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

Running Title: Tea UGTs in biosynthesis of aroma *β*-primeverosides

Author to whom correspondence should be sent:

Toshiyuki Ohnishi

Graduate School of Agriculture, Shizuoka University, 829, Ohya, Suruga-ku, Shizuoka 422-8529, Japan

Telephone: + 81 54 238 3082

Fax: + 81 54 238 3082

e-mail: dtonish@ipc.shizuoka.ac.jp

Eiichiro Ono

Research Institute, Suntory Global Innovation Center Ltd., 1-1-1 Wakayamadai, Shimamoto, Mishima, Osaka 618-8503, Japan

Telephone: +81 75 962 2244

Fax: +81 75 962 2244

e-mail: eiichiro_ono@suntory.co.jp

Research Area 1: Biochemistry and Metabolism

Research Area 2: Genes, Development and Evolution

Volatile glycosylation in tea plants: Sequential glycosylations for the biosynthesis of aroma *β*-primeverosides are catalyzed by two *Camellia sinensis* glycosyltransferases

Shoji Ohgami[2], Eiichiro Ono[2,*], Manabu Horikawa, Jun Murata, Koujirou Totsuka, Hiromi Toyonaga, Yukie Ohba[3], Hideo Dohra, Tatsuo Asai, Kenji Matsui, Masaharu Mizutani, Naoharu Watanabe and Toshiyuki Ohnishi[*]

Graduate School of Agriculture, Shizuoka University, Shizuoka 422-8529, Japan (SO, KT, YO, TA, TO); Research Institute, Suntory Global Innovation Center Ltd., Shimamoto, Mishima, Osaka 618-8503, Japan (EO, HT); Bioorganic Research Institute, Suntory Foundation for Life Sciences, Shimamoto, Mishima, Osaka 618-8503, Japan (MH, JM); Research Institute of Green Science and Technology, Shizuoka University, Shizuoka 422-8529, Japan (HD, TO); Graduate School of Medicine, Yamaguchi University, Yamaguchi 753-8515, Japan (KM); Graduate School of Agricultural Science, Kobe University, Kobe 657-8501, Japan (MM); Graduate School of Engineering, Shizuoka University, Hamamatsu 432-8561, Japan (WN)

One-Sentence Summary:

Two glycosyltransferases catalyze sequential glycosylations of volatiles important for tea aroma quality, leading to stable accumulation of the volatiles as the water-soluble *β*-primeverosides.

Footnotes

[1]This work was supported by JSPS KAKENHI Grant Number 26870252 to T.O., and also supported by Suntory Holdings Limited to E.O.

[2]These authors contributed equally to this work.

[3]Present address: Bioorganic Research Institute, Suntory Foundation for Life Sciences, Shimamoto, Mishima, Osaka 618-8503, Japan

[*] Corresponding authors: dtonish@ipc.shizuoka.ac.jp (TO); eiichiro_ono@suntory.co.jp (EO)

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

Abstract

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

Introduction

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

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

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

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

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

genes in *C. sinensis*.

Results

Organ-specific composition in aroma monoglycosides and diglycosides

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

Identification of *Arabidopsis* UGT85A3 showing trans-glucosylation activity toward volatiles

Concurrent occurrence of monoglucosides and primeverosides of the corresponding volatiles in tea leaves suggested that primeverosides are biosynthesized via two sequential glycosylations steps, an initial glucosylation, followed by xylosylation rather than by direct conjugation of the primeverosyl moiety to the volatiles. We therefore searched for glucosyltransferases potentially responsible for the first glucosylation step in aroma *β*-primeveroside biosynthesis. For several classes of specialized metabolites, the biosynthetic genes are often found co-expressed (Fukushima et al., 2011). Transcriptome expression profiles and co-expression analysis became powerful tools for prediction of the biosynthetic genes constituting the metabolic pathway (Fukushima et al., 2011, Ginglinger et al., 2013, Usadel et al., 2009). However, a co-expression analytical tool for *C. sinensis*, a non-model plant, are not yet available. For identification of the glucosyltransferases catalyzing the first glucosylation of volatiles, we surveyed UGTs co-expressing with structural genes for monoterpene biosynthesis in Arabidopsis by ATTED II (http://atted.jp), which is a database developed to identify functionally related genes by co-expression. By using geraniol/nerol 10-hydroxylase gene (*At2g45580; CYP76C3*) and linalool synthase gene (*At1g61680; AtLIS*) (Ginglinger et al., 2013; Mizutani et al., 1997; Obayashi et al., 2011) as probes, we found that the expression profiles of Arabidopsis *UGT85A3* (At_UGT85A3, At1g22380; r = 0.873) exhibits relatively high correlation

with *AtLIS and CYP76C3* (Supplemental Fig. S1). *In vitro* functional characterization of At_UGT85A3 was performed using UDP-glucose as a sugar donor and geraniol or (*Z*)-3-hexenol as a sugar acceptor revealed that At_UGT85A3 produced geranyl-glc from geraniol and hexenyl-glc from (*Z*)-3-hexenol (Supplemental Fig. S2 and Supplemental Fig. S3). These data demonstrate that At_UGT85A3 is capable of catalyzing the glucosylation of monoterpene alcohols and aliphatic alcohols.

Identification of a *C. sinensis* UGT catalyzing the first glucosylation step for volatile *β*-primeveroside.

To isolate *C. sinensis* UGTs responsible for the first glucosylation step in volatile *β*-primeveroside biosynthesis, a cDNA library constructed from a mixture of leaves, stem and roots of *C. sinensis* (Mizutani et al., 2002) was screened with digoxigenin (DIG)-labeled *At_UGT85A3*. Two rounds of screening identified four novel UGTs, which were individually expressed in *Escherichia coli* and subjected to enzyme activity assays using UDP-glucose as a sugar donor and a variety of volatile alcohol acceptors. We found that one of the UGTs, named CsGT1, catalyzes glucosylation of geraniol as shown by the appearance of a product peak at the retention time of 10.2 min with *m/z* 361 ($[M + HCOO]^-$), both values of which correspond to those of authentic geranyl-glc (Fig. 3A and Fig. 3B). CsGT1 was assigned as Cs_UGT85K11 by the committee responsible for naming UDP-glucuronosyltransferases (Mackenzie et al., 1997).

The maximum velocity (V_{max}) and estimated apparent K_m values of CsGT1 for geraniol were 332.1 ± 8.1 nkat mg^{-1} protein and 44.2 ± 3.0 μM, respectively (Table 1).

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

Identification of orthologous UGTs of *CsGT1* from various plants

Volatile glycosides are reported in different plant species, including apricot, peach, yellow plum (Krammer et al., 1991), grape berries (Günata et al., 1985), kiwi (Young and Paterson, 1995), strawberry (Roscher, et al., 1997), raspberry (Pabst et al., 1991), and tomato (Marlatt, et al., 1992). Based on these general observations, we searched for UGTs in the NCBI Genbank (www.ncbi.nlm.nih.gov) based on amino acid sequence similarities with CsGT1 (accession number: AB847092). We found homologous UGTs broadly represented throughout the angiosperm plant lineages. We cloned another five UGTs from grapevine (*Vitis vinifera*: Vv_UGT85A33, Vv_UGT85A28, Vv_UGT85A30), sweet potato (*Ipomoea batatas*: Ib_UGT85A32), and snapdragon (*Antirrhinum majus*: Am_UGT85A13), and experimentally characterized the

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

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

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

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

Biochemical characterization of the xylosyltransferase

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

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

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

Homology modeling and mutagenesis analysis of CsGT2

The sugar donor specificity of UGTs is dictated by a small number of amino acid residues (Osmani et al., 2008, Noguchi et al., 2009, Ono et al. 2010a). The residues are located in three distinct domains: N-terminal, middle, and C-terminal (PSPG-box) (Sayama et al., 2012). To gain insights into the molecular mechanism of UDP-xylose specificity of CsGT2, we constructed a structural model of CsGT2 by homology modeling (Discovery Studio 3.5, Accelrys). The crystal structures of the glycosyltransferases, At_UGT72B1 (PDB code: 2vce) and Mt_UGT85H2 (PDB code: 2pq6), were selected as templates for their similarities to CsGT2. In addition, we used the three-dimensional (3D) structure of the sugar donor, UDP-2F-Glc, and the crystal structure of grape UDP-glucose:flavonoid 3-*O*-glycosyltransferase *Vv*GT1 (PDB code: 2c1z) (Offen et al., 2006). In the constructed homology model, Ile141 was identified as a candidate residue for the control of sugar donor specificity because it is located

proximal to UDP-xylose (Fig. 5A and Fig. 5E). This unique Ile141 was found to be conserved in two xylosyltransferases specific for flavonoid glycosides, kiwi F3GGT1 (Ile136) and *Arabidopsis* UGT79B1 (Ile142) (Montefiori et al., 2011, Yonekura-Sakakibara et al., 2012) (Table II). In contrast, various amino acid residues occupy this position in other structurally similar glycosyltransferases, including *Ipomoea purpurea* UGT79G16 (Thr138), *Sesamum indicum* UGT94D1 (Ser140), *Veronica persica* UGT94F1 (Ala144), and *Solanum lycopersicum* Nonsmoky glycosyltransferase 1 (Sl_NSGT1) (Val145) (Table II) (Morita et al., 2005, Noguchi et al., 2008, Ono et al., 2010b, Tikunov et al., 2013). To assess the functional relevance of Ile141 for the specificity toward UDP-xylose, a CsGT2-I141S mutant was generated by site-specific mutagenesis, in which Ile141 was replaced by a Ser residue. CsGT2-I141S was heterologously expressed in *E. coli* (Supplemental Figure S6B). Compared with wild-type CsGT2, the mutant exhibited significantly lower activity with UDP-xylose but higher activity with UDP-glucose (Fig. 5C and Fig. 5D). These experiments identified Ile141 as the crucial residue responsible for the sugar donor specificity of CsGT2 for UDP-xylose.

Gene expression and phylogenetic analysis of CsGT1 and CsGT2

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

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

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

Tissue localization of a selected aroma glycosidic precursor

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

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

Discussion

CsGT1 and CsGT2 catalyze the two glycosylation steps of volatile monoterpenes and alcohols

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

365 and grape (AdGT4 for (*Z*)-3-hexenol: 57.0 μM; VvGt14, VvGT15a, VvGT15b,
366 VvGT15c, and VvGT16 for citronellol: 9, 29, 55, 20, and 108 μM, 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 *β*-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

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

Sugar donor specificity of CsGT2 for UDP-xylose

UGTs usually show exclusive sugar donor specificity, which is determined by a small number of amino acid residues proximal to the bound sugar donor in the substrate binding pocket (Noguchi et al. 2009; Ono et al., 2010a; Osmani et al., 2008). These crucial residues for sugar donor specificity are located in three distinguishable regions such as the N-terminal, middle, and C-terminal regions (Sayama et al., 2012). Mutagenesis experiments (Ile141→Ser141) revealed that the unique Ile141 of CsGT2 located in the middle region is a residue determining the specificity toward UDP-xylose. The fact that the CsGT2-I141S mutant had considerably higher specificity for UDP-glucose supports the notion that the xylosyltransferase evolved from a glucosyltransferase by the acquisition of the crucial Ile141. It should be noted that an Ile residue, which corresponds to Ile141 in CsGT2, is also present in the mid-region of the two flavonoid 3-*O*-glycoside:2″-*O*-xylosyltransferases (kiwi Ac_F3GGT1 and Arabidopsis UGT79B1), but not in Arabidopsis flavonol 3-*O*-glucoside:2″-*O*-glucosyltransferase (UGT79B6) (Table II) (Yonekura-Sakakibara et al., 2012, 2014). These data suggest that the recognition mechanism for UDP-xylose by 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^{a} (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^{a} 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^{a} for UDP-xylose results from convergent evolution
430 because i) Sg-1^{a} and CsGT2 are phylogenetically remote GGTs specialized for

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

Putative physiological roles of CsGT1 and CsGT2 in volatile metabolism

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

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

Various plant species use the glycosylation of specialized metabolites to increase their solubility, which constitutes a primary strategy to accumulate metabolites operating as defensive compounds against herbivores and pathogens. At the same time, glycosylation of the metabolites facilitates the transport of the resulting glycosides to specific compartments, where they are stored separately from hydrolyzing enzymes 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 *β*-glucosidase that hydrolyzes the stored form
477 DIMBOA-glucoside (Frey et al., 2009). Similarly, glucosinolates
478 (*β*-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

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

Conclusion

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

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

Materials and Methods

Chemicals

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

Plant Materials

Young leaves of *C. sinensis* var. *sinensis* cv. Yabukita were harvested at the Center for Education and Research of Field Sciences, Shizuoka University and at Tea Research Center, Shizuoka Prefectural Research Institute of Agriculture and Forestry (Shizuoka, Japan). 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 (Fig. S8). All plant materials were quickly frozen in liquid nitrogen and stored at −80 °C prior to use.

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

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

Identification of *CsGT1*

Full-length cDNA clone of *CsGT1* was isolated by a screen with full-length DIG-labeled mixed probes for At_UGT85A, in a cDNA library derived from *C. sinensis* previously described (Mizutani et al., 2002). Library screening was performed under a low stringency condition as described in Yonekura-Sakakibara et al. (2000), and Noguchi et al. (2008, 2009). The screening probes of At_UGT85A were DIG-labeled by

PCR using gene-specific oligonucleotides (Supplemental Table S3). More than 50 positive clones were obtained in approximately 5,000,000 plaques after two rounds of screening. The cDNA fragments of positive clones were sequenced by conventional primer walking method with BigDye-terminator version 3.1 cycle sequencing kit (Life Technologies; www.lifetechnologies.com). Among these clones, a full-length cDNA for the *CsGT1* gene was identified by blastx search based on sequence similarities with At_UGT85A3.

Enzyme purification of a xylosyltransferase specific for monoglucoside-bound volatiles

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

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

Identification of the peptide sequence of CsGT2 by LC-MS/MS

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

Identification of full-length *CsGT2* cDNA

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

using the CloneJET kit (Thermo Fisher Scientific; www.thermoscientificbio.com).

Heterologous expression of recombinant UGT proteins

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

Enzyme assay of CsGT1 and CsGT2

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

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

Homology modeling

The construction of a 3D model according to CsGT2 was conducted using the Discovery Studio (DS) 3.5 (http://accelrys.com/). The crystal structures of At_UGT72B1 (PDB code: 2vce), Mt_UGT85H2 (PDB code: 2pq6), and VvGT1 (PDB code: 2c1z), and the 3D structure of the sugar donor UDP-2-F-Glc, were used as templates. The initial structure of CsGT2 was constructed using the multiple homology modeling protocols of the DS3.5 MODELER module. The resulting CsGT2 structure was inserted in the UDP-2F-Glc bound in VvGT1, and the sugar moiety of UDP-2F-Glc 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 *β-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*, *β-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 ± 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

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

GenBank accession numbers

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

Supplemental Data Files

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

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

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

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

Supplemental Figure S5. Partial peptide sequence of CsGT2.

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

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

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

CsGT2 and *β-PD*.

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

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

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

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.

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

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

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

Acknowledgments

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

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

References

Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel WJ, Verstappen FW, Verhoeven HA, Jongsma MA, Schwab W, Bouwmeester HJ (2003) Terpenoid metabolism in wild-type and transgenic Arabidopsis plants. Plant Cell **15**: 2866–2884

Arimura G, Matsui K, Takabayashi J (2009) Chemical and molecular ecology of herbivore-induced plant volatiles: proximate factors and their ultimate functions. Plant Cell Physiol **50**: 911–923

Balentine DA, Wiseman SA, Bouwens LC (1997) The chemistry of tea flavonoids. Crit Rev Food Sci Nutr **37**: 693–704

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72**: 248–254

Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M, Bitz O, Schwab W (2014a) Activity based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from Vitis vinifera. Plant Physiol **166**: 23–39

Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M, Bitz O, Schwab W (2014b) A UDP-Glucose:Monoterpenol Glucosyltransferase Adds to the Chemical Diversity of the Grapevine Metabolome. Plant Physiol **165**: 561–581

Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A (2009) Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants.

Phytochemistry **70**: 1645–1651

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

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

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

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

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

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

monoterpenols and phenylurea herbicides. PMID: 25082892

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Roscher R, Herderich M, Steffen JP, Schreier P, Schwab W (1997) 2,5-dimethyl-4-hydroxy-3(2H)-furanone 6′-O-malonyl-*β*-D-glucopyranoside in strawberry fruits. Phytochemistry **43**: 155–159

Sawada S, Suzuki H, Ichimaida F, Yamaguchi MA, Iwashita T, Fukui Y, Hemmi H, Nishino T, Nakayama, T (2005) UDP-glucuronic acid:anthocyanin glucuronosyltransferase from red daisy (Bellis perennis) flowers. Enzymology and phylogenetics of a novel glucuronosyltransferase involved in flower pigment biosynthesis. J Biol Chem **280**: 899–906

Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S, Tatsuzaki N, Tsukamoto C, Ishimoto M (2012) The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell **24**: 2123–2138

Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in Glycine max. FEBS Lett **584**: 2258–2264

Sugimoto K, Matsui K, Iijima Y, Akakabe Y, Muramoto S, Ozawa R, Uefune M, Sasaki R, Alamgir KM, Akitake S, Nobuke T, Galis I, Aoki K, Shibata D, Takabayashi J (2014) Intake and transformation to a glycoside of (*Z*)-3-hexenol from infested neighbors reveals a mode of plant odor reception and defense. Proc. Natl. Acad. Sci. U S A., **111**: 7144–7149

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6:

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

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

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

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

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

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

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

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

Yonekura-Sakakibara, K, Tanaka Y, Fukuchi-Mizutani M, Fujiwara H, Fukui Y, Ashikari T, Murakami Y, Yamaguchi M, Kusumi T (2000) Molecular and Biochemical Characterization of a Novel Hydroxycinnamoyl-CoA: Anthocyanin 3-*O*-Glucoside-6”-*O*-Acyltransferase from *Perilla frutescens*. Plant Cell Physiol **41**: 495–502

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

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

Yonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi M, Kitajima M, Takayama H, Saito K (2014) A flavonoid 3-*O*-glucoside:2”-*O*-glucosyltransferase responsible for terminal modification of pollen-specific flavonols in Arabidopsis thaliana. Plant J **79**: 769–782

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

Figure legends

Figure 1. Metabolism of volatile organic compounds in *Camellia sinensis.* A) Photo of young leaves. (B) Biosynthesis pathway of geranyl *β*-primeveroside (geranyl 6-*O*-*β*-D-xylopyranosyl-*β*-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 *β*-D-glucopyranoside (benzyl-glc), benzyl *β*-primeveroside (benzyl-pri), geranyl *β*-D-glucopyranoside (geranyl-glc), geranyl *β*-primeveroside (geranyl-pri), (*Z*)-3-hexenyl *β*-D-glucopyranoside (hexenyl-glu), (*Z*)-3-hexenyl *β*-primeveroside (hexenyl-pri), linalyl *β*-D-glucopyranoside (linalyl-glc), linalyl *β*-primeveroside (linalyl-pri), 2-phenylethyl *β*-D-glucopyranoside (2PE-glc), and 2-phenylethyl *β*-primeveroside (2PE-pri). Data are presented as mean ± SD (n = 3).

Figure 3. Biochemical characterization of CsGT1 and At_UGT85As. A) CsGT1 catalyzes the glucosylation of geraniol to produce geranyl *β*-D-glucopyranoside (geranyl-glc). B) LC-MS analysis of the enzymatic product of CsGT1 (UGT85K11) 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-PE, 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 ± SD (n = 3).

Figure 4. Biochemical characterization of CsGT2. A) CsGT2 (UGT94P1) catalyzes the xylosylation of geranyl-glc into geranyl *β*-primeveroside (geranyl 6-*O*-*β*-D-xylopyranosyl-*β*-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 *p*NP-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 geranyl-glc as a sugar acceptor. D) Relative activity of CsGT2 and CsGT2 (I141S) toward UDP-glucose with geranyl-glc as a sugar acceptor. Data are presented as mean ± 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 *β-PD* in various organs (young leaves, matures 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 ± 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 of CsGT1, CsGT2, and *β*-PD in the volatile metabolism in *C. sinensis*.

1022

1023

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

enzymes	substrate	K_m	V_{max}	V_{max}/K_m
		μM	nkat g^{-1} protein	nkat g^{-1} μM^{-1}
CsGT1 (UGT85K11)	geraniol	44.2 ± 3.0	332.1 ± 8.1	7.5 ± 0.5
CsGT2 (UGT94P1)	geranyl β-D-glucopyranoside	78.1 ± 19.6	60.0 ± 4.8	0.77 ± 0.2

Data are presented as mean ± 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	
CsGT2 (UGT94P1)	Tea (*C. sinensis*)	**UDP-xylose**	Volatile mono-glucoside	6'	IPAVQLM I TGAT 141
AcF3GGT1	Kiwifruit (*A. chinensis*)	**UDP-xylose**	Flavonoid galactoside	2'	IKSVNYC I ISPA 136
UGT79B1	*A. thaliana*	**UDP-xylose**	Flavonoid glucoside	2'	AKTVCFN I VSAA 142
UGT79B6	*A. thaliana*	UDP-glucose	Flavonoid glucoside	2'	VKSVNFI T ISAA 135
UGT94D1	Sesame (*S. indicum*)	UDP-glucose	Lignan glucoside	6'	IPAMVFL S TGAA 140
UGT94F1	*V. persica*	UDP-glucose	Flavonoid glucoside	2'	SPSVWFM A SGAT 144
Dusky (UGT79G16)	Morning glory (*I. purpurea*)	UDP-glucose	Flavonoid glucoside	2'	IKSVFYS T ISPL 138
NSGT1	Tomato (*S. lycopersicum*)	UDP-glucose	Volatile di-glycoside	2'	IHAIMFY V SSTS 145

1028

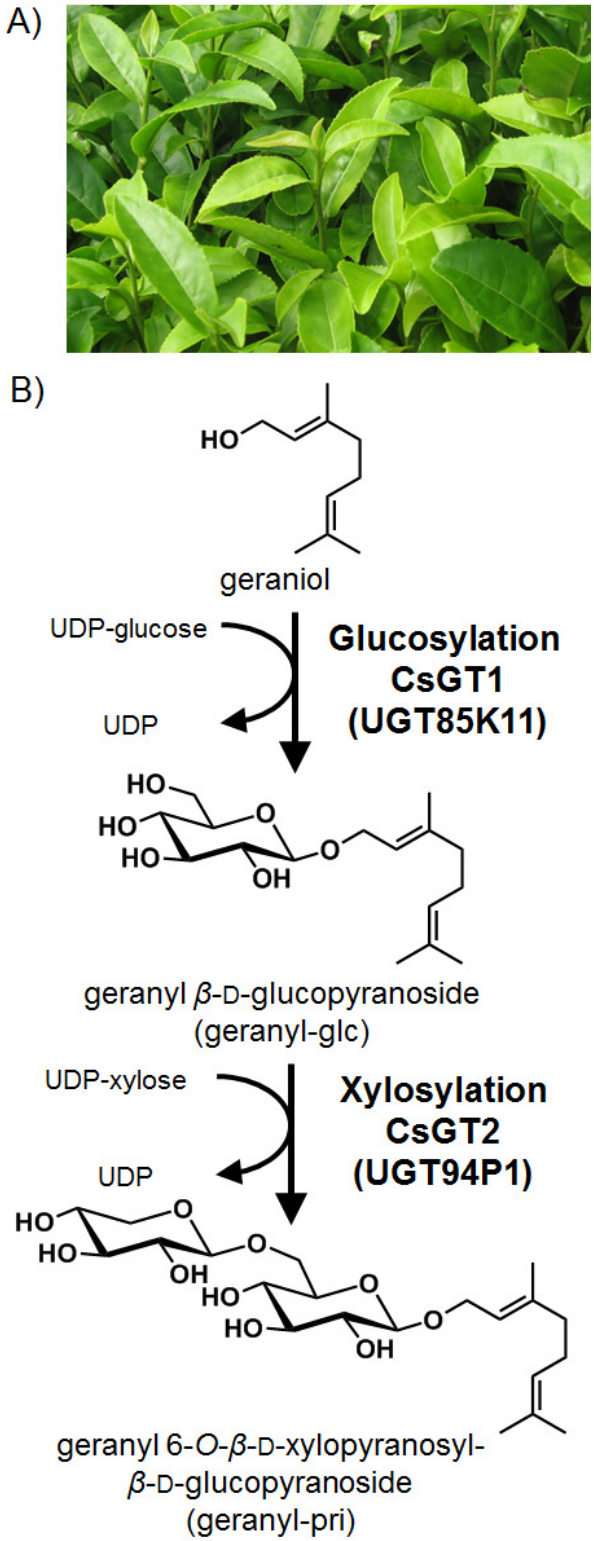


Figure 1. Metabolism of volatile organic compounds in *Camellia sinensis.* A) Photo of young leaves. (B) Biosynthesis pathway of geranyl *β*-primeveroside (geranyl 6-*O*-*β*-D-xylopyranosyl-*β*-D-glucopyranoside) from geraniol. CsGT1 (UGT85K11) and CsGT2 (UGT94P1) are the two glycosyltransferases that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

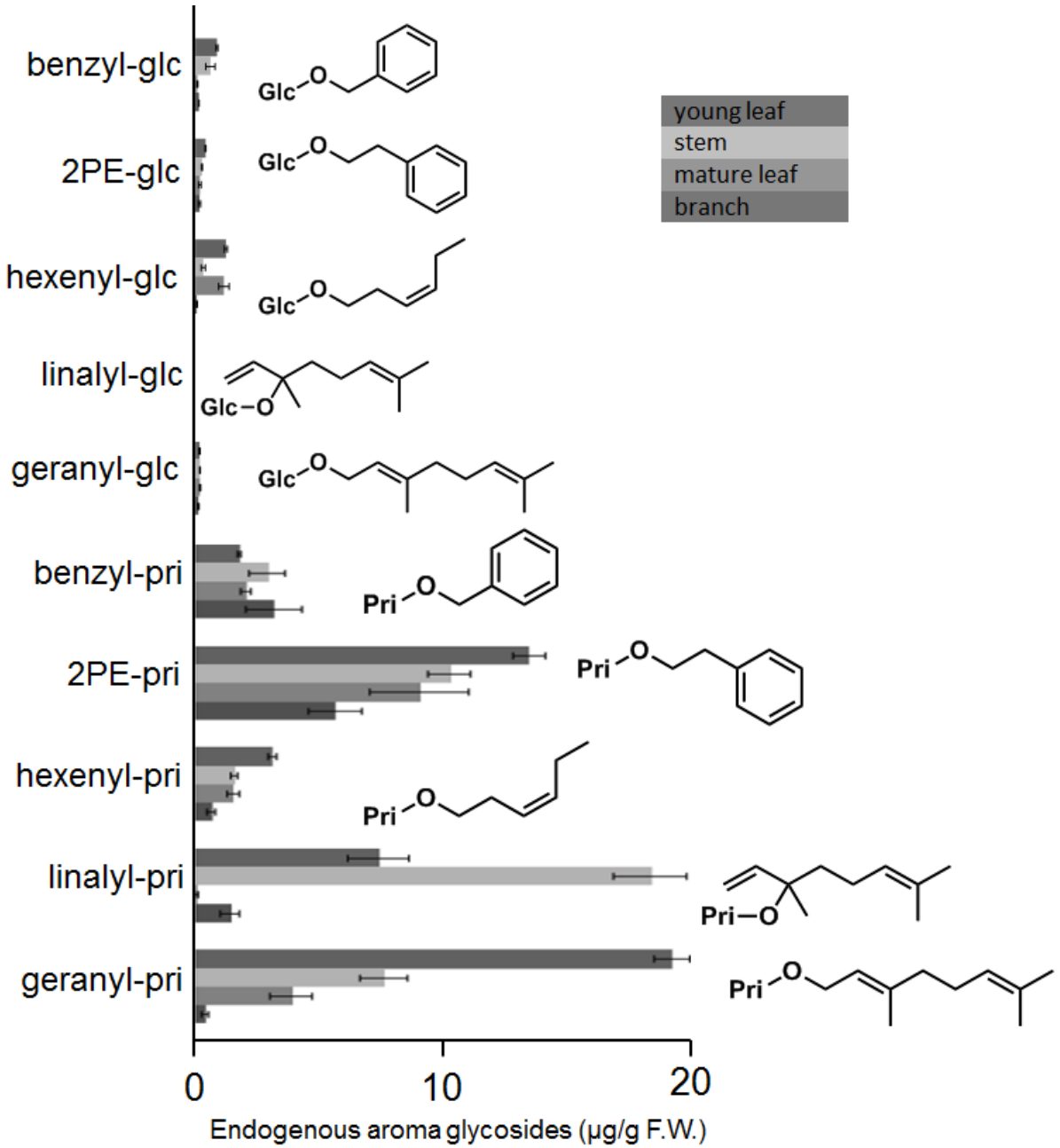


Figure 2. Quantification of aroma monoglycosides and diglycosides in fresh tea leaves Ten glycosides were used as authentic standards to quantify endogenous glycoside in young leaves, young stems, mature leaves, and mature stems of *C. sinensis*: benzyl *β*-D-glucopyranoside (benzyl-glc), benzyl *β*-primeveroside (benzyl-pri), geranyl *β*-D-glucopyranoside (geranyl-glc), geranyl *β*-primeveroside (geranyl-pri), (*Z*)-3-hexenyl *β*-D-glucopyranoside (hexenyl-glu), (*Z*)-3-hexenyl *β*-primeveroside (hexenyl-pri), linalyl *β*-D-glucopyranoside (linalyl-glc), linalyl *β*-primeveroside (linalyl-pri), 2-phenylethyl *β*-D-glucopyranoside (2PE-glc), and 2-phenylethyl *β*-primeveroside (2PE-pri). Data are presented as mean ± SD (n = 3).

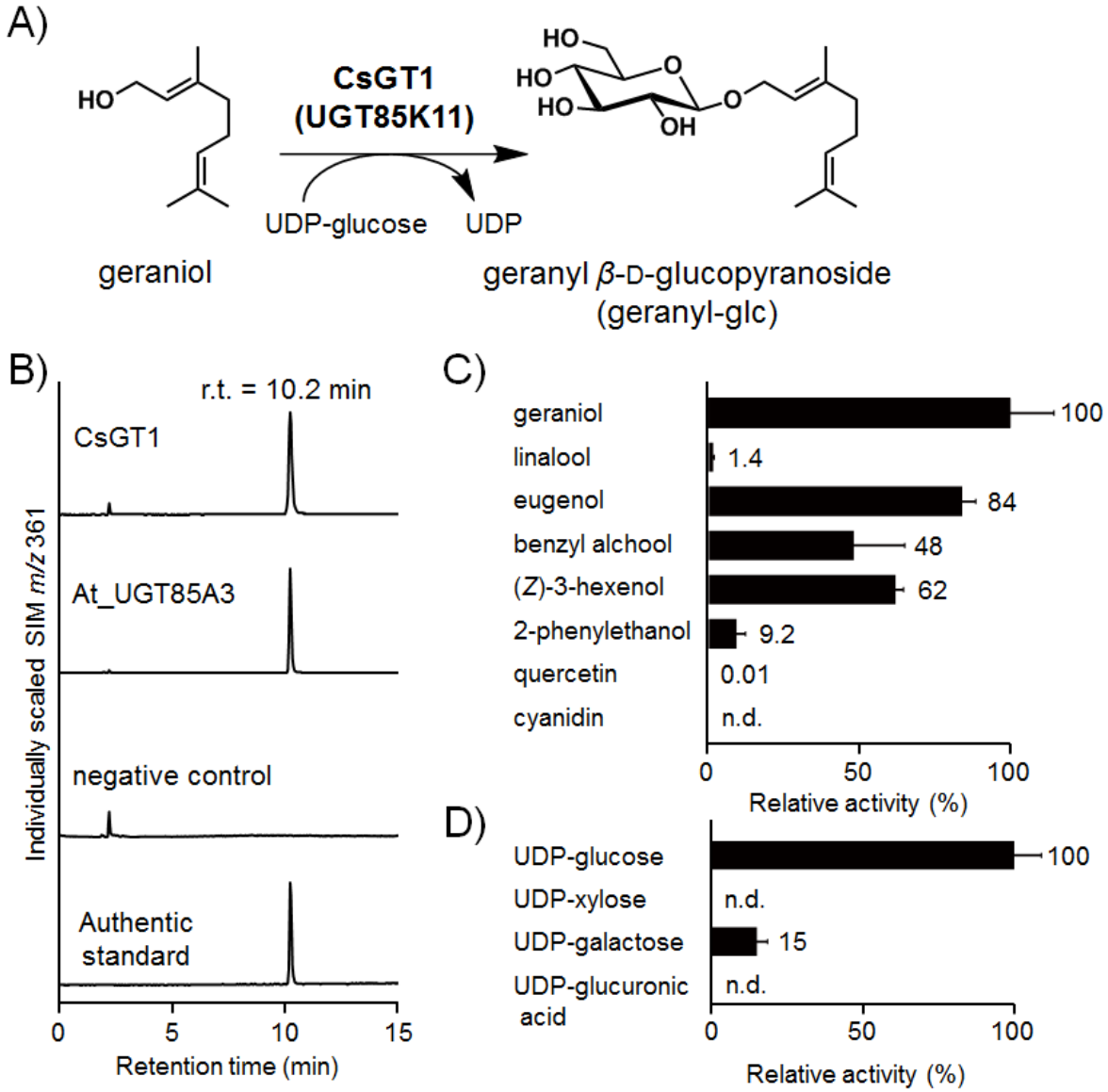


Figure 3. Biochemical characterization of CsGT1 and At_UGT85As. A) CsGT1 catalyzes the glucosylation of geraniol to produce geranyl *β*-D-glucopyranoside (geranyl-glc). B) LC-MS analysis of the enzymatic product of CsGT1 (UGT85K11), At_UGT85A1 (At1g22400), and At_UGT85A3 (At1g22380) compared with the 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 ± SD (n = 3).

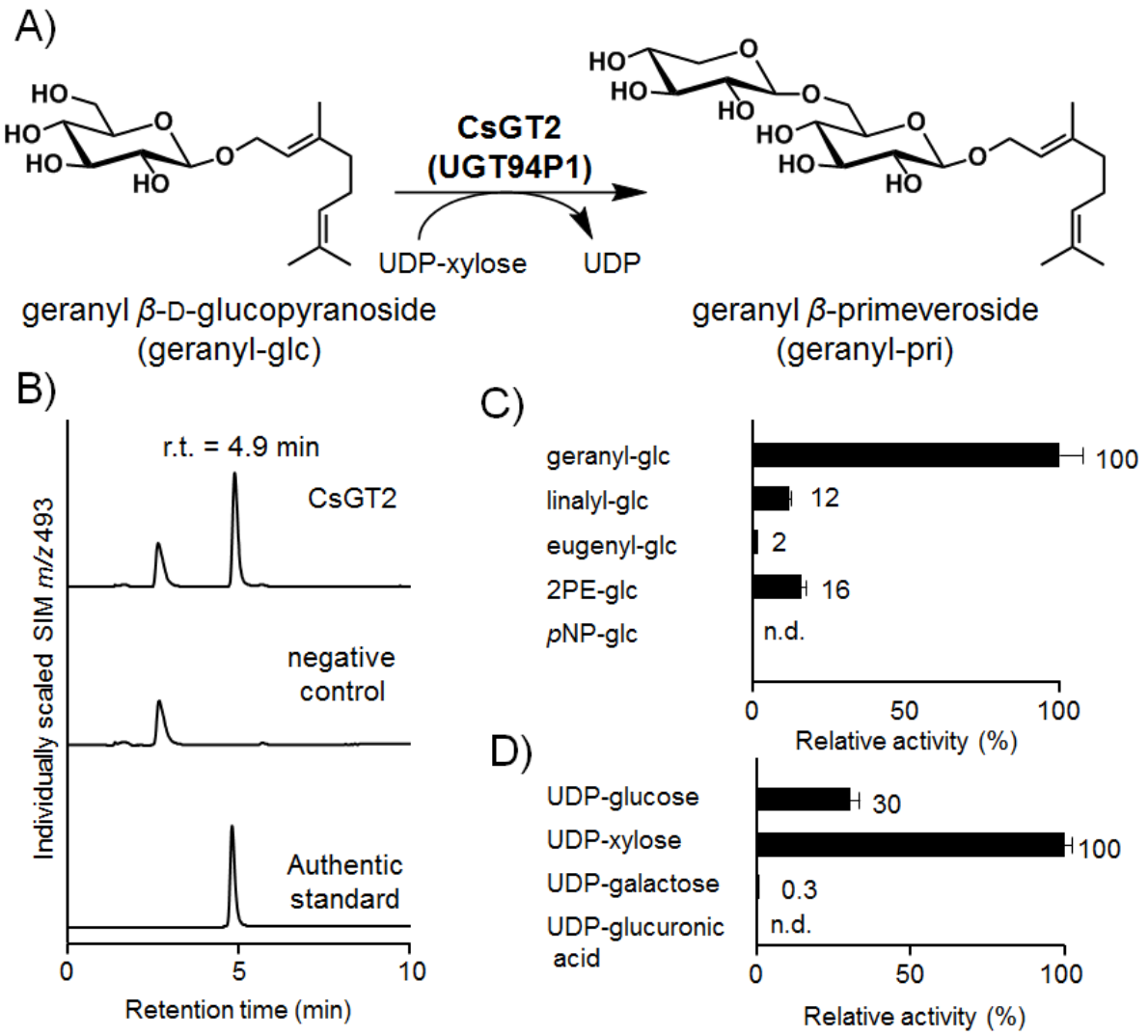


Figure 4. Biochemical characterization of CsGT2. A) CsGT2 (UGT94P1) catalyzes the xylosylation of geranyl-glc into geranyl *β*-primeveroside (geranyl 6-*O*-*β*-D-xylopyranosyl-*β*-D-glucopyranosides). B) LC-MS analysis of the enzymatic product of CsGT2 (UGT94P1) compared with an authentic standard (geranyl-pri)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

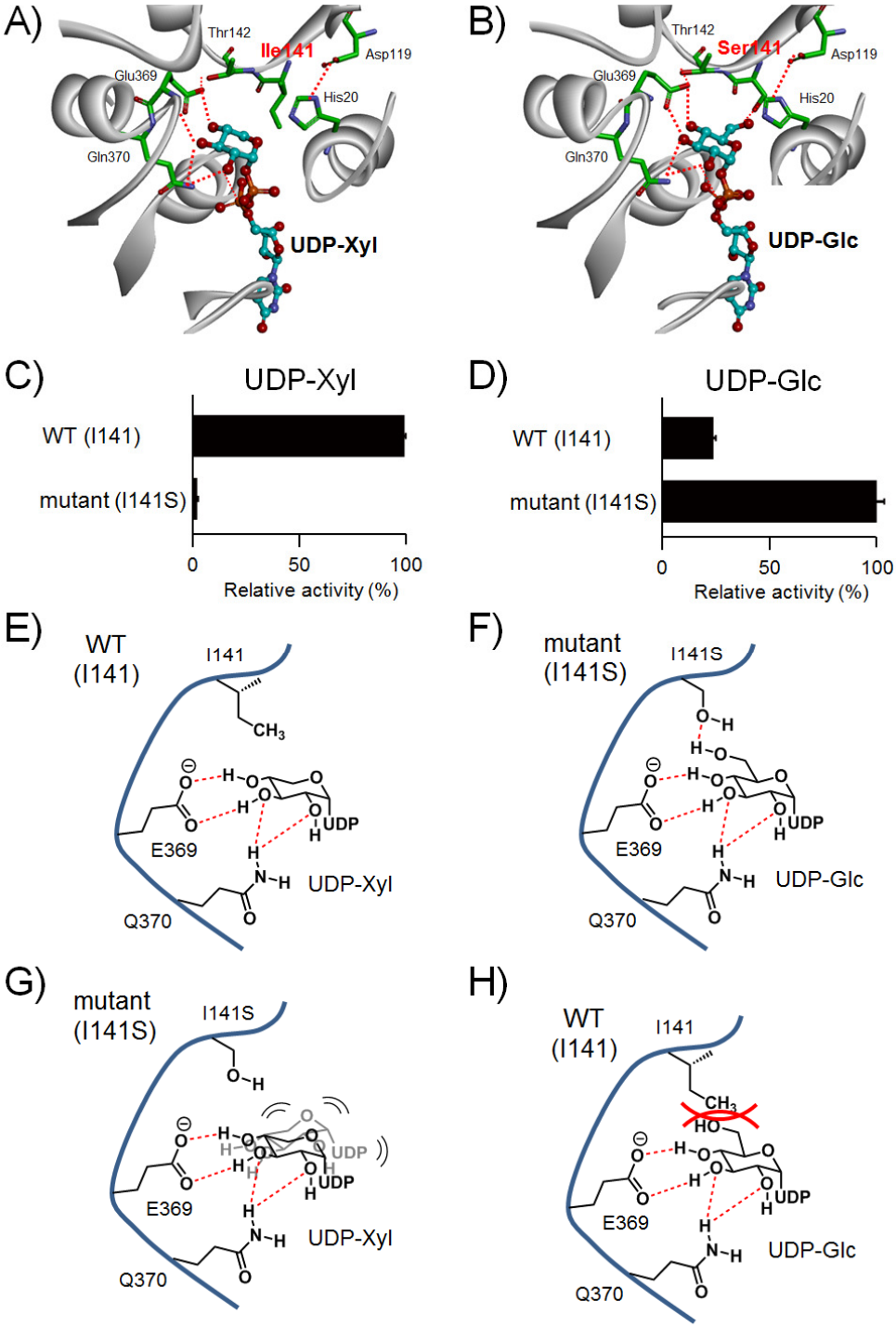


Figure 5. Structural Comparison on the sugar-donor specificity of CsGT2 and its mutant, CsGT2 (I141S). A) Homology model of UDP-xylose-bound CsGT2. B) Homology model of UDP-glucose-bound CsGT2 (I141S). For homology models, important amino acid residues on the active site are drawn at the stick form and 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 geranyl-glc as a sugar acceptor. D) Relative activity of CsGT2 and CsGT2 (I141S) toward UDP-glucose with geranyl-glc as a sugar acceptor. Data are presented as mean ± 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.

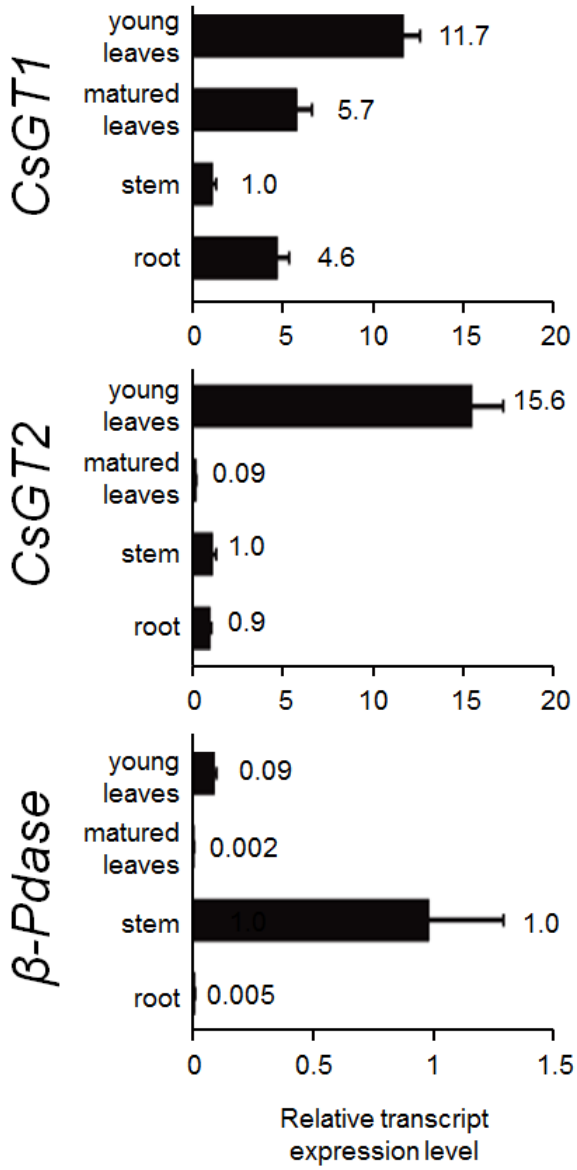


Figure 6. Relative transcript abundance of *CsGT1, CsGT2,* and *β*-PD in various organs (young leaves, matures leaves, stem, root, and flower) of *C. sinensis* . Transcript abundance was measured by qRT-PCR and normalized to the internal reference *ribosomal 18S*. The expression level of each gene in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

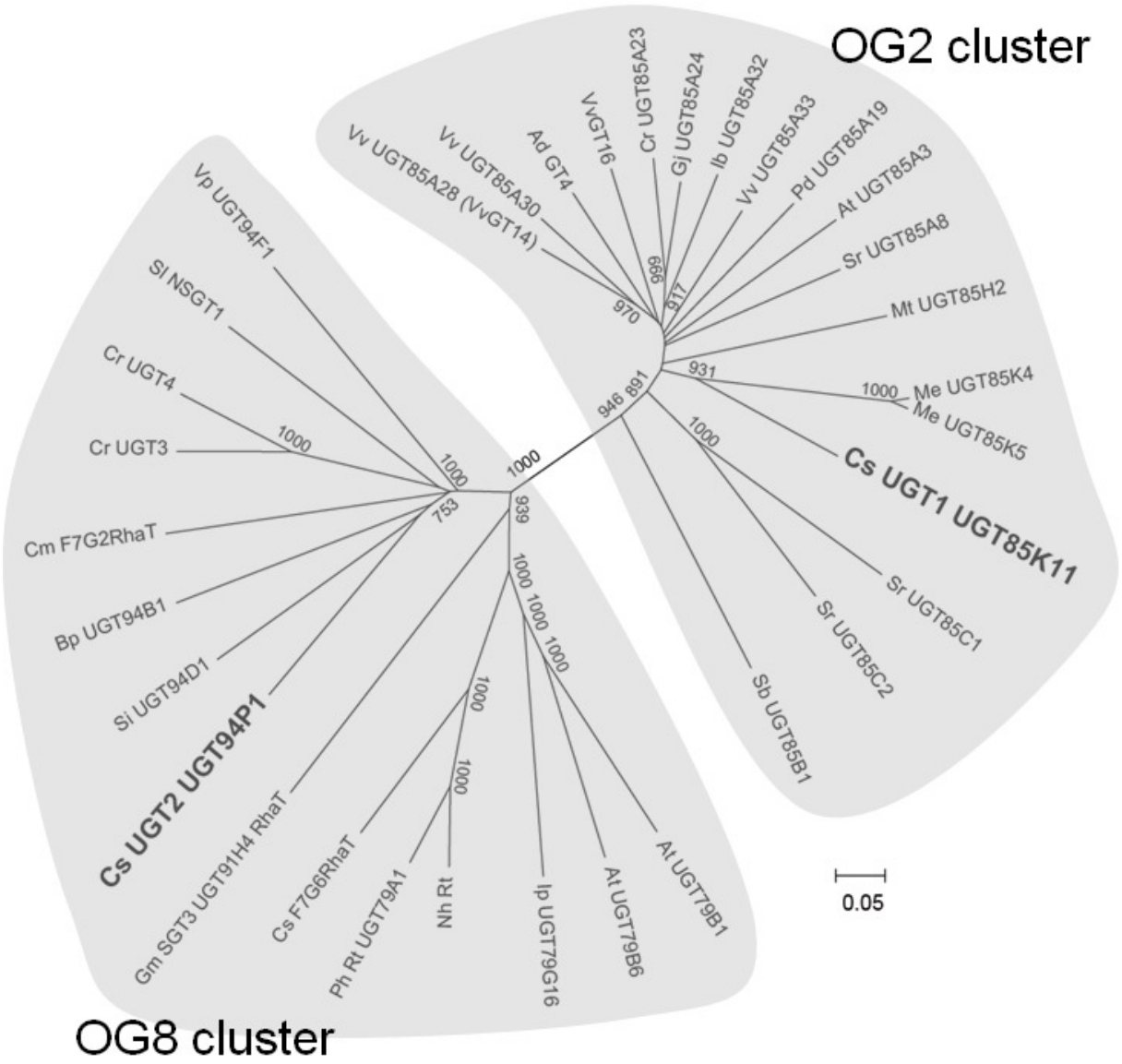


Figure 7. Phylogeny of UGT-glucosyltransferase OG2 and OG8 family. All other UGT sequences were for *Arabidopsis thaliana* available on The Arabidopsis Information Resource (TAIR) website. High bootstrap values (>750) are indicated on the branches (1000 replicates). The Vv_UGT85A28 described here was found to be identical to VvGT14 (Bönisch et al., 2014b).

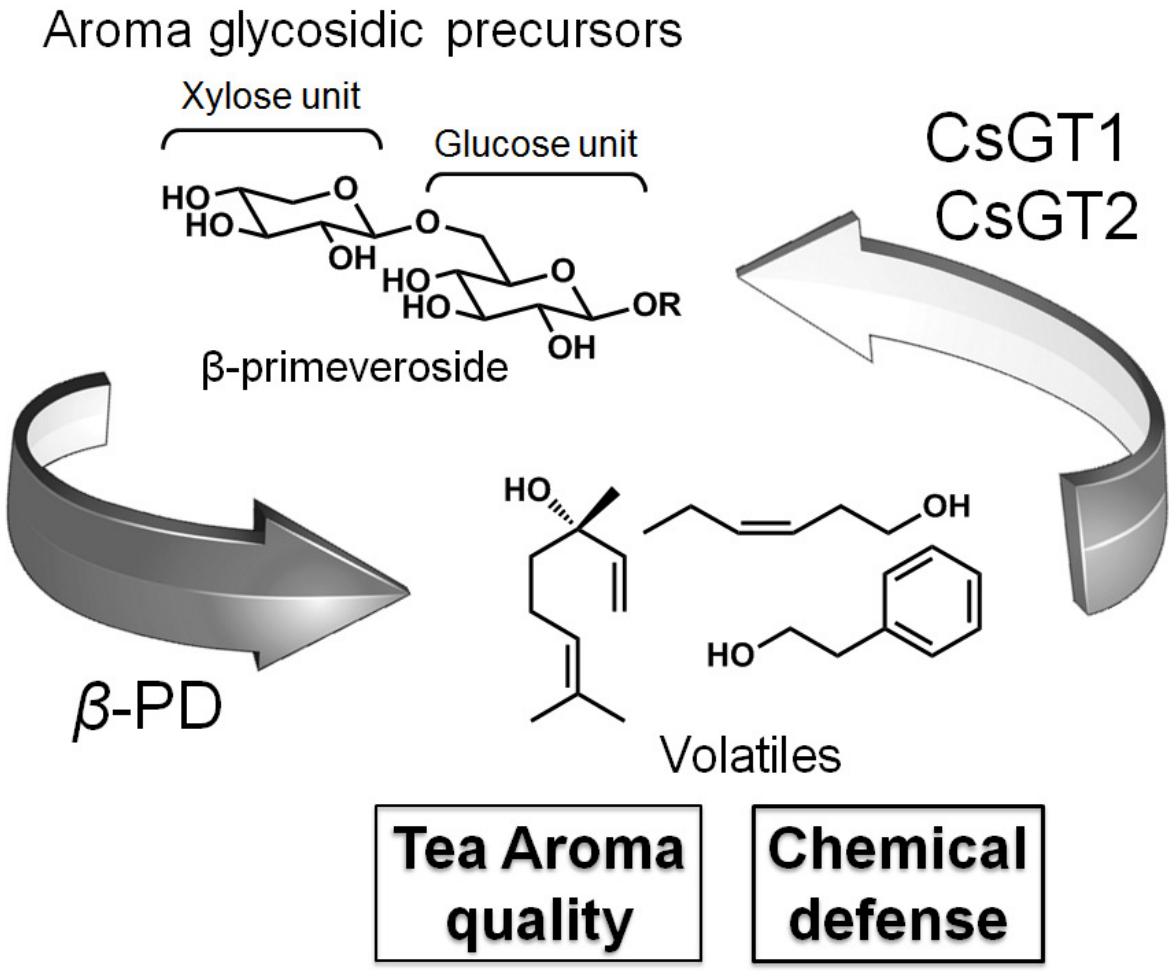


Figure 8. Schematic illustration of a mode-of-action for the metabolism of aroma volatiles by CsGT1, CsGT2, and *β*-PD in *C. sinensis.*

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) equipped with Capcell Pak UG120 C_{18} reversed phase column (2.0 mm i.d. x 150 mm, 5 μm, Shiseidou; 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) 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 μl) was injected, and the peptides were trapped on a C18 column, Monolith Trap (50 μm × 150 mm, Hitachi High-Technologies). Peptide separation was achieved using a packed nano-capillary column (capillary-Ex nano mono cap, 0.05 × 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) and the MASCOT database (Matrix Science, 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) system with negative electrospray ionization mode; dry temperature, 250°C; a flow rate, 0.2 ml/min and column oven temperature, 40 ^{o}C. HPLC separation was performed according to the following conditions: Analysis of CsGT1 enzymatic products was performed using a CAPCELL PAK C_{18} UG120 column (2.0 mm I.D.×150 mm; 5 μm) (Shiseido, 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_2O containing 0.05% (v/v) formic acid, B: acetonitrile) Analysis of CsGT2 enzymatic products was performed using a CAPCELL PAK C_{18} 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_2O 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), and then digested with *Xho*I and *Bgl*II for CsGT1, and *Nde*I and *BamH*I for CsGT2 and the resulting DNA fragments were ligated into a pET-15b vector (Merck Millipore; www.merckmillipore.com) previously digested with *Xho*I and *BamH*I for CsGT1, and *Nde*I and *BamH*I 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 Hisx6 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).

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 *Nde*I and *Bgl*II, and the resulting DNA fragments were ligated with pET-15b at *Nde*I and *BamH*I 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 (7000*g*, 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 (7,000*g*, 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 (7,000*g*, 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) 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/) in Molecular Imager ChemiDoc XRS+ (BioRad; 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) in a Tissue-Tek Cryomold (Sakura Finetek, 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) 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). 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 µm; raster, 15 µm; 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 *β*-primeverosides in young leaves and mature leaves of *C.sinensis.*

	aroma glycosides (μg g^{-1} fresh weight)		total amounts of aroma glycosides (μg)	
	young leaves	mature leaves	young leaves (FW 0.192 ± 0.016 g)	mature leaves (FW 0.808± 0.057 g)
benzyl-glc	0.46 ± 0.022	0.044 ±0.01	0.088 ± 0.008	0.036 ± 0.008
2PE-glc	0.22 ± 0.010	0.098 ±0.03	0.042 ± 0.004	0.079 ± 0.025
hexenyl-glc	0.65 ± 0.030	0.609 ±0.11	0.125 ± 0.012	0.493 ± 0.096
linalyl-glc	n.d.	n.d.	nd	nd
geranyl-glc	0.092 ± 0.015	0.101 ±0.01	0.018 ± 0.003	0.081 ± 0.010
benzyl-pri	1.75 ± 0.082	2.031 ±0.20	0.336 ± 0.032	1.64 ± 0199
2PE-pri	13.0 ± 0.61	8.78 ±1.9	2.50 ± 0.24	7.09 ± 1.63
hexenyl-pri	3.04 ± 0.14	1.505 ±0.23	0.584 ± 0.056	1.22 ± 0.21
linalyl-pri	7.19 ± 1.20	0.079 ±0.06	1.38 ± 0.26	0.064 ± 0.049
geranyl-pri	18.6 ± 0.67	3.789 ±0.84	3.57 ± 0.32	3.06 ± 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^{-1})	Yield (%)	Purification fold (-fold)
1. Crude extract	681	457	0.67	100.0	1.0
2. 30-70%$(NH_4)_2SO_4$	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 μmol of geranyl *β*-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'-ATGGGATCTCAGATCATTCATAACT-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-*Xho*I-CsGT1-FW	5'-CACCCTCGAGATGGGTAGCAGAAAGCAG-3'
CsGT1-*Bgl*II-stop-RV	5'-AGATCTTTAGTATTGCTCACAATAGTGAAGAGC-3'
CACC-*Nde*I-CsGT2-FW	5'-CACCCATATGGATTCAAAAAAGAGCAAAATG-3'
CsGT2-*BamH*I-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-*β-PD* -FW	5'-AAGGATCCCCAGAGGGTCTA-3'
qRT-*β-PD* -RV	5'-TCCGAACCTTTGGGTGTAAC -3'
qRT-Cs18SrRNA-FW	5'-CACGGGGAGGTAGTGACAAT-3'
qRT-Cs18SrRNA-RV	5'-CCTCCAATGGATCCTCGTTA-3'

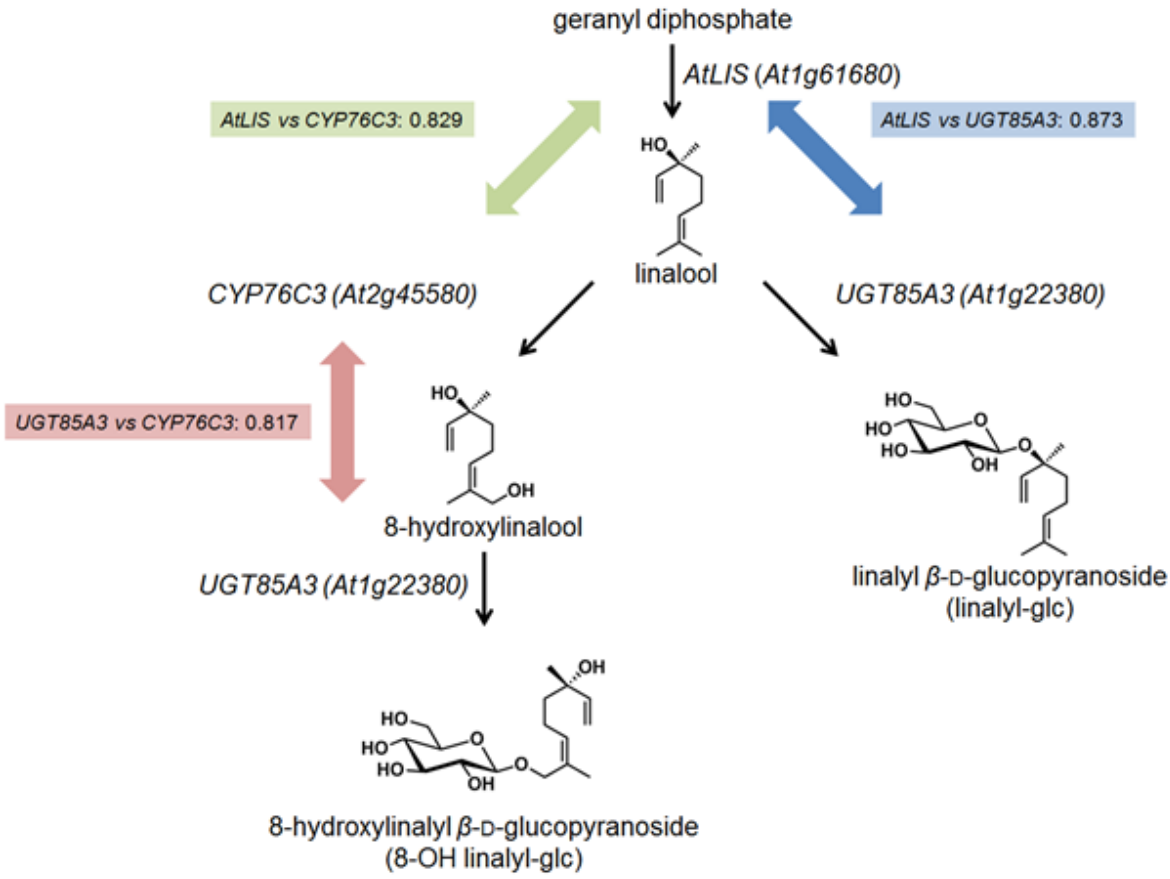
Supplemental Table S4. Composition of the buffers and solutions for purification and identification of CsGT2 protein

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) polyvinylpolypyrrolidone and 10% (w/v) DOWEX1x2(Cl^-)
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_4HCO_3/acetonitrile (1:1 v/v)
Solution G	10 mM dithiothreitol/50 mM NH_4HCO_3
Solution H	55 mM iodoacetoamide/50 mM NH_4HCO_3
Solution I	50% acetonitrile containing 1% formic acid

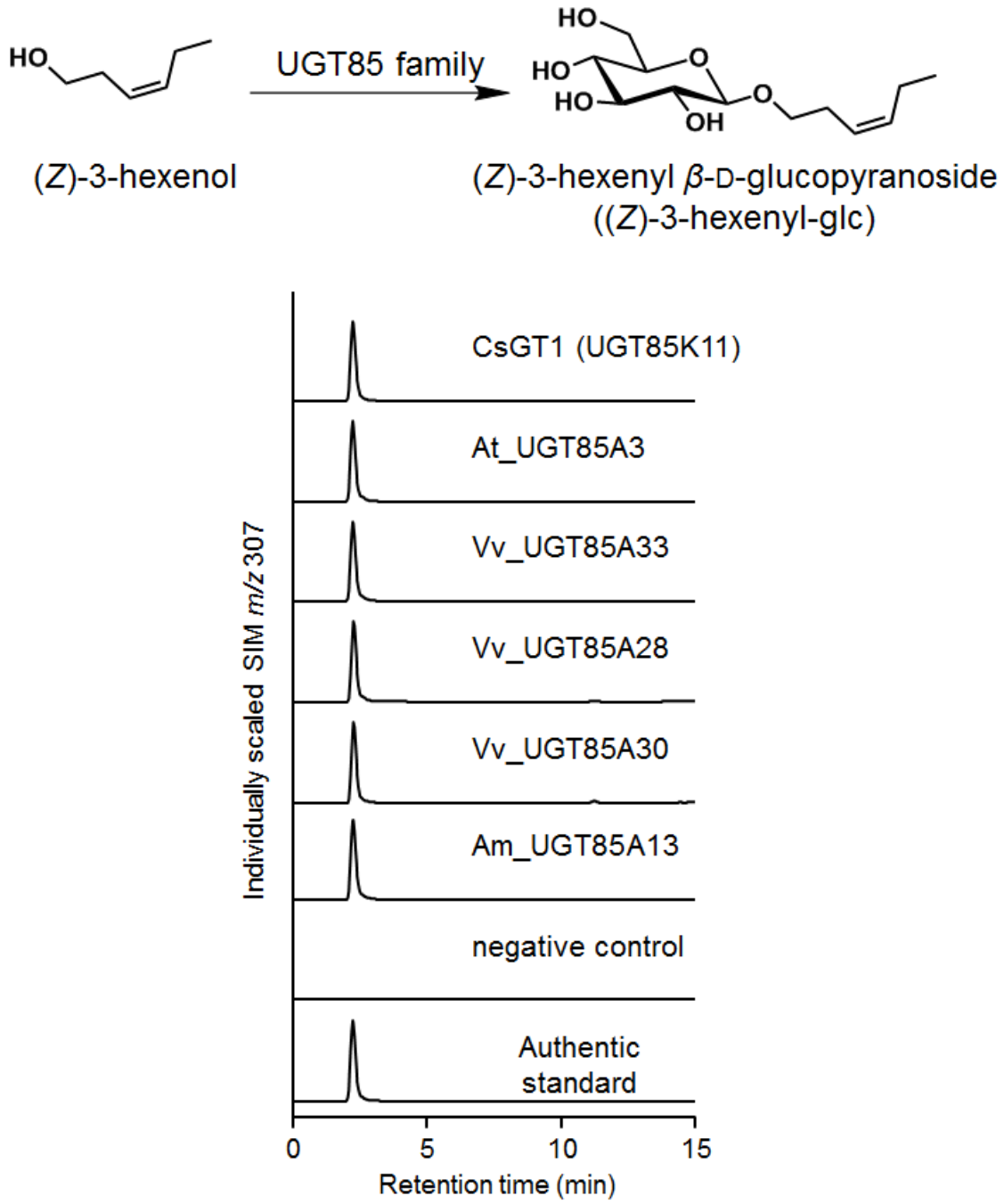
Supplemental Table S5. GenBank accession numbers used for construction of the phylogenetic tree in Figure 7.

Abbreviation	Plant species		Accession NO
Ad_GT4	*Actinidia deliciosa*	fuzzy kiwifruit	AIL51400
At_UGT79B1	*Arabidopsis thaliana*	Arabidopsis	NP_200217
At_UGT79B6	*Arabidopsis thaliana*	Arabidopsis	NP_200212
At_UGT85A3	*Arabidopsis thaliana*	Arabidopsis	NP_173655
Bp_UGT94B1	*Bellis perennis*	daisy	Q5NTH0
Cm_F7G2RhaT	*Citrus maxima*	pomelo	Q8GVE3
Cr_UGT3	*Catharanthus roseus*	madagascar periwinkle	BAH80312
Cr_UGT4	*Catharanthus roseus*	madagascar periwinkle	BAH80313
Cr_UGT85A23	*Catharanthus roseus*	madagascar periwinkle	F8WLS6
Cs_F7G6RhaT	*Citrus sinensis*	orange	NP_001275829
Cs_UGT1_UGT85K11	*Camellia sinensis*	tea plant	BAO51834
Cs_UGT2_UGT94P1	*Camellia sinensis*	tea plant	BAO51835
Gj_UGT85A24	*Gardenia jasminoides*	gardenia	F8WKW1
Gm_SGT3_UGT91H4_RhaT	*Glycine max*	soybean	NP_001240857
Ib_UGT85A32	*Ipomoea batatas*	sweet potato	BAO51842
Ip_UGT79G16	*Ipomoea purpurea*	morning glory	Q53UH5
Me_UGT85K4	*Manihot esculenta*	cassava	AEO45781
Me_UGT85K5	*Manihot esculenta*	cassava	AEO45782
Mt_UGT85H2	*Medicago truncatula*	barrel clover	XP_003618665
Nh_Rt	*Nierembergia sp. NB17*	cupflower	BAC10994
Pd_UGT85A19	*Prunus dulcis*	almond	ABV68925
Ph_Rt (UGT79A1)	*Petunia x hybrida*	petunia	CAA50376
Sb_UGT85B1	*Sorghum bicolor*	sorghum	XP_002463518
Si_UGT94D1	*Sesamum indicum*	sesame	BAF99027
Sl_NSGT1	*Solanum lycopersicum*	tomato	AGO03777
Sr_UGT85A8	*Stevia rebaudiana*	stevia	AAR06913
Sr_UGT85C1	*Stevia rebaudiana*	stevia	AAR06922
Sr_UGT85C2	*Stevia rebaudiana*	stevia	AAR06916
Vp_UGT94F1	*Veronica persica*	persian speedwell	BAI44133
VvGT16	*Vitis vinifera*	grapevine	XP_002263158
Vv_UGT85A28 (VvGT14)	*Vitis vinifera*	grapevine	BAO51844
Vv_UGT85A30	*Vitis vinifera*	grapevine	NP_001277170
Vv_UGT85A33	*Vitis vinifera*	grapevine	NP_001277168

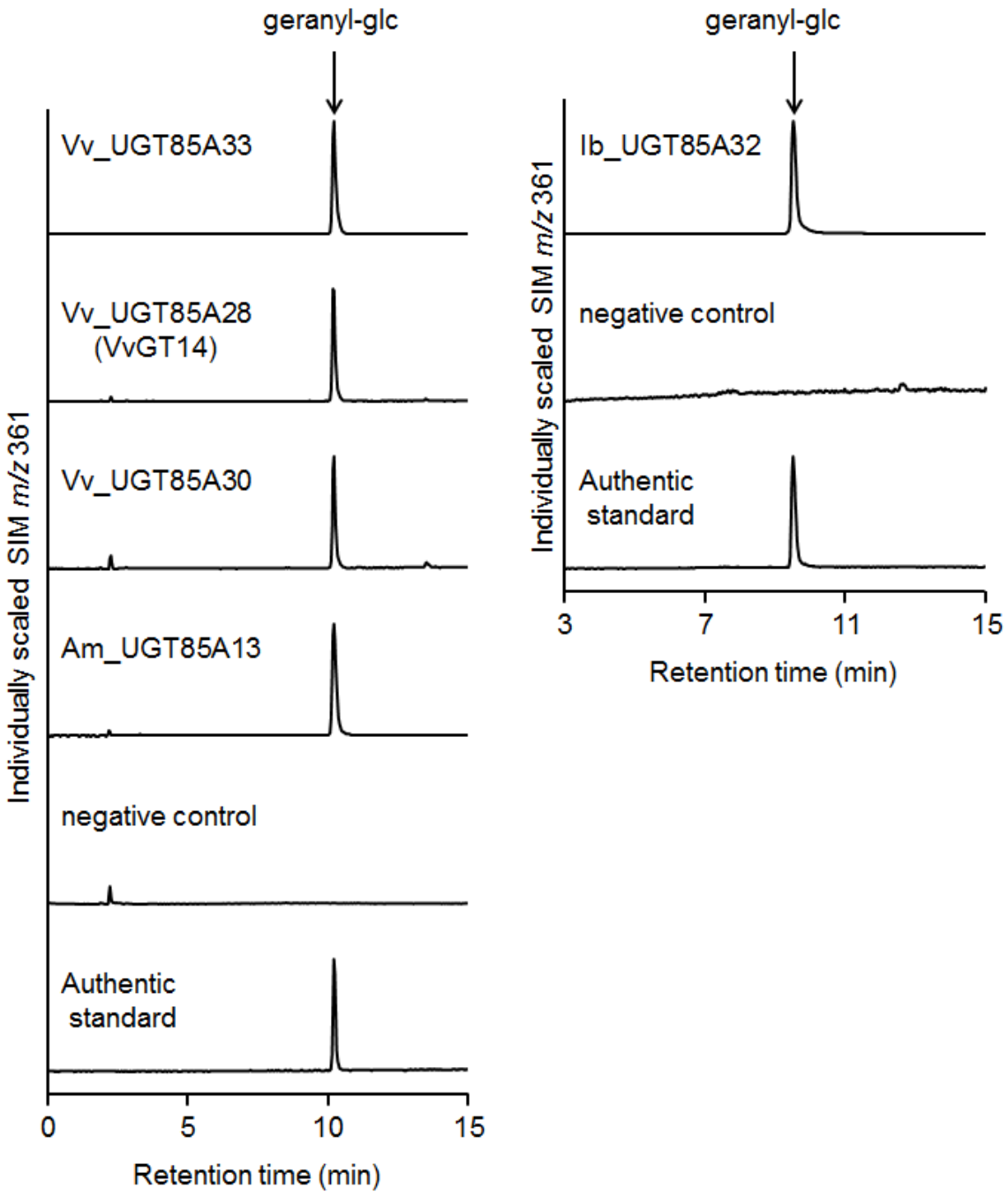
Supplemental Figures



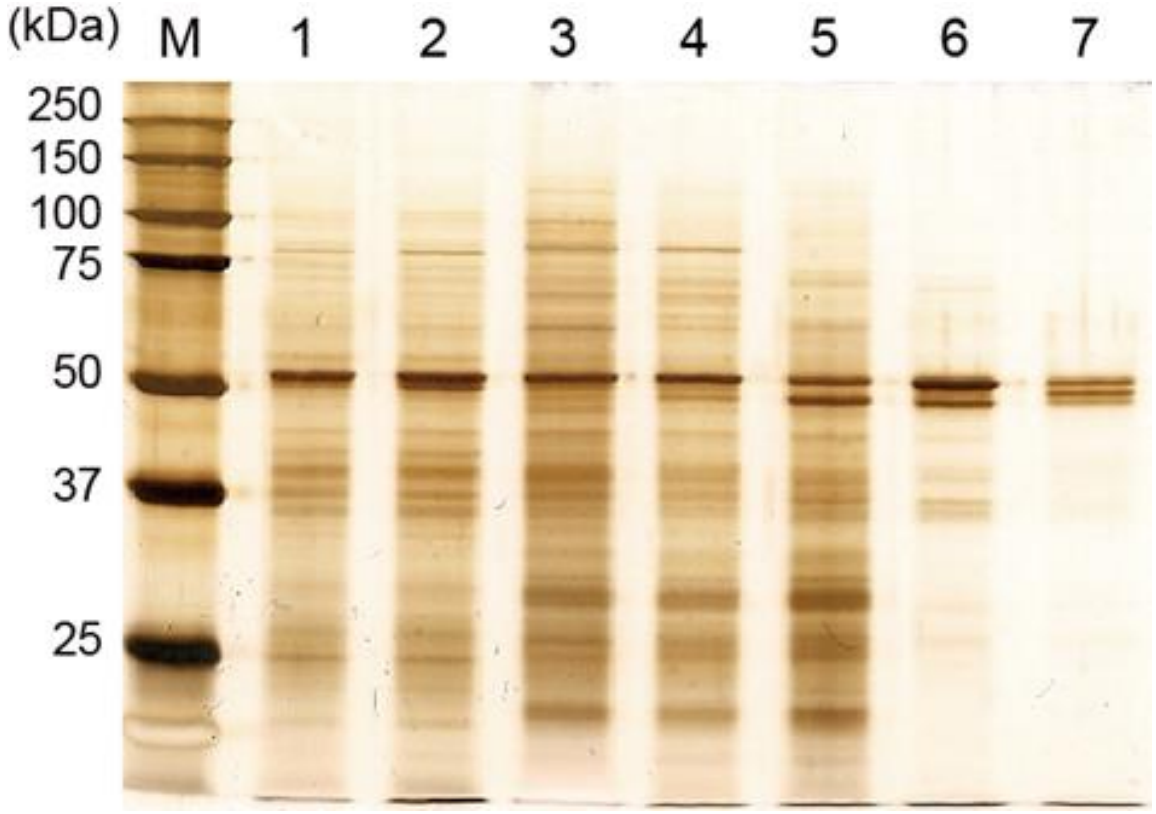
Supplemental Figure S1. **Co-expression analysis of AtLIS, CYP76Cs and UGT85A3 by ATTDII (Gene coexpression database, http://atted.jp/) with Ver. C4.1.**



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 staining by a silver staining methods (Sil-best stain one, Nacalai, Kyoto, Japan)

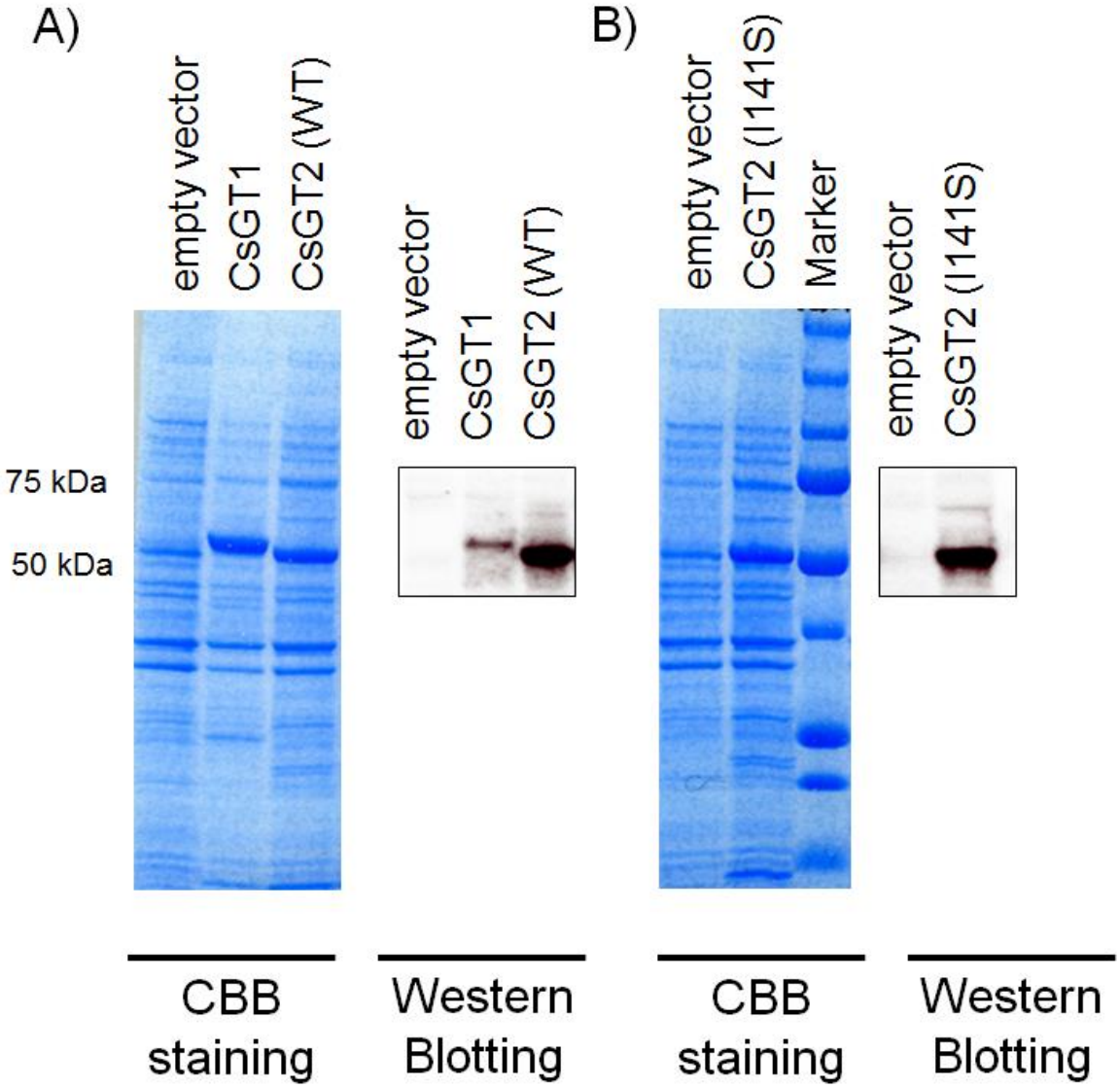
A)

Sequence	Charge	Observed *m/z*	*Mr* (Calc)
FPEVEKVELEEALPK	+3	586.3224	1755.9294
GLVVEGWAPQAR	+2	641.852	1281.6829
EEIEEIAHGLELSMVNFIWVVRFPEVEK	+4	840.4036	3357.6948

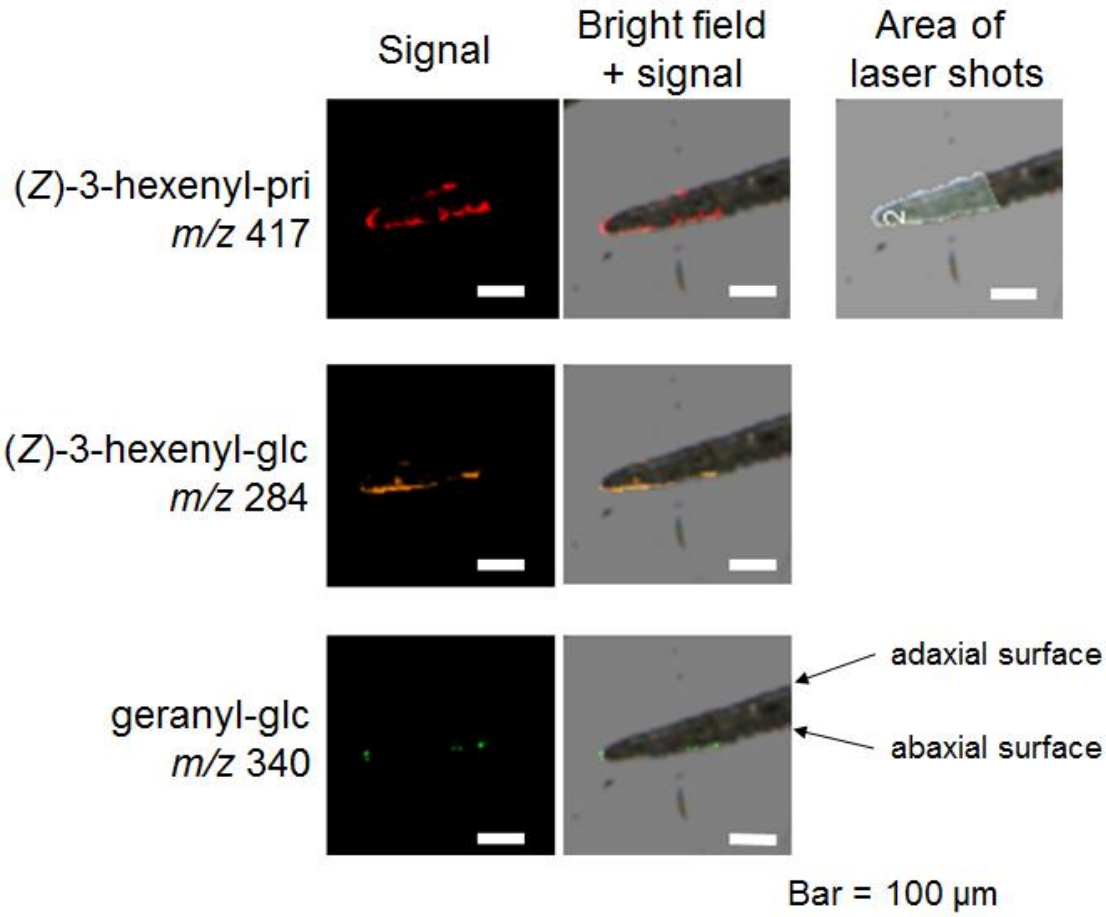
B)

```
  1 ENEHSEIIQW  LDNKGEYSTL  FVSFGSEYFM  SKEEIEEIAH  GLELSMVNFI
 51 WVVRFPEVEK  VELEEALPKG  FIDRVGERGL  VVEGWAPQAR  ILTHSSTGGF
101 VSHCEWNSVL  ESLKFGVPMV  AIPMQYEQPL  NAKLVEEVGV  AAEVNRDING
151 RLNREEIAQV  IRKVVVEKSG  EDIRIKARIF  GDKIRMKGDE  EIDEAVEVLL
201 QLCKDVKLLK  NSNFIEIFKW  VEKVFYFYFI  LFYWVVWYTN  SLIRLYTSRR
251 STNIVFDHVV  LVLDLVALFF  IYIYRFTIIG  VIVSSFCSID  TGGILNMV
```

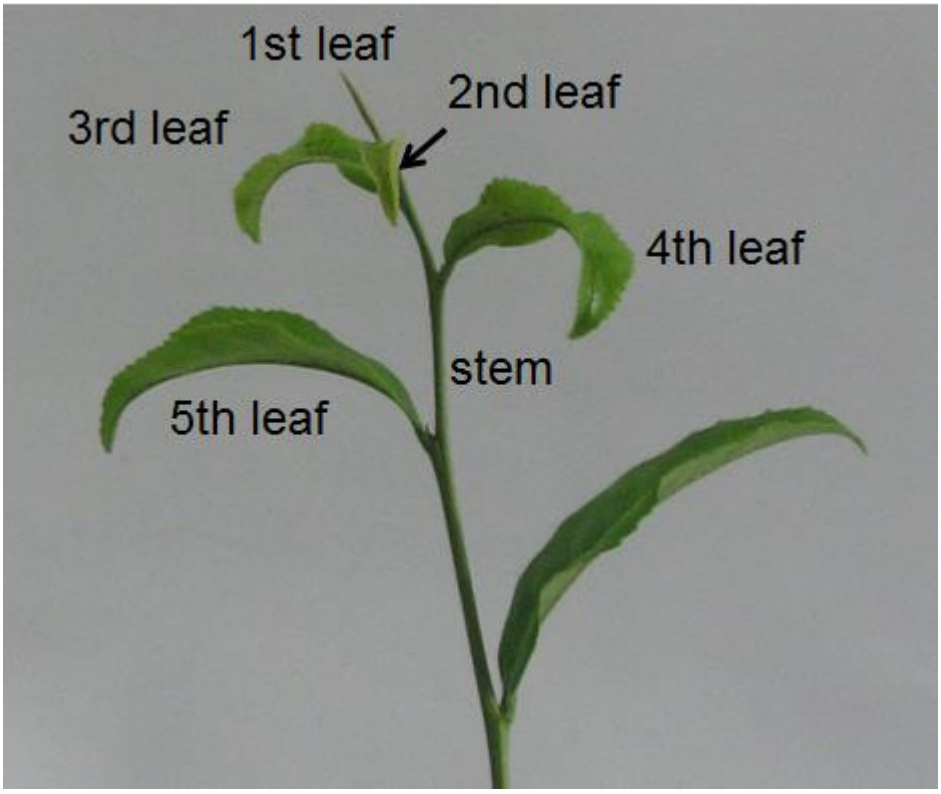
Supplemental Figure S5. Partial peptide sequence of the CsGT2. LC-MS/MS analysis of the partial peptide sequences of CsGT2 (UGT94P1) that purified from tea leaves. A) Three peptide sequences were exactly identified with tea EST database contig 13. B) Predicted amino acid sequence of tea EST database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



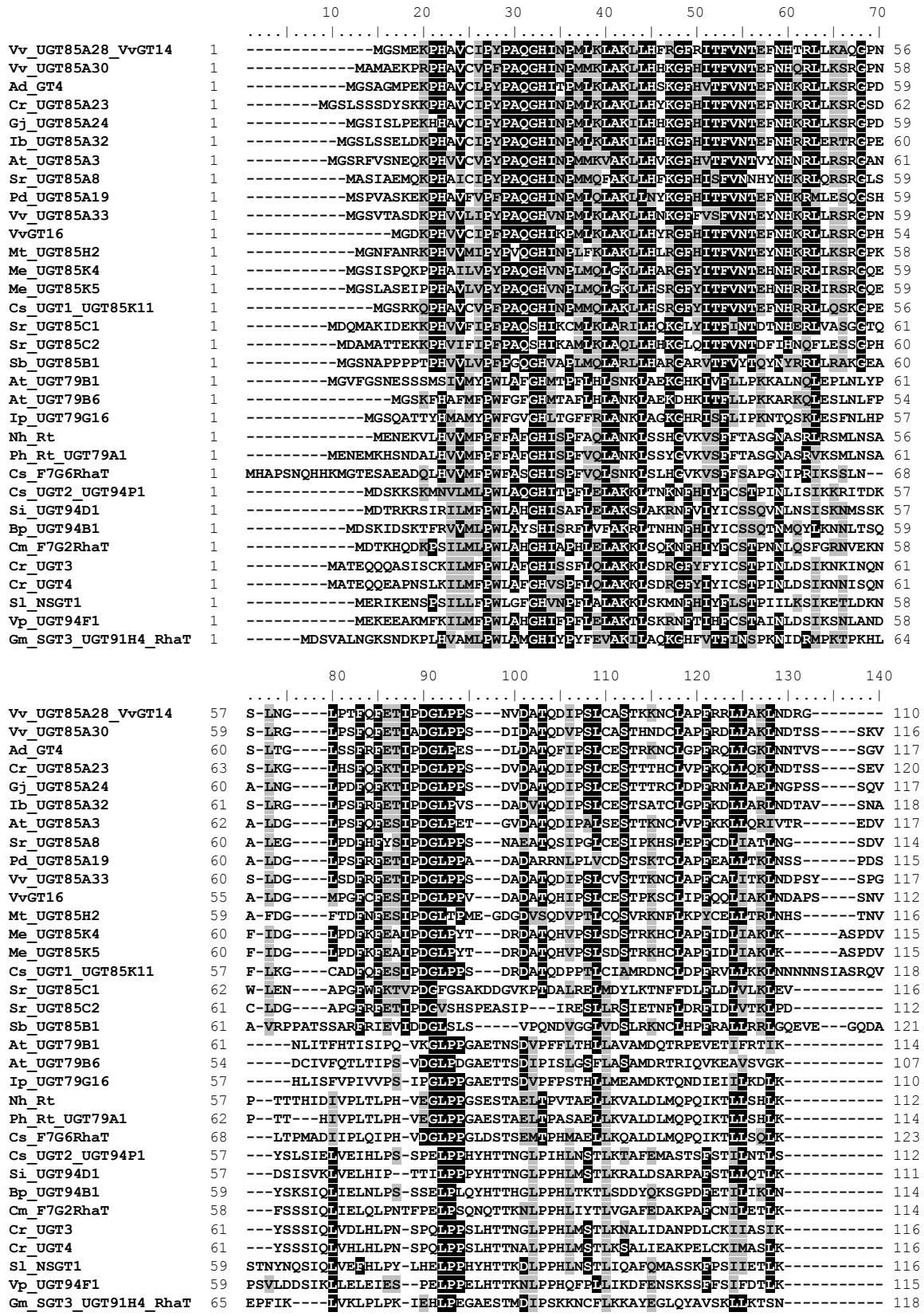
Supplemental Figure S6. Recombinant proteins of a series of CsGTs. Specific bands responsible of CsGT1 and CsGT2 (A), and CsGT2-I141S proteins (B) were detected around the size marker of 50kDa in SDS-PAGE, are roughly consistent with the estimated molecular size of CsGT1 (55.7 kDa) and CsGT2 (51.1 kDa),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



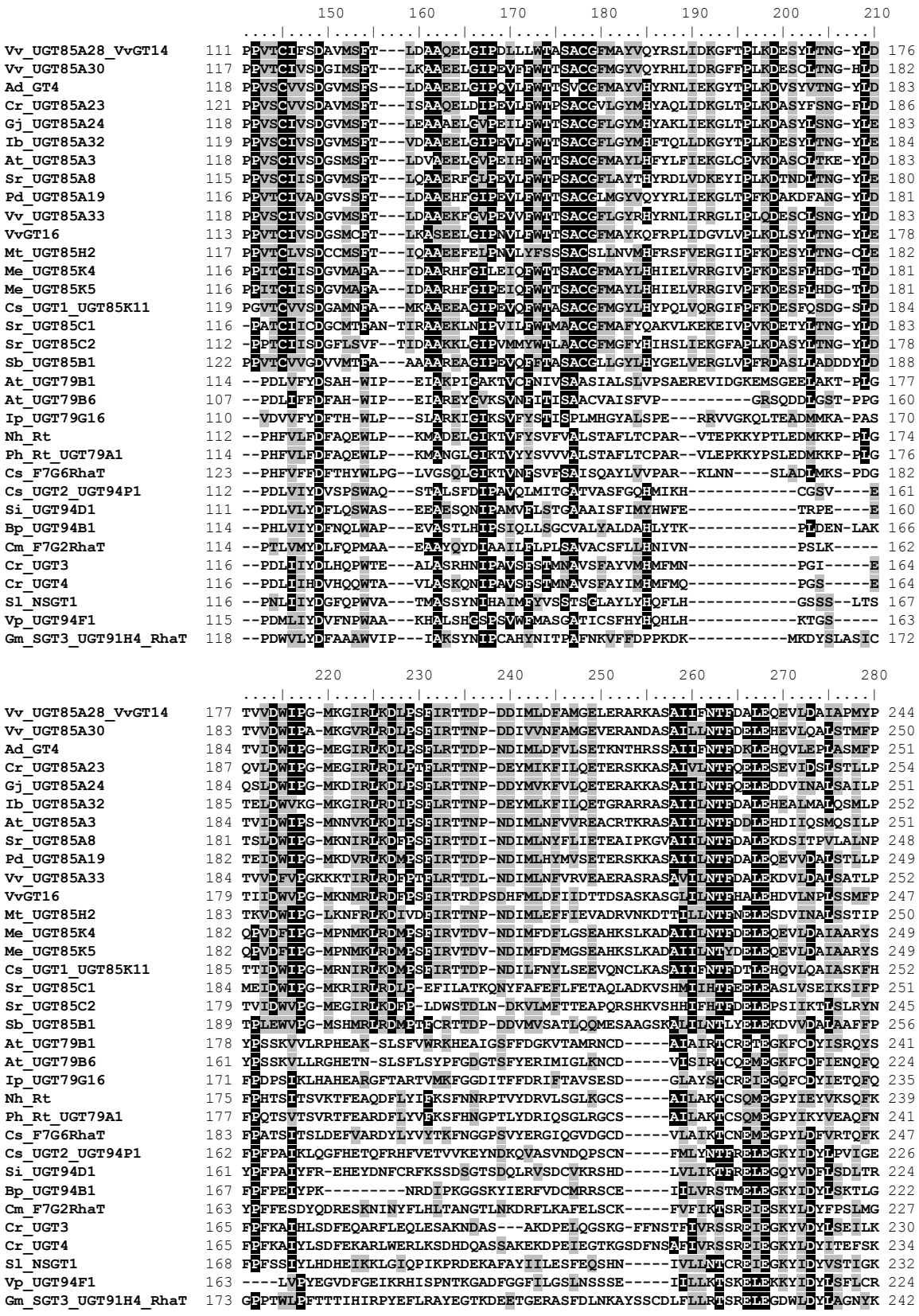
Supplemental Figure S7. Imaging MS of young fresh leaves of *C.sinensis*. Specific signal (red; *m/z* 417) corresponding to (*Z*)-3-hexenyl-pri. Specific signal (orange; *m/z* 284) corresponding to (*Z*)-3-hexenyl-glc. Specific signal (green; *m/z* 340) corresponding to geranyl-glc.



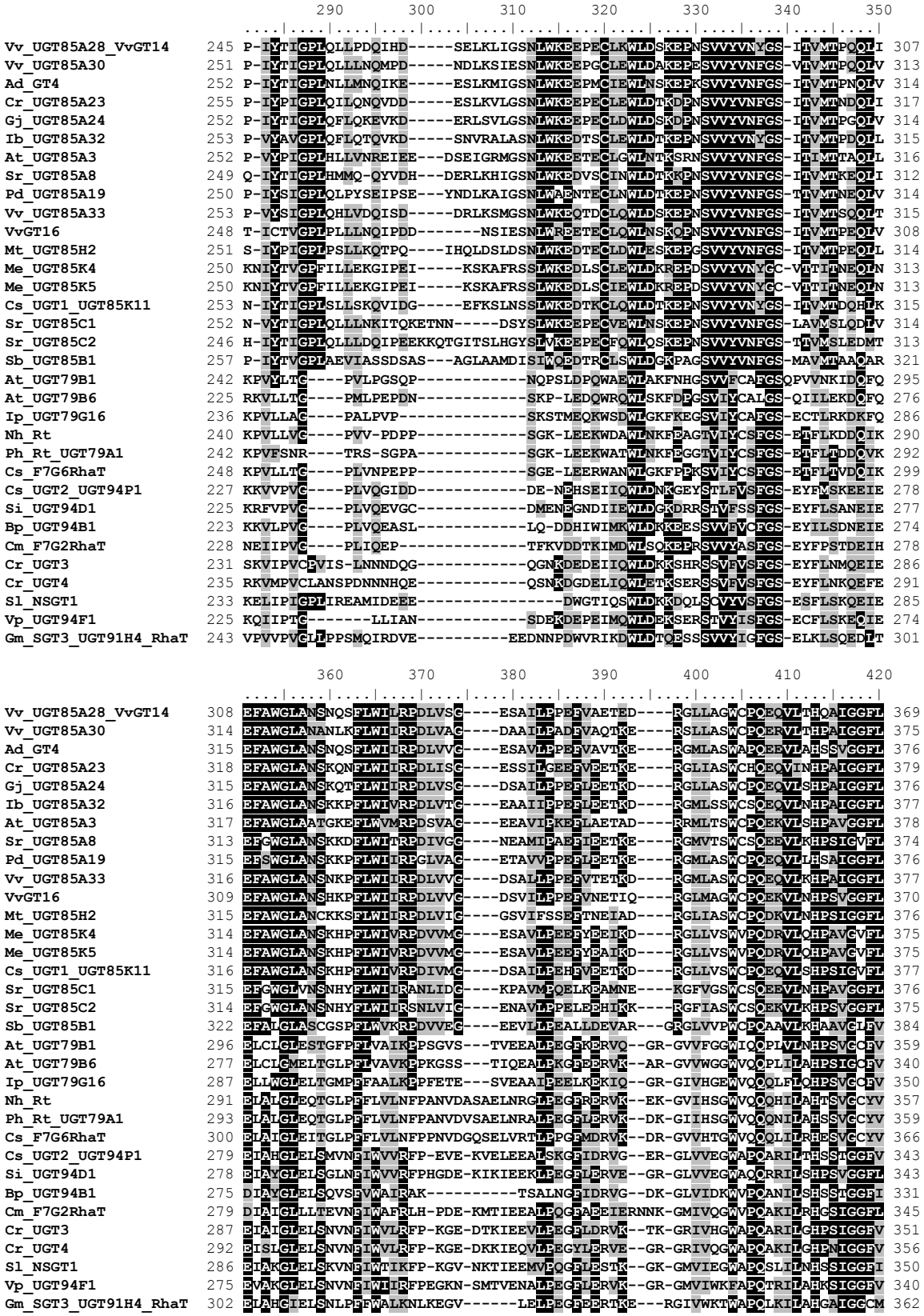
Supplemental Figure S8. Harvesting individual tissues for quantification of endogenous aroma glycosides and quantitative real-time PCR (qRT-PCR) of CsGTs and β-PD. The young and mature developmental stages of leaves and stems were defined as follows: young leaves, which are the first, second, and third 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.



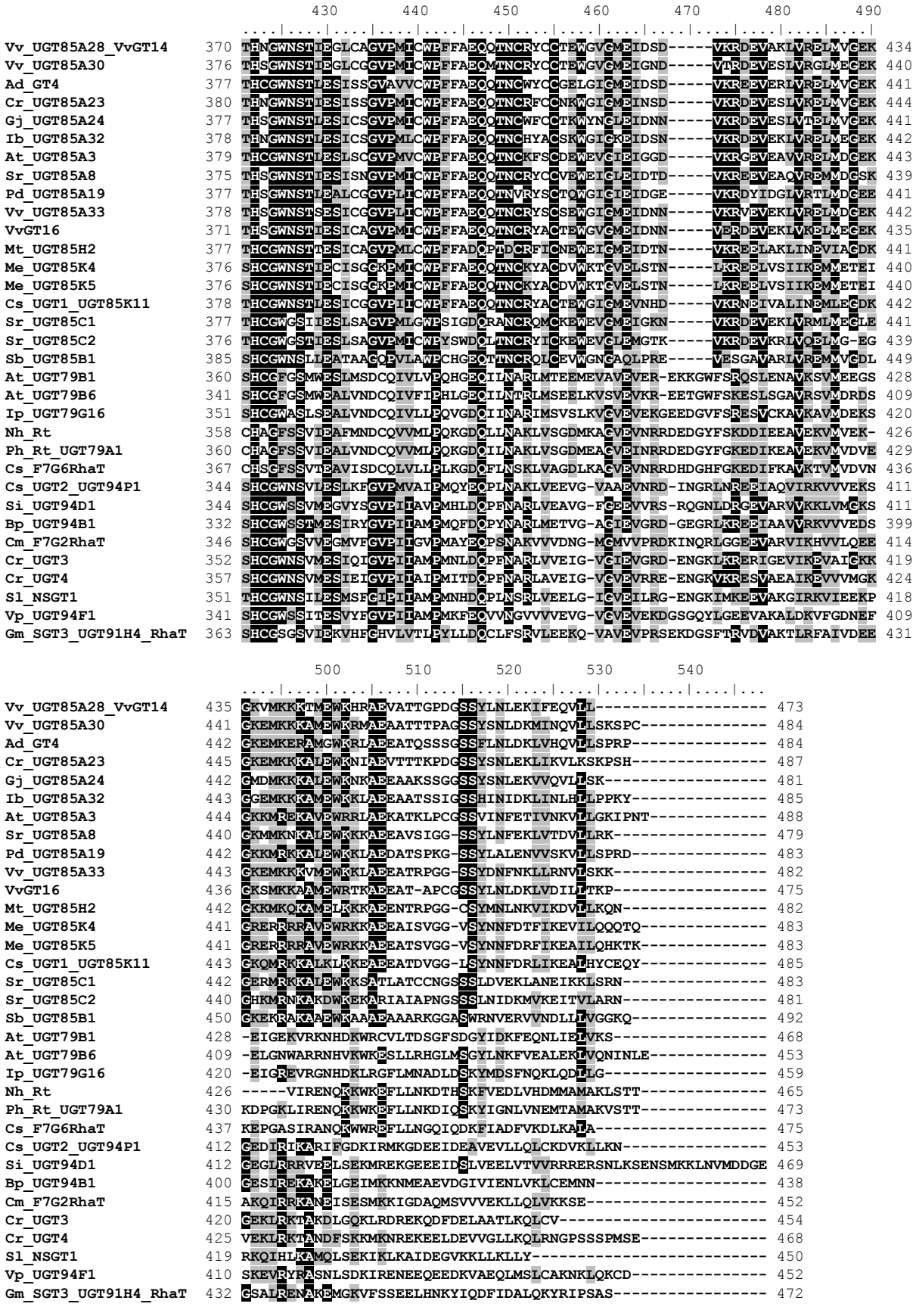
Supplemental Figure S9. Multiple sequence alignment of protein sequences of UGT-glucosyltransferase OG2 and OG8 family. The alignment was performed using ClustalW2.1.



Supplemental Figure S9. *continued.*



Supplemental Figure S9. *continued.*



Supplemental Figure S9. *continued.*