Ephedra foliate*. *Indo American Journal of Pharmaceutical Sciences, 4,* 399-406.

An, D.G., Yang, S.M., Kim, B.G., & Ahn, J.H. (2016). Biosynthesis of two quercetin *O*-diglycosides in *Escherichia coli*. *Journal of Industrial Microbiology & Biotechnology, 43,* 841-849.

Bringmann, G., & Irmer, A. (2008). Acetogenic anthraquinones: biosynthetic convergence and chemical evidence of enzymatic cooperation in nature. *Phytochemistry Reviews, 7,* 499-511.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research, 42,* W252-W258.

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., & Martens. S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *The Plant Journal, 69,* 1030-1042.

Chang, A., Singh, S., Phillips, G.N., Jr., & Thorson, J.S. (2011). Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Current Opinion in Biotechnology, 22,* 800-808.

Chen, Q.F. (2018). The status of buckwheat production and recent progresses of breeding on new type of cultivated buckwheat. *Journal of Guizhou Normal University Natural Sciences*, *362*, 1-7.

Dhaubhadel, S., Farhangkhoee, M., & Chapman, R. (2008). Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. *Journal of Experimental Botany, 59,* 981-994.

Fabjan, N., Rode, J., Kosir, I.J., Wang, Z., Zhang, Z., & Kreft, I. (2003). Tartary buckwheat *Fagopyrum tataricum* Gaertn. as a source of dietary rutin and quercitrin. *Journal of Agricultural and Food Chemistry, 51,* 6452-6455.

Frydman, A., Weisshaus, O., Bar-Peled, M., Huhman, D.V., Sumner, L.W., Marin, F.R., Lewinsohn, E., Fluhr, R., Gressel, J., & Eyal, Y. (2004). Citrus fruit bitter flavors: isolation and functional

528 characterization of the gene Cm1,2Rha1 encoding a 1,2 rhamnosyltransferase, a key enzyme
529 in the biosynthesis of the bitter flavonoids of citrus. *The Plant Journal, 40,* 88-100.
530 Funaki, A., Waki, T., Noguchi, A., Kawai, Y., Yamashita, S., Takahashi, S., & Nakayama, T.
531 (2015). Identification of a highly specific isoflavone 7-*O*-glucosyltransferase in the soybean
532 *Glycine max* (L.) Merr. *Plant and Cell Physiology, 56,* 1512-1520.
533 Ghimire, G.P., Koirala, N., Pandey, R.P., Jung, H.J., & Sohng, J.K. (2015). Modification of emodin
534 and aloe-emodin by glycosylation in engineered *Escherihia coli*. *World Journal of*
535 *Microbiology and Biotechnology, 31,* 611-619.
536 Griesser, M., Vitzthum, F., Fink, B., Bellido, M.L., Raasch, C., Munoz-Blanco, J., & Schwab, W.
537 (2008). Multi-substrate flavonol *O*-glucosyltransferases from strawberry *Fragaria* x *ananassa*.
538 achene and receptacle. *Journal of Experimental Botany, 59,* 2611-2625.
539 Jing, R., Li, H.Q., Hu, C.L., Jiang, Y.P., Qin, L.P., & Zheng, C.J. (2016). Phytochemical and
540 pharmacological profiles of three Fagopyrum Buckwheats. *International Journal of Molecular*
541 *Sciences, 17,* E589.
542 Koja, E., Ohata, S., Maruyama, Y., Suzuki, H., Shimosaka, M., & Taguchi, G. (2018).
543 Identification and characterization of a rhamnosyltransferase involved in rutin biosynthesis in
544 *Fagopyrum esculentum* (common buckwheat). *Bioscience Biotechnology and Biochemistry, 4,*
545 1-13.
546 Kovinich, N., Saleem, A., Arnason, J.T., & Miki, B. (2010). Functional characterization of a
547 UDP-glucose:flavonoid 3-*O*-glucosyltransferase from the seed coat of black soybean *Glycine*
548 *max* L. Merr. *Phytochemistry, 71,* 1253-1263.
549 Kubo, A., Arai, Y., Nagashima, S., & Yoshikawa, T. (2004). Alteration of sugar donor specificities
550 of plant glycosyltransferases by a single point mutation. *Archives of Biochemistry and*
551 *Biophysics, 429,* 198-203.
552 Lee, W., Ku, S.K. & Bae, J.S. (2013). Emodin-6-*O*-*β*-D-glucoside down-regulates endothelial
553 protein C receptor shedding. *Archives of Pharmacal Research, 36,* 1160-1165.

554 Liu, C.L., Chen, Y.S., Yang, J.H., & Chiang, B.H. (2008). Antioxidant activity of Tartary
555 *Fagopyrum tataricum* (L.) Gaertn. and common *Fagopyrum esculentum* Moench. buchwheat
556 sprouts. *Journal of Agricultural and Food Chemistry, 56,* 173-178.
557 Mizohata, E., Okuda, T., Hatanaka, S., Nakayama, T., Horikawa, M., Nakayama, T., Ono, E., &
558 Inoue, T. (2013). Crystallization and preliminary X-ray crystallographic analysis of
559 UDP-glucuronic acid:flavonol-3-*O*-glucuronosyltransferase VvGT5 from the grapevine *Vitis*
560 *vinifera*. *Acta Crystallographica Section F Structural Biology Communications, 69,* 65-68.
561 Nagatomo, Y., Usui, S., Ito, T., Kato, A., Shimosaka, M., & Taguchi, G. (2014). Purification,
562 molecular cloning and functional characterization of flavonoid C-glucosyltransferases from
563 *Fagopyrum esculentum* M. buckwheat. cotyledon. *The Plant Journal, 80,* 437-448.
564 Owens, D.K., & McIntosh, C.A. (2009). Identification, recombinant expression, and biochemical
565 characterization of a flavonol 3-*O*-glucosyltransferase clone from *Citrus paradisi*.
566 *Phytochemistry, 70,* 1382-1391.
567 Pang, Y.Z., Peel, G.J., Sharma, S.B., Tang, Y.H., & Dixon, R.A. (2008). A transcript profiling
568 approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of
569 *Medicago truncatula*. *Proceedings of the National Academy of Sciences of the United States of*
570 *America, 105,* 14210-14215.
571 Peng, L.X., Wang, J.B., Hu, L.X., Zhao, J.L., Xiang, D.B., Zou, L., & Zhao, G. (2013). Rapid and
572 simple method for the determination of emodin in Tartary buckwheat *Fagopyrum tataricum*.
573 by high-Performance liquid chromatography coupled to a diode array detector. *Journal of*
574 *Agricultural and Food Chemistry, 61,* 854-857.
575 Ravi, K.P., &Mukash, G. (2012). NGS QC Toolkit: a toolkit for quality control of next generation
576 sequencing data. *PLoS One, 7,* e30619.
577 Rautengarten, C., Ebert, B., Moreno, I., Temple, H., Herter, T., Link, B., Doñas-Cofré, D., Moreno,
578 A., Saéz-Aguayo, S., Blanco, F., Mortimer, J.C., Schultink, A., Reiter, W. D., Dupree, P.,
579 Pauly, M., Heazlewood, J.L., Scheller, H.V., & Orellana, A. (2014) The Golgi localized

bifunctional UDP-rhamnose/UDP-galactose transporter family of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 11563-11568.

Rodas, F.R., Rodriguez, T.O., Murai, Y., Iwashina, T., Sugawara, S., Suzuki, M., Nakabayashi, R., Yonekura-Sakakibara, K., Saito, K., Kitajima, J., Toda, K, & Takahashi, R. (2014). Linkage mapping, molecular cloning and functional analysis of soybean gene *Fg2* encoding flavonol 3-*O*-glucoside 1> 6. rhamnosyltransferase. *Plant Molecular Biology, 84,* 287-300.

Sharma, R., Panigrahi, P., & Suresh, C.G. (2014). In-silico analysis of binding site features and substrate selectivity in plant flavonoid-3-*O* glycosyltransferases F3GT through molecular modeling, docking and dynamics simulation studies. *Plos One, 9,* e92636.

Sun, W., Leng, L., Yin, Q., Xu, M., Huang, M., Xu, Z., Zhang, Y., Yao, H., Wang, C., Xiong, C., Chen, S., Jiang, C., Xie, N., Zheng, X., Wang, Y., Song, C., Peters. R.J., & Chen, S. (2019). The genome of the medicinal plant *Andrographis paniculata* provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *The Plant Journal, 97,* 841-857

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution, 24,* 1596-1599.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology, 28,* 511-555.

Tsuji, K., & Ohnishi, O. (2000). Origin of cultivated Tatary buchwheat *Fagopyrum tataricum* Gaertn. revealed by RAPD analysis. *Genetic Resources and Crop Evolution, 47*, 431-438.

Wu, B., Gao, L., Gao, J., Xu, Y., Liu, H., Cao, X., Zhang, B., & Chen, K. (2017). Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach *Prunus persica* L. Batsch. *Frontiers in Plant Science, 8,* 389.

Xie, K.B., Chen, R.D., Li, J.H., Wang, R.S., Chen, D.W., Dou, X.X., & Dai, J.G. (2014). Exploring the catalytic promiscuity of a new glycosyltransferase from *Carthamus tinctorius*. *Organic Letter, 16,* 4874-4877.

Xu, J., Chu, Y., Liao, B., Xiao, S., Yin, Q., Bai, R., Su, H., Dong, L., Li, X., Qian, J., Zhang, J., Zhang, Y., Zhang, X., Wu, M., Zhang, J., Li, G., Zhang, L., Chang, Z., Zhang, Y., Jia, Z., Liu, Z., Afreh, D., Nahurira, R., Zhang, L., Cheng, R., Zhu, Y., Zhu, G., Rao, W., Zhou, C., Qiao, L., Huang, Z., Cheng, Y.C., & Chen, S. (2017). *Panax ginseng* genome examination for ginsenoside biosynthesis. *GigaScience, 6,* 1-15.

Yin, Q.G., Shen, G.A., Chang, Z.Z., Tang, Y.H., Gao, H.W., & Pang, Y.Z. (2017). Involvement of three putative glucosyltransferases from the UGT72 family in flavonol glucoside/rhamnoside biosynthesis in *Lotus japonicus* seeds. *Journal of Experimental Botany, 68,* 594-609.

Zhang, D., Jiang, C., Huang, C., Wen, D., Lu, J., Chen, S., Zhang, T., Shi, Y., Xue, J., Ma, W., Xiang, L., Sun, W., Chen, S. (2019) The light - induced transcription factor FtMYB116 promotes accumulation of rutin in *Fagopyrum tataricum*. *Plant Cell Environment*, 42, 1340-1351.

Zhang, L.J., Li, X.X., Ma, B., Gao, Q., Du, H.L., Han, Y.H., Li, Y., Cao, Y., Qi, M., Zhu, Y., Lu, H., Ma, M., Liu, L., Zhou, J., Nan, C., Qin, Y., Wang, J., Cui, L., Liu, H., Liang, C., Qiao, Z. (2017). The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Molecular Plant, 10,* 1224-1237.

Zhang, W., Ye, M., Zhan, J., Chen, Y., &Guo, D. (2004). Microbial glycosylation of four free anthraquinones by *Absidia coerulea*. *Biotechnology Letters, 26,* 127-131.

Zhou, J., Li, C.L., Gao, F., Luo, X.P., Li, Q.Q., Zhao, H.X., Yao, H.P., Chen, H., Wang, A.H., & Wu, Q. (2016). Characterization of three glucosyltransferase genes in Tartary buckwheat and their expression after cold stress. *Journal of Agricultural and Food Chemistry, 64,* 6930-6938.

Figure Captions

Figure 1 Phylogenetic analysis of UGTs in *F. tataricum*

Amino acid sequences were analyzed using Clustal X2 and the bootstrap consensus tree was constructed using MEGA 4.0 with the neighbor-joining method and 1000 bootstrap replicates. The respective protein names and numbers are listed in Supplementary Data S1. UGTs from *F. tataricum* are marked in green and those from *Arabidopsis* in black. The groups are marked in different colors.

Figure 2 Hierarchical clustering analysis of transcript levels of UGT genes in different tissues and seed development stages of *F. tataricum*.

FtUGTs marked green were the candidates for further study. R was the abbreviation of roots; St, stem; L, leaf; F, flower; Sd1-Sd3 means prophase, mid-term, later period of ovary bulge, Sd4-Sd6 means prophase, mid-term, later period of grout, and Sd7-Sd9 means prophase, mid-term, later period of mature. The clustering was performed using Heatmap2. The heat map shows relative transcript level of FtUGT genes in various tissues. The color scale (-3 to 3 in green to red color) represents Z-score-normalized gene expression. Dendrograms along the top and left sides of the heat map indicate the hierarchical clustering of tissues and genes.

Figure 3 Identification of the enzymatic product of recombinant FtUGT73BE5 protein by HPLC and UPLC/MS/MS

Diagram A and D represent HPLC chromatographs of the enzymatic products containing quercetin (Q), quercetin 3-*O*-glucoside (QG) and emodin (E) with recombinant FtUGT73EB5, respectively. Diagrams B, C and E are the tandem mass spectrum of the enzymatic products of recombinant FtUGT73EB5 protein with Q, QG and E determined by UPLC/MS/MS, respectively. G, Glucoside.

Figure 4 The characteristics of FtUGT73BE5 proteins

(A) Kinetic parameters of the recombinant FtUGT73BE5 and FtUGT75R2 proteins with flavonol aglycones or emodin as acceptor substrate and UDP-glucose as donor substrate. Values represent the means ± SD from triplicate enzymatic assays. (B) the proposal glucosylation of rutin and emodin 6-*O*-glucoside. (C) Representative binding domain (left panel) and the overall view of the structural model of FtUGT73BE5 docking with UDP-glucose and quercetin 3-*O*-glucose. The amino acids in the form of ligands with UDP-glucose donor and acceptor are highlighted in black. UPG was abbreviated from UDP-glucose (blue stick); Q, quercetin; QG, quercetin 3-*O*-glucose (green stick). KG, kaempferol 3-*O*-glucose; E, emodin.

Figure 5 Identification of FtUGT73BE5 *in vivo*

(A) Transgenic hairy roots of *F. tataricum* detected under bright field and GFP florescence. (B) Transcript levels of FtUGT73BE5 in control (CK) and transgenic hairy roots (overexpression) detected by qRT-PCR. (C) Flavonol glucosides content in overexpressing hairy roots; left, the content of quercetin 3-O-glucoside (QG); right, the content of rutin. (D) Representative MS chromatographs of the metabolic extracts from transgenic line (first panel) and empty vector (CK, second panel), the third and last panel shows the MS chromatographs of enzymatic product and emodin (E). (E) Contents of emodin 6-*O*-glucoside (E 6-*O*-G) in vector control (CK) and transgenic hairy roots supplied with emodin.

Figure 6 Expression profiles of FtUGTs and its relationship with hormones

(A) Relative expression of *FtUGT73BE5* and *FtUGT75R2*, and rutin contents in various tissues of *F. tataricum.* (B) The relative expression of *FtUGT73BE5* and *FtUGT75R2* in seedlings after treated by ABA(4h, 14h, 24h) and MeJA (4h, 14h, 24h). (C) The contents of rutin, quercetin and quercetin 3-*O*-glucoside in seedlings after treated by ABA (4h, 14h, 24h) and MeJA (4h, 14h, 24h).

SUPPLEMENTARY DATA

Supplementary Table S1 Candidate FtUGTs by phylogenetic and transcriptomics analysis

The bold UGTs were cloned successfully.

Supplementary Table S2 The enzymatic activity of five recombinant FtUGT proteins toward flavonoids and emodin with UDP-glucose

4-coumaric acid (4-C), naringenin (Na), kaempferol (Ka), quercetin (Q), quercetin 3-O-glucoside (QG), quercitrin (QG'), catechin (C), flavone (Fl, CAS number, 525-82-6) , apigenin(A) , luteolin (L) and emodin (E). A dash (-) indicates the assay was not done.

Supplementary Table S3 The information of ion fragment of flavonoids and emodin for QqQ quantification

Supplementary Data S1 Transcripts and names of FtUGTs

Supplementary Data S2 Primer sequences used in this study

Supplementary Figure S1 Key genes involved in rutin and emodin glucoside schematic biosynthetic pathway of *F. tataricum*

(A) the general biosynthetic pathway of rutin. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; FLS, flavonol synthase; UGT, UDP glycosyltransferase. (B) the general biosynthetic pathway of emodin. PKS, polyketone synthase; CYP450, cytochrome P450; E6G, emodin 6-O-glucose; Gly, glucose. Solid lines mean one step; dashed lines mean multiple steps.

Supplementary Figure S2 Intron distribution of 106 UGT genes in *F.tataricum*

Yellow column representative CDS (coding sequence); purple column, UTR (untranslated coding regions); line, intron; red column, 5'UTR; green column, starting codon; orange column, stop codon; grey column, 3'UTR.

Supplementary Figure S3 Deduced chromosomal positions of FtUGT genes

Supplementary Figure S4 Sequence analysis and protein expression of candidate FtUGTs

(A) Represents of Molecular phylogenetic tree of identified FtUGTs. (B) PSPG boxes alignment of cloned FtUGTs. (C) The SDS-PAGE gel map of FtUGTs possessing activity to the tested substrates. Multiple sequences were aligned by Clustal X2 and used for tree constructions with the maximum likelihood method using MEGA4. Bootstrap values from 1000 replicates are shown on branches (values > 50% are shown). Bar represents 0.5 amino acid substitutions per site. GenBank accessions are as follows: At3G7GT, NP_181217.1[*Arabidopsis thaliana*]; At5GT, NP_193146.1[*Arabidopsis*

thaliana]; At7GT, NP_567955.1[*Arabidopsis thaliana*]; Cm1,2Rha1, Q8GVE3.2[*Citrus maxima*]; Ct3GT, 3WC4_A[*Clitoria Ternatea*]; GelF7GT, BAC78438.1[*Glycyrrhiza echinata*]; Ih3GT, BAD83701.1[*Iris* x *hollandica*]; Ip3GGT, Q53UH5.1[*Ipomoea purpurea*]; PfA5GT, Q9ZR27.1[*Perilla frutescens*]; VhA5GT, Q9ZR25.1[*Glandularia x hybrida*]; VmF3GT, BAA36972.1[*Vigna mungo*]; Vv3GT, P51094.2[*Vitis vinifera*]; Zm3GT, P16165.1[*Zea mays*]; Sb7GT, BAA83484.1*[Scutellaria baicalensis*]; ThA5GT, BAC54093.1[*Torenia* hybrid cultivar]; PhA3G6"RT, CAA50376.1[*Petunia x hybrida*]; PhA5GT, BAA89009.1[*Petunia x hybrida*]; Fi3GT, AAD21086.1[*Forsythia x intermedia*]; Hv3GT, P14726.1[*Hordeum vulgare*]; Cm7G2"RT, ACX70154.1[*Citrus maxima*]; IpA3G2"GT, Q53UH5.1[*Ipomoea purpurea*]; At7RT, NP_563756.1[*Arabidopsis thaliana*]; At3RT, NP_564357.1[*Arabidopsis thaliana*]; Lb7GlcT, BAG80536.1[*Lycium barbarum*]; Am7GT, BAG16513.1*[Antirrhinum majus*]; CtUGT73AE1, AJT58578.1[*Carthamus tinctorius*]. FtUFGT2, AOS85163.1[*Fagopyrum tataricum*]; FtUGT714B1,AOS85162.1[*Fagopyrum tataricum*]; FtUGT714C1, AOS85164.1[*Fagopyrum tataricum*]; FtUGT71AG1, APY21160.1[*Fagopyrum tataricum*]; FtUGT71U4, APY21161.1[*Fagopyrum tataricum*]; FeUGT708C1, BAP90360.1[*Fagopyrum esculentum*]; FeUGT708C2,BAP90361.1[*Fagopyrum esculentum*]; FeCGTv, BAP90374.1[*Fagopyrum esculentum*];

Supplementary Figure S5 MS detection of Products catalyzed by rUGT73BE5 with others substrates

Supplementary Figure S6 MS detection of Products catalyzed by rUGT75R2 with others substrates

Supplementary Figure S7 *In vitro* activity identification of FtUGT73BE5 related to emodin

Glucosylation site verification of products by HPLC (A) and MS (B); panels of HPLC (A) are FtUGT73BE5 with E (emodin), E (authentic emodin), E 1-*O*-G (authentic emodin 1-O-glucoside), and E 8-*O*-G (authentic emodin 8-O-glucoside) from top to bottom; MS chromatograms of E 1-*O*-G (first panel), E 8-*O*-G (second panel), product (third panel) and E (last panel) are shown in B.

Supplementary Figure S8 The characterization of rUGT73BE5

(A) Represents optimum temperature for rUGT73BE5 for quercetin 3-*O*-glucoside and emodin; the assay mix (100 μL) contained 2.5 μg purified protein, 2 mM UDP-glucose, 10 mM DTT, 100 μM quercetin or emodin and 50 mM Tris-HCl (pH 7.0) for 30 min. (B) shows optimum pH value for rUGT73BE5, the reaction was carried out at 37 ℃ for 30 min. (C) The reaction rates of rUGT73BE5 toward 10 substrates, the reaction was carried out at 37 ℃ for 30 min with 50 mM Tris-HCl (pH 7.0). naringenin (Na), kaempferol (Ka), quercetin (Q), quercetin 3-O-glucoside (QG), quercitrin (QG'), catechin (C), flavone (Fl, CAS number, 525-82-6) , apigenin(A) , luteolin (L) and emodin (E).

Supplementary Figure S9 The inhibition of rutin biosynthesis in *FtUGT73BE5*-RNAi hairy roots

(A) Transcript levels of FtUGT73BE5 in control (CK) and RNAi hairy roots detected by qRT-PCR. B&C represent contents of quercetin 3-*O*-glucoside and rutin in transgenic lines, respectively.

Credit Author Statement

QY, XH, WS and SC: Supervision, conceived and designed the study. QY, XH, ZH, and CX: Investigation, performed the experiments and analyzed the data. QY: Writing-Original draft preparation, Writing- Reviewing and Editing. QC, YS, GD and CS: Writing-Reviewing and Editing. QC and GD: Resources, provided the seeds of tartary buckwheat. HG and TZ: Data curation, Formal analysis, analyzed the UGTs information from transcriptome and genome. All authors read and approved the manuscript.

1. Temporal and spatial transcripts of 106 FtUGTs were analyzed in genome level.

2. FtUGT73BE5 was identified to dominate the first step of rutin glycosylation.

3. FtUGT73BE5 was demonstrated to catalyze emodin to form emodin 6-*O*-glucoside.

4. MeJA up-regulates transcript of *FtUGT73BE5*; it was related to anti-adversity.