C-Glycosyltransferases catalyzing the formation of di-*C*-glucosyl flavonoids in citrus plants

Takamitsu Ito, Shunsuke Fujimoto, Fumiaki Suito[†], Makoto Shimosaka and Goro Taguchi[*]

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

***For correspondence:** Goro Taguchi, Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, 3-15-1 Tokida, Ueda 386-8567, Japan. Tel: +81-268-215342, Fax: +81-268-215331, e-mail: gtagtag@shinshu-u.ac.jp

E-mail addresses; Takamitsu Ito, field112233445566778899blood@yahoo.co.jp; Shunsuke Fujimoto, sf2044shinshu@gmail.com; Fumiaki Suito, suito@cb.k.u-tokyo.ac.jp; Makoto Shimosaka, mashimo@shinshu-u.ac.jp.

[†]Present address: Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Running title: di-*C*-glycosyltransferases from citrus plants

Keywords: di-*C*-glucosylation; flavonoid *C*-glycosyltransferase; *Fortunella crassifolia*; *Citrus unshiu*; phloretin di-*C*-glucoside; vicenin-2.

GenBank/EMBL/DDBJ accession numbers:

FcCGT, LC131333; CuCGT, LC131334; ChCGT, LC131335.

Word count: 6,459 (Summary: 217; Significance Statement: 70;

Introduction: 789; Results: 2,113; Discussion: 672; Experimental Procedures: 1,783;

Acknowledgements: 68; Table titles: 84; Figure legends: 733; References, 1,052).

SUMMARY

Citrus plants accumulate many kinds of flavonoids, including di-*C*-glucosyl flavonoids, which have attracted considerable attention owing to their health benefits. However, biosynthesis of di-*C*-glucosyl flavonoids has not been elucidated at the molecular level. Here, we identified the *C*-glycosyltransferases (CGTs) FcCGT (UGT708G1) and CuCGT (UGT708G2) as the primary enzymes involved in the biosynthesis of di-*C*-glucosyl flavonoids in the citrus plants kumquat (*Fortunella crassifolia*) and satsuma mandarin (*Citrus unshiu*), respectively. The amino acid sequences of these CGTs were 98% identical, indicating that CGT genes are highly conserved in the citrus family. The recombinant enzymes FcCGT and CuCGT utilized 2-hydroxyflavanones, dihydrochalcone, and their mono-*C*-glucosides as sugar acceptors and produced corresponding di-*C*-glucosides. The K_m and k_{cat} values of FcCGT toward phloretin were <0.5 μM and 12.0 s^{-1}, and those toward nothofagin (3′-*C*-glucosylphloretin) were 14.4 μM and 5.3 s^{-1}, respectively; these values are comparable to those of other glycosyltransferases reported to date. Transcripts of both CGT genes were found to concentrate in various plant organs, and particularly in leaves. Our results suggest that di-*C*-glucosyl flavonoid biosynthesis proceeds via a single enzyme using either 2-hydroxyflavanones or phloretin as a substrate in citrus plants. In addition, *Escherichia coli* cells expressing CGT genes were found to be capable of producing di-*C*-glucosyl flavonoids, which is promising for commercial production of these valuable compounds.

SIGNIFICANCE STATEMENT

Although di-*C*-glucosyl flavonoids have attracted a considerable amount of attention owing to their health benefits, the plant glycosyltransferase enzymes responsible for their biosynthesis are not yet known. We identified the *C*-glycosyltransferases responsible for di-*C*-glucosyl flavonoid biosynthesis in the citrus plants kumquat and satsuma mandarin and demonstrated that a single enzyme catalyzes both first and second *C*-glucosylation of flavonoids using either 2-hydroxyflavanones or phloretin as substrates in these plants.

INTRODUCTION

Flavonoids are phytochemicals that are widely distributed in plants, with more than 8,000 flavonoids, including their derivatives, identified thus far (Pietta, 2000). Found in various parts of plant organs as pigments or signal compounds, flavonoids serve several functions; for example, they are involved in UV-B protection, plant–microbe interactions, pollinator guidance, male fertilization, and antifungal and antimicrobial activities (Shirley, 1996). In addition, the antioxidant activities of these compounds have attracted considerable attention due to the benefits they have on human health (Pietta, 2000).

Flavonoids are often modified by glycosylation, resulting in changes to their physiological properties such as solubility and stability, and therefore glycosylation is thought to play a role in the modification of biological activities and accumulation of flavonoids in plants (Xiao *et al*., 2014). Most glycosylation reactions are catalyzed by uridine diphosphate (UDP)-sugar dependent glycosyltransferases (UGTs), which utilize UDP sugars as sugar donors. Polypeptides of UGTs possess a conserved region for UDP-sugar binding in their C-terminus and are classified into family 1 glycosyltransferases (Paquette *et al*., 2003; Osmani *et al*., 2009; Wang, 2009).

Most flavonoid glycosides are *O*-glycosides, in which the sugar moiety is linked to the flavonoid skeleton by oxygen. In *C*-glycosides, however, the anomeric carbon of the sugar moiety is directly bound to the aromatic ring carbon of flavonoids through a C–C bond (Franz and Grün, 1983; Hultin, 2005); these compounds occur in many plants, including species of bryophytes, pteridophytes, gymnosperms, and angiosperms (Jay *et al.*, 2005; Talhi and Silva, 2012). In addition, di-*C*-glycosides of flavonoids, which have two sugar moieties connected though *C*-glycosidic bonds, as

well as mono-*C*-glycosides, have been found in many plants (Jay *et al*., 2005). The sugar moieties of these *C*-glycosides are usually glucose, but other sugars, such as arabinose, xylose, and rhamnose, are also present (Jay *et al*., 2005). Since *C*-glycosidic bonds are resistant to acid hydrolysis and glycosidase cleavage, *C*-glycosides often exhibit different properties from other glycosides (Talhi and Silva, 2012, Xiao *et al.*, 2016). *C*-Glycosylflavonoids are reported to encompass a wide range of compounds that have bioactive functions *in planta*, such as the antifungal agent phytoalexin (McNally *et al.*, 2003), as well as allelochemicals that act as seed germination inhibitors (Hooper *et al.*, 2010), a component of oviposition stimulants in swallowtail butterfly (Ohsugi *et al*., 1985), and those that confer antibiosis to the maize earworm (Byrne *et al.*, 1996). They are also often found in flowers as co-pigments conferring flower color (Jay, 1994). In addition, *C*-glycosylflavonoids are known to have medicinal properties, including antioxidant, antitumor, hepatoprotective, antidiabetic, and anti-inflammatory activities, indicating the potential health benefits of these compounds (reviewed by Xiao *et al.*, 2016). Of these, di-*C*-glycosylflavonoids are of particular interest because they are readily absorbed in the intestine and are delivered to several organs with no change to their structures (Xiao *et al*., 2016).

The biosynthesis of *C*-glycosylflavonoids is catalyzed by *C*-glycosyltransferases (CGTs). CGTs have been reported from rice (OsCGT), buckwheat (FeCGTa and FeCGTb), and soybean (UGT708D1), where they utilize open-circular forms of 2-hydroxyflavanones as substrates and produce *C*-glucosylflavones through successive dehydration (Brazier-Hicks *et al.*, 2009, Nagatomo *et al.*, 2014, Hirade *et al.*, 2015). In contrast, CGT from maize (UGT708A6) is considered to utilize close-circular forms of 2-hydroxyflavanones as substrates

(Falcone Ferreyra *et al*., 2014), and CGT from mango (MiCGT) catalyzes the glucosylation of xanthones (Chen *et al*., 2015). Recently, another type of CGT was found from gentian (GtUF6CGT) that utilizes flavones as substrates to produce *C*-glucosylflavones directly (Sasaki *et al*., 2015). However, genes coding for enzymes catalyzing the formation of di-*C*-glycosides from mono-*C*-glycosides have not yet been identified.

Citrus fruits are cultivated worldwide (Khan and Kender, 2007), and most commonly used as table fruits and for juice production. Global net production of citrus fruits in 2014 amounted to 1.38×10^8 t (FAOSTAT, http://www.fao.org/faostat/en/#data/QC). Citrus plants contain a large amount of *C*-glycosylflavonoids in their fruits and leaves, including the mono-*C*-glucosides vitexin, isovitexin, orientin, and isoorientin, and the di-*C*-glucosides 6,8-di-*C*-glucosylapigenin (vicenin-2), 6,8-di-*C*-glucosyl luteolin (lucenin-2), 3′,5′-di-*C*-glucosylphloretin, 3,6-di-*C*-glucosylapigenin, and 3,8-di-*C*-glucosylapigenin (Kumamoto *et al*., 1985; Manthey *et al*., 2000; Ogawa *et al*., 2001; Jay *et al*., 2005; Gattuso *et al*., 2007; Talhi and Silva, 2012; Barreca *et al*., 2016). Thus, we hypothesized that citrus plants also possess CGTs that form di-*C*-glycosides using mono-*C*-glycosylflavonoids as sugar acceptors.

To examine this possibility, we identified CGTs from two citrus plants common in Japan, kumquat (*Fortunella crassifolia* Swingle*,* syn. *Citrus japonica* Thunb. subf. *crassifolia* [Swingle] Hiröe) and satsuma mandarin (*Citrus unshiu* Marcow. var. *praecox* Tanaka ‘Miyagawa wase’), both of which are known to accumulate *C*-glycosylflavonoids in various organs (Ohsugi *et al*., 1985; Ogawa *et al*., 2001). Here, we identify and characterize the recombinant CGT enzymes that produce

131 di-*C*-glucosylflavones from their aglycons, and then summarize the biosynthesis of

132 di-*C*-glucosyl flavonoids in citrus plants.

RESULTS

Flavonoids of *F. crassifolia* and *C. unshiu*

Citrus plants are known to contain several flavonoid *C*-glycosides, including di-*C*-glycosides, in their fruits and leaves (Manthey *et al*., 2000; Gattuso *et al*., 2007). We used HPLC-MS to analyze methanol extracts of kumquat and satsuma mandarin leaves to identify possible *C*-glycosides (Figure S1); methanol extracts of kumquat leaves contained a compound with a major peak at a retention time of 2.56 min (peak 2), the ultraviolet absorption spectrum of which was similar to that of phloretin (Figure S1a). The mass spectrum of this compound showed the presence of an $[M-H]^-$ ion at 597, which corresponded to that of diglucoside of phloretin (3′,5′-di-*C*-glucosylphloretin, Mw = 598.6), a compound previously reported to be a major component of kumquat leaves (Ogawa *et al.*, 2001). For further confirmation, the extract was treated with 2 M HCl at 60 °C for 3 h to hydrolyze *O*-glycosides to their aglycons, whereas *C*-glycosides were stable for this treatment. The peak did not change as a result of this treatment (Figure S1b), suggesting that the compound was *C*-glucoside; it was eventually identified as 3′,5′-di-*C*-glucosylphloretin through comparisons with a synthesized compound in this work (Figure S1e). In addition to 3′,5′-di-*C*-glucosylphloretin, we found a second acid-tolerant compound in extracts of kumquat leaves at a retention time of 2.06 min (peak 1), the ultraviolet absorption spectrum of which was similar to that of apigenin (Figure S1a, b). The mass spectrum of this compound exhibited an $[M-H]^-$ ion at 593, which corresponded to di-*C*-glucosyl-apigenin (Mw = 594.5); the compound was later identified as vicenin-2 by comparing it with an authentic compound (Figure S1d). Mono-*C*-glucosides of apigenin (vitexin and isovitexin) were not detected in kumquat. Methanol extracts of

157 satsuma mandarin leaves also contained vicenin-2 (Figure S1c). Our results revealed the
158 presence of a flavonoid di-*C*-glucoside in both types of plants, thus corroborating the
159 results of other studies (Ohsugi *et al*., 1985; Ogawa *et al*., 2001). To examine whether
160 CGT activities are high enough in citrus plants to produce these *C*-glucosides, crude
161 enzyme preparations of kumquat and satsuma mandarin leaves were examined for
162 enzymatic activity using 2-hydroxynaringenin as a substrate (Figure S2). Both types of
163 plants showed *C*-glucosylation activities; notably, the reactions involving kumquat
164 extracts exhibited evidence of di-*C*-glucoside formation, suggesting that these citrus
165 plants were suitable for the study of CGTs responsible for the formation of flavonoid
166 mono- and/or di-*C*-glucosides produced by *C*-glucosylation reaction, as shown in
167 Figure 1.
168
169 **Isolation of candidate CGT genes from citrus plants**
170 A search of the EST database of citrus plants in TIGR (conducted in September 2012, at
171 http://www.jcvi.org) using a buckwheat CGT gene (*FeCGTa*) as a query produced
172 several CGT candidates in citrus plants, including TA5377_37690 (*Poncirus trifoliata*),
173 TA14910_2711 (*Citrus sinensis*), TA4403_85681 (*C. clementine*), and TA829_55188
174 (*C. unshiu*). The deduced amino acid sequences had high degrees of similarity, with
175 greater than 95% sequence identity (Figure S3), indicating that CGT sequences in citrus
176 plants were highly conserved beyond the genus level. We designed a set of primers
177 based on the conserved sequences corresponding to the N-terminal and C-terminal
178 regions and performed PCR using cDNAs synthesized from the leaves of kumquat,
179 satsuma mandarin, and hanaju (*Citrus hanaju* Siebold ex Shirai). The amplified
180 fragments were cloned into a pCR4-TOPO vector and their sequences were determined.

The full-length cDNAs of CGTs from kumquat, satsuma mandarin, and hanaju were designated as *FcCGT*, *CuCGT*, and *ChCGT*, respectively.

FcCGT (1,419 bp), *CuCGT* (1,419 bp), and *ChCGT* (1,419 bp) coded for the proteins FcCGT (UGT708G1), CuCGT (UGT708G2), and ChCGT (UGT708G3), each of which had a molecular mass of 51.6 kDa. As expected, the deduced amino acid sequences of the three CGTs shared a high level of identity (>98%), and had 48% and 43–44% identity to sequences of OsCGT and FeCGTa, respectively (Figure S4). In addition, the nucleotide sequence of *CuCGT* shared 99% identity with an UGT708-like sequence (NC_023052) in the genomic sequence of *C. sinensis* that was recently released in GenBank. The citrus CGTs also contained the PSPG motif required for UDP-sugar binding (Osmani *et al*., 2009), as well as His_{23} and Asp_{117}, which are required for the nucleophilic reaction toward sugar acceptor conserved in UGTs forming the acceptor-His-Asp triad (Wang, 2009).

Properties of recombinant citrus CGTs expressed in *Escherichia coli*

To confirm enzymatic properties of the three citrus CGTs (FcCGT, CuCGT, and ChCGT), their full-length cDNAs were cloned into a pET28a(+) vector and introduced into *Escherichia coli* Rosetta 2(DE3). The expressed recombinant proteins had a fused tag of 6× histidine and a molecular weight of 53.7 kDa when analyzed with SDS-PAGE. All of the resulting recombinant proteins (FcCGT, CuCGT, and ChCGT) displayed *C*-glucosylation activity against 2-hydroxynaringenin (Figure 2, Figure S5), indicating that they are CGTs derived from citrus plants.

The typical reaction pattern catalyzed by FcCGT is shown in Figure 2. A product peak (peak 6) was observed for a short period when phloretin (peak 7) was used

as the substrate (Figure 2a), and was coincident with nothofagin (Figure 2c, peak N). Nothofagin levels decreased as the reaction continued, whereas levels of the other product exhibiting the $[M-H]^-$ ions at *m/z* 597 (peak 2) increased. This compound was coincident with synthesized 3′,5′-di-*C*-glucosylphloretin (Figure 2c, peak Dp). The retention time of it was different from that observed in Figure S1, because we used different HPLC condition to separate nothofagin and 3′,5′-di-*C*-glucosylphloretin clearly. Moreover, FcCGT also catalyzed the *C*-glucosylation of 2-hydroxynaringenin (Figure 2b). Dehydrated derivatives of the products were detected in HPLC chromatographs constructed after the reactions were stopped by the addition of HCl and further incubation at 60 ºC for 30 min to complete dehydration. Two peaks, peaks 3 and 4, were observed, which corresponded to the compounds vitexin and isovitexin, respectively (Figure 2d, peak Vt and Iv). As the reaction continued, levels of vitexin and isovitexin decreased, whereas the level of the other product exhibiting the $[M-H]^-$ ions at *m/z* 593 increased (Figure 2b, peak1). This peak corresponded to the compound vicenin-2 (Figure 2d, peak Vc). CuCGT and ChCGT also produced vicenin-2 following reaction with 2-hydroxynaringenin and after treatment with HCl (Figure S5a, b). These results clearly indicated that citrus CGTs alone catalyze the first *C*-glucosylation and the successive second *C*-glucosylation of flavonoids.

The enzymatic properties of FcCGT and CuCGT were examined further. The pH preference of FcCGT and CuCGT were determined using 2-hydroxynaringenin as a substrate, with both enzymes exhibiting greater than 65% of the maximum level of activity at pHs of 8.0–11.0 (Figure S6). The optimum reaction temperatures of FcCGT and CuCGT differed somewhat, although their amino acid sequences shared 98% sequence identity. Maximum FcCGT activity occurred at 50 °C, with significant

229 activity maintained up to 60 °C, whereas maximum CuCGT activity occurred at 45 °C,
230 and it was deactivated at 55 °C (Figure S6).
231 Substrate specificities of FcCGT and CuCGT were examined using several
232 flavonoid-related compounds as substrates (Figure S7), with UDP-glucose as a sugar
233 donor (Table 1). Both enzymes displayed evidence of *C*-glucosylation activity toward
234 2-hydroxyflavanones, such as 2-hydroxynaringenin and 2-hydroxypinocembrin, which
235 are in equilibrium with their open-circular form, or dibenzoylmethane form. They also
236 reacted with flavonoids having 2′,4′,6′-trihydroxyacetophenone-like structures such as
237 dihydrochalcone (phloretin); 2-phenyl-2′,4′,6′-trihydroxyacetophenone;
238 2′,4′,6′-trihydroxyacetophenone; 2,4,6-trihydroxybenzaldehyde; and maclurin. In
239 contrast, they did not catalyze the *C*-glucosylation of other compounds, including
240 naringenin, apigenin, quercetin, isoliquiritigenin, propyl gallate, and
241 2′,4′-dihydroxyacetophenone, suggesting that 2′,4′,6′-trihydroxyacetophenone-like
242 structures are necessary for substrates to be acceptable to these enzymes.
243 Both enzymes showed strong activity toward nothofagin and
244 2-hydroxynaringenin *C*-glucoside, resulting in a formation of corresponding
245 di-*C*-glucosides (Table 1), but both also exhibited less potent activity toward phloretin
246 2′-*O*-glucoside (phloridzin), with activity levels 0.1% of that for nothofagin. These
247 enzymes produced mono-*C*-glucosides but not di-*C*-glucosides when
248 2′,4′,6′-trihydroxyacetophenone, 2,4,6-trihydroxybenzaldehyde, and maclurin were used
249 as substrates. Thus, FcCGT and CuCGT appear to have the same catalytic activities.
250 Sugar-donor specificities of FcCGT and CuCGT were also examined using
251 2-hydroxynaringenin as a sugar acceptor. Both of the enzymes primarily utilized
252 UDP-glucose as the sugar-donor, but also used UDP-xylose as well; little activity with

253 UDP-galactose and no activity with UDP-glucuronic acid were observed. This pattern
254 of sugar-donor specificity in utilizing UDP-xylose was similar to the sugar-donor
255 specificities of CGTs deriving from buckwheat (Nagatomo *et al*., 2014).
256 Kinetic parameters of FcCGT reaction were calculated using phloretin,
257 2-hydroxynaringenin, and their mono-*C*-glucosides as substrates (Table 2). Since the
258 initial reaction rate toward phloretin was high and saturated even at substrate
259 concentrations of 0.5 μM, we were unable to calculate the exact K_m value; however, the
260 K_m and k_{cat} values toward 2-hydroxynaringenin were 0.85 μM and 7.6 s^{-1}, respectively.
261 The enzyme also showed adequate affinity to nothofagin and
262 6-*C*-glucosyl-2-hydroxynaringenin, with K_m and k_{cat} values of 14.4 μM and 5.3 s^{-1}, and
263 112.5 μM and 5.7 s^{-1}, respectively. The K_m and k_{cat} values for both substrates were
264 comparable to those reported for other UGTs, such as FeCGTa and OsCGT, whose
265 values toward 2-hydroxynaringenin were 4.4 μM and 2.5 μM for K_m, and 3.2 s^{-1} and 3.1
266 s^{-1} for k_{cat}, respectively (Brazier-Hicks *et al*., 2009; Nagatomo *et al*., 2014). These
267 results suggested that FcCGT catalyzes both mono-*C*-glucoside formation and
268 di-*C*-glucoside formation of flavonoids with high affinity.
269
270 **Expression of citrus CGTs in plant organs**
271 To investigate the role of FcCGT and CuCGT *in planta,* total RNAs were extracted
272 from several organs of kumquat and satsuma mandarin (flower, immature fruit, peel of
273 mature fruit, young leaves, and mature leaves), and analyzed using quantitative reverse
274 transcription (qRT)-PCR (Figure 3). Expression of these CGTs was detected in all of
275 the plant organs included in the analyses, but particularly in the leaves of both plants
276 and in the flowers of kumquat. Flavonoid di-*C*-glucoside (i.e.,

3′,5′-di-*C*-glucosylphloretin and vicenin-2) amounts were also estimated in each organ (Figure 3), with large amounts of 3′,5′-di-*C*-glucosylphloretin found in kumquat, especially in the leaves, with lower amounts in immature fruit and flowers, whereas this compound was not detected in satsuma mandarin. Both plants accumulated vicenin-2, mainly in their leaves, although at far lower levels than that of 3′,5′-di-*C*-glucosylphloretin in kumquat. The accumulation patterns of these flavonoids were similar to those of the *FcCGT* or *CuCGT* transcripts in these plants. Thus, the evidence suggested that FcCGT and CuCGT play roles in the biosynthesis of flavonoid-*C*-glucosides.

Bioconversion of 2-hydroxynaringenin and phloretin into their di-*C*-glucoside by *E. coli* expression system

Bioconversion of substrates into their *C*-glucosides was examined using *E. coli* cells that express CuCGT, with the compounds having 2′,4′,6′-trihydroxyacetophenone-like structures used as the substrates. Substrates added to the culture media were mostly converted into their *C*-glucosides, which were recovered from the media after 2 h (Figures 4 and S8). When 2-hydroxynaringenin (Figures 4a, peak 10) was administered to the medium, it was converted into di-*C*-glucoside (peak 8) and mono-*C*-glucoside (peak 9) as main products, and their dehydrated compounds (vicenin-2 [peak 1], vitexin [peak 3], and isovitexin [peak 4]) were detected in trace amounts (Figure 4b). After treatment with HCl, these compounds were completely dehydrated and converted into apigenin-*C*-glucosides (vicenin-2, vitexin, and isovitexin; Figure 4c). In the case of phloretin (Figures 4d–f), di-*C*-glucosylphloretin (peak 2) and nothofagin (peak 6) were detected as main products. We purified the former product (peak 2) and used NMR for

301 analysis of the product, determining that the NMR data corresponded to that reported
302 for 3′,5′-di-*C*-glucosylphloretin (Ogawa *et al*., 2001), thus confirming that the product
303 was 3′,5′-di-*C*-glucosylphloretin. In these conversions, quantities of the di-*C*-glucosides
304 exceeded 50% of their initial amounts, suggesting that this system would be useful for
305 the production of di-*C*-glucosides. These conversions were also available for other
306 compounds that can be recognized as substrates by the enzymes, and can be used for *E.*
307 *coli* cells expressing FcCGT (Figure S8). Syntheses of flavonoid *C*-glucosides using
308 plant CGTs have been elucidated (Brazier-Hicks and Edwards, 2013; Bungaruang *et al*.,
309 2013; Ito *et al*., 2014), but the process of di-*C*-glucosides synthesis has yet to be
310 determined. Thus, our results may represent a cost-effective approach to the production
311 of flavonoid di-*C*-glucosides.
312
313 **Molecular phylogenetic analysis of citrus CGTs**
314 A molecular phylogenetic tree was constructed based on the deduced amino acid
315 sequences of citrus CGTs, other reported CGTs, and several types of UGTs; these are
316 summarized in Table S1 (Figure 5). Plant CGTs utilizing
317 2′,4′,6′-trihydroxyacetophenone-like structures formed a clade (UGT708), suggesting
318 that these CGTs have evolved from the same ancestral gene as proposed previously
319 (Yonekura-Sakakibara and Hanada, 2011; Nagatomo *et al*., 2014). Those UGTs
320 displaying CGT activities with different substrate specificities, such as gentian CGT
321 (GtUF6CGT) showing 6-*C*-glucosylation activity to flavone substrates (Sasaki *et al.*,
322 2015) and *Arabidopsis* UGT73B4 showing *O*- and *C*-glucosylation activities toward
323 xenobiotics (Gandia-Herrero *et al*., 2008), were resolved in different clades—an

indication that these enzymes evolved independently and in parallel, and derived from different ancestral genes that had diverged from OGTs (Sasaki *et al.*, 2015).

DISCUSSION

Information regarding the biosynthesis of *C*-glycosides of flavonoids and their related compounds is limited due to insufficient CGT identification, despite several CGTs having been identified from rice, maize, buckwheat, soybean, mango, and gentian over the past several years (Brazier-Hicks *et al*., 2009; Falcone Ferreyra *et al*., 2014; Nagatomo *et al*., 2014; Chen *et al*., 2015; Hirade *et al*., 2015; Sasaki *et al*., 2015). Available information regarding CGTs is especially inadequate considering the diversity of these *C*-glycosides, particularly those CGTs involved in di-*C*-glycoside formation. In this study, we isolated *CGT* genes from citrus plants and characterized their gene products (*C*-glycosyltransferase enzymes), including those involved in the mono-*C*-glucosylation of several aglycons and further *C*-glucosylation of mono-*C*-glucosyl flavonoids.

Biosynthesis of di-*C*-glycosylflavonoids has been examined in a few studies. In rice, candidate genes responsible for the accumulation of di-*C*-arabinosylapigenin have been narrowed down by metabolome–genome-wide association studies (Matsuda *et al*., 2015), so that enzymatic characterization is expected to elucidate *C*-arabinosylation. Leaves of *Desmodium* are known to accumulate di-*C*-glycosides of apigenin, and their cell-free extracts showed both *C*-glucosylation activities to 2-hydroxynaringenin and subsequent second glycosylation of *C*-glucosyl-2-hydroxynaringenin. Thus, the first and the second *C*-glycosylation reactions are catalyzed by different enzymes, as indicated by the use of different UDP-sugars, such as UDP-glucose, UDP-arabinose, or UDP-galactose (Hamilton *et al*., 2012, Hao *et al*., 2015). In contrast, our results clearly showed that a single enzyme can catalyze both the first and second *C*-glycosylation reactions in the biosynthesis of

3′,5′-di-*C*-glucosylphloretin and vicenin-2 in citrus plants (Figures 1, 2, and S5). These enzymes showed strong activities toward the aglycons, such as phloretin and 2-hydroxynaringenin, and toward flavonoid mono-*C*-glucosides, such as nothofagin and 6-*C*-glucosyl-2-hydroxynaringenin (Table 1). The k_{cat}/K_m values for FcCGT suggested that the reaction rate of the first *C*-glycosylation is much faster than that of the second *C*-glycosylation (Table 2). However, the parameters of the second *C*-glycosylation were also comparable to other reported UGT enzymes and sufficient to catalyze these reactions *in vivo* (Tables 1 and 2). That FcCGT can catalyze both first and second *C*-glucosylation of flavonoids would be consistent with the fact that kumquat accumulates mostly di-*C*-glucosides but not mono-*C*-glucosides of phloretin and apigenin (Figures 3 and S1).

Both FcCGT and CuCGT exhibited slight activity with phloridzin, a 2′-*O*-glucoside of phloretin, although this amounted to only 0.1% of the activity with nothofagin (Table 1), and both reacted with 6-*C*-glucosyl-2-hydroxynaringenin but not with vitexin or isovitexin. These results suggested that, similar to the first *C*-glycosylation step, the open-circular form of 2-hydroxyflavanone is required for substrate recognition in the second *C*-glycosylation. These enzymes showed levels of activity with 2′,4′,6′-trihydroxyacetophenone, 2,4,6-trihydroxybenzaldehyde, and maclurin, but did not produce their di-*C*-glucosides, suggesting high substrate specificity at the di-*C*-glucosylation step but less so during the mono-*C*-glucosylation in terms of the distance between the two aromatic rings. In addition, neither enzyme produced 3-*C*-glucosides of flavonoids like those found in citrus plants (Jay *et al*., 2005), indicating the existence of another enzyme or enzymes catalyzing 3-*C*-glucosylation of flavonoids.

374 The reaction mechanism of bacterial CGT has been intensely studied,
375 especially that of UrdGT2 from *Streptomyces fradiae*, in which a conversion of related
376 OGT to CGT is achieved via substitution of several amino acid residues involved in
377 substrate recognition, as inferred from its crystal structure (Härle *et al*., 2011). In
378 contrast, research on the reaction mechanism of plant CGTs is lacking, mainly due to
379 the low number of plant CGTs that have been identified and characterized. In general,
380 the amino acid residues at the sugar-acceptor binding pockets of UGTs are not highly
381 conserved (Osmani *et al*., 2009). Hirade *et al*. (2015) concluded from analysis of
382 homology modeling and the substitution of these residues into alanine that the two
383 amino acid residues of UGT708D1 located in the active site, Asp_{85} and Arg_{292}, are
384 important for *C*-glucosylation activity. However, the molecular bases of the CGT
385 reaction mechanism remain unclear. We were unable to determine the structural
386 differences between the CGTs catalyzing mono-*C*-glucosylation alone and those
387 catalyzing mono- and di-*C*-glucosylation. Further studies are warranted to elucidate in
388 detail the CGT reactions.

EXPERIMENTAL PROCEDURES

Plant materials

Kumquat (*Fortunella crassifolia*), Satsuma mandarin (*Citrus unshiu* var. *praecox* 'Miyagawa wase'), and hanaju (*Citrus hanaju*) were used in this study. Leaves, flowers, and fruits were collected from potted plants grown under natural conditions, immediately frozen with liquid nitrogen, and stored at −80 °C until use.

Reagents

The following substrates were used: apigenin and luteolin (Indofine Chemical Company, http://indofinechemical.com/); chrysin (Across Organics, http://www.acros.com/); quercetin, UDP-glucose, and UDP-glucuronic acid (Nacalai Tesque, http://www.nacalai.co.jp/global/index.html); vicenin-2, naringenin, maclurin, 2-phenyl-2′,4′,6′-trihydroxyacetophenone, UDP-xylose, and UDP-galactose (Sigma-Aldrich https://www.sigmaaldrich.com/); phloretin, phloridzin, 2′,4′-dihydroxyacetophenone, 2′,4′,6′-trihydroxyacetophenone, and isoliquiritigenin (Tokyo Chemical Industries, http://www.tcichemicals.com/en/ap/); vitexin and isovitexin (Extrasynthèse, http://www.extrasynthese.com/); 2,4,6-trihydroxybenzaldehyde (Alfa Aesar, https://www.alfa.com/en/); and propyl gallate (Wako Pure Chemical Industries, http://www.wako-chem.co.jp/english/). Unless otherwise specified, all other chemicals were obtained from Sigma-Aldrich, Nacalai Tesque, Wako Pure Chemical Industries, and Kanto Kagaku (http://www.kanto.co.jp/en/). 2-Hydroxypinocembrin and 2-hydroxynaringenin were synthesized as previously described by Nagatomo *et al*. (2014). 6-*C*-Glucosyl-2-hydroxynaringenin, 6-*C*-glucosyl-2-hydroxypinocembrin, 3′-*C*-glucosyl-2-phenyl-

413 2′,4′,6′-trihydroxyacetophenone, and nothofagin were synthesized by a
414 biotransformation using *E. coli* expressing FeCGTa enzyme as previously described by
415 Ito *et al*. (2014).
416
417 **Analysis of phenolic compounds in plants**
418 Frozen samples (100 mg fresh weight) of kumquat and satsuma mandarin ground with
419 mortar and pestle were extracted with methanol (1 mL) overnight at 4 °C. The resulting
420 extracts were centrifuged at 17,000 × *g* for 10 min to remove cellular debris. Aliquots of
421 the samples were treated with 2 M HCl at 60 °C for 3 h to hydrolyze *O*-glycosidic
422 bonds as described previously (Nagatomo *et al.*, 2014). Both acid-treated and
423 non-treated samples were filtrated through a 0.2-μm polytetrafluoroethylene filter
424 (Merck Millipore, http://www.merckmillipore.com/) and analyzed with HPLC.
425
426 **HPLC conditions**
427 HPLC-MS was performed with a Waters UPLC ACQUITY SQD system (Waters,
428 http://www.waters.com/) equipped with an ODS column (2.1 mm i.d. × 50 mm, 1.9 μm;
429 J-Pak UPX Supero C_{18}, JASCO, https://www.jasco.co.jp/), an electron-splay ionization
430 probe (negative mode), and a diode array detector, as described previously by
431 Nagatomo *et al*. (2014). Briefly, the column was eluted with 20% solvent B (methanol
432 supplemented with 0.1% formic acid) and 80% solvent A (0.1% formic acid) for 0.5
433 min, followed by a gradient from 20% to 60% solvent B for 2 min, 60% solvent B and
434 40% solvent A for 2 min, and finally 20% solvent B and 80% solvent A for 2 min at a
435 flow rate of 0.25 mL min^{-1} and a temperature of 40 °C. To separate the reaction
436 products of phloretin, the column was eluted with a gradient from 20% to 30% solvent

437 B for 1 min, 40% solvent B and 60% solvent A for 0.5 min, a gradient from 40% to
438 60% solvent B for 1 min, 60% solvent B and 40% solvent A for 2 min, and then 20%
439 solvent B in 80% solvent A for 2 min. Methanolic extracts of citrus plants were also
440 analyzed with an LC10Avp system (Shimadzu, http://www.shimadzu.co.jp/) equipped
441 with an ODS column (4.6 mm i.d. × 150 mm; KINETEX C18, Phenomenex,
442 http://www.phenomenex.com/). The column was eluted with 25% solvent B in solvent
443 A for 2 min, followed by a linear gradient of 25% to 55% solvent B in solvent A for 18
444 min, then by 70% solvent B and 30% solvent A for 4 min at a flow rate of 1 ml min^{-1}
445 and a temperature of 40 °C. The eluate was monitored at 320 nm using a diode array
446 detector (SPD-M10Avp, Shimadzu).
447
448 **Detection of CGT enzyme activities in leaves**
449 Frozen leaves (0.3 g fresh weight) of both kumquat and satsuma mandarin were
450 separately ground with a mortar and pestle under liquid nitrogen and extracted with 0.6
451 mL of ice-cold buffer A (50 mM Tris-HCl, pH 8.0, 1 mM dithiothreitol) containing 5%
452 (w/v) polyvinylpolypyrrolidone. The mixture was sonicated for 20 cycles of 3 s each at
453 an amplitude of 24% (Vibra Cell VCX500, Sonic & Materials, Inc.,
454 http://www.sonics.com/lp-vibra.htm) on ice and centrifuged at 17,000 × *g* for 15 min.
455 The supernatant was desalted using a NAP-5 desalting column (GE Healthcare Japan,
456 http://www3.gehealthcare.co.jp) equilibrated with buffer A. The desalted supernatant
457 was concentrated using Amicon Ultra-15 Ultracel-10k (Merck Millipore) and the
458 concentrate was utilized as crude enzymes. *C*-Glucosylation activity was assayed in a
459 reaction mixture (100 μL) composed of crude enzymes, 100 μM phloretin or
460 2-hydroxynaringenin as glucose acceptor, and 2 mM UDP-glucose as glucose donor in

461 buffer A. The reaction was initiated by adding the substrate, then incubated at 30 °C for
462 2 h, and stopped by adding 20 μL of 1M HCl and 100 μL of methanol.
463
464 **Cloning and sequencing of CGT genes**
465 Candidate genes coding for CGTs were determined by searching the former EST
466 database of citrus plants in the TIGR Plant Transcript Assemblies
467 (http://plantta.jcvi.org/index.shtml, conducted in September 2012) using the FeCGTa
468 amino acid sequence as the query sequence. PCR primers were constructed from the
469 candidate genes (Figure S3) along with the addition of restriction enzyme cutting sites,
470 namely, citrusCGT-Fw_Nde 5′- GTCATATGTCAGATTCCGGCGGCTTTG-3′ and
471 citrusCGT-Rv_Xho 5′- GACTCGAGTTAATGGGTGTTGTTGTTGCAC-3′. Total
472 RNAs of kumquat, satsuma mandarin, and hanaju were extracted from frozen leaves by
473 grinding leaves with a mortar and pestle under liquid nitrogen using a Sepasol RNA I
474 SuperG (Nacalai Tesque) in accordance with the manufacturer's instructions, and then
475 treating the resultant RNA solutions with DNase I (Takara Bio, Inc.,
476 http://www.takara-bio.com/). First-strand cDNAs were synthesized from the total RNAs
477 (500 ng) using a ReverTra Ace (TOYOBO,
478 http://www.toyobo-global.com/seihin/xr/lifescience/) and a dT-T3 primer
479 (5′-ATTAACCCTCACTAAAGGGTTTTTTTTTTTTTTTVV-3′, 100 pmol) at 42 °C
480 for 60 min; mixtures were diluted by adding 180 μL of TE buffer. The PCR was
481 performed using 2 μL of first-strand cDNA as a template, citrusCGT-Fw_Nde and
482 citrusCGT-Rv_Xho as primers, and iProof Hi-Fidelity DNA polymerase (Bio-Rad,
483 http://www.bio-rad.com/) under the following conditions: 98 °C for 30 s, 40 cycles of
484 98 °C for 10 s, 60 °C for 20 s, and 72 °C for 45 s, followed by 72 °C for 8 min. The

amplified fragments were cloned into pCR4Blunt-TOPO (Thermo Fisher Scientific, http://www.thermofisher.com/) and sequenced using an ABI-PRISM 3130xl Genetic Analyzer (Thermo Fisher Scientific).

Phylogenetic analysis

A phylogenetic tree was created using the neighbor-joining method with 1,000 bootstrap replicates. Construction of the tree was performed with MEGA6 software (http://www.megasoftware.net/) using sequences aligned with ClustalW.

Heterologous expression of CGTs in *E. coli*

Full-length cDNAs of CGTs from citrus plants (kumquat, satsuma mandarin, and hanaju) cloned in pCR4Blunt-TOPO were digested with *Nde*I and *Xho*I and subcloned into pET28a(+) (Merck Millipore) to produce pET-FcCGT, pET-CuCGT, and pET-ChCGT. The recombinant proteins were expressed in *E. coli* Rosetta™ 2(DE3) (Merck Millipore) and extracted in accordance with the manufacturer's instructions, with minor modifications: for an induction of expression of the objective genes, the culture was supplemented with 0.4 mM isopropyl thiogalactoside and incubated at 22 °C for 18 h. The recombinant proteins were purified using a nickel-affinity column (His-GraviTrap, GE Healthcare Japan) and concentrated using Amicon-Ultra-15 Ultracel-10k (Merck Millipore); concentrated proteins were used as purified recombinant enzymes.

Characterization of the recombinant enzymes

Enzyme reactions were performed with reaction mixtures (50 μL) composed of 0.5–500 ng purified enzyme, 200 μM substrate, 2 mM UDP-glucose, and 50 mM potassium phosphate buffer (pH 8.0) containing 5 mM 2-mercaptoethanol and 0.01% BSA. The reaction was initiated by adding the substrate and then incubated at 30 °C for 5–30 min. The reaction was terminated by adding 10 μL 2 M HCl and subjected to HPLC-MS analysis after the addition of methanol (100 μL) and an internal standard (1 mM umbelliferone, 2,4,6-trihydroxybenzaldehyde, chrysin or naringenin, 5 μL). The products were detected by measuring absorbance at 290 nm or 340 nm. To verify the effect of pH on enzyme activity, 100 mM potassium phosphate (pH 5.0–8.5) and 100 mM ethanolamine-HCl (pH 8.0–11.0) were used as a buffer, and 2-hydroxynaringenin was used as a substrate. To determine the optimal temperature, the reactions were performed at 20–65 °C. Substrate preferences of the enzymes were confirmed using 200 μM flavonoids and related compounds listed in Table 1 and Figure S7, and 2 mM UDP-glucose. Preference of sugar donor was confirmed using 500 μM UDP-sugars and 100 μM 2-hydroxynaringenin. Kinetic parameters of the enzyme reaction were obtained by fitting the kinetics data to the Michaelis–Menten equation using Hyper 32 software (http://homepage.ntlworld.com/john.easterby/software.html).

Quantitative RT-PCR analysis

Total RNAs were extracted from flowers, leaves, immature fruit, and fruit peels of kumquat and satsuma mandarin with SepasolG, as described above, and the first-strand cDNAs were synthesized from the total RNAs (200 ng each) using a PrimeScript RT Reagent Kit (Perfect Real Time) (Takara Bio) according to the manufacturer's instructions. The mixtures were each diluted by adding 10 μL of water, and PCR was

532 performed in a Thermal Cycler Dice Real Time System TP800 (Takara Bio) using 2 μL
533 of the first-strand cDNA as a template and a SYBR *Premix Ex Taq II* in accordance
534 with the manufacturer's instructions. *CGT*s fragments were amplified using a primer set
535 of citrus-CGT-735F (5′-TGGACCTCTTTTGCCTTGTG-3′) and citrus-CGT-880R (5′-
536 CCTTCGTTTGCTCCATTGAC-3′). For normalization, glyceraldehyde-3-phosphate
537 dehydrogenase (*gapdh*, XM006476919) was used as a housekeeping gene with a primer
538 set citrus-GAPDH-762F (5′-GTCTTGCCTGCTTTGAATGG-3′) and
539 citrus-GAPDH-862R (5′-GCATCCTTCTCCAGCCTCAC-3′). The transcript level was
540 calculated using Real Time System (Takara Bio) software via the ddCt method based on
541 the 2nd derivative maximum and three biological replicates (average ± SD). Statistical
542 analyses were performed with Tukey's test on MAPHAS software
543 (http://www.gen-info.osaka-u.ac.jp/MEPHAS/tukey-e.html).
544
545 **Bioconversion of flavonoids by *E. coli* expressing citrus plant CGTs**
546 Bioconversion of flavonoids was performed as described previously by Ito *et al*. (2014)
547 using *E. coli* Rosetta 2(DE3) harboring pET-CuCGT or pET-FcCGT. Briefly, *E. coli*
548 cells expressing CGTs were collected by centrifugation at 3,000 × *g* for 10 min and then
549 suspended in M9 basal media containing 2% glucose at a cell density of OD_{600} = 3.0.
550 The suspension was incubated at 30 ºC and shaken at 150 rpm, then added to the
551 substrate at a final concentration of 100 μM.
552
553 **NMR analysis**
554 NMR spectra were recorded on a Bruker Avance 400 spectrometer (Bruker BioSpin,
555 https://www.bruker.com/about-us/offices/offices/bruker-biospin.html) and compared

with the reported NMR spectrum (Ogawa *et al*., 2001). 3′,5′-di-*C*-glucosylphloretin: ^{1}H NMR (400 MHz, DMSO-d6): δ 11.74 (1H, s, Ar-OH), 9.13 (2H, d, J = 4.7 Hz, Ar-OH), 7.03 (2H, d, J = 8.4 Hz, H-2 and H-6), 6.66 (2H, d, J = 8.4 Hz, H-3 and H-5), 3.27 (2H, m, H-α), 2.79 (2H, t, J = 7.6 Hz, H-β); sugar moiety [5.00 (1H, Sugar-OH), 4.96 (1H, Sugar-OH), 4.76 (1H, Sugar-OH), 4.72 (1H, d, J = 9.8 Hz, H-1″), 4.70 (1H, Sugar-OH), 3.62 (2H, m, H-6″), 3.47 (1H, m, H-2″), 3.35–3.23 (3H, m, H-3″–H-5″)] × 2; ^{13}C NMR (100 MHz, DMSO-d6): δ 205.5 (C = O), 163.4 (C-4′), 161.5 (C-2′ and C-6′), 155.8 (C-4), 131.9 (C-1), 129.6 (C-2 and C-6), 115.5 (C-3 and C-5), 105.0 (C-1′), 104.4 (C-3′ and C-5′), 81.5 (C-5″), 78.2 (C-3″), 75.0 (C-1″), 72.4 (C-2″), 69.6 (C-4″), 60.3 (C-6″), 46.6 (C-α), 29.6 (C-β).

ACCESSION NUMBERS

FcCGT, LC131333; CuCGT, LC131334; ChCGT, LC131335.

ACKNOWLEDGEMENTS

We would like to thank the UGT Nomenclature Committee for suggesting the UGT names and Sachiko Yoshioka of Shinshu University for assistance with the NMR analysis. We are indebted to the Division of Gene Research, Research Center for Human and Environmental Sciences, Shinshu University, for providing facilities, and Editage (www.editage.jp) for English-language editing. This work was supported in part by JSPS KAKENHI Grant Number JP26450120 to G.T.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

SUPPORTING INFORMATION

Figure S1. Distribution of *C*-glucosides in kumquat and satsuma mandarin leaves.

Figure S2. Detection of *C*-glucosylation activities in cell-free extracts of kumquat and satsuma mandarin leaves.

Figure S3. Alignment of deduced amino acid sequences of *C*-glycosyltransferases (CGT) candidates (EST) from citrus plants.

Figure S4. Comparison of the amino acid sequences of *C*-glycosyltransferases (CGTs) from citrus and other types of plants.

Figure S5. HPLC analysis of the recombinant CuCGT and ChCGT reaction products from 2-hydroxynaringenin.

Figure S6. Properties of recombinant FcCGT and CuCGT.

Figure S7. Structures of the compounds used for the enzyme assay.

Figure S8. Bioconversion of phenolic compounds using *E. coli* expressing CuCGT or FcCGT.

Table S1. List of UDP-sugar dependent glycosyltransferases (UGTs) used in the phylogenetic analysis.

REFERENCES

Barreca, D., Gattuso, G., Laganà, G., Leuzzi, U. and Bellocco, E. (2016) *C*- and *O*-glycosyl flavonoids in Sanguinello and Tarocco blood orange (*Citrus sinensis* (L.) Osbeck) juice: Identification and influence on antioxidant properties and acetylcholinesterase activity. *Food Chem.* **196**, 619–627.

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., McMullen, M.D., Snook, M.E., Musket, T.A., Theuri, J.M., Widstrom, N.W., Wiseman, B.R. and Coe, E.H. (1996) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc. Natl. Acad. Sci. USA*, **93**, 8820–8825.

Chen, D., Chen, R., Wang, R., Li, J., Xie, K., Bian, C., Sun, L., Zhang, X., Liu, J., Yang, L., Ye, F., Yu, X. and Dai, J. (2015) Probing the catalytic promiscuity of a regio- and Stereospecific *C*-Glycosyltransferase from *Mangifera indica*. *Angew. Chem. Int. Ed. Engl.* **54**, 12678–12682.

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.

621 **FAOSTAT** (Statistical analysis website of Food and Agriculture Organization of the United
622 Nations) http://www.fao.org/faostat/en/#data/QC.
623 **Franz, G. and Grün, M.** (1983) Chemistry, occurrence and biosynthesis of *C*-glycosyl
624 compounds in plants. *Planta Med.* **47**, 131–140.
625 **Gandia-Herrero, F., Lorenz, A., Larson, T., Graham, I.A., Bowles, D.J., Rylott, E.L. and**
626 **Bruce, N.C.** (2008) Detoxification of the explosive 2,4,6-trinitrotoluene in Arabidopsis:
627 discovery of bifunctional *O*- and *C*-glucosyltransferases. *Plant J.* **56**, 963-974.
628 **Gattuso, G., Barreca, D., Gargiulli, C., Leuzzi, U. and Caristi, C.** (2007) Flavonoid
629 composition of *Citrus* juices. *Molecules*, **12**, 1641–1673.
630 **Hamilton, M., Kuate, S., Brazier-Hicks, M., Caulfield, J., Rose, R., Edwards, R., Torto, B.,**
631 **Pickett, J. and Hooper, A.** (2012) Elucidation of the biosynthesis of the
632 di-*C*-glycosylflavone isoschaftoside, an allelopathic component from *Desmodium* spp.
633 that inhibits *Striga* spp. development. *Phytochemistry*, **84**, 169–176.
634 **Hao, B., Caulfield, J.C., Hamilton, M.L., Pickett, J.A., Midega, C.A., Khan, Z.R., Wang,**
635 **J.R. and Hooper, A.M.** (2015) The biosynthesis of allelopathic di-*C*-glycosylflavones
636 from the roots of *Desmodium incanum* (G. Mey.) DC. *Org. Biomol. Chem.* **13**, 11663–
637 11673.
638 **Hirade, Y., Kotoku, N., Terasaka, K., Saijo-Hamano, Y., Fukumoto, A. and Mizukami, H.**
639 (2015) Identification and functional analysis of 2-hydroxyflavanone
640 *C*-glucosyltransferase in soybean (*Glycine max*). *FEBS Lett.* **589**, 1778–1786.
641 **Hooper, A.M., Tsanuo, M.K., Chamberlain, K., Tittcomb, K., Scholes, J., Hassanali, A.,**
642 **Khan, Z.R. and Pickett, J.A.** (2010) Isoschaftoside, a *C*-glycosylflavonoid from
643 *Desmodium uncinatum* root exudate, is an allelochemical against the development of
644 *Striga*. *Phytochemistry*, **71**, 904–908.

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.

Ito, T., Fujimoto, S., Shimosaka, M. and Taguchi, G. (2014) Production of *C*-glucosides of flavonoids and related compounds by *Escherichia coli* expressing buckwheat *C*-glucosyltransferase. *Plant Biotechnol.* **31**, 519–524.

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.

Khan I.A. and Kender, W.J. (2007) *Citrus* breeding: introduction and objectives. In *Citrus* Genetics, Breeding and Biotechnology (Khan, I.A. eds). Oxfordshare, UK: CABI, pp. 1–8.

Kumamoto, H., Matsubara, Y., Iizuka, Y., Okamoto, K. and Yokoi, K. (1986) Structure and hypotensive effect of flavonoid glycosides in orange (*Citrus sinensis* Osbeck) peelings. *Agric. Biol. Chem.* **50**, 781–783.

Manthey, J.A., Grohmanna, K., Berhowb, M.A. and Tisseratb, B. (2000) Changes in citrus leaf flavonoid concentrations resulting from blight induced zinc-deficiency. *Plant Physiol. Biochem.* **38**, 333–343.

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J., Ebana, K., Yano, M. and Saito, K. (2015) Metabolome-genome-wide association study dissects genetic

669 architecture for generating natural variation in rice secondary metabolism. *Plant J.* **81**,
670 13–23.
671 **McNally, D.J., Wurms, K.V., Labbé, C., Quideau, S. and Bélanger, R.R.** (2003) Complex
672 *C*-glycosyl flavonoid phytoalexins from *Cucumis sativus*. *J. Nat. Prod.* **66**, 1280–1283.
673 **Nagatomo, Y., Usui, S., Ito, T., Kato, A., Shimosaka, M. and Taguchi, G.** (2014)
674 Purification, molecular cloning and functional characterization of flavonoid *C*-
675 glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon. *Plant J.*
676 **80**, 437–448.
677 **Ogawa, K., Kawasaki, A., Omura, M., Yoshida, T., Ikoma, Y. and Yano, M.** (2001)
678 3',5'-Di-*C*-beta-glucopyranosylphloretin, a flavonoid characteristic of the genus
679 *Fortunella*. *Phytochemistry*, **57**, 737–742.
680 **Ohsugi, T.**, Nishida, R., Fukami, H. (1985) Oviposition stimulant of *Papilio xuthus*, a
681 citrus-feeding swallowtail butterfly. *Agric. Biol. Chem.* **49**, 1897–1900.
682 **Osmani, S.A., Bak, S. and Møller, B.L.** (2009) Substrate specificity of plant UDP-dependent
683 glycosyltransferases predicted from crystal structures and homology modeling.
684 *Phytochemistry*, **70**, 325–347.
685 **Paquette, S., Møller, L.B., Bak, S.** (2003) On the origin of family 1 plant glycosyltransferases.
686 *Phytochemistry*, **62**, 399–413.
687 **Pietta, P.G.** (2000) Flavonoids as antioxidants. *J. Nat. Prod.* **63**, 1035–1042.
688 **Sasaki, N., Nishizaki, Y., Yamada, E., Tatsuzawa, F., Nakatsuka, T., Takahashi, H. and**
689 **Nishihara, M.** (2015) Identification of the glucosyltransferase that mediates direct
690 flavone *C*-glucosylation in *Gentiana triflora*. *FEBS Lett.* **589**, 182–187.
691 **Shirley, B.W.** (1996) Flavonoid Biosynthesis: 'New' functions for an 'old' pathway. *Trends.*
692 *Plant Sci.* **1**, 377–382.

693 **Talhi, O. and Silva, A.** (2012) Advances in *C*-glycosylflavonoid research. *Curr. Org. Chem.*
694 **16**, 859–896.
695 **Wang, X.** (2009) Structure, mechanism and engineering of plant natural product
696 glycosyltransferases. *FEBS Lett.* **583**, 3303–3309.
697 **Xiao, J., Capanoglu, E., Jassbi, A.R. and Miron, A.** (2016) Advance on the flavonoid
698 *C*-glycosides and health benefits. *Crit. Rev. Food Sci. Nutr.*, **56**, S29–S45.
699 **Xiao, J., Muzashvili, T.S. and Georgiev, M.I.** (2014) Advances in the biotechnological
700 glycosylation of valuable flavonoids. *Biotechnol. Adv.* **32**, 1145–1156.
701 **Yonekura-Sakakibara, K. and Hanada, K.** (2011) An evolutionary view of functional
702 diversity in family 1 glycosyltransferases. *Plant J.* **66**, 182–193.
703

Table 1. Substrate specificities of recombinant *C*-glycosyltransferases (CGTs) from citrus plants.

Substrates	Enzyme activity (nkat mg protein^{-1}) FcCGT	CuCGT
Sugar acceptors[a]		
2-Hydroxyflavanones		
2-hydroxypinocembrin	67.6 ± 4.3	26.0 ± 0.4
2-hydroxynaringenin	63.4 ± 3.6	21.2 ± 2.1
Flavones, flavonols, and flavanones		
naringenin	ND[b]	ND
quercetin	ND	ND
apigenin	ND	ND
Dihydrochalcone and chalcone		
phloretin	34.2 ± 4.2	6.6 ± 0.1
isoliquiritigenin	ND	ND
Others		
2-phenyl-2′,4′,6′-trihydroxyacetophenone	37.9 ± 0.8	18.9 ± 1.9
maclurin	0.8 ± 0.0	1.3 ± 0.1
2′,4′,6′-trihydroxyacetophenone	5.6 ± 0.2	2.9 ± 0.2
2,4,6-trihydroxybenzaldehyde	1.9 ± 0.0	0.9 ± 0.0
2′,4′-dihydroxyactophenone	ND	ND
propyl gallate	ND	ND

Glucosides		
6-*C*-glucosyl-2-hydroxynaringenin	24.4 ± 0.3	12.4 ± 0.2
nothofagin (3′-*C*-glucosylphloretin)	59.6 ± 0.7	22.2 ± 1.0
3′-*C*-glucosyl-2-phenyl-2′,4′,6′-trihydroxyacetophenone	9.8 ± 0.2	21.2 ± 0.4
phloridzin	0.062	0.021
vitexin	ND	ND
isovitexin	ND	ND
Sugar donors[c]		
UDP-glucose	40.6 ± 5.1	15.3 ± 0.6
UDP-xylose	10.7 ± 0.8	5.7 ± 0.3
UDP-galactose	0.94 ± 0.07	0.24 ± 0.04
UDP-glucuronic acid	ND	ND

Values are averages ± SD (n = 3). [a]UDP-glucose was used as a sugar donor. [b]ND: not detected. [c]2-Hydroxynaringenin was used as a sugar acceptor.

Table 2. Kinetic parameters of the recombinant FcCGT.

Substrates	K_m (μM)	k_{cat} (sec^{-1})	k_{cat}/K_m (M^{-1} sec^{-1})
Sugar acceptors [a]			
phloretin	< 0.5[c]	12.0 ± 0.9	>2.4 × 10^7
2-hydroxynaringenin	0.85 ± 0.1	7.6 ± 0.5	8.9 × 10^6
nothofagin (3′-*C*-glucosylphloretin)	14.4 ± 1.8	5.3 ± 0.4	3.7 × 10^5
6-*C*-glucosyl-2-hydroxynaringenin	112.5 ± 22.0	5.7 ± 1.1	5.1 × 10^4
Sugar donor [b]			
UDP-glucose	71.5 ± 13.4	0.8 ± 0.1	1.1 × 10^4

[a]UDP-glucose was used as a sugar donor. [b]Nothofagin was used as a sugar acceptor. [c]K_m value could not be determined because of detection limits; the initial reaction speed remained constant even when concentrations of the substrate phloretin were 0.5 μM.

766 **FIGURE LEGENDS**

767 **Figure 1.** A proposed pathway of *C*-glycosylation toward flavonoids in citrus plants.
768 A single *C*-glycosyltransferase (CGT) catalyzes the first and second *C*-glucosylation of
769 flavonoids. (a) Vicenin-2; (b) 3′,5′-di-*C*-glucosylphloretin. Carbon positions are
770 indicated by the numbers beside the structures.

771

772 **Figure 2.** HPLC-MS analysis of the recombinant FcCGT reaction products of phloretin
773 and 2-hydroxynaringenin.
774 Each panel shows a chromatogram of the following conditions: reaction using (a)
775 phloretin and (b) 2-hydroxynaringenin as substrates incubated for 0 min, 5 min, and 30
776 min; standard compounds of (c) 3′,5′-di-*C*-glucosylphloretin and nothofagin; and (d)
777 vicenin-2, vitexin, and isovitexin. As the reaction was stopped by adding HCl,
778 2-hydroxynaringenin and its *C*-glucosides were dehydrated and thus detected as
779 apigenin and its *C*-glucosides. 3′,5′-Di-*C*-glucosylphloretin was synthesized by the
780 enzymatic reaction as described in the Experimental Procedures and confirmed by NMR.
781 The eluates were monitored at 290 nm (a, c) or 340 nm (b, d) using a diode array
782 detector. The negative electron-splay ionization (ES^-) MS spectra corresponding to the
783 di-*C*-glucosylated products are shown. The retention time of MS peaks was delayed by
784 about 0.08 min over that of the diode array. Peak identification: 1 and Vc, vicenin-2; 2
785 and Dp, 3′,5′-di-*C*- glucosylphloretin; 3 and Vt, vitexin; 4 and Iv, isovitexin; 5,
786 apigenin; 6 and N, nothofagin; 7, phloretin.

787

788

789 **Figure 3.** qRT-PCR analyses of *FcCGT* and *CuCGT,* and the accumulation of flavonoid
790 di-*C*-glucosides in kumquat and satsuma mandarin plants.
791 (a, b) Quantitative reverse transcription (qRT)-PCR analyses of *FcCGT* and *CuCGT*
792 transcript levels in kumquat and satsuma mandarin plants, respectively, were performed
793 as described in the Experimental Procedures using total RNAs extracted from the
794 flowers (FL), immature fruit (IMF), peel of matured fruit (PE), young leaves (YL), and
795 mature leaves (ML) of these plants. Transcript levels were estimated via the ddCt
796 method based on the 2nd derivative maximum, and were shown relative to that of
797 glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) with three biological replicates
798 (average ± SD). (a) *FcCGT* transcript in kumquat; (b) *CuCGT* transcript in satsuma
799 mandarin. (c, d) Quantities of flavonoid di-*C*-glucosides in each organ of (c) kumquat
800 and (d) satsuma mandarin were analyzed as described in the Experimental Procedures.
801 The average ± SD amounts (μmol per g fresh weight [gFW]) of three biological
802 replicates are shown. Filled and empty bars indicate the amount of vicenin-2 and
803 3′,5′-di-*C*-glucosylphloretin, respectively. Statistical analyses were performed with
804 Tukey's test; different letters above the error bars indicate significant differences ($p <$
805 0.05); n.d., not detected.
806
807 **Figure 4.** Production of di-*C*-glucosides by bioconversion using *E. coli* expressing
808 CuCGT.
809 *E. coli* harboring CuCGT cDNA were used for the bioconversion of flavonoid
810 substrates into their *C*-glucosides. Induction of CuCGT expression and feeding of the
811 2-hydroxynaringenin (a–c) or phloretin (d, e) were performed as described in the
812 Experimental Procedures. The culture media were sampled immediately (0 h) and 2 h

813 after substrate addition and subjected to HPLC-MS analysis following the addition of
814 methanol. Chromatograms reflect the following conditions: (a) immediately and (b) 2 h
815 after addition of 2-hydroxynaringenin; (c) the 2-h samples following treatment with
816 HCl; (d) immediately and (e) 2 h following addition of phloretin; and (f) negative
817 electron-splay ionization (ES^-) MS spectra of the products for 2-hydroxynaringenin
818 (peaks 8 and 9). The retention time of MS peaks was delayed by about 0.08 min over
819 that of the diode array. Peak identification: 1, vicenin-2; 2, 3′,5′-di-*C*-glucosylphloretin;
820 3, vitexin; 4, isovitexin; 5, apigenin; 6, nothofagin; 7, phloretin; 8,
821 6,8-di-*C*-glucosyl-2-hydroxynaringenin; 9, 6-*C*-glucosyl-2-hydroxynaringenin; 10,
822 2-hydroxynaringenin.
823
824 **Figure 5.** Neighbor-joining tree inferred from the amino acid sequences of citrus plant
825 *C*-glycosyltransferases (CGTs) and related glycosyltransferases.
826 A molecular phylogenetic tree was constructed using the neighbor-joining method in the
827 MEGA6 software (http://www.megasoftware.net/) based on the deduced amino acid
828 sequences of UGTs (citrus CGTs, other reported CGTs, and several types of UGTs were
829 selected as representative) that were aligned with ClustalW. Bar indicates 0.2 amino
830 acid substitutions per site. Orthologous groups (OG) of plant UGTs are shown in
831 accordance with those proposed by Yonekura-Sakakibara and Hanada (2011).
832 Abbreviations and GenBank accession numbers: FcCGT (UGT708G1), LC131333;
833 CuCGT (UGT708G2), LC131334; ChCGT (UGT708G3), LC131335; MiCGT,
834 ALD83754; GmCGT (UGT708D1), I1L3T1; FeCGTa (UGT708C1), BAP90360;
835 FeCGTb (UGT708C2), BAP90361; ZmCGT (UGT708A6), A0A096SRM5; OsCGT,
836 CAQ77160; PcPGT (UGT88F2), ACZ44838; RhGT1, BAD99560; Sb7GT (UGT88D1),

837 BAC98300; CrYMb4 (UGT71E2), BAF75901; UGT72B1, Q9M156; UGT73B4,
838 Q7Y232; NtIS5a (UGT73A1), AAB36653; AtF7RhamT (UGT89C1), Q9LNE6;
839 VvGT1 (UGT78A5), AAB81683; UGT85H2, ABI94024; AtNGT (UGT76C1),
840 AED90934; AtSAGT (UGT74F2), O22822; AtHCAGT2 (UGT84A1), Q5XF20;
841 GtUF6CGT, BAQ19550; UrdGT2, AAF00209. Detailed information about the UGTs is
842 presented in Table S1.

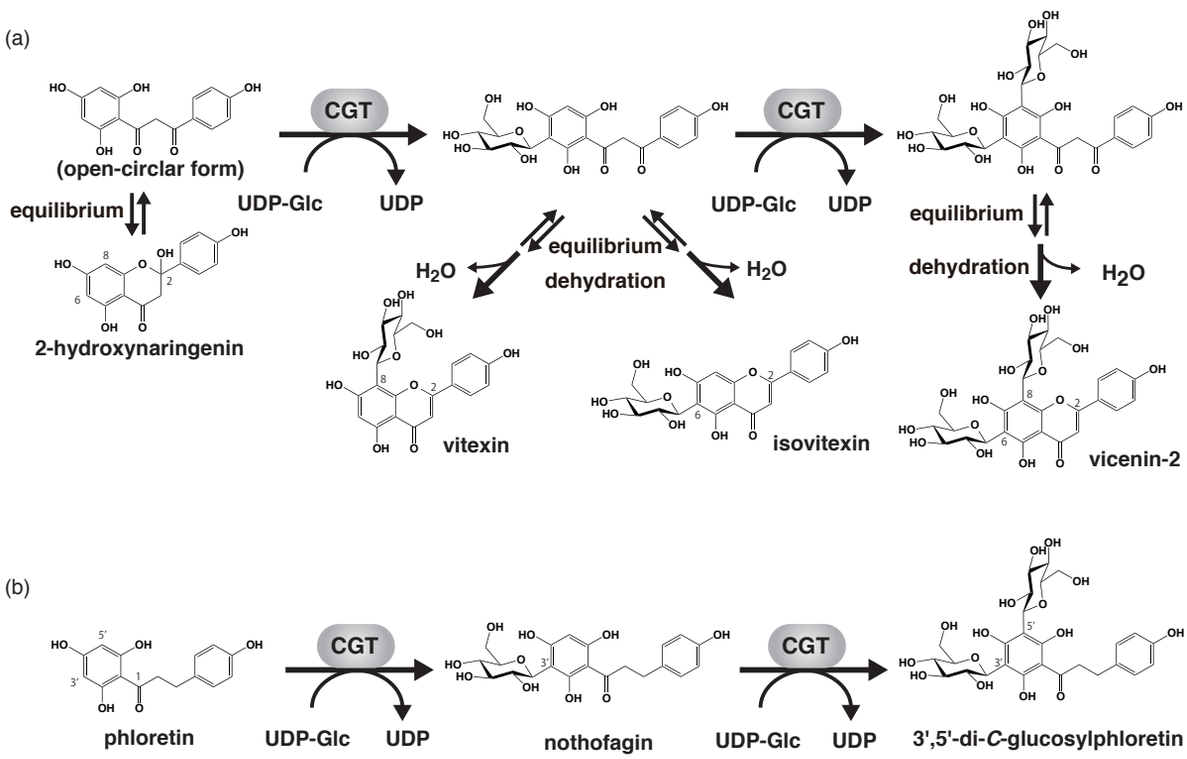


Figure 1. A proposed pathway of C-glycosylation toward flavonoids in citrus plants.
A single C-glycosyltransferase (CGT) catalyzes the first and second C-glucosylation of flavonoids. (a) Vicenin-2; (b) 3′,5′-di-C-glucosylphloretin. Carbon positions are indicated by the numbers beside the structures.

Figure 1, Ito et al.

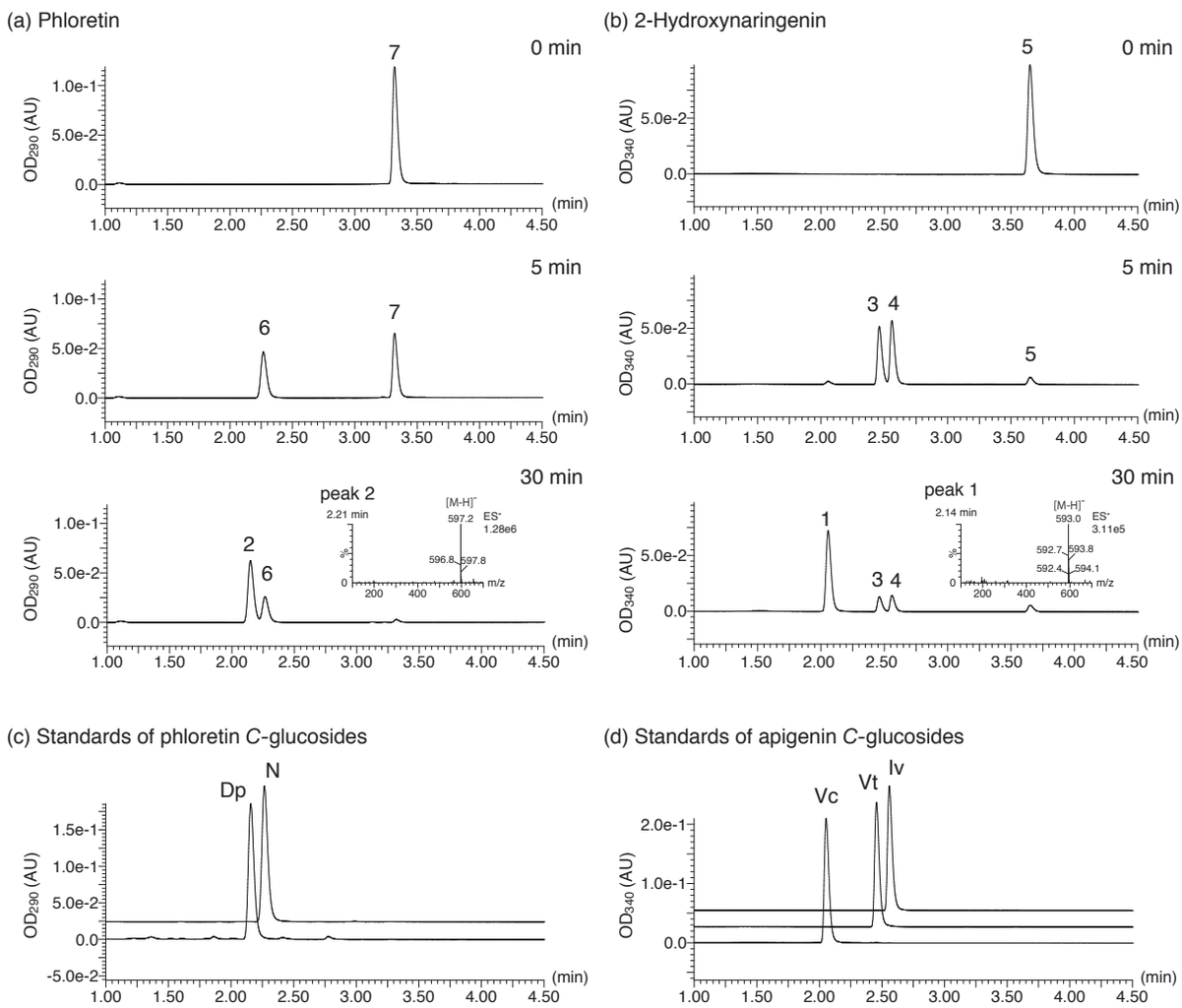


Figure 2. HPLC-MS analysis of the recombinant FcCGT reaction products of phloretin and 2-hydroxynaringenin. Each panel shows a chromatogram of the following conditions: reaction using (a) phloretin and (b) 2-hydroxynaringenin as substrates incubated for 0 min, 5 min, and 30 min; standard compounds of (c) 3′,5′-di-C-glucosylphloretin and nothofagin; and (d) vicenin-2, vitexin, and isovitexin. As the reaction was stopped by adding HCl, 2-hydroxynaringenin and its C-glucosides were dehydrated and thus detected as apigenin and its C-glucosides. 3′,5′-Di-C-glucosylphloretin was synthesized by the enzymatic reaction as described in the Experimental Procedures and confirmed by NMR. The eluates were monitored at 290 nm (a, c) or 340 nm (b, d) using a diode array detector. The negative electron-splay ionization (ES–) MS spectra corresponding to the di-C-glucosylated products are shown. The retention time of MS peaks was delayed by about 0.08 min over that of the diode array. Peak identification: 1 and Vc, vicenin-2; 2 and Dp, 3′,5′-di-C- glucosylphloretin; 3 and Vt, vitexin; 4 and Iv, isovitexin; 5, apigenin; 6 and N, nothofagin; 7, phloretin.

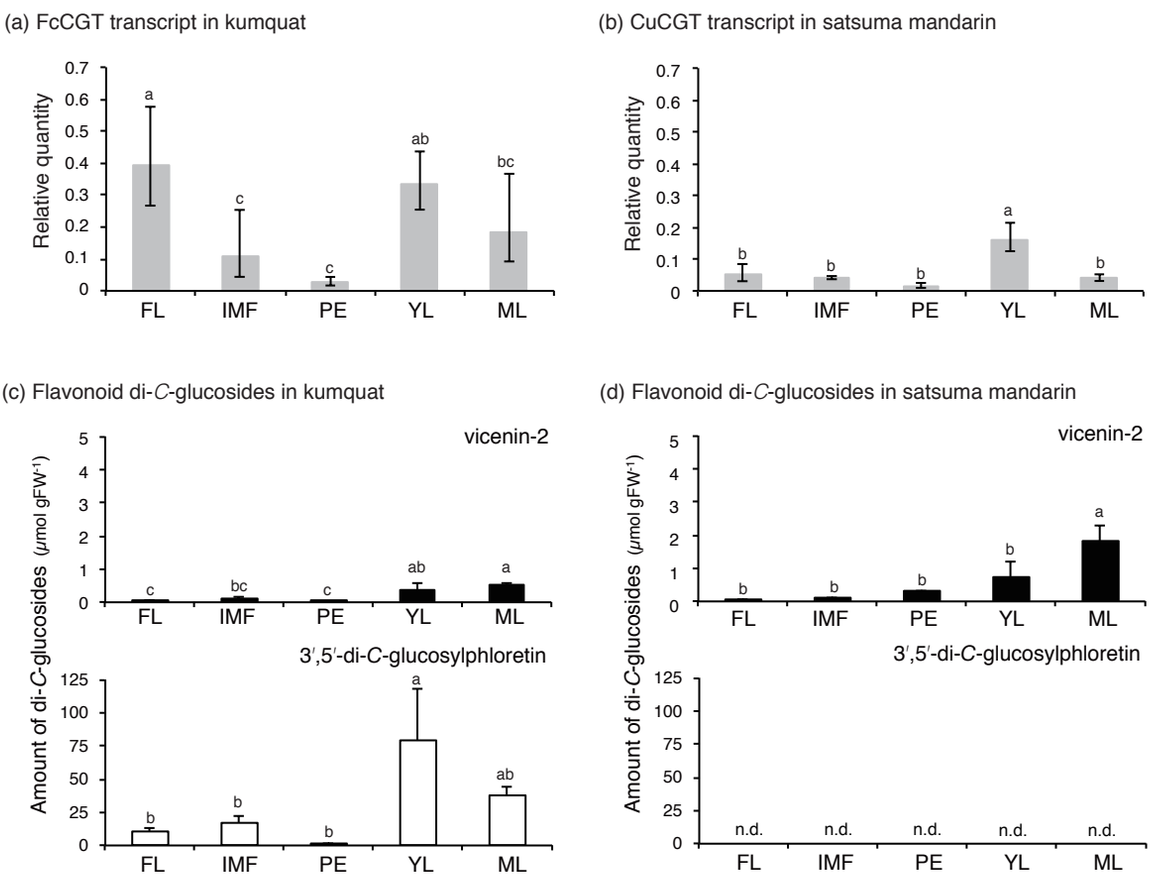


Figure 3. qRT-PCR analyses of FcCGT and CuCGT, and the accumulation of flavonoid di-C-glucosides in kumquat and satsuma mandarin plants.

(a, b) Quantitative reverse transcription (qRT)-PCR analyses of FcCGT and CuCGT transcript levels in kumquat and satsuma mandarin plants, respectively, were performed as described in the Experimental Procedures using total RNAs extracted from the flowers (FL), immature fruit (IMF), peel of matured fruit (PE), young leaves (YL), and mature leaves (ML) of these plants. Transcript levels were estimated via the ddCt method based on the 2nd derivative maximum, and were shown relative to that of glyceraldehyde-3-phosphate dehydrogenase (gapdh) with three biological replicates (average ± SD). (a) FcCGT transcript in kumquat; (b) CuCGT transcript in satsuma mandarin. (c, d) Quantities of flavonoid di-C-glucosides in each organ of (c) kumquat and (d) satsuma mandarin were analyzed as described in the Experimental Procedures. The average ± SD amounts (μmol per g fresh weight [gFW]) of three biological replicates are shown. Filled and empty bars indicate the amount of vicenin-2 and 3′,5′-di-C-glucosylphloretin, respectively. Statistical analyses were performed with Tukey' s test; different letters above the error bars indicate significant differences ($p < 0.05$); n.d., not detected.

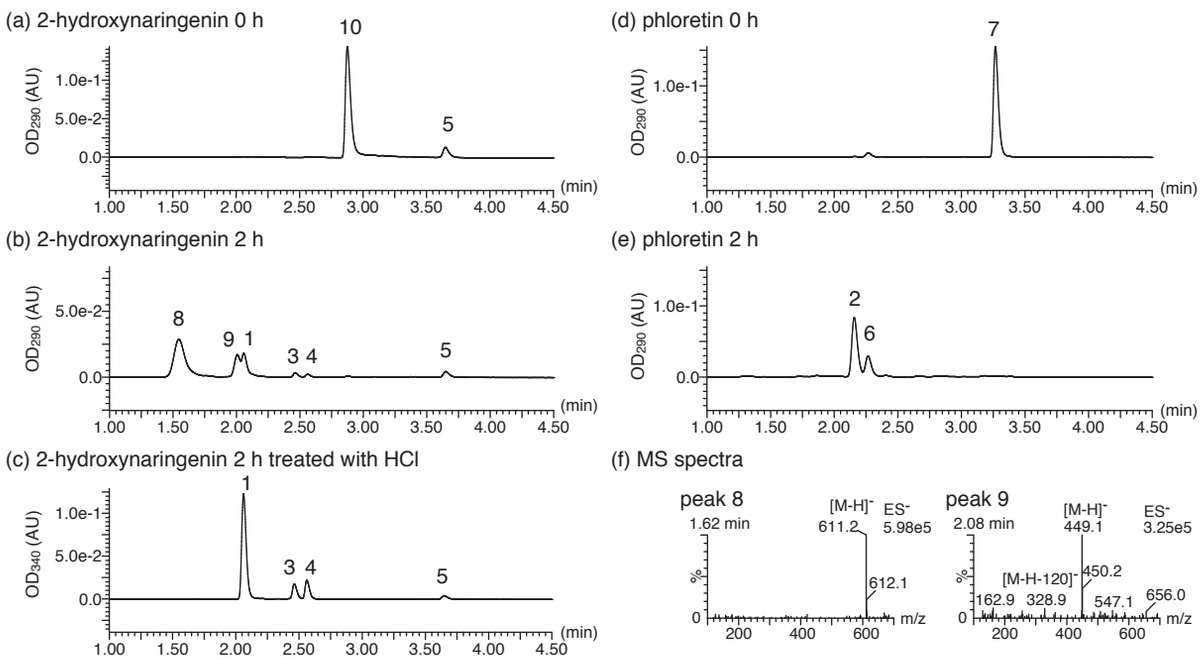


Figure 4. Production of di-C-glucosides by bioconversion using E. coli expressing CuCGT.
E. coli harboring CuCGT cDNA were used for the bioconversion of flavonoid substrates into their C-glucosides. Induction of CuCGT expression and feeding of the 2-hydroxynaringenin (a–c) or phloretin (d, e) were performed as described in the Experimental Procedures. The culture media were sampled immediately (0 h) and 2 h after substrate addition and subjected to HPLC-MS analysis following the addition of methanol. Chromatograms reflect the following conditions: (a) immediately and (b) 2 h after addition of 2-hydroxynaringenin; (c) the 2-h samples following treatment with HCl; (d) immediately and (e) 2 h following addition of phloretin; and (f) negative electron-splay ionization (ES–) MS spectra of the products for 2-hydroxynaringenin (peaks 8 and 9).
The retention time of MS peaks was delayed by about 0.08 min over that of the diode array. Peak identification: 1, vicenin-2; 2, 3′,5′-di-C-glucosylphloretin; 3, vitexin; 4, isovitexin; 5, apigenin; 6, nothofagin; 7, phloretin; 8, 6,8-di-C-glucosyl-2-hydroxynaringenin; 9, 6-C-glucosyl-2-hydroxynaringenin; 10, 2-hydroxynaringenin.

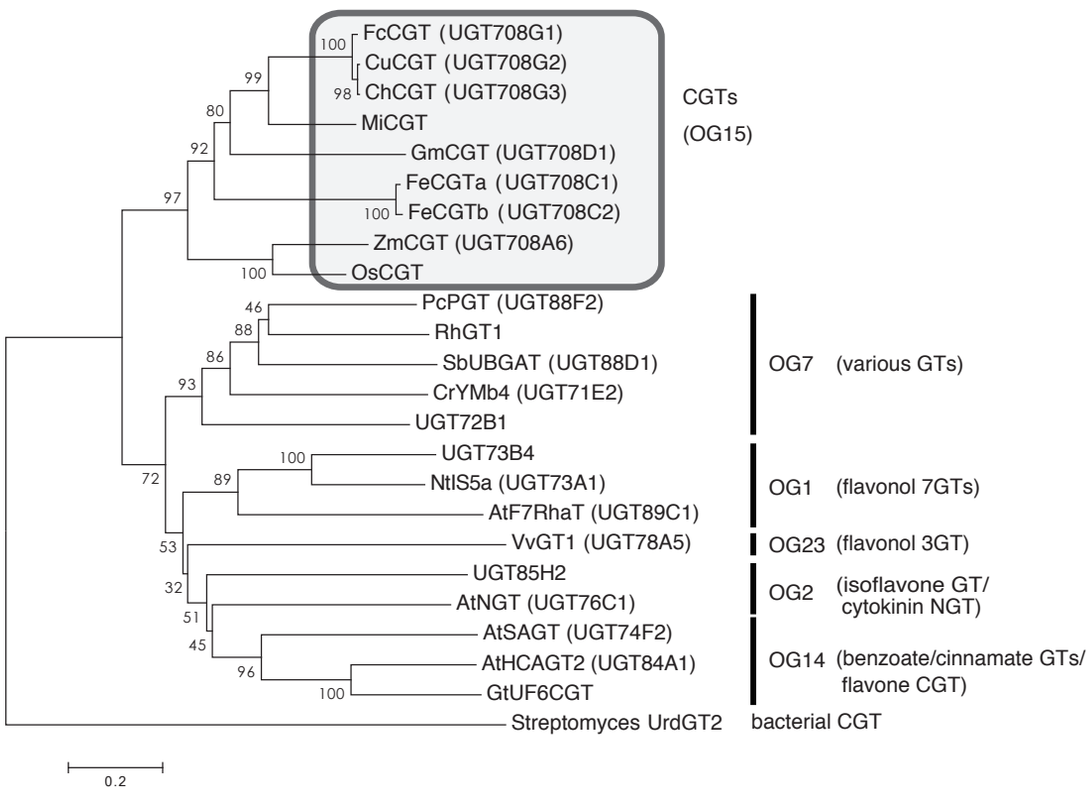


Figure 5. Neighbor-joining tree inferred from the amino acid sequences of citrus plant C-glycosyltransferases (CGTs) and related glycosyltransferases.

A molecular phylogenetic tree was constructed using the neighbor-joining method in the MEGA6 software (http://www.megasoftware.net/) based on the deduced amino acid sequences of UGTs (citrus CGTs, other reported CGTs, and several types of UGTs were selected as representative) that were aligned with ClustalW. Bar indicates 0.2 amino acid substitutions per site. Orthologous groups (OG) of plant UGTs are shown in accordance with those proposed by Yonekura-Sakakibara and Hanada (2011). Abbreviations and GenBank accession numbers: FcCGT (UGT708G1), LC131333; CuCGT (UGT708G2), LC131334; ChCGT (UGT708G3), LC131335; MiCGT, ALD83754; GmCGT (UGT708D1), I1L3T1; FeCGTa (UGT708C1), BAP90360; FeCGTb (UGT708C2), BAP90361; ZmCGT (UGT708A6), A0A096SRM5; OsCGT, CAQ77160; PcPGT (UGT88F2), ACZ44838; RhGT1, BAD99560; Sb7GT (UGT88D1), BAC98300; CrYMb4 (UGT71E2), BAF75901; UGT72B1, Q9M156; UGT73B4, Q7Y232; NtIS5a (UGT73A1), AAB36653; AtF7RhamT (UGT89C1), Q9LNE6; VvGT1 (UGT78A5), AAB81683; UGT85H2, ABI94024; AtNGT (UGT76C1), AED90934; AtSAGT (UGT74F2), O22822; AtHCAGT2 (UGT84A1), Q5XF20; GtUF6CGT, BAQ19550; UrdGT2, AAF00209. Detailed information about the UGTs is presented in Table S1.

SUPPORTING INFORMATION

Table S1. List of UDP-sugar dependent glycosyltransferases (UGTs) used in the phylogenetic analysis.

Figure S1. Distribution of *C*-glucosides in kumquat and satsuma mandarin leaves. Chromatograms reflect the following conditions: methanol extract of (a) kumquat leaves and (b) kumquat leaves treated with 2 M HCl at 60 ºC for 3 h; (c) satsuma mandarin leaves; (d) standard compound of vicenin-2; and (e) 3′,5′-di-*C*-glucosylphloretin synthesized in this study. The eluates were detected by measuring absorbance at 290 nm and/or 340 nm. Peaks of the UV-spectra and negative electron-splay ionization (ES^-) MS spectra corresponding to di-*C*-glucosyl flavonoids are shown. The retention time of MS peaks was delayed by about 0.08 min over that of the diode array. Peak identifications: 1 and Vc, vicenin-2; 2 and Dp, 3′,5′-di-*C*-glucosylphloretin.

Figure S2. Detection of *C*-glucosylation activities in cell-free extracts of kumquat and satsuma mandarin leaves.
Cell-free extracts of kumquat and satsuma mandarin leaves were incubated with 2-hydroxynaringenin and uridine diphosphate (UDP)-glucose at 30 °C for 2 h. The reaction was stopped by adding HCl; the reaction mixtures were further incubated at 60 ºC for 30 min to complete the dehydration of the compounds, then analyzed by HPLC-MS as described in the Experimental Procedures. Chromatograms reflect the enzymatic reaction of (a) kumquat and (b) satsuma mandarin leaves before (0 h; upper

panel) and after (2 h; lower panel) incubation. The eluates were detected by measuring absorbance at 340 nm. Peak identifications: 1, vicenin-2; 3, vitexin; 4, isovitexin; 5, apigenin.

Figure S3. Alignment of deduced amino acid sequences of *C*-glycosyltransferases (CGT) candidates (EST) from citrus plants.

Deduced amino acid sequences of citrus plant ESTs found in the TIGR database (http://blast.jcvi.org/euk-blast/plantta-blast.cgi/, as of September 2012) using *FeCGTa* (Accession No. BAP90360) as a query were aligned via ClustalW (http://clustalw.ddbj.nig.ac.jp) and visualized with BoxShade Server (http://www.ch.embnet.org/software/BOX_form.html). The EST sequences consisted of TA5377_37690 (*Poncirus trifoliate*), TA14910_2711 (*Citrus sinensis*), TA4403_85681 (*C. clementine*), and TA829_55188 (*C. unshiu*).

Figure S4. Comparison of the amino acid sequences of *C*-glycosyltransferases (CGTs) from citrus and other types of plants.

Deduced CGT amino acid sequences were aligned via the ClustalW program in MEGA6 (http://www.megasoftware.net/) and visualized with BoxShade Server (http://www.ch.embnet.org/software/BOX_form.html). Conserved regions among plant UGTs (PSPG-motif) are indicated in the figure. Asterisks indicate the amino acid residues required for the nucleophilic reaction toward sugar acceptors conserved in UDP-sugar dependent glycosyltransferases (UGTs). Accession numbers of the CGTs were: CuCGT (UGT708G2), LC131334; ChCGT (UGT708G3), LC131335; FcCGT (UGT708G1), LC131333; GmCGT (UGT708D1), I1L3T1; MiCGT, ALD83754;

FeCGTa (UGT708C1), BAP90360; FeCGTb (UGT708C2), BAP90361; OsCGT, CAQ77160; ZmCGT (UGT708A6), A0A096SRM5.

Figure S5. HPLC analysis of the recombinant CuCGT and ChCGT reaction products from 2-hydroxynaringenin.

Recombinant (a) CuCGT and (b) ChCGT were incubated with 2-hydroxynaringenin and UDP-glucose for 0 min, 5 min, and 30 min, following which reactions were stopped by the addition of HCl; further incubation at 60 ºC for 30 min was then performed to dehydrate the compounds completely, after which they were analyzed with HPLC. The eluates were monitored at 340 nm using a diode array detector. Peak identification: 1, vicenin-2; 3, vitexin; 4, isovitexin; 5, apigenin.

Figure S6. Properties of recombinant FcCGT and CuCGT.

To verify pH preference (a, b) and optimal temperature (c, d) for reactions (FcCGT [a, c] and CuCGT [b, d]), reactions were examined using UDP-glucose and 2-hydroxynaringenin as substrates, as described in the Experimental Procedures. Values of the relative activities are average ± SD (n = 3), with maximum activity levels assumed to be 100%. Filled squares and filled circles represent the activities using a potassium phosphate buffer and an ethanolamine-HCl buffer, respectively.

Figure S7. Structures of the compounds used for the enzyme assay.

Numbers beside the structures indicate the carbon positions.

Figure S8. Bioconversion of phenolic compounds using *E. coli* expressing CuCGT or FcCGT.

E. coli cells expressing CuCGT or FcCGT were used for the bioconversion of flavonoid substrates into their *C*-glucosides. The culture media were sampled immediately after (0 h) and 2 h after substrate addition, then subjected to HPLC-MS analysis. Panels show chromatograms constructed from the following conditions: bioconversion of (a) 2-phenyl-2′,4′,6′-trihydroxyacetophenone or (b) phloretin into their *C*-glucosides by *E. coli* expressing (a) CuCGT or (b) FcCGT. The eluates were monitored at 290 nm using a diode array detector. Products were compared with the standard compounds except for peak 11. The negative electron-splay ionization (ES^-) MS spectra of the products for 2-phenyl-2′,4′,6′-trihydroxyacetophenone (peaks 11 and 12) are shown in the small panels. The retention time of MS peaks was delayed by about 0.08 min over that of the diode array. Peak identifications: 2, 3′,5′-di-*C*-glucosylphloretin; 6, nothofagin; 7, phloretin; 11, 2-phenyl-2′,4′,6′-trihydroxyacetophenone-di-*C*-glucoside; 12, 2-phenyl-3′-*C*-glucosyl- 2′,4′,6′-trihydroxyacetophenone; 13, 2-phenyl-2′,4′,6′-trihydroxyacetophenone.

Table S1. List of UDP-sugar dependent glycosyltransferases (UGTs) used in the phylogenetic analysis.

Name	Function	Accession No.
FcCGT (UGT708G1)	flavonoid *C*-glucosyltransferase from *Fortunella crassifolia*	LC131333 [a]
CuCGT (UGT708G2)	flavonoid *C*-glucosyltransferase from *Citrus unshiu*	LC131334 [a]
ChCGT (UGT708G3)	flavonoid *C*-glucosyltransferase from *Citrus hanaju*	LC131335 [a]
MiCGT	*C*-glucosyltransferase from *Mangifera indica*	ALD83754
GmCGT (UGT708D1)	flavonoid *C*-glucosyltransferase from *Glycine max*	I1L3T1
FeCGTa (UGT708C1)	flavonoid *C*-glucosyltransferase from *Fagopyrum esculentum*	BAP90360
FeCGTb (UGT708C2)	flavonoid *C*-glucosyltransferase from *Fagopyrum esculentum*	BAP90361
UGT708A6	flavonoid *C*-glucosyltransferase from *Zea mays*	A0A096SRM5
OsCGT	flavonoid *C*-glucosyltransferase from *Oryza sativa*	CAQ77160
PcPGT (UGT88F2)	phloretin 2'-*O*-glucosyltransferase *Pyrus communis*	ACZ44838
RhGT1	anthocyanidin 5,3-*O*-glucosyltransferase from *Rosa* hybrid cultivar	BAD99560
SbUBGAT (UGT88D1)	baicalein 7-*O*-glucuronosyltransferase from *Scutellaria baicalensis*	BAC98300
CrYMb4 (UGT71E2)	tetrahydroxychalcone 2'-glucosyltransferase from *Catharanthus roseus*	BAF75901
UGT72B1	xenobiotic glucosyltransferase from *Arabidopsis thaliana*	Q9M156
UGT73B4	xenobiotic *O*- and *C*-glucosyltransferase from *A. thaliana*	Q7Y232
NtIS5a	flavonol and coumarin glucosyltransferase from *N. tabacum*	AAB36653
AtF7RhamT (UGT89C1)	flavonol-7-*O*-rhamnosyltransferase from *A. thaliana*	Q9LNE6
VvGT1 (UGT78A5)	flavonoid-3-*O*-glucosyltransferases from *Vitis vinifera*	AAB81683
UGT85H2	(iso)flavonoid glycosyltransferase from *Medicago truncatula*	ABI94024
AtSAGT (UGT74F2)	salicylic acid glucosyltransferase from *A. thaliana*	O22822
AtNGT (UGT76C1)	cytokinin-*N*-glucosyltransferase from *A. thaliana*	AED90934
AtHCGT2 (UGT84A1)	hydroxycinnamate glucosyltransferase 2 from *A. thaliana*	Q5XF20
GtUF6CGT	flavone 6-*C*-glucosyltransferase from *Gentiana triflora*	BAQ19550
UrdGT2	angucycline *C*-glycosyltransferase from *Streptomyces fradiae*	AAF00209

[a] Accession No. for gene sequences.

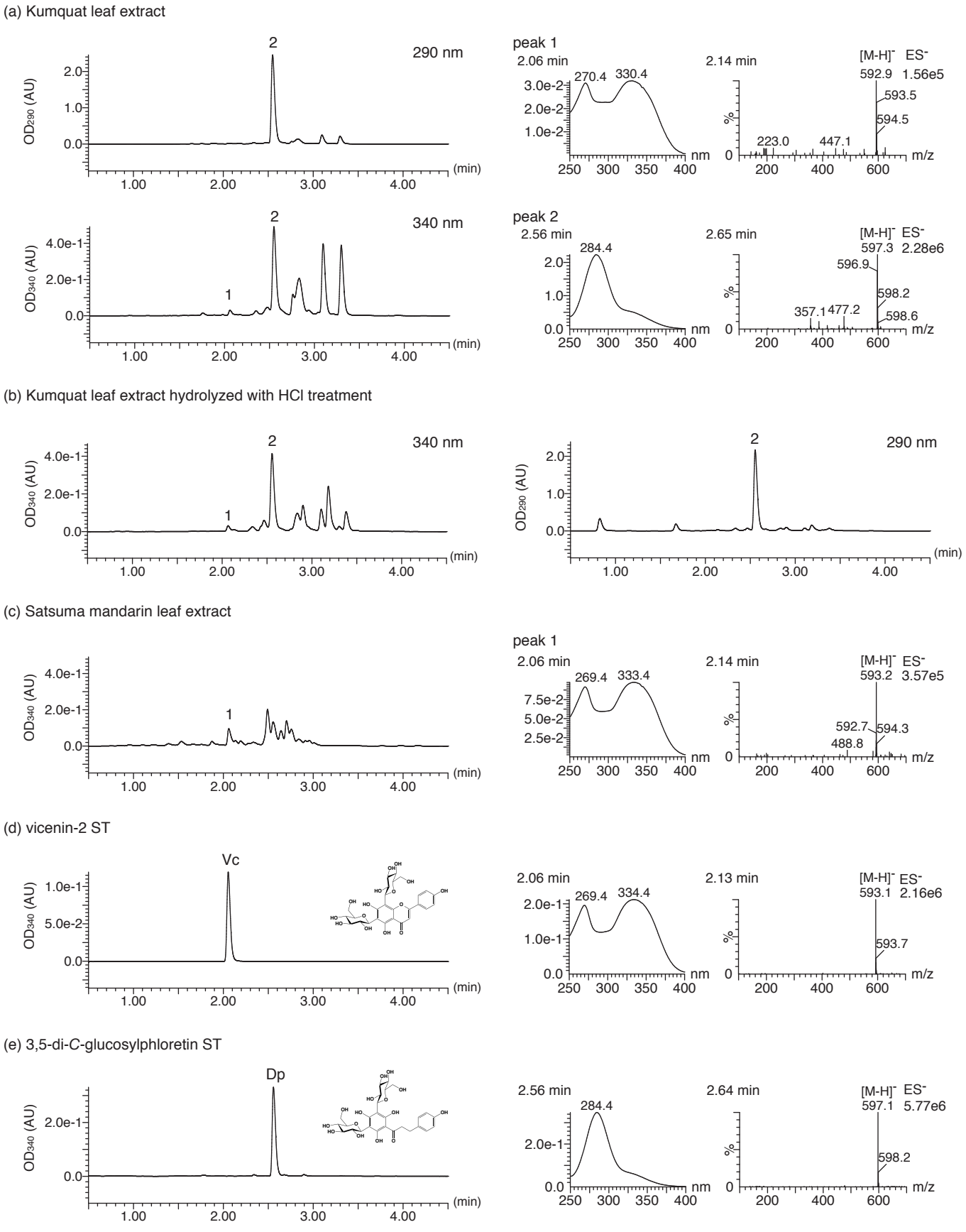
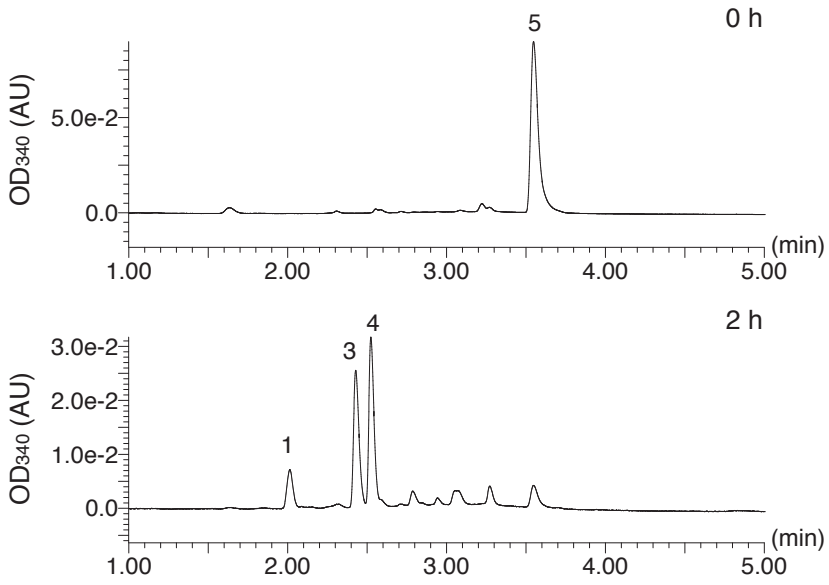


Figure S1. Distribution of *C*-glucosides in kumquat and satsuma mandarin leaves.
Chromatograms reflect the following conditions: methanol extract of (a) kumquat leaves and (b) kumquat leaves treated with 2 M HCl at 60 ºC for 3 h; (c) satsuma mandarin leaves; (d) standard compound of vicenin-2; and (e) 3′,5′-di-*C*-glucosylphloretin synthesized in this study. The eluates were detected by measuring absorbance at 290 nm and/or 340 nm. Peaks of the UV-spectra and negative electron-splay ionization (ES^-) MS spectra corresponding to di-*C*-glucosyl flavonoids are shown. The retention time of MS peaks was delayed by about 0.08 min over that of the diode array. Peak identifications: 1 and Vc, vicenin-2; 2 and Dp, 3′,5′-di-*C*-glucosylphloretin.

(a) Kumquat leaf enzyme



(b) Satsuma mandarin leaf enzyme

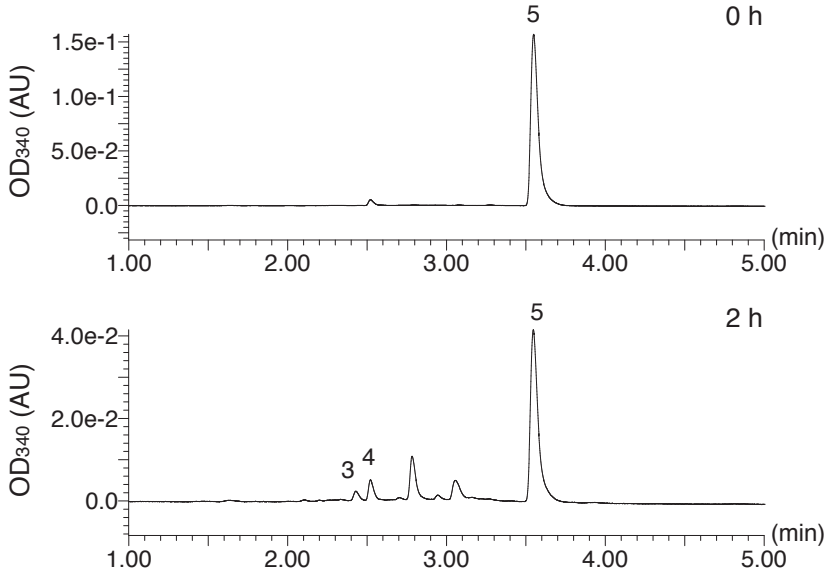


Figure S2. Detection of *C*-glucosylation activities in cell-free extracts of kumquat and satsuma mandarin leaves. Cell-free extracts of kumquat and satsuma mandarin leaves were incubated with 2-hydroxynaringenin and uridine diphosphate (UDP)-glucose at 30 °C for 2 h. The reaction was stopped by adding HCl; the reaction mixtures were further incubated at 60 ºC for 30 min to complete the dehydration of the compounds, then analyzed by HPLC-MS as described in the Experimental Procedures. Chromatograms reflect the enzymatic reaction of (a) kumquat and (b) satsuma mandarin leaves before (0 h; upper panel) and after (2 h; lower panel) incubation. The eluates were detected by measuring absorbance at 340 nm. Peak identifications: 1, vicenin-2; 3, vitexin; 4, isovitexin; 5, apigenin.

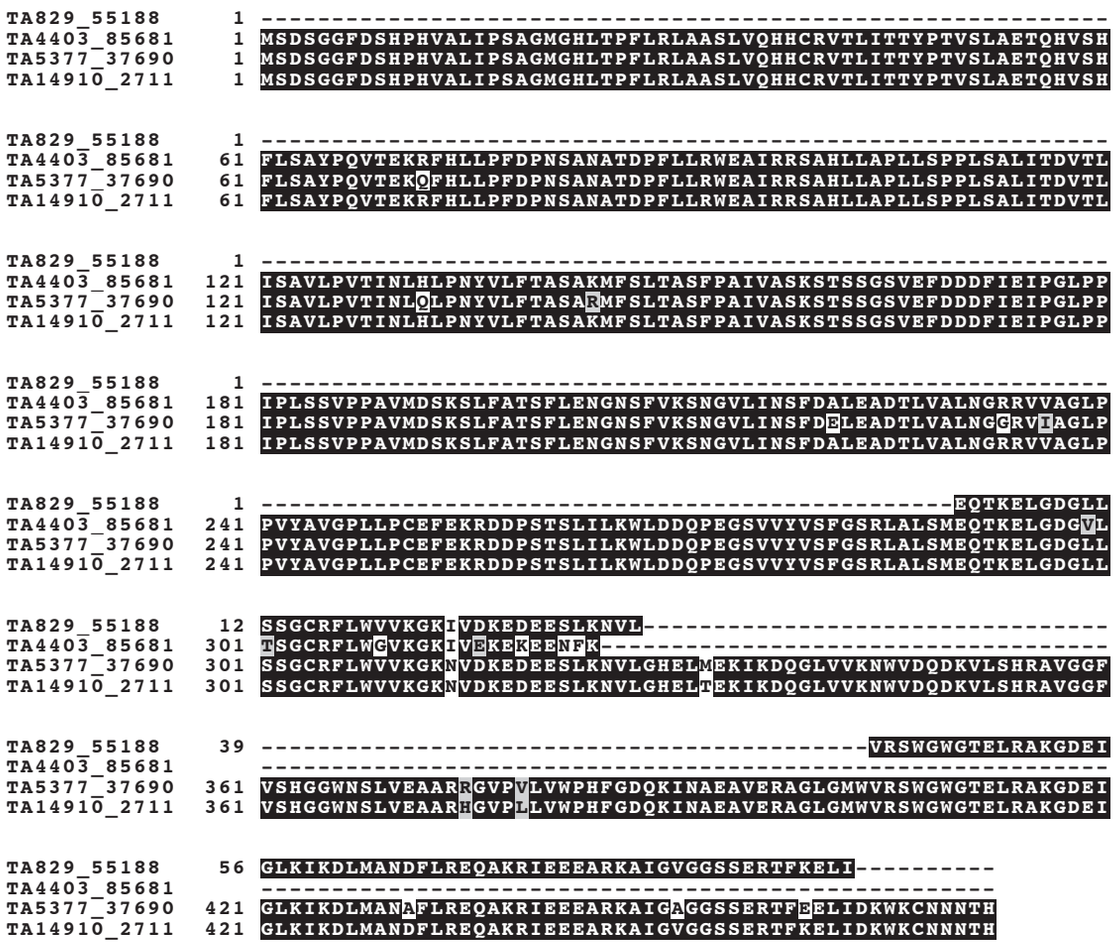


Figure S3. Alignment of deduced amino acid sequences of *C*-glycosyltransferases (CGT) candidates (EST) from citrus plants.

Deduced amino acid sequences of citrus plant ESTs found in the TIGR database (http://blast.jcvi.org/euk-blast/plantta-blast.cgi/, as of September 2012) using FeCGTa (Accession No. BAP90360) as a query were aligned via ClustalW (http://clustalw.ddbj.nig.ac.jp) and visualized with BoxShade Server (http://www.ch.embnet.org/software/BOX_form.html). The EST sequences consisted of TA5377_37690 (*Poncirus trifoliate*), TA14910_2711 (*Citrus sinensis*), TA4403_85681 (*C. clementine*), and TA829_55188 (*C. unshiu*).

Figure S4. Comparison of the amino acid sequences of *C*-glycosyltransferases (CGTs) from citrus and other types of plants. Deduced CGT amino acid sequences were aligned via the ClustalW program in MEGA6 (http://www.megasoftware.net/) and visualized with BoxShade Server (http://www.ch.embnet.org/software/BOX_form.html). Conserved regions among plant UGTs (PSPG-motif) are indicated in the figure. Asterisks indicate the amino acid residues required for the nucleophilic reaction toward sugar acceptors conserved in UDP-sugar dependent glycosyltransferases (UGTs). Accession numbers of the CGTs were: CuCGT (UGT708G2), LC131334; ChCGT (UGT708G3), LC131335; FcCGT (UGT708G1), LC131333; GmCGT (UGT708D1), I1L3T1; MiCGT, ALD83754; FeCGTa (UGT708C1), BAP90360; FeCGTb (UGT708C2), BAP90361; OsCGT, CAQ77160; ZmCGT (UGT708A6), A0A096SRM5.

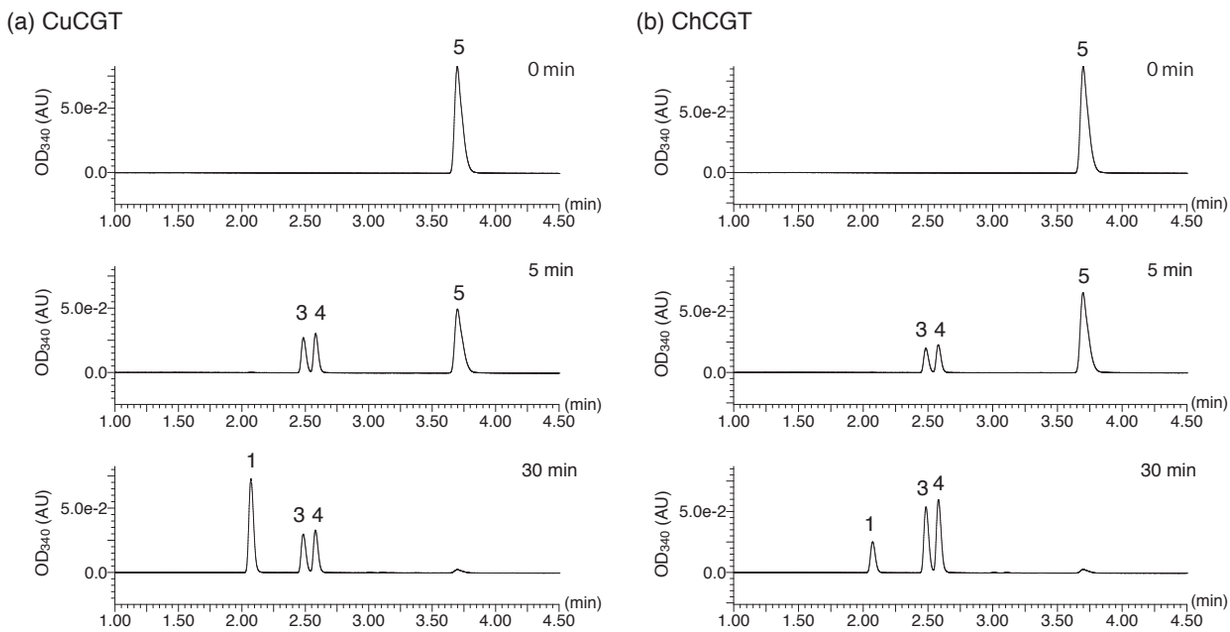


Figure S5. HPLC analysis of the recombinant CuCGT and ChCGT reaction products from 2-hydroxynaringenin. Recombinant (a) CuCGT and (b) ChCGT were incubated with 2-hydroxynaringenin and UDP-glucose for 0 min, 5 min, and 30 min, following which reactions were stopped by the addition of HCl; further incubation at 60 ºC for 30 min was then performed to dehydrate the compounds completely, after which they were analyzed with HPLC. The eluates were monitored at 340 nm using a diode array detector. Peak identification: 1, vicenin-2; 3, vitexin; 4, isovitexin; 5, apigenin.

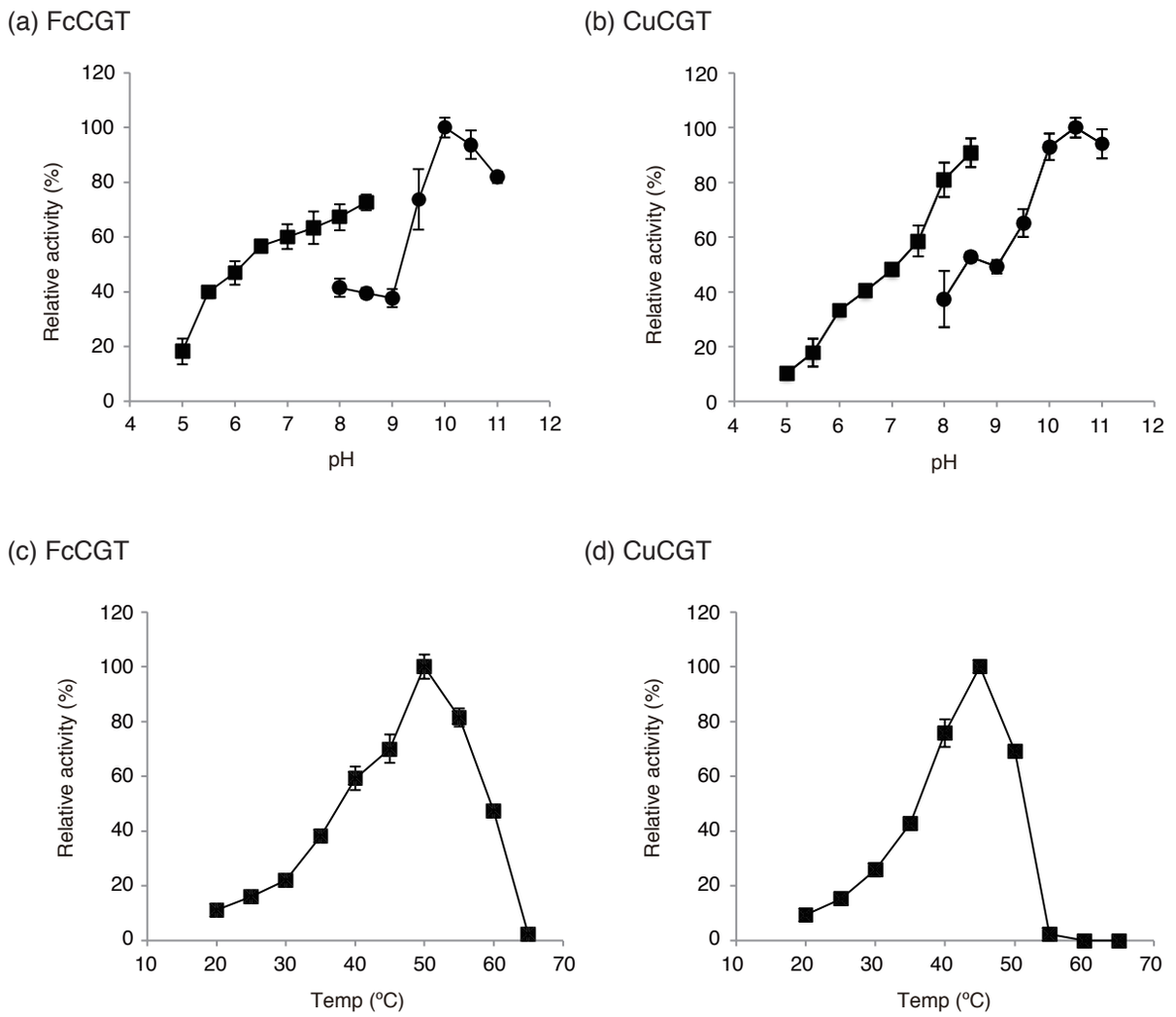
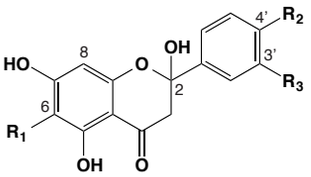
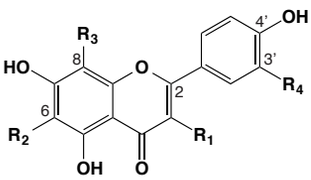


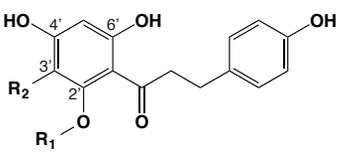
Figure S6. Properties of recombinant FcCGT and CuCGT.
To verify pH preference (a, b) and optimal temperature (c, d) for reactions (FcCGT [a, c] and CuCGT [b, d]), reactions were examined using UDP-glucose and 2-hydroxynaringenin as substrates, as described in the Experimental Procedures. Values of the relative activities are average ± SD (n = 3), with maximum activity levels assumed to be 100%. Filled squares and filled circles represent the activities using a potassium phosphate buffer and an ethanolamine-HCl buffer, respectively.



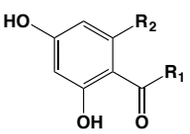
R_1=R_2=R_3=H 2-hydroxypinocembrin
R_1=R_3=H, R_2=OH 2-hydroxynaringenin
R_1=Glc, R_2=OH, R_3=H 6-*C*-glucosyl-2-hydroxynaringenin



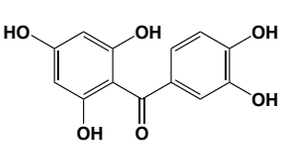
R_1=R_2=R_3=R_4=H apigenin
R_1=R_4=OH, R_2=R_3=H quercetin
R_1=R_2=R_4=H, R_3=Glc vitexin
R_1=R_3=R_4=H, R_2=Glc isovitexin



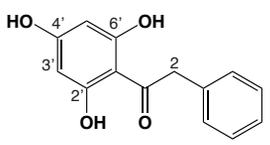
R_1=R_2=H phloretin
R_1=H, R_2=Glc nothofagin
R_1=Glc, R_2=H phloridzin



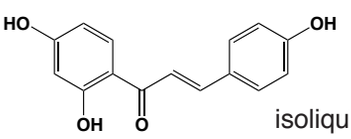
R_1=CH_3, R_2=H 2',4'-dihydroxyactophenone
R_1=H, R_2=OH 2,4,6-trihydroxybenzaldehyde
R_1=CH_3, R_2=OH 2',4',6'-trihydroxyacetophenone



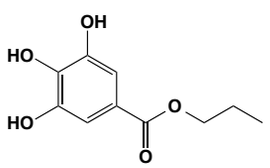
maclurin



2-phenyl-2',4',6'-trihydroxyactophenone



isoliquiritigenin



propyl gallate

Figure S7. Structures of the compounds used for the enzyme assay.
Numbers beside the structures indicate the carbon positions.

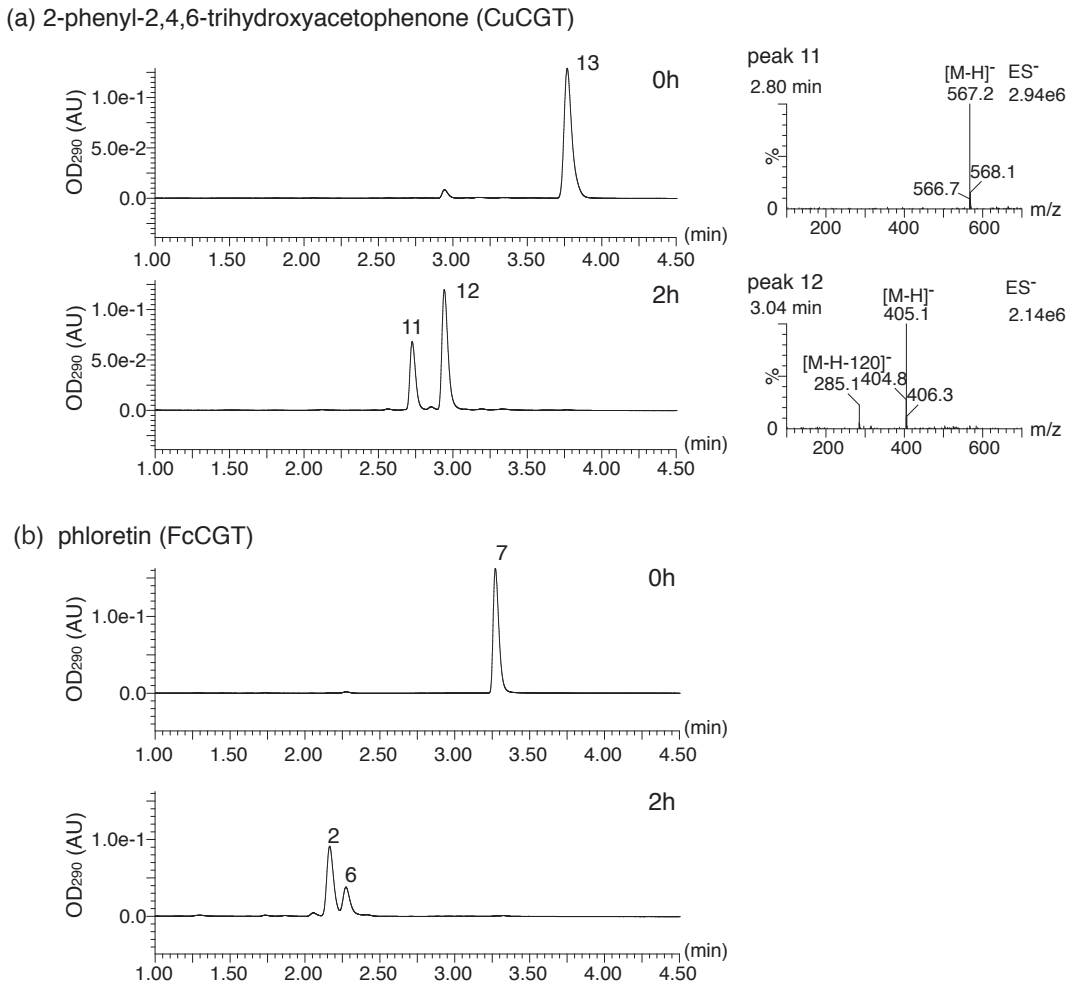


Figure S8. Bioconversion of phenolic compounds using E. coli expressing CuCGT or FcCGT.
E. coli cells expressing CuCGT or FcCGT were used for the bioconversion of flavonoid substrates into their *C*-glucosides. The culture media were sampled immediately after (0 h) and 2 h after substrate addition, then subjected to HPLC-MS analysis. Panels show chromatograms constructed from the following conditions: bioconversion of (a) 2-phenyl-2′,4′,6′-trihydroxyacetophenone or (b) phloretin into their *C*-glucosides by *E. coli* expressing (a) CuCGT or (b) FcCGT. The eluates were monitored at 290 nm using a diode array detector. Products were compared with the standard compounds except for peak 11. The negative electron-splay ionization (ES^-) MS spectra of the products for 2-phenyl-2′,4′,6′-trihydroxyacetophenone (peaks 11 and 12) are shown in the small panels. The retention time of MS peaks was delayed by about 0.08 min over that of the diode array. Peak identifications: 2, 3′,5′-di-*C*-glucosylphloretin; 6, nothofagin; 7, phloretin; 11, 2-phenyl-2′,4′,6′-trihydroxyacetophenone-di-*C*-glucoside; 12, 2-phenyl-3′-*C*-glucosyl- 2′,4′,6′-trihydroxyacetophenone; 13, 2-phenyl-2′,4′,6′-trihydroxyacetophenone.