Title: Purification, molecular cloning, and functional characterization of flavonoid *C*-glucosyltransferases from buckwheat (*Fagopyrum esculentum*) cotyledon.

Authors: Yoshihisa Nagatomo, Shiori Usui, Takamitsu Ito, Akira Kato, Makoto Shimosaka, Goro Taguchi

Addresses: Division of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, 3-15-1 Tokida, Ueda 386-8567, Japan.

For correspondence: Goro Taguchi, Division of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, 3-15-1 Tokida, Ueda 386-8567, Japan. Tel: +81-268-21-5342, Fax: +81-268-21-5331, E-mail address: gtagtag@shinshu-u.ac.jp

Running title: Flavonoid *C*-glucosyltransferase from buckwheat

Key words: Flavonoid *C*-glucosyltransferase; buckwheat; *Fagopyrum esculentum*; cotyledon; 2-hydroxyflavanone; germination; vitexin

Genbank/EMBL/DDBJ accession nos.

FeCGTa (AB909375), FeCGTb (AB909376), FeCGT_variant1 (AB909389), FeCGT_variant2 (AB909390), Fe-gapdh (AB919116).

Summary

C-glycosides are characterized by their C–C bonds in which the anomeric carbon of the sugar moieties is directly bound to the carbon atom of aglycon. *C*-glycosides are remarkably stable, as their C–C bonds are resistant to glycosidase or acid hydrolysis. A variety of plant species are known to accumulate *C*-glycosylflavonoids; however, the genes encoding for enzymes that catalyze *C*-glycosylation of flavonoids have been identified only from *Oriza sativa* (rice) and *Zea mays* (maize), and have not been identified from dicot plants. In this study, we identified *C*-glucosyltransferase gene from a dicot plant, buckwheat (*Fagopyrum esculentum* M.). We purified two isozymes from buckwheat seedlings that catalyze *C*-glucosylation of 2-hydroxyflavanones, which are expressed specifically in the cotyledon during seed germination. Following purification we isolated the cDNA corresponding to each isozyme [*FeCGTa* (UGT708C1) and *FeCGTb* (UGT708C2)]. When expressed in *Escherichia coli*, both proteins demonstrated *C*-glucosylation activity toward 2-hydroxyflavanones, dihydrochalcone, trihydroxyacetophenones, and other related compounds with chemical structures similar to 2′,4′,6′-trihydroxyacetophenone.

Molecular phylogenetic analysis of plant glycosyltransferases shows that flavonoid *C*-glycosyltransferases form a different clade with other functionally analyzed plant glycosyltransferases.

Introduction

Glycosylation is one of major modifications of phytochemicals and is crucial for multiple physiological properties and functions, including increased solubility and stability, regulation of bioactivity, transport and accumulation, and detoxification of harmful compounds (Gachon *et al*., 2005; Bowles *et al*., 2005). The majority of plant glycosides are *O*-linked, with their sugar moieties linked to the hydroxyl group of aglycons. However, additional glycosides *i.e.,* carboxyl-, *N*-, *S*-, or *C*-glycosides, exist (Jones and Vogt, 2000). Glycosylation reactions are usually catalyzed by uridine diphosphate (UDP)-sugar: glycosyltransferases (UGTs) belonging to family 1 glycosyltransferases, which utilize UDP-sugar as an activated donor of sugar moieties (Paquette *et al*., 2003; Yonekura-Sakakibara and Hanada, 2011).

C-glycosides' sugar moieties are directly bound to the carbon atom of aglycon, forming a C–C bond at the anomeric carbon (Franz and Grün, 1983; Hultin, 2005). Compare to other glycosides such as *O*-glycosides, *C*-glycosides are remarkably stable, as their C–C bonds are resistant to glycosidase or acid hydrolysis (Jay, 1994).

C-glycosides are widely distributed in nature and function as bioactive

molecules in bacteria, insects and plants (Hultin, 2005). For examples, a *C*-glycosylpeptide (enterobactin) works as a bacterial siderophore in iron transport in *E. coli* (Fischbach *et al*., 2005); an angucycline-*C*-glycoside (urdamycin) works as an antibiotic in *Streptomyces fradiae* (Faust *et al.* 2000); an anthoraquinone *C*-glucoside (carminic acid) works as a feeding deterrent in *Dactylopius coccus* (the cochineal insect; Eisner, *et al*., 1980), *C*-glycosylflavonoids function as co-pigments of flower color formation in plants (Jay, 1994).

Many different *C*-glycosides have been isolated from plants, including flavonoid, anthrone, xanthone, chromone, and gallic acid *C*-glycosides (Franz and Grün, 1983). Among them, flavonoid-*C*-glycosides are prevalent compounds and have been identified in bryophytes, pteridophytes, gymnosperms, and angiosperms (Jay, 1994; Jay *et al.*, 2005). They have several biological functions, such as anti-oxidant activities (Zhang *et al.*, 2005; Watanabe, 2007), antifungal activities as a phytoalexin (McNally *et al.*, 2003), and inhibition of caterpillar growth and radicle growth (Byrne *et al*., 1996; Khan *et al.*, 2008). They also have several medicinal properties, including those against hypertension (Prabhakar *et al.*, 1981), inflammation (Prabhakar *et al.*, 1981), oxidation

(Kim *et al.*, 2005), and obesity (Kim *et al.*, 2010). Thus, from a pharmaceutical standpoint, there has been much focus on the development of an enzymatic synthesis of flavonoid *C*-glycosides.

A reaction mechanism for *C*-glycosyltransferase (CGT)-catalyzed *C*-glycosylation of flavonoids has been proposed as follows (Kerscher and Franz, 1987): Flavanone substrates are hydroxylated at the 2-position of the flavanone skeleton (Du *et al*., 2010). The resulting 2-hydroxylflavanone is in equilibrium with its open-circular form (dibenzoylmethane-form), which is the direct substrate of CGT (Figure 1a). After conjugation of the sugar-moiety, closed-circular form of the product (2-hydroxylflavanone *C*-glycoside) is dehydrated to produce flavone *C*-glycoside either spontaneously or enzymatically.

A gene coding for flavonoid *C*-glucosyltransferase was identified and characterized from *Oryza sativa* (rice; *OsCGT*) and it was demonstrated that the 2-hydroxyflavanones are potential substrates of this enzyme (Brazier-Hicks *et al.*, 2009). This study also reported an involvement of dehydratase activity toward the product in rice, suggesting that a successive dehydration would enzymatically proceed. More

recently, a CGT was reported from *Zea mays* (maize; UGT708A6), exhibiting *O*-glucosylation activity toward naringenin in addition to *C*-glucosylation activity toward 2-hydroxynaringenin (Falcone Ferreyra *et al.*, 2013). Several groups have attempted to isolate *CGT* genes (He *et al*., 2011; Hamilton *et al*., 2012); however, they have not yet been identified from dicot plants. Thus, information regarding the CGT responsible for plant *C*-glycosylation is still limited

Buckwheat (*Fagopyrum* sp.) is a pseudocereal, and buckwheat achenes are used as food such as buckwheat noodles ("soba" in Japanese) or buckwheat groats. They are cultivated world-wide, particularly in China, Russia, USA, and Europe (FAOSTAT; http://faostat3.fao.org/faostat-gateway/go/to/home/E). Buckwheat is known to accumulate large amounts of *C*-glucosylflavones in their cotyledons, such as vitexin, isovitexin, orientin, and isoorientin (Margna *et al*., 1990, Figure S1). The flavonoid *C*-glucosyltransferase was investigated using partially purified enzymes (Kerscher and Franz, 1988), but their coding genes have not been identified.

In this work, we characterized flavonoid *C*-glucosyltransferases from buckwheat (*F. esculentum*). We purified two isozymes from etiolated buckwheat

seedlings to near homogeneity, isolated their genes, and characterized the recombinant enzymes heterologously expressed in *E. coli*.

Results

Purification of C-glucosyltransferases from buckwheat cotyledon

We confirmed that *C*-glucosylation activity toward 2-hydroxynaringenin, a substrate for flavonoid *C*-glucosylation, was present mostly in the developing cotyledon during germination. In contrast, *O*-glucosylation activity toward quercetin, a major flavonoid aglycon in buckwheat, was identified in flowers, leaves, cotyledon, hypocotyl, and stem (Table S1). These results suggested that the expression of CGT genes would be cotyledon-specific. Initially, we attempted to isolate a gene coding for CGT from a cDNA library of buckwheat cotyledon by PCR using degenerate primers (Taguchi *et al.*, 2003) designed from the conserved motif of UGTs (Paquette *et al*., 2003). However, we did not identify UGT genes coding for proteins with CGT activity (isolated genes are listed in Table S2). We then attempted to purify CGT enzymes from etiolated buckwheat seedlings (2 kg fresh weight). We monitored CGT activity toward 2-hydroxypinocembrin (Figure 1b) because it is more stable than other 2-hydroxyflavanone substrates.

Following separation using Mono Q chromatography, we identified two peaks of CGT activity, which eluted between 100 mM and 150 mM of the NaCl gradient

(Figure S2a). SDS-PAGE analysis of fractions at the two peaks identified the major protein bands closely related with CGT activity with apparent molecular masses of approximately 50 kDa (Figure S2b). These two proteins were likely to be CGTs and were thus named CGTa (first peak) and CGTb (second peak). The full purification process resulted in an approximately 3,500- and 7,300-fold purification and 4% and 9% recovery for CGTa and CGTb, respectively (Figure 2, Table S3).

Characterization of the purified CGTs

We examined the substrate specificity of the purified enzymes using several phenolic compounds as substrates (Figure 1b). Both enzymes catalyzed *C*-glucosylation of 2-hydroxyflavones, phloretin, and structurally-related compounds thereof (Table 1). Neither enzyme utilized flavanones, flavones, or flavonols as glucose acceptors. Both enzymes showed maximum activities at pH 6.5–7.0 and each was stable at an alkaline pH (more than 70% level of the maximum) (Figure S3). CGTa and CGTb were active at an optimum temperature of 45–50°C and 50–55°C, respectively (Figure S3).

Determination of CGTs peptide sequences and isolation of their coding genes

Both CGTa and CGTb proteins were eluted from the SDS-polyacrylamide gel, treated with lysyl endopeptidase, and analyzed by Ultra Performance Liquid Chromatography (UPLC)-MS/MS. MS/MS data were analyzed by Proteinlynx (Waters) and the candidate peptide sequences of both enzymes were subjected to BLASTP search in the Genbank protein database. Two sequences (PNVDPFFLRYK and QEELLPWLDQQPEK) showed significant similarity to the internal sequences of UGTs in the database (Figure S4a), suggesting that CGTa and CGTb would be members of the UGT enzyme family.

The cDNAs coding for CGTa and CGTb were subsequently obtained by PCR using a cDNA library constructed from buckwheat cotyledon as a template. The deduced amino acid sequences were accord with the partial peptide sequences determined from each CGT. The two genes were named *FeCGTa* and *FeCGTb*, corresponding to the gene products CGTa and CGTb, respectively (Figure S4b).

FeCGTa and *FeCGTb* (both 1,374-bp) coded for proteins [FeCGTa (UGT708C1) and FeCGTb (UGT708C2), also referred to herein as FeCGTs] comprising 457 amino acid

residues with the calculated molecular masses of 50.37 kDa and 50.43 kDa, respectively. The enzymes showed a high degree of identity (96%) throughout the entire length of the protein sequences. The primary sequence of FeCGTs showed approximately 41% identity with that of OsCGT and 30%–35% identity with that of UGT88 and UGT72 enzymes. FeCGTs showed low identity with that of UGT73B4, which could explain the *C*-glucosylation of xenobiotics such as trinitrophenol in *Arabidopsis* (Gandia-Herrero *et al*., 2008). They showed very low identities with that of bacterial CGTs such as IroB (8%) and UrdGT2 (10%).

Properties of recombinant CGTs expressed in *E. coli*

The cDNA fragments corresponding to the entire coding regions of *FeCGTa* and *FeCGTb* were subcloned into the pET28c(+) vector and introduced into *E. coli* Rosetta 2(DE3) to study the enzymatic properties of the gene products. Both recombinant proteins showed glucosylation activities toward 2-hydroxynaringenin. This resulted in its conversion to two different glucosylated compounds, vitexin and isovitexin, which were separated into two peaks exhibiting the same pattern of $[M-H]^-$ ion at *m/z* 431

(Figure 3). Retention times of these two peaks completely coincided with those of authentic compounds. The glucosidic bonds of the products were stable for 2 h during 2 M HCl treatment at 60°C, in which quercetin 3-*O*-glucoside was hydrolyzed (Figure S5g). The reaction was stopped by the addition of methanol rather than acid, we observed a main peak that exhibited $[M-H]^-$ ion at *m/z* 449, corresponding to a glucoside of 2-hydroxynaringenin (Figure 3b). A fragment ion of $[M-H-120]^-$ at *m/z* 329 was also observed in the mass spectrum (Figure 3b), which is a fragment typical of *C*-glucoside (Li *et al*., 1991; Jay, 1994), suggesting that the product is 2-hydroxynaringenin *C*-glucoside. These results clearly indicate that *FeCGTs* code for flavonoid *C*-glucosyltransferases and that they do not catalyze the successive dehydration of the products to *C*-glucoside.

The recombinant FeCGTa displayed enzyme activity that is essentially independent of pH in the pH range of 6.0–11.0. FeCGTb displayed a similar trend with the activity showing the maximum at pH 6.0 and essentially leveling-off to a 70–80% level of the maximum activity between pH range of 6.0–11.0 (Figure S6a). The optimal reaction temperatures for FeCGTa and FeCGTb were 45°C and 50–55°C, respectively (Figure

S6b). Although most properties of FeCGTa and FeCGTb were very similar, their thermostabilities were slightly different. At 50°C/pH 6.5 for 30 min, FeCGTb was stable whereas FeCGTa lost 50% of its activity (Figure S6c). Thus, the activities and stabilities of recombinant FeCGTs were very similar to those of purified buckwheat purified enzymes.

The substrate preferences of the recombinant FeCGTs were examined using several phenolic compounds (Figure 1b) as substrates (Table 1). Both enzymes showed *C*-glucosylation activity toward the expected natural substrates, 2-hydroxylflavanones (2-hydroxylnaringenin, 2-hydroxyeriodictyol, and 2-hydroxypinocembrin). Upon termination of the reaction by acid addition, the product was dehydrated to *C*-glucosylflavones. Among the compounds tested, they also showed significant activity towards a dihydrochalcone (phloretin), and artificial compounds (2-phenyl-2′,4′,6′-trihydroxyacetophenone, 2′,4′,6′-trihydroxyacetophenone, and 2,4,6-trihydroxybenzaldehyde). All glucosidic bonds of these products were stable to acid treatment at 60°C for 2 h, suggesting that they are *C*-glucosides (Figure S5). In contrast, they did not utilize flavanones, flavones, or flavonols as glucose acceptors,

indicating that FeCGTs are unable to catalyze *O*-glucosylation.

We also examined the sugar donor preferences of the enzymes using 2-hydroxypinocembrin as a sugar acceptor (Table 1). Both recombinant FeCGTs utilized UDP-glucose most efficiently of all the UDP-sugars tested. They poorly utilized UDP-xylose (1%–4% activity as compared to UDP-glucose) and did not utilize UDP-galactose and UDP-glucuronic acid at all. The kinetic parameters of recombinant FeCGTs were determined using UDP-glucose as a glucose donor and either 2-hydroxynaringenin, 2-hydroxypinocembrin, or 2-phenyl-2′,4′,6′-trihydroxyacetophenone as a glucose acceptor (Table 2). The results demonstrated that both CGTs showed significant activities toward all of these compounds.

Organ-specific expression of FeCGTs

In order to determine where *FeCGTs* transcript was expressed, total RNA was extracted from several buckwheat organs, including the flower, leaf, stem, root, hypocotyl and cotyledon, and then RT-PCR was performed (Figure 4a). *FeCGTs* were expressed only

in cotyledons, which is in good agreement with the observation that *C*-glucosylflavones accumulate in this organ (Figure S1). We further examined the accumulation of these transcripts in the cotyledon during seed germination (Figure 4b, pictures of each stage were shown in Figure S7). The transcripts began to accumulate when seeds started to germinate, remained stable during cotyledon formation, and diminished when the cotyledon was developed. These results clearly indicate that FeCGTs are *C*-glucosyltransferases responsible for the biosynthesis of *C*-glucosylflavones in buckwheat cotyledon.

Discussion

C-glycosides have been identified in diverse plant species; however, the UGTs responsible for *C*-glycosylation were poorly understood because of a lack of information regarding CGT. Only two genes coding for CGT have been reported so far from rice (*OsCGT*, Brazier-Hicks *et al.*, 2009) and maize (*UGT708A6*, Falcone Ferreyra *et al*., 2013). In this study, we purified two isozymes of flavonoid CGTs, which catalyze *C*-glucosylation of 2-hydroxyflavanone from buckwheat cotyledons. Moreover, we isolated the two genes (*FeCGTa* and *FeCGTb*), which likely correspond to the purified enzymes (CGTa and CGTb). The deduced amino acid sequences of *FeCGTs* were quite similar (96% identical) and the properties of both recombinant enzymes were also similar except for their thermostability. It should be noted that hybrid genes arising through a possible recombination event between *FeCGTa* and *FeCGTb* were also observed in the cDNA library (Figure S8). These results led us to hypothesize that the two genes may be allelic and result from highly-frequent recombination, which occurs in allogamy plants such as buckwheat.

The properties of the recombinant and purified FeCGTs were similar to each other, except in the relative activity on some substrates (Table 1). The purified CGTs

showed relatively lower activities against 2-hydroxynaringenin and 2-hydroxyeriodictyol than the recombinant FeCGTs. As the purification steps required more than two weeks, enzymes could not be freed from partial inactivation causing a decrease of specific activities. Thus lower activities of the purified enzymes against 2-hydroxynaringenin seem to be reasonable, however, we have no rational explanation for their high activity against 2-hydroxypinocembrin.

The properties of FeCGTs were slightly different from those of *C*-glucosyltransferase partially purified from buckwheat by Kerscher and Franz (1988). They reported that the enzyme had a molecular mass of 41 kDa and the activity peaked at pH 10, which is contrary to our data that FeCGTs had molecular masses of 50 kDa and a broad pH preference between the pH range of 6.0–11.0. The increased activity in the alkaline condition for the partially purified enzyme could be related to the solubility of the substrate. We have no rational explanation for these conflicting results, but FeCGTs are likely to work even in neutral to alkaline condition up to pH 11. This study also demonstrated that the enzyme utilized UDP-galactose as a sugar donor, but FeCGTs did not.

FeCGTs utilized not only 2-hydroxyflavanones that are plausible natural substrates in buckwheat but also structurally related compounds, including dihydrochalcones and 2′,4′,6′-trihydroxyacetophenones. In contrast, they did not utilize 2′,4′-dihydroxyacetophenone, suggesting that a 2′,4′,6′-trihydroxyacetophenone-like structure (Figure 1b) would be essential for substrate recognition. A 2′,4′-dihydroxyacetophenone lacking one of the hydroxyl-groups on the benzene ring would not be sufficient to form the nucleophilic aryl group that attacks the anomeric carbon of UDP-glucose (Gutmann and Nidetzky, 2012).

Both FeCGTs and reported OsCGT (Brazier-Hicks *et al.*, 2009) catalyze the *C*-glucosylation of flavonoids, though they showed only 40% identities in their amino acid sequences. The k_{cat}/K_m of FeCGTs on 2-hydroxyflavanones values were similar to that of OsCGT, suggesting their similar roles *in planta*. FeCGTs and OsCGT showed strong activities toward phloretin; however, only FeCGTs demonstrated a strong activity toward 2′,4′,6′-trihydroxyacetophenone-like compounds (Table 2, Brazier-Hicks *et al.*, 2009), suggesting that FeCGTs has broader substrate specificity than OsCGT. Recently, a production of flavone *C*-glucosides from yeast cells was reported, in which

recombinant OsCGT was expressed (Brazier-Hicks *et al*., 2013). In addition, OsCGT was used for *in vitro* production of phloretin *C*-glycosides (Bungaruang *et al*., 2013). Similarly, FeCGT would also be a useful tool for the production of various *C*-glucosides.

The plausible biosynthetic pathway of flavone *C*-glucosides is shown in Figure 1a. Recently, Falcone Ferreyra *et al.* (2013) proposed that the actual substrate of maize CGT (UGT708A6) is a closed form of 2-hydroxyflavanones. This may indeed be the case because UGT708A6 was able to catalyze 7-*O*-glucosylation of flavanone, a compound related to a closed form of 2-hydroxyflavanones. This study also suggested that the dehydration of the glucosylated product spontaneously proceeds, because they could not detect the 2-hydroxyflavanone *C*-glycosides following the reaction. In contrast, our results clearly showed that FeCGT produced 2-hydroxyflavanone *C*-glucosides (Figure 3) similar to that demonstrated with OsCGT (Brazier-Hicks *et al*., 2009). Our results also suggest that an open form of 2-hydroxyflavanones is an actual substrate of FeCGT, because it is able to catalyze *C*-glycosylation of several compounds possessing a 2′,4′,6′- trihydroxyacetophenone-like structure and is not able to catalyze

C-glycosylation of naringenin (Table 2, Figure 1a). We cannot rule out the possibility that each enzyme has a different reaction mechanism towards substrates with similar structures, because the enzymatic characteristics of UGT708A6 toward these compounds have not been reported. In addition, a ratio of 8-*C*-glucosides/6-*C*-glucosides (vitexin/isovitexin) was determined to be 1/1.87 (n=10) in buckwheat cotyledons by HPLC analysis (Figure S1). This ratio was significantly different ($p<0.01$, Student's t-test) from that of the products obtained by the enzyme reaction of recombinant FeCGTa after acid treatment (1/1.37, n=10) (Figure 3). This difference cannot be explained by an involvement of spontaneous dehydration. The dehydration process in buckwheat is likely to proceed by enzymatic catalysis that is able to regulate the regioisomer's ratio as proposed in rice (Brazier-Hicks *et al*., 2009).

The reaction mechanism of bacterial CGT (UrdGT2) has been precisely characterized, along with the enzyme's crystal structure (Faust *et al.*, 2000; *Baig et al.*, 2006; Mittler *et al.*, 2007). In addition to CGT activity, the enzyme showed OGT activity on some artificial substrates (Dürr *et al.*, 2004). Moreover, an OGT (LanGT2) whose structure and substrate are similar to UrdGT2, was successfully converted to

CGT by *in vitro* mutagenesis, based on the structural information from UrdGT2 (Härle *et al.*, 2011). Knowledge of *C*-glycosylation in other organisms, including plants, however, is very limited. Recently, Gutmann and Nidetzky (2012) reported that a mutual exchange of the two amino acid residues interacting with the catalytic His residue of OsCGT (Asp-Ile motif) and pear phloretin OGT (Ile-Asp motif) resulted in a switch between CGT and OGT activity of these enzymes; although, the mutant enzyme showed only slight activity (less than 0.1%) compared to that of the wild type. They proposed that this Ile-Asp motif would be important for the CGT activity, but FeCGTs do not conserve this Ile-Asp motif. The active site essential for CGT activity in plants remains unclear, and further analysis is required for the elucidation of the molecular basis of the CGT reaction mechanism.

We constructed a molecular phylogenetic tree based on the deduced amino acid sequences of FeCGTs with those of other plant UGTs and bacterial CGTs (listed in Table S2) characterized so far (Figure 5). FeCGTs (UGT708C1 and UGT708C2), OsCGT, and UGT708A6 are grouped into one clade (UGT708), in which their amino acid sequences show about 40% identity to one another, and no reported OGT was

included in this clade. This result suggested that flavonoid *C*-glucosyltransferases in monocot and dicot have evolved from the same ancestral gene. This hypothesis is consistent with the prediction that flavonoid CGTs would form an independent orthologous group (OG15) in UGT, which was determined through an evolutionary analysis of plant glycosyltransferases (Yonekura-Sakakibara and Hanada, 2011).

In buckwheat, it was shown that large amounts of flavone *C*-glucosides accumulated in the cotyledons but not in seeds. In accordance with this, the transcripts of *FeCGT* increased in the cotyledon during germination (Figure 4b). The accumulated *C*-glucosides were retained in mature cotyledons. As *C*-glucosylflavones have general antifungal and antibacterial activity (Jay *et al.*, 2005), they could play a role in the defense mechanism of buckwheat during the early growth stage. We currently have no explanation as to why *C*-glucosides accumulate only in the cotyledons but not in other parts. To answer this question, we are now attempting to suppress *FeCGT* expression by RNA interference, which would clarify the physiological role of flavonoid *C*-glucosides in buckwheat.

Experimental procedures

Plant materials

A diploid cultivar of buckwheat (*Fagopyrum esculentum* cv. Shinano No. 1) selected from Japanese indigenous variety was used in this study. For enzyme purification, seeds were sown on a wet paper towel and germinated in the dark at 22°C–28°C. Four days after sowing, etiolated seedlings were collected, immediately frozen with liquid nitrogen, and stored at −80°C prior to use. For RNA extraction, cotyledons, hypocotyls, and roots were collected separately and stored at −80°C. For the preparation of other plant materials, seeds were sown on culture soil and were cultured at 22°C under 14-h light/10-h dark conditions.

Reagents

The following substrates were obtained: quercetin (Nacalai Tesque, Kyoto, Japan), kaempferol, 2′,4′-dihydroxyacetophenone and 2′,4′,6′-trihydroxyacetophenone (Tokyo Chemical Industries, Tokyo, Japan), phloretin, naringenin, quercetin 3-*O*-glucoside and 2-phenyl-2′,4′,6′-trihydroxyacetophenone (Sigma-Aldrich, St-Louis, MO, USA),

apigenin and luteolin (Indofine Chemicals, Hillsborough, NJ, USA), chrysin (Across Organics, Geel, Belgium), vitexin and isovitexin (Extrasynthèse, Genay, France), 2,4,6-trihydroxybenzaldehyde (Alfa Aesar, Heysham, England), UDP-glucose and UDP-glucuronic acid (Nacalai Tesque), and UDP-galactose and UDP-xylose (Sigma-Aldrich). Unless otherwise specified, all other chemicals were obtained from Sigma-Aldrich, Nacalai Tesque, and Wako Pure Chemical Industries (Osaka, Japan).

2-Hydroxypinocembrin, 2-hydroxynaringenin, and 2-hydroxyeriodictyol were synthesized from chrysin, apigenin, and luteolin, respectively, as previously described (Kerscher and Franz, 1987) with some modification. For the synthesis of 2-hydroxynaringenin, apigenin（100 mg）dissolved in pyridine (10 ml) was added with KOH (3 g) ground to a powder with mortar and pestle in the presence of diethylether, and then reacted at 120 °C for 2 h with reflux. After cooling, the solvent (pyridine) was removed and the residual brown pellet with KOH powder was dissolved in water. The solution was neutralized with addition of acetic acid until white precipitate was appeared, and then extracted with ethyl acetate. After the solvent (ethyl acetate) removal by evaporation, the obtained crude product was loaded onto an octadesylsilyl

(ODS) column (Wako-gel 50C18, 15 mm i.d. × 100 mm, Wako Pure Chemical Industry) equilibrated with 25% methanol, and eluted with a linear gradient of 25–50% methanol (300 ml). The fraction containing the product was concentrated, and dried by freeze dryer.

HPLC conditions

HPLC was performed using an ODS column (4.6 mm i.d. × 150 mm: LUNA 5u C18(2), Phenomenex, Torrance, CA) with an LC10Avp system (Shimadzu, Kyoto, Japan) as previously described with some modifications (Taguchi *et al.*, 2005). The column was eluted using a linear gradient of 20% to 60% solvent B (methanol containing 0.1% formic acid) in solvent A (0.1% formic acid) for 12 min, followed by 60% solvent B in A for 8 min at a flow rate of 1 ml min^{-1} at 40°C. The eluate was monitored at 350 nm using a diode array detector (SPD-M10Avp, Shimadzu).

UPLC-MS was performed using a Waters UPLC ACQUITY SQD system (Waters, Milford, MA, USA) with an electron-spray ionization probe. Samples were eluted from an ODS column (2.1 mm i.d. × 50 mm: ACQUITY UPLC BEH C18 Column, Waters)

at a flow rate of 0.25 ml min^{-1} at 40°C as follows: 20% solvent B in solvent A for 0.5 min, followed by a 20% to 60% solvent B in A for 4 min, and then 60% solvent B in A for 1 min. For the separation of 2',4',6'-trihydroxyacetophenone and 2,4,6-trihydroxybenzaldehyde, the column was eluted using 5% solvent B in A for 0.5 min, followed by a 5% to 45% solvent B in A for 4 min, and then 45% solvent B in A for 1 min.

Extraction and purification of C-glucosyltransferase enzymes

A frozen seedling (2,000 g fresh weight) was added to 4 litre of ice-cold buffer A [100 mM Tris-HCl, pH 8.0, 1 mM dithiothreitol (DTT)] containing 5% (w/v) polyvinylpolypyrrolidone and 0.1 mM phenylmethylsulfonyl fluoride. The mixture was blended in a food processor, followed by sonication for 24 × 5 sec at an amplitude of 25% (Vibra Cell VCX400, Sonic & Materials, Inc., Newtown, CT, USA). Enzyme purification was carried out at 4°C unless otherwise stated. The homogenate was filtered through a disposable, non-woven bag to remove cellular debris and the filtrate was centrifuged at 13,000 × *g* for 20 min. The supernatant was used as crude extract,

the proteins from which were fractionated by ammonium sulfate precipitation (35%–65% saturation). The resulting pellet was dissolved in buffer A and centrifuged at 93,000 × *g* for 30 min to remove insoluble lipids and the membrane fraction. The supernatant containing CGT activity was concentrated by ammonium sulfate precipitation (65% saturation) and loaded onto a Phenyl-Sepharose CL-4B column (2 cm i.d. × 10 cm; GE Healthcare, Tokyo, Japan) equilibrated with buffer A containing 1M ammonium sulfate. The column was washed with 100 mL of the same buffer and proteins were eluted with a linear gradient of ammonium sulfate (300 ml, 1 to 0 M in buffer A). Fractions with CGT activity were collected and concentrated by 65% saturation of ammonium sulfate precipitation and loaded onto a Sephadex G-100 column (2.0 cm i.d. × 50 cm; GE Healthcare) equilibrated with buffer A. Proteins were eluted with 250 mL of buffer A at a flow rate of 20 ml h^{-1}. The resulting fractions with CGT activity were loaded onto a DEAE-Sepharose CL-6B column (1.5 cm i.d. × 20 cm; GE Healthcare) equilibrated with buffer A. After washing with 50 ml of buffer A, proteins were eluted with a linear gradient of NaCl (200 ml, 20 mM to 200 mM NaCl in buffer A). The fractions with CGT activity were collected and concentrated and solvent

exchanged with buffer B (5 mM sodium phosphate, pH 7.5, containing 1 mM DTT) using an Amicon Ultra-15 Ultracel-10k (Merck Millipore, Billerica, MA, USA). The resulting fraction was loaded onto Cibacron Blue-3GA column (1.5 cm i.d. × 3.5 cm; Sigma) equilibrated with buffer A. The column was washed with 10 ml of buffer A, and then eluted stepwise with 6 ml each of buffer A supplemented with 0.25 mM UDP-glucose and 10, 20, 50, 100, 200, 500, 1000 or 2000 mM NaCl. The fractions with CGT activity (500 to 2000 mM NaCl) were pooled, concentrated, and desalted by ultracentrifugation using Amicon Ultra-15. After this step, UDP-glucose was added to the buffer to prevent enzyme inactivation.

The CGT was further purified by HPLC (Multi-station LC-8020 Model II, Tosoh, Tokyo, Japan) using a Superdex G-200 HR 10/30 (GE Healthcare). The sample was loaded onto the column equilibrated with buffer A supplemented with 50 μM UDP-glucose. The column was then eluted with buffer A at a flow rate of 0.25 ml min^{-1}. The active fraction was concentrated and purified by a Mono Q™ 5/50 GL column (GE Healthcare). The sample was loaded onto the column equilibrated with buffer C (buffer A supplemented with 50 μM UDP-glucose and 50 mM NaCl). The column was then

eluted with buffer C for 5 min., followed by a linear gradient of 0%–100% buffer D (buffer A supplemented with 50 μM UDP-glucose and 150 mM NaCl) in C for 35 min at a flow rate of 1 ml min^{-1}. The active fraction was concentrated using Vivaspin 500-10K (Sartorius Stedim Biotech GmbH, Goettingen, Germany). Protein concentration was quantitated by Protein Assay (Bio-Rad, Hercules, CA, USA) using BSA as the standard.

CGT activity assay

CGT was assayed in a reaction mixture (50 μl) composed of 100 μM 2-hydroxypinocembrin as a glucose acceptor, 1 mM UDP-glucose, and an appropriate amount of enzyme (0.2–10 μl of the fraction) in buffer A. The reaction was initiated by adding 2-hydroxypinocembrin and incubated at 30°C for 10 min–30 min. The reaction was stopped by adding 10 μl of 1M HCl, and the reaction mixture was subjected to HPLC after addition of internal standard.

Sequencing analysis of peptides

The purified CGT protein bands from the SDS-polyacrylamide (15%) gel were blotted onto a PVDF membrane (Immobilon PSQ, Merck Millipore). The membrane with the protein band was soaked in a peptidase buffer (50 mM Tris HCl, pH 9.0, 10% acetonitrile), and treated with a lysyl endopeptidase (final conc. 2 ng μl^{-1}, Wako Pure Chemical) for 16 h at 37°C. Peptide fragments were desalted and concentrated using a Zip Tip C18 Pipette Tip (Merck Millipore) and separated by a 0%–40% linear acetonitrile gradient for 60 min with a Waters UPLC Xevo Qtof system (Waters). The data obtained were processed using ProteinLynx Global Server 2.5 (Waters) and MASCOT Wizard (Matrix Science, http://www.matrixscience.com/), and searched against the Genbank database (https://www.ncbi.nlm.nih.gov/) using the BLASTP program.

cDNA library construction

Total RNA was extracted from etiolated buckwheat cotyledons by phenol-SDS methods in accordance with those of Shirzadegan *et al.* (1991) with modifications. In short, to remove viscous polysaccharides from the sample, phenol-chloroform extraction and

chloroform-isoamylalcohol treatment were repeated at least 5 times. mRNAs were then purified using Poly(A) Purist™-MAG (Life Technologies, Grand Island, NY, USA). The cDNA library was constructed in *E. coli* using the CloneMiner™ II cDNA Library Construction Kit (Life Technologies).

PCR primers

PCR primers used in this work are listed in Table S4.

Cloning and sequencing of CGT genes

Degenerate primers (Primer-Fw1, Primer-Fw2, and Primer-Rv) were designed on the basis of the amino acid sequences of the peptide fragments from the purified CGTs, which showed similarity to those from other UGTs (Figure S4). The DNA fragments corresponding to partial CGT genes were amplified from the cDNA library with Primer-Fw1 and vector (pDONR222) sequence (M13-Rv) using *Ex Taq* (Takara Bio, Kusatsu, Japan) at the following conditions: 94°C for 5 min, 35 cycles of 94°C for 30 sec, 50°C for 30 sec, and 72°C for 1 min, and followed by 72°C for 5 min. A 500-fold

dilution of the amplified fragment was used as the template for the second nested-PCR with Primer-Fw2 and Primer-Rv under the same conditions. The resulting fragments were cloned into a dT-protruding pBlueScript SK-vector. Full-length 3′-cDNAs of CGTs were amplified by nested-PCR in the same manner, using CGT-F1 and M13-Rv, followed by CGT-F2 and T7-Pro. The 5′- end of the CGT genes were amplified by an RNA ligase-mediated rapid amplification of the 5′-cDNA end (GeneRacer™ Kit, Life Technologies) with CGT-R2 and GeneRacer 5′-Primer (Life Technologies). We utilized iProof™ High-Fidelity DNA polymerase (Bio-Rad) and the PCR was run using the following conditions: 98°C for 30 sec, 40 cycles of 98°C for 10 sec, 55°C for 20 sec, and 72°C for 30 sec, followed by 72°C for 8 min. The fragments containing the full-length cDNA were amplified from the cDNA library with CGT-Fw-Nde and CGT-Rv-Eco using iProof DNA polymerase. The fragments were cloned into pCR4Blunt-TOPO (Life Technologies) and sequenced using a DNA sequencer (Genetic Analyzer 3100, Life Technologies).

Heterologous expression of FeCGT in E. coli

The full-length cDNAs of FeCGTs were subcloned into the pET28a(+) vector (Merck, Darmstadt, Germany). The recombinant proteins were expressed in *E. coli* Rosetta™ 2(DE3) (Merck) and extracted in accordance with the manufacturer's instruction. The recombinant proteins were purified using a nickel-affinity column (His-GraviTrap, GE Healthcare). The active fractions were concentrated using Amicon Ultra-15.

Characterization of the recombinant FeCGT

We performed the enzyme assay with the recombinant FeCGTs in a reaction mixture (50 μl) composed of 200 μM 2-hydoroxypinocembrin, 1 mM UDP-glucose, and 5–500 ng of the purified enzymes in to the reaction buffer [50 mM potassium phosphate, pH 6.5, containing 0.01% BSA and 5 mM 2-mercaptoethanol]. The reaction was initiated by adding 2-hydoroxypinocembrin and incubated at 30°C for 5–10 min. The reaction was stopped by adding 10 μl of 1M HCl. To evaluate the effect of pH on enzyme activity, 100 mM potassium phosphate buffer (pH 5.0 to 8.5) or 100 mM triethanolamine-HCl buffer (pH 8.5 to 11.0) were used. In order to determine the enzymes' thermal stabilities, the enzymes were prepared in reaction buffer

supplemented with UDP-glucose and incubated at 20°C to 70°C for 30 min, after which the temperature was shifted to 30°C for the reaction. To determine their optimal reaction temperatures, the reactions were performed at 20°C to 70°C. Substrate preferences were confirmed using 200 μM phenolic compounds listed in Table 1 and 1 mM UDP-glucose as substrates. UDP-sugar preferences were confirmed using 200 μM 2-hydroxypinocembrin and multiple UDP-sugars (400 μM) as substrates.

Kinetic parameters of the enzyme reaction were obtained by fitting the kinetics data to the Michaelis–Menten equation by using the Hyper 32 program (http://homepage.ntlworld.com/john.easterby/software.html). Reactions performed appear to be single-substrate saturable enzyme reactions in the presence of the second substrate in amounts essentially saturating the enzyme (2000 μM UDP-glucose or 200 μM 2-hydroxypinocembrin). To determine the K_m values of the substrates, we utilized a range of concentrations of 2-hydroxynaringenin, 2-hydroxypinocembrin, and 2,4,6-trihydroxyacetophenone (0.5 to 50 μM, 2.5 to 200 μM and 2.5 to 200 μM, respectively) at a fixed UDP-glucose concentration (2,000 μM). To determine the K_m values of UDP-glucose, reactions were performed with concentrations from 5 μM to

2,000 μM at a fixed 2-hydroxypinocembrin concentration of 200 μM. After incubation at 30°C for 5 min, the reaction was stopped by adding 10 μl of 1 M HCl and the reaction mixture was subjected to UPLC-MS after the addition of an internal standard (umbelliferone).

RT-PCR analysis

Total RNAs from buckwheat flowers, leaves, etiolated cotyledons, hypocotyls, stems, roots, germinating seeds, and developing cotyledons were extracted by the phenol-SDS method described above. First-strand cDNAs were synthesized from these total RNAs (0.5 μg each) using ReverTra Ace (TOYOBO, Tokyo, Japan) with a dT-T3 primer in a 20 μl reaction at 42°C for 60 min, followed by a 10-fold dilution with distilled water. Subsequently, 2 μl of the sample was used for PCR in a 20 μl reaction using *Ex Taq* under the following conditions: 95°C for 3 min, 28 cycles of 95°C for 30 sec, 52°C (55°C for *gapdh*) for 30 sec, and 72°C for 30 sec, followed by a final 5 min extension at 72°C. CGT-F2 and CGT-R3 primers were used to amplify a 509-bp internal fragment of *FeCGTs*. FeGAPDH-F and FeGAPDH-R were used as primers for

the amplification of the 584-bp internal fragment of *Fe-gapdh* (Genbank accession No. AB919116), which served as an endogenous control. The amplified fragments (5 μl) were separated on an 1.5% agarose gel, and stained with ethidium bromide.

Acknowledgements

We would like to thank Dr. Takaomi Nomura (Shinshu University) for his support on protein purification, Dr. Yoshinori Nishii (Shinshu University) for his useful discussion on the synthesis of substrates, Ms. Eri Ishikawa (Shinshu University) for sequencing analysis of peptide fragments, and the UGT nomenclature committee for suggesting the UGT names, Enago (www.enago.jp) for the English language review. We are indebted to the Division of Gene Research and Division of Instrumental analysis, Research Center for Human and Environmental Sciences, Shinshu University, for providing facilities. This work was supported in part by JSPS KAKENHI Grant Number 26450120 to G.T.

Short legends for supporting information

Figure S1. Distribution of phenolic compounds in buckwheat plants.

Figure S2. Purification of CGT from buckwheat seedlings by Mono Q anion exchange chromatography.

Figure S3. Properties of purified CGTa and CGTb.

Figure S4. Analysis of peptide sequences of the purified CGTs.

Figure S5. UPLC-MS analysis of the recombinant FeCGTa reaction against several phenolic substrates.

Figure S6. Properties of recombinant FeCGTa and FeCGTb.

Figure S7. Pictures of developmental stages during germination and cotyledon development of buckwheat seed used for RT-PCR in Figure 4.

Figure S8. Comparison of the amino acid sequences of FeCGTa, FeCGTb and two variants obtained by cDNA screening.

Table S1. Glucosylation activities in cell-free extracts of buckwheat organs.

Table S2. Buckwheat UGT genes isolated by PCR-based cloning.

Table S3. Purification of *C*-glucosyltransferases from buckwheat seedlings.

Table S4. PCR Primers used in this work.

Table S5. List of UGTs used for the phylogenetic analysis in Figure 5.

References

Baig, I., Kharel, M., Kobylyanskyy, A., Zhu, L., Rebets, Y., Ostash, B., Luzhetskyy, A., Bechthold, A., Fedorenko, V.A. and Rohr, J. (2006) On the acceptor substrate of C-glycosyltransferase UrdGT2: three prejadomycin C-Glycosides from an engineered mutant of *Streptomyces globisporus* 1912 DeltalndE (*urdGT2*). *Angew Chem Int Ed Engl*, **45**, 7842-7846.

Bowles, D., Isayenkova, J., Lim, E.K. and Poppenberger, B. (2005) Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant Biol.*, **8**, 254-263.

Brazier-Hicks, M. and Edwards, R. (2013) Metabolic engineering of the flavone-*C*-glycoside pathway using polyprotein technology. *Metab. Eng.*, **16**, 11-20.

Brazier-Hicks, M., Evans, K.M., Gershater, M.C., Puschmann, H., Steel, P.G. and Edwards, R. (2009) The *C*-glycosylation of flavonoids in cereals. *J. Biol. Chem.*, **284**, 17926-17934.

Bungaruang, L, Gutmann, A. and Nidetzky, B. (2013) Leloir glycosyltransferases and natural product glycosylation: biocatalytic synthesis of the *C*-glucoside nothofagin, a major antioxidant of redbush herbal tea. *Adv. Synth. Catal.*, **355**, 2757-2763.

Byrne, P.F., Darrah, L.L., Snook, M.E., Wiseman, B.R., Widstrom, N.W., Moellenbeck, D.J. and Barry, B.D. (1996) Maize silk-browning, maysin content, and antibiosis to the maize earworm, *Helicoverpa zea* (Boddie). *Maydica*, **41**, 13-18.

Du, Y., Chu, H., Chu, I.K. and Lo, C. (2010) CYP93G2 is a flavanone 2-hydroxylase required for *C*-glycosylflavone biosynthesis in rice. *Plant Physiol.*, **154**, 324-333.

Dürr, C., Hoffmeister, D., Wohlert, S.E., Ichinose, K., Weber, M., Von Mulert, U., Thorson, J.S. and Bechthold, A. (2004) The glycosyltransferase UrdGT2 catalyzes both *C*- and *O*-glycosidic sugar transfers. *Angew Chem. Int. Ed. Engl.*, **43**, 2962-2965.

Eisner, T., Nowicki, S., Goetz, M. and Meinwald, J. (1980) Red cochineal dye (carminic Acid): its role in nature. *Science*, **208**, 1039-1042.

Falcone Ferreyra, M.L., Rodriguez, E., Casas, M.I., Labadie, G., Grotewold, E. and Casati, P. (2013) Identification of a bifunctional maize *C*- and *O*-glucosyltransferase. *J Biol Chem*., **288**, 31678-31688.

FAOSTAT (Statistical analysis website of Food and Agriculture Organization of the United Nations) http://faostat3.fao.org/home/index.html.

Faust, B., Hoffmeister, D., Weitnauer, G., Westrich, L., Haag, S., Schneider, P., Decker, H., Künzel, E., Rohr, J. and Bechthold, A. (2000) Two new tailoring enzymes, a glycosyltransferase and an oxygenase, involved in biosynthesis of the angucycline antibiotic urdamycin A in *Streptomyces fradiae* Tü2717. *Microbiology*, **146 (Pt 1)**, 147-154.

Fischbach, M.A., Lin, H., Liu, D.R. and Walsh, C.T. (2005) *In vitro* characterization of *IroB*, a pathogen-associated *C*-glycosyltransferase. *Proc. Natl. Acad. Sci, U S A*, **102**, 571-576.

Franz, G. and Grün, M. (1983) Chemistry, occurrence and biosynthesis of *C*-glycosyl compounds in plants. *Planta Med.*, **47**, 131-140.

Gachon, C.M., Langlois-Meurinne, M. and Saindrenan, P. (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci.* **10**, 542-549.

Gandia-Herrero, F., Lorenz, A., Larson, T., Graham, I.A., Bowles, D.J., Rylott, E.L. and Bruce, N.C. (2008) Detoxification of the explosive 2,4,6-trinitrotoluene in Arabidopsis: discovery of bifunctional *O*- and *C*-glucosyltransferases. *Plant J.*, **56**, 963-974.

Gutmann, A. and Nidetzky, B. (2012) Switching between *O*- and *C*-glycosyltransferase through exchange of active-site motifs. *Angew Chem. Int. Ed. Engl.*, **51**, 12879-12883.

Hamilton, M., Kuate, S., Brazier-Hicks, M., Caulfield, J., Rose, R., Edwards, R., Torto, B., Pickett, J. and Hooper, A. (2012) Elucidation of the biosynthesis of the di-*C*-glycosylflavone isoschaftoside, an allelopathic component from *Desmodium* spp. that inhibits *Striga* spp. development. *Phytochemistry*, **84**, 169-176.

He, X., Blount, J.W., Ge, S., Tang, Y. and Dixon, R.A. (2011) A genomic approach to isoflavone biosynthesis in kudzu (*Pueraria lobata*). *Planta*, **233**, 843-855.

Hultin, P.G. (2005) Bioactive *C*-glycosides from bacterial secondary metabolism. *Curr. Top. Med. Chem.*, **5**, 1299-1331.

Härle, J., Günther, S., Lauinger, B., Weber, M., Kammerer, B., Zechel, D.L., Luzhetskyy, A. and Bechthold, A. (2011) Rational design of an aryl-*C*-glycoside catalyst from a natural product *O*-glycosyltransferase. *Chem. Biol.*, **18**, 520-530.

Jay, M. (1994) *C*-Glycosylflavonoids. In *The Flavonoids: advances in research since 1986* (Harborne, J.B. eds). London, UK: Chapman & Hall, pp. 57-93.

Jay, M., Viricel, M.R. and Gonnet, J.F. (2005) *C*-Glycosylflavonoids. In *Flavonoids*: *Chemistry, Biochemistry and Applications* (Andersen, O.M. and Markham, K.R. eds). Boca Raton, FL, USA: CRC Press, pp. 857-915.

Jones, P. and Vogt, T. (2001) Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta*, **213**, 164-174.

Kerscher, F. and Franz, G. (1987) Biosynthesis of vitexin and isovitexin: enzymic synthesis of the *C*-glucosylflavones vitexin and isovitexin with an enzyme preparation from *Fagopyrum esculentum* M. seedlings. *Z. Naturforsch.*, **42c**, 519-524.

Kerscher, F. and Franz, G. (1988) Isolation and some properties of an UDP-glucose: 2-hydroxyflavanone-6 (or 8)-*C*-glucosyltransferase from *Fagopyrum esculentum* M. cotyledon. *J. Plant Physiol.*, **132**, 110-115.

Khan, Z.R., Pickett, J.A., Hassanali, A., Hooper, A.M. and Midega, C.A.O. (2008) *Desmodium* species and associated biochemical traits for controlling *Striga* species: present and future prospects. *Weed Res.*, **48**, 302-306.

Kim, J.H., Lee, B.C., Kim, J.H., Sim, G.S., Lee, D.H., Lee, K.E., Yun, Y.P. and Pyo, H.B. (2005) The isolation and antioxidative effects of vitexin from *Acer palmatum*. *Arch. Pharm. Res.*, **28**, 195-202.

Kim, J., Lee, I., Seo, J., Jung, M., Kim, Y., Yim, N. and Bae, K. (2010) Vitexin, orientin and other flavonoids from *Spirodela polyrhiza* inhibit adipogenesis in 3T3-L1 cells. *Phytother. Res.*, **24**, 1543-1548.

Li, Q.M., van den Heuvel, H., Delorenzo, O., Corthout, J., Pieters, L.A., Vlietinck, A.J. and Claeys, M. (1991) Mass spectral characterization of *C*-glycosidic flavonoids isolated from a medicinal plant (*Passiflora incarnata*). *J. Chromatogr.*, **562**, 435-446.

Li, Y., Baldauf, S., Lim, E.K., Bowles, D.J. (2001) Phylogenic analysis of the UDP-glycosyltransferase multigene family of *Arabidpsis thaliana. J. Biol. Chem.*, **276**, 4338-4343.

Margna, U., Margna, E., Puluteder, A. (1990) Localization and distribution of flavonoids in buckwheat seedling cotyledons. *J. Plant Physiol.*, **136**, 166-171.

McNally, D.J., Wurms, K.V., Labbé, C., Quideau, S. and Bélanger, R.R. (2003) Complex *C*-glycosyl flavonoid phytoalexins from *Cucumis sativus*. *J. Nat. Prod.*, **66**, 1280-1283.

Mittler, M., Bechthold, A. and Schulz, G.E. (2007) Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin. *J. Mol. Biol.*, **372**, 67-76.

Paquette, S., Møller, L.B., Bak, S. (2003) On the origin of family 1 plant glycosyltransferases. Phytochemistry, **62**, 399-413.

Prabhakar, M.C., Bano, H., Kumar, I., Shamsi, M.A. and Khan, M.S. (1981) Pharmacological investigations on vitexin. *Planta Med.*, **43**, 396-403.

Shirzadegan, M., Christie, P., Seemann, J.R. (1991) An efficient methods for isolation of RNA from tissue cultured plant cells, *Nucleic Acids Res.*, **19**, 6055.

Taguchi, G., Nakamura, M., Hayashida, N. and Okazaki, M. (2003) Exogenously added naphthols induce three glucosyltransferases, and are accumulated as glucosides in tobacco cells. *Plant Sci*., **164**, 231-240.

Taguchi, G., Shitchi, Y., Shirasawa, S., Yamamoto, H., Hayashida, N. (2005) Molecular cloning, characterization, and downregulation of an acyltransferase that catalyzes the malonylation of flavonoid and naphthol glucosides in tobacco cells. *Plant J.*, **42**, 481-491.

Watanabe, M. (2007) An anthocyanin compound in buckwheat sprouts and its contribution to antioxidant capacity. *Biosci. Biotechnol. Biochem.*, **71**, 579-582.

Yonekura-Sakakibara, K. and Hanada, K. (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.*, **66**, 182-193.

Zhang, Y., Bao, B., Lu, B., Ren, Y. and Tie, X. (2005) Determination of flavone *C*-glucosides in antioxidant of bamboo leaves (AOB) fortified foods by reversed-phase high-performance liquid chromatography with ultraviolet diode array detection. *J. Chromatogr. A*, **1065**, 177-185.

Table 1. Substrate specificities of purified and recombinant CGTs from buckwheat.

Substrates	Purified enzyme		Recombinant enzyme	
	CGTa	CGTb	FeCGTa	FeCGTb
			nkat mg_protein^{-1}	
Sugar acceptors [a]				
2-Hydroxyflavanones				
2-hydroxyeriodictyol	1.8 ± 0.4	2.1 ± 0.4	28.1±1.7	23.9±1.0
2-hydroxynaringenin	5.8 ± 1.4	4.5 ± 0.8	35.7±1.9	34.5±1.3
2-hydroxypinocembrin	41.6 ± 12.1	83.2 ± 9.6	34.2±2.7	55.1±2.5
Flavones, flavonol and flavanones				
Naringenin	ND [b]	ND	ND	ND
Quercetin	ND	ND	ND	ND
Luteorin	ND	ND	ND	ND
Apigenin	ND	ND	ND	ND
Chrysin	NT [c]	NT	ND	ND
Dihydrochalcone				
Phloretin	4.3 ± 0.4	4.7 ± 1.8	5.2±0.3	8.4±0.2
Others				
2-phenyl-2',4',6'-trihydroxyactophenone	6.9 ± 0.4	31.1 ± 4.0	12.7±1.1	45.0±0.4
2',4',6'-trihydroxyacetophenone	NT	NT	10.1±0.8	12.5±0.7
2,4,6-trihydroxybenzaldehyde	NT	NT	2.2±0.4	2.5±0.1
2',4'-dihydroxyactophenone	NT	NT	ND	ND
Sugar donors [d]				
UDP-glucose	41.6 ± 12.1	83.2 ± 9.6	35.5±1.2	25.6±0.1
UDP-xylose	NT	NT	1.8±0.2	0.7±0.0
UDP-galactose	NT	NT	ND	ND
UDP-glucuronic acid	NT	NT	ND	ND

[a]UDP-glucose was used as a sugar donor. [b]ND: not detected. [c]NT: not tested. [d]2-hydroxypinocembrin was used as a sugar acceptor. Data was shown as average ± SD (n = 3).

Table 2. Kinetic parameters of recombinant FeCGTa and FeCGTb.

	FeCGTa			FeCGTb		
Substrates	K_m (μM)	k_{cat} (sec^{-1})	k_{cat} / K_m (M^{-1} sec^{-1})	K_m (μM)	k_{cat} (sec^{-1})	k_{cat} / K_m (M^{-1} sec^{-1})
Sugar acceptors[a]						
2-hydroxynaringenin	4.4±0.9	3.2	7.2×10^5	3.7±1.8	3.1	8.3×10^5
2-hydroxypinocembrin	40.0±3.3	10.2	2.5×10^5	6.5±1.2	2.3	3.6×10^5
2-phenyl-2',4',6'-trihydroxyactophenone	36.5±4.4	1.9	5.3×10^4	38.5±4.8	4.0	1.0×10^5
Sugar donor[b]						
UDP-glucose	58.1±5.1	4.5	7.8×10^4	36.0±3.7	2.3	6.4×10^4

[a]UDP-glucose was used as a sugar donor. [b]2-hydroxypinocembrin was used as a sugar acceptor.

Figure legends

Figure 1. Structures of phenolic compounds used in this study.

(a) A proposed pathway of *C*-glycosylation in flavonoid biosynthesis (Kerscher and Franz, 1988; Brazier-Hicks *et al*., 2009). (b) Structures of phenolic compounds: [1] 2-hydroxypinocembrin; [2] 2-hydroxynaringenin; [3] 2-hydroxyeriodictyol; [4] naringenin; [5] chrysin; [6] apigenin; [7] luteorin; [8] quercetin; [9] phloretin; [10] 2-phenyl-2′,4′,6′-trihydroxyacetophenone; [11] 2′,4′,6′-trihydroxyacetophenone; [12] 2,4,6-trihydroxybenzaldehyde; [13] 2′,4′-dihydroxyacetophenone.

Figure 2. SDS-PAGE analysis of proteins from buckwheat purified *C*-glucosyltransferase.

Samples from all steps of the CGT purification process were separated on a 12.5% SDS-polyacrylamide gel. Lanes 1–5 contain 10 μg each of crude extract, ammonium sulfate precipitation, ultracentrifugation, Phenyl Sepharose CL-4B, and Sephadex G-100, respectively; Lane 6 contains 5 μg of DEAE-Sepharose CL-6B; and lanes 7–10 contain 3 μg each of Cibacron Blue3GA, Superdex G-200 HR 10/30 (lane 8), Mono Q fraction A (lane 9), and B (lane 10), respectively. Following separation, the gel was

stained with coomassie brilliant blue R-250. Arrowheads indicate the purified CGTs. Standard proteins were loaded on lane M.

Figure 3. HPLC-MS analysis of the recombinant FeCGTa reaction products of 2-hydroxynaringenin.

The reaction was performed as described in Experimental Procedures. Each panel shows a chromatogram from the following reaction conditions: (a) without enzyme; (b) stopped with methanol instead of HCl; (c) 2 M HCl-treated; authentic compounds of (d) vitexin, and (e) isovitexin. The eluates were monitored at 290 nm using a diode array detector. The ES-MS^- corresponding to the indicated peak is shown inside each figure. The retention time of MS peaks were delayed by 0.08 min compared to that of the diode array. Peak identification: S1, 2-hydroxynaringenin (substrate); S2, apigenin (dehydrated substrate); P1, 2-hydroxynaringenin *C*-glucoside; P2 and V, vitexin; P3 and Iv, isovitexin.

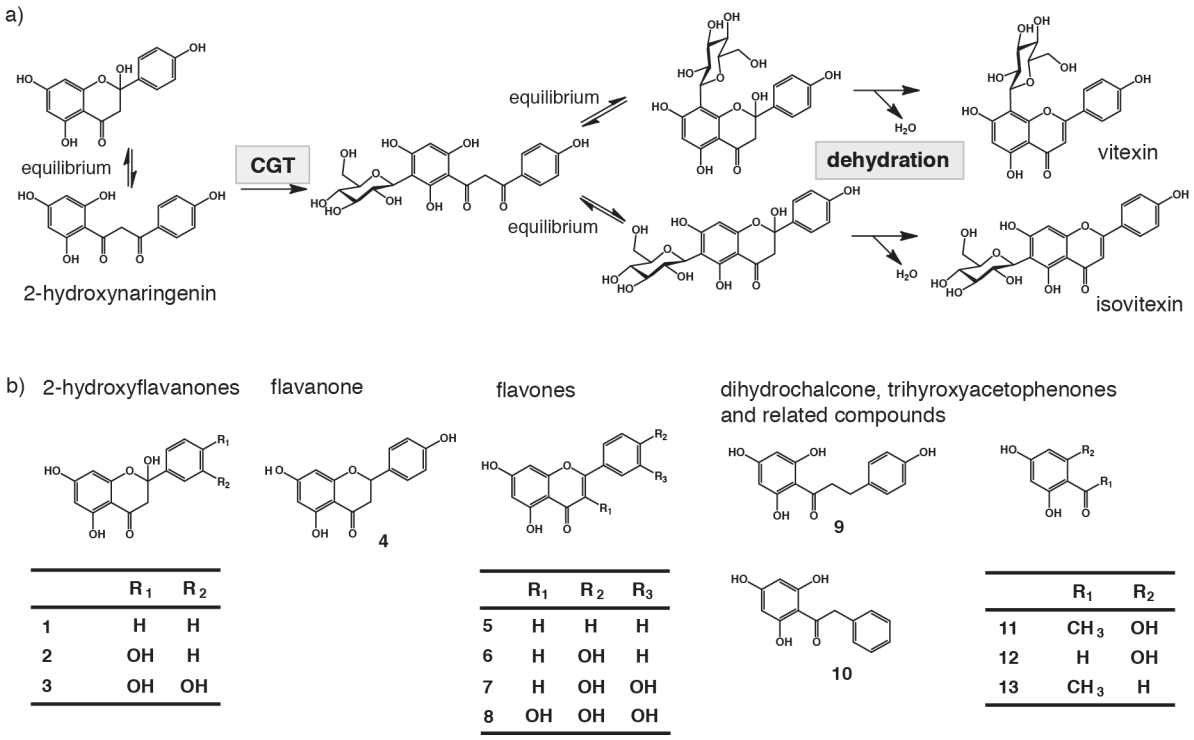
Figure 4. RT-PCR analysis of the *FeCGT* transcripts.

Total RNAs were extracted from various organs of buckwheat plants, followed by

RT-PCR as described in Experimental Procedures. A 509-bp cDNA fragment was expected for *FeCGT* mRNA, and 584-bp fragment for glyceraldehyde-3-phosphate dehydrogenese (*gapdh*) mRNA. (a) Agarose gel electrophoresis demonstrating the organ-specific expression of *FeCGTs* in flower (lane 1), leaf (lane 2), stem (lane 3), root (lane 4), hypocotyl (lane 5), and cotyledon (lane 6). (b) Changes in the amounts of *FeCGT* transcripts in the cotyledon during seed germination and cotyledon development. Total RNAs were extracted from seeds or cotyledons after cutting out hypocotyl and root. As the amount of *gapdh* transcript also changed during germination, total RNA (400 ng) of each sample was also shown as reference. Agarose gel electrophoresis demonstrating the expression of *FeCGTs* in seeds before sowing (lane 1); 2.5, 6, and 10 h after sowing (lane 2–4); emergence of radicle (1 day after sowing; lane 5); elongation of root and root hair formation (1.5 and 2 days after sowing; lanes 6, 7); splitting of seed coat and formation of the cotyledon (3 and 5 days after sowing; lanes 8, 9); development of the cotyledon and removal of seed coat (7 days after sowing; lanes 9, 10); mature cotyledon (3 weeks and 2 months on soil; lane 11, 12). Pictures of each germination stage are shown in Figure S7.

Figure 5. Molecular phylogenetic tree constructed based on the deduced amino acid sequences of FeCGTs and related glycosyltransferases.

The tree was constructed by the neighbor-joining method using ClustalW (http://clustalw.ddbj.nig.ac.jp/). The lengths of lines indicate the relative distance between nodes. The grouping of plant UGTs is shown in accordance with a proposal by Li *et al*. (2001) and also, for orthologous groups by Yonekura-Sakakibara and Hanada (2011). Abbreviations and Genbank accession nos. of UGTs are as follows: FeCGTa (UGT708C1), AB909375; FeCGTb (UGT708C2), AB909376; OsCGT, FM179712; UGT708A6, GRMZM2G162783; UGT71E2, AB294401; NtGT1a (UGT71A6), AB052557; ScUGT2, AB537179; CbB6GT, AF374004; UGT707B1, HE793682; RhGT1, AB201048; GmIF7GT (UGT88E3), GLYMA16G29400; UGT88D6, AB362990; MdPGT (UGT88F1), EU246349; PcPGT (UGT88F2), FJ854496; UGT72B1, Q9M156; IroB, CAE55724; UrdGT2, AAF00209; PfA5GT, AB013596; NtGT2 (UGT75L3), AB072919; ZmIAAGT, L34847; AtSAGT (UGT74F2), O22822; AtNGT (UGT76C1), AED90934; VvGT1, P51094; UGT78G1, A6XNC6; UGT73B4, Q7Y232; NtIS5a, U32644; UGT89C1, Q9LNE6; UGT94D1, AB333799; CaUGT, AB443870; UGT79A2, AY345985. Detailed information is shown in Table S5.



	R_1	R_2
1	H	H
2	OH	H
3	OH	OH

	R_1	R_2	R_3
5	H	H	H
6	H	OH	H
7	H	OH	OH
8	OH	OH	OH

	R_1	R_2
11	CH_3	OH
12	H	OH
13	CH_3	H

Figure 1. Structures of phenolic compounds used in this study.
(a) A proposed pathway of C-glycosylation in flavonoid biosynthesis (Kerscher and Franz, 1988; Brazier-Hicks et al., 2009).
(b) Structures of phenolic compounds: [1] 2-hydroxypinocembrin; [2] 2-hydroxynaringenin; [3] 2-hydroxyeriodictyol; [4] naringenin; [5] chrysin; [6] apigenin; [7] luteorin; [8] quercetin; [9] phloretin; [10] 2-phenyl-2′,4′,6′-trihydroxyacetophenone; [11] 2′,4′,6′-trihydroxyacetophenone; [12] 2,4,6-trihydroxybenzaldehyde; [13] 2′,4′-dihydroxyacetophenone.

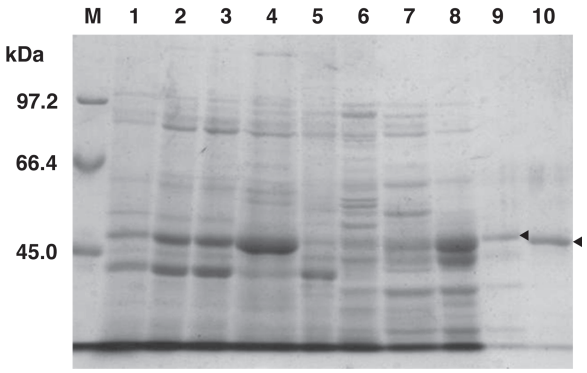


Figure 2. SDS-PAGE analysis of proteins from buckwheat purified C-glucosyltransferase.
Samples from all steps of the CGT purification process were separated on a 12.5% SDS-polyacrylamide gel. Lanes 1–5 contain 10 μg each of crude extract, ammonium sulfate precipitation, ultracentrifugation, Phenyl Sepharose CL-4B, and Sephadex G-100, respectively; Lane 6 contains 5 μg of DEAE-Sepharose CL-6B; and lanes 7–10 contain 3 μg each of Cibacron Blue3GA, Superdex G-200 HR 10/30 (lane 8), Mono Q fraction A (lane 9), and B (lane 10), respectively. Following separation, the gel was stained with coomassie brilliant blue R-250. Arrowheads indicate the purified CGTs. Standard proteins were loaded on lane M.

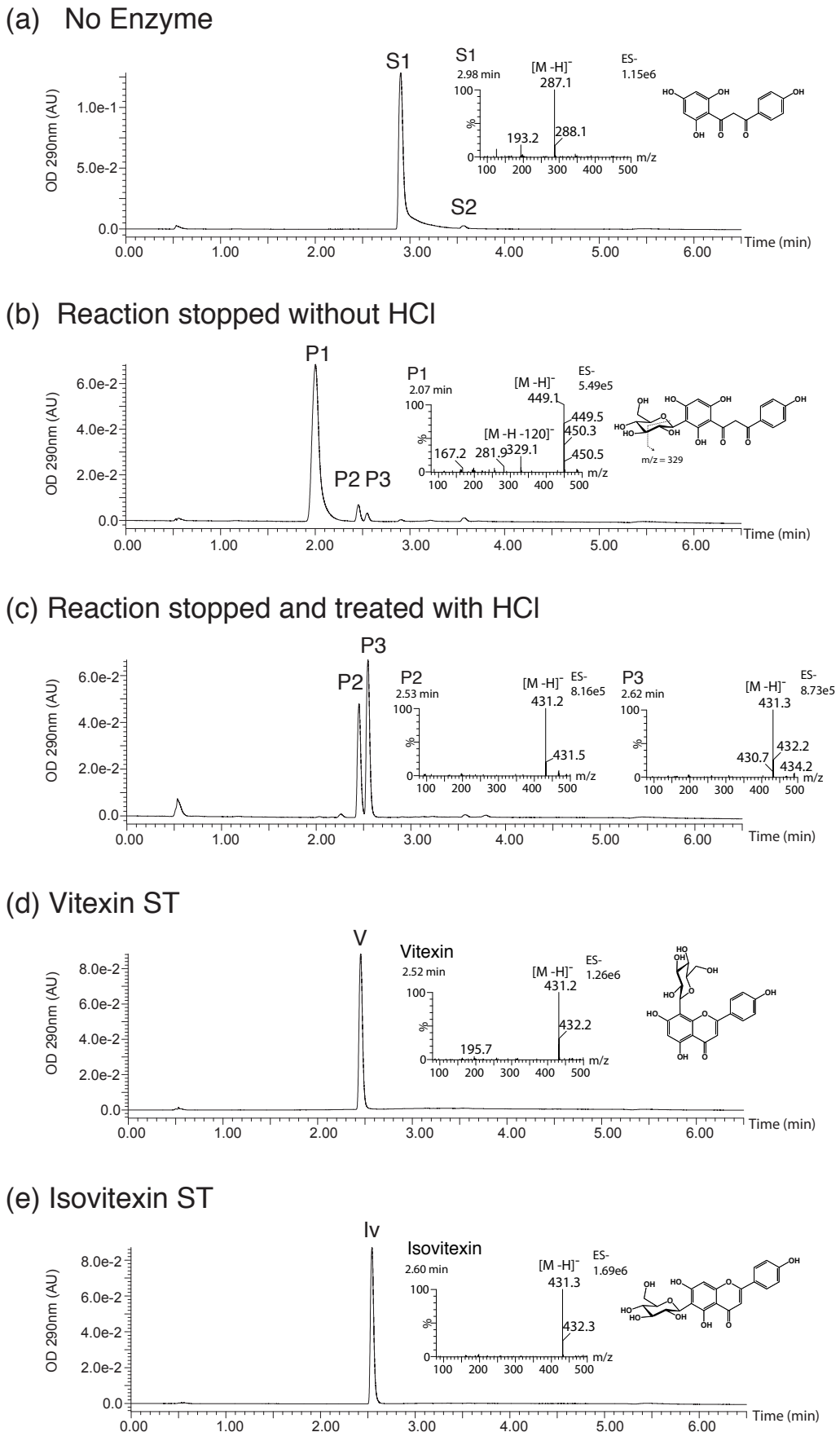


Figure 3. HPLC-MS analysis of the recombinant FeCGTa reaction products of 2-hydroxynaringenin.
The reaction was performed as described in Experimental Procedures. Each panel shows a chromatogram from the following reaction conditions: (a) without enzyme; (b) stopped with methanol instead of HCl; (c) 2 M HCl-treated; authentic compounds of (d) vitexin, and (e) isovitexin. The eluates were monitored at 290 nm using a diode array detector. The ES-MS– corresponding to the indicated peak is shown inside each figure. The retention time of MS peaks were delayed by 0.08 min compared to that of the diode array. Peak identification: S1, 2-hydroxynaringenin (substrate); S2, apigenin (dehydrated substrate); P1, 2-hydroxynaringenin C-glucoside; P2 and V, vitexin; P3 and Iv, isovitexin.

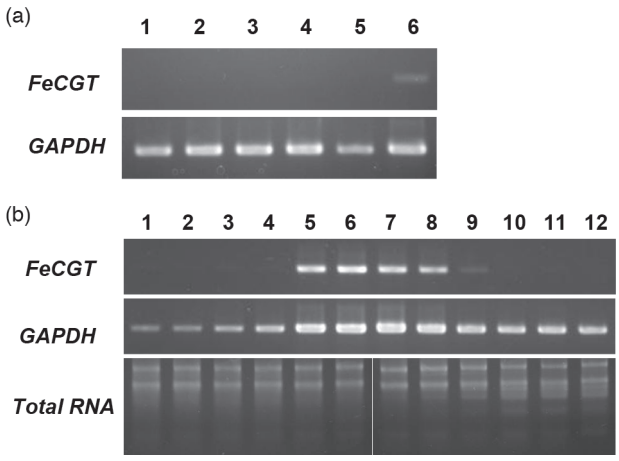


Figure 4. RT-PCR analysis of the FeCGT transcripts.
Total RNAs were extracted from various organs of buckwheat plants, followed by RT-PCR as described in Experimental Procedures. A 509-bp cDNA fragment was expected for *FeCGT* mRNA, and 584-bp fragment for glyceraldehyde-3-phosphate dehydrogenese (*gapdh*) mRNA. (a) Agarose gel electrophoresis demonstrating the organ-specific expression of *FeCGTs* in flower (lane 1), leaf (lane 2), stem (lane 3), root (lane 4), hypocotyl (lane 5), and cotyledon (lane 6). (b) Changes in the amounts of *FeCGT* transcripts in the cotyledon during seed germination and cotyledon development. Total RNAs were extracted from seeds or cotyledons after cutting out hypocotyl and root. As the amount of *gapdh* transcript also changed during germination, total RNA (400 ng) of each sample was also shown as reference. Agarose gel electrophoresis demonstrating the expression of *FeCGTs* in seeds before sowing (lane 1); 2.5, 6, and 10 h after sowing (lane 2–4); emergence of radicle (1 day after sowing; lane 5); elongation of root and root hair formation (1.5 and 2 days after sowing; lanes 6, 7); splitting of seed coat and formation of the cotyledon (3 and 5 days after sowing; lanes 8, 9); development of the cotyledon and removal of seed coat (7 days after sowing; lanes 9, 10); mature cotyledon (3 weeks and 2 months on soil; lane 11, 12). Pictures of each germination stage are shown in Figure S7.

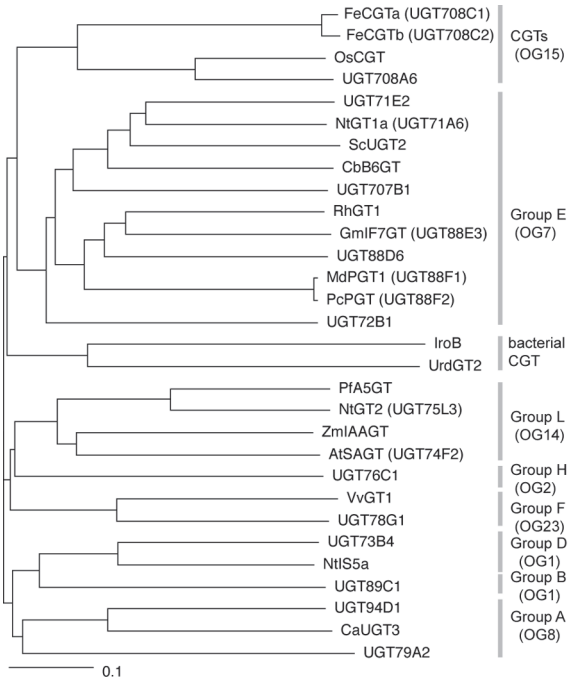


Figure 5. Molecular phylogenetic tree constructed based on the deduced amino acid sequences of FeCGTs and related glycosyltransferases.

The tree was constructed by the neighbor-joining method using ClustalW (http://clustalw.ddbj.nig.ac.jp/). The lengths of lines indicate the relative distance between nodes. The grouping of plant UGTs is shown in accordance with a proposal by Li et al. (2001) and also, for orthologous groups by Yonekura-Sakakibara and Hanada (2011). Abbreviations and Genbank accession nos. of UGTs are as follows: FeCGTa (UGT708C1), AB909375; FeCGTb (UGT708C2), AB909376; OsCGT, FM179712; UGT708A6, GRMZM2G162783; UGT71E2, AB294401; NtGT1a (UGT71A6), AB052557; ScUGT2, AB537179; CbB6GT, AF374004; UGT707B1, HE793682; RhGT1, AB201048; GmIF7GT (UGT88E3), GLYMA16G29400; UGT88D6, AB362990; MdPGT (UGT88F1), EU246349; PcPGT (UGT88F2), FJ854496; UGT72B1, Q9M156; IroB, CAE55724; UrdGT2, AAF00209; PfA5GT, AB013596; NtGT2 (UGT75L3), AB072919; ZmIAAGT, L34847; AtSAGT (UGT74F2), O22822; AtNGT (UGT76C1), AED90934; VvGT1, P51094; UGT78G1, A6XNC6; UGT73B4, Q7Y232; NtIS5a, U32644; UGT89C1, Q9LNE6; UGT94D1, AB333799; CaUGT, AB443870; UGT79A2, AY345985. Detailed information is shown in Table S5.

Table S1. Glucosylation activities in cell-free extracts of buckwheat organs.

Organ	*O*-Glucosylation activity (pkat mg_protein^{-1})	*C*-Glucosylation activity (pkat mg_protein^{-1})
Seed	0.3 ± 0.1	3.3 ± 0.3
Hypocotyl	25.8 ± 8.1	11.3 ± 4.9
Cotyledon (seedlings)	62.9 ± 16.2	62.4 ± 10.1
Cotyledon (mature plant)	37.7 ± 7.8	ND[a]
Leaf	25.3 ± 11.8	ND
Root	ND	ND
Stem	29.2 ± 8.1	ND
Flower	769.3 ± 358.9	3.8 ± 4.5

A frozen materials was ground in a mortar under liquid nitrogen, then an aliquot (0.3 g fresh weight) was added with 1 mL of ice-cold buffer A (100 mM Tris-HCl, pH 8.0, 1 mM DTT) containing 5% (w/v) polyvinylpolypyrrolidone. The mixture was further homogenized by sonication for 12 × 5 sec at an amplitude of 20% (Vibra Cell VCX400, Sonic & Materials, Inc., Newtown, CT, USA). The homogenate was centrifuged at 17,000 × *g* for 10 min, and the supernatant was desalted using a NAP-5 columns (GE Healthcare) equilibrated with buffer A and was used as an enzyme fraction. The reaction was performed as described in the experimental procedure, using 100 μM quercetin and 100 μM 2-hydroxypinocembrin as substrates for *O*-glucosylation and *C*-glucosylation, respectively. Data was shown as average ± SD (n = 3–4). [a]ND, not detected.

Table S2. Buckwheat UGT genes isolated by PCR-based cloning.

Gene Name	UGT name	Gene Accession No.
FeGT1	UGT92H1	AB909377
FeGT2	UGT92J1	AB909378
FeGT3	UGT714A1	AB909379
FeGT4	UGT85Z1	AB909380
FeGT5	UGT88J1	AB909381
FeGT6	UGT89G1	AB909382
FeGT7	UGT72AC1	AB909383
FeGT8	UGT71U1	AB909384
FeGT9	UGT75R1	AB909385
FeGT10	UGT72B19	AB909386
FeGT11	UGT89D6	AB909387
FeGT12	UGT709H1	AB909388

Table S3. Purification of *C*-glucosyltransferases from buckwheat seedlings.

Purification Step	Protein (mg)	Total activity (nkat)	Specific activity (nkat mg_protein^{-1})	Purity (-fold)	Yield (%)
Crude extract	16,969	118.8	0.007	1	100
$(NH_4)_2SO_4$	8,160	163.2	0.02	2.9	137
Ultracentrifuge	7,706	146.4	0.019	2.7	123
Phenyl-Sepharose	2,150	139.8	0.065	9.3	118
Sephadex G-100	706	81.2	0.115	16	68
DEAE-Sepharose	87	134.9	1.55	230	118
Cibacron Blue 3GA	14.2	93	6.55	936	78
Superdex G-200	2.0	68	34	4,857	57
Mono Q (Fraction A)	0.21	5.25	25	3,571	4
Mono Q (Fraction B)	0.21	10.75	51.2	7,314	9

Table S4. PCR Primers used in this work.

Primer name	Sequence
Primer-Fw1	5'-CCNAAYGTNGAYCCNTTYTT-3'
Primer-Fw2	5'-GTNGAYCCNTTYTTYYTNMG-3'
Primer-Rv	5'-GGYTGYTGRTCNARCCANGG-3'
M13-Rv	5'- CAGGAAACAGCTATGACCAT -3'
M13-40U	5'- GTTTTCCCAGTCACGAC -3'
M13-21U	5'- TGTAAAACGACGGCCAGT-3'
T7-Pro	5'- TAATACGACTCACTATAGGG -3'
CGT-F1	5'- AACTACACCTTCACCACCAC -3'
CGT-R1	5'- GTGGTGGTGAAGGTGTAGTT -3'
CGT-F2	5'- GTCAACACATTCGACTCCTT -3'
CGT-R2	5'- AAGGAGTCGAATGTGTTGAC -3'
CGT-R3	5'- TTCTCTCTCTGATCCCCGTG -3'
CGT-Fw-Nde	5'-CCGGCATATGATGGGAGATTTAACAACTTC -3'
CGT-Rv-Eco	5'- CCTCGAGAATTCAACGTTTAAGACTTCCGA -3'
FeGAPDH-F	5'- ATCAACGGTTTCGGAAGAAT -3'
FeGAPDH-R	5'- CTTCCACCTCTCCAGTCCTT -3'
dT-T3 primer	5'- ATTAACCCTCACTAAAGGGTTTTTTTTTTTTTTTVV-3'

Table S5. List of UGTs used for the phylogenetic analysis in Figure 5.

Name of UGT	Description	Accession No.
FeCGTa (UGT708C1)	flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum*	AB909375
FeCGTb (UGT708C2)	flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum*	AB909376
OsCGT	flavonoid *C*-glucosyltransferases from *Oryza sativa*	FM179712
UGT708A6	flavonoid *C*-glucosyltransferases from *Zea mays*	GRMZM2G162783
UGT71E2	tetrahydroxychalcone 2'-glucosyltransferase from *Catharanthus roseus*	AB294401
NtGT1a (UGT71A6)	flavonol and coumarin glucosyltransferase from *Nicotiana tabacum*	AB052557
ScUGT2	3-deoxyanthocyanidin 5-*O*-glucosyltransferase from *Sinningia cardinalis*	AB537179
CbB6GT	betanidin 6-*O*-glucosyltransferase from *Cleretum bellidiforme*	AF374004
UGT707B1	flavonol glucosyltransferase from *Crocus sativus*	HE793682
RhGT1	anthocyanidin 5,3-*O*-glucosyltransferase from *Rosa* hybrid	AB201048
GmIF7GT (UGT88E3)	isoflavone 7-*O*-glucosyltransferase from *Glycine max*	GLYMA16G29400
UGT88D6	flavonol glucosyltransferase from *Sesamum indicum*	AB362990
MdPGT (UGT88F1)	phloretin 2'-*O*-glucosyltransferase from *Malus pumila*	EU246349
PcPGT (UGT88F2)	phloretin 2'-*O*-glucosyltransferase *Pyrus communis*	FJ854496
UGT72B1	xenobiotic glucosyltransferase from *Arabidopsis thaliana*	Q9M156
IroB	salmochelin producing *C*-glycosyltransferase from *Escherichia coli*	CAE55724[a]
UrdGT2	angucycline *C*-glycosyltransferase from *Streptomyces fradiae*	AAF00209[a]
PfA5GT	anthocyanin-5-*O*-glucosyltransferase from *Perilla*	AB013596
NtGT2 (UGT75L3)	flavonol-7-*O*-glucosyltransferase from *N. tabacum*	AB072919
ZmIAAGT	indole-3-acetic acid glucosyltransferase from maize	L34847
AtSAGT(UGT74F2)	salicylic acid glucosyltransferase from *A. thaliana*	O22822
AtNGT (UGT76C1)	cytokinin-*N*-glucosyltransferase from *A. thaliana*	AED90934[a]
VvGT1	flavonoid-3-*O*-glucosyltransferases from *Vitis vinifera*	P51094
UGT78G1	flavonoid-3-*O*-glucosyltransferases from *Medicago truncatula*	A6XNC6
UGT73B4	flavonol and coumarin glucosyltransferase from *A. thaliana*	Q7Y232
NtIS5a	flavonol and coumarin glucosyltransferase from *N. tabacum*	U32644
UGT89C1	flavonol-7-*O*-rhamnosyltransferase from *A. thaliana*	Q9LNE6
UGT94D1	2'-*O*-glucoside-*O*-glucosyltransferase from *Sesamum indicum*	AB333799
CaUGT	flavonoid glucoside 1,6-glucosyltransferase from *Catharanthus roseus*	AB443870
UGT79A2	stevioside glucosyltransferase from *Stevia rebaudiana*	AY345985

[a] Accession Nos. for protein sequences.

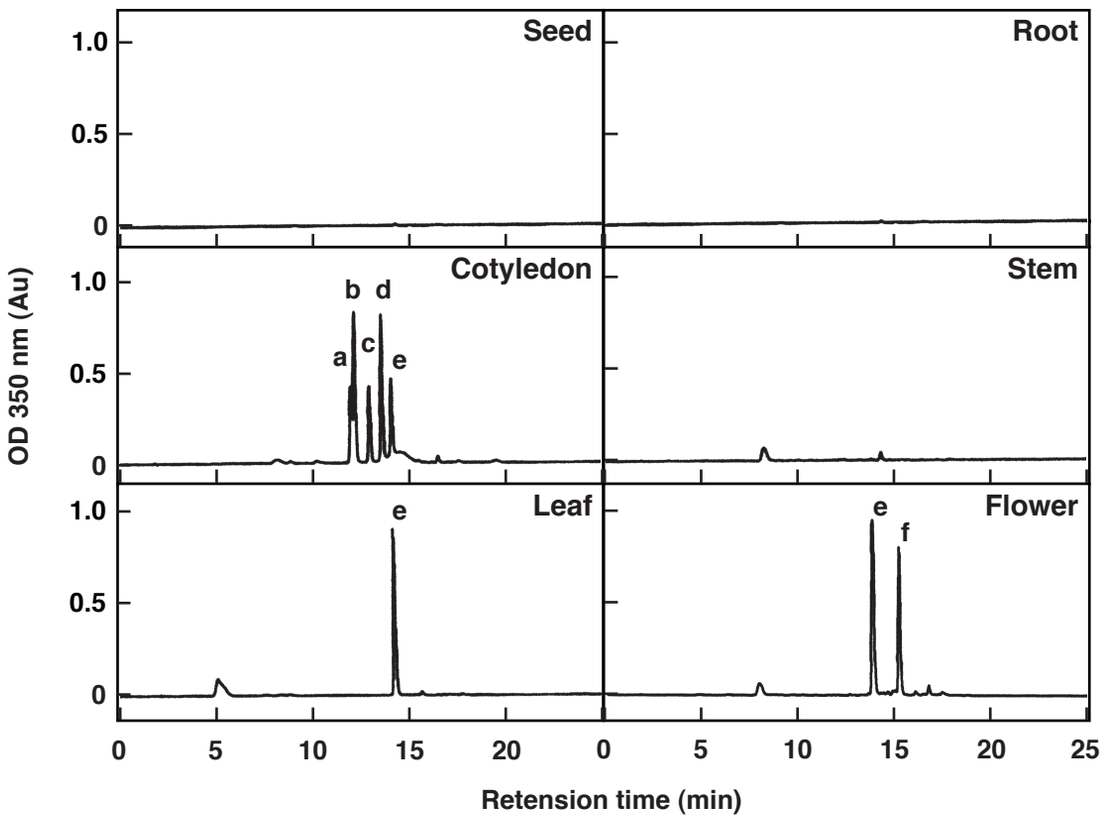


Figure S1. Distribution of phenolic compounds in buckwheat plants.

Methanol extracts of buckwheat plant organs were analyzed by HPLC. The column was eluted using a linear gradient of 20% to 60% solvent B (methanol containing 0.1% formic acid) in solvent A (0.1% formic acid) for 12 min, followed by 60% solvent B in A for 8 min at a flow rate of 1 ml min^{-1} at 40°C. The eluate was monitored at 350 nm using a diode array detector (SPD-M10Avp, Shimadzu). Peak identification: a, orientin; b, isoorientin; c, vitexin; d, isovitexin; e, rutin; f, quercetin pentoside (unidentified).

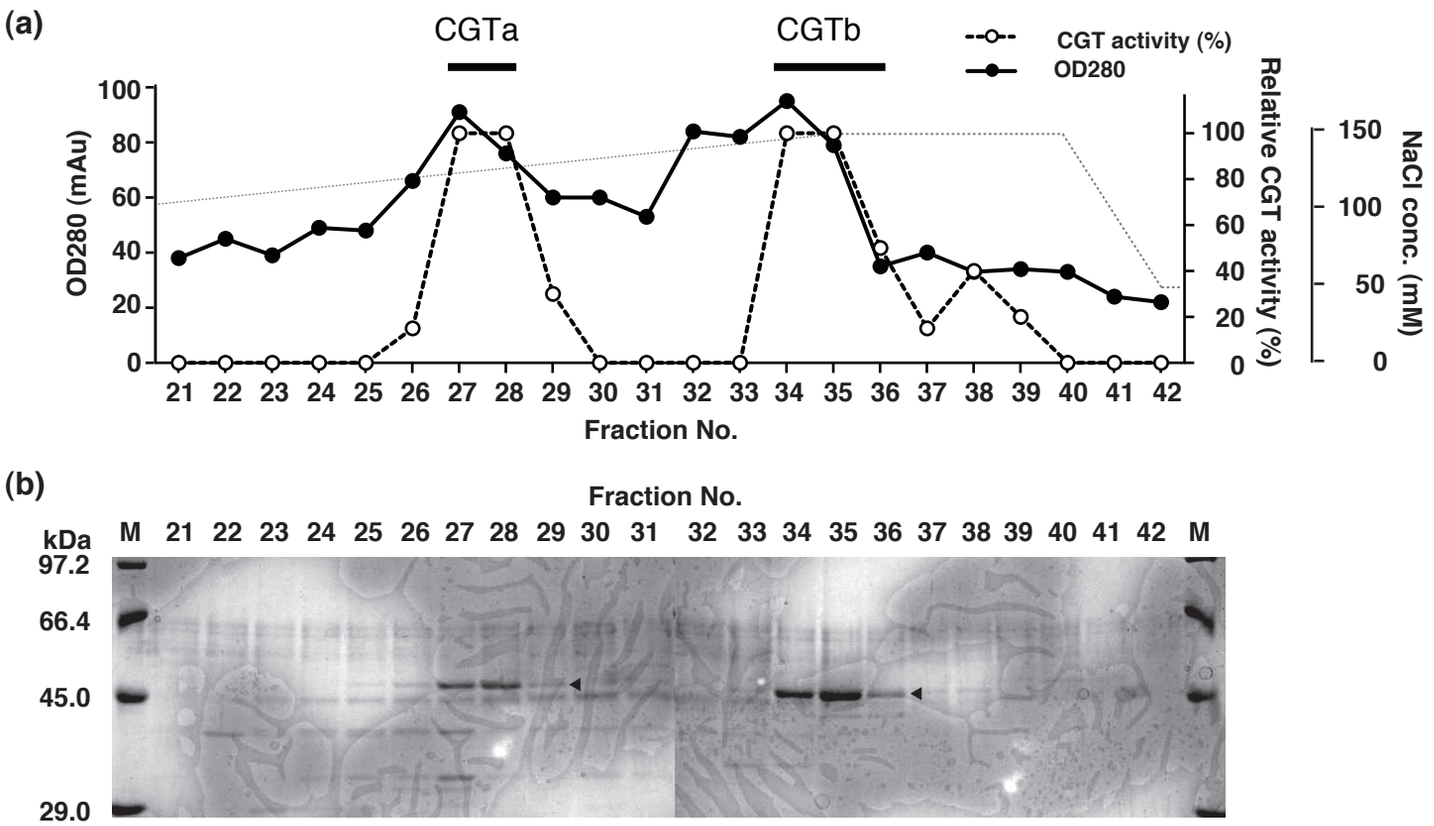


Figure S2. Purification of CGT from buckwheat seedlings by Mono Q anion exchange chromatography.
(a) Chromatogram of the purified CGT fraction on Mono Q. Solid line with filled circles and dotted line with open circles indicate protein concentration and CGT activity, respectively. The dashed gray line shows the NaCl gradient in the elution buffer. Active fractions separately pooled as CGTa and CGTb are indicated by the bars on top.
(b) SDS-PAGE analysis of the purified CGT fractions. The numbers indicate the fraction number on Mono Q shown in (a). Arrowheads indicate the purified CGTa and CGTb. Standard proteins were loaded on lane M.

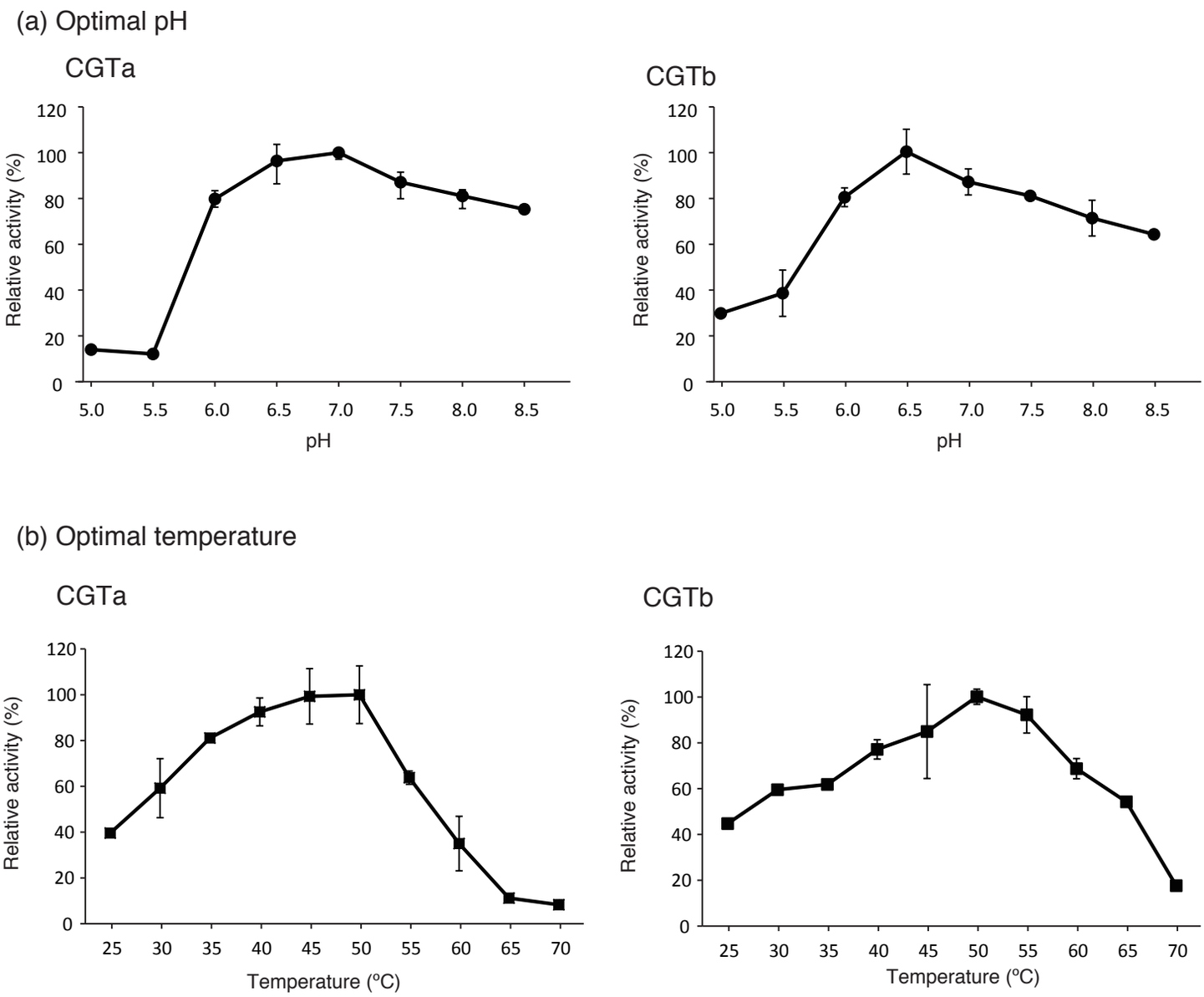


Figure S3. Properties of purified CGTa and CGTb.

The reaction of was performed as described in the "Experimental procedure" using 200 μM 2-hydroxypinocembrin and 2 mM UDP-glucose as substrates. Panels show the effects of pH (a) and temperature (b) for on the activity. For determination of pH optimum, the reaction was performed in the 100 mM potassium phosphate buffer containing 0.01 % BSA and 1 mM DTT, with pH of 5.0 to 8.5. For evaluation of the effect of temperature on the enzymatic reaction, the enzyme was incubated between 25 and 70 ºC in 100 mM potassium phosphate buffer (pH 6.5) containing 1 mM DTT.

(a)

```
FeCGT fr1          1   PNVDPFFLRY  10
                       *  ******.
OsCGT             86   PGADPFFLRF  95
```

Identities 7/10 (70%), Positives 8/10 (80%)

```
FeCGT fr2          3  ELLPWLDQQP   12
                      *** ******
UFOG5_MANES      260  ELLDWLDQQP  269
```

Identities 9/10 (90%), Positives 9/10 (90%)

(b)

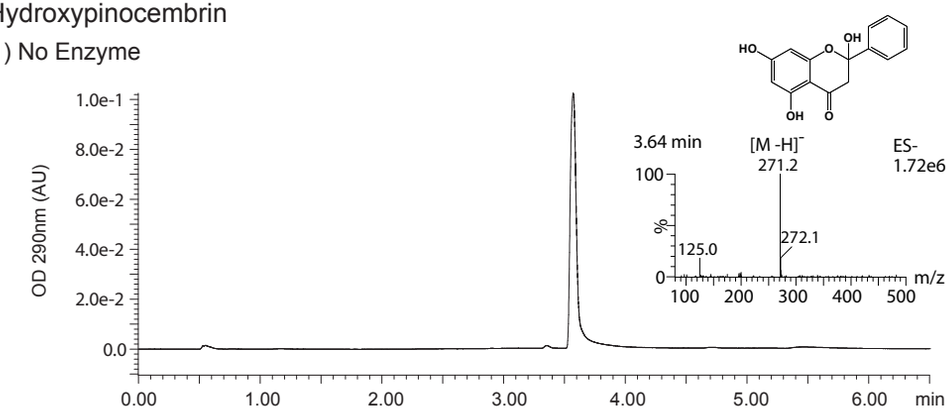
```
FeCGTa     1  MMGDLTTSFP ATTLTTNDQP HVVVCSGAGM GHLTPFLNLA SALSSAPYNC KVTLLIVIPL   60
              ********** *******.** ********** ***.****** *.*******.* **********
FeCGTb     1  MMGDLTTSFP ATTLTTNEQP HVVVCSGAGM GHLIPFLNLA STLSSAPYRC KVTLLIVIPL   60

FeCGTa    61  ITDAESHHIS SFFSSHPTIH RLDFHVNLPA PKPNVDPFFL RYKSISDSAH RLPVHLSALS   120
              ********** ********** ********** ********** ********** *******.*.
FeCGTb    61  ITDAESHHIS SFFSSHPTIH RLDFHVNLPA PKPNVDPFFL RYKSISDSAH RLPVHLSTLA   120

FeCGTa   121  PPISAVFSDF LFTQGLNTTL PHLPNYTFTT TSARFFTLMS YVPHLAKSSS SSPVEIPGLE   180
              ********** ********** ********** ********** ********** **********
FeCGTb   121  PPISAVFSDF LFTQGLNTTL PHLPNYTFTT TSARFFTLMS YVPHLAKSSS SSPVEIPGLE   180

FeCGTa   181  PFPTDNIPPP FFNPEHIFTS FTISNAKYFS LSKGILVNTF DSFEPETLSA LNSGDTLSDL   240
              ********** ****.***** ******.*.* *****.**** ********** *****.*.**
FeCGTb   181  PFPTDNIPPP FFNPDHIFTS FTISNANYLS LSKGIIVNTF DSFEPETLSA LNSGDSLPDL   240

FeCGTa   241  PPVIPIGPLN ELEHNKQEEL LPWLDQQPEK SVLYVSFGNR TAMSSDQILE LGMGLERSDC   300
              ********** ********** ********** ********** ********** **********
FeCGTb   241  PPVIPIGPLN ELEHNKQEEL LPWLDQQPEK SVLYVSFGNR TAMSSDQILE LGMGLERSDC   300

FeCGTa   301  RFIWVVKTSK IDKDDKSELR KLFGEELYLK LSEKGKLVKW VNQTEILGHT AVGGFLSHCG   360
              ********** ********** ********.* ********** ********** **********
FeCGTb   301  RFIWVVKTSK IDKDDKSELR KLFGEELYVK LSEKGKLVKW VNQTEILGHT AVGGFLSHCG   360

FeCGTa   361  WNSVMEAARR GVPILAWPQH GDQRENAWVV EKAGLGVWER EWASGIQAAI VEKVKMIMGN   420
              ********** ********** ********** ********** **.****.** **********
FeCGTb   361  WNSVMEAARR GVPILAWPQH GDQRENAWVV EKAGLGVWER EWSSGIQVAI VEKVKMIMGN   420

FeCGTa   421  NDLRKSAMKV GEEAKRACDV GGSSATALMN IIGSLKR  457
              ****.**..* ********** ********** *******
FeCGTb   421  NDLRNSAVRV GEEAKRACDV GGSSATALMN IIGSLKR  457
```

Figure S4. Analysis of peptide sequences of the purified CGTs.

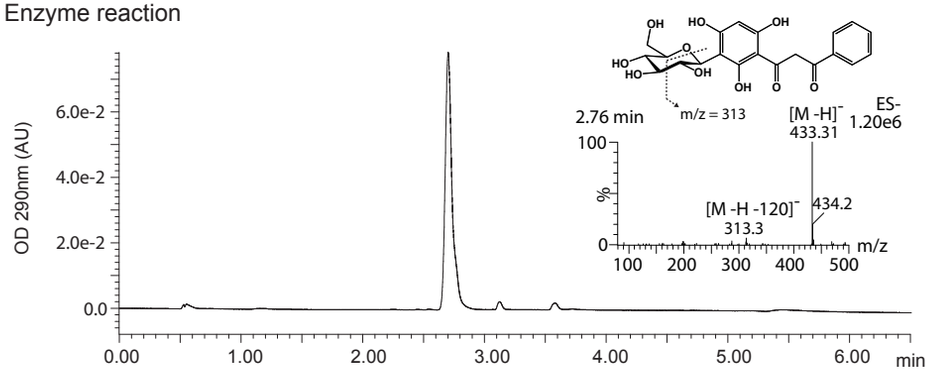
(a) Results of the BLASTP search using the peptide sequences obtained by LC-MS/MS analysis of CGTs. Two peptide fragments with sequence homology against the reported glycosyltransferases are shown. The accession Nos. are OsCGT, CAQ77160; UFOG5_MANES, Q40287. (b) Comparison of the amino acid sequences of FeCGTa and FeCGTb. Asterisks and dots represent the identical and similar amino acids residues between two proteins.The sequence observed in the peptide sequence of CGTa and CGTb are highlighted by blue and orange boxes, respectively.

(a) 2-Hydroxypinocembrin

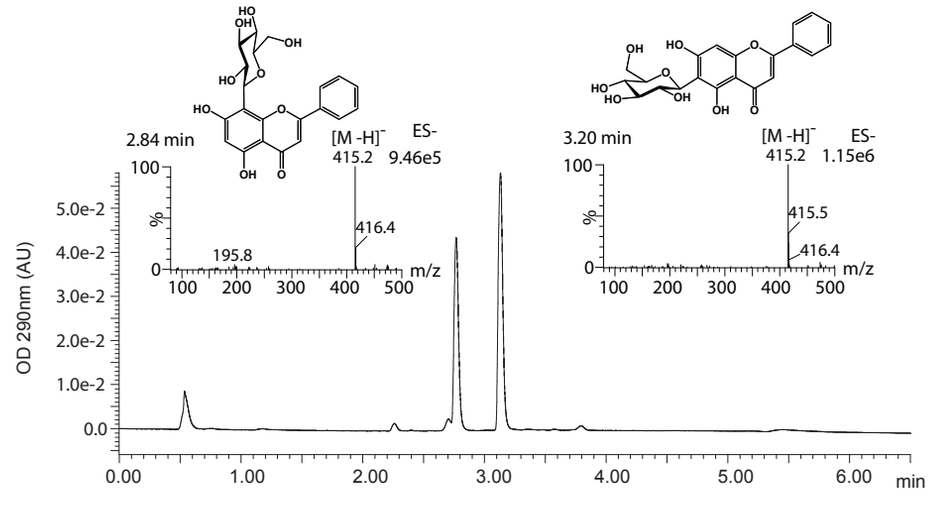
(1) No Enzyme



(2) Enzyme reaction

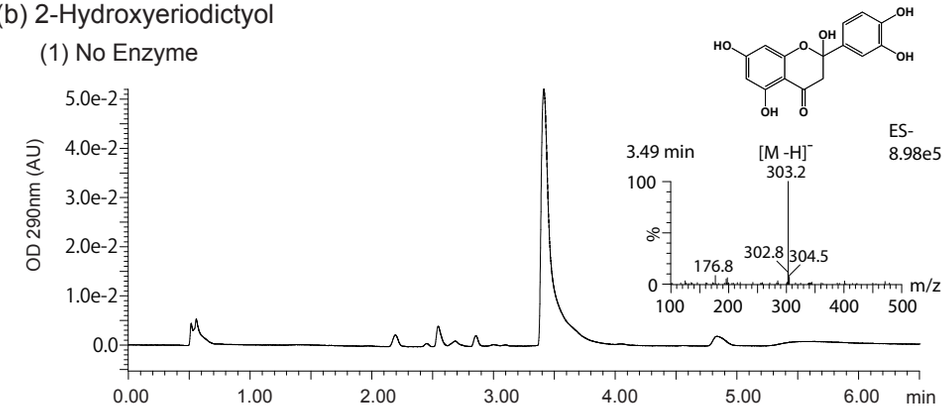


(3) Enzyme reaction + acid hydrolysis

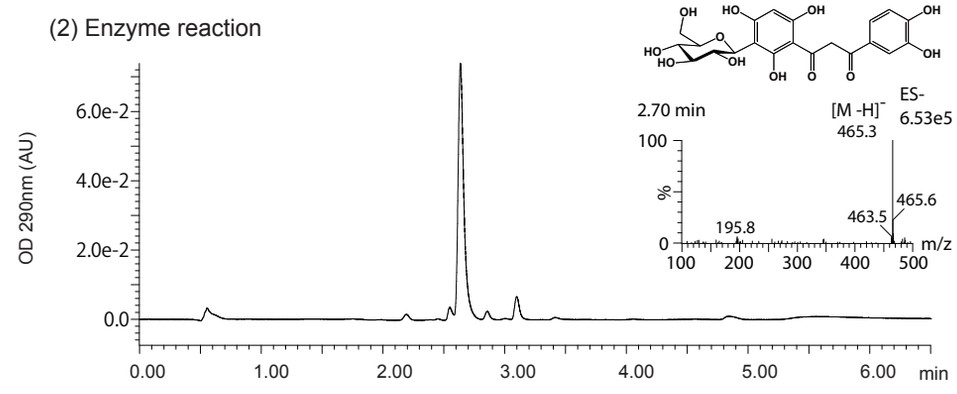


(b) 2-Hydroxyeriodictyol

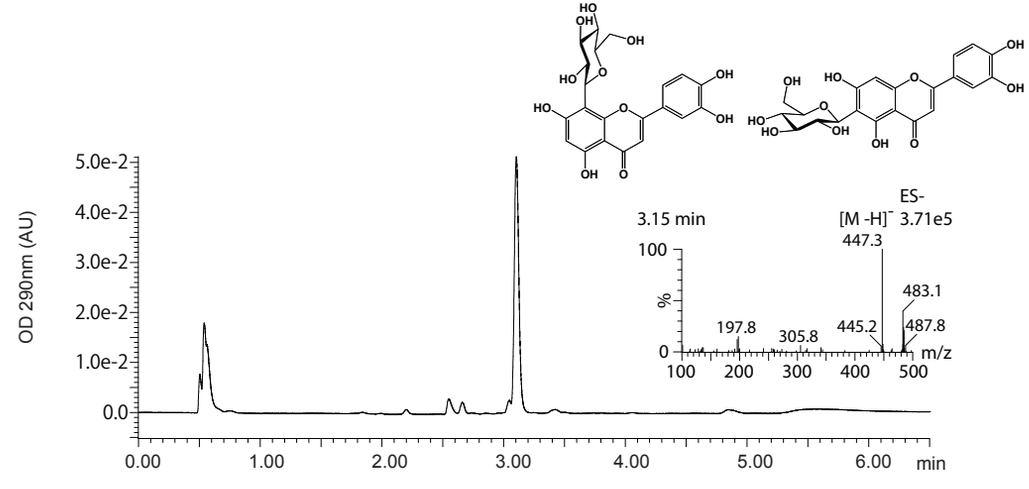
(1) No Enzyme



(2) Enzyme reaction



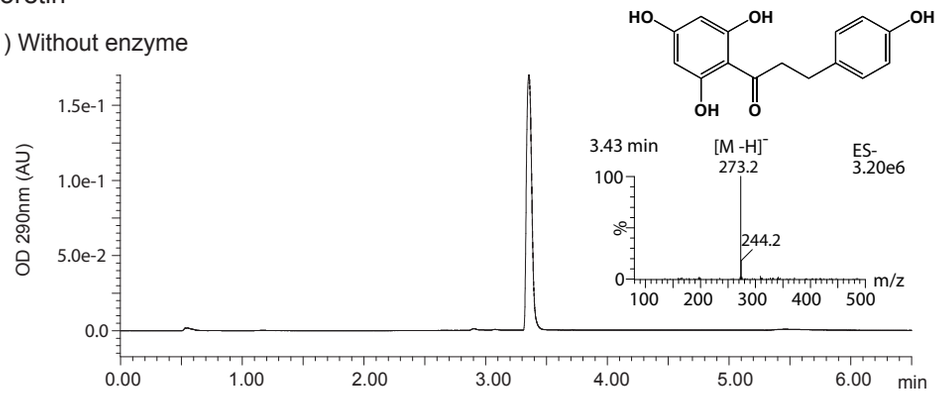
(3) Enzyme reaction + acid hydrolysis [a]



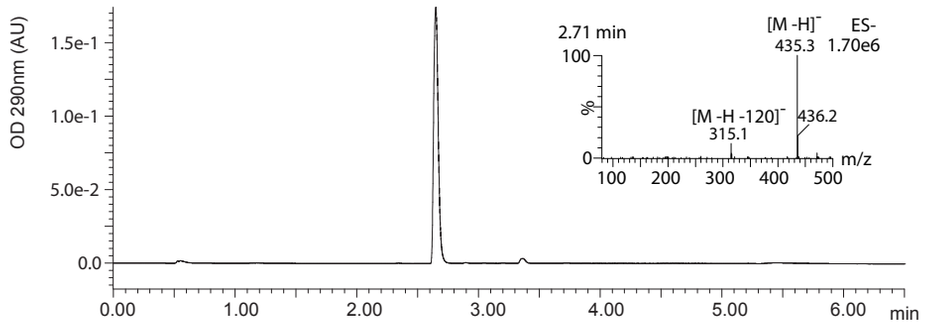
[a] Two *C*-glucosylated products (orientin and isoorientin) could not be separated under this condition.

(c) Phloretin

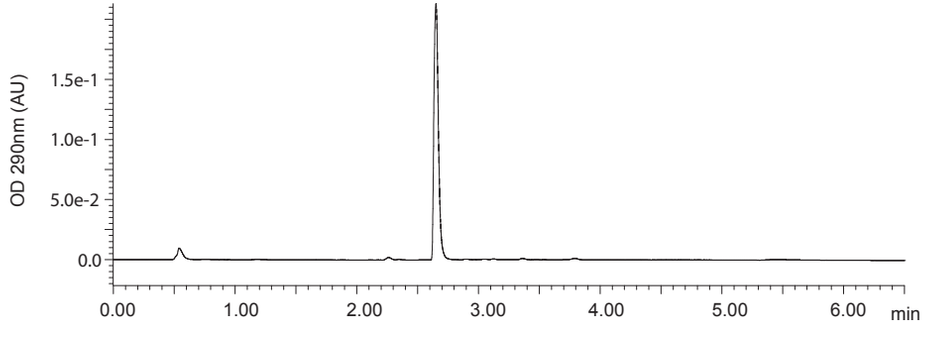
(1) Without enzyme



(2) Enzyme reaction

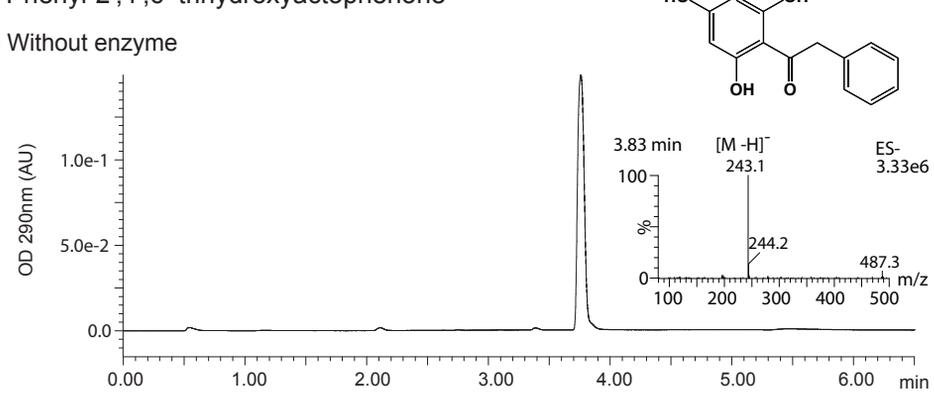


(3) Enzyme reaction + acid hydrolysis

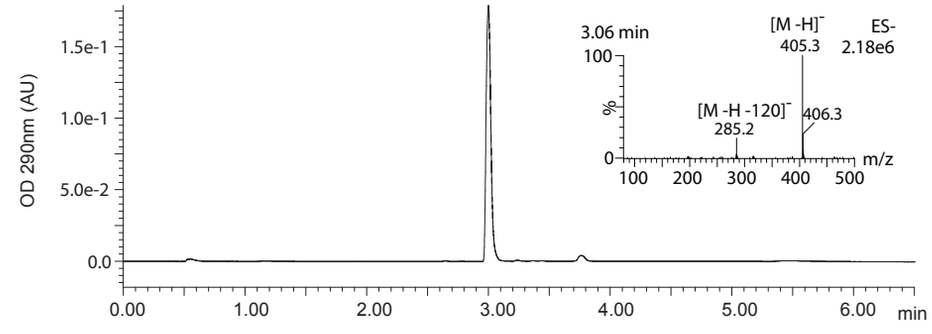


(d) 2-Phenyl-2',4',6'-trihydroxyactophenone

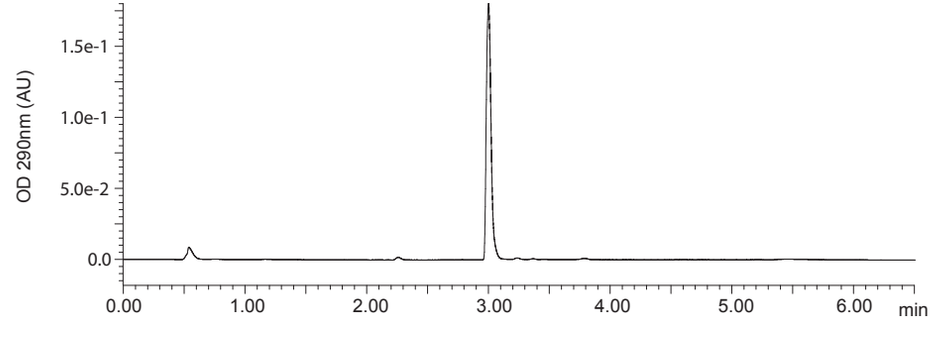
(1) Without enzyme



(2) Enzyme reaction

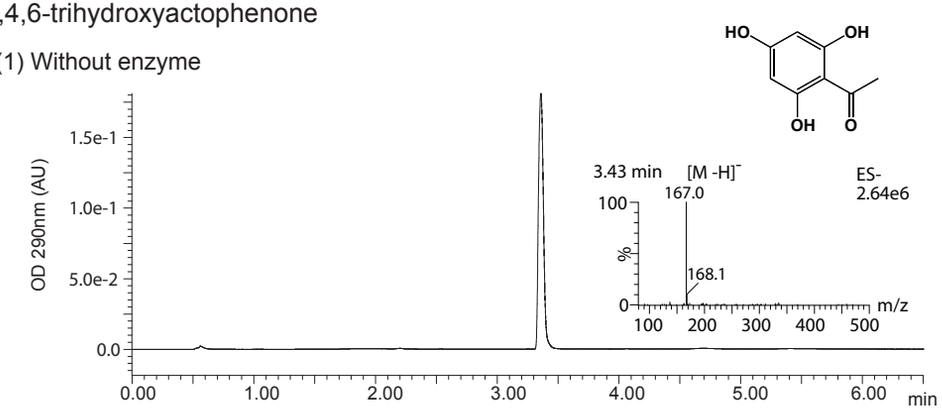


(3) Enzyme reaction + acid hydrolysis

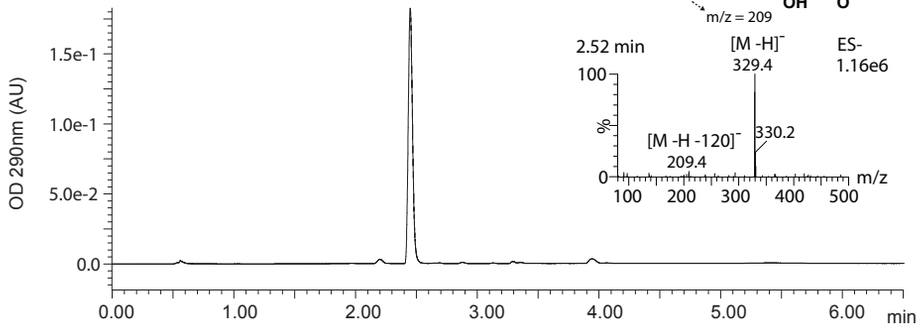


(e) 2,4,6-trihydroxyactophenone

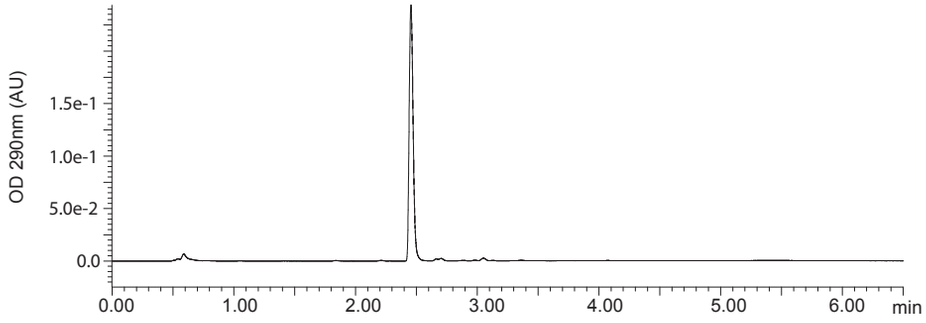
(1) Without enzyme



(2) Enzyme reaction

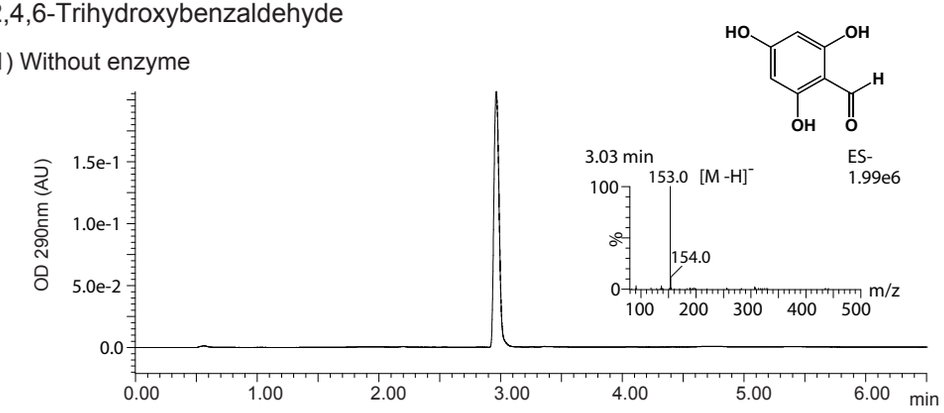


(3) Enzyme reaction + acid hydrolysis

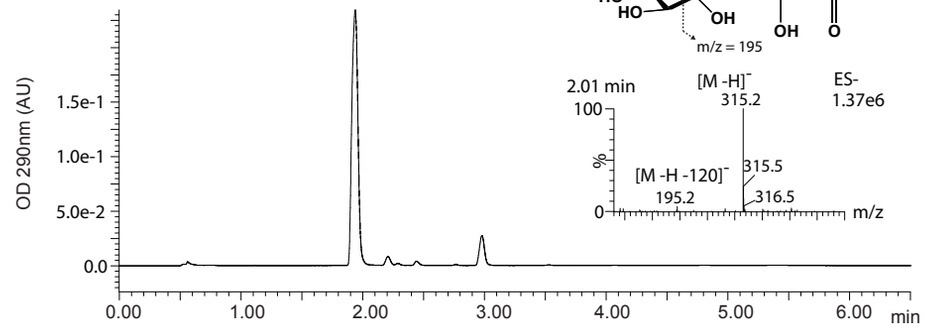


(f) 2,4,6-Trihydroxybenzaldehyde

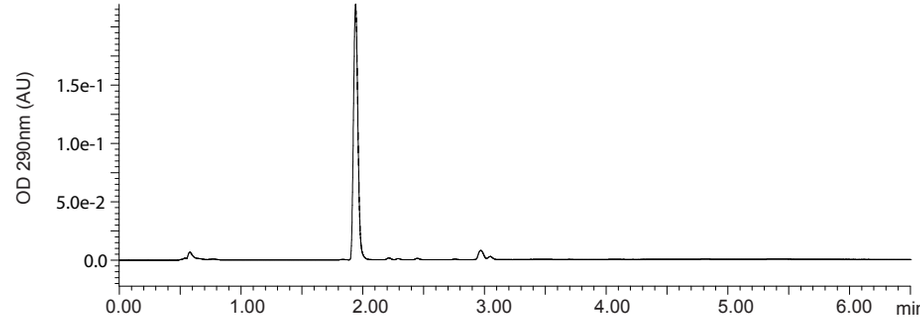
(1) Without enzyme



(2) Enzyme reaction

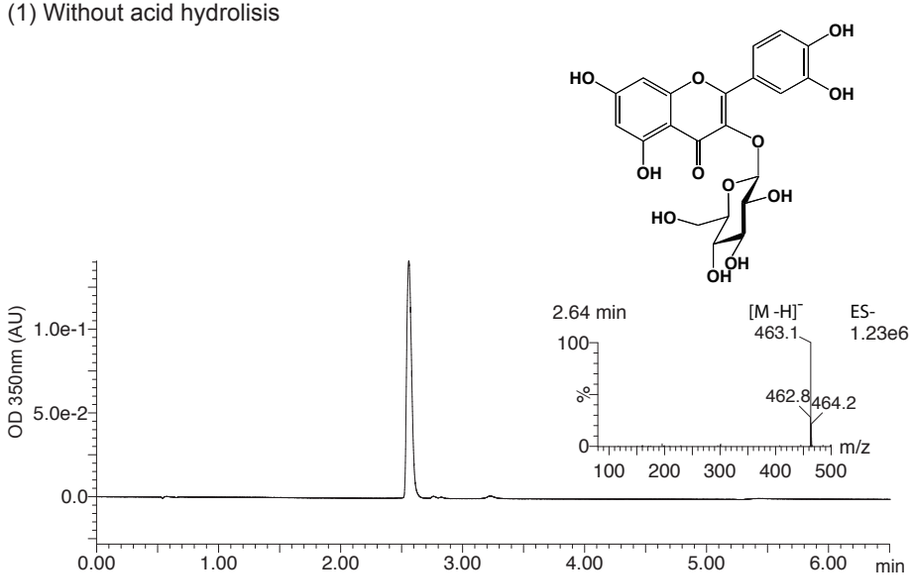


(3) Enzyme reaction + acid hydrolysis



(g) Quercetin 3-*O*-glucoside

(1) Without acid hydrolisis



(2) With acid hydrolysis

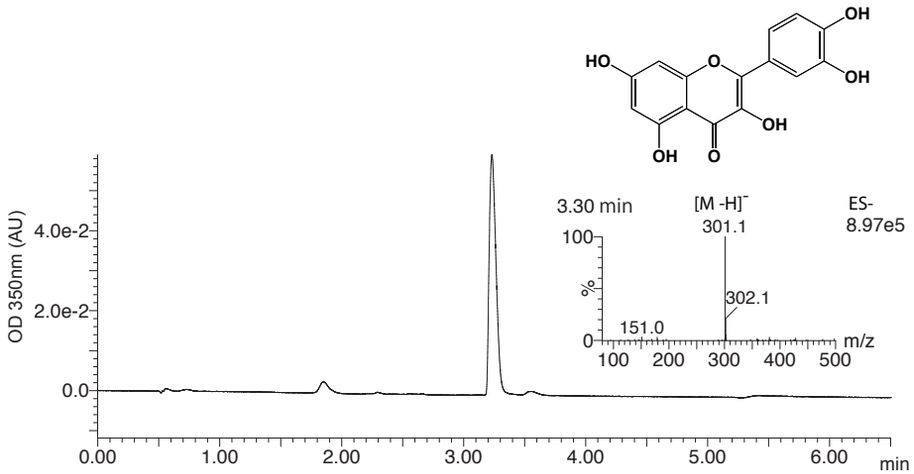


Figure S5. UPLC-MS analysis of the recombinant FeCGTa reaction against several phenolic substrates.

UPLC chromatograms of the enzyme reaction without enzyme, with enzyme, and with enzyme followed by acid hydrolysis were shown. ES-MS$^-$ corresponding to the indicated peak and the structure of compounds were shown in the figure. The retention time of MS peaks were delayed by 0.08 min compared to that of the diode array. The substrate of each reaction is as follows: (a) 2-hydroxypinocembrin; (b) 2-hydroxyeriodictyol; (c) phloretin; (d) 2-phenyl-2′,4′,6′-trihydroxyacetophenone; (e) 2′,4′,6′-trihydroxyacetophenone; (f) 2,4,6-trihydroxybenzaldehyde;(g) Acid hydrolysis of authentic quercetin-3-*O*-glucoside. [a]Two *C*-glucosylated products (orientin and isoorientin) could not be separated under this condition.

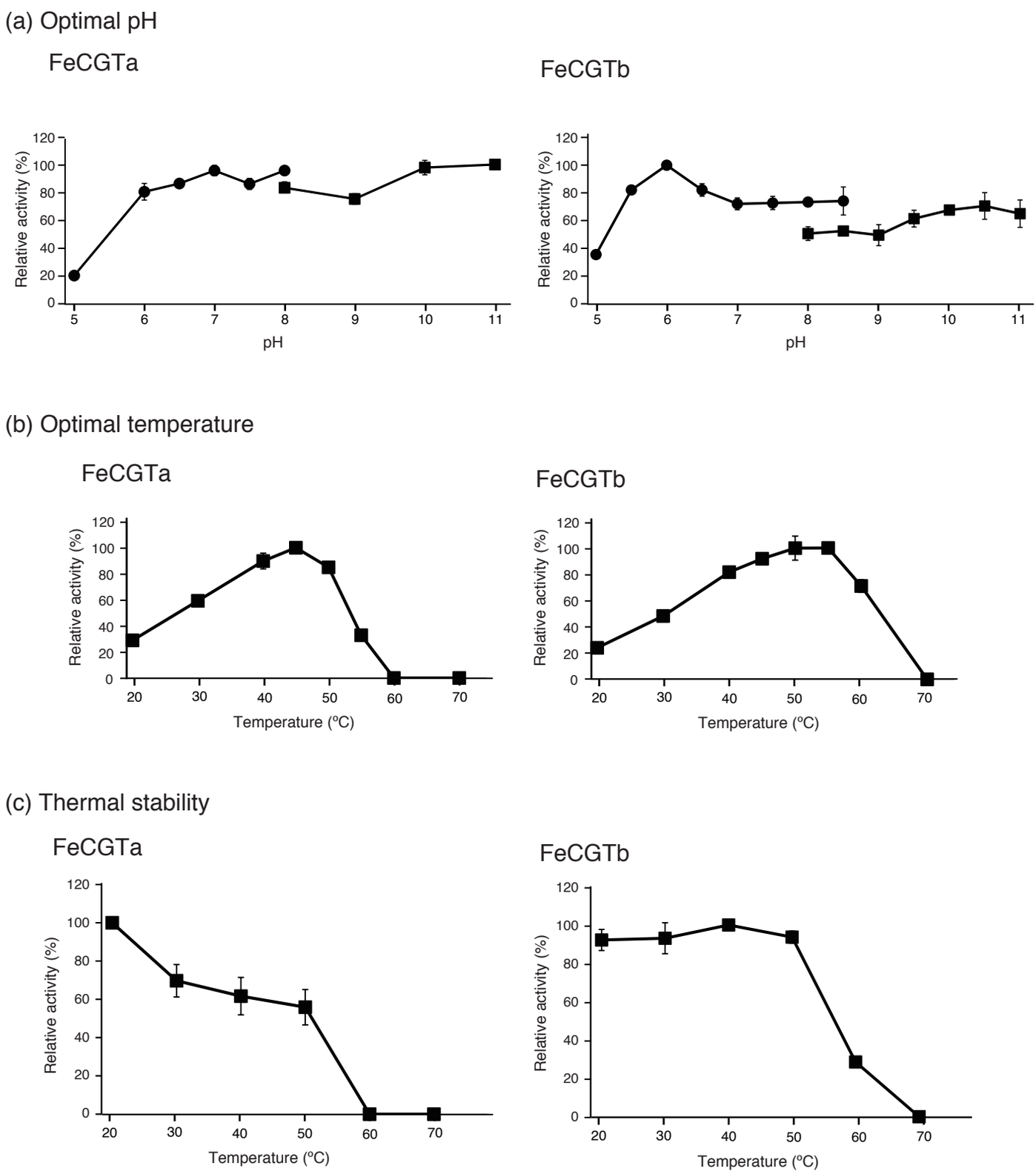


Figure S6. Properties of recombinant FeCGTa and FeCGTb.

The reaction was performed as described in the "Experimental procedure" using 2-hydroxypinocembrin and UDP-glucose as substrates. For each enzyme, we determined their (a) optimal temperature for activity; (b) thermal stability properties; and (c) optimal pH for activity.

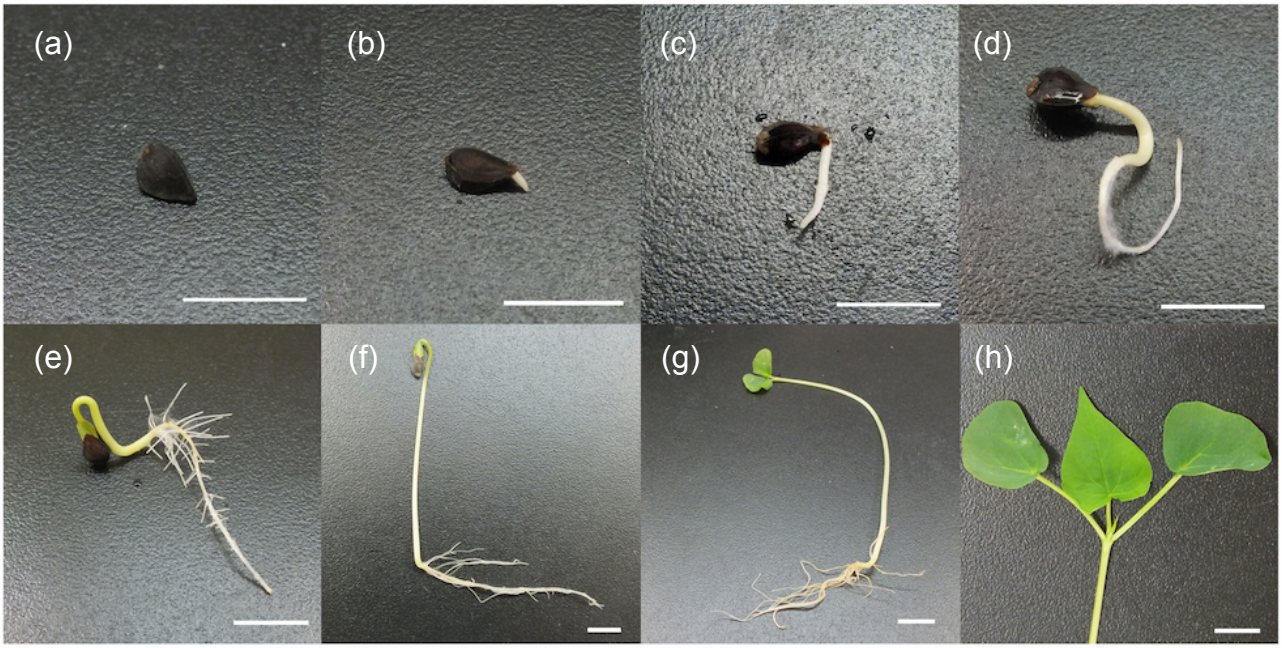


Figure S7. Pictures of developmental stages during germination and cotyledon development of buckwheat seed used for RT-PCR in Figure 4.
Each panels show the developmental stages of seed germinations: (a) the seed 6 h after sowing; (b) emergence of radicle 1 day after sowing; (c, d) elongation of root and root hair formation after 1.5 days (C) and 2 days (D); (e, f) splitting of seed coat and formation of cotyledon after 3 days (E) and 5 days (F); (g) development of cotyledon and removal of seed coat after 7 days; (h) mature cotyledon after 3 weeks. White scale bar indicates 1 cm.

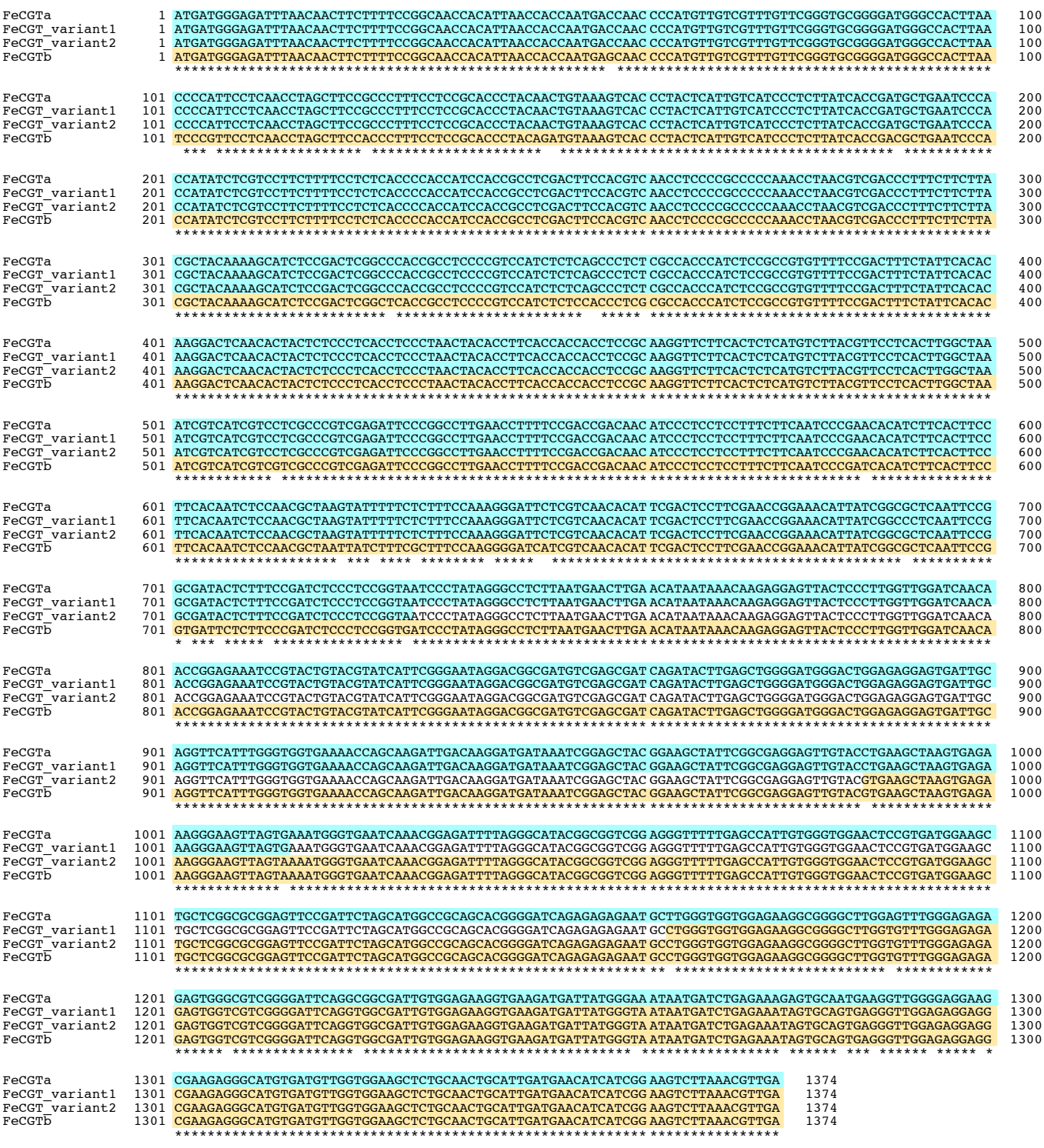


Figure S8. Comparison of the DNA sequences of FeCGTa, FeCGTb and two variants obtained by cDNA screening.
The sequences that would be derived from *FeCGTa* and *FeCGTb* are highlighted by blue and orange boxes, respectively. Asterisks represent the identical nucleotides among four sequences.The accession nos. are as follows: FeCGTa, AB909375; FeCGTb, AB909376; FeCGT_variant1, AB909389; FeCGT_variant2, AB909390.