

Volatile Glycosylation in Tea Plants: Sequential  
Glycosylations for the Biosynthesis of Aroma  
 $\beta$ -Primeverosides Are Catalyzed by Two *Camellia  
sinensis* Glycosyltransferases

メタデータ	言語: eng 出版者: 公開日: 2015-06-18 キーワード (Ja): キーワード (En): 作成者: Ohgami, Shoji, Ono, Eiichiro, Horikawa, Manabu, Murata, Jun, Totsuka, Koujirou, Toyonaga, Hiromi, Ohba, Yukie, Dohra, Hideo, Asai, Tatsuo, Matsui, Kenji, Mizutani, Masaharu, Watanabe, Naoharu, Ohnishi, Toshiyuki メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/10297/8723">http://hdl.handle.net/10297/8723</a>

1 **Running Title: Tea UGTs in biosynthesis of aroma  $\beta$ -primeverosides**

2

3 **Author to whom correspondence should be sent:**

4 Toshiyuki Ohnishi

5 Graduate School of Agriculture, Shizuoka University, 829, Ohya, Suruga-ku,

6 Shizuoka 422-8529, Japan

7 Telephone: + 81 54 238 3082

8 Fax: + 81 54 238 3082

9 e-mail: [dtonish@ipc.shizuoka.ac.jp](mailto:dtonish@ipc.shizuoka.ac.jp)

10

11 Eiichiro Ono

12 Research Institute, Suntory Global Innovation Center Ltd., 1-1-1

13 Wakayamadai, Shimamoto, Mishima, Osaka 618-8503, Japan

14 Telephone: +81 75 962 2244

15 Fax: +81 75 962 2244

16 e-mail: [eiichiro\\_ono@suntory.co.jp](mailto:eiichiro_ono@suntory.co.jp)

17

18 Research Area 1: Biochemistry and Metabolism

19

20 Research Area 2: Genes, Development and Evolution

21

22 **Volatile glycosylation in tea plants: Sequential glycosylations for the biosynthesis of aroma**

23 ***β*-primeverosides are catalyzed by two *Camellia sinensis* glycosyltransferases**

24

25 Shoji Ohgami<sup>2</sup>, Eiichiro Ono<sup>2,\*</sup>, Manabu Horikawa, Jun Murata, Koujiro Totsuka,

26 Hiromi Toyonaga, Yukie Ohba<sup>3</sup>, Hideo Dohra, Tatsuo Asai, Kenji Matsui, Masaharu

27 Mizutani, Naoharu Watanabe and Toshiyuki Ohnishi\*

28

29 Graduate School of Agriculture, Shizuoka University, Shizuoka 422-8529, Japan (SO,

30 KT, YO, TA, TO); Research Institute, Suntory Global Innovation Center Ltd.,

31 Shimamoto, Mishima, Osaka 618-8503, Japan (EO, HT); Bioorganic Research Institute,

32 Suntory Foundation for Life Sciences, Shimamoto, Mishima, Osaka 618-8503, Japan

33 (MH, JM); Research Institute of Green Science and Technology, Shizuoka University,

34 Shizuoka 422-8529, Japan (HD, TO); Graduate School of Medicine, Yamaguchi

35 University, Yamaguchi 753-8515, Japan (KM); Graduate School of Agricultural Science,

36 Kobe University, Kobe 657-8501, Japan (MM); Graduate School of Engineering,

37 Shizuoka University, Hamamatsu 432-8561, Japan (WN)

38

39 One-Sentence Summary:

40 Two glycosyltransferases catalyze sequential glycosylations of volatiles important for

41 tea aroma quality, leading to stable accumulation of the volatiles as the water-soluble

42 *β*-primeverosides.

43 **Footnotes**

44

45 <sup>1</sup>This work was supported by JSPS KAKENHI Grant Number 26870252 to T.O., and  
46 also supported by Suntory Holdings Limited to E.O.

47 <sup>2</sup>These authors contributed equally to this work.

48 <sup>3</sup>Present address: Bioorganic Research Institute, Suntory Foundation for Life Sciences,  
49 Shimamoto, Mishima, Osaka 618-8503, Japan

50 \* Corresponding authors: dtonish@ipc.shizuoka.ac.jp (TO); eiichiro\_ono@suntory.co.jp  
51 (EO)

52

53 The author responsible for distribution of materials integral to the findings presented in  
54 this article in accordance with the policy described in the Instructions for Authors  
55 (www.plantphysiol.org) is Toshiyuki Ohnishi (dtonish@ipc.shizuoka.ac.jp) and Eiichiro  
56 Ono (eiichiro\_ono@suntory.co.jp).

57

58 **Abstract**

59 Tea plants (*Camellia sinensis*) store volatile organic compounds (VOCs; monoterpene,  
60 aromatic, and aliphatic alcohols) in the leaves in the form of water-soluble diglycosides,  
61 primarily as  $\beta$ -primeverosides (6-*O*- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranosides). These  
62 VOCs play a critical role in plant defenses and tea aroma quality, yet little is known  
63 about their biosynthesis and physiological roles *in planta*. Here we identified two  
64 UDP-glycosyltransferases (UGTs) from *C. sinensis*: UGT85K11 (CsGT1) and  
65 UGT94P1 (CsGT2), converting VOCs into  $\beta$ -primeverosides by sequential  
66 glucosylation and xylosylation, respectively. CsGT1 exhibits a broad substrate  
67 specificity toward monoterpene, aromatic and aliphatic alcohols to produce the  
68 respective glucosides. On the other hand, CsGT2 specifically catalyzes the xylosylation  
69 of the 6'-hydroxy group of the sugar moiety of geranyl  $\beta$ -D-glucopyranoside, producing  
70 geranyl  $\beta$ -primeveroside. Homology modeling, followed by site-directed mutagenesis of  
71 CsGT2, identified a unique isoleucine 141 residue playing a crucial role in sugar donor  
72 specificity toward UDP-xylose. The transcripts of both *CsGTs* were mainly expressed in  
73 young leaves, along with  $\beta$ -primeverosidase ( $\beta$ -PD) encoding a diglycoside-specific  
74 glycosidase. In conclusion, our findings reveal the mechanism of aroma  
75  $\beta$ -primeverosides biosynthesis in *C. sinensis*. This information can be used to preserve  
76 tea aroma better during the manufacturing process and to investigate the mechanism of  
77 plant chemical defenses.

78

## 79 **Introduction**

80 Plants emit volatile organic compounds (VOCs), such as monoterpenes (C10),  
81 sesquiterpenes (C15), phenylpropanoids (C9), norisoprenoids (C16), aromatic esters, or  
82 green leaf alcohols (C6), in response to attacks by insect herbivores, mechanical  
83 wounding, or endogenous developmental cues. In general, VOCs are considered not  
84 only to be chemical defense compounds transmitting biological signals to the  
85 environment (Arimura et al., 2009), but also important commercial products because  
86 they influence the quality and character of dietary foods and beverages as aromas. Tea,  
87 manufactured from *Camellia sinensis* (Fig. 1A), is the most popular beverage in the  
88 world, and is classified as black, green or oolong tea based on the manufacturing  
89 processes (withering, rolling, and fermentation), which affects the composition and  
90 quantity of aroma compounds (Balentine et al., 1997; Graham, 1992). For instance,  
91 black tea is produced by full fermentation, during which tea metabolites artificially react  
92 with endogenous enzymes (e.g., polyphenol oxidases and  $\beta$ -glucosidases). Floral tea  
93 aroma is one of the crucial components to evaluate the value and quality of tea products.  
94 The floral aroma caused by linalool, geraniol, 2-phenylethanol (2-PE) and benzyl  
95 alcohol are predominant flavor volatile compounds in oolong tea and black tea, whereas  
96 (*Z*)-3-hexenol adds a grassy note to green tea (Kawakami et al., 1995; Kumazawa and  
97 Masuda, 2002). However, fresh leaves of *C. sinensis* barely emit a slightly green note  
98 before they are processed for tea. This is because aroma volatiles in tea leaves are  
99 accumulated typically as water-soluble glycoside form. Since the first report on the  
100 isolation of benzyl  $\beta$ -D-glucopyranoside (benzyl-glc) (Yano et al., 1991), various aroma

101 glycosides have been identified in fresh tea leaves (Kobayashi et al., 1994). The  
102 chemical structure of most glycosides was shown to be  $\beta$ -primeveroside  
103 (6-*O*- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranoside; Guo et al., 1993, 1994; Moon et al.,  
104 1994, 1996), suggesting that a common biosynthetic machinery for the conjugation of  
105  $\beta$ -primeveroside to aroma volatiles exists. Previous quantitative analysis of aroma  
106 glycosides in tea leaves demonstrated that levels of aroma  $\beta$ -primeverosides are  
107 three-fold higher than monoglycosides (Wang et al., 2000), indicating that sequential  
108 sugar conjugating reactions to aroma volatiles (glucosylation followed by xylosylation)  
109 occur in tea leaves. Highly diverse aroma volatiles, such as benzyl alcohol, 2-PE,  
110 (*Z*)-3-hexenol, linalool and geraniol, are stored as  $\beta$ -primeverosides in the leaves of *C.*  
111 *sinensis* (Guo et al., 1993, 1994; Moon et al., 1994, 1996). Aromatic alcohols, such as  
112 benzyl alcohol and 2-PE, serve as attractants for both parasitic and predatory insects for  
113 herbivores and (*Z*)-3-hexenol released from herbivore-damaged tissue has also been  
114 found to induce defense responses in neighboring plants (Pichersky and Gershenzon,  
115 2002; Sugimoto et al., 2014). Monoterpene alcohols such as geraniol and linalool have  
116 potential activity toward microorganisms and fungi and geraniol also has a potent  
117 apoptosis-inducing activity in plant cells (Pattnaik et al., 1997; Izumi et al., 1999).  
118 Mizutani *et al.* (2002) reported that a  $\beta$ -primeverosidase ( $\beta$ -PD) from *C. sinensis*  
119 specifically hydrolyzes aroma  $\beta$ -primeverosides into primeverose (disaccharide unit)  
120 and aroma volatile (aglycone unit). The data support the idea that aroma  $\beta$ -PDs are the  
121 key enzymes responsible for the production of chemical defense compounds against  
122 pathogens and herbivores as well as for the characteristic aromas of tea products. Thus,

123 it is of particular interest to understand the biosynthesis and physiological role of aroma  
124  $\beta$ -primeverosides in *C. sinensis*. However, corresponding genes for biosynthesis of  
125 volatile  $\beta$ -primeverosides have so far not been reported.

126 Previous attempts to over-produce volatile compounds in plants by over-expressing  
127 genes that are responsible for the biosynthesis of aglycones often resulted in the  
128 accumulation of respective glycosides. For example, ectopic expression of *Clarkia*  
129 *breweri* (*S*)-linalool synthase (LIS) in *Petunia hybrida* resulted in the accumulation of  
130 (*S*)-linalyl  $\beta$ -D-glucopyranosides ((*S*)-linalyl-glc) (Lücker et al., 2001) and transgenic  
131 Arabidopsis plants expressing a strawberry terpene synthase (*FaNES1*) produced  
132 (*S*)-linalool, nerolidol, and the glycosylated derivatives (Aharoni et al., 2003). Moreover,  
133 relieving a bottleneck in the endogenous eugenol pathway by heterologous  
134 over-expression of a *P. hybrida* coniferyl alcohol acetyltransferase (*PhCFAT*) gene  
135 resulted in up to 7- and 22-fold increase in the levels of eugenol, and its glycoside  
136 (eugenyl-glc), respectively, in leaves of transgenic aspen plants (Koeduka et al., 2013).  
137 These results suggest that glycosylation of volatiles is a general phenomenon in land  
138 plants.

139 Here we demonstrate the biochemical and molecular characteristics of two  
140 UDP-glycosyltransferases (UGTs) from *C. sinensis*, UGT85K11 (CsGT1) and  
141 UGT94P1 (CsGT2), responsible for the sequential glucosylation and xylosylation in the  
142 biosynthesis of volatile  $\beta$ -primeverosides (Fig. 1B). In addition, we discuss the  
143 physiological roles that volatile metabolites might play in plants, based on the  
144 distribution of aroma precursors and spatiotemporal expression pattern of these UGT

145 genes in *C. sinensis*.

146

147

## 148 **Results**

### 149 **Organ-specific composition in aroma monoglycosides and diglycosides**

150 Aroma monoglycosides and diglycosides were extracted from fresh leaves and  
151 stems at two developmental stages (young and mature) of *C. sinensis*. Various  
152  $\beta$ -primeverosides, as well as monoglycosides of aroma compounds in tea leaves, were  
153 quantified by liquid chromatography–mass spectrometry (LC–MS) (Fig. 2). The results  
154 show that geranyl  $\beta$ -primeveroside (geranyl-pri) and linalyl  $\beta$ -primeveroside  
155 (linalyl-pri) were the two primary aroma glycosides that were detected mainly in young  
156 organs, leaves and stems, respectively. These data also suggest that the metabolic  
157 activity of the glycosylation machinery responsible for the biosynthesis of aroma  
158  $\beta$ -primeverosides is higher in growing young tissues. As the tea leaves grew, the total  
159 amounts of 2-phenylethyl  $\beta$ -primeveroside (2PE-pri), benzyl  $\beta$ -primeveroside  
160 (benzyl-pri), and (*Z*)-3-hexenyl  $\beta$ -primeveroside (hexenyl-pri) increased in the mature  
161 leaves, whereas those of geranyl-pri and linalyl-pri decreased (Supplemental Table S1).  
162 Since the overall fresh weight of the mature leaves was approximately four times larger  
163 than in young leaves, the apparent concentrations of geranyl-pri and linalyl-pri were  
164 substantially decreased in mature leaves. The results suggested that these two  
165  $\beta$ -primeverosides were further metabolized to unknown chemical forms or were  
166 transferred from young leaves to other parts of the plant.

167

168 **Identification of *Arabidopsis* UGT85A3 showing trans-glycosylation activity**  
169 **toward volatiles**

170 Concurrent occurrence of monoglucosides and primeverosides of the  
171 corresponding volatiles in tea leaves suggested that primeverosides are biosynthesized  
172 via two sequential glycosylations steps, an initial glucosylation, followed by  
173 xylosylation rather than by direct conjugation of the primeverosyl moiety to the  
174 volatiles. We therefore searched for glucosyltransferases potentially responsible for the  
175 first glucosylation step in aroma  $\beta$ -primeveroside biosynthesis. For several classes of  
176 specialized metabolites, the biosynthetic genes are often found co-expressed  
177 (Fukushima et al., 2011). Transcriptome expression profiles and co-expression analysis  
178 became powerful tools for prediction of the biosynthetic genes constituting the  
179 metabolic pathway (Fukushima et al., 2011, Ginglinger et al., 2013, Usadel et al.,  
180 2009). However, a co-expression analytical tool for *C. sinensis*, a non-model plant,  
181 are not yet available. For identification of the glucosyltransferases catalyzing the first  
182 glucosylation of volatiles, we surveyed UGTs co-expressing with structural genes for  
183 monoterpene biosynthesis in *Arabidopsis* by ATTED II (<http://atted.jp>), which is a  
184 database developed to identify functionally related genes by co-expression.  
185 By using geraniol/nerol 10-hydroxylase gene (*At2g45580*; *CYP76C3*) and linalool  
186 synthase gene (*At1g61680*; *AtLIS*) (Ginglinger et al., 2013; Mizutani et al., 1997;  
187 Obayashi et al., 2011) as probes, we found that the expression profiles of *Arabidopsis*  
188 *UGT85A3* (*At\_UGT85A3*, *At1g22380*;  $r = 0.873$ ) exhibits relatively high correlation

189 with *AtLIS* and *CYP76C3* (Supplemental Fig. S1). *In vitro* functional characterization of  
190 *At\_UGT85A3* was performed using UDP-glucose as a sugar donor and geraniol or  
191 (*Z*)-3-hexenol as a sugar acceptor revealed that *At\_UGT85A3* produced geranyl-glc  
192 from geraniol and hexenyl-glc from (*Z*)-3-hexenol (Supplemental Fig. S2 and  
193 Supplemental Fig. S3). These data demonstrate that *At\_UGT85A3* is capable of  
194 catalyzing the glucosylation of monoterpene alcohols and aliphatic alcohols.

195

196 **Identification of a *C. sinensis* UGT catalyzing the first glucosylation step for**  
197 **volatile  $\beta$ -primeveroside.**

198 To isolate *C. sinensis* UGTs responsible for the first glucosylation step in volatile  
199  $\beta$ -primeveroside biosynthesis, a cDNA library constructed from a mixture of leaves,  
200 stem and roots of *C. sinensis* (Mizutani et al., 2002) was screened with digoxigenin  
201 (DIG)-labeled *At\_UGT85A3*. Two rounds of screening identified four novel UGTs,  
202 which were individually expressed in *Escherichia coli* and subjected to enzyme activity  
203 assays using UDP-glucose as a sugar donor and a variety of volatile alcohol acceptors.  
204 We found that one of the UGTs, named CsGT1, catalyzes glucosylation of geraniol as  
205 shown by the appearance of a product peak at the retention time of 10.2 min with *m/z*  
206 361 ( $[M + HCOO]^-$ ), both values of which correspond to those of authentic geranyl-glc  
207 (Fig. 3A and Fig. 3B). CsGT1 was assigned as Cs\_UGT85K11 by the committee  
208 responsible for naming UDP-glucuronosyltransferases (Mackenzie et al., 1997).

209 The maximum velocity ( $V_{max}$ ) and estimated apparent  $K_m$  values of CsGT1 for  
210 geraniol were  $332.1 \pm 8.1$  nkat  $mg^{-1}$  protein and  $44.2 \pm 3.0$   $\mu M$ , respectively (Table 1).

211 The sugar acceptor specificity of CsGT1 was surveyed using six aroma alcohols and  
212 two flavonoids found in the leaves of *C. sinensis*. CsGT1 was active toward all six  
213 volatiles with relative activities: geraniol (100%), eugenol (84%), (Z)-3-hexenol (62%),  
214 benzyl alcohol (48%), 2-PE (9.2%), and linalool (1.4%), whereas CsGT1 did not accept  
215 quercetin or cyanidine as substrates (Fig. 3C). On the other hand, CsGT1 showed clear  
216 preference to UDP-glucose (100%) as a sugar donor compared to UDP-galactose (15%),  
217 UDP-xylose (n.d.), and UDP-glucuronic acid (n.d.) when geraniol was used as a sugar  
218 acceptor (Fig. 3D). Taken together, these data indicate that CsGT1 preferentially  
219 glucosylates volatiles using UDP-glucose as a specific sugar donor but exhibits a broad  
220 substrate specificity for sugar acceptors.

221

## 222 **Identification of orthologous UGTs of *CsGT1* from various plants**

223 Volatile glycosides are reported in different plant species, including apricot, peach,  
224 yellow plum (Krammer et al., 1991), grape berries (Günata et al., 1985), kiwi (Young  
225 and Paterson, 1995), strawberry (Roscher, et al., 1997), raspberry (Pabst et al., 1991),  
226 and tomato (Marlatt, et al., 1992). Based on these general observations, we searched for  
227 UGTs in the NCBI Genbank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) based on amino acid sequence  
228 similarities with CsGT1 (accession number: AB847092). We found homologous UGTs  
229 broadly represented throughout the angiosperm plant lineages. We cloned another five  
230 UGTs from grapevine (*Vitis vinifera*: Vv\_UGT85A33, Vv\_UGT85A28,  
231 Vv\_UGT85A30), sweet potato (*Ipomoea batatas*: Ib\_UGT85A32), and snapdragon  
232 (*Antirrhinum majus*: Am\_UGT85A13), and experimentally characterized the

233 recombinant enzymes by the procedure used for CsGT1. They exhibited volatile  
234 glycosylating activities similar to CsGT1, including the production of geranyl-glc and  
235 hexenyl-glc (Supplemental Fig. S2 and S3). These data show that UGTs with structural  
236 similarities, capable of catalyzing the first glucosylation step of aroma diglycosides  
237 such as  $\beta$ -primeverosides are widely conserved in various plant lineages.

238

### 239 **Purification of *C. sinensis* UGT catalyzing the second xylosylation step**

240 To identify the UGT that is responsible for the second step (6-*O*-xylosylation for  
241 glucose moiety of aroma monoglucoside) which is the conversion of aroma glucosides  
242 to aroma  $\beta$ -primeverosides, xylosyltransferase from young tea leaves was purified based  
243 on the xylosylation activity using geranyl-glc as a substrate at each step. The  
244 purification of the xylosyltransferase through seven purification steps resulted in  
245 13.0-fold enrichment (Supplemental Table S2). Protein purity was assessed by  
246 SDS-PAGE, followed by silver staining (Supplemental Fig. S4). Each excised protein  
247 band was subjected to LC-MS/MS to determine the partial amino acid sequence. *De*  
248 *novo* analysis (PEAKS Software, [www.bioinform.com](http://www.bioinform.com)) identified three peptide sequences  
249 (Fig. S 5A). Using the three peptide sequences (FPEVEKVELEEALPK,  
250 GLVVEGWAPQAR, and EEIEEIAHGLELSMVNFIWVVRFPVEK) obtained from a  
251 single protein band, the corresponding cDNA was surveyed by a tBLASTn search in a  
252 *C. sinensis* EST database constructed by 454 GS-FLX (Roche) (Ohgami et al. 2014).  
253 Contig134, encoding a partial UGT gene, was identified as the most likely candidate  
254 gene. A cDNA clone was isolated, carrying the sequence of contig134 in a

255 1362-bp ORF encoded a polypeptide of 453 amino acid residues (calculated M.W:  
256 51.3 kDa). The encoded polypeptide was named CsGT2, which was assigned as  
257 UGT94P1 by the committee responsible for naming UDP-glucuronosyltransferases  
258 (Mackenzie et al., 1997).

259

### 260 **Biochemical characterization of the xylosyltransferase**

261 To test whether CsGT2 catalyzes the xylosylation of aroma glucosides into  
262 aroma  $\beta$ -primeverosides (Fig. 4A), we performed heterologous expression of CsGT2 in  
263 *E. coli* (Supplemental Fig. S6A) and *in vitro* enzymatic assays with recombinant CsGT2,  
264 UDP-xylose as a sugar donor and geranyl-glc as a sugar acceptor. Figure 4B shows that  
265 CsGT2 produced a new peak with a retention time at 4.9 min. This peak was identical to  
266 the authentic geranyl-pri, which was structurally determined to be xylosylated at the  
267 C-6' position of the glucoside moiety by nuclear magnetic resonance spectroscopy  
268 (NMR) (Guo et al, 1993). These results demonstrate that CsGT2 specifically catalyzes  
269 the xylosylation toward the C-6' position of geranyl-glc. The  $V_{\max}$  and estimated  
270 apparent  $K_m$  values of CsGT2 were determined to be  $60.0 \pm 4.8$  nkat  $\text{mg}^{-1}$  protein and  
271  $78.1 \pm 19.6$   $\mu\text{M}$ , respectively (Table I). The substrate specificity of CsGT2 was  
272 determined using four aroma glucosides and one non-natural glucoside as sugar  
273 acceptors. CsGT2 was active toward all four aroma glucosides with the following  
274 relative activities: geranyl-glc (100%), 2-phenylethyl  $\beta$ -D-glucopyranoside (2PE-glc)  
275 (16%), linalyl-glc (12%), and eugenyl-glc (2%), but not toward the non-natural  
276 *p*-nitrophenyl  $\beta$ -D-glucopyranoside (*p*NP-glc) (Fig. 4C). It is important to mention that

277 CsGT2 did not exhibit any activity toward monoterpene alcohols (volatile aglycones).

278           On the other hand, investigation of the specificity of CsGT2 toward various  
279 sugar donors using geranyl-glc as a sugar acceptor revealed that CsGT2 preferentially  
280 used UDP-xylose (100%) as a sugar donor, while a weak activity was detected with  
281 UDP-glucose (30%) and no apparent activity for UDP-glucuronic acid or  
282 UDP-galactose (Fig. 4D). These results demonstrate that CsGT2 preferentially catalyzes  
283 the xylosylation of aroma glucosides, leading to the formation of aroma  
284  $\beta$ -primeverosides.

285

## 286 **Homology modeling and mutagenesis analysis of CsGT2**

287           The sugar donor specificity of UGTs is dictated by a small number of amino  
288 acid residues (Osmani et al., 2008, Noguchi et al., 2009, Ono et al. 2010a). The residues  
289 are located in three distinct domains: N-terminal, middle, and C-terminal (PSPG-box)  
290 (Sayama et al., 2012). To gain insights into the molecular mechanism of UDP-xylose  
291 specificity of CsGT2, we constructed a structural model of CsGT2 by homology  
292 modeling (Discovery Studio 3.5, Accelrys). The crystal structures of the  
293 glycosyltransferases, At\_UGT72B1 (PDB code: 2vce) and Mt\_UGT85H2 (PDB code:  
294 2pq6), were selected as templates for their similarities to CsGT2. In addition, we used  
295 the three-dimensional (3D) structure of the sugar donor, UDP-2F-Glc, and the crystal  
296 structure of grape UDP-glucose:flavonoid 3-*O*-glycosyltransferase VvGT1 (PDB code:  
297 2c1z) (Offen et al., 2006). In the constructed homology model, Ile141 was identified as  
298 a candidate residue for the control of sugar donor specificity because it is located

299 proximal to UDP-xylose (Fig. 5A and Fig. 5E). This unique Ile141 was found to be  
300 conserved in two xylosyltransferases specific for flavonoid glycosides, kiwi F3GGT1  
301 (Ile136) and *Arabidopsis* UGT79B1 (Ile142) (Montefiori et al., 2011,  
302 Yonekura-Sakakibara et al., 2012) (Table II). In contrast, various amino acid residues  
303 occupy this position in other structurally similar glycosyltransferases, including  
304 *Ipomoea purpurea* UGT79G16 (Thr138), *Sesamum indicum* UGT94D1 (Ser140),  
305 *Veronica persica* UGT94F1 (Ala144), and *Solanum lycopersicum* Nonsmoky  
306 glycosyltransferase 1 (Sl\_NSQT1) (Val145) (Table II) (Morita et al., 2005, Noguchi et  
307 al., 2008, Ono et al., 2010b, Tikunov et al., 2013). To assess the functional relevance of  
308 Ile141 for the specificity toward UDP-xylose, a CsGT2-I141S mutant was generated by  
309 site-specific mutagenesis, in which Ile141 was replaced by a Ser residue. CsGT2-I141S  
310 was heterologously expressed in *E. coli* (Supplemental Figure S6B). Compared with  
311 wild-type CsGT2, the mutant exhibited significantly lower activity with UDP-xylose  
312 but higher activity with UDP-glucose (Fig. 5C and Fig. 5D). These experiments  
313 identified Ile141 as the crucial residue responsible for the sugar donor specificity of  
314 CsGT2 for UDP-xylose.

315

### 316 **Gene expression and phylogenetic analysis of CsGT1 and CsGT2**

317 The tissue specificity of the glycosylation of volatiles was assessed by  
318 quantitative real-time polymerase chain reaction (qRT-PCR) performed with specific  
319 organs of *C. sinensis* (leaves from young to fully mature stage, stems, roots and  
320 flowers). Both *CsGT1* and *CsGT2* were highly expressed in young leaves, where  $\beta$ -PD

321 was also found to be highly expressed (Fig. 6). During leaf maturation, the expression  
322 of *CsGT1* and *CsGT2* decreased, which is consistent with the accumulation profile of  
323  $\beta$ -primeverosides (Fig. 2).

324           Sequence analysis indicated that *CsGT1* and *CsGT2* only share 27% amino  
325 acid identity. Phylogenetic analysis showed that *CsGT1* and *CsGT2* belong to different  
326 clades, OG2 and OG8, respectively (Yonekura-Sakakibara and Hanada, 2011) (Fig. 7).  
327 *CsGT1* showed high similarity to cassava UGT85K4, and UGT85K5 involved in the  
328 biosynthesis of cyanogenic glucosides (Kannangara et al., 2011). In contrast, *CsGT2*  
329 constitutes a new member of the so-called “sugar–sugar UGT” or glycoside-specific  
330 glycosyltransferase (GGT) group that specifically catalyzes glycosylation at the sugar  
331 moiety of various phytochemical glycosides (but not aglycones), including the morning  
332 glory Dusky (UGT79G16), and tomato NSGT1 involved in the glucosylation of  
333 anthocyanin and volatile glycosides, respectively (Morita et al., 2005, Tikunov et al.,  
334 2013).

335

### 336 **Tissue localization of a selected aroma glycosidic precursor**

337           The preferential expression of the two *CsGTs* in young leaves is consistent  
338 with the accumulation of  $\beta$ -primeverosides in plant tissue (Fig. 2). To gain further  
339 insights into the biological role of aroma glucosides and aroma  $\beta$ -primeverosides, MS  
340 imaging analysis was conducted to localize the metabolites in the young leaves at the  
341 cellular level. The specific signals, with a molecular weight of  $m/z$  417, 284 and 340  
342 corresponding to hexenyl-pri, hexenyl-glc and geranyl-glc respectively, were

343 preferentially detected in the epidermal layer of *C. sinensis* leaves, indicating a highly  
344 regulated distribution of the aroma glycosidic precursor (Supplemental Fig. S7).

345

346

## 347 **Discussion**

### 348 **CsGT1 and CsGT2 catalyze the two glycosylation steps of volatile monoterpenes**

### 349 **and alcohols**

350           The broad substrate specificity of CsGT1 for sugar acceptors substantiates the  
351 structural diversity of the  $\beta$ -primeverosides of monoterpenes and primary alcohols  
352 known to accumulate in leaves of *C. sinensis* (Fig. 2). CsGT1 belongs to the UGT85  
353 family and shows similarities in structure and function to kiwi AdGT4 and grape  
354 VvGT14, VvGT16, VvGT17, and VvGT19 that were recently shown to catalyze the  
355 glucosylation of small terpenes and primary alcohols that are accumulated as glycosides  
356 in ripe kiwi and grapes (Bönisch et al., 2014a; Bönisch et al., 2014b; Yauk et al., 2014).  
357 It is noteworthy that AdGT4, VvGT14, and VvGT16 also have broad substrate  
358 specificities toward volatile sugar acceptors (Bönisch et al., 2014a; Yauk et al., 2014).  
359 Taken together, our data and these studies support the notion that the machinery behind  
360 the glucosylation of monoterpenes and primary alcohols is fairly conserved among  
361 phylogenetically discrete various plant species (tea, kiwi, grapevine, Arabidopsis,  
362 snapdragon, and sweet potato) (Supplemental Fig. S2 and Supplemental Fig. S3)  
363 (Bönisch et al., 2014a; Yauk et al., 2014). The estimated apparent  $K_m$  of CsGT1 for  
364 geraniol (44  $\mu$ M) was comparable to those of other volatile UGTs isolated from kiwi

365 and grape (AdGT4 for (*Z*)-3-hexenol: 57.0  $\mu$ M; VvGt14, VvGT15a, VvGT15b,  
366 VvGT15c, and VvGT16 for citronellol: 9, 29, 55, 20, and 108  $\mu$ M, respectively). These  
367 UGTs were found highly expressed in young tea leaf, ripe kiwi, and grape where  
368 aroma glucosides are dominantly accumulated. Furthermore, the concentration of  
369 substrates, (*Z*)-3-hexenol in ripe kiwi, citronellol in grape, and geraniol in young tea  
370 leaves were determined to be at least 0.8, 1.0, and 5.8 mM, respectively (Bönisch et al.,  
371 2014a; Yauk et al., 2014). Their relatively lower substrate specificity toward volatile  
372 sugar acceptors, compared to those of previously characterized non-volatile UGTs,  
373 might reflect their promiscuous biochemical nature, recognizing structurally diverse  
374 substrates. Taken together, UGT85-related enzymes play a role in the formation of  
375 aroma glucosides in various plants.

376 On the other hand, CsGT2 was identified as a novel UGT that specifically  
377 catalyzes 6-*O*-xylosylation of the sugar moiety of aroma monoglucosides, the second  
378 step of glycosylation in the biosynthesis of  $\beta$ -primeverosides (Fig. 4A). CsGT2 belongs  
379 to the GGT cluster (OG8), with regio-specificity for the C-2 hydroxy or C-6 hydroxy  
380 group of sugar moieties of various phytochemical glycosides (Noguchi et al., 2008).  
381 CsGT2 is phylogenetically related to tomato NSGT1 (Sl\_NSGT1), which catalyzes the  
382 third 2-*O*-glucosylation of volatile-derived diglycosides (Tikunov et al., 2013) (Fig. 7).  
383 The majority of UGTs within this GGT cluster catalyze sugar–sugar glycosylation of  
384 various phytochemicals e.g., flavonoids, triterpenoids, and lignans (Morita et al., 2005;  
385 Noguchi et al., 2008; Sawada et al., 2005; Shibuya et al., 2010; Yonekura-Sakakibara et  
386 al., 2014). Therefore, it is conceivable that an ancestral GGT has adapted to

387 accommodate structurally diverse specialized metabolites that often exist only in  
388 particular plant lineages while maintaining its unique regio-specificity for the sugar  
389 moiety. The biochemical activities of GT1 and GT2 suggest their participation  
390 in the biosynthesis of aroma  $\beta$ -primeverosides in tea plants.

391

### 392 *Sugar donor specificity of CsGT2 for UDP-xylose*

393 UGTs usually show exclusive sugar donor specificity, which is determined by  
394 a small number of amino acid residues proximal to the bound sugar donor in the  
395 substrate binding pocket (Noguchi et al. 2009; Ono et al., 2010a; Osmani et al., 2008).  
396 These crucial residues for sugar donor specificity are located in three distinguishable  
397 regions such as the N-terminal, middle, and C-terminal regions (Sayama et al., 2012).  
398 Mutagenesis experiments (Ile141→Ser141) revealed that the unique Ile141 of CsGT2  
399 located in the middle region is a residue determining the specificity toward UDP-xylose.  
400 The fact that the CsGT2-I141S mutant had considerably higher specificity for  
401 UDP-glucose supports the notion that the xylosyltransferase evolved from a  
402 glucosyltransferase by the acquisition of the crucial Ile141. It should be noted that an Ile  
403 residue, which corresponds to Ile141 in CsGT2, is also present in the mid-region of the  
404 two flavonoid 3-*O*-glycoside:2''-*O*-xylosyltransferases (kiwi Ac\_F3GGT1 and  
405 Arabidopsis UGT79B1), but not in Arabidopsis flavonol  
406 3-*O*-glucoside:2''-*O*-glucosyltransferase (UGT79B6) (Table II) (Yonekura-Sakakibara et  
407 al., 2012, 2014). These data suggest that the recognition mechanism for UDP-xylose by  
408 CsGT2 is similar to those of xylosyltransferases from kiwifruit and Arabidopsis,

409 although these two UGTs specifically recognize flavonoid glycosides as their sugar  
410 acceptors.

411           The hydrophobic bulky side chain of the Ile residue possibly contributes to  
412 the unique sugar donor specificity of CsGT2 for UDP-xylose, by hindering the access of  
413 sugar donors with a functional group at the C6 position (UDP-glucose, UDP-galactose,  
414 and UDP-glucuronate) to the substrate pocket of GGT xylosyltransferases. In contrast,  
415 the CsGT2-I141S mutant showed a preference for UDP-glucose instead of UDP-xylose,  
416 as a sugar donor (Fig. 5C and Fig. 5D). Therefore, the hydroxy group of Ser141 could  
417 contribute to the recognition of UDP-glucose via a hydrogen bond with the C6 hydroxy  
418 group of UDP-glucose (Fig. 5B and Fig. 5F). Conversely, the fact that the CsGT2-I141S  
419 mutant failed to use UDP-xylose as a sugar donor suggests that the hydroxy group of  
420 the side chain of the Ser141 prevents binding to UDP-xylose because of a lack of the C6  
421 functional group in the sugar donor binding pocket, probably via hydrophilic properties.  
422 The bulky side chain of the Ile may structurally be required to form an appropriate sugar  
423 donor-binding pocket for UDP-xylose (Fig. 5G and Fig. 5H).

424           Soybean Sg-1<sup>a</sup> (Gm\_UGT73F4) of the OG1 cluster is a xylosyltransferase  
425 involved in the biosynthesis of triterpenoid soyasaponins but has no Ile residue  
426 corresponding to Ile141 of CsGT2. Instead, Sg-1<sup>a</sup> has a unique Ser residue (Ser138) in  
427 the middle domain found essential for its sugar donor specificity towards UDP-xylose  
428 (Sayama et al., 2012). Therefore, our findings indicate that the typical sugar donor  
429 specificity of CsGT2 and Sg-1<sup>a</sup> for UDP-xylose results from convergent evolution  
430 because i) Sg-1<sup>a</sup> and CsGT2 are phylogenetically remote GGTs specialized for

431 structurally different substrates (glycosides of large triterpenes vs. small volatile  
432 monoterpenes, respectively) but with the same sugar donor specificity for UDP-xylose,  
433 and ii) replacement of Ile141 with a Ser residue in the CsGT2-I141S mutant resulted in  
434 a significant decrease in specificity for UDP-Xylose (Fig. 5C), whereas replacement of  
435 Ser138 with Gly residue in Sg-1<sup>a</sup> (S138G) converted the xylosylating activity into a  
436 glucosylating activity (Sayama et al., 2012). These findings not only highlight the  
437 plasticity of sugar donor specificity of UGTs, but demonstrate that the metabolic  
438 specialization is a consequence of the lineage-specific differentiation of the enzymes.

439

#### 440 *Putative physiological roles of CsGT1 and CsGT2 in volatile metabolism*

441 The predominant gene expression of the two *CsGTs* into young tea leaves, together  
442 with the localization of geranyl-pri, 2PE-pri, and hexenyl-pri in young leaves of *C.*  
443 *sinensis*, suggests that these VOCs play physiological roles in this tissue. Given that  
444 VOCs are chemical defense precursors against fungi and herbivores, *CsGTs* and  $\beta$ -PD  
445 play vital roles in the storage and release of *C. sinensis* VOCs, respectively. Based on  
446 these data, we propose a plausible volatile metabolism in tea plant where tissue damage  
447 caused by herbivores would allow geranyl-pri to encounter extracellular  $\beta$ -PD, resulting  
448 in the rapid release of geraniol into the air, without *de novo* biosynthesis (Fig. 8). It is of  
449 particular interest to reveal the sophisticated defense system and whether the molecular  
450 evolution of the second enzyme *CsGT2* is coupled to the evolution of  $\beta$ -PD, which  
451 specifically hydrolyzes diglycosides but is inactive against monoglycosides (Mizutani et  
452 al., 2002).

453 Linalyl-pri dominantly accumulated in stems, while geranyl-pri accumulated in  
454 leaves, (Fig. 2). However, CsGT1 exhibited less activity towards linalool compared to  
455 geraniol while CsGT2 showed weaker activity for linalyl-glc than geranyl-glc.  
456 Moreover, both CsGT1 and CsGT2 were predominantly expressed in young leaves  
457 rather than stem (Fig. 6). These data suggest that geraniol and linalool are separately  
458 biosynthesized in leaves and stems of tea plants, respectively and that members of an  
459 unknown class of glucosyltransferases specifically catalyze glucosylation of linalool  
460 and linalyl-glc. This notion is supported by recent reports that kiwi AdGT4 and grape  
461 VvGT14, which are UGT85-class glucosyltransferases for volatiles, also showed little  
462 enzyme activity towards linalool in contrast to their preferred substrate, geraniol  
463 (Bönisch et al., 2014b; Yauk et al., 2014). Since CsGT1, AdGT4 and VvGT14  
464 commonly exhibited substrate preference to the primary alcohols geraniol and  
465 (Z)-3-hexenol over the tertiary alcohol linalool, there might be a structural feature of  
466 substrate recognition that is shared by these enzymes. All together, these data suggest  
467 the possibility that the molecular machinery for volatile glycosides could be distinct in  
468 leaves and stems of the tea plant.

469 Various plant species use the glycosylation of specialized metabolites to increase  
470 their solubility, which constitutes a primary strategy to accumulate metabolites  
471 operating as defensive compounds against herbivores and pathogens. At the same time,  
472 glycosylation of the metabolites facilitates the transport of the resulting glycosides to  
473 specific compartments, where they are stored separately from hydrolyzing enzymes  
474 glycosidases. For instance, tissue damage by insect attack or pathogen infection triggers

475 the production of the phytotoxin 4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) and its  
476 7-methoxy derivative benzoxazinoid by a  $\beta$ -glucosidase that hydrolyzes the stored form  
477 DIMBOA-glucoside (Frey et al., 2009). Similarly, glucosinolates  
478 ( $\beta$ -thioglucoside-*N*-hydroxysulfates) are precursors of the isothiocyanates and nitriles  
479 known as “mustard oil bomb” against insects. These glucosinolates are  
480 compartmentalized into S-cells, whereas the myrosinases that hydrolyze glucosinolates  
481 are only expressed in myrosin cells (Koroleva et al., 2010). Tissue damage by insects  
482 elicits the hydrolysis of glucosinolates by myrosinases and the accumulation of  
483 isothiocyanates and nitriles.

484 On the other hand, volatile alcohols and monoterpenes are mainly stored in the  
485 form of diglycosides. Interestingly, various sugars (i.e., xylose, arabinose, apiose, and  
486 rhamnose) could be attached to monoglycosides of VOCs as a second sugar molecule  
487 (Tikunov et al., 2013, Bönisch et al., 2014b). This observation suggests that the second  
488 sugar moiety of the glyco-conjugated VOCs further increases stability that leads to the  
489 accumulation of such glyco-conjugated VOCs. Calculation of the cLogP (low value  
490 indicates high water solubility) revealed that the two sequential glycosylation reactions  
491 supporting the conversion of geraniol into geranyl-pri are associated with a stepwise  
492 increase in hydrophilicity, from 2.97 to 2.00 (geranyl-glc) and from 2.00 to 0.46  
493 (geranyl-pri). The cLogP value significantly depends on the sugar type at the  
494 non-reducing end, as well as the sugar number (Tsukada et al. 2006). Furthermore, the  
495 existence of exoglycosidases that cleave disaccharide primeverose into glucose and  
496 xylose has not been established in tea or other plants. Therefore, the xylosylation of

497 geranyl-glc by CsGT2 shown in this study should substantially contribute to the  
498 increase in water solubility and the endurance against exoglycosidases. In addition, the  
499 increase in water solubility of geraniol through glycosylation may be related to the fact  
500 that geraniol and other monoterpene alcohols exhibit high apoptosis-inducing activity in  
501 plant cells (Izumi et al., 1999). This toxic nature of geraniol necessitates precise control  
502 of its biosynthesis and conversion into mono- and diglycosides by CsGT1 and CsGT2,  
503 respectively for the accumulation of defensive geranyl-pri upon potential attacks by  
504 herbivores.

505

## 506 **Conclusion**

507  $\beta$ -Primeverosides are the most abundant form of aroma diglycosides in *C. sinensis*, and  
508 they are commercially and physiologically important for tea aroma quality in dietary  
509 beverages, and for chemical defense against herbivores in tea plant. Here we  
510 demonstrated through metabolic profiling of aroma glycosides in the plant, the  
511 enzymatic characterization and the transcript analysis that CsGT1 and CsGT2 catalyze  
512 the sequential glucosylation and xylosylation of aromas, respectively, leading to  
513 production of aroma  $\beta$ -primeverosides. The transcripts of the two *CsGTs* and  $\beta$ -*PD* in  
514 young tea leaves, together with the localization of geranyl-pri in epidermal cells of  
515 young tea leaves, strongly support the potent physiological role of VOCs in chemical  
516 defense primarily in this tender tissue. Here we report identification and  
517 characterization of molecular machineries for the biosynthesis of  $\beta$ -primeverosides of  
518 tea aroma. Our findings provided not only molecular insights into volatile metabolism

519 in *C. sinensis*, but also crucial molecular tools for controlling tea aroma quality and for  
520 understanding of a sophisticated chemical defense system elaborated during plant  
521 evolution.

522

523

## 524 **Materials and Methods**

### 525 **Chemicals**

526 *p*NP-pri, eugenyl-pri, and 2PE-pri were kindly provided by Prof. Usui and Dr.  
527 Murata (Shizuoka University, Japan) (Murata et al., 1999). Geranyl-pri and linalyl-pri  
528 were prepared as described (Guo et al., 1993, 1999). *p*NP-glc, cyanidin 3-*O*-glucoside  
529 chloride, and quercetin 3-*O*-glucoside were purchased from Sigma-Aldrich  
530 (www.sigmaaldrich.com). Other glucosides were chemically synthesized using the  
531 procedure for 2PE-glc as previously described (Ma et al., 2001). UDP-xylose was  
532 obtained from CarboSource Services at the University of Georgia (Athens, USA). Other  
533 chemicals were purchased from Sigma-Aldrich and Wako Pure Chemical Industries  
534 (www.wako-chem.co.jp/english/).

535

### 536 **Plant Materials**

537 Young leaves of *C. sinensis* var. *sinensis* cv. Yabukita were harvested at the  
538 Center for Education and Research of Field Sciences, Shizuoka University and at Tea  
539 Research Center, Shizuoka Prefectural Research Institute of Agriculture and Forestry  
540 (Shizuoka, Japan). The young and mature developmental stages of leaves and stems

541 were defined as follows: young leaves, which are the first, second, and third leaves  
542 (plucking part for high grade tea products); mature leaves, which are the fourth and fifth  
543 leaves; stem, which are green young non-lignified part between the 3rd leaf and 5th  
544 leaf; branch, which are brown lignified old part (Fig. S8). All plant materials were  
545 quickly frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  prior to use.

546

#### 547 **Quantification of endogenous aroma glycosides in *C. sinensis*.**

548 Three point five grams (fresh weight) of leaves and stems at two  
549 developmental stages (young and mature) of *C. sinensis* were finely crushed in a tissue  
550 mill (TK-AM5, www.e-taitec.com) suspended in 80% methanol (30 ml) and filtrated.  
551 The filtrate was concentrated *in vacuo* and separated with *n*-hexane. The aqueous layer  
552 was concentrated *in vacuo*, dissolved in distilled water and purified with Oasis HLB  
553 (3cc, Waters; www.waters.com). The glycosidic fractions were concentrated *in vacuo*  
554 and dissolved in distilled water prior to GC-MS analysis. Detailed procedures for  
555 LC-MS analysis condition are given in Supplemental Materials and Methods S1.

556

#### 557 **Identification of *CsGT1***

558 Full-length cDNA clone of *CsGT1* was isolated by a screen with full-length  
559 DIG-labeled mixed probes for At\_UGT85A, in a cDNA library derived from *C. sinensis*  
560 previously described (Mizutani et al., 2002). Library screening was performed under a  
561 low stringency condition as described in Yonekura-Sakakibara et al. (2000), and  
562 Noguchi et al. (2008, 2009). The screening probes of At\_UGT85A were DIG-labeled by

563 PCR using gene-specific oligonucleotides (Supplemental Table S3). More than 50  
564 positive clones were obtained in approximately 5,000,000 plaques after two rounds of  
565 screening. The cDNA fragments of positive clones were sequenced by conventional  
566 primer walking method with BigDye-terminator version 3.1 cycle sequencing kit (Life  
567 Technologies; www.lifetechnologies.com). Among these clones, a full-length cDNA for  
568 the *CsGT1* gene was identified by blastx search based on sequence similarities with  
569 *At\_UGT85A3*.

570

571 **Enzyme purification of a xylosyltransferase specific for monoglucoside-bound**  
572 **volatiles**

573 All procedures were performed at 4°C. The composition of the purification  
574 buffers and solutions for peptide analysis is described in Supplemental Table S4. Tea  
575 leaves (100 g) were finely chopped, crushed in a tissue mill (Taitec) suspended in 100  
576 ml buffer A and centrifuged (20,000 ×g; 30 min). The supernatant was collected, and  
577 ammonium sulfate was added up to 30% saturation. The mixture was centrifuged  
578 (20,000 ×g; 30 min), the supernatant collected, and ammonium sulfate was added up to  
579 70% saturation, followed by another centrifugation (20,000 ×g; 30 min). The pellet was  
580 dissolved in buffer B and dialyzed against buffer B for complete removal of the  
581 ammonium sulfate. Purification of active fractions was performed using the following  
582 columns;: HiTrap DEAE FF (5 ml, GE Healthcare; www.gelifesciences.com), HiTrap Q  
583 FF (5 ml), Macro-prep Ceramic Hydroxyapatite Type III (5 ml, Bio-Rad;  
584 www.bio-rad.com), HiTrap Blue HP (1 ml), and Mono Q 5/50 GL (1 ml). At each

585 purification step, the eluted fractions were tested for xylosyltransferase activity toward  
586 geranyl-glc by LC-MS, and the active fractions were pooled before the next purification  
587 step. Detailed procedures are given in Supplemental Materials and Methods.

588

#### 589 **Identification of the peptide sequence of CsGT2 by LC-MS/MS**

590 Purified proteins were separated by SDS-PAGE and stained by Silver staining,  
591 the major bands were excised from the gel and destained with solution F. Proteins in the  
592 gel pieces were reduced and alkylated in solutions G and H, respectively, followed by  
593 solution H. After serial washes with the wash solution F and acetonitrile, the proteins  
594 were digested with trypsin (Promega; www.promega.com). The tryptic peptides were  
595 extracted from the gel pieces with solution I, and the extract was concentrated *in vacuo*.  
596 The concentrated solution was centrifuged (20,000 ×g; 10 min) and the supernatant was  
597 analyzed by LC-MS/MS (Supplemental Materials and Methods). All peptide mass data  
598 were analyzed using the Peaks software (Bioinformatics Solutions).

599

#### 600 **Identification of full-length CsGT2 cDNA**

601 Since contig134 had a partial ORF of *CsGT2*, the full-length sequence of  
602 *CsGT2* cDNA was obtained from fresh young tea leaves using gene-specific  
603 *CsGT2*-Race-FW and *CsGT2*-Race-RV oligonucleotides, using the SMARTer cDNA  
604 RACE cDNA amplification kit (Clontech; www.clontech.com) and PrimeStar HS  
605 polymerase (TAKARA BIO; www.takara-bio.com), according to manufacturer's  
606 instructions. The amplified products were gel purified and ligated into pJET 1.2 vector

607 using the CloneJET kit (Thermo Fisher Scientific; [www.thermoscientificbio.com](http://www.thermoscientificbio.com)).

608

### 609 **Heterologous expression of recombinant UGT proteins**

610 Total RNA was extracted from fresh young leaves of *C. sinensis* using  
611 RNeasy Plant Mini kit (Qiagen; [www.qiagen.com](http://www.qiagen.com)), according to the manufacturer's  
612 instructions. cDNA was reverse-transcribed from 1 µg of total RNA with SuperScript III  
613 (Life Technologies). Full-length cDNA fragments of *CsGT1* and *CsGT2* genes were  
614 amplified from cDNA of *C. sinensis* cv. Yabukita by RT-PCR, using gene-specific  
615 oligonucleotides (Supplemental Table S3). *In vitro* mutagenesis of the *CsGT2* gene was  
616 performed by recombinant PCR with specific mutagenic oligonucleotides  
617 (Supplemental Table S3) as previously described (Noguchi et al. 2009). For their  
618 expression in *E. coli*, the generated amplicons were ligated into the pENTR/D-TOPO  
619 vector (Life Technologies), and the sequence was verified. They were subcloned into the  
620 pET15b expression vector (Merck Millipore; [www.merckmillipore.com](http://www.merckmillipore.com)) and  
621 transformed into *E. coli* BL21 (DE3) (TOYOBO; [www.toyobo-global.com](http://www.toyobo-global.com)). The  
622 recombinant proteins produced by *E. coli* BL21 were quantified by Bradford method  
623 (Bradford, 1976) with BSA as the standard, and separated by SDS-PAGE. The  
624 expressed recombinant proteins were immunologically detected in the gels by western  
625 blotting as described previously (Sayama et al., 2012).

626

### 627 **Enzyme assay of CsGT1 and CsGT2**

628 For relative activity assays of CsGT1 and CsGT2, the enzymatic reaction

629 mixture (50  $\mu$ L) consisted of 100 mM sugar acceptor, 2 mM sugar donor, 50 mM  
630 potassium phosphate buffer, pH 8.0, and enzyme. The enzyme assays were initiated  
631 after pre-incubation of the mixture without the enzyme at 30°C for 5 min. After  
632 incubation at 30°C for 10 min, the reaction was stopped by the addition of 50  $\mu$ L of  
633 ice-cold methanol. The same assay conditions were used for determination of the kinetic  
634 parameters of CsGT1 and CsGT2 except that the sugar acceptor geranyl-glc was used  
635 instead of geraniol and at six different concentrations from 1.25-250  $\mu$ M. The enzymatic  
636 products were analyzed by LC-MS analysis (Materials and Methods S1). The apparent  
637  $K_m$  and  $V_{max}$  values for each sugar donor and the sugar acceptor (geraniol) were  
638 determined by a saturating substrate concentration by fitting the initial linear velocity  
639 data to a Michaelis-Menten equation using nonlinear regression analysis in the  
640 Kaleidagraph software ([www.synergy.com](http://www.synergy.com)).

641

#### 642 **Homology modeling**

643 The construction of a 3D model according to CsGT2 was conducted using the  
644 Discovery Studio (DS) 3.5 (<http://accelrys.com/>). The crystal structures of  
645 At\_UGT72B1 (PDB code: 2vce), Mt\_UGT85H2 (PDB code: 2pq6), and VvGT1 (PDB  
646 code: 2c1z), and the 3D structure of the sugar donor UDP-2-F-Glc, were used as  
647 templates. The initial structure of CsGT2 was constructed using the multiple homology  
648 modeling protocols of the DS3.5 MODELER module. The resulting CsGT2 structure  
649 was inserted in the UDP-2F-Glc bound in VvGT1, and the sugar moiety of UDP-2F-Glc  
650 was replaced with a xylose. Structure optimization of the initial complex model

651 (CsGT2-UDP-Xyl) was performed using molecular mechanics and dynamics simulation  
652 with the CHARMM force field of DS3.5. On the other hand, the 3D structure of the  
653 CsGT2 (I141S) mutant was first constructed by replacing Ile141 with a serine residue.  
654 After insertion of UDP-2F-Glc bound to VvGT1, the fluoride atom of UDP-2F-Glc was  
655 converted to a hydroxy group to generate the model complex, CsGT2 (I141S)-UDP-Glc,  
656 which was optimized by the same procedure used for CsGT2-UDP-Xyl.

657

#### 658 **Quantitative real-time PCR (qRT-PCR) of *CsGTs* and $\beta$ -PD**

659 The qRT-PCR was performed as previously described (Noguchi et al., 2008).  
660 In brief, the cDNA was prepared from multiple organs and tissues of *C. sinensis*. The  
661 *CsGT1*, *CsGT2*,  $\beta$ -PD, and *18S rRNA* were quantified by real-time PCR using specific  
662 primers (Supplemental Table S3) and a Power SYBR Green PCR kit (Qiagen) on a 7500  
663 Real-Time PCR system (Life Technologies). The transcription levels were quantified  
664 using the ddCT threshold cycle method, and normalized to the expression level of an  
665 internal standard (*18S rDNA*). The results are presented as the means  $\pm$  SE of three  
666 independent experiments.

667

#### 668 **Phylogenetic analysis**

669 The amino acid sequences of UGTs (Supplemental Table S5) were aligned  
670 based on codon position using ClustalW bundled in MEGA6 (Tamura et al. 2013). All  
671 sites containing gaps and missing data were eliminated from the remaining analysis.  
672 Unrooted phylogenetic trees were reconstructed by neighbor-joining methods from the

673 translated amino acid sequences. The neighbor-joining tree was reconstructed by  
674 MEGA6, and the matrix of evolutionary distances was calculated by Poisson correction  
675 for multiple substitutions. The reliability of the reconstructed tree was evaluated by a  
676 bootstrap test for 1,000 replicates.

677

#### 678 **GenBank accession numbers**

679           Sequence data from this article can be found in GenBank/EMBL under the  
680 following accession numbers: UGT85K11 (AB847092) and UGT94P1 (AB847093).

681

682

#### 683 **Supplemental Data Files**

684 **Supplemental Figure S1.** Co-expression analysis of *AtLIS*, *CYP76C3* and *UGT85A3*  
685 by ATTEDII (Gene coexpression database, <http://atted.jp/>) with Ver. C4.1.

686 **Supplemental Figure S2.** Enzymatic activity of CsGT1 homologs for (*Z*)-3-hexenol.

687 **Supplemental Figure S3.** Enzymatic activity of CsGT1 homologs for geraniol.

688 **Supplemental Figure S4.** Purified enzymes catalyzing the second xylosyltransferase  
689 (CsGT2).

690 **Supplemental Figure S5.** Partial peptide sequence of CsGT2.

691 **Supplemental Figure S6.** Recombinant proteins of a series of CsGTs.

692 **Supplemental Figure S7.** Imaging MS of young fresh leaves of *C. sinensis*.

693 **Supplemental Figure S8.** Harvesting individual tissues for quantification of  
694 endogenous aroma glycosides and quantitative real-time PCR (qRT-PCR) of *CsGT1*,

695 *CsGT2* and  $\beta$ -*PD*.

696 **Supplemental Figure S9.** Multiple sequence alignment of protein sequences of  
697 UGT-glucosyltransferase OG2 and OG8 family. The alignment was performed using  
698 ClustalW2.1.

699 **Supplemental Table S1.** Summary of fresh weight and amounts of aroma  
700  $\beta$ -primeverosides in young leaves and mature leaves of *C. sinensis*.

701 **Supplemental Table S2.** Summary of the purification of *CsGT2* from young fresh  
702 leaves of *C. sinensis*.

703 **Supplemental Table S3.** Gene specific primers used for 5', 3'-RACE, amplification of  
704 full-length genes from *C. sinensis*, construction of *CsGT2*-I141S or real-time PCR.

705 **Supplemental Table S4.** Composition of the buffers and solutions for purification and  
706 identification of *CsGT2* protein.

707 **Supplemental Table S5.** GenBank accession numbers used for the construction of the  
708 phylogenetic tree in Figure 7.

709

710 **Supplemental Materials and Methods.** Detailed description of the experimental  
711 procedure in this study.

712

### 713 **Acknowledgments**

714 We thank Professor Emeritus Kanzo Sakata (Kyoto University), Drs. Tsuyoshi  
715 Katsuno (Tea Research Center, Shizuoka Prefectural Research Institute of Agriculture  
716 and Forestry), Masahiro Nakao, Nobuo Tsuruoka, Yoshihide Matsuo, Misa

717 Inohara-Ochiai, Masako Fukuchi-Mizutani, Kensuke Tanaka, Naofumi Yoshida, Hitoshi  
718 Onuki, Yuko Fujimori (Suntory Ltd), Atsushi Nesumi (Tea Research Center,  
719 Makurazaki, Kagoshima Prefecture), Seiji Takahashi, Toru Nakayama (Tohoku  
720 University), Hiroto Kihara, Takao Koedusa (Yamaguchi University), Koichi  
721 Sugimoto and Junji Takabayashi (Kyoto University) for their technical support and  
722 useful discussions. We are grateful to Professor Taichi Usui and Dr. Takeomi Murata  
723 (Shizuoka University) for kindly providing *p*NP-pri, eugenyl-pri, and 2PE-pri. We also  
724 thank Dr. Bjoern Hamberger (Copenhagen University, Denmark) for the critical  
725 proofreading of the manuscript, as well as the Molecular Structure Analysis Section and  
726 Functional Genomics Section (Research Institute of Green Science and Technology,  
727 Shizuoka University) for LC-MS (LTQ Orbitrap Discovery, Thermo Fisher Scientific),  
728 LC-MS/MS (NanoFrontier eLD, Hitachi High-Technology) and GC-MS  
729 (JMS-T100GCV, JEOL)  
730  
731  
732

733 **References**

- 734 **Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel WJ, Verstappen FW,**  
735 **Verhoeven HA, Jongsma MA, Schwab W, Bouwmeester HJ** (2003)  
736 Terpenoid metabolism in wild-type and transgenic Arabidopsis plants. *Plant*  
737 *Cell* **15**: 2866–2884
- 738 **Arimura G, Matsui K, Takabayashi J** (2009) Chemical and molecular ecology of  
739 herbivore-induced plant volatiles: proximate factors and their ultimate  
740 functions. *Plant Cell Physiol* **50**: 911–923
- 741 **Balentine DA, Wiseman SA, Bouwens LC** (1997) The chemistry of tea flavonoids.  
742 *Crit Rev Food Sci Nutr* **37**: 693–704
- 743 **Bradford MM** (1976) A rapid and sensitive method for the quantitation of microgram  
744 quantities of protein utilizing the principle of protein-dye binding. *Anal*  
745 *Biochem* **72**: 248–254
- 746 **Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M, Bitz O, Schwab W** (2014a)  
747 Activity based profiling of a physiologic aglycone library reveals sugar  
748 acceptor promiscuity of family 1 UDP-glucosyltransferases from *Vitis*  
749 *vinifera*. *Plant Physiol* **166**: 23–39
- 750 **Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M, Bitz O, Schwab W** (2014b) A  
751 UDP-Glucose:Monoterpenol Glucosyltransferase Adds to the Chemical  
752 Diversity of the Grapevine Metabolome. *Plant Physiol* **165**: 561–581
- 753 **Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A** (2009) Benzoxazinoid  
754 biosynthesis, a model for evolution of secondary metabolic pathways in plants.

755           Phytochemistry **70**: 1645–1651

756   **Ginglinger JF, Boachon B, Höfer R, Paetz C, Köllner TG, Miesch L, Lugan R,**  
757           **Baltenweck R, Mutterer J, Ullmann P, Beran F, Claudel P, Verstappen F,**  
758           **Fischer MJ, Karst F, Bouwmeester H, Miesch M, Schneider B,**  
759           **Gershenzon J, Ehlting J, Werck-Reichhart D** (2013) Gene coexpression  
760           analysis reveals complex metabolism of the monoterpene alcohol linalool in  
761           Arabidopsis flowers. *Plant Cell* **25**: 4640–4657

762   **Graham HN** (1992) Green tea composition, consumption, and polyphenol chemistry.  
763           *Prev Med* **21**: 334–350

764   **Günata YZ, Bayonove CL, Baumes RL, Cordonnier RE** (1985) The aroma of  
765           grapes: I. Extraction and determination of free and glycosidically bound  
766           fractions of some grape aroma components. *J.Chromatogr A* **331**: 83–90

767   **Guo W, Hosoi R, Sakata K, Watanabe N, Yagi A, Ina K, Luo S** (1994) (S)-linalyl,  
768           2-phenylethyl, and benzyl disaccharide glycosides isolated as aroma  
769           precursors from oolong tea leaves. *Biosci Biotechnol Biochem* **58**:  
770           1532–1534

771   **Guo W, Sakata K, Watanabe N, Nakajima R, Yagi A, Ina K, Luo S** (1993) Geranyl  
772           6-O- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranoside isolated as an aroma precursor  
773           from tea leaves for oolong tea. *Phytochemistry* **33**: 1373–1375

774   **Höfer R, Boachon B, Renault H, Gavira C, Miesch L, Iglesias J, Ginglinger JF,**  
775           **Allouche L, Miesch M, Grec S, Larbat R, Werck-Reichhart D** (2014) Dual  
776           function of the CYP76 family from *Arabidopsis thaliana* in the metabolism of

777           monoterpenols and phenylurea herbicides. PMID: 25082892

778   **Izumi S, Takashima O, Hirata T** (1999) Geraniol is a potent inducer of apoptosis-like  
779           cell death in the cultured shoot primordia of *Matricaria chamomilla*. *Biochem*  
780           *Biophys Res Commun* **259**: 519-522

781   **Kannangara R, Motawia MS, Hansen NK, Paquette SM, Olsen CE, Møller BL,**  
782           **Jørgensen K** (2011) Characterization and expression profile of two  
783           UDP-glucosyltransferases, UGT85K4 and UGT85K5, catalyzing the last step  
784           in cyanogenic glucoside biosynthesis in cassava. *Plant J* **68**: 287–301

785   **Kawakami M, Ganguly SN, Banerjee J, Kobayashi A** (1995) Aroma Composition of  
786           Oolong Tea and Black Tea by Brewed Extraction Method and Characterizing  
787           Compounds of Darjeeling Tea Aroma. *J Agric Food Chem* **43**: 200–207

788   **Kobayashi A, Kubota K, Joki Y, Wada E, Wakabayashi M** (1994) (Z)-3-Hexenyl  
789            $\beta$ -D-glucopyranoside in fresh tea leaves as a precursor of green odor. *Biosci*  
790           *Biotechnol Biochem* **58**: 592–593

791   **Koeduka T, Suzuki S, Iijima Y, Ohnishi T, Suzuki H, Watanabe B, Shibata D,**  
792           **Umezawa T, Pichersky E, Hiratake J** (2013) Enhancement of production of  
793           eugenol and its glycosides in transgenic aspen plants via genetic engineering.  
794           *Biochem Biophys Res Commun* **21**: 73–78

795   **Koroleva OA, Gibson TM, Cramer R, Stain C** (2010) Glucosinolate-accumulating  
796           S-cells in *Arabidopsis* leaves and flower stalks undergo programmed cell  
797           death at early stages of differentiation. *Plant J* **64**: 456–469

798   **Krammer G, Winterhalter P, Schwab M, Schreier P** (1991) Glycosidically bound

799            aroma compounds in the fruits of Prunus species: apricot (*P. armeniaca*, *L.*),  
800            peach (*P. persica*, *L.*), yellow plum (*P. domestica*, *L. ssp. syriaca*) *J. Agric.*  
801            *Food Chem* **39**: 778–781

802    **Kumazawa K and Masuda H** (2002) Identification of Potent Odorants in Different  
803            Green Tea Varieties Using Flavor Dilution Technique. *J Agric Food Chem* **50**:  
804            5660–5663

805    **Lücker J, Bouwmeester HJ, Schwab W, Blaas J, van der Plas LH, Verhoeven HA**  
806            (2001) Expression of Clarkia S-linalool synthase in transgenic petunia plants  
807            results in the accumulation of S-linalyl- $\beta$ -D-glucopyranoside. *Plant J* **27**:  
808            315–324

809    **Ma SJ, Mizutani M, Hiratake J, Hayashi K, Yagi K, Watanabe N, Sakata K** (2001)  
810            Substrate specificity of  $\beta$ -primeverosidase, a key enzyme in aroma formation  
811            during oolong tea and black tea manufacturing. *Biosci Biotechnol Biochem*  
812            **65**: 2719–2729

813    **Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Bélanger A,**  
814            **Fournel-Gigleux S, Green M, Hum DW, Iyanagi T, Lancet D, Louisot P,**  
815            **Magdalou J, Chowdhury JR, Ritter JK, Schachter H, Tephly TR, Tipton**  
816            **KF, Nebert DW** (1997) The UDP glycosyltransferase gene superfamily:  
817            Recommended nomenclature update based on evolutionary divergence.  
818            *Pharmacogenetics* **7**: 255–269

819    **Marlatt C, Ho C, Chien MJ** (1992) Tomato: Studies of aroma constituents bound as  
820            glycosides in tomato. *J Agric Food Chem* **40**: 249–252

821 **Mizutani M, Nakanishi H, Ema J, Ma SJ., Noguchi E, Inohara-Ochiai M,**  
822 **Fukuchi-Mizutani M, Nakao M, Sakata K** (2002) Cloning of  
823  $\beta$ -primeverosidase from tea leaves, a key enzyme in tea aroma formation.  
824 *Plant Physiol* **130**: 2164–76

825 **Mizutani M, Ward E, Ohta D** inventors. March 27, 1997. Plant geraniol/nerol  
826 10-hydroxylase and DNA coding therefor. European Patent Application No.  
827 96931058.0

828 **Montefiori M, Espley RV, Stevenson D, Cooney J, Datson PM, Saiz A, Atkinson RG,**  
829 **Hellens RP, Allan AC** (2011) Identification and characterisation of F3GT1  
830 and F3GGT1, two glycosyltransferases responsible for anthocyanin  
831 biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*). *Plant J* **65**:  
832 106–118

833 **Moon JH, Watanabe N, Ijima Y, Yagi A, Sakata K** (1996) *Cis*- and *trans*-linalool  
834 3,7-oxides and methyl salicylate glycosides and (*Z*)-3-hexenyl  
835  $\beta$ -D-glucopyranoside as aroma precursors from tea leaves for oolong tea.  
836 *Biosci Biotechnol Biochem* **60**: 1815–1819

837 **Moon JH, Watanabe N, Sakata K, Yagi A, Ina K, Luo S** (1994) *Trans*- and  
838 *cis*-linalool 3,6-oxide 6-O- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranosides isolated  
839 as aroma precursors from leaves for oolong tea. *Biosci Biotechnol Biochem*  
840 **58**: 1742–1744

841 **Morita Y, Hoshino A, Kikuchi Y, Okuhara H, Ono E, Tanaka Y, Fukui Y, Saito N,**  
842 **Nitasaka E, Noguchi H, Iida S** (2005) Japanese morning glory dusky

843 mutants displaying reddish-brown or purplish-gray flowers are deficient in a  
844 novel glycosylation enzyme for anthocyanin biosynthesis,  
845 UDP-glucose:anthocyanidin 3-*O*-glucoside-2''-*O*-glucosyltransferase, due to  
846 4-bp insertions in the gene. *Plant J* **42**: 353–363

847 **Murata T, Shimada M, Watanabe N, Sakata K, Usui T** (1999) Practical enzymatic  
848 synthesis of primeverose and its glycoside. *J Appl Glycosci.* 46:431–437

849 **Noguchi A, Fukui Y, Iuchi-Okada A, Kakutani S, Satake H, Iwashita T, Nakao M,**  
850 **Umezawa T, Ono E** (2008) Sequential glucosylation of a furofuran lignan,  
851 (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1  
852 glucosyltransferases. *Plant J* **54**: 415–427

853 **Noguchi A, Horikawa M, Fukui Y, Fukuchi-Mizutani M, Iuchi-Okada A, Ishiguro**  
854 **M, Kiso Y, Nakayama T, Ono E** (2009) Local differentiation of sugar donor  
855 specificity of flavonoid glucosyltransferase in Lamiales. *Plant Cell* **21**:  
856 1556–1572

857 **Obayashi T, Nishida K, Kasahara K, Kinoshita K** (2011) ATTED-II updates:  
858 condition-specific gene coexpression to extend coexpression analyses and  
859 applications to a broad range of flowering plants. *Plant Cell Physiol.*  
860 **52**:213-219

861 **Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford**  
862 **CM, Bowles DJ, Davies GJ** (2006) Structure of a flavonoid  
863 glucosyltransferase reveals the basis for plant natural product modification.  
864 *EMBO J* **25**: 1396–1405

865 **Ohgami S, Ono E, Toyonaga H, Watanabe N, Ohnishi, T** (2014) Identification and  
866 characterization of *Camellia sinensis* glucosyltransferase, UGT73A17: A  
867 possible role in flavonol glucosylation. *Plant Biotech.* **31**: 573–578

868 **Ono E, Homma Y, Horikawa M, Kunikane-Doi S, Imai H, Takahashi S, Kawai Y,**  
869 **Ishiguro M, Fukui Y, Nakayama T** (2010a) Functional differentiation of the  
870 glycosyltransferases that contribute to the chemical diversity of bioactive  
871 flavonol glycosides in grapevines (*Vitis vinifera*). *Plant Cell* **22**: 2856–2871

872 **Ono E, Ruike M, Iwashita T, Nomoto K, Fukui Y** (2010b) Co-pigmentation and  
873 flavonoid glycosyltransferases in blue *Veronica persica* flowers.  
874 *Phytochemistry* **71**: 726–735

875 **Osmani S, Bak S, Imberty A, Olsen CE, Møller BL** (2008) Catalytic key amino acids  
876 and UDP–sugar donor specificity of a plant glucuronosyltransferase,  
877 UGT94B1. *Plant Physiol* **148**: 1295–1308

878 **Pabst A, Barron A, Eti evant P, Schreier P** (1991) Studies on the enzymatic  
879 hydrolysis of bound aroma constituents from raspberry fruit pulp. *J Agric*  
880 *Food Chem* **39**: 173–175

881 **Pattnaik S, Subramanyam VR, Bapaji M, Kole CR** (1997) Antibacterial and  
882 antifungal activity of aromatic constituents of essential oils. *Microbios* **89**:  
883 39–46

884 **Pichersky E and Gershenzon J** (2002) The formation and function of plant volatiles:  
885 perfumes for pollinator attraction and defense. *Curr Opin Plant Biol* **5**:  
886 237–243

- 887 **Roscher R, Herderich M, Steffen JP, Schreier P, Schwab W** (1997)  
888 2,5-dimethyl-4-hydroxy-3(2H)-furanone 6'-O-malonyl- $\beta$ -D-glucopyranoside  
889 in strawberry fruits. *Phytochemistry* **43**: 155–159
- 890 **Sawada S, Suzuki H, Ichimaida F, Yamaguchi MA, Iwashita T, Fukui Y, Hemmi H,**  
891 **Nishino T, Nakayama, T** (2005) UDP-glucuronic acid:anthocyanin  
892 glucuronosyltransferase from red daisy (*Bellis perennis*) flowers. *Enzymology*  
893 and phylogenetics of a novel glucuronosyltransferase involved in flower  
894 pigment biosynthesis. *J Biol Chem* **280**: 899–906
- 895 **Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A,**  
896 **Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S,**  
897 **Tatsuzaki N, Tsukamoto C, Ishimoto M** (2012) The Sg-1  
898 glycosyltransferase locus regulates structural diversity of triterpenoid  
899 saponins of soybean. *Plant Cell* **24**: 2123–2138
- 900 **Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y** (2010) Identification and  
901 characterization of glycosyltransferases involved in the biosynthesis of  
902 soyasaponin I in *Glycine max*. *FEBS Lett* **584**: 2258–2264
- 903 **Sugimoto K, Matsui K, Iijima Y, Akakabe Y, Muramoto S, Ozawa R, Uefune M,**  
904 **Sasaki R, Alamgir KM, Akitake S, Nobuke T, Galis I, Aoki K, Shibata D,**  
905 **Takabayashi J** (2014) Intake and transformation to a glycoside of  
906 (*Z*)-3-hexenol from infested neighbors reveals a mode of plant odor reception  
907 and defense. *Proc. Natl. Acad. Sci. U S A.*, **111**: 7144–7149
- 908 **Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S** (2013) MEGA6:

909 Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology*  
910 *and Evolution* **30**: 2725-2729

911 **Tikunov YM, Molthoff J, de Vos RC, Beekwilder J, van Houwelingen A, van der**  
912 **Hooft JJ, Nijenhuis-de Vries M, Labrie CW, Verkerke W, van de Geest H,**  
913 **Viquez Zamora M, Presa S, Rambla JL, Granell A, Hall RD, Bovy AG**  
914 (2013) Non-smoky glycosyltransferase1 prevents the release of smoky aroma  
915 from tomato fruit. *Plant Cell* **25**: 3067–3078

916 **Tsukada T, Igarashi K, Yoshida M, Samejima M** (2006) Molecular cloning and  
917 characterization of two intracellular  $\beta$ -glucosidases belonging to glycoside  
918 hydrolase family 1 from the basidiomycete *Phanerochaete chrysosporium*.  
919 *Appl Microbiol Biotechnol* **73**: 807–814

920 **Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A,**  
921 **Steinhauser D, Persson S, Provarnt NJ** (2009) Co-expression tools for plant  
922 biology: opportunities for hypothesis generation and caveats. *Plant Cell*  
923 *Environ.* **32**: 1633–1651

924 **Wang D, Yoshimura T, Kubota K, Kobayashi A** (2000) Analysis of glycosidically  
925 bound aroma precursors in tea leaves: I. Qualitative and quantitative analyses  
926 of glycosides with aglycons as aroma compounds. *J Agric Food Chem* **48**:  
927 5411–5418

928 **Yano M, Joki Y., Mutoh, H., Kubota K, Kobayashi A** (1991) Benzyl glucoside from  
929 tea leaves. *Agric Biol Chem* **55**: 1205–1206

930 **Yauk YK, Ged C, Wang MY, Matich AJ, Tessarotto L, Cooney JM, Chervin C,**

931           **Atkinson RG** (2014) Manipulation of flavour and aroma compound  
932           sequestration and release using a glycosyltransferase with specificity for  
933           terpene alcohols. *Plant J* **80**: 317–330

934   **Yonekura-Sakakibara, K, Tanaka Y, Fukuchi-Mizutani M, Fujiwara H, Fukui Y,**  
935           **Ashikari T, Murakami Y, Yamaguchi M, Kusumi T** (2000) Molecular and  
936           Biochemical Characterization of a Novel Hydroxycinnamoyl-CoA:  
937           Anthocyanin 3-*O*-Glucoside-6''-*O*-Acyltransferase from *Perilla frutescens*.  
938           *Plant Cell Physiol* **41**: 495–502

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

941   **Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada, K, Matsuda F,**  
942           **Sugawara, S, Inoue E, Kuromori T, Ito T, Shinozaki K, Wangwattana B,**  
943           **Yamazaki M, Saito K** (2012) Two glycosyltransferases involved in  
944           anthocyanin modification delineated by transcriptome independent  
945           component analysis in *Arabidopsis thaliana*. *Plant J* **69**: 154–167

946   **Yonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi M,**  
947           **Kitajima M, Takayama H, Saito K** (2014) A flavonoid  
948           3-*O*-glucoside:2''-*O*-glucosyltransferase responsible for terminal modification  
949           of pollen-specific flavonols in *Arabidopsis thaliana*. *Plant J* **79**: 769–782

950   **Young H and Paterson VJ** (1995) Characterization of bound flavour components in  
951           kiwi fruit. *J Sci Food Agric* **68**: 257–260

952

953

954 **Figure legends**

955 **Figure 1. Metabolism of volatile organic compounds in *Camellia sinensis*.** A) Photo  
956 of young leaves. (B) Biosynthesis pathway of geranyl  $\beta$ -primeveroside (geranyl  
957 6-*O*- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranoside) from geraniol. CsGT1 (UGT85K11) and  
958 CsGT2 (UGT94P1) are the two glycosyltransferases that are shown to catalyze the  
959 sequential glucosylation and xylosylation of geraniol, respectively, in this work (Bold  
960 arrows).

961

962 **Figure 2. Quantification of aroma monoglycosides and diglycosides in fresh tea**  
963 **leaves** Ten glycosides were used as authentic standards to quantify endogenous  
964 glycoside in young leaves, young stems, mature leaves, and mature stems of *C. sinensis*:  
965 benzyl  $\beta$ -D-glucopyranoside (benzyl-glc), benzyl  $\beta$ -primeveroside (benzyl-pri), geranyl  
966  $\beta$ -D-glucopyranoside (geranyl-glc), geranyl  $\beta$ -primeveroside (geranyl-pri),  
967 (*Z*)-3-hexenyl  $\beta$ -D-glucopyranoside (hexenyl-glu), (*Z*)-3-hexenyl  $\beta$ -primeveroside  
968 (hexenyl-pri), linalyl  $\beta$ -D-glucopyranoside (linalyl-glc), linalyl  $\beta$ -primeveroside  
969 (linalyl-pri), 2-phenylethyl  $\beta$ -D-glucopyranoside (2PE-glc), and 2-phenylethyl  
970  $\beta$ -primeveroside (2PE-pri). Data are presented as mean  $\pm$  SD (n = 3).

971

972 **Figure 3. Biochemical characterization of CsGT1 and At\_UGT85As.** A) CsGT1  
973 catalyzes the glucosylation of geraniol to produce geranyl  $\beta$ -D-glucopyranoside  
974 (geranyl-glc). B) LC-MS analysis of the enzymatic product of CsGT1 (UGT85K11) and

975 At\_UGT85A3 (At1g22380) compared with the authentic standard (geranyl-glu). C)  
976 Relative activity of CsGT1 toward sugar acceptors (geraniol, linalool, eugenol, benzyl  
977 alcohol, (Z)-3-hexenol, 2-PE, quercetin, and cyanidin). D) Relative activity of CsGT1  
978 toward sugar donors (UDP-glucose, UDP-xylose, UDP-galactose, and UDP-glucuronic  
979 acid). Data are presented as mean  $\pm$  SD (n = 3).

980

981 **Figure 4. Biochemical characterization of CsGT2.** A) CsGT2 (UGT94P1) catalyzes  
982 the xylosylation of geranyl-glc into geranyl  $\beta$ -primeveroside (geranyl  
983 6-O- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranosides). B) LC-MS analysis of the enzymatic  
984 product of CsGT2 (UGT94P1) compared with an authentic standard (geranyl-pri) (r.t. =  
985 4.9 min). C) Relative activity of CsGT2 toward sugar acceptors (geranyl-glc, linalyl-glc,  
986 eugenyl-glc, 2PE-glc, and pNP-glc). D) Relative activity of CsGT2 toward sugar donors  
987 (UDP-glucose, UDP-xylose, UDP-galactose, and UDP-glucuronic acid).

988

989 **Figure 5. Structural Comparison on the sugar-donor specificity of CsGT2 and its**  
990 **mutant, CsGT2 (I141S).** A) Homology model of UDP-xylose-bound CsGT2. B)  
991 Homology model of UDP-glucose-bound CsGT2 (I141S). For homology models,  
992 important amino acid residues on the active site are drawn at the stick form and  
993 UDP-sugars at the ball and stick form. Carbon atoms are colored in green for UGT  
994 amino acid residues and in cyan for UDP-sugars. Oxygen atoms are red, nitrogen atoms  
995 are blue, and phosphorus atoms are orange. Plausible hydrogen bonds are indicated by  
996 the red dotted lines. To simplify the visibility of the models, the structure of a sugar

997 acceptor is removed. C) Relative activity of CsGT2 and CsGT2 (I141S) toward  
998 UDP-xylose with geranyl-glc as a sugar acceptor. D) Relative activity of CsGT2 and  
999 CsGT2 (I141S) toward UDP-glucose with geranyl-glc as a sugar acceptor. Data are  
1000 presented as mean  $\pm$  SD (n = 3). E) Schematic representations of UDP-sugar  
1001 recognition of with CsGT2 (WT) and UDP-xylose (schematic model of Fig. 5A). F)  
1002 Schematic representations of UDP-sugar recognition of with CsGT2 (I141S) and  
1003 UDP-glucose (schematic model of Fig. 5B). G) Schematic representations of  
1004 UDP-sugar recognition of with CsGT2 (I141S) and UDP-xylose. H) Schematic  
1005 representations of UDP-sugar recognition of with CsGT2 (WT) and UDP-glucose.

1006

1007 **Figure 6. Relative transcript abundance of *CsGT1*, *CsGT2*, and  $\beta$ -PD in various**  
1008 **organs (young leaves, matures leaves, stem, root, and flower) of *C. sinensis*.**

1009 Transcript abundance was measured by qRT-PCR and normalized to the internal  
1010 reference *ribosomal 18S*. The expression level of each gene in the stem was set at 1.0.  
1011 Data are presented as mean  $\pm$  SD (n = 3).

1012

1013 **Figure 7. Phylogeny of UGT-glucosyltransferase OG2 and OG8 family.** All other  
1014 UGT sequences were for *Arabidopsis thaliana* available on The Arabidopsis  
1015 Information Resource (TAIR) website. High bootstrap values (>750) are indicated on  
1016 the branches (1000 replicates). The Vv\_UGT85A28 described here was found to be  
1017 identical to VvGT14 (Bönisch et al., 2014b).

1018

1019 **Figure 8. Schematic illustration of a mode-of-action of CsGT1, CsGT2, and  $\beta$ -PD**  
1020 **in the volatile metabolism in *C. sinensis*.**  
1021

1022

1023

Table 1. Kinetic parameters of CsGT1 (UGT85K11) and CsGT2 (UGT94P1).

enzymes	substrate	$K_m$	$V_{max}$	$V_{max}/K_m$
		$\mu\text{M}$	$\text{nkat g}^{-1} \text{protein}$	$\text{nkat g}^{-1} \mu\text{M}^{-1}$
CsGT1 (UGT85K11)	geraniol	$44.2 \pm 3.0$	$332.1 \pm 8.1$	$7.5 \pm 0.5$
CsGT2 (UGT94P1)	geranyl $\beta$ -D- glucopyranoside	$78.1 \pm 19.6$	$60.0 \pm 4.8$	$0.77 \pm 0.2$

Data are presented as mean  $\pm$  SD (n = 3)

1024

1025

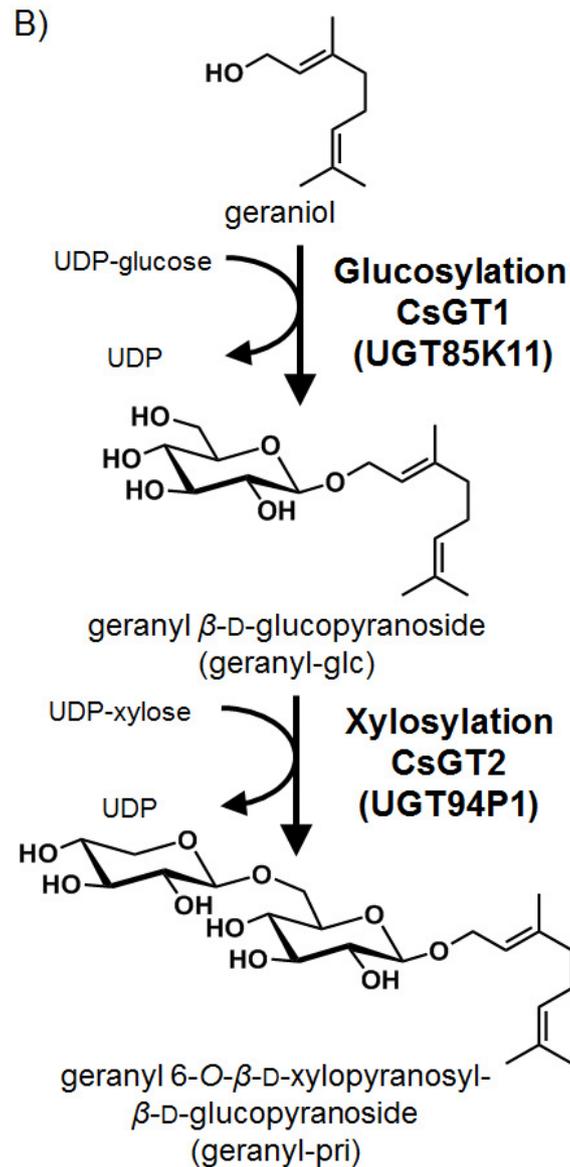
1026

1027

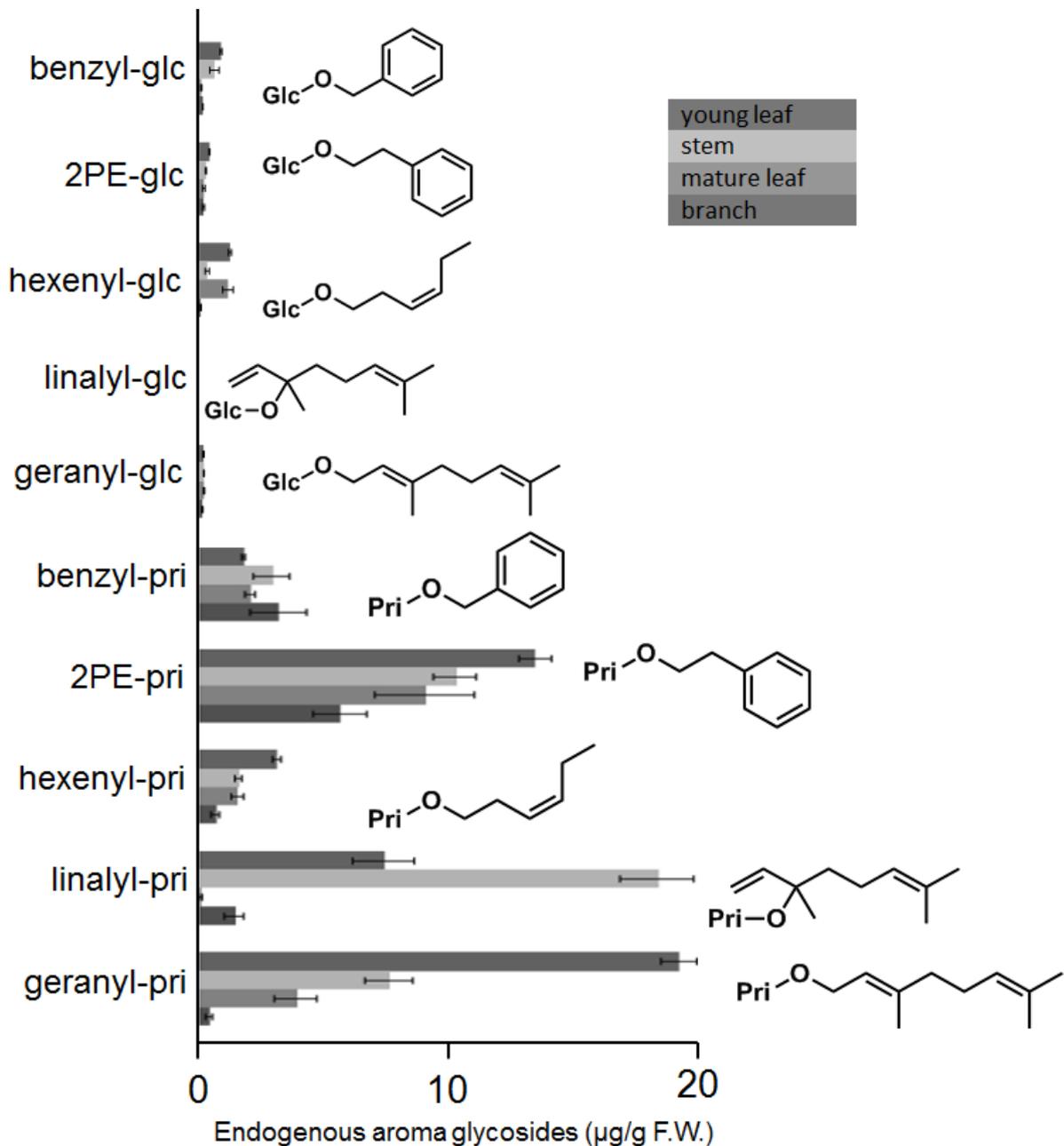
Table 2. Comparison of substrate specificity of GGTs in OG8 cluster

GGT (OG8)	species	substrate specificity			sequence alignment
		sugar donor	sugar acceptor	position	
<b>CsGT2</b> <b>(UGT94P1)</b>	Tea ( <i>C. sinensis</i> )	<b>UDP-xylose</b>	Volatile mono-glucoside	6'	IPAVQLM <u>I</u> TGAT 141
AcF3GGT1	Kiwifruit ( <i>A. chinensis</i> )	<b>UDP-xylose</b>	Flavonoid galactoside	2'	IKSVNYC <u>I</u> ISPA 136
UGT79B1	<i>A. thaliana</i>	<b>UDP-xylose</b>	Flavonoid glucoside	2'	AKTVCFN <u>I</u> VSAA 142
UGT79B6	<i>A. thaliana</i>	UDP-glucose	Flavonoid glucoside	2'	VKSVNFI <u>T</u> ISAA 135
UGT94D1	Sesame ( <i>S. indicum</i> )	UDP-glucose	Lignan glucoside	6'	IPAMVFL <u>S</u> TGAA 140
UGT94F1	<i>V. persica</i>	UDP-glucose	Flavonoid glucoside	2'	SPSVWFM <u>A</u> SGAT 144
Dusky (UGT79G16)	Morning glory ( <i>J. purpurea</i> )	UDP-glucose	Flavonoid glucoside	2'	IKSVFYYS <u>T</u> ISPL 138
NSGT1	Tomato ( <i>S. lycopersicum</i> )	UDP-glucose	Volatile di-glycoside	2'	IHAIMFY <u>V</u> SSTS 145

1028

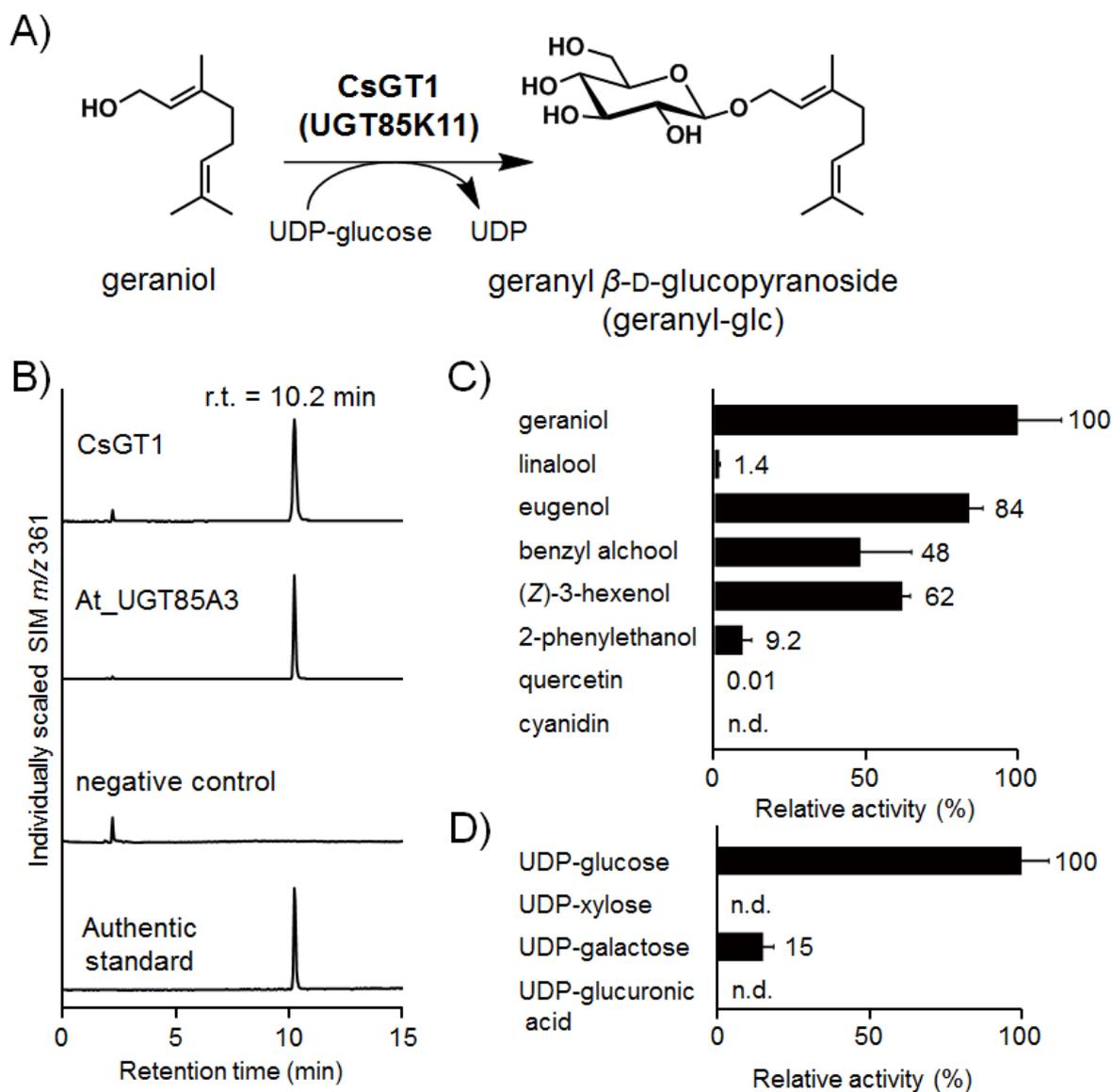


**Figure 1. Metabolism of volatile organic compounds in *Camellia sinensis*.** A) Photo of young leaves. (B) Biosynthesis pathway of geranyl  $\beta$ -primeveroside (geranyl 6-O- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranoside) from geraniol. CsGT1 (UGT85K11) and CsGT2 (UGT94P1) are the two glycosyltransferases that are shown to catalyze the sequential glucosylation and xylosylation of geraniol, respectively, in this work (Bold arrows).

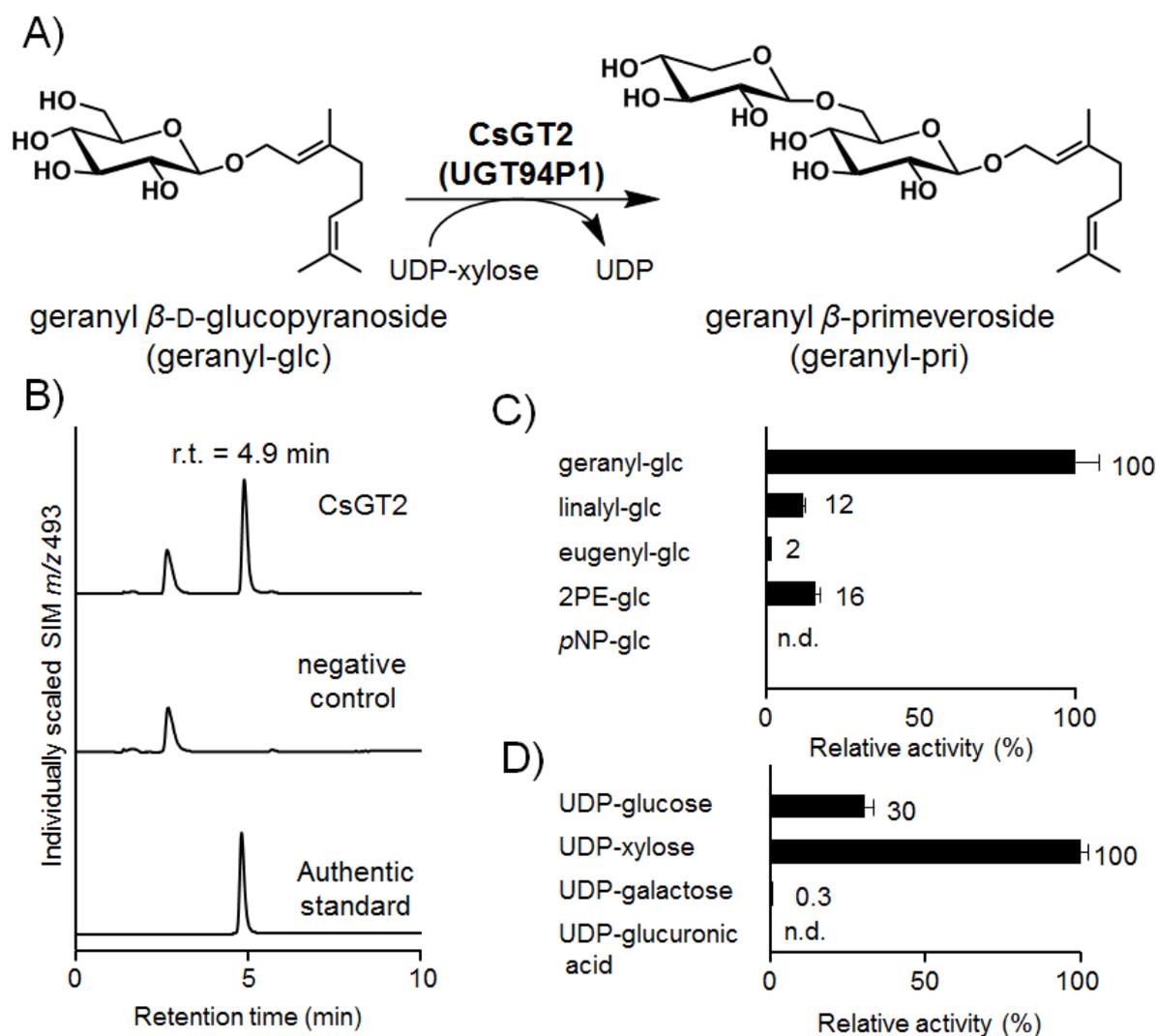


**Figure 2. Quantification of aroma monoglycosides and diglycosides in fresh tea leaves**

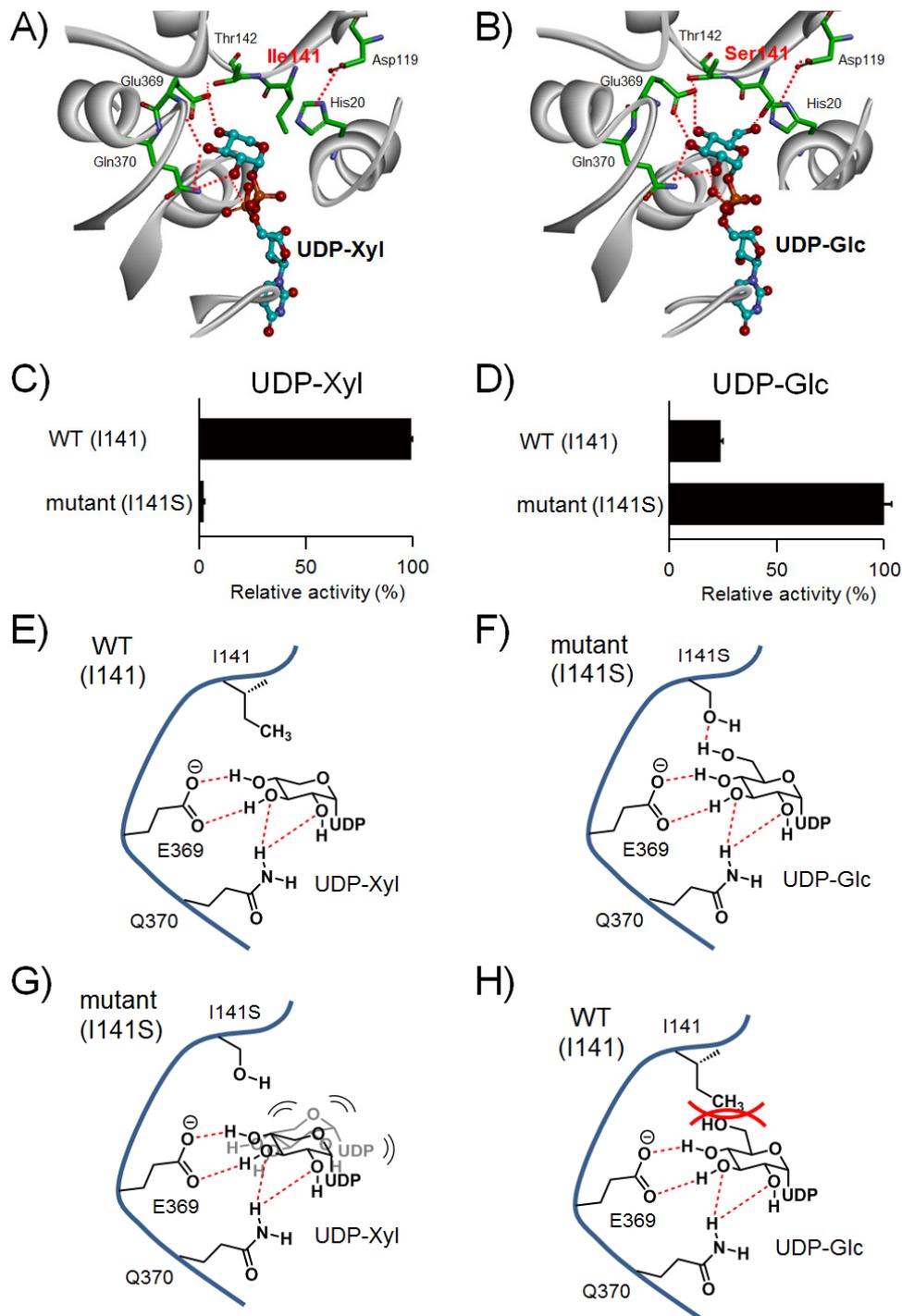
Ten glycosides were used as authentic standards to quantify endogenous glycoside in young leaves, young stems, mature leaves, and mature stems of *C. sinensis*: benzyl  $\beta$ -D-glucopyranoside (benzyl-glc), benzyl  $\beta$ -primeveroside (benzyl-pri), geranyl  $\beta$ -D-glucopyranoside (geranyl-glc), geranyl  $\beta$ -primeveroside (geranyl-pri), (Z)-3-hexenyl  $\beta$ -D-glucopyranoside (hexenyl-glc), (Z)-3-hexenyl  $\beta$ -primeveroside (hexenyl-pri), linalyl  $\beta$ -D-glucopyranoside (linalyl-glc), linalyl  $\beta$ -primeveroside (linalyl-pri), 2-phenylethyl  $\beta$ -D-glucopyranoside (2PE-glc), and 2-phenylethyl  $\beta$ -primeveroside (2PE-pri). Data are presented as mean  $\pm$  SD (n = 3).



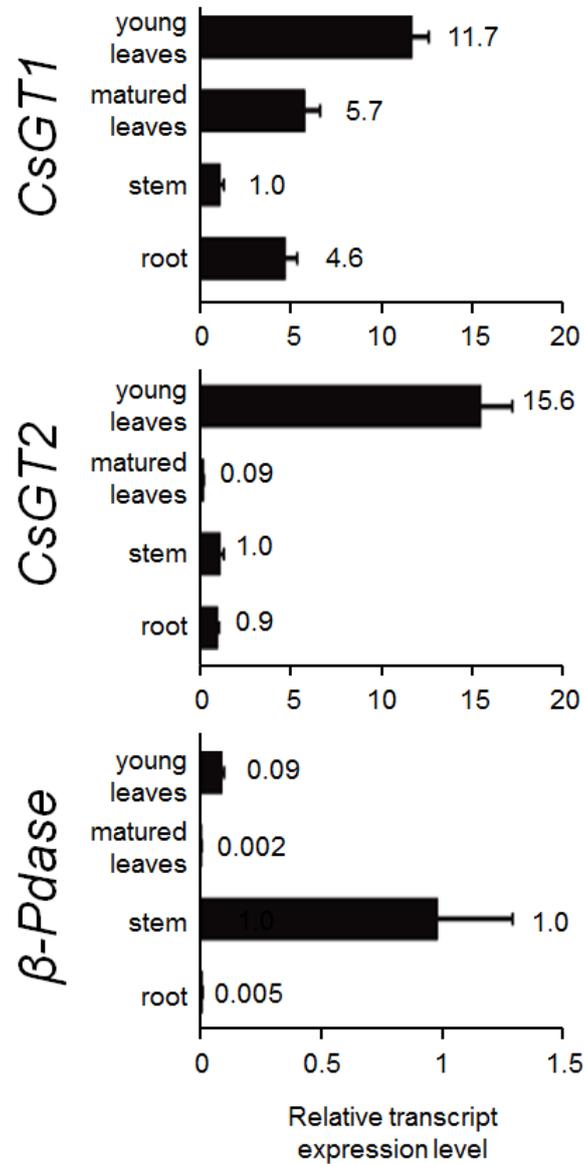
**Figure 3. Biochemical characterization of CsGT1 and At\_UGT85As.** A) CsGT1 catalyzes the glucosylation of geraniol to produce geranyl  $\beta$ -D-glucopyranoside (geranyl-glc). B) LC-MS analysis of the enzymatic product of CsGT1 (UGT85K11), At\_UGT85A1 (At1g22400), and At\_UGT85A3 (At1g22380) compared with the authentic standard (geranyl-glu). C) Relative activity of CsGT1 toward sugar acceptors (geraniol, linalool, eugenol, benzyl alcohol, (Z)-3-hexenol, 2-phenylethanol, quercetin, and cyanidin). D) Relative activity of CsGT1 toward sugar donors (UDP-glucose, UDP-xylose, UDP-galactose, and UDP-glucuronic acid). Data are presented as mean  $\pm$  SD ( $n = 3$ ).



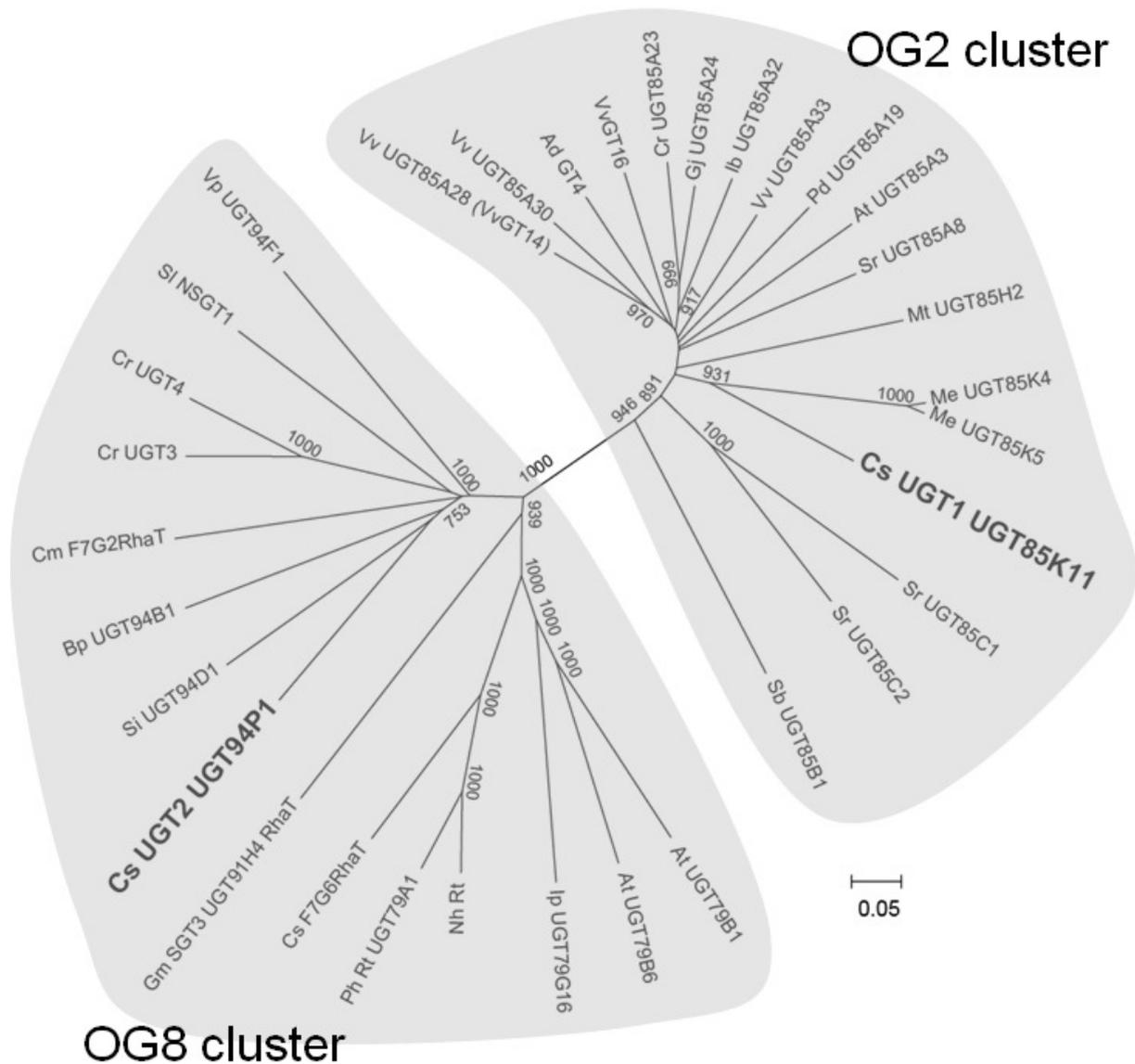
**Figure 4. Biochemical characterization of CsGT2.** A) CsGT2 (UGT94P1) catalyzes the xylosylation of geranyl-glc into geranyl  $\beta$ -primeveroside (geranyl 6-O- $\beta$ -D-xylopyranosyl- $\beta$ -D-glucopyranosides). B) LC-MS analysis of the enzymatic product of CsGT2 (UGT94P1) compared with an authentic standard (geranyl-pri) (r.t. = 4.9 min). C) Relative activity of CsGT2 toward sugar acceptors (geranyl-glc, linalyl-glc, eugenyl-glc, 2PE-glc, and pNP-glc). D) Relative activity of CsGT2 toward sugar donors (UDP-glucose, UDP-xylose, UDP-galactose, and UDP-glucuronic acid).



**Figure 5. Structural Comparison on the sugar-donor specificity of CsGT2 and its mutant, CsGT2 (I141S).** A) Homology model of UDP-xylose-bound CsGT2. B) Homology model of UDP-glucose-bound CsGT2 (I141S). For homology models, important amino acid residues on the active site are drawn at the stick form and UDP-sugars at the ball and stick form. Carbon atoms are colored in green for UGT amino acid residues and in cyan for UDP-sugars. Oxygen atoms are red, nitrogen atoms are blue, and phosphorus atoms are orange. Plausible hydrogen bonds are indicated by the red dotted lines. To simplify the visibility of the models, the structure of a sugar acceptor is removed. C) Relative activity of CsGT2 and CsGT2 (I141S) toward UDP-xylose with geranylglc as a sugar acceptor. D) Relative activity of CsGT2 and CsGT2 (I141S) toward UDP-glucose with geranylglc as a sugar acceptor. Data are presented as mean  $\pm$  SD ( $n = 3$ ). E) Schematic representations of UDP-sugar recognition of with CsGT2 (WT) and UDP-xylose (schematic model of Fig. 5A). F) Schematic representations of UDP-sugar recognition of with CsGT2 (I141S) and UDP-glucose (schematic model of Fig. 5B). G) Schematic representations of UDP-sugar recognition of with CsGT2 (I141S) and UDP-xylose. H) Schematic representations of UDP-sugar recognition of with CsGT2 (WT) and UDP-glucose.



**Figure 6. Relative transcript abundance of *CsGT1*, *CsGT2*, and  $\beta$ -PD in various organs (young leaves, matured leaves, stem, root, and flower) of *C. sinensis*.** Transcript abundance was measured by qRT-PCR and normalized to the internal reference *ribosomal 18S*. The expression level of each gene in the stem was set at 1.0. Data are presented as mean  $\pm$  SD (n = 3).



**Figure 7. Phylogeny of UGT-glucosyltransferase OG2 and OG8 family.** All other UGT sequences were for *Arabidopsis thaliana* available on The Arabidopsis Information Resource (TAIR) website. High bootstrap values (>750) are indicated on the branches (1000 replicates). The Vv\_UGT85A28 described here was found to be identical to VvGT14 (Bönisch et al., 2014b).

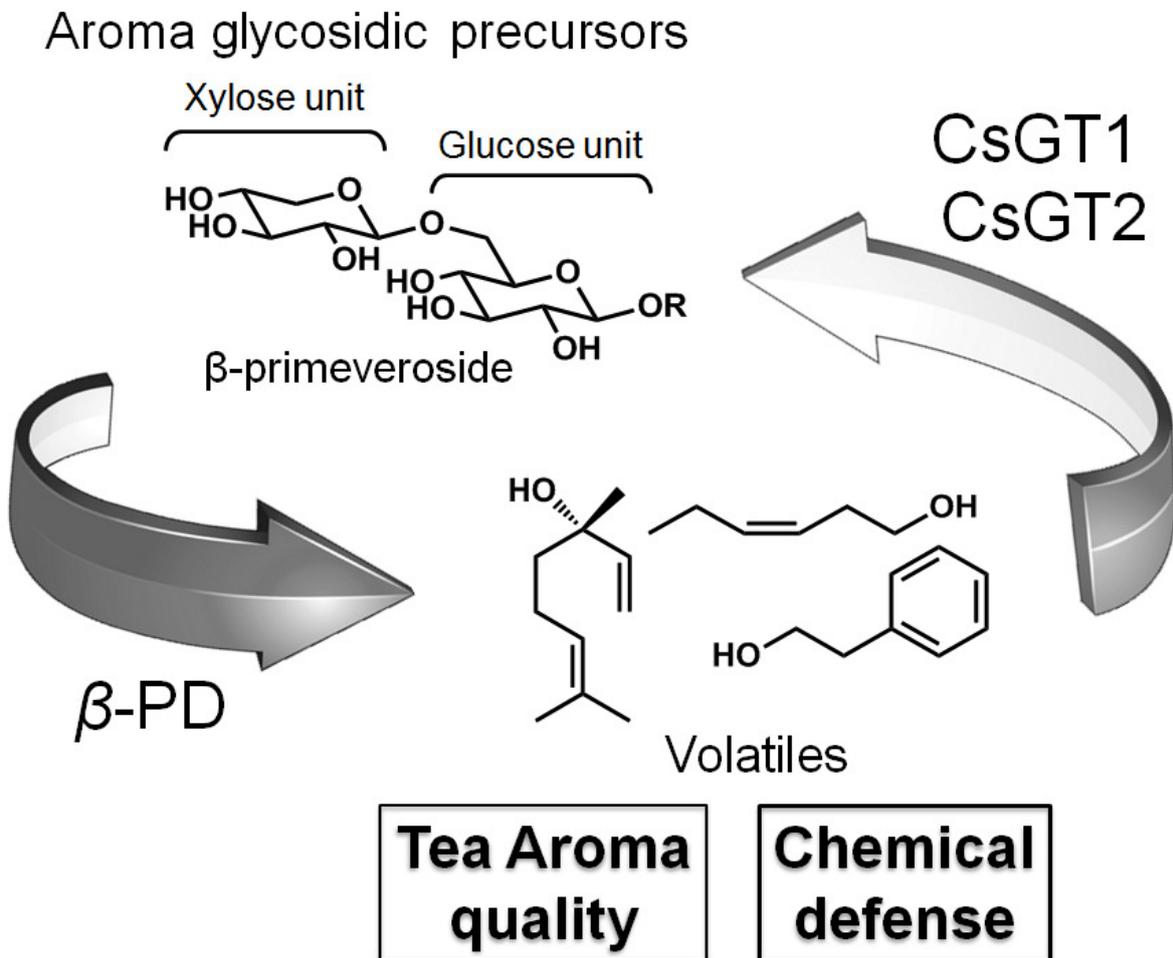


Figure 8. Schematic illustration of a mode-of-action for the metabolism of aroma volatiles by CsGT1, CsGT2, and  $\beta$ -PD in *C. sinensis*.

## Supplemental Materials and Methods

### LC-MS conditions for quantification of endogenous aroma glycosides in *C. sinensis*.

LC-MS analysis of ten aroma glycosidic precursors was performed on LCMS-2010A system (Shimadzu; [www.shimadzu.com](http://www.shimadzu.com)) equipped with Capcell Pak UG120 C<sub>18</sub> reversed phase column (2.0 mm i.d. x 150 mm, 5 μm, Shiseidou; [hplc.shiseido.co.jp/e](http://hplc.shiseido.co.jp/e)) and electrospray operating conditions were used the followings: dry gas 1.5 l/min, capillary voltage 1.5 kV, dry gas temperature 250°C. The endogenous aroma glycosides were quantified with the SIM negative mode. For benzyl-glc, benzyl-pri, 2PE-glc, 2PE-pri, (Z)-3-hexenyl-glc and (Z)-3-hexenyl-pri, the glycosidic fractions of were subjected to the LC-MS using gradient elution with aqueous formic acid (0.1%,v/v) as solvent A and acetonitrile as solvent B at a flow rate of 0.2 ml/min at 40°C. The gradient condition started with isocratic conditions of 16% of solvent B for 5 min and then increased up to 29% of solvent B for 24.5 min. The *m/z* 293 ion for benzyl-glc (*tR* 10.5 min), the *m/z* 425 ion for benzyl-pri (*tR* 14.9 min) the *m/z* 307 ion for 2PE-glc (*tR* 18.3 min), the *m/z* 439 ion for 2PE-pri (*tR* 22.4 min) The *m/z* 285 ion for (Z)-3-hexenyl-glc (*tR* 21.6 min), the *m/z* 417 ion for (Z)-3-hexenyl-pri (*tR* 26.3 min) and the *m/z* 456 ion for *p*NP-pri (*tR* 8.2 min) as a internal standard. For geranyl-glc, geranyl-pri, linalyl-glc and linalyl-pri, the glycosidic fractions of were subjected to the LC-MS using gradient elution with aqueous formic acid (0.1%,v/v) as solvent A and acetonitrile as solvent B at a flow rate of 0.2 ml/min at 40°C. The gradient condition started with 7% of solvent B then increased up to 30% of solvent B for 4 min and keep 30% of solvent B for 13 min. The *m/z* 339 ion for geranyl-glc (*tR* 14.5 min) and linalyl-glc (*tR* 13.0 min), and the *m/z* 471 ion for geranyl-pri (*tR* 11.9 min) and linalyl-glc (*tR* 11.0 min)

### Enzyme purification of a xylosyltransferase specific for monoglucoside-bound

## **volatiles**

The enzyme solution was applied to a HiTrap DEAE FF (5 ml, GE Healthcare) equilibrated with buffer B and eluted by a linear gradient of NaCl (0 to 340 mM) of buffer B. The active fraction was equilibrated with buffer C using an Amicon Ultra-15 centrifugal filter device (Mr 30,000 cutoff; Millipore). Then, the enzyme fraction was applied to a HiTrap Q FF (5 ml, GE Healthcare) and eluted with a linear gradient of NaCl (0 to 1 M) of buffer C. The active fractions was equilibrated with buffer D and applied to a Macro-prep Ceramic Hydroxyapatite Type III (5ml, Bio-Rad), and eluted with a linear gradient of NaCl (0 to 1 M) in buffer D. The active fraction was equilibrated with buffer E, applied to a HiTrap Blue HP (1 ml, GE Healthcare) and eluted with a linear gradient of NaCl (0 to 2 M) of buffer E. The active fraction was equilibrated with buffer C and applied to a Mono Q 5/50 GL (1 ml, GE Healthcare) and eluted with a linear gradient of NaCl (0 to 1 M) of buffer C. Peptide sequences were analyzed by Nano-LC-TOF MS.

## **Nano-LC-MS/MS analysis of partial peptide sequences of purified CsGT2**

Peptide assignments were performed using an LC-ESI-LIT-q-TOF mass spectrometer equipped with NanoFrontier eLD System (Hitachi High-Technologies, [www.hitachi-hitec.com](http://www.hitachi-hitec.com)) and a nano-flow HPLC, NanoFrontier nLC (Hitachi High-Technologies). The LIT-TOF and CID modes were used for MS detection and peptide fragmentation, respectively. The trypsin-treated sample (10  $\mu$ l) was injected, and the peptides were trapped on a C18 column, Monolith Trap (50  $\mu$ m  $\times$  150 mm, Hitachi High-Technologies). Peptide separation was achieved using a packed nano-capillary column (capillary-Ex nano mono cap, 0.05  $\times$  150 mm, GL Science, Japan) at a flow rate of 200 nl/min. The peptides were eluted using an acetonitrile gradient (A: 2% acetonitrile containing 0.1% formic acid; B: 98% acetonitrile containing 0.1% formic acid; 0 min with A = 98%, B = 2%, followed by 60 min with A = 60%, B = 40%). All peptide mass data were analyzed using

Peaks software (Bioinformatics Solutions Inc., [www.bioinfor.com](http://www.bioinfor.com)) and the MASCOT database (Matrix Science, [www.matrixscience.com](http://www.matrixscience.com)).

### **LC-MS analysis of CsGTs enzymatic products**

Enzymatic reaction mixtures by UGT enzymes were analyzed by LC-MS using Shimadzu LCMS-2010A ([www.shimadzu.com](http://www.shimadzu.com)) system with negative electrospray ionization mode; dry temperature, 250°C; a flow rate, 0.2 ml/min and column oven temperature, 40 °C. HPLC separation was performed according to the following conditions: Analysis of CsGT1 enzymatic products was performed using a CAPCELL PAK C<sub>18</sub> UG120 column (2.0 mm I.D.×150 mm; 5 µm) (Shiseido, [hplc.shiseido.co.jp/e/](http://hplc.shiseido.co.jp/e/)) and elution was started with 15% solvent B, following a linear gradient flow up to 90% in 15 min (solvent A: H<sub>2</sub>O containing 0.05% (v/v) formic acid, B: acetonitrile) Analysis of CsGT2 enzymatic products was performed using a CAPCELL PAK C<sub>18</sub> UG120 column (2.0 mm I.D.×75 mm; 5 µm) (Shiseido) and elution was started with 15% solvent B (0-2 min) following a linear gradient flow up to 60% in 8 min (solvent A: H<sub>2</sub>O containing 0.05% (v/v) formic acid, B: acetonitrile)

### **Subcloning of CsGT1 and CsGT2 to pET15b vector for *E.coli* expression**

The amplified cDNA fragments of CsGT1 and CsGT2 genes were subcloned into pENTR/D-TOPO vector (Life Technologies; [www.lifetechnologies.com](http://www.lifetechnologies.com)), and then digested with *XhoI* and *BglII* for CsGT1, and *NdeI* and *BamHI* for CsGT2 and the resulting DNA fragments were ligated into a pET-15b vector (Merck Millipore; [www.merckmillipore.com](http://www.merckmillipore.com)) previously digested with *XhoI* and *BamHI* for CsGT1, and *NdeI* and *BamHI* for CsGT2, respectively. The resulting plasmids were named pET-15b-CsGT1 and pET-15b-CsGT2, which encoded an N-terminal in-frame fusion of CsGT1 and CsGT2, respectively, with a His<sub>6</sub> tag. The inserted fragments were sequenced to confirm the absence of PCR errors and transformed into *E. coli* BL21(DE3) (TOYOBO; [www.toyobo-global.com](http://www.toyobo-global.com)).

### **Subcloning of CsGT2 (I141S) to pET15b vector**

According to the previously published method of Noguchi *et al.* (2007), *in vitro* mutagenesis of the CsGT2 gene was performed using recombinant PCR with the pENTR-Directional-TOPO vector (Life Technologies) containing the wild-type CsGT2 cDNAs as the templates and the specific mutagenic oligonucleotide primer set (CsGT2-134-I141S-FW and CsGT2-134-I141S-Rv) (Supplemental Table S2) to obtain the CsGT2-I141S mutant. The amplified fragment containing the I141S mutation was digested with *NdeI* and *BglII*, and the resulting DNA fragments were ligated with pET-15b at *NdeI* and *BamHI* site as described above. The introduced mutation was verified by DNA sequencing of both strands. The resulting CsGT2-I141S fragment was inserted in the expression vector pET-15b as described above.

### ***E. coli* expression and purification of recombinant CsGT1 and CsGT2**

The transformant cells were precultured at 37°C for 16 hr in a Luria-Bertani broth containing 50 µg/mL ampicillin. Twenty milliliters of the pre-culture was then inoculated into 200 mL, then further incubated at 22°C for 20 hr with Overnight Express Autoinduction System 1 (Merck Millipore) for expression of recombinant proteins. All subsequent operations were conducted at 0 to 4°C. The recombinant *E. coli* cells were harvested by centrifugation (7000g, 15 min), washed with distilled water, and resuspended in buffer F containing 20 mM imidazole. The cells were disrupted at 4°C by five cycles of ultrasonication. The cell debris was removed by centrifugation (7,000g, 15 min). Polyethyleneimine was slowly added to the supernatant solution to a final concentration of 0.12% (v/v). The mixture was allowed to stand at 4°C for 30 min, followed by centrifugation (7,000g, 15 min). The supernatant was applied to a HisTrap HP column (1 mL; GE Healthcare) equilibrated with buffer F containing 20 mM imidazole. The column was washed with buffer F containing 20 mM imidazole, and

the enzyme was eluted with buffer F containing 200 mM imidazole. The active column-bound fractions were concentrated and desalted using Vivaspin 30,000 MWCO (GE Healthcare), followed by substitution with buffer G. The expressed recombinant proteins in the gels were separated by SDS-PAGE and visualized by using PAGE Blue 83 (Cosmo Bio; [www.cosmobio.co.jp/index\\_e](http://www.cosmobio.co.jp/index_e)) and His · Tag monoclonal antibody (mouse, Merck Millipore) and ECL Anti-mouse IgG Horseradish Peroxidase-linked Whole antibody (sheep, GE Healthcare) respectively. Chemical luminescence derived from the recombinant proteins was detected by using Chemi-Lumi One Ultra (Nacalai tesque; [www.nacalai.co.jp/english/](http://www.nacalai.co.jp/english/)) in Molecular Imager ChemiDoc XRS+ (BioRad; [www.bio-rad.com/](http://www.bio-rad.com/))

### **Preparation of cryosections of tea leaf for MS imaging**

For MS imaging analysis, cryosections of the young tea leaf were prepared as described elsewhere. Briefly, tea leaves were embedded in SCEM embedding medium (Leica, [www.leica-microsystems.com](http://www.leica-microsystems.com)) in a Tissue-Tek Cryomold (Sakura Finetek, [www.sakura-finetek.com/](http://www.sakura-finetek.com/)) and desiccated in *vacuo* for 5 min at ambient temperature. The leaves were frozen by placing the mold containing tea leaves embedded in the medium on a block of dry ice for 15 min. Longitudinal sections of the leaves in 5 µm thickness were cut using a cryotome (Cryostar NX70, Thermo Scientific; [www.thermosci.jp](http://www.thermosci.jp)) with blade and block temperatures of -20 and -25 °C, respectively, and the sections were transferred onto a glass slide coated with indium tin oxide (Bruker Daltonics; [www.bruker.jp](http://www.bruker.jp)). The sections were dried in 50 ml tubes containing dried silica gel. The cryosections were uniformly sprayed with a solution of 50 mg/ml dihydrobenzoic acid (DHBA) in 70% methanol containing 0.1% trifluoroacetic acid using a 0.2-mm caliber airbrush (Procon Boy FWA Platinum; Mr. Hobby, Japan), air-dried for 10 min and mounted on a MALDI target plate.

### **MALDI-MS imaging analysis**

Longitudinal sections of seeds were selected for MS imaging analysis, performed by MALDI-TOF-MS/MS (Ultraflextream III, Bruker Daltonics). Parameters of the laser shots were as follows: laser diameter, 15  $\mu\text{m}$ ; raster, 15  $\mu\text{m}$ ; laser power, 3\_mid; laser power boost, 50%. Data from 100 laser shots were obtained at each position using the FlexImaging software program (Bruker Daltonics). The spectra were acquired in the positive mode. For MS/MS analysis, the parent ion of (*Z*)-3-hexenyl-pri ( $m/z$  417), (*Z*)-3-hexenyl- glu ( $m/z$  284) and geranyl-glc ( $m/z$  340) were monitored for a peak tolerance of 0.5 kDa. A cutoff value of 50% was used for the production of the images. The instrument was calibrated in the positive ion mode in the low-mass range using a mixture of selected chemical standards (DHBA and 4-hydroxycinnamic acid) prior to data acquisition. Authentic (*Z*)-3-hexenyl-pri was used as a reference compound. The images of the parent ion were overlaid onto bright-field images of the corresponding leaf sections with the use of FlexImaging software.

## Supplemental Tables

**Supplemental Table S1.** Summary of fresh weight and amounts of aroma  $\beta$ -primeverosides in young leaves and mature leaves of *C.sinensis*.

	aroma glycosides ( $\mu\text{g g}^{-1}$ fresh weight)		total amounts of aroma glycosides ( $\mu\text{g}$ )	
	young leaves	mature leaves	young leaves (FW $0.192 \pm 0.016$ g)	mature leaves (FW $0.808 \pm 0.057$ g)
benzyl-glc	$0.46 \pm 0.022$	$0.044 \pm 0.01$	$0.088 \pm 0.008$	$0.036 \pm 0.008$
2PE-glc	$0.22 \pm 0.010$	$0.098 \pm 0.03$	$0.042 \pm 0.004$	$0.079 \pm 0.025$
hexenyl-glc	$0.65 \pm 0.030$	$0.609 \pm 0.11$	$0.125 \pm 0.012$	$0.493 \pm 0.096$
linalyl-glc	n.d.	n.d.	nd	nd
geranyl-glc	$0.092 \pm 0.015$	$0.101 \pm 0.01$	$0.018 \pm 0.003$	$0.081 \pm 0.010$
benzyl-pri	$1.75 \pm 0.082$	$2.031 \pm 0.20$	$0.336 \pm 0.032$	$1.64 \pm 0.199$
2PE-pri	$13.0 \pm 0.61$	$8.78 \pm 1.9$	$2.50 \pm 0.24$	$7.09 \pm 1.63$
hexenyl-pri	$3.04 \pm 0.14$	$1.505 \pm 0.23$	$0.584 \pm 0.056$	$1.22 \pm 0.21$
linalyl-pri	$7.19 \pm 1.20$	$0.079 \pm 0.06$	$1.38 \pm 0.26$	$0.064 \pm 0.049$
geranyl-pri	$18.6 \pm 0.67$	$3.789 \pm 0.84$	$3.57 \pm 0.32$	$3.06 \pm 0.72$

**Supplemental Table S2.** Summary of the purification of CsGT2 from young fresh leaves of *C.sinensis*

	Total protein (mg)	Total activity (mU)	Specific activity (mU mg <sup>-1</sup> )	Yield (%)	Purification fold (-fold)
1. Crude extract	681	457	0.67	100.0	1.0
2. 30-70%(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	616	414	0.67	90.5	1.0
3. HiTrap DEAE	63	39	0.62	8.5	0.9
4. HiTrap Q	30	29	0.97	6.3	1.4
5. Hydroxyapatite	5.4	8.7	1.61	1.9	2.4
6. HiTrap Blue	0.34	1.1	3.23	0.24	4.8
7. Mono Q	0.15	1.3	8.67	0.28	13

One unit activity was defined as the amount of enzyme biosynthesized 1  $\mu$ mol of geranyl  $\beta$ -primeveroside per minute at 30°C. Specific activity and purification fold were calculated from the activity data.

**Supplemental Table S3.** Gene specific primers used for 5', 3'-RACE, amplification of full-length genes from *C.sinensis*, construction of CsGT2-I141S or real-time PCR.

gene specific primers	Primer sequences
At_UGT85A1-Dig-Fw	5'-ATGGGATCTCAGATCATTTCATAACT-3'
At_UGT85A1-Dig-Rv	5'-TTAATCCTGTGATTTTTGTCCCA-3'
At_UGT85A3-Dig-Fw	5'-ATGGGATCCCGTTTTGTTTCTAACGAA-3'
At_UGT85A3-Dig-Rv	5'-TTACGTGTTAGGGATCTTTCCCAA-3'
CsGT2-Race-FW	5'- GGCTGCTCATATTGCATGGGTATGGCTA-3'
CsGT2-Race-RV	5'-CGACAGGGTAGGAGAGAGAGGACTGGTTGT-3'
CACC- <i>Xho</i> I-CsGT1-FW	5'-CACCTCGAGATGGGTAGCAGAAAGCAG-3'
CsGT1- <i>Bgl</i> II-stop-RV	5'-AGATCTTTAGTATTGCTCACAATAGTGAAGAGC-3'
CACC- <i>Nde</i> I-CsGT2-FW	5'-CACCCATATGGATTCAAAAAGAGCAAAATG-3'
CsGT2- <i>Bam</i> HI-RV	5'-GGATCCTAATTTTTGAGCAACTTCACATC-3'
CsGT2-134-I141S-FW	5'-GCAGTCCAGCTCATGAGTACCGGAGCCACG-3'
CsGT2-134-I141S-RV	5'-CGTGGCTCCGGTACTCATGAGCTGGACTGC-3'
qRT-CsGT1-FW2	5'-TGTCCAAAGAGGCATTTTCC-3'
qRT-CsGT1-RV2	5'- AAGGATGGCATGTCCTTGAG-3'
qRT-CsGT2-FW2	5'-CGCAGTCCAGCTCATGATTA-3'
qRT-CsGT2-RV2	5'-TCAACAAAGTGGCGAAACTG-3'
qRT- $\beta$ -PD -FW	5'-AAGGATCCCCAGAGGGTCTA-3'
qRT- $\beta$ -PD -RV	5'-TCCGAACCTTTGGGTGTAAC -3'
qRT-Cs18SrRNA-FW	5'-CACGGGGAGGTAGTGACAAT-3'
qRT-Cs18SrRNA-RV	5'-CCTCCAATGGATCCTCGTTA-3'

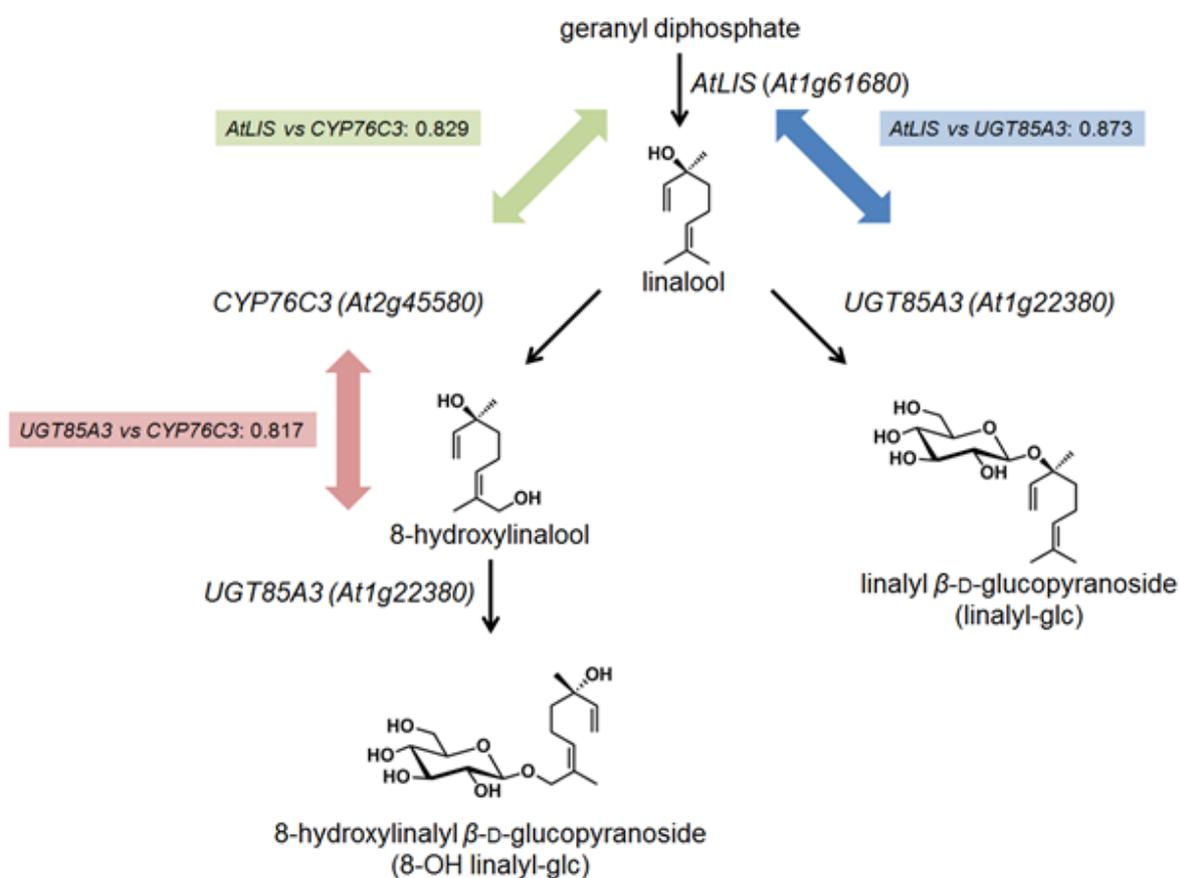
**Supplemental Table S4.** Composition of the buffers and solutions for purification and identification of CsGT2 protein

Buffer/Solution	Composition
Buffer A	100 mM Tris-HCl (pH 7.5) containing 10 mM sodium ascorbate, 5 mM dithiothreitol, 1 mM phenylmethylsulfonyl, 1 mM EDTA, 0.1% CHAPS, 0.1% (v/v) 2-mercaptoethanol, 5% (w/v) polyvinylpyrrolidone and 10% (w/v) DOWEX1x2(Cl <sup>-</sup> )
Buffer B	20 mM Tris-HCl (pH 8.0) containing 0.1% (v/v) 2-mercaptoethanol
Buffer C	20 mM borate buffer (pH 9.0) containing 0.1% (v/v) 2-mercaptoethanol
Buffer D	5 mM potassium phosphate buffer (pH 6.8) containing 0.1% (v/v) 2-mercaptoethanol
Buffer E	20 mM potassium phosphate buffer (pH 6.8) containing 0.1% (v/v) 2-mercaptoethanol
Solution F	25 mM NH <sub>4</sub> HCO <sub>3</sub> /acetonitrile (1:1 v/v)
Solution G	10 mM dithiothreitol/50 mM NH <sub>4</sub> HCO <sub>3</sub>
Solution H	55 mM iodoacetoamide/50 mM NH <sub>4</sub> HCO <sub>3</sub>
Solution I	50% acetonitrile containing 1% formic acid

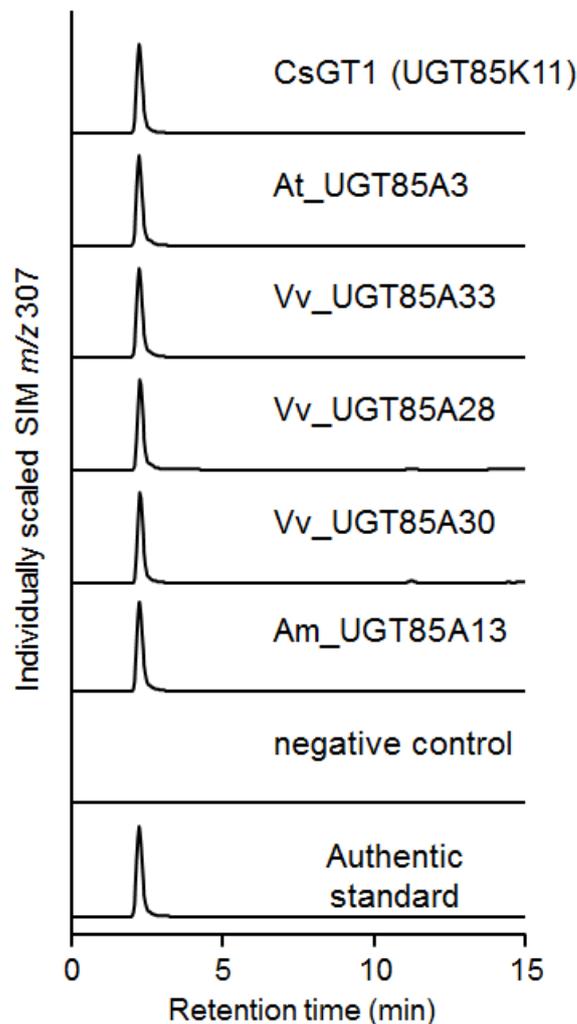
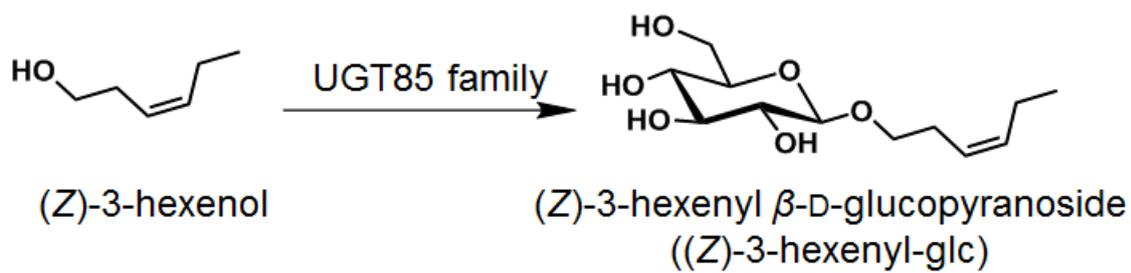
**Supplemental Table S5.** GenBank accession numbers used for construction of the phylogenetic tree in Figure 7.

Abbreviation	Plant species		Accession NO
Ad_GT4	<i>Actinidia deliciosa</i>	fuzzy kiwifruit	AIL51400
At_UGT79B1	<i>Arabidopsis thaliana</i>	Arabidopsis	NP_200217
At_UGT79B6	<i>Arabidopsis thaliana</i>	Arabidopsis	NP_200212
At_UGT85A3	<i>Arabidopsis thaliana</i>	Arabidopsis	NP_173655
Bp_UGT94B1	<i>Bellis perennis</i>	daisy	Q5NTH0
Cm_F7G2RhaT	<i>Citrus maxima</i>	pomelo	Q8GVE3
Cr_UGT3	<i>Catharanthus roseus</i>	madagascar periwinkle	BAH80312
Cr_UGT4	<i>Catharanthus roseus</i>	madagascar periwinkle	BAH80313
Cr_UGT85A23	<i>Catharanthus roseus</i>	madagascar periwinkle	F8WLS6
Cs_F7G6RhaT	<i>Citrus sinensis</i>	orange	NP_001275829
Cs_UGT1_UGT85K11	<i>Camellia sinensis</i>	tea plant	BAO51834
Cs_UGT2_UGT94P1	<i>Camellia sinensis</i>	tea plant	BAO51835
Gj_UGT85A24	<i>Gardenia jasminoides</i>	gardenia	F8WKW1
Gm_SGT3_UGT91H4_RhaT	<i>Glycine max</i>	soybean	NP_001240857
Ib_UGT85A32	<i>Ipomoea batatas</i>	sweet potato	BAO51842
Ip_UGT79G16	<i>Ipomoea purpurea</i>	morning glory	Q53UH5
Me_UGT85K4	<i>Manihot esculenta</i>	cassava	AEO45781
Me_UGT85K5	<i>Manihot esculenta</i>	cassava	AEO45782
Mt_UGT85H2	<i>Medicago truncatula</i>	barrel clover	XP_003618665
Nh_Rt	<i>Nierembergia sp. NB17</i>	cupflower	BAC10994
Pd_UGT85A19	<i>Prunus dulcis</i>	almond	ABV68925
Ph_Rt (UGT79A1)	<i>Petunia x hybrida</i>	petunia	CAA50376
Sb_UGT85B1	<i>Sorghum bicolor</i>	sorghum	XP_002463518
Si_UGT94D1	<i>Sesamum indicum</i>	sesame	BAF99027
SI_NSQT1	<i>Solanum lycopersicum</i>	tomato	AGO03777
Sr_UGT85A8	<i>Stevia rebaudiana</i>	stevia	AAR06913
Sr_UGT85C1	<i>Stevia rebaudiana</i>	stevia	AAR06922
Sr_UGT85C2	<i>Stevia rebaudiana</i>	stevia	AAR06916
Vp_UGT94F1	<i>Veronica persica</i>	persian speedwell	BAI44133
VvGT16	<i>Vitis vinifera</i>	grapevine	XP_002263158
Vv_UGT85A28 (VvGT14)	<i>Vitis vinifera</i>	grapevine	BAO51844
Vv_UGT85A30	<i>Vitis vinifera</i>	grapevine	NP_001277170
Vv_UGT85A33	<i>Vitis vinifera</i>	grapevine	NP_001277168

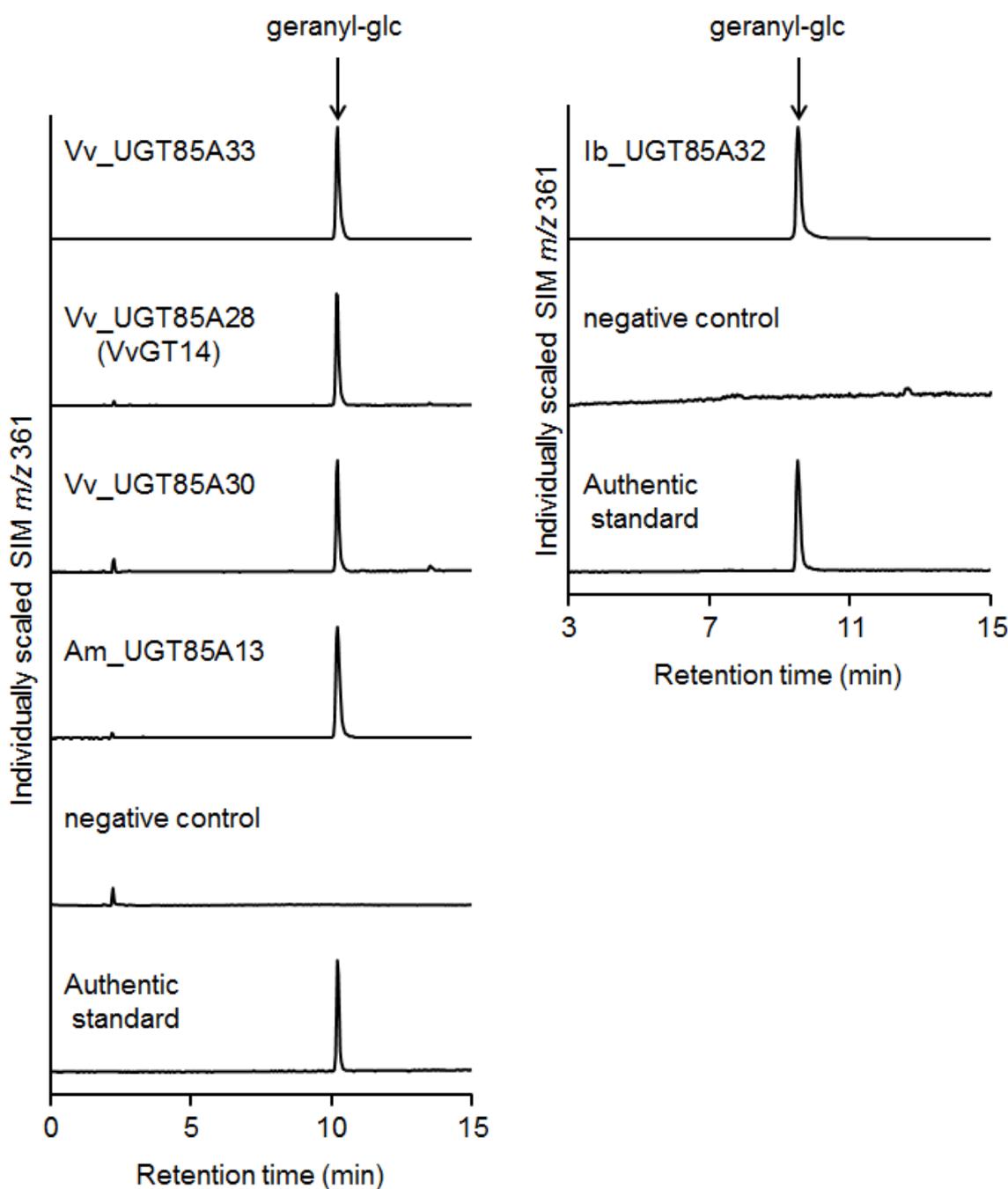
## Supplemental Figures



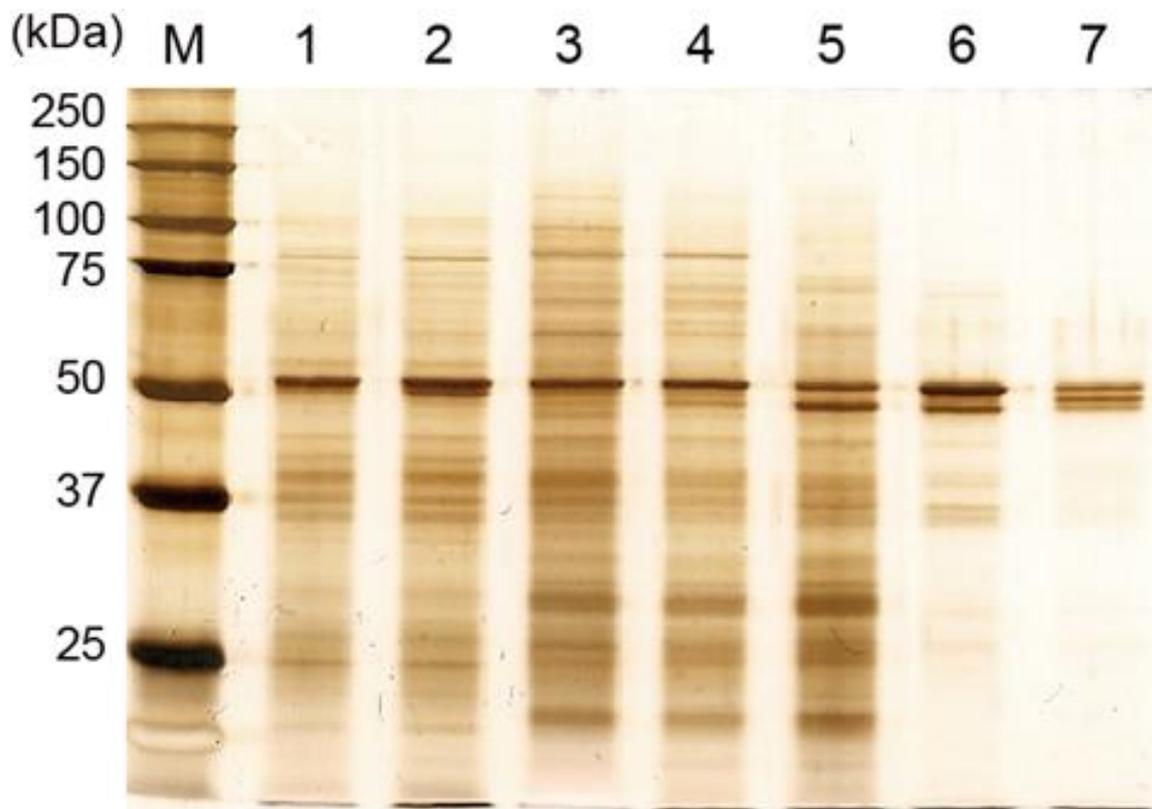
Supplemental Figure S1. Co-expression analysis of *AtLIS*, *CYP76Cs* and *UGT85A3* by ATTDII (Gene coexpression database, <http://atted.jp/>) with Ver. C4.1.



**Supplemental Figure S2. Enzymatic activity of CsGT1 homologs for (Z)-3-hexenol.** *In vitro* characterization of UGT85s toward (Z)-3-hexenol with tea plant (CsGT1; UGT85K11), Arabidopsis (At\_UGT85A3), grapevine (*Vitis vinifera*: Vv\_UGT85A33, Vv\_UGT85A28, Vv\_UGT85A30) and snapdragon (*Antirrhinum majus*: Am\_UGT85A13). LC-MS analysis of the enzymatic product of CsGT1 (UGT85K11), At\_UGT85A1, At\_UGT85A3, Vv\_UGT85A33, Vv\_UGT85A28, Vv\_UGT85A30 and Am\_UGT85A13 compared to the authentic standard ((Z)-3-hexenol).



**Supplemental Figure S3. Enzymatic activity of CsGT1 homologs for geraniol.** *In vitro* characterization of UGT85As toward geraniol with grapevine (*Vitis vinifera*: Vv\_UGT85A33, Vv\_UGT85A28, Vv\_UGT85A30), sweet potato (*Ipomoea batatas*: Ib\_UGT85A32), snapdragon (*Antirrhinum majus*: Am\_UGT85A13). LC-MS analysis of the enzymatic product of Vv\_UGT85A33, Vv\_UGT85A28, Vv\_UGT85A30 and Am\_UGT85A13 and Ib\_UGT85A32 compared to the authentic standard (geranyl-glu).



**Supplemental Figure S4. Purified enzymes catalyzing the second xylosyltransferase (CsGT2).** SDS-PAGE analysis of each active fraction obtained from purification steps. Lane 1; crude enzyme, lane 2; ammonium sulfate, lane 3; HiTrap DEAE FF, lane 4; HiTrap Q FF, lane 5; Macro-prep Ceramic Hydroxyapatite Type III, lane 6; HiTrap Blue HP, lane 7; Mono Q 5/50 GL. SDS-PAGE was stained by a silver staining method (Sil-best stain one, Nacalai, Kyoto, Japan)

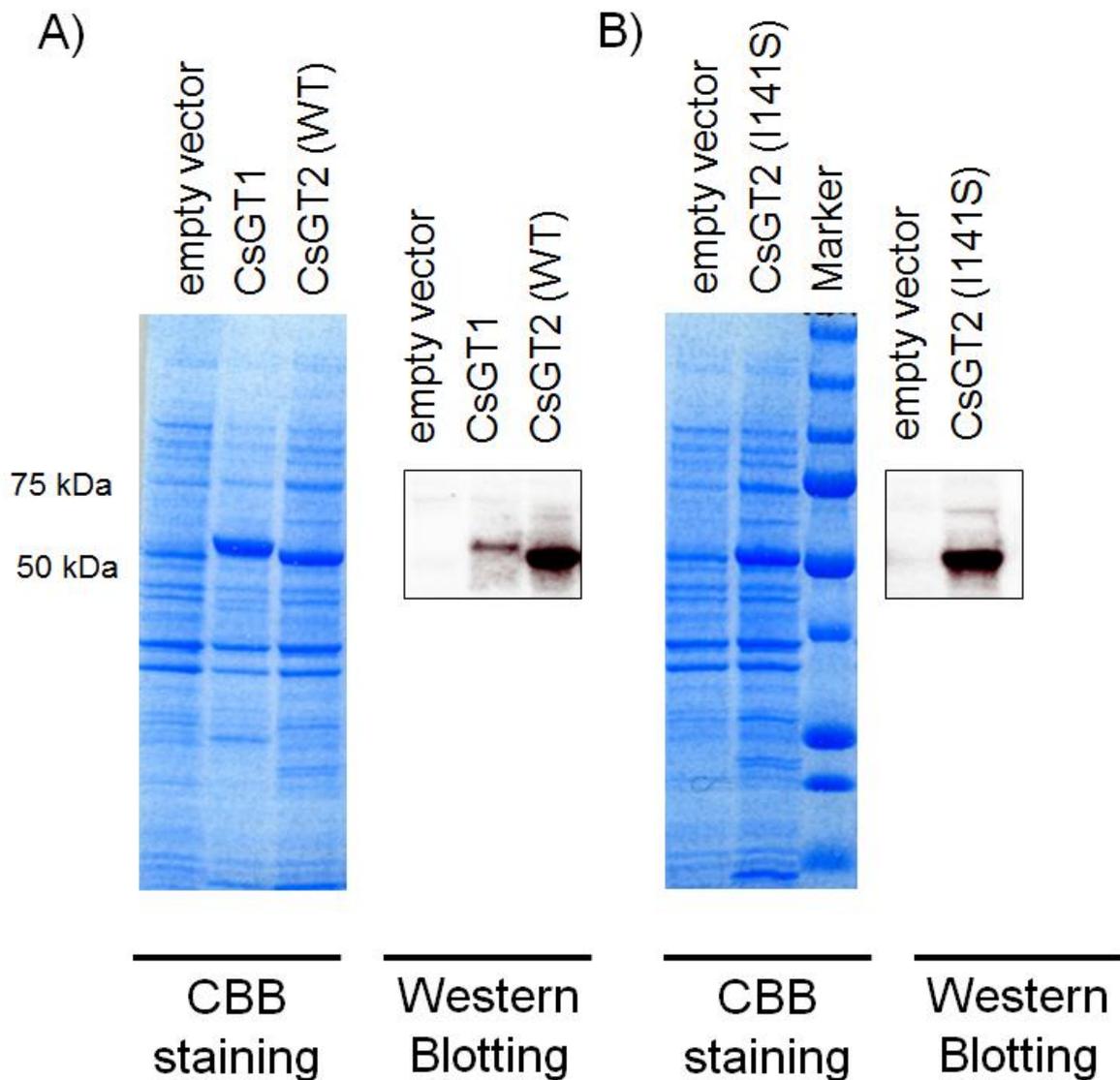
A)

Sequence	Charge	Observed <i>m/z</i>	<i>Mr</i> (Calc)
FPEVEKVELEEEALPK	+3	586.3224	1755.9294
GLVVEGWAPQAR	+2	641.852	1281.6829
EEIEEIAHGLELSMVNFIWVRFPEVEK	+4	840.4036	3357.6948

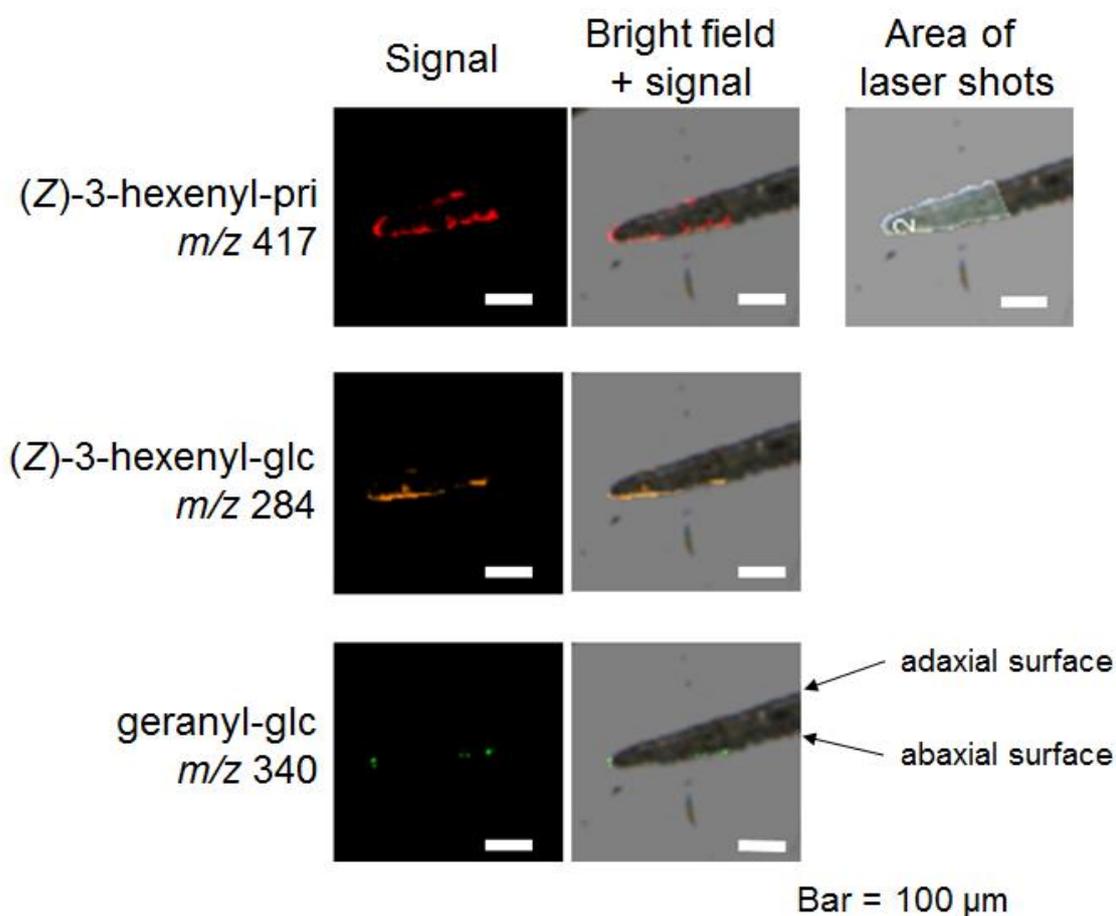
B)

1 ENEHSEIIQW LDNKGEYSTL FVSEFGSEYFM SK**EEIEEIAH** **GLELSMVNFI**  
 51 **WVRFPEVEK** **VELEEALPKG** FIDRVGERGL **VVEGWAPQAR** ILTHSSTGGF  
 101 VSHCEWNSVL ESLKFGVPMV AIPMQYEQPL NAKLVEEVGV AAEVNRDING  
 151 RLNREEIAQV IRKVVVEKSG EDIRIKARIF GDKIRMGDE EIDEAVEVLL  
 201 QLCKDVKLLK NSNFIEIFKW VEKVFYFYFI LFYWVVWYTN SLIRLYTSRR  
 251 STNIVFDHVV LVLDLVALFF IYIYRFTIIG VIVSSFCSID TGGILNMV

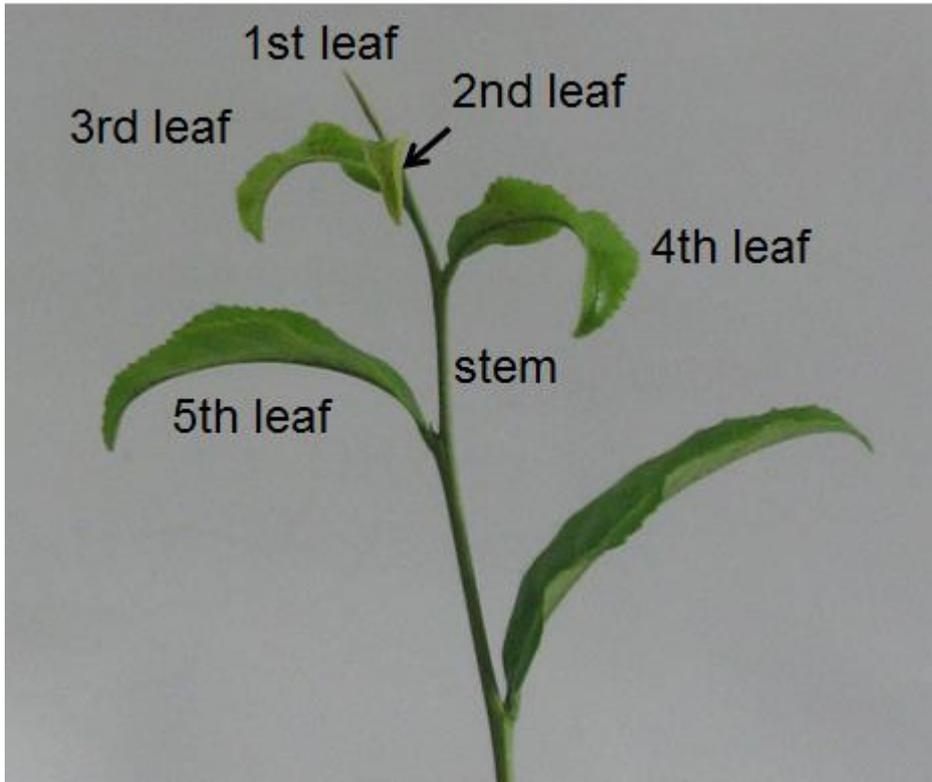
**Supplemental Figure S5. Partial peptide sequence of the CsGT2.** LC-MS/MS analysis of the partial peptide sequences of CsGT2 (UGT94P1) that purified from tea leaves. A) Three peptide sequences were exactly identified with tea EST database contig 13. B) Predicted amino acid sequence of tea EST database contig 134. Red letters show identical peptides with tea EST database and LC-MS/MS analytical results.



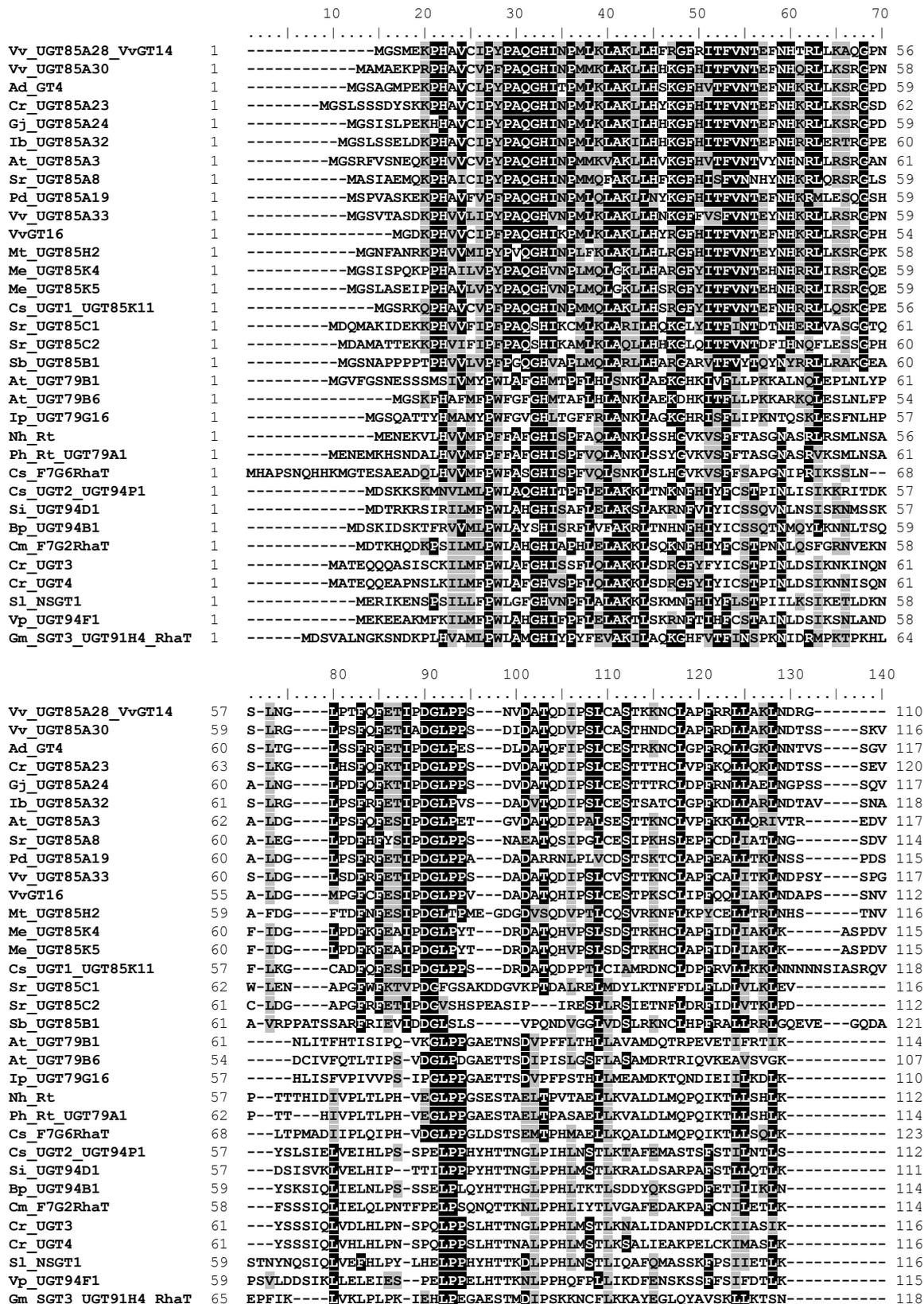
**Supplemental Figure S6. Recombinant proteins of a series of CsGTs.** Specific bands responsible of CsGT1 and CsGT2 (A), and CsGT2-I141S proteins (B) were detected around the size marker of 50kDa in SDS-PAGE, are roughly consistent with the estimated molecular size of CsGT1 (55.7 kDa) and CsGT2 (51.1 kDa), respectively. These specific bands were immunologically confirmed to be the recombinant CsGTs by Western blotting analysis.



**Supplemental Figure S7. Imaging MS of young fresh leaves of *C.sinensis*.** Specific signal (red; *m/z* 417) corresponding to (Z)-3-hexenyl-pri. Specific signal (orange; *m/z* 284) corresponding to (Z)-3-hexenyl-glc. Specific signal (green; *m/z* 340) corresponding to geranyl-glc.



**Supplemental Figure S8.** Harvesting individual tissues for quantification of endogenous aroma glycosides and quantitative real-time PCR (qRT-PCR) of CsGTs and  $\beta$ -PD. The young and mature developmental stages of leaves and stems were defined as follows: young leaves, which are the first, second, and third leaves (plucking part for high grade tea products); mature leaves, which are the fourth and fifth leaves; stem, which are green young non-lignified part between the 3rd leaf and 5th leaf; branch, which are brown lignified old part.



**Supplemental Figure S9.** Multiple sequence alignment of protein sequences of UGT-glucosyltransferase OG2 and OG8 family. The alignment was performed using ClustalW2.1.

	150	160	170	180	190	200	210								
Vv UGT85A28_VvGT14	111	PEVTCI	FSDAVMS	ET---	LDAAQEL	GLIPD	LLWTA	SACCFMAYVQYRSLIDR	KGFTL	KDES	YITNG	-YTD			
Vv UGT85A30	117	PEVTCI	VSDGIMS	ET---	LKAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-HTD			
Ad GT4	118	PEVSCV	VSDGVMS	SFS---	LDAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Cr UGT85A23	121	PEVSCV	VSDAVMS	ET---	ISAAQEL	LDLPE	YVFW	TSACCFMAYVQYRSLIDR	KGFTL	KDES	YITNG	-FTD			
Gj UGT85A24	118	PEVSCV	VSDGVMS	SFS---	LEAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Ib UGT85A32	119	PEVSCV	VSDGVMS	SFS---	VDAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
At UGT85A3	118	PEVSCV	VSDGVMS	SFS---	LDVAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Sr UGT85A8	115	PEVSCV	VSDGVMS	SFS---	LQAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Pd UGT85A19	116	PEVTCI	VADGVSS	ET---	LDAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Vv UGT85A33	118	PEVSCV	VSDGVMS	SFS---	LDAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
VvGT16	113	PEVTCI	VSDGSM	ET---	LKAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Mt UGT85H2	117	PEVTCI	VSDGSM	ET---	LQAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Me UGT85K4	116	PEVTCI	VSDGVMA	FA---	LDAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Me UGT85K5	116	PEVTCI	VSDGVMA	FA---	LDAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Cs UGT1 UGT85K11	119	PGVTCV	VSDGAMN	FA---	MKAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Sr UGT85C1	116	PEVTCI	ICDGCMT	FRAN	TIRAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Sr UGT85C2	112	PEVTCI	VSDGVMS	SFS---	TIDAAEEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
Sb UGT85B1	122	PEVTCV	VSDGVMS	SFS---	AAAAREEL	GLIPE	YVFW	TSACCFMAYVQYRHLIDR	GFPEL	KDES	YITNG	-YTD			
At UGT79B1	114	--PDLV	VYD	SAH	-WIP---	EIAKPI	CAKTY	CENIVS	SAASIALS	LVPSAERE	VIDGKEMS	GEBAKT	-PTG		
At UGT79B6	107	--PDL	VYD	FAH	-WIP---	EIAREY	CGKTY	CENIVS	SAASIALS	LVPSAERE	VIDGKEMS	GEBAKT	-PPG		
Ip UGT79G16	110	--VDV	VYD	FTH	-WLP---	SLARKI	CGKTY	CENIVS	SAASIALS	LVPSAERE	VIDGKEMS	GEBAKT	-PAS		
Nh Rt	112	--PHF	VYD	FAQEW	LP---	KMANG	CGKTY	CENIVS	SAASIALS	LVPSAERE	VIDGKEMS	GEBAKT	-PTG		
Ph Rt UGT79A1	114	--PHF	VYD	FAQEW	LP---	KMANG	CGKTY	CENIVS	SAASIALS	LVPSAERE	VIDGKEMS	GEBAKT	-PTG		
Cs F7G6RhaT	123	--PHF	VYD	FTH	WLP---	LVGQ	CGKTY	CENIVS	SAASIALS	LVPSAERE	VIDGKEMS	GEBAKT	-PDG		
Cs UGT2 UGT94P1	112	--PDL	VYD	VSPSWAQ	---	STALS	SFD	FAVQ	LMITG	TVAS	FQGM	MIKH	-----CGSV-----E		
Si UGT94D1	111	--PDL	VYD	FLQSWAS	---	EEAESQ	N	FAVQ	LMITG	TVAS	FQGM	MIKH	-----TRPE-----E		
Bp UGT94B1	114	--PHL	VYD	FNQLWAP	---	EVASL	HTE	SIQ	LVV	PAR	-KLN	---	SLAD		
Cm F7G2RhaT	114	--PTL	VYD	LFQPM	AA---	EAAQY	D	FAVQ	LMITG	TVAS	FQGM	MIKH	-----PSLK-----E		
Cr UGT3	116	--PDL	VYD	DLHQ	PWTE	---	ALASR	HNE	FAVQ	LMITG	TVAS	FQGM	MIKH	-----PGI-----E	
Cr UGT4	116	--PDL	VYD	HLHQ	PWTE	---	ALASR	HNE	FAVQ	LMITG	TVAS	FQGM	MIKH	-----PGS-----E	
Sl NSGT1	116	--PDL	VYD	HLHQ	PWTE	---	ALASR	HNE	FAVQ	LMITG	TVAS	FQGM	MIKH	-----LTS	
Vp UGT94F1	115	--PDL	VYD	VFN	PWAA	---	KHALS	HNE	FAVQ	LMITG	TVAS	FQGM	MIKH	-----KTGS-----E	
Gm SGT3 UGT91H4_RhaT	118	--PDW	VYD	FAAA	WVIP	---	IARSYN	IP	CAHYNITP	FAVQ	LMITG	TVAS	FQGM	MIKH	-----MKDYS

	220	230	240	250	260	270	280																																																						
Vv UGT85A28_VvGT14	177	TVVDW	WVPG	-MKG	IRL	RDL	PESE	FIR	TTDP	-DDIM	LD	FAM	GELER	RARKAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																					
Vv UGT85A30	183	TVVDW	WVPG	-MKG	IRL	RDL	PESE	FIR	TTDP	-DDIV	VN	FAM	GELER	RARKAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																					
Ad GT4	184	TVVDW	WVPG	-MEG	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FVL	SET	RN	THRSS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Cr UGT85A23	187	QVLDW	WVPG	-MEG	IRL	RDL	PESE	FIR	TTDP	-DEY	MIK	F	ILQ	ET	ERS	KKAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																			
Gj UGT85A24	184	QSLDW	WVPG	-MKD	IRL	RDL	PESE	FIR	TTDP	-DDY	MV	K	FV	LQ	ET	ERAKKAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																			
Ib UGT85A32	185	TELDW	WVPG	-MKG	IRL	RDL	PESE	FIR	TTDP	-DEY	MLK	F	ILQ	ET	GRARRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
At UGT85A3	184	TVVDW	WVPG	-MKN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Sr UGT85A8	181	TSLDW	WVPG	-MKN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Pd UGT85A19	182	TELDW	WVPG	-MKN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Vv UGT85A33	184	TVVDW	WVPG	-MKN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
VvGT16	179	TIIDW	WVPG	-MKN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Mt UGT85H2	183	TKVDW	WVPG	-LKN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Me UGT85K4	182	QVLDW	WVPG	-MPN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Me UGT85K5	182	QVLDW	WVPG	-MPN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Cs UGT1 UGT85K11	185	TTIDW	WVPG	-MRN	IRL	RDL	PESE	FIR	TTDP	-NDIM	LD	FV	RE	ACT	RKRAS	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																																				
Sr UGT85C1	184	MEIDW	WVPG	-MKR	IRL	RDL	PESE	FIR	TTDP	-EFIL	AT	QNY	FA	FE	FL	FETA	Q	AD	KV	SH	II	H	FE	LE	AS	LV	SE	IK	S	I	F	P																													
Sr UGT85C2	179	TVVDW	WVPG	-MEG	IRL	RDL	PESE	FIR	TTDP	-LDW	ST	DLN	-DKV	LM	FT	TEAP	QR	SH	KV	SH	II	H	FE	LE	AS	LV	SE	IK	S	I	F	P																													
Sb UGT85B1	189	TPLEW	WVPG	-MSH	IRL	RDL	PESE	FIR	TTDP	-DDV	M	V	S	AT	L	Q	M	E	S	A	G	S	K	AT	I	FN	FD	AL	EQE	VLD	DAI	APMP	Y																												
At UGT79B1	178	YESSK	V	LR	PHEAK	-SLS	F	V	R	K	H	E	A	I	G	S	F	F	D	G	K	V	T	A	M	R	N	C	---	-ATA	I	R	T	C	R	E	T	E	G	K	F	O	D	Y	I	S	R	O	Y	S											
At UGT79B6	161	YESSK	V	LR	GHE	TN	-SLS	F	L	S	Y	P	F	D	G	T	S	F	Y	E	R	I	M	I	G	L	K	N	C	---	-V	S	I	R	T	C	O	E	M	E	G	K	F	O	D	Y	I	S	R	O	Y	S									
Ip UGT79G16	171	FDDPS	I	K	L	H	A	H	E	A	R	G	F	T	A	R	T	V	M	K	F	G	D	I	T	F	D	R	I	E	T	A	V	S	E	S	---	-GLA	Y	S	T	C	R	E	T	E	G	K	F	O	D	Y	I	S	R	O	Y	S			
Nh Rt	175	FHTS	I	T	S	V	K	T	F	E	A	O	D	F	L	Y	I	E	K	S	F	N	R	P	T	Y	D	R	V	L	S	G	L	K	G	C	S	---	-ATA	I	R	T	C	O	E	M	E	G	K	F	O	D	Y	I	S	R	O	Y	S		
Ph Rt UGT79A1	177	FQTS	V	S	V	R	T	F	E	A	R	D	F	L	Y	I	E	K	S	F	N	R	P	T	Y	D	R	V	L	S	G	L	K	G	C	S	---	-ATA	I	R	T	C	O	E	M	E	G	K	F	O	D	Y	I	S	R	O	Y	S			
Cs F7G6RhaT	183	FVATS	I	T	S	L	D	E	F	A	R	D	F	L	Y	I	E	K	S	F	N	R	P	T	Y	D	R	V	L	S	G	L	K	G	C	S	---	-VLA	I	R	T	C	O	E	M	E	G	K	F	O	D	Y	I	S	R	O	Y	S			
Cs UGT2 UGT94P1	162	FPPA	I	K	L	Q	F	H	E	T	Q	F	R	H	F	V	E	T	V	K	E	N	D	K	Q	V	A	S	V	N	D	Q	P	S	C	N	---	-F	M	L	V	N	T	F	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S	
Si UGT94D1	161	YFPFA	I	H	Y	F	R	-E	H	E	Y	D	N	F	C	R	F	K	S	S	D	S	D	Q	L	R	V	S	D	C	V	K	R	S	H	D	---	-L	V	L	I	K	T	F	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S	
Bp UGT94B1	167	FPPFA	I	H	Y	F	R	-E	H	E	Y	D	N	F	C	R	F	K	S	S	D	S	D	Q	L	R	V	S	D	C	V	K	R	S	H	D	---	-L	V	L	I	K	T	F	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S	
Cm F7G2RhaT	163	YFFFS	D	Y	Q	D	R	E	S	K	N	I	N	F	L	H	L	T	A	N	G	T	L	N	K	D	R	F	L	K	A	F	E	L	S	C	K	---	-F	V	F	I	K	T	S	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S
Cr UGT3	165	FFFKA	I	H	L	S	D	F	E	Q	A	R	F	L	E	Q	L	E	S	K	N	D	A	S	---	-AKD	P	E	L	Q	G	S	R	G	-F	N	S	T	F	I	V	R	S	S	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S	
Cr UGT4	165	FFFKA	I	H	L	S	D	F	E	Q	A	R	F	L	E	Q	L	E	S	K	N	D	A	S	---	-AKD	P	E	L	Q	G	S	R	G	-F	N	S	T	F	I	V	R	S	S	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S	
Sl NSGT1	168	FFFSS	I	H	L	D	H	E	I	K	L	G	I	Q	I	P	R	D	E	K	A	F	A	Y	I	L	E	S	F	Q	S	H	N	---	-I	V	L	I	N	T	C	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S				
Vp UGT94F1	163	---	-L	V	E	G	V	D	F	G	E	I	K	R	H	I	S	P	N	T	K	G	A	D	F	G	G	F	I	L	G	S	L	N	S	S	E	---	-I	L	L	K	T	S	R	E	L	E	G	K	F	O	D	Y	I	S	R	O	Y	S	
Gm SGT3 UGT91H4_RhaT	173	GPTN	L	P	F	T	T	T	H	I	R	P	E	F	L	R	A	Y	E	G	T	D	E	E	T	G	E	R	A	S	F	D	L	N	K	A	Y	S	C	D	L	F	L	L	R	S	R	E	L	E	G	D	W	D	Y	I	A	G	N	Y	K

Supplemental Figure

	290	300	310	320	330	340	350																																																	
Vv UGT85A28_VvGT14	245	P-IYTI	GPPLQLLPDQIHD	----	SELKLGISN	LWKKEEPECLK	WLD	SKEE	NSVYVYVNF	GS-ITVMT	PQQLI	307																																												
Vv UGT85A30	251	P-IYTI	GPLOLLLNQMPD	----	NDLKSIEN	SNLWKEEP	GCLEWLD	AKEE	ESVYVYVNF	GS-VITVMT	PQQLV	313																																												
Ad GT4	252	P-IYTI	GPPLNLLMNQIKE	----	ESLKMIGS	NLWKEEPM	CLEWLN	SKEE	PKSVYVYVNF	GS-ITVMT	PQQLV	314																																												
Cr UGT85A23	255	P-IYTI	GPLOILNQVDD	----	ESLKVLSG	NLWKEEPE	CLEWLD	KD	ENS	SVYVYVNF	GS-ITVMT	NDQIL	317																																											
Gj UGT85A24	252	P-IYTI	GPLOFLQKEVKD	----	ERLSVLGS	NLWKEEPE	CLDWLD	SKEE	NSVYVYVNF	GS-ITVMT	PQQLV	314																																												
Ib UGT85A32	253	P-VYAV	GPLOFLQTVKDD	----	SNVRALAS	NLWKEETS	CLEWLD	TK	KE	NSVYVYVNF	GS-ITVMT	PDQLL	315																																											
At UGT85A3	252	P-VVPI	GPPLHLLVNRTEIE	---	DSEIGRMS	NLWKEETE	CLQWLN	TK	SRNSVYVYVNF	GS-ITVMT	TAAQLL	316																																												
Sr UGT85A8	249	Q-IYTI	GPPLHMMQ-QYVDH	---	DERLKHIG	SNLWKEE	DVSCIN	WLD	TK	KE	NSVYVYVNF	GS-ITVMT	KEQLI	312																																										
Pd UGT85A19	250	P-IYSI	GPLOLPYSEIPSE	---	YNDLKAIG	SNLWKEETE	CLN	WLD	TK	KE	NSVYVYVNF	GS-ITVMT	NEQLV	314																																										
Vv UGT85A33	253	P-VYSI	GPLOHLVDQISD	---	DRLKSMGS	NLWKEQTD	CLQWLD	SKEE	NSVYVYVNF	GS-ITVMT	SQQLT	315																																												
VvGT16	248	T-ICTV	GPPLLLNQIPDD	----	NSIESNL	LRREETE	CLQWLN	SK	KE	NSVYVYVNF	GS-ITVMT	PEQLV	308																																											
Mt UGT85H2	251	S-IYPI	GPPLSLLKQTPQ	---	IHQLDSD	SNLWKEDE	TECLDWL	ESKEE	PSVYVYVNF	GS-ITVMT	PEQLL	314																																												
Me UGT85K4	250	KNIY	TVGPFILLEKGIPEI	----	KSKAFRSS	LWKEEDLS	CLEWLD	KRE	ESVYVYVNF	GS-VITVMT	NEQLN	313																																												
Me UGT85K5	250	KNIY	TVGPFILLEKGIPEI	----	KSKAFRSS	LWKEEDLS	CLEWLD	KRE	ESVYVYVNF	GS-VITVMT	NEQLN	313																																												
Cs UGT1 UGT85K11	253	N-IYTI	GPPLSLLSKQVIDG	----	EFKLNSL	NLWKEETK	CLQWLD	TK	KE	NSVYVYVNF	GS-ITVMT	DQHLV	315																																											
Sr UGT85C1	252	N-VYTI	GPLOLLLNKITQKETT	----	DSYSLN	KEEPECE	WLN	SKEE	NSVYVYVNF	GS-LAVMS	LQDITV	314																																												
Sr UGT85C2	246	H-IYTI	GPLOLLDQIPEEKQV	---	SGITSLV	KEEPECF	WLN	OSKEE	NSVYVYVNF	GS-ITVMT	KEEDMT	313																																												
Sb UGT85B1	257	P-IYTV	GPPLAEVLIASSDSAS	---	AGLAAMD	ISITQED	TRCL	SWLD	CK	PAG	SVYVYVNF	GS-MAVMT	AAQAR	321																																										
At UGT79B1	242	KPVY	LTG---PVL	LP	SGSQP	-----	NQPSLD	PQ	AW	E	AKF	NHGS	VVFC	AFGS	QPVN	NKID	QFQ	295																																						
At UGT79B6	225	RKVL	LTG---PML	PE	PDN	-----	SKP-LE	DQ	WR	WL	SK	FDE	GS	VY	CA	FGS	QI	LE	KD	QFQ	276																																			
Ip UGT79G16	236	KPVLL	AG---PAL	P	V	-----	SKSTME	Q	K	W	S	D	W	L	G	F	K	E	G	S	V	I	C	A	F	G	S	E	C	T	L	R	K	D	K	F	Q	286																		
Nh Rt	240	KPVLL	AG---P	V	PD	PP	-----	SGK-LE	E	K	W	A	W	L	N	K	F	E	A	G	T	V	I	C	S	F	G	S	E	E	T	L	K	D	D	Q	I	K	290																	
Ph Rt UGT79A1	242	KPV	FSNR---TR	S	GP	A	-----	SGK-LE	E	K	W	A	W	L	N	K	F	E	A	G	T	V	I	C	S	F	G	S	E	E	T	L	T	D	D	Q	I	K	292																	
Cs F7G6RhaT	248	KPVLL	AG---P	L	V	NE	PE	PP	-----	SGE-LE	E	R	W	A	W	L	N	K	F	E	A	G	T	V	I	C	S	F	G	S	E	E	T	L	T	D	D	Q	I	K	299															
Cs UGT2 UGT94P1	227	KRVV	PV---P	L	V	Q	I	D	-----	DE-NE	H	S	E	I	I	Q	W	L	N	K	G	E	S	T	L	F	V	S	F	G	S	E	F	Y	M	S	K	E	E	I	278															
Si UGT94D1	225	KRVV	PV---P	L	V	Q	E	V	G	-----	DMEN	E	G	N	D	I	E	W	L	D	G	R	R	S	A	V	S	S	F	G	S	E	F	L	S	N	E	E	I	277																
Bp UGT94B1	223	KFLV	PV---P	L	V	Q	E	A	S	L	-----	LQ-DD	H	I	W	I	M	K	W	L	D	K	K	E	S	S	V	V	F	V	F	G	S	E	F	L	S	N	E	E	I	274														
Cm F7G2RhaT	228	NEI	I	P	V	G	---P	L	I	Q	E	P	-----	TFKV	D	D	T	K	I	M	D	W	L	S	O	K	E	E	R	S	V	V	A	S	F	G	S	E	F	Y	P	S	T	D	E	I	H	278								
Cr UGT3	231	SKV	I	P	V	C	---V	I	S	-L	N	N	D	Q	G	-----	QGN	K	D	E	D	E	I	I	Q	W	L	D	K	K	S	H	R	S	S	V	E	V	S	F	G	S	E	F	L	N	Q	E	E	I	286					
Cr UGT4	235	RKM	P	V	C	L	A	N	S	P	D	N	N	H	O	E	-----	QSN	K	D	G	D	E	L	I	Q	W	L	D	K	K	S	H	R	S	S	V	E	V	S	F	G	S	E	F	L	N	Q	E	E	I	291				
Sl NSGT1	233	KELI	P	C	P	T	I	R	E	A	M	I	D	E	E	E	-----	DWGT	I	Q	S	W	L	D	K	K	D	Q	L	S	C	V	V	V	S	F	G	S	E	S	L	S	K	Q	E	I	E	285								
Vp UGT94F1	225	KQI	P	T	P	C	---L	L	I	A	N	-----	SDE	K	D	E	P	E	I	M	Q	W	L	D	E	K	S	E	R	S	T	V	V	I	S	F	G	S	E	C	F	L	S	K	Q	E	I	E	274							
Gm SGT3 UGT91H4_RhaT	243	VV	V	V	P	V	G	L	---P	P	S	M	Q	I	R	D	V	E	-----	EED	N	N	P	D	W	V	R	I	K	D	W	L	D	T	E	S	S	S	V	V	I	C	F	G	S	E	L	K	L	S	Q	E	D	I	T	301

	360	370	380	390	400	410	420																																														
Vv UGT85A28_VvGT14	308	EFAW	GLANS	NQ	S	F	L	W	I	R	P	D	L	V	S	---	ESAIL	P	P	E	V	A	E	T	E	---	R	G	L	L	A	G	W	C	P	O	E	V	L	T	H	O	A	I	G	G	F	L	369				
Vv UGT85A30	314	EFAW	GLANS	NK	F	L	W	I	R	P	D	L	V	A	G	---	DAAIL	P	A	D	E	V	A	Q	T	K	E	---	R	S	L	L	A	S	W	C	P	O	E	V	L	T	H	P	A	I	G	G	F	L	375		
Ad GT4	315	EFAW	GLANS	NQ	S	F	L	W	I	R	P	D	L	V	G	---	ESAVL	P	P	E	V	A	V	T	K	E	---	R	G	M	L	A	S	W	A	P	O	E	V	L	T	H	S	S	V	G	G	F	L	376			
Cr UGT85A23	318	EFAW	GLANS	S	K	Q	F	L	W	I	R	P	D	L	I	S	---	ESSIL	G	E	E	V	E	E	T	K	E	---	R	G	L	I	A	S	W	C	H	O	E	V	L	T	H	P	A	I	G	G	F	L	379		
Gj UGT85A24	315	EFAW	GLANS	S	K	Q	F	L	W	I	R	P	D	L	V	S	---	DSAIL	P	P	E	V	E	E	T	K	D	---	R	G	L	L	A	S	W	C	P	O	E	V	L	T	H	P	A	I	G	G	F	L	376		
Ib UGT85A32	316	EFAW	GLANS	S	K	K	F	L	W	I	R	P	D	L	V	T	---	EAAI	I	P	E	E	E	T	K	D	---	R	G	M	S	S	W	C	S	O	E	V	L	T	H	P	A	I	G	G	F	L	377				
At UGT85A3	317	EFAW	GLA	A	T	K	E	F	L	W	M	R	D	S	V	A	---	EEAVI	P	K	E	L	A	E	T	A	D	---	R	R	M	L	T	S	W	C	P	O	E	V	L	T	H	P	A	I	G	G	F	L	378		
Sr UGT85A8	313	EFG	W	GLANS	S	K	K	F	L	W	I	R	P	D	I	V	G	---	NEAMI	P	A	E	E	E	T	K	E	---	R	G	M	V	T	S	W	C	S	O	E	V	L	T	H	P	S	I	G	V	F	L	374		
Pd UGT85A19	315	EFS	W	GLANS	S	K	K	F	L	W	I	R	E	G	L	V	A	---	EATAV	P	P	E	E	E	T	K	E	---	R	G	M	L	A	S	W	C	P	O	E	V	L	T	H	P	A	I	G	G	F	L	376		
Vv UGT85A33	316	EFAW	GLANS	N	K	F	L	W	I	R	P	D	L	V	G	---	DSAIL	P	P	E	V	E	T	K	D	---	R	G	M	L	A	S	W	C	P	O	E	V	L	T	H	P	A	I	G	G	F	L	377				
VvGT16	309	EFAW	GLANS	S	H	K	F	L	W	I	R	P	D	L	V	G	---	DSVIL	P	P	E	V	N	E	T	I	Q	---	R	G	L	M	A	G	W	C	P	O	E	V	L	T	H	P	S	V	G	G	F	L	370		
Mt UGT85H2	315	EFAW	GLANS	C	K	S	F	L	W	I	R	P	D	L	V	I	---	GSVIF	S	S	B	E	T	N	E	I	A	D	---	R	G	L	I	A	S	W	C	P	O	D	R	V	L	N	H	P	S	I	G	G	F	L	376
Me UGT85K4	314	EFAW	GLANS	S	K	H	F	L	W	I	R	P	D	V	V	M	---	ESAVL	E	E	E	V	E	E	T	K	D	---	R	G	L	L	V	S	W	C	P	O	D	R	V	L	O	H	P	A	V	G	V	F	L	375	
Me UGT85K5	314	EFAW	GLANS	S	K	H	F	L	W	I	R	P	D	V	V	M	---	ESAVL	E	E	E	V	E	E	T	K	D	---	R	G	L	L	V	S	W	C	P	O	D	R	V	L	O	H	P	A	V	G	V	F	L	375	
Cs UGT1 UGT85K11	316	EFAW	GLANS	S	K	H	F	L	W	I	R	P	D	I	V	M	---	DSAIL	E	E	E	V	E	E	T	K	D	---	R	G	L	L	V	S	W	C	P	O	E	V	L	T	H	P	S	I	G	V	F	L	377		
Sr UGT85C1	315	EFG	W	GLV	N	S	N	H	F	L	W	I	R	A	N	L	I	D	---	K																																	

		430	440	450	460	470	480	490	
Vv_UGT85A28_VvGT14	370	THNGWNSTDEGLCA	GVPMLCWF	FFAECQ	TNCRYCCTE	WGVGME	IDS	---	VKRDEVAKLVRRLMVEEK 434
Vv_UGT85A30	376	THSGWNSTDEGLCC	GVPLCWF	FFAECQ	MTNCRYCCTE	WGVGME	IGND	---	VPRDEVESLVRRLMVEEK 440
Ad_GT4	377	THCGWNSTDESIS	SGVAVVCWF	FFAECQ	TNCWYCCG	GELGIC	MEIDS	---	VKRDEVERLVRRLMVEEK 441
Cr_UGT85A23	380	THNGWNSTDESIS	SGVPMICWF	FFAECQ	TNCRFC	KNKVGIC	MEINS	---	VKRDEVESLVRRLMVEEK 444
Gj_UGT85A24	377	THSGWNSTDESIC	SGVPLCWF	FFAECQ	TNCWFC	CTKWNGL	MEIDN	---	VKRDEVESLVRRLMVEEK 441
Ib_UGT85A32	378	THNGWNSTDESIC	SGVPLCWF	FFAECQ	TNCHYAC	SKWGI	MEIDSN	---	VKRDEVEKLVRLMVEEK 442
At_UGT85A3	379	THCGWNSTDESL	SGVPMVCWF	FFAECQ	TNCKFS	ODEWV	MEIGD	---	VKRGEVEAVVRRLMVEEK 443
Sr_UGT85A8	375	THSGWNSTDESIS	SGVPMICWF	FFAECQ	TNCRYC	CVMEIG	MEIDT	---	VKRDEVEAQLVRRLMVEEK 439
Pd_UGT85A19	377	THSGWNSTDEAL	CGVPLCWF	FFAECQ	TNRYSC	TOWGIC	MEIDGE	---	VKRDYIDGLVRLMVEEK 441
Vv_UGT85A33	378	THSGWNSTDESIC	CGVPLCWF	FFAECQ	TNCRYSC	SEWGI	MEIDN	---	VKRVEVEKLVRLMVEEK 442
VvGT16	371	THSGWNSTDESIC	AGVPLCWF	FFAECQ	TNCRYC	CTE	WGVGMEIDN	---	VPRDEVEKLVRLMVEEK 435
Mt_UGT85H2	377	THCGWNSTDESIS	AGVPLCWF	FFAECQ	PTDCRF	CNEWIG	MEIDT	---	VKRDEELAKLVRLMVEEK 441
Me_UGT85K4	376	SHCGWNSTDECIS	CGKPLCWF	FFAECQ	TNCKYAC	DDVWKT	MEVLS	---	LKREELVSIIRKEMMETEI 440
Me_UGT85K5	376	SHCGWNSTDECIS	CGKPLCWF	FFAECQ	TNCKYAC	DDVWKT	MEVLS	---	LKREELVSIIRKEMMETEI 440
Cs_UGT1_UGT85K11	378	THCGWNSTDESIC	CGVPLCWF	FFAECQ	TNCRYC	CTE	WGVGMEVNH	---	VKRNEIVALINEMLEEK 442
Sr_UGT85C1	377	THCGWNSTDESIS	AGVPLCWF	FFAECQ	PTDCRF	CNEWIG	MEIDT	---	VKRDEELAKLVRLMVEEK 441
Sr_UGT85C2	376	THCGWNSTDESIS	AGVPLCWF	FFAECQ	TNCRYC	CKE	WVLEMGTK	---	VKRDEVKRLVRLMVEEK-EG 439
Sb_UGT85B1	385	SHCGWNSLDEATA	AGQVPLAWP	CHG	BOTNCR	QCEV	WGNCAOLPRE	---	VESGAVARLVRLMVEEK 449
At_UGT79B1	360	SHCCGSMWESLMS	DCQIVLVPH	QGC	ILNARL	MTE	EEMVAVEVER-EKKGWFS	---	QOSLENAVRSVMEEGS 428
At_UGT79B6	341	SHCCGSMWESLMS	DCQIVLVPH	QGC	ILNARL	MTE	EEMVAVEVER-EETGWFS	---	KESLSGAVRSVMEEGS 409
Ip_UGT79G16	351	SHCCGASLSL	EALVNDQIVL	LVLLQV	GDQILN	ARIMS	SVSLKVEVEKGEED	---	GVFSRESVCKAVKAVMDEKS 420
Nh_Rt	358	CHACFSSVIE	EALVNDQIVL	VLLQV	GDQILN	AKLV	SGDMKACVEVNRDED	---	GYFSKDDIEAEVKKVMEK- 426
Ph_Rt_UGT79A1	360	CHACFSSVIE	EALVNDQIVL	VLLQV	GDQILN	AKLV	SGDMKACVEVNRDED	---	GYFGKEDIKEAEVKKVMDVE 429
Cs_F7G6RhaT	367	CHSCFSSVTE	EAVIDCQ	LVVFLH	KGDFL	NSKL	VAGDLKACVEVNRDHD	---	GHFGEKDIKFAVKITVMDVN 436
Cs_UGT2_UGT94P1	344	SHCGWNSVLES	LKFGVPMVA	IPMOY	SOPLN	AKLV	VEEVG-VAE	---	EVNRD-INGRLNREETAQVIRKVVVEKS 411
Si_UGT94D1	344	SHCGWNSVLE	GVYSGVPL	IAVEM	MHLDP	PNARL	VEAVG-FGE	---	EVVRS-RQGNLDRGEVARVVRKLVMEKS 411
Bp_UGT94B1	332	SHCGWNSVLES	IRYGVPL	IAM	MOFOP	PNARL	MEVTG-AG	---	TEVGRD-GEGLRREETAQVIRKVVVEDS 399
Cm_F7G2RhaT	346	SHCGWNSVLE	GMVFGVPL	IAM	MAYE	OPSN	AKVVVNDNG-MGM	---	VVPRDKINQRLGGEVARVIRKVVVEE 414
Cr_UGT3	352	SHCGWNSVLES	IQIGVPL	IAM	MNLD	OPNARL	VVEIG-VG	---	TEVGRD-ENGLKREERIEVIRKVEATCKK 419
Cr_UGT4	357	SHCGWNSVLES	IEIGVPL	IAM	MITD	OPNARL	VEIG-VG	---	EVVRR-ENGLKREESVAEIRKVVVMGK 424
S1_NSGT1	351	THCGWNSVLES	MSGTP	IAM	MNHD	OPNARL	VEELG-IC	---	VEILRG-ENGLKREESVAEIRKVVVEEK 418
Vp_UGT94F1	341	SHCGWNSVLES	VYFGVPL	IAM	MKFE	QVNV	GVVVEVG-VG	---	VEKDGSGQYLGEVAKLDRKVFGEDEF 409
Gm_SGT3_UGT91H4_RhaT	363	SHCGWNSVLES	KVHGHV	LVTL	YLLD	CLFS	RVLEEKQ-VAV	---	EVPRSEKDGSGFTVVDVAKTLRFAIVDEE 431

		500	510	520	530	540			
Vv_UGT85A28_VvGT14	435	GVVMK	KKTEW	KHRAE	VATTG	PDGSS	VLNLEKIFEQVTL	---	473
Vv_UGT85A30	441	GKEMK	KKAEW	KRMAE	AATTP	PAGSS	YSNLDKMINQVLLSKSPC	---	484
Ad_GT4	442	GKEMK	ERAMG	WRRAE	EAATQ	SSSSG	SLNLDKLVHQQVLLSPRP	---	484
Cr_UGT85A23	445	GKEMK	KKAEW	KNIAE	VTTK	PDGSS	YSNLEKLIKVLKSKPSH	---	487
Gj_UGT85A24	442	GMDMK	KKAEW	KNAE	EAASG	SSYSN	LEKVVQVLLSK	---	481
Ib_UGT85A32	443	GKEMK	KKAEW	KKIAE	EAATSS	IGSSH	INIDKLINLHLLPPKY	---	485
At_UGT85A3	444	GKEMK	KKAEW	WRRAE	KATK	PCGSS	VINFEITVNVKVLGKIPNT	---	488
Sr_UGT85A8	440	GKMMK	KKAEW	KKRAE	EAVSIG	SSVLN	FELVTDVLLRK	---	479
Pd_UGT85A19	442	GKMRK	KKAEW	KKLAE	ADATSP	KGSS	VLALENVVSKVLLSPRD	---	483
Vv_UGT85A33	443	GKEMK	KKVME	KKIAE	EAATP	PGSS	VDFNFKLLRNVLGSK	---	482
VvGT16	436	GKSMK	KAAMW	RTKAE	EAAT	APCGSS	VLNLDKLVDIILTKP	---	475
Mt_UGT85H2	442	GKMKK	KAAMW	ELKKA	EENTR	PGGSS	VMLNKNKVIKDVLLKQ	---	482
Me_UGT85K4	441	GRERR	RRAVE	WRKKA	EAEIS	VGGSS	VYNNFDFTFKEVILQQQTQ	---	483
Me_UGT85K5	441	GRERR	RRAVE	WRKKA	EAEIS	VGGSS	VYNNFDRFKEALLQHKTK	---	483
Cs_UGT1_UGT85K11	443	GKQMR	KKALK	KKBAE	EAATD	VGGSS	LSYNNFDRLKEALLHYCEQY	---	485
Sr_UGT85C1	442	GKMRK	KKAEW	KKKSA	TLATC	CNGSS	LDVEKLANEIKKLSRN	---	483
Sr_UGT85C2	440	GKMRK	KKAKD	WKEK	ARIAI	APNGSS	SLNIDKMVKEITVLARN	---	481
Sb_UGT85B1	450	GKMKK	KAAMW	KAAMW	AAARK	GKGS	WRNVERVNDLLVGGKQ	---	492
At_UGT79B1	428	-EIG	KVRKN	HDK	WRCV	LTD	SFSDGYIDKFEQNLIELVKS	---	468
At_UGT79B6	409	-ELG	WARRN	HV	KWKES	LLRH	GLMSGYLNKFVEALEKIVONINLE	---	453
Ip_UGT79G16	420	-EIG	REV	RGN	HDK	LRG	FLMNADLLSKYMSDFNQKLDLIG	---	459
Nh_Rt	426	-----	VIREN	Q	KKW	KE	FLLNKDTHSKFVEDLVHDMMAKLTST	---	465
Ph_Rt_UGT79A1	430	KDP	GLIRE	NQ	KKW	KE	FLLNKDIQSKYIGNLVNMETAMAKVSTT	---	473
Cs_F7G6RhaT	437	KEP	GA	SIRAN	Q	KKW	REFLLNQIQDKFIADFKDLKATA	---	475
Cs_UGT2_UGT94P1	412	GEDI	R	IKARI	F	GDK	IRMKGDEEIDEAVEVLLQLCKDVLLKN	---	453
Si_UGT94D1	412	GEGL	R	RRV	DEL	SEK	MRKEGEEIDSLVEELVTVRRRERSNLKSENSMKKLVNMDGGE	---	469
Bp_UGT94B1	400	GES	IR	FKAK	ELGE	IMK	KNMAEVDGIVIENLVKLCMN	---	438
Cm_F7G2RhaT	415	AKQ	IR	KKAN	GIS	ES	MKKIGDAQMSVVVEKLLQLVKKE	---	452
Cr_UGT3	420	GKLR	KTAK	DLG	QKLR	DREK	QDFDELAATLKQLCV	---	454
Cr_UGT4	425	VEK	LR	K	TAN	DF	SKMKNREKEELDEVGLLQQLRNGPSSPMSE	---	468
S1_NSGT1	419	RQ	I	HL	KAM	QL	SEKIKKLAIDGVKLLKLLY	---	450
Vp_UGT94F1	410	SKE	V	RY	PAS	N	SDKIRENEEQEEDKVAEQMLSLCAKNLQKCD	---	452
Gm_SGT3_UGT91H4_RhaT	432	GSA	L	REN	AK	EM	GKVFSSSEELHNKYIQDFIDALQKYRIPSAS	---	472

Supplemental Figure S9. continued.