

1 **UDP-glucose:anthocyanidin 3-*O*-glucoside-2''-*O*-glucosyltransferase catalyzes**  
2 **further glycosylation of anthocyanins in purple *Ipomoea batatas***

3 **Running Title: Further anthocyanin glycosylation in sweetpotato**

4 Hongxia Wang<sup>1</sup>, Chengyuan Wang<sup>2</sup>, Weijuan Fan<sup>1,3</sup>, JunYang<sup>3</sup>, Ingo Appelhagen<sup>4</sup>,  
5 Yinliang Wu<sup>1,5</sup> and Peng Zhang<sup>1,5</sup>

6 <sup>1</sup>National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in  
7 Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai  
8 Institutes for Biological Sciences, Chinese Academy of Science, Shanghai 200032,  
9 China

10 <sup>2</sup>Center for Computational Medicine and Bioinformatics, University of Michigan,  
11 Ann Arbor, MI 48109-2218, USA

12 <sup>3</sup>Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai  
13 Chenshan Plant Science Research Center, Chinese Academy of Science, Shanghai  
14 Chenshan Botanical Garden, Shanghai 201602, China

15 <sup>4</sup>John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United  
16 Kingdom

17 <sup>5</sup>University of Chinese Academy of Sciences, Beijing 100049, China

18 Authors for correspondence:

19 Hongxia Wang, Tel: +86 2154924097, Email: hxwang@sibs.ac.cn

20 Peng Zhang, Tel: +86 2154924096, Email: zhangpeng@sibs.ac.cn

21 **Abstract**

22 Glycosylation contributes to the diversity and stability of anthocyanins in plants. The  
23 process is catalyzed by various glucosyltransferases using different anthocyanidin  
24 aglycones and glycosyl donors. An anthocyanidin  
25 3-*O*-glucoside-2''-*O*-glucosyltransferase (3GGT) from purple sweetpotato (cv.  
26 Ayamurasaki) served for the catalytic conversion of anthocyanidin 3-*O*-glucoside into

27 anthocyanidin 3-*O*-sophoroside, which is functionally different from the 3GGT  
28 ortholog of *Arabidopsis*. The phylogenetic analysis indicates regioselectivity of 3GGT  
29 using UDP-xylose or UDP-glucose as the glycosyl is divergent between  
30 *Convolvulaceae* and *Arabidopsis*. Homology-based protein modeling and site-directed  
31 mutagenesis of Ib3GGT and At3GGT suggested that the Thr-138 of Ib3GGT is a key  
32 amino acid residue for UDP-glucose recognition and plays a major role in sugar donor  
33 selectivity. The wild type and *ugt79b1* mutants of *Arabidopsis* plants overexpressing  
34 *Ib3GGT* produced the new component cyanidin 3-*O*-sophoroside. Moreover, *Ib3GGT*  
35 expression was associated with anthocyanin accumulation in different tissues during  
36 Ayamurasaki plant development and was regulated by the transcription factor  
37 IbMYB1. The localization assay of Ib3GGT showed that further glycosylation occurs  
38 in the cytosol and not endoplasmic reticulum. The present study revealed the function  
39 of Ib3GGT in further glycosylation of anthocyanins and its Thr-138 is the key amino  
40 acid residue for UDP-glucose recognition.

41 **Key words:** *Ipomoea batatas*, anthocyanins, further glycosylation, UDP-glucose,  
42 glucosyltransferase, regioselectivity.

### 43 **Introduction**

44 Anthocyanins are major secondary metabolites responsible for color variation in  
45 plants, exhibiting health-promoting properties (de Pascual-Teresa and  
46 Sanchez-Ballesta, 2008; He and Giusti, 2010). The basic structures of anthocyanins  
47 are mono- and di-glycosylated forms in common anthocyanidins, which include  
48 cyanidin, delphinidin, malvidin, pelargonidin, peonidin, and petunidin (Moglia et al.,  
49 2014). Different sugar moieties, i.e., glucose, galactose, xylose, arabinose, or fructose  
50 can be linked to hydroxyl groups at 3, 5, 7, 3', and 5' positions, with the glycosylation  
51 at the 3<sup>rd</sup> position on the C-ring ubiquitously (Andersen and Jordheim, 2010).  
52 Glycosylation of 3-OH is catalyzed by a series of UDP carbohydrate-dependent  
53 glycosyltransferases (UGTs), which utilize the nucleotide-activated sugars as donor  
54 substrates and anthocyanidin aglycones or anthocyanins as acceptors. These activities  
55 increase the structural diversity of anthocyanins by adding different types and/or  
56 numbers of sugar moieties on various positions (Gachon et al., 2005). The  
57 glycosylation of anthocyanin is speculated to occur on the cytoplasmic surface of the

58 endoplasmic reticulum (ER), and may serve as a signal for the transport of  
59 anthocyanins to vacuoles via multiple pathways; this transport is essential for the  
60 stable storage of anthocyanins in vacuoles (Ono et al., 2006; Matsuba et al., 2010; Sun  
61 et al., 2012; Zhao et al., 2011; Zhao, 2015). Glycosylation also participates in the fine  
62 adjustment and stabilization of flower pigmentation in ornamental plants  
63 (Yonekura-Sakakibara et al., 2012).

64 Monoglycosylation of anthocyanidins produces anthocyanidin 3-*O*-glucosides,  
65 the first major stable colored pigments in the anthocyanin biosynthesis pathway  
66 (Griesser et al., 2008a; Montefiori et al., 2011). Deficiency of the activity of the  
67 corresponding UDP-glucose:flavonoid 3-*O*-glycosyltransferase (UF3GT), in maize  
68 *bronze1* and *Arabidopsis anl1*, results in a significantly suppressed accumulation of  
69 anthocyanin (Fedoroff et al., 1984; Kubo et al., 2007). Until now, UF3GT is one of  
70 the well-characterized UGTs related to anthocyanin biosynthesis (Gachon et al., 2005;  
71 Yonekura-Sakakibara and Hanada, 2011). Further glycosylation of anthocyanidin  
72 3-*O*-glucosides involves diverse sugars in different species, such as UDP-rhamnose,  
73 UDP-glucose, UDP-xylose, and UDP-arabinose, as donor substrates to be added at a  
74 species-specific position to the glycosides of mono 3-*O*-glycosylated anthocyanins  
75 (Yonekura-Sakakibara et al., 2012). The mutants affected in this further glycosylation  
76 function may be impacted in the anthocyanin accumulation, as reported in petunia and  
77 Japanese morning glory (Kroon et al., 1994; Morita et al., 2005). Since all the UGT  
78 proteins are highly similar in their secondary and tertiary structures, with a defined  
79 fold structure and highly conserved putative secondary product glycosyltransferase  
80 (PSPG) motif (Breton et al., 2006; Lairson et al., 2008; Osmani et al., 2009),  
81 structure-based modeling have identified the key residues of UF3GT responsible for  
82 sugar donor specificity (Kubo et al., 2004) in *Arabidopsis* (Kim et al., 2013), *Freesia*  
83 *hybrid* (Sun et al., 2016), grapes (Offen et al., 2006; Ono et al., 2010), lamiales  
84 (Noguchi et al., 2009), perilla (Noguchi et al., 2009), and red daisy (Osmani et al.,  
85 2009). Nevertheless, the residues involved in UDP-sugar selectivity in 3GGT are yet  
86 unknown.

87 Although the anthocyanidin decoration by glycosylation is progressive, it  
88 commonly begins with 3-*O*-glycosylation to ensure the stability of the aglycon.  
89 Additional glycosylation leads to the compound and functional diversity, thereby

90 contributing to several varieties of anthocyanins in the plant (Gachon et al., 2005;  
91 Caputi et al., 2012). To date, more than 600 anthocyanins or their derivatives have  
92 been identified in nature (Glover and Martin, 2012); however, only a limited number  
93 of genes encoding UFGTs in different species have been well characterized. Several  
94 flavonoid 3-*O*-glycosyltransferases have been characterized in *Arabidopsis* (Kubo et  
95 al., 2007; Saito et al., 2013), strawberry (Griesser et al., 2008a, b), grapes (Offen et al.,  
96 2006), and maize (Fedoroff et al., 1984). In addition, for further flavonoid  
97 glycosylation multiple UGTs were also characterized, including anthocyanidin  
98 3-*O*-glucoside 6''-*O*-rhamnosyltransferase in *Petunia hybrida* (Kroon et al., 1994),  
99 anthocyanidin 3-*O*-glucoside 2''-*O*-glucuronosyltransferase in red daisy flowers  
100 (Sawada et al., 2005), and flavonol 3-*O*-glucoside 2''-*O*-glucosyltransferase in  
101 *Arabidopsis* (Yonekura-Sakakibara et al., 2014). Apparently, the divergence towards  
102 different glycosylation types occurs at this step. At the same 2'' position, different  
103 glycosylation types, i.e., glycosylation or xylosylation, are found in various plant  
104 species. In morning glory, anthocyanidin 3-*O*-glucoside 2''-*O*-glucosyltransferase  
105 catalyzes the addition of a glucose molecule to anthocyanidin 3-*O*-glucosides on the 2''  
106 position to form anthocyanidin 3-*O*-sophorosides (Morita et al., 2005). In *Arabidopsis*,  
107 further glycosylation of the 3-*O*-glucoside is catalyzed by anthocyanidin  
108 3-*O*-glucoside 2''-*O*-xylosyltransferase (AtA3G2XylT, i.e., At3GGT) to add one  
109 xylose molecule specifically to the first glucose residue (Yonekura-Sakakibara et al.,  
110 2012). Further decorations, for example, to the diversity or functionality,  
111 malonylation and aromatic acylation rely on the glycosylation of anthocyanidins  
112 (Sasaki et al., 2014).

113 Purple sweetpotato (*Ipomoea batatas*) accumulates a lot of anthocyanins in  
114 storage roots. Anthocyanidin 3-*O*-glucoside-2''-*O*-glucoside (anthocyanin  
115 3-*O*-sophoroside) and derivatives are the major anthocyanin compounds (Tian et al.,  
116 2005). So far, at least 26 components, mostly caffeoylated, coumarylated or  
117 feruloylated anthocyanidin glucosides have been identified (Truong et al., 2009; Lee  
118 et al., 2013). In contrast, 11 anthocyanins have been identified in *Arabidopsis*; all of  
119 them derived from cyanidin 3-*O*-glucoside-2''-*O*-xyloside (Tohge et al., 2005;  
120 Yonekura-Sakakibara et al., 2012; Kovinich et al., 2014). Therefore, unlike  
121 *Arabidopsis*, purple sweetpotato uses UDP-glucose as sugar donor for further

122 glycosylation of anthocyanidin 3-*O*-glucosides to form anthocyanidin  
123 3-*O*-sophorosides. In the present study, we characterized a UFGT, termed as  
124 UDP-glucose:anthocyanidin 3-*O*-glucoside-2''-*O*-glucosyltransferase (IbA3G2GluT,  
125 i.e., Ib3GGT) that catalyzes the anthocyanin glycosylation in purple sweetpotato and  
126 its key amino acid for sugar donor selectivity.

## 127 **Materials and methods**

### 128 **Plant materials**

129 The purple-fleshed sweetpotato (*Ipomoea batatas* Lam.) cultivar Ayamurasaki was  
130 used in this study. The *in vitro* shoot cultures were subcultured on SBM medium (MS  
131 salts including vitamins + 0.3 mg/L VB1 + 30 g/L sucrose, pH 5.8) in plant growth  
132 chambers under a 16 h photoperiod provided by cool-white fluorescent tubes (~50  
133  $\mu\text{mol}/\text{m}^2/\text{s}$ ), at 25°C and 50% relative humidity. One-month-old plantlets were  
134 transplanted into plastic pots containing well-mixed soil (soil:peat:perlite, 1:1:1) and  
135 grown in the greenhouse (16 h/8 h light/dark cycle, 25°C day/night). Various tissues  
136 including leaves, stems, fibrous roots, and storage roots of sweetpotato plants at  
137 different developmental stages were harvested from the pot- or field-grown plants for  
138 multiple analyses. The *Arabidopsis* plants were grown under a 16 h/8 h light/dark  
139 cycle, at 22°C in the growth chamber.

### 140 **Plasmid construction and production of transgenic sweetpotato**

141 The open reading frame of *Ib3GGT* (1380 bp) was amplified from the cDNA of  
142 sweetpotato Ayamurasaki using the primers Ib3GGTF  
143 (5'-CGGGGTACCATGGGTCTCAAGCAACAAC-3', KpnI site underlined) and  
144 Ib3GGTR (5'-AATGTCGACTCATCCAAGGAGATCCTGCA-3', SalI site  
145 underlined). This fragment was inserted into the KpnI/SalI sites of the  
146 pCAMBIA1301-based plant expression vector to generate the binary vector  
147 pOE-Ib3GGT containing the expression cassette of *Ib3GGT* driven by the CaMV 35S  
148 promoter. The pRNAi-Ib3GGT binary vector was manipulated to express  
149 double-stranded hairpin RNA of the 252 bp *Ib3GGT* fragment (382–633 bp) based on  
150 the pRNAi-DFR vector (Wang et al., 2013). Then, pOE-Ib3GGT and pRNAi-Ib3GGT  
151 were introduced into *Agrobacterium tumefaciens* strain LBA4404 for sweetpotato  
152 transformation, as described previously (Yang et al., 2011). Transgenic plants were  
153 produced and verified for *Ib3GGT* expression by real-time RT-PCR. For total *Ib3GGT*

154 expression, an internal primer pair of *Ib3GGT* was designed for detecting the *Ib3GGT*  
155 expression in WT, OE-*Ib3GGT* and RNAi-*Ib3GGT* plants by real-time RT-PCR  
156 (Table S2). The *Actin* gene of sweetpotato was used as an internal control for gene  
157 amplification.

### 158 **Transformation and analysis of *Ib3GGT*-overexpressing *Arabidopsis***

159 Two independent UGT79B1 *Arabidopsis* transposon mutants, *ugt79b1-1* and  
160 *ugt79b1-2* (Kuromori et al., 2004; Ito et al., 2005), along with the WT Nossen and  
161 ecotype Col-0 were transformed with *A. tumefaciens* LB4404 harboring pOE-*Ib3GGT*,  
162 using the floral dip method (Clough and Bent, 1998). The transformants were selected  
163 on 1/2 MS medium containing 50 mg/L hygromycin for Nossen and mutants or 25  
164 mg/L hygromycin for Col-0 plants. The RNA extracted from T3 homozygous  
165 *Arabidopsis* seedlings was used for RT-PCR analysis. The primer pairs used to detect  
166 the expression of *At3GGT* and *Ib3GGT* in WT and transgenic *Arabidopsis* plants were  
167 designed using software Primer 3.0 and listed in Table S2. *At3GGT* were amplified a  
168 223-bp fragment from position +369 to +591 bp and *Ib3GGT* were amplified a 189-bp  
169 fragment from +1009 to +1197 bp. The *Actin* gene of *Arabidopsis* was used as a  
170 reference gene.

### 171 **Phylogenetic Analysis**

172 To construct a phylogenetic tree, 16 UGT protein sequences obtained from NCBI  
173 GenBank were aligned by ClustalW and implemented in MEGA6 (Tamura et al.,  
174 2013). Ten closely related UGTs were used to illustrate the relationship. The  
175 maximum likelihood method was used to obtain the alignment results (Stamatakis,  
176 2014). Bootstrap values were obtained with 1000 replications.

### 177 **Site-directed mutagenesis and *in vitro* enzymatic assay of recombinant *Ib3GGT* 178 and *At3GGT***

179 The full-length sequence of the *Ib3GGT* gene was amplified by PCR using the  
180 primers *IbGGT-FP* (5'-CCCAAGCTTATGGGTTCTCAAGCAACAAC-3', HindIII  
181 site underlined) and *IbGGT-RP*  
182 (5'-CGCGGATCCTCACATCACCATCACCATCACTCCAAGGAGATCCTGCA-3',  
183 BamHI site and 6 His sites underlined). The full-length *At3GGT* was amplified by  
184 PCR using the primers *AtGGT-FP*

185 (5'-GGGGTACCATGGGTGTTTTTGGATCGAA-3', KpnI site underlined) and  
186 AtGGT-RP  
187 (5'-CGGAATTCCTCACATCACCATCACCATCACTGACTTCACAAGTTCAATTA  
188 AATT-3', EcoRI site and 6 His sites underlined). Site-directed mutations were  
189 generated by changing the Thr-138 nucleotide ACC into ATT in *Ib3GGT* and Ile-142  
190 ATC into ACT in *At3GGT* using PCR-based amplification with a Phusion  
191 Site-Directed Mutagenesis Kit (Thermo Scientific). The sequence fragments, with or  
192 without the mutation of the 3GGTs, were cloned into the pYES2 vector and  
193 introduced in *Saccharomyces cerevisiae* BY4742 according to the manufacturer's  
194 instructions (Cat# V825-20, Invitrogen). The recombinant 3GGT proteins were  
195 induced by replacing the carbon source from 2% glucose to 2% galactose in the SC-U  
196 medium. The reaction mixture for the 3GGT enzymatic assay consisted of 100 mM  
197 phosphate buffer (pH 7.0), 0.6 mM flavonoid aglycones (cyanidin,  
198 cyanidin3-*O*-glucoside, cyanidin 3,5-*O*-diglucoside, or flavonol 3-*O*-glucoside), 1  
199 mM UDP-glucose, and 20  $\mu$ L of crude yeast extract as the enzymatic solution in a  
200 reaction volume of 100  $\mu$ L. After incubation for 2 h at 37°C, the reaction was  
201 terminated by centrifugation. The enzymatic activity of mutant 3GGT was assessed  
202 by cyanidin 3-*O*-glucoside as the acceptor substrate and different UDP-sugars  
203 (UDP-glucose, UDP-xylose, UDP-galactose or UDP-arabinose) as a sugar donor.

#### 204 **LC-MS analyses of metabolites obtained by enzymatic reaction**

205 Ten  $\mu$ L of filtered supernatants were analyzed on an Agilent  
206 HPLC1200-MSD/Q-TOF 6520 system (Agilent, Waldbronn, Germany) as described  
207 previously (Wang et al., 2013). Briefly, the mobile phase consisted of 0.5% (v/v)  
208 acetic acid in water (eluent A) and 100% acetonitrile (eluent B). The samples eluted at  
209 a flow rate of 0.2 mL/min passed through a reverse-phase C18 column (Agilent  
210 ZORBAX Eclipse XDB, 4.6  $\times$  50 mm ID, 1.8  $\mu$ m), and a DAD detector at 530 nm  
211 monitored the anthocyanin. Subsequently, an ESI interfaced Q-TOF mass detector  
212 ( $m/z$  40–1500) collected the mass  $m/z$  data that were processed by Agilent Mass  
213 Hunter Qualitative Analysis (version 3.0) for the estimation of accurate molecular  
214 mass as well as spectrum evaluation. Cyanidin 3-*O*-sophoroside (Tongtian, Shanghai,  
215 China) was used as a standard.

#### 216 **Subcellular localization of *Ib3GGT* in plant cells**

217 The *Ib3GGT* gene was amplified by PCR using Pfu polymerase (Takara, Shanghai,  
218 China) to obtain a non-stop coding sequence using the primers FPGGT\_L  
219 (5'-AATGTCGACATGGGTTCTCAAGCAACAAC-3', SalI site underlined) and  
220 RPPGT\_L (5'-GGACTAGTCCAAGGAGATCCTGCAGTT-3', SpeI site underlined).  
221 *Ib3GGT*-eGFP was constructed by inserting the *Ib3GGT* fragment into the  
222 corresponding sites of a modified pCambia1300 to fuse with the eGFP coding  
223 sequence. The construction of the ER-marker (Nelson et al., 2007) and the expression  
224 construct for mRFP (Claudia et al., 2017) has been described elsewhere. The ER  
225 marker, ER-mCherry, contains a signal peptide of AtWAK2 at the N-terminal and a  
226 synthetic HDEL at the C-terminal (He et al., 1999; Nelson et al., 2007). All constructs  
227 were introduced into *A. tumefaciens* GV3101 (pMP90). The growth conditions for *N.*  
228 *benthamiana* and *A. tumefaciens*, as well as the agro-infiltration procedure, were  
229 described previously (Leuzinger et al., 2013). The images were acquired 36 h  
230 post-infiltration with a Leica SP8X confocal microscope equipped with a Leica HC  
231 PL APO CS2 63x/1.20 water immersion objective. The GFP fluorescence was  
232 detected by hybrid detector HyD1 in the range of 500–540 nm and excited using the  
233 488-nm line of an argon ion laser. mCherry and mRFP fluorescence were detected in  
234 the range of 580–630 nm by HyD2 after excitation at 561 nm with a diode-pumped  
235 solid-state laser. Both fluorophores were recorded line-by-line sequentially at a 3- to  
236 4-fold average in a background noise-dependent manner. The Leica Application Suite  
237 X software was used for image acquisition and intensity estimations.

### 238 **Anthocyanin measurement and detection**

239 Total anthocyanins in the WT and transgenic lines were extracted using previously  
240 described methods with slight modifications (Wang et al., 2013). The total content of  
241 anthocyanin in the WT and transgenic lines was quantified as cyanidin  
242 3-*O*-sophoroside equivalent. The anthocyanin autofluorescence in epidermal cells of  
243 sweetpotato leaves was examined using a PCM-2000/Nikon Eclipse 600  
244 laser-scanning microscope (Nikon, Japan) equipped with an argon and helium-neon  
245 laser (excitation 488 nm, emission 544 nm).

### 246 **Luciferase assay**

247 The *Ib3GGT* promoter (2000 bp) was amplified by the primers *Ib3GGT*prFP  
248 (3'-AACTGCAGTTCAGTCAGGCAATCACAGG-5', PstI site underlined) and



249 *Ib3GGT*prRP (3'-CGCGGATCCAATAATACCTAGCTAGCT-5', BamHI site  
250 underlined) and cloned into the pLL00R vector to generate the luciferase reporter  
251 vector. The *IbMYB1* gene was amplified by the primers *IbMYB1FP*  
252 (3'-GGGGTACCATGGTTATTTTCATCTGTATG, KpnI site underlined) and  
253 *IbMYB1RP* (3'-AACTGCAGTTAGCTTAAACAGTTCTGAC-5', PstI site underlined)  
254 and subcloned into pCAMBIA1300 to generate the CaMV35S-*IbMYB1* effector  
255 plasmid. *A. tumefaciens* strain GV3101 harboring the *Ib3GGT* promoter-LUC reporter  
256 and CaMV 35S-*IbMYB1* effector was infiltrated into the 5-week-old *N. benthamiana*  
257 leaves using a needleless syringe for assessing the luciferase activity. The plants were  
258 grown for 48 h (16 h/8 h light/dark cycle, 25°C day/night), followed by injecting the  
259 leaves with 0.94 mM luciferin as substrate. The leaves were collected in the dark after  
260 3 min and luciferase signals detected on a Tanon-5200 image system. The LUC  
261 reporter empty vector with 35S-*IbMYB1* or *Ib3GGT* promoter-LUC reporter with  
262 empty effector vector was also co-infiltrated as a negative control. These experiments  
263 were repeated at least three times, and similar results were obtained.

#### 264 **Molecular modeling of *Ib3GGT* and *At3GGT* active sites**

265 The 3D models of *Ib3GGT* and *At3GGT* were generated using SWISS-MODEL  
266 workspace (Biasini et al., 2014; Wetterhorn et al., 2016) and I-TASSER server  
267 (Hiromoto et al., 2006) based on the structure of *N-/O*-glucosyltransferase of *A.*  
268 *thaliana* that served as a template (UGT72B1 PDB ID: 2VCE, Brazier-Hicks et al.,  
269 2007). The substrate binding sites were predicted by superposing both models to  
270 UGT72B1 using the COOT program (Emsley et al., 2010).

#### 271 **Statistical analyses**

272 All data were represented as mean  $\pm$  SD from at least three biological replicates.  
273 One-way ANOVA analyses were performed by using SPSS Statistics 17.0 to Duncan's  
274 multiple comparison tests. A value of  $P < 0.05$  was considered as statistically  
275 significant difference.

## 276 **Results**

### 277 **Comparison of anthocyanins indicates different further glycosylation patterns in** 278 **sweetpotato and *Arabidopsis***

279 In purple sweetpotato cv. Ayamurasaki, anthocyanins include aromatically acylated

280 anthocyanidin 3-*O*-sophoroside and derivatives, whereas in *Arabidopsis* Col-0,  
281 anthocyanin components are anthocyanidin 3-*O*-glucoside-2''-*O*-xylosyl derivatives  
282 (Table S1). This phenomenon implies that, although the first glycosylation step of  
283 anthocyanins is similar for the production of anthocyanidin 3-*O*-glucosides and is  
284 catalyzed by UDP-glucose:flavonoid 3-*O*-glucosyltransferases, further modifications  
285 of anthocyanidin 3-*O*-glucosides diverge based on the utilization of different sugar  
286 donors, i.e. glucose in sweetpotato and xylose in *Arabidopsis*. In *Arabidopsis*,  
287 glycosyltransferase UGT79B1 (*At3GGT*) catalyzes the conversion of UDP-xylose  
288 and cyanidin 3-*O*-glucoside into cyanidin 3-*O*-glucoside-2''-*O*-xyloside (Tohge et al.,  
289 2005; Saito et al., 2013). In sweetpotato, a novel glucosyltransferase,  
290 UDP-glucose:anthocyanidin 3-*O*-glucoside-2-*O*-glucosyltransferase (*Ib3GGT*), was  
291 predicted to participate in further glycosylation.

### 292 **Cloning and phylogenetic characterization of *Ib3GGT***

293 The full-length *Ib3GGT* CDS sequence (GenBank accession number EF108571) was  
294 identified from a sweetpotato cDNA library by comparison with the *At3GGT*  
295 sequence. The 1380-bp *Ib3GGT* gene harbors an open-reading frame encoding 459  
296 amino acids (aa) with a calculated molecular mass of 50.87 kDa and an isoelectric  
297 point 6.537. Further sequence analysis of *Ib3GGT* showed that its amino acid  
298 sequence shared the common domain of PSPG box (334–377 aa, Fig. 1A) with other  
299 UF3GGTs in the C-terminal region (Osmani et al., 2009). In addition, although the  
300 sugar donor specificity was reported to be partially determined by the last amino acid  
301 residue of the PSPG box, i.e., glutamine (Gln) for UDP-glucose and histidine (His)  
302 for UDP-galactose (Kubo et al., 2004), the last residue of PSPG in *Ib3GGT* (at 377  
303 aa), *At3GGTF* (UDP-glucose:flavonol 3-*O*-glucoside-2''-*O*-glucosyltransferase),  
304 *At3GGT*, and *Ip3GGT* are Gln that is conserved among these glycosyltransferases.  
305 This phenomenon indicated that other amino acid residues in their sequences might  
306 contribute towards sugar donor specificity, which necessitates further elucidation.

307 Phylogenetic analysis showed that *Ib3GGT* belongs to a cluster of typical further  
308 glycosyltransferases, and is most closely related to *Ip3GGT* of *Ipomoea purpurea*  
309 (Morita et al., 2005), showing 94.3% identity (Fig. 1B). *Ib3GGT* is also homologous  
310 to *At3GGT* and *At3GGTF* with 45.7% and 45.6% identity, respectively.

311 **Ib3GGT is an enzyme that catalyzes the glycosylation of anthocyanidin**  
312 **3-*O*-glucoside into anthocyanidin 3-*O*-sophoroside**

313 To further examine the function of Ib3GGT *in vitro*, recombinant His-tag fusion  
314 Ib3GGT and At3GGT proteins expressed in the yeast expression vector pYES2  
315 (Invitrogen, USA), were used for assessing the enzymatic activity. The specificity of  
316 Ib3GGT was examined using different sugar acceptors and donors. The recombinant  
317 Ib3GGT protein only catalyzed the conversion of cyanidin 3-*O*-glucoside into  
318 cyanidin 3-*O*-sophoroside using UDP-glucose as a sugar donor (Fig. 2A). Other  
319 glucosyl acceptors such as cyanidin, cyanidin 3,5-*O*-diglucoside, and flavonol  
320 3-*O*-glucoside could not serve as substrates, and hence, no product was detected (Fig.  
321 2B, 2C, 2D), similar to the negative control (empty vector) (Fig. 2E). In addition, the  
322 Ib3GGT protein could use peonidin 3-*O*-glucoside as the glycosyl acceptor to form  
323 peonidin 3-*O*-sophoroside (Fig. 2F). These findings indicated that Ib3GGT used  
324 anthocyanidin 3-*O*-glucoside as the glycosyl acceptor.

325 Ib3GGT specificity was also confirmed using four different UDP-sugars (Table  
326 1). No detectable or predominant UGT activity was detected by UDP-sugars except  
327 UDP-glucose, indicating that Ib3GGT is highly specific to UDP-glucose. The weak  
328 utilization of UDP-xylose to produce cyanidin 3-*O*-glucoside-2''-*O*-xyloside  
329 indicated the low affinity to this substrate. In contrast, the At3GGT protein was  
330 capable of using only UDP-xylose (not for UDP-glucose) as the sugar donor to  
331 catalyze cyanidin 3-*O*-glucoside into cyanidin 3-*O*-glucoside-2''-*O*-xyloside (Table 1,  
332 Yonekura-Sakakibara et al., 2012), showing the divergence in the specificity of sugar  
333 donors by the two UF3GGTs from different species.

334 **Thr-138 of Ib3GGT contributes to sugar donor preference**

335 To further identify the key amino acid residue of Ib3GGT responsible for sugar donor  
336 recognition, docking experiments were performed based on the 3D structures of over  
337 10 different glycosyltransferases enzymes from various plants (Brazier-Hicks et al.,  
338 2007; Hiromoto et al., 2006; Hiromoto et al., 2013; Modolo et al., 2009; Offen et al.,  
339 2006; Shao et al., 2005; Wetterhorn et al., 2016). The overall structures of these  
340 glycosyltransferases share a similar folding topology: two Rossmann-like domains  
341 formed a cleft which contains two substrates binding sites and one functional  
342 conserved histidine residues located between these sites (Supplementary Fig.S1). By

343 using DALI services 8 (Holm and Laakso, 2016), more UGT homologous structures  
344 were analyzed and the root mean square deviations (RMSDs) between each other  
345 ranged from 1.1 to 2.5 Å over the core structure region (Supplementary Fig. S2).  
346 Although the sugar acceptor ligands in these structures are diverse, the sugar donors  
347 (most of them are UDP-glucose) are similar and share a group of conserved residues  
348 in their binding pockets (Supplementary Fig. S2).

349 Two different methods, SWISS-MODEL (Biasini et al., 2014) and I-TASSER  
350 Suite (Yang et al., 2015), were used for Ib3GGT structure modeling building. To  
351 qualify the modeling results, the overall structures of Ib3GGT with the two service  
352 models were compared with proteins 2VG8 and 2VCH; then the corresponding  
353 residues in the modeling structures were also checked. Both sets of results were  
354 similar, especially in the ligand binding sites (Supplementary Fig. S3). The RMSD  
355 between two Ib3GGT modeling structures is 1.72Å and two At3GGT modeling  
356 structures is 1.92Å, showing similar structure folding of the two services. Thus, the  
357 results of SWISS-MODEL service were used to compare the Ib3GGT/At3GGT  
358 modeling structures with an *Arabidopsis thaliana* O-glucosyltransferase (PDB ID:  
359 2VCE, Brazier-Hicks et al., 2007) by superposing all three structures together. The  
360 UDP-glucose in the template structure (2VCE) was also well bound in  
361 *Ib3GGT/At3GGT* modeling structures. In Ib3GGT modeling structure,  
362 Thr22/Ser276/Glu360/Gln337/Gln338/Trp334 form a binding pocket and interact  
363 with the uridine group of UDP-glucose (Fig. 3A, Supplementary Fig.S1B);  
364 Glu277/His352/Ser357 show tightly interactions with diphosphate group  
365 (Supplementary Fig.S1C). All these residues are extremely conserved in UGTs (Hsu  
366 et al., 2018; Thompson et al., 2017).

367 The two modeling structures shared a common group of residues in binding the  
368 sugar donors (Glu277/Asp376/Gln377 in Ib3GGT and Gln285/Glu385/Gln386 in  
369 At3GGT). The only difference in sugar binding pocket is the Thr-138 in Ib3GGT  
370 which is equivalent to Ile-142 in At3GGT. The distance between O-6-glucose and  
371 Thr-138 is 2.7Å (Supplementary Fig. S4), which can form a tight interaction.

372 However, the distance in At3GGT is 1.7Å (Supplementary Fig.S4B), which is too  
373 short to bind the UDP-glucose. While replacing UDP-glucose with UDP-xylose in the  
374 same position may lead to weak interaction between the xylose group and Thr-138 in  
375 Ib3GGT (Supplementary Fig.S4C) and forms a 3.9Å hydrophobic interaction in  
376 At3GGT (Supplementary Fig. S4D). These modeling observations are consistent with  
377 our enzymatic activity assays (Table 1). Therefore, we hypothesized that the residue  
378 Thr-138 in Ib3GGT and its equivalence Ile-142 in At3GGT are the key residues for  
379 sugar donor specificity in purple sweetpotato and Arabidopsis, respectively.

380 To further attest to our hypothesis, firstly the corresponding site-directed mutants,  
381 namely Ib3GGT<sup>T138I</sup> (Thr-138 changed to Ile-138) and At3GGT<sup>I142T</sup> (Ile-142 changed  
382 to Thr-142), were constructed. Their enzyme activity showed that both the intact  
383 protein and Ib3GGT<sup>T138I</sup> could catalyze UDP-xylose to cyanidin  
384 3-*O*-glucoside-2''-*O*-xyloside. However, the Ib3GGT<sup>T138I</sup> mutant failed to use  
385 UDP-glucose (Fig. 3B). On the other hand, At3GGT<sup>I142T</sup> could not only primarily  
386 catalyze UDP-xylose to produce cyanidin 3-*O*-glucoside-2''-*O*-xyloside but also use  
387 UDP-glucose to synthesize cyanidin 3-*O*-sophorosides. These findings confirmed that  
388 T138 is a key residue for sugar (glucose/xylose) recognition in Ib3GGT.

389 To check whether the Thr-138 is a key residue for other sugar recognition, 3D  
390 models generated for UDP-galactose and UDP-arabinose were compared in the same  
391 position with UDP-glucose and UDP-xylose. The only difference between  
392 UDP-galactose and UDP-glucose is the direction of O<sub>4</sub>H moiety which changes the  
393 distance between O<sub>4</sub>H moiety with the main chain N from 3.1Å to 4.93Å, thus  
394 UDP-galactose should have less binding affinity than UDP-glucose (Supplementary  
395 Fig. S5A, S5B). The UDP-arabinose also has less binding affinity than UDP-xylose as  
396 the disappearance of the interaction between O<sub>5</sub> with His-20 (Supplementary Fig.  
397 S5C, S5D). As expected, no enzymatic activities were detected for Ib3GGT or  
398 Ib3GGT<sup>T138I</sup> using UDP-galactose and UDP-arabinose as sugar donors (Table 1).

399 To verify whether other species have the same mechanism of sugar donor  
400 selectivity, two 3GGT proteins containing Thr-138 residue from *Prunus persica*  
401 (Pp3GGT, XP\_007213494) and *Lupinus angustifolius* (La3GGT, XP\_019424989 )

402 and one containing Ile-138 from *Camelina sativa* (Cs3GGT, XP\_018450414) were  
403 cloned (Supplementary Fig. S6). Both Pp3GGT and La3GGT prefer UDP-glucose  
404 rather than UDP-xylose as sugar donor (Table 1). Their weak utilization of  
405 UDP-xylose indicated the low affinity to this substrate. On the other hand, the  
406 Cs3GGT protein was capable of using only UDP-xylose as the sugar donor (Table 1).  
407 Therefore, Thr-138 residue plays a key role in specificity of UDP-glucose donors by  
408 the two kinds of UF3GGTs. These results indicated that plant UGTs may share a same  
409 mechanism in sugar donor selectivity.

#### 410 ***Ib3GGT* expression in *Arabidopsis* produces new anthocyanin molecules**

411 To further validate the activity of *Ib3GGT* in planta, the *Ib3GGT* gene driven by the  
412 CaMV 35S promoter was overexpressed in *Arabidopsis* Col-0 and the UGT79B1  
413 transposon insertion mutants, *ugt79b1-1* and *ugt79b1-2* (Kuromori et al., 2004; Ito et  
414 al., 2005). More than 10 independent transgenic plant lines were produced for each  
415 transformation event and their T3 homozygous lines (Fig. 4A). Further RT-PCR  
416 analysis confirmed the overexpression of *Ib3GGT* in these transgenic lines of Col-0,  
417 *ugt79b1-1* and *ugt79b1-2* (Fig. 4B , 4C). In the T3 homozygous *Ib3GGT*-OE  
418 transgenic lines, a new peak with an *m/z* value corresponding to cyanidin  
419 3-*O*-sophoroside was detected in comparison to WT Col-0 by HPLC-electrospray  
420 ionization (ESI)-tandem mass spectrometry (MS/MS) analysis (Fig. 4D), although the  
421 purple-color phenotype and anthocyanin content at the cotyledon-stage seedling was  
422 indistinguishable (Fig. 4A, 4E. Moreover, the seedlings of *ugt79b1-1* and *ugt79b1-2*  
423 lines, which lacked the purple coloration as compared to the WT Nossen on 4.5%  
424 sucrose containing media, showed recovered anthocyanin accumulation when  
425 overexpressing the *Ib3GGT* gene (Fig.4A, bottom panel, Supplementary Fig. S7). The  
426 comparison of anthocyanin profiles among WT, *ugt79b1-1*, and *ugt79b1-1*  
427 overexpressing *Ib3GGT* showed the production of cyanidin 3-*O*-sophoroside in  
428 transgenic lines (Fig. 4D; Supplementary Fig. S7). These results confirmed that  
429 *Ib3GGT* could specifically catalyze the conversion of cyanidin 3-*O*-glucoside to  
430 cyanidin 3-*O*-sophoroside in *Arabidopsis*, a biological process absent in this plant.

#### 431 ***Ib3GGT* expression is associated with anthocyanin accumulation and organ 432 development and is regulated by *IbMYB1***

433 Anthocyanin accumulation in Ayamurasaki plants showed an organ-dependent pattern.  
434 The immature leaves and mature storage roots contained maximum levels of  
435 anthocyanins (Fig. 5A); while mature leaves and fibrous roots had the least amounts.  
436 Among leaves, Lf1 reached a concentration of 0.6324 mg/g, approximately 7-fold that  
437 of Lf5. The *Ib3GGT* expression analyzed by real-time PCR in different organs also  
438 showed a similar pattern — high expression was found in immature leaves as well as  
439 developing and mature storage roots (Fig. 5A). *Ib3GGT* was expressed abundantly in  
440 developing storage roots (Dt, Fig. 5A) as compared to mature roots (Mt), which  
441 accumulated 30% more anthocyanins (0.4276 mg/g, Fig. 5A). Overall, *Ib3GGT*  
442 expression was associated with anthocyanin accumulation in different organs of  
443 Ayamurasaki plants.

444 The transcription factor IbMYB1 predominantly regulates the anthocyanin  
445 biosynthesis in purple sweetpotato (Mano et al., 2007), and its expression in  
446 Ayamurasaki plants was associated with anthocyanin accumulation. Therefore,  
447 *Ib3GGT* could potentially be a target gene of IbMYB1 that regulates the expression of  
448 downstream genes by binding the G-box element (CACGTG) in their promoters  
449 (Mano et al., 2007). Nonetheless, the promoter region of *Ib3GGT* showed a G-box  
450 element at position 992 as analyzed by the PlantCARE software (Supplementary Fig.  
451 S8). To confirm that *Ib3GGT* expression was stimulated by IbMYB1, a luciferase  
452 gene reporter driven by a 2000 bp promoter of *Ib3GGT* was assayed for luciferase  
453 activity in tobacco leaves after co-agroinfiltration with the effector, which harbors the  
454 *CaMV 35S::IbMYB1* expression cassette. Interestingly, a strong luciferase activity  
455 was detected. Reporter only or effector with a corresponding empty vector failed to  
456 detect the luminescent signals (Fig. 5B). These findings indicated that IbMYB1  
457 regulates the *Ib3GGT* expression in sweetpotato Ayamurasaki plants.

#### 458 **Regulation of *Ib3GGT* expression in sweetpotato Ayamurasaki alters the** 459 **anthocyanin content but not the overall component profile**

460 To further elucidate the role of *Ib3GGT* in sweetpotato, *Ib3GGT*-overexpressing  
461 (OE-*Ib3GGT*) or -RNAi (RNAi-*Ib3GGT*) transgenic Ayamurasaki plants were  
462 analyzed. Multiple independent transgenic plant lines were produced and propagated  
463 in the greenhouse. Compared to the WT, RNAi-*Ib3GGT* lines showed reduced  
464 anthocyanin levels in the leaves of pot-grown plants, whereas OE-*Ib3GGT* plants

465 showed an increased anthocyanin accumulation in the top leaves (Fig. 6A, 6B). The  
466 expression of *Ib3GGT* was down-regulated in the RNAi-*Ib3GGT* lines and  
467 up-regulated in the OE-*Ib3GGT* lines by real-time PCR analyses (Fig. 6C).  
468 Anthocyanin level in the third leaf was reduced to 28.5% of WT in the  
469 RNAi-*Ib3GGT*-2 and increased up to 112% of WT in the OE-*Ib3GGT*-2 (Fig. 6A, 6B,  
470 6D). The changes in the anthocyanin levels were correlated to *Ib3GGT* expression in  
471 these plants (Fig. 6C, 6D). Nevertheless, the overall profile of anthocyanins did not  
472 alter in these plants (Fig. 6E), indicating that *Ib3GGT* is involved in an early stage of  
473 anthocyanin modifications. Furthermore, the auto-fluorescence assayed in leaf  
474 epidermal cells showed a dramatic reduction in the fluorescent intensity in  
475 RNAi-*Ib3GGT* lines, while WT and OE-*Ib3GGT* transgenic plants displayed strong  
476 signals (Fig. 6F). Similar trends of altered anthocyanin accumulation were also  
477 observed in the field-grown corresponding plants (Supplementary Fig. S9).

#### 478 ***Ib3GGT* functions in the cytosol**

479 Anthocyanins have been suggested to be synthesized on the outer surface of the ER.  
480 To determine the location of the glycosylation of anthocyanins by *Ib3GGT*, the  
481 putative transit peptides were predicted *in silico* by Signal IP3.0, but none was found  
482 in the full *Ib3GGT* protein sequence. To test whether *Ib3GGT* is associated with ER,  
483 the N- and C-terminal fusions of *Ib3GGT* to eGFP regulated by the CaMV 35S  
484 promoter, together with an ER-marker or a soluble mRFP, were transiently expressed  
485 in *Nicotiana benthamiana* leaves (Fig. 7). Both, the N- and C-terminal *Ib3GGT* fusion  
486 proteins were found to localize in the cytosol, similar to the soluble mRFP (Fig. 7C,  
487 7D). However, some signal of the eGFP-*Ib3GGT* fusions was also observed in the  
488 nucleus. Additionally, when *Ib3GGT*-eGFP was expressed together with ER-mCherry,  
489 no co-localization was found (Fig. 7A, 7B). Thus, *Ib3GGT* is a soluble protein in the  
490 cytosol and not associated with the ER (Poustka et al., 2007).

#### 491 **Discussion**

##### 492 **Further glycosylation modifications of anthocyanins in the cytosol**

493 Further glycosylation of anthocyanins is an essential step in their biosynthesis,  
494 accumulation, and stability (Yonekura-Sakakibara et al., 2008; Zhang et al., 2014). In  
495 purple sweetpotato, the major anthocyanins include cyanidin and peonidin



496 3-sophorosides as well as their acylated derivatives (Truong et al., 2009; Lee et al.,  
497 2013), indicating that further glycosylation is required for the conversion from  
498 anthocyanidin 3-*O*-glucosides into anthocyanidin 3-*O*-sophorosides. In this study, we  
499 found that Ib3GGT was responsible for the reaction using UDP-glucose as the sugar  
500 donor. Unlike sweetpotato, further glycosylation of anthocyanins is xylosylation  
501 catalyzed by UGT79B1 in *Arabidopsis* (Saito et al., 2013). Apparently, *Arabidopsis*  
502 lacks the enzymes to form cyanidin 3-*O*-sophoroside using cyanidin 3-*O*-glucosides  
503 as the substrate for further modification, as the overexpressing Ib3GGT in  
504 *Arabidopsis* only showed a peak of cyanidin 3-*O*-sophoroside without a change in  
505 other anthocyanin components. Therefore, Ib3GGT is a key player of anthocyanin for  
506 further glycosylation modifications in sweetpotato.

507 Further glycosylation is a critical step determining the subsequent anthocyanin  
508 modifications, such as malonylation and acylation (Yonekura-Sakakibara et al., 2008;  
509 Andersen and Jordheim, 2010). The Ib3GGT as UDP-glycosyltransferase can add a  
510 sugar residue to anthocyanidin 3-*O*-glucosides but not anthocyanidin  
511 3,5-*O*-diglucosides (Fig. 2A, 2C). Thus, we conclude that 5GT catalyzes a glucose  
512 molecule into the 5<sup>th</sup> position of C-ring that can hinder the transfer of a glucose  
513 molecule into the 2<sup>nd</sup> position of anthocyanidin 3-*O*-glucosides. In certain order  
514 modification of anthocyanin, this phenomenon might be a crucial factor for providing  
515 the condition for subsequent modification in anthocyanin. Ib3GGT cannot catalyze  
516 flavonol-3-*O*-glucoside as acceptor substrate, demonstrating that Ib3GGT has  
517 substrate specificity in purple sweetpotato (Fig. 2D). However, At3GGT can catalyze  
518 flavonol-3-*O*-glucoside and anthocyanin-3-*O*-glucoside as a broad substrate in  
519 *Arabidopsis* (Saito et al., 2013).

520 As a primary sedative mechanism that maintains metabolic homeostasis in plants,  
521 glycosylation contributes to the diversity in synthesizing various secondary plant  
522 metabolites, thereby altering the biological functions of these metabolites (Jones and  
523 Vogt, 2001; Gachon et al., 2005). Apparently, divergence occurs among species by  
524 adaption of glycosyltransferase substrate specificity. In peach, the PpUGT79B is  
525 responsible for glycosylation by adding a rhamnoside molecule to anthocyanidin  
526 3-*O*-glucosides forming the anthocyanidin 3-*O*-rutinoside (Cheng et al., 2014).  
527 Ib3GGT was functional in transgenic *Arabidopsis* by producing

528 anthocyanin-3-sophoroside; while the overexpression of UGT79B1 (At3GGT) in  
529 sweetpotato did not catalyze the production of anthocyanin  
530 3-*O*-glucoside-2''-*O*-xylose (Supplementary Fig. S10). These findings indicated that  
531 anthocyanin glycosylation in sweetpotato diverges from that of *Arabidopsis* towards  
532 the specific sugar acceptor. Interestingly, the further glycosylation of anthocyanin  
533 occurs in the cytosol, not like other UGTs that are mainly ER membrane-bound  
534 enzymes (Poustka et al., 2007; Zhao, 2015).

### 535 **Key amino acids in UGTs affect both sugar donor preference and regioselectivity**

536 The phylogenetic comparisons of flavonoid GGTs suggested that the potentially  
537 conserved amino acid residues are involved in further substrate-selectivity. Four  
538 amino acid residues (Trp-334/Gln-337/Glu-360/His-352 in Ib3GGT) are generally  
539 conserved across all known flavonoid 3-*O*-glucoside-2''-*O*-glycosyltransferases. The  
540 close relationship between Ib3GGT and UGT79B1 in the phylogenetic tree also  
541 indicated that the sugar donor selectivity of flavonoid GGTs was established after  
542 species differentiation (Saito et al., 2013). In sweetpotato, Ib3GGT accepts  
543 UDP-glucose as sugar donor to conjugate to anthocyanins, such as cyanidin  
544 3-*O*-glucoside or peonidin 3-*O*-glucoside. Interestingly, *Arabidopsis* also has a  
545 UDP-glucose:flavonoid3-*O*-glucoside-2''-*O*-glucosyltransferase (At3GGTF), which  
546 preferentially uses flavonol 3-*O*-glucoside and UDP-glucose as substrates (Kubo et al.,  
547 2007). Thr-138 as the key residue for UDP-glucose recognition is also conserved in  
548 glucosyltransferases that uses UDP-glucose as sugar donor in morning glory,  
549 *Arabidopsis* (At3GGTF), *Ricinus communis*, and *Glysin max* (Supplementary Fig. S6).  
550 Corresponding to the Thr-138 residue, the Ile-142 is the residue for UDP-xylose  
551 recognition in *Arabidopsis*. Instead, the corresponding sites of Ile in *Camelina sativa*,  
552 Thr in *Tarenaya hassleriana* and *Brassica napus*, and Val in *Eucalyptus grandis* are  
553 also responsible for recognizing UDP-xylose (Supplementary Fig. S6).

### 554 **Anthocyanin glycosyltransferases are regulated by transcription factors (TFs)**

555 Anthocyanin biosynthesis is a finely regulated system involving multiple TFs  
556 associated with plant development (Pireyre and Burow, 2015; Xu et al., 2015). For  
557 example, the temporal and spatial regulation of anthocyanin production in flowers  
558 mediated by TFs, R2R3-MYB, basic Helix-Loop-Helix (bHLH), or WD40 type  
559 (reviewed in Davies et al., 2012) brings the colorful variation in the world. Hitherto,

560 the only well-characterized TF in sweetpotato is R2R3-MYB type IbMYB1, which  
561 controls anthocyanin biosynthesis specifically in tuberous roots by inducing all the  
562 structural anthocyanin genes (Mano et al., 2007). In this study, the accumulation of  
563 anthocyanin in different organs was strongly associated with the *Ib3GGT* expression,  
564 which indicates its divergent regulation during plant development. Importantly, the  
565 activation of the *Ib3GGT* promoter by IbMYB1 confirmed that *Ib3GGT* is highly  
566 regulated by the TF in storage roots. The relatively low level of *IbMYB1* transcript in  
567 the leaves might reflect its tissue specificity. In *Arabidopsis*, it is well-documented  
568 that the R2R3-MYB TF can induce glycosyltransferases such as UGT79B1 (Tohge et  
569 al., 2005; Yonekura-Sakakibara et al., 2008; Stracke et al., 2010).

570 In summary, sweetpotato Ib3GGT catalyzes the anthocyanidin 3-*O*-glucosides  
571 into anthocyanidin 3-*O*-sophorosides using UDP-glucose as a sugar donor. The  
572 Thr-138 of Ib3GGT is a key residue for sugar donor selectivity in the further  
573 glycosylation that contributes to the stability and diversity of anthocyanins. The  
574 Ib3GGT glycosylation occurs in the cytosol and is regulated by IbMYB1 TF. The  
575 present study provides further insights regarding the glycosylation enzymes involved  
576 in secondary metabolism in divergence that can assist in developing a useful approach  
577 to diversifying certain flavonoids in crops.

#### 578 **Supplementary data**

579 **Fig. S1.** UDP-glucose binding sites of Ib3GGT. (A) Chemical structure of  
580 UDP-glucose. (B,C) Close-up views of the interactions of uridine moiety (B), and  
581 diphosphate moiety. (D) Overall structure of Ib3GGT modeling structure.  
582 UDP-glucose is showed with ball-and-stick model and the protein structure is showed  
583 with cartoon; the helix/sheet/loops are showed in cyan/red/magenta, respectively.

584 **Fig. S2.** Structure alignment of UGT homologs. (A) All structures are showed with  
585 ribbon and superposed by Coot (Emsley et al., 2010). The PDB numbers of these  
586 structures are 2ACW, 2C1Z, 2PQ6, 2VG8, 3HBF, 3WC4, 5GL5, 5NLM, 5TMB and  
587 5U6M. (B) Ligands in superposed structures are showed with sticks.

588 **Fig. S3.** Sequence alignment of Ib3GGT with 2VG8 and 2VCH proteins using  
589 SWISS-MODEL and I-TASSER.

590 **Fig. S4.** Sugar donor binding sites in Ib3GGT and At3GGT. (A, B, C, D) Close-up  
591 views of the interactions of UDP-glucose with Ib3GGT (A), UDP-glucose with  
592 At3GGT (B), UDP-xylose with Ib3GGT (C), and UDP-xylose with At3GGT (D).  
593 UDP-glucose and UDP-xylose are showed with stick-and-ball in magenta and red,  
594 respectively; side chain residues are showed with stick; the hydrogen bonds are  
595 indicated by dashed lines.

596 **Fig. S5.** Difference in binding affinity of sugar analogies. (A, B, C, D)  
597 Ball-and-stick model of UDP-galactose (A), UDP-glucose (B), UDP-arabinose (C),  
598 and UDP-xylose (D). They are showed in grey, magenta, pink, and yellow,  
599 respectively. Chemical structure of galactose, glucose, arabinose, and xylose are  
600 showed in the corresponding positions.

601 **Fig. S6.** Amino acid sequence comparison of GGT analogies with Ib3GGT. The  
602 Thr-138 site is boxed.

603 **Fig. S7.** Anthocyanin pigmentation and component profiles in seedlings of the  
604 *ugt79b1-2* mutant and *Ib3GGT*-overexpressing *ugt79b1-2* transgenic line  
605 (*ugt19b1-2*+*Ib3GGT*-OE).

606 **Fig. S8.** *Ib3GGT* promoter sequence showing the two G-box sites.

607 **Fig. S9.** Leaf and root phenotypes of field-grown wild-type (WT), RNAi-*Ib3GGT*-2  
608 and OE-*Ib3GGT*-2 plant lines.

609 **Fig. S10.** Analysis of anthocyanin compounds in wild-type (WT) and  
610 *At3GGT*-overexpressing (OE-*At3GGT*) sweetpotato by HPLC-MS.

611 **Table S1:** Anthocyanin compounds in sweetpotato (Tian et al., 2005) and *Arabidopsis*  
612 (Tohge et al., 2005)

613 **Table S2:** List of primers for gene expression analysis in *Arabidopsis* and sweetpotato  
614 plant lines.

#### 615 **Acknowledgments**

616 This work was supported by the grants from the National Natural Science Foundation  
617 of China (31501357, 31771854), the Collaborative Innovation Action – Agricultural

618 Science and Technology Innovation Program of Chinese Academy of Agricultural  
619 Sciences (CAAS-XTCX2016009), and XDPB0402 of CAS. We thank Prof. Cathie  
620 Martin from JIC for providing helpful suggestions and corrections for the manuscript.  
621 We also thank our colleagues Prof. Peng Zhang for protein structure analysis, Miss  
622 Yuanhong Shan for HPLC-MS-MS analysis and Mr. XiaoshuGao for confocal  
623 microscopy.

## 624 **Reference**

625 **Andersen ØM, Jordheim M.** 2010. Anthocyanins. In: Encyclopedia of Life Sciences  
626 (ELS). Chichester: John Wiley and Sons Ltd.

627 **Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F,**  
628 **Gallo CT, Bertoni M, Bordoli L.** 2014. SWISS-MODEL: modelling protein tertiary  
629 and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**,  
630 W252-W258.

631 **Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ,**  
632 **Davies G J, Edwards R.** 2007. Characterization and engineering of the bifunctional  
633 N- and O-glucosyltransferase involved in xenobiotic metabolism in plants.  
634 *Proceedings of the National Academy of Sciences, USA* **104**, 20238-20243.

635 **Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A.** 2006. Structures and  
636 mechanisms of glycosyltransferases. *Glycobiology* **16**, 29R-37R.

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

641 **Cheng J, Wei G, Zhou H, Gu C, Vimolmangkang S, Liao L, Han Y.** 2014.  
642 Unraveling the mechanism underlying the glycosylation and methylation of  
643 anthocyanins in peach. *Plant Physiology* **166**, 1044-1058.

644 **Claudia C, Philippe R, Edith F, Vincent B, Yumei Z, Sharyn EP, Juan J R,**  
645 **Martin Y, Christophe D.** 2017. The class III peroxidase PRX17 is a direct target of  
646 the MADS-box transcription factor AGAMOUS-LIKE15 (AGL15) and participates in

- 647 lignified tissue formation. *New Phytologist* **213**, 250-263.
- 648 **Clough SJ, Bent AF.** 1998. Floral dip: a simplified method for  
649 *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal* **16**,  
650 735-743.
- 651 **Davies KM, Albert NW, Schwinn KE.** 2012. From landing lights to mimicry: the  
652 molecular regulation of flower colouration and mechanisms for pigmentation  
653 patterning. *Funct. Plant Biology* **39**, 619-638.
- 654 **de Pascual-Teresa S, Sanchez-Ballesta MT.** 2008. Anthocyanins: from plant to  
655 health. *Phytochemistry Reviews* **7**, 281-299.
- 656 **Emsley P, Lohkamp B, Scott WG, Cowtan K.** 2010. Features and development of  
657 Coot. *Acta Crystallographica Section D: Structural Biology* **66**, 486-501.
- 658 **Fedoroff NV, Furtak DB, Nelson OE.** 1984. Cloning of the bronze locus in maize by  
659 a simple and generalizable procedure using the transposable controlling element  
660 Activator (Ac). *Proceedings of the National Academy of Sciences, USA* **81**,  
661 3825-3829.
- 662 **Gachon CM, Langlois-Meurinne M, Saindrenan P.** 2005. Plant secondary  
663 metabolism glycosyltransferases: the emerging functional analysis. *Trends in Plant*  
664 *Science* **10**, 542-549.
- 665 **Glover BJ, Martin C.** 2012. Anthocyanins. *Current Biology* **22**, R147-R150.
- 666 **Griesser M, Hoffmann T, Bellido ML, Rosati C, Fink B, Kurtzer R, Aharoni A,**  
667 **Munoz-Blanco J, Schwab W.** 2008a. Redirection of flavonoid biosynthesis through  
668 the down-regulation of an anthocyanidin glucosyltransferase in ripening strawberry  
669 fruit. *Plant Physiology* **146**, 1528-1539.
- 670 **Griesser M, Vitzthum F, Fink B, Bellido ML, Raasch C, Munoz-Blanco J,**  
671 **Schwab W.** 2008b. Multi-substrate flavonol O-glucosyltransferases from strawberry  
672 (*Fragaria* × *ananassa*) achene and receptacle. *Journal of Experimental Botany* **59**,  
673 2611-2625.
- 674 **He J, Giusti MM.** 2010. Anthocyanins: natural colorants with health-promoting  
675 properties. *Annual Review of Food Science and Technology* **1**, 163-187.

- 676 **He ZH, Cheeseman I, He D, Kohorn BD.** 1999. A cluster of five cell  
677 wall-associated receptor kinase genes, Wak1-5, are expressed in specific organs of  
678 Arabidopsis. *Plant Molecular Biology* **39**, 1189-1196.
- 679 **Hiroto T, Honjo E, Noda N, Tamada T, Kazuma K, Suzuki M, Blaber M,**  
680 **Kuroki R.** 2015. Structural basis for acceptor-substrate recognition of  
681 UDP-glucose:anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea*. *Protein*  
682 *Science* **24**, 395-407.
- 683 **Hiroto T, Honjo E, Tamada T, Noda N, Kazuma K, Suzuki M, Kuroki R.**  
684 2013. Crystal structure of UDP-glucose:anthocyanidin 3-O-glucosyltransferase from  
685 *Clitoria ternatea*. *Journal of Synchrotron Radiation* **20**, 894-898.
- 686 **Holm L, Laakso LM.** 2016. Dali server update. *Nucleic Acids Research* **44**, 351-355.
- 687 **Hsu TM, Welner DH, Russ ZN, Cervantes B, Prathuri RL, Adams PD, Dueber**  
688 **JE.** 2018. Employing a biochemical protecting group for a sustainable indigo dyeing  
689 strategy. *Nature Chemical Biology* **14**, 256-261.
- 690 **Ito T, Motohashi R, Kuromori T, Noutoshi Y, Seki M, Kamiya A, Mizukado S,**  
691 **Sakurai T, Shinozaki K.** 2005. A resource of 5,814 dissociation transposon-tagged  
692 and sequence-indexed lines of Arabidopsis transposed from start loci on chromosome  
693 5. *Plant Cell Physiology* **46**, 1149-1153.
- 694 **Jones P, Vogt T.** 2001. Glycosyltransferases in secondary plant metabolism:  
695 tranquilizers and stimulant controllers. *Planta* **213**, 164-174.
- 696 **Kim HS, Kim B, Sung S, Kim M, Mok H, Chong Y, Ahn J.** 2013. Engineering  
697 flavonoid glycosyltransferases for enhanced catalytic efficiency and extended  
698 sugar-donor selectivity. *Planta* **238**, 683-693.
- 699 **Kovinich N, Kayanja G, Chanoca A, Riedl K, Otegui MS, Grotewold E.** 2014.  
700 Not all anthocyanins are born equal: distinct patterns induced by stress in Arabidopsis.  
701 *Planta* **240**, 931-940.
- 702 **Kroon J, Souer E, de Graaff A, Xue Y, Mol J, Koes R.** 1994. Cloning and structural  
703 analysis of the anthocyanin pigmentation locus *Rt* of *Petunia hybrida*: characterization  
704 of insertion sequences in two mutant alleles. *Plant Journal* **5**, 69-80.

- 705 **Kubo A, Arai Y, Nagashima S, Yoshikawa T.** 2004. Alteration of sugar donor  
706 specificities of plant glycosyltransferases by a single point mutation. Archives of  
707 Biochemistry and Biophysics **429**, 198-203.
- 708 **Kubo H, Nawa N, Lupsea SA.** 2007. Anthocyaninless1 gene of Arabidopsis thaliana  
709 encodes a UDP-glucose:flavonoid-3-O-glucosyltransferase. Journal of Plant Research  
710 **120**, 445-449.
- 711 **Kuromori T, Hirayama T, Kiyosue Y, Takabe H, Mizukado S, Sakurai T,**  
712 **Akiyama K, Kamiya A, Ito T, Shinozaki K.** 2004. A collection of 11800 single-copy  
713 Ds transposon insertion lines in Arabidopsis. Plant Journal **37**, 897-905.
- 714 **Lairson LL, Henrissat B, Davies GJ, Withers SG.** 2008. Glycosyltransferases:  
715 structures, functions, and mechanisms. Annual Review of Biochemistry **77**, 521-555.
- 716 **Leuzinger K, Dent M, Hurtado J, Stahnke J, Lai H, Zhou X, Chen Q.** 2013.  
717 Efficient agroinfiltration of plants for high-level transient expression of recombinant  
718 proteins. Journal of Visualized Experiments **77**, 50521.
- 719 **Lee MJ, Park JS, Choi DS, Jung MY.** 2013. Characterization and quantitation of  
720 anthocyanins in purple-fleshed sweet potatoes cultivated in Korea by HPLC-DAD and  
721 HPLC-ESI-QTOF-MS/MS. Journal of Agricultural and Food Chemistry **61**,  
722 3148-3158.
- 723 **Mano H, Ogasawara F, Sato K, Higo H, Minobe Y.** 2007. Isolation of a regulatory  
724 gene of anthocyanin biosynthesis in tuberous roots of purple-fleshed sweetpotato.  
725 Plant Physiology **143**, 1252-1268.
- 726 **Matsuba Y, Sasaki N, Tera M, Okamura M, Abe Y, Okamoto E, Nakamura H,**  
727 **Funabashi H, Takatsu M, Saito M, Matsuoka H, Nagasawa K, Ozeki Y.** 2010. A  
728 novel glucosylation reaction on anthocyanins catalyzed by acyl-glucose-dependent  
729 glucosyltransferase in the petals of carnation and delphinium. Plant Cell **22**,  
730 3374-3389.
- 731 **Modolo LV, Li L, Pan HY, Blount JW, Dixon RA, Wang XQ.** 2009. Crystal  
732 structures of glycosyltransferase UGT78G1 reveal the molecular basis for  
733 glycosylation and deglycosylation of (iso)flavonoids. Journal of Molecular Biology



734 **392**, 1292-1302.

735 **Moglia A, Lanteri S, Comino C, Hill L, Knevitt D, Cagliero C, Rubiolo P,**  
736 **Bornemann S, Martin C.** 2014. Dual catalytic activity of  
737 hydroxycinnamoyl-Coenzyme A quinate transferase from tomato allows it to  
738 moonlight in the synthesis of both mono-and dicaffeoylquinic acids. *Plant Physiology*.  
739 **166**, 1777-1787.

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

744 **Morita Y, Hoshino A, Kikuchi Y, Okuhara H, Ono E, Tanaka Y, Fukui Y, Saito N,**  
745 **Nitasaka E, Noguchi H.** 2005. Japanese morning glory dusky mutants displaying  
746 reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme  
747 for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2  
748 "-O-glycosyltransferase, due to 4-bp insertions in the gene. *Plant Journal* **42**, 353-363.

749 **Nelson BK, Cai X, Nebenfuhr A.** 2007. A multicolored set of in vivo organelle  
750 markers for co-localization studies in *Arabidopsis* and other plants. *Plant Journal* **51**,  
751 1126-1136.

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

755 **Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford**  
756 **CM, Bowles DJ, Davies GJ.** 2006. Structure of a flavonoid glucosyltransferase  
757 reveals the basis for plant natural product modification. *EMBO Journal* **25**,  
758 1396-1405.

759 **Ono E, Fukuchi-Mizutani M, Nakamura N, Fukui Y, Yonekura-Sakakibara K,**  
760 **Yamaguchi M, Nakayama T, Tanaka T, Kusumi T, Tanaka Y.** 2006. Yellow  
761 flowers generated by expression of the aurone biosynthetic pathway. *Proceedings of*  
762 *the National Academy of Sciences, USA* **103**, 11075-11080.

- 763 **Ono E, Homma Y, Horikawa M, Kunikane-Doi S, Imai H, Takahashi S, Kawai Y,**  
764 **Ishiguro M, Fukui Y, Nakayama T.** 2010. Functional differentiation of the  
765 glycosyltransferases that contribute to the chemical diversity of bioactive flavonol  
766 glycosides in grapevines (*Vitis vinifera*). *Plant Cell* **22**, 2856-2871.
- 767 **Osmani SA, Bak S, Møller BL.** 2009. Substrate specificity of plant UDP-dependent  
768 glycosyltransferases predicted from crystal structures and homology modeling.  
769 *Phytochemistry* **70**, 325-347.
- 770 **Pireyre M, Burow M.** 2015. Regulation of MYB and bHLH transcription factors: a  
771 glance at the protein level. *Molecular Plant* **8**, 378-388.
- 772 **Poustka F, Irani N G, Feller A, Lu Y, Pourcel L, Frame K, Grotewold E.** 2007. A  
773 trafficking pathway for anthocyanins overlaps with the endoplasmic  
774 reticulum-to-vacuole protein-sorting route in *Arabidopsis* and contributes to the  
775 formation of vacuolar inclusions. *Plant Physiology* **145**, 1323-1335.
- 776 **Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M,**  
777 **Tohge T, Fernie AR.** 2013. The flavonoid biosynthetic pathway in *Arabidopsis*:  
778 structural and genetic diversity. *Plant Physiology and Biochemistry* **72**, 21-34.
- 779 **Sasaki N, Nishizaki Y, Ozeki Y, Miyahara T.** 2014. The role of acyl-glucose in  
780 anthocyanin modifications. *Molecules* **19**, 18747-18766.
- 781 **Sawada S, Suzuki H, Ichimaida F, Yamaguchi MA, Iwashita T, Fukui Y, Hemmi**  
782 **H, Nishino T, Nakayama T.** 2005. UDP-glucuronic acid:anthocyanin  
783 glucuronosyltransferase from red daisy (*Bellis perennis*) flowers. Enzymology and  
784 phylogenetics of a novel glucuronosyltransferase involved in flower pigment  
785 biosynthesis. *Journal of Biological Chemistry* **280**, 899-906.
- 786 **Shao H, He XZ, Achnine L, Biount JW, Dixon RA, Wang XQ.** 2005. Crystal  
787 structures of a multifunctional triterpene/flavonoid glycosyltransferase from  
788 *Medicago truncatula*. *Plant Cell* **17**, 3141-3154.
- 789 **Shen Y, Chen Y, Wu J, Shaner NC, Robert E, Campbell RE.** 2017. Engineering of  
790 mCherry variants with long Stokes shift, red-shifted fluorescence, and low  
791 cytotoxicity. *PLoS One* **12(2)**, e0171257.

- 792 **Stamatakis A.** 2014. RAxML version 8: a tool for phylogenetic analysis and  
793 post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.
- 794 **Stracke R, Jahns O, Keck M, Tohge T, Niehaus K, Fernie AR, Weisshaar B.** 2010.  
795 Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol  
796 glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and  
797 MYB111-independent flavonol glycoside accumulation. *New Phytologist* **188**,  
798 985-1000.
- 799 **Sun Y, Li H, Huang JR.** 2012. *Arabidopsis* TT19 functions as a carrier to transport  
800 anthocyanin from the cytosol to tonoplasts. *Molecular Plant* **5**, 387-400.
- 801 **Sun W, Liang L, Meng X, Li Y, Gao F, Liu X, Wang L.** 2016. Biochemical and  
802 molecular characterization of a flavonoid 3-O-glycosyltransferase responsible for  
803 anthocyanins and flavonols biosynthesis in *Freesia hybrida*. *Front Plant Science* **7**,  
804 410.
- 805 **Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S.** 2013. MEGA6: molecular  
806 evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**,  
807 2725-2729.
- 808 **Thompson AMG, Iancu CV, Neet KE, Dean JV, Choe JY.** 2017. Differences in  
809 salicylic acid glucose conjugations by UGT74F1 and UGT74F2 from *Arabidopsis*  
810 *thaliana*. *Scientific Reports* **7**, 46629.
- 811 **Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E,**  
812 **Takahashi H, Goodenowe DB, Kitayama M.** 2005. Functional genomics by  
813 integrated analysis of metabolome and transcriptome of *Arabidopsis* plants  
814 over-expressing an MYB transcription factor. *Plant Journal* **42**, 218-235.
- 815 **Truong VD, Deighton N, Thompson RT, McFeeters RF, Dean LO, Pecota KV,**  
816 **Yencho GC.** 2009. Characterization of anthocyanins and anthocyanidins in  
817 purple-fleshed sweetpotatoes by HPLC-DAD/ESI-MS/MS. *Journal of Agricultural*  
818 *and Food Chemistry* **58**, 404-410.
- 819 **Tian Q, Konczak I, Schwartz SJ.** 2005. Probing anthocyanin profiles in purple

820 sweet potato cell line (*Ipomoea batatas* L. Cv. Ayamurasaki) by high-performance  
821 liquid chromatography and electrospray ionization tandem mass spectrometry. *Journal*  
822 *of Agricultural and Food Chemistry* **53**, 6503-6509.

823 **Wang H, Fan W, Li H, Yang J, Huang J, Zhang P.** 2013. Functional  
824 characterization of dihydroflavonol-4-reductase in anthocyanin biosynthesis of purple  
825 sweet potato underlies the direct evidence of anthocyanins function against abiotic  
826 stresses. *PLoS One* **8**, e78484.

827 **Wetterhorn KM, Newmister SA, Caniza RK, Busman K, McCormick SP,**  
828 **Berthiller F, Adam G, Rayment I.** 2016. Crystal structure of Os79 (Os04g0206600)  
829 from *Oryza sativa*: A UDP-glucosyltransferase involved in the detoxification of  
830 deoxynivalenol. *Biochemistry* **55**, 6175-6186.

831 **Xu W, Dubos C, Lepiniec L.** 2015. Transcriptional control of flavonoid biosynthesis  
832 by MYB–bHLH–WDR complexes. *Trends Plant Science* **20**, 176-185.

833 **Yang J, Bi HP, Fan WJ, Zhang M, Wang HX, Zhang P.** 2011. Efficient  
834 embryogenic suspension culturing and rapid transformation of a range of elite  
835 genotypes of sweet potato (*Ipomoea batatas* [L.] Lam.). *Plant Science* **181**, 701-711.

836 **Yang JY, Yan RX, Roy A, Xu D, Poissonl J, Zhang Y.** 2015. The I-TASSER Suite:  
837 protein structure and function prediction. *Nature Methods* **12**, 7-8.

838 **Yonekura-Sakakibara K, Nakayama T, Yamazaki M, Saito K.** 2008. Modification  
839 and stabilization of anthocyanins. In: *Anthocyanins*. Springer New York, 169-190.

840 **Yonekura-Sakakibara K, Fukushima AR, Nakabayashi K, Hanada F, Matsuda S,**  
841 **Sugawara E, Inoue T, Kuromori T, Ito K, Shinozaki B.** 2012. Two  
842 glycosyltransferases involved in anthocyanin modification delineated by  
843 transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant Journal*  
844 **69**, 154-167.

845 **Yonekura-Sakakibara K, Hanada K.** 2011. An evolutionary view of functional  
846 diversity in family 1 glycosyltransferases. *Plant Journal* **66**, 182-193.

847 **Yonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi**  
848 **M, Saito K.** 2014. A flavonoid 3-O-glucoside:2''-O-glucosyltransferase responsible

849 for terminal modification of pollen-specific flavonols in *Arabidopsis thaliana*. *Plant*  
850 *Journal* **79**, 769-782.

851 **Zhang Y, Butelli E, Martin C.** 2014. Engineering anthocyanin biosynthesis in plants.  
852 *Current Opinion in Plant Biology* **19**, 81-90.

853 **Zhao J, Huhman D, Shadle G, He XZ, Sumner LW, Tang Y, Dixon RA.** 2011.  
854 MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside  
855 malonates in *Medicago truncatula*. *Plant Cell* **23**, 1536-1555.

856 **Zhao J.** 2015. Flavonoid transport mechanisms: how to go, and with whom. *Trends*  
857 *Plant Science* **20**, 576-585.

858

859

860

861

862

863

864

865

866

867

868

869

870

871 **Tables**

872 **Table 1:** Sugar substrate specificity of different 3GGT proteins

Sugar donor	Relative activity (%)					
	Ib3GGT	At3GGT	Ib3GGT <sup>T1</sup> <sub>38I</sub>	Pp3GGT	La3GG <sub>T</sub>	Cs3GGT
UDP-glucose	100 ± 8.5	ND	ND	100 ± 10.5	100 ± 9.3	ND
UDP-xylose	10.4 ± 1.2	100 ± 15.0	6.1 ± 0.79	9.4 ± 1.4	11.4 ± 1.2	100 ± 11.7
UDP-galactose	ND	ND	ND	--	--	--
UDP-arabinose	ND	ND	ND	--	--	--

873 Note: The reactions were performed with cyanidin 3-*O*-glucoside as the sugar  
874 acceptor. ND, not detected; --, not tested. The method for calculation of sugar donor  
875 specificity is according to Yonekura-Sakakibara et al. (2012). Abbreviations for  
876 species: At, *Arabidopsis thaliana*; Ib, *Ipomoea batatas*; Pp, *Prunus persica*; La,  
877 *Lupinus angustifolius*; Cs, *Camelina sativa*.

878 **Figure legends**

879 **Figure 1** Alignment of amino acid sequences and phylogenetic tree of flavonoid  
880 glycosyltransferases. **(A)** Multiple alignments of amino acid sequences of sweetpotato  
881 Ib3GGT, morning glory Ip3GGT, and *Arabidopsis* At3GGT. The underlined  
882 nucleotides represent the putative C-terminal UDP binding motif for  
883 glycosyltransferases (PSPG). **(B)** Non-rooted molecular phylogenetic tree of  
884 flavonoid glycosyltransferases from selected plant UDP-glycosyltransferases. All  
885 amino acids were aligned using CLUSTALW. Bootstrap values from 100 retrials are  
886 indicated at each branch. The scale shows 0.2 amino acid substitution per site. The  
887 GenBank accession numbers or genome sequence codes for the sequences are shown  
888 in parentheses: AtA3G2"XylT (NP\_200217); AtF3G2"GlcT (NP\_200212);  
889 BpA3G2"GlcAT (AB190262); CmF7G2"RhaT (AY048882); CrF3G6"GlcT  
890 (BAH80312); CsF3G2"XylT(XP\_018450414); CsaF3G2"GlcT(CCG85331);  
891 CsiF7G6"RhaT (NP\_001275829); Ib3GGT (ABL74480); IbF3G2"XylT  
892 (XP\_019151635); IpA3G2"GlcT (AB192315); InF3G2"GlcT (XP\_019194233);  
893 ItF3G2"GlcT (itf02g12970.t1); ItF3G2"GlcT (itb02g08330.t1); ItF3G2"XylT  
894 (itb03g28310.t1); ItF3G2"XylT (itf03g22690.t2); LaF3G2"GlcT(XP\_019424989);  
895 PhA3G6"RhaT (CAA81057); PpF3G2"Glc (XP\_007213494). Abbreviations for  
896 species: Ac, *Actinidia chinensis*; At, *Arabidopsis thaliana*; Bp, *Bellis perennis*; Cm,  
897 *Citrus maxima*; Cr, *Catharanthus roseus*; Cs, *Camelina sativa*; Csa, *Crocus sativus*;  
898 Csi, *Citrus sinensis*; Ib, *Ipomoea batatas*; In, *Ipomoea nil*; Ip, *Ipomoea purpurea*; Itf,  
899 *Ipomoea trifida*; Itl, *Ipomoea triloba*; La, *Lupinus angustifolius*; Ph, *Petunia hybrida*;  
900 Pp, *Prunus persica*.

901 **Figure 2** Functional assays of Ib3GGT recombinant protein using UDP-glucose and  
902 different acceptor substrates by HPLC. **(A)** Cyanidin 3-*O*-glucoside as acceptor  
903 substrate; **(B)** Cyanidin as acceptor substrate. **(C)** Cyanidin 3,5-*O*-diglucoside as  
904 acceptor substrate. **(D)** Quercetin 3-*O*-glucoside as acceptor substrate. **(E)** Cyanidin  
905 3-*O*-glucoside as acceptor substrate without Ib3GGT protein treatment. **(F)** Peonidin  
906 3-*O*-glucoside as acceptor substrate.

907 **Figure 3** Three-dimensional modeling of Ib3GGT and At3GGT interacting with a  
908 sugar donor and a glycone acceptor. **(A)** Active center of Ib3GGT showing the key  
909 amino acid residues for sugar donor and acceptor positions. **(B)** Docking illustration  
910 of sugar donors and a glycone acceptor in the binding pocket of WT and mutant  
911 Ib3GGT and At3GGT. The performance of their reactions using cyanidin

912 3-*O*-glucoside with sugar nucleotide UDP-glucose or UDP-xylose is shown in the  
913 bottom panel. The percentage of relative enzyme activity is indicated in the  
914 parentheses.

915 **Figure 4** Anthocyanin characterization of transgenic *Arabidopsis* plants  
916 overexpressing *Ib3GGT* gene. (A) Anthocyanin pigmentation in seedlings of WTs  
917 (Col0 and Nossen), *Ib3GGT*-overexpressing Col0 lines (Ib3GGT-OE), *ugt79b1*  
918 mutant (*ugt79b1-1*), and *Ib3GGT*-overexpressing *ugt79b1* line  
919 (*ugt79b1-1*+Ib3GGT-OE). (B) The expression of *At3GGT* and *Ib3GGT* in the WT  
920 Col0 and *Ib3GGT*-overexpressing Col0 lines (Ib3GGT-OE1 and Ib3GGT-OE2) by  
921 real-time RT-PCR analysis. (C) RT-PCR detection of *At3GGT* and *Ib3GGT*  
922 expression in the WTs (Col0 and Nossen), two independent *Ib3GGT*-overexpressing  
923 Col0 lines (Ib3GGT-OE1 and Ib3GGT-OE2), *ugt79b1* mutant, and two  
924 *Ib3GGT*-overexpressing *ugt79b1* lines (*ugt79b1-1*+Ib3GGT-OE1 and  
925 *ugt79b1-1*+Ib3GGT-OE2). Different letters indicate significant differences (one-way  
926 ANOVA,  $P < 0.05$ ). (D) Anthocyanin component profiles by HPLC/PDA/MS in the  
927 seedlings of WTs (Col0 and Nossen), *ugt79b1* mutant, and *Ib3GGT*-overexpressing  
928 *ugt79b1* line (*ugt79b1-1*+Ib3GGT-OE). Blue arrows indicate the new peaks of  
929 cyanidin 3-*O*-sophoroside. (E) Anthocyanin content in the WTs (Col0 and Nossen),  
930 two independent *Ib3GGT*-overexpressing Col0 lines (Ib3GGT-OE1 and  
931 Ib3GGT-OE2), *ugt79b1* mutant, and two *Ib3GGT*-overexpressing *ugt79b1* line  
932 (*ugt79b1-1*+Ib3GGT-OE1 and *ugt79b1-1*+Ib3GGT-OE2).

933 **Figure 5** Correlation of anthocyanin accumulation and gene expression in various  
934 organs of sweetpotato cv. Ayamurasaki. (A) Profiles of anthocyanin accumulation,  
935 and *Ib3GGT* and *IbMYB1* transcript levels as detected by qRT-PCR in different organs.  
936 Lf1, Lf2, Lf3, and Lf4 represent leaves of different developmental stages: St, stem;  
937 Ft1, white fibrous root; Ft2, red fibrous root; Dt, developing root; Mt, mature root.  
938 Values are mean  $\pm$  SD (n = 6). (B) Luciferase assay of *Ib3GGT* promoter activity  
939 (reporter) regulated by IbMBY1 (effector) in agroinfiltrated tobacco leaves.

940 **Figure 6** Anthocyanin characterization of wildtype and *Ib3GGT* transgenic  
941 sweetpotato plants. (A) Pot-grown plant phenotypes. WT, wild type; OE-Ib3GGT line,  
942 transgenic plants overexpressing *Ib3GGT*; RNAi-Ib3GGT line, *Ib3GGT* RNAi



943 transgenic plants. **(B)** Anthocyanin pigmentation in top leaves of WT,  
944 RNAi-Ib3GGT-2, and OE-Ib3GGT-2 plants. Both the adaxial (left) and abaxial (right)  
945 leaf surfaces are shown. **(C)** Relative transcription levels of native *Ib3GGT* and the  
946 *Ib3GGT* transgene in WT and transgenic lines assessed by qRT-PCR. **(D)**  
947 Anthocyanin content in WT, RNAi-Ib3GGT, and OE-Ib3GGT plant lines. Different  
948 letters indicate significant differences (one-way ANOVA,  $P < 0.05$ ). **(E)** Component  
949 profiles of anthocyanins in WT, RNAi-Ib3GGT-2, and OE-Ib3GGT-2 plants, as  
950 assessed by HPLC. **(F)** Anthocyanin autofluorescence in leaf epidermal cells of  
951 RNAi-Ib3GGT-2 and OE-Ib3GGT-2 plants.

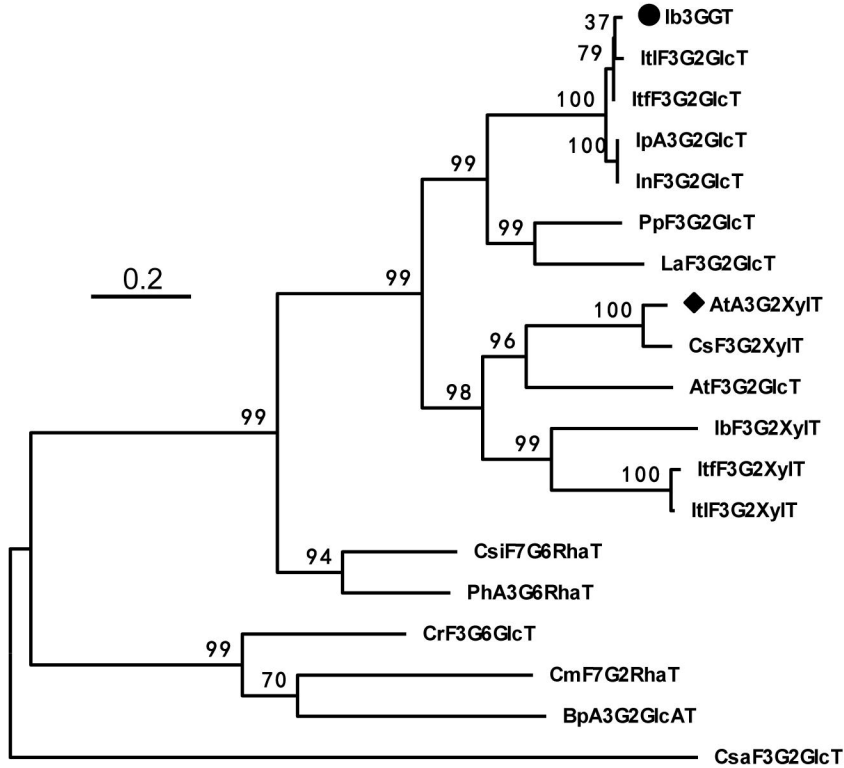
952 **Figure 7** Subcellular localization of Ib3GGT after transient expression in *Nicotiana*  
953 *benthamiana* leaves. **(A)** Optical section through a pavement cell, co-expressing  
954 Ib3GGT-eGFP (left panel) and ER-mCherry (middle panel). GFP and mCherry signals  
955 are distinct in the merged image (right panel). The outlined region is shown at a  
956 higher magnification in **(B)**. **(B)** Relative fluorescence intensity along the axis is  
957 marked by the dotted arrow in the right panel. **(C)** Co-expression of eGFP-Ib3GGT  
958 (left panel) and soluble mRFP (middle panel). White pixels of the merged image show  
959 an overlay of both channels (right panel). The marked region was enlarged in **(D)**. **(D)**  
960 Intensity plot, as presented before. eGFP-Ib3GGT and soluble mRFP were  
961 co-localized in the cytoplasm. Scale bar, 15  $\mu$ m.

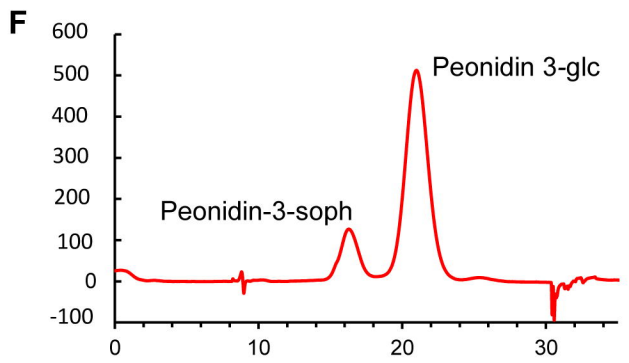
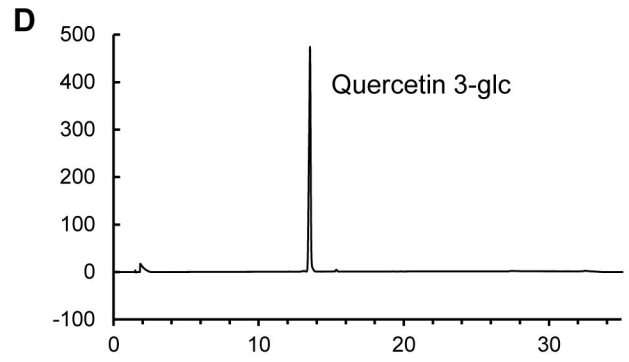
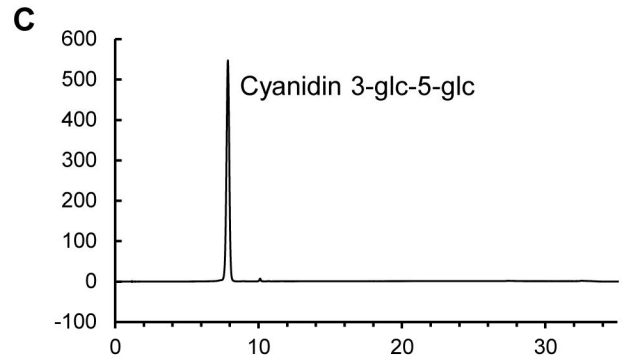
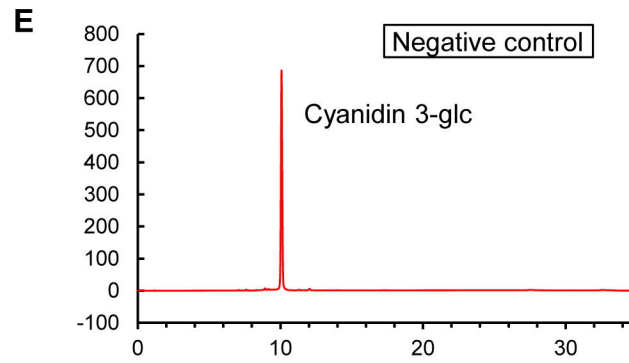
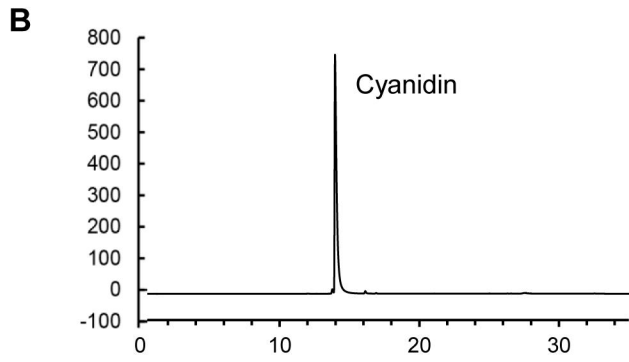
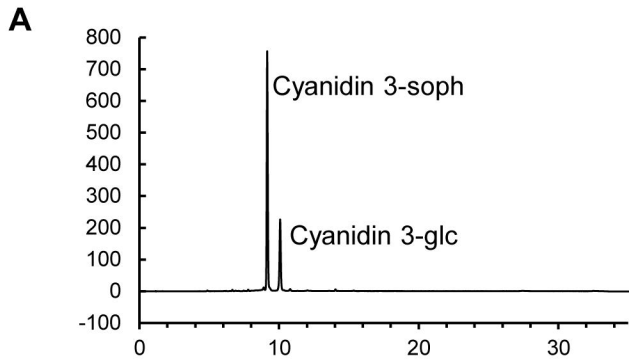
962

**A**

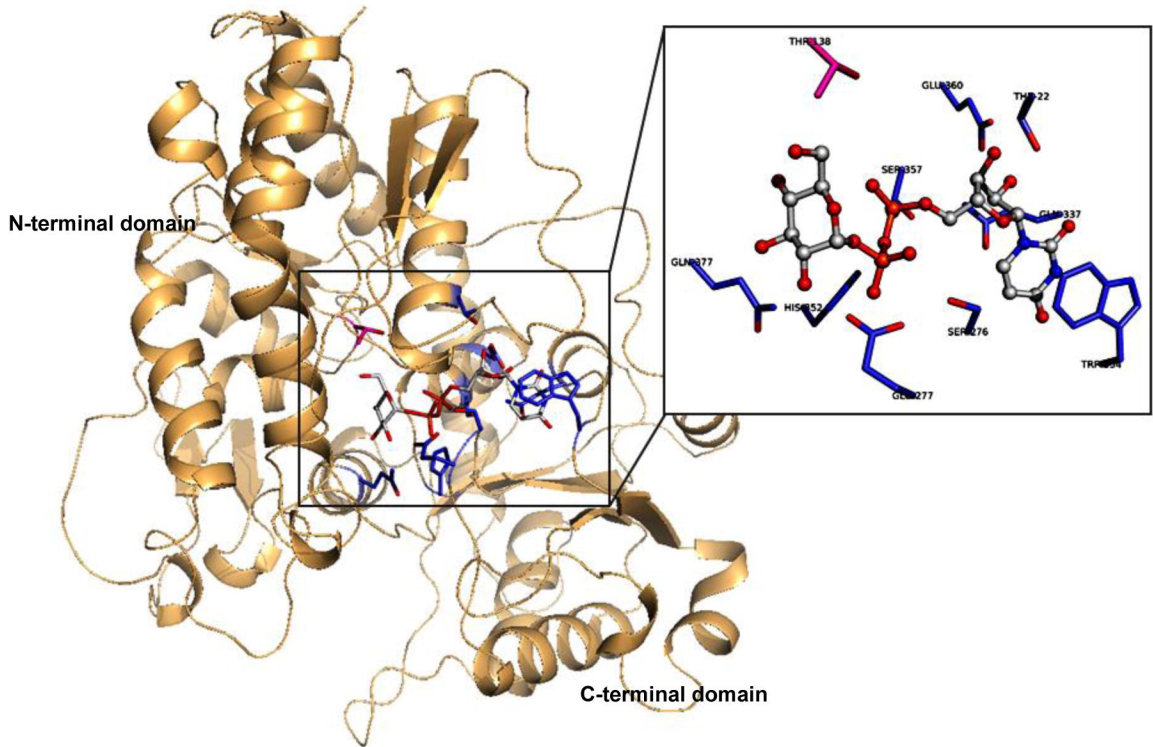
SWEET_POTATO_3GGT	MG . . . . SQATTHHWAMYPWFGVGHLLTAFFRLANKLASKGHRISFLI PKNTQSKLASFNLFPHLVSFVPI TVPSI PGLPPG	76
MORNING_GLOXY_3GGT	MG . . . . SQATTYHWAMYPWFGVGHLLTGFFRLANKLAGKGRISFLI PKNTQSKLESFNLFPHLSFVPI VVPSI PGLPPG	76
ARABIDOPSIS_3GGTF	MGVFGSNESSSMSI VWPVLAFFGHMTFFLLHLSNKLAEKGRHIVFLLPKALNQLLEPLNLYFNLI TFHTI SI PQVKGLPPG	80
ARABIDOPSIS_3GGT	MG . . . . . SKFHAFMFPWFGRGHMTAFLHLANKLAEKDKIKTFLLPKARKKLESLNLFDFDCI VFQTLTI PAVDGLPCG	73
SWEET_POTATO_3GGT	AETTSDVFPFSSTHLLMEAMDKTQTDI EI I LKNLEVDVVFDFETHWLPGLARKI GI KSVFYSTI SPLMHGFALSPERR . .	153
MORNING_GLOXY_3GGT	AETTSDVFPFSTHLLMEAMDKTQNDI EI I LKDLKVDVVFDFETHWLPGLARKI GI KSVFYSTI SPLMHGYALSPERR . .	153
ARABIDOPSIS_3GGTF	AETTNSDVFPFLLTHLLAVAMDKTRPEVETI FRTI KPDLVFPYSAHWI PEI AKPI GAKTVCFNI VSAASI ALSLVPSAEREV	160
ARABIDOPSIS_3GGT	AETTSDI PI SLGSLFASAMDRTIRI QVKEAVSVGKPDLI FDFFAHWI PEI AREYGVKSVNFI TI SAACVAI SFVP . . . . .	147
SWEET_POTATO_3GGT	VAGKQLTEADNWKAPASFPDPSI KLHAHEARGF TARTVMKFGGDI TFFDRI FTAVSESDGLAYSTCREI EGCFCDYI ETQ	233
MORNING_GLOXY_3GGT	VVGKQLTEADNWKAPASFPDPSI KLHAHEARGF TARTVMKFGGDI TFFDRI FTAVSESDGLAYSTCREI EGCFCDYI ETQ	233
ARABIDOPSIS_3GGTF	LDGKEMSGEELAKTPLGYPSKVVLRPHEAKLSF. VVRKHEAI GSFDFGKVTAMRNDAI AI RTCRETEGKFCDYI SRQ	239
ARABIDOPSIS_3GGT	..GR..SQDDLGSTPGYPSKVVLRGHEINLSLF. LSYPEGDGTSFYERI MI GLKNCQVI SI RTCCMEGKFCDI ENQ	222
SWEET_POTATO_3GGT	FKKPVLLAGPALPVP. . SKSTMEQKVSDDLCKFKEGSVI YCAFGSECTLRK. EQFQELLVGLLEL TGMPPFAALKAPFGTD	310
MORNING_GLOXY_3GGT	FQKPVLLAGPALPVP. . SKSTMEQKVSDDLCKFKEGSVI YCAFGSECTLRK. DKFQELLVGLLEL TGMPPFAALKAPPFETE	310
ARABIDOPSIS_3GGTF	YSKPVYLLTGPVLPQSGPNQPSLDPCWAEWLAKFNHGVSVPFCAFGSPVVNKI DQFQELCLGLESTGFPFLVAI KPPSGVS	319
ARABIDOPSIS_3GGT	FQRKVVLLTGPMLPEPDNSKP. LEDQVRQWLSKFDPSVI YCALGSGI I LEK. DQFQELCLGMLTGLPFLVAVKPKGSS	300
SWEET_POTATO_3GGT	SI EAAI PEELREKI HGKGI VHGQWQCQLFLGHPSVGCVFVSHCGMASLSEALVNDCCI VLLPQVGDQI INARI MSVSLKV	390
MORNING_GLOXY_3GGT	SVEAAI PEELREKI QGRGI VHGQWQCQLFLGHPSVGCVFVSHCGMASLSEALVNDCCI VLLPQVGDQI INARI MSVSLKV	390
ARABIDOPSIS_3GGTF	TVEEALPEGFKERVQGRVVFQWVGGVLPVNLHPSVGCVFVSHCGFSGNWESLMSDCCI VLVVPHGEQI LNARLMTTEEMEV	399
ARABIDOPSIS_3GGT	TI QEALPKGFEBRVKARVWVGWQCPLI LAHPSI GCFVSHCGFSGNWESLMSDCCI VFI PHLGEQI LNTRLMSEELKV	380
SWEET_POTATO_3GGT	GVEVEKGEEDCVFSRESVCKAVKAVVDEKSEI GREVRGNHDKLRGFLLNADLDSKYVDSFNQKIQDILG. . . .	459
MORNING_GLOXY_3GGT	GVEVEKGEEDCVFSRESVCKAVKAVVDEKSEI GREVRGNHDKLRGFLLNADLDSKYVDSFNQKIQDILG. . . .	459
ARABIDOPSIS_3GGTF	AVEVER. EKKWFSRGSLENAVKSVMEEGSEI GEKVRKNHDKWRCVLTDSGSDGI DKFEQNI ELVK. . . .	467
ARABIDOPSIS_3GGT	SVEVKR. EETCVFSKESLSGAVRSVNRDSELCNWARRNHVWKESLLRHGLMSGVNLNKAVEALEKLVQNI NL	452

**B**

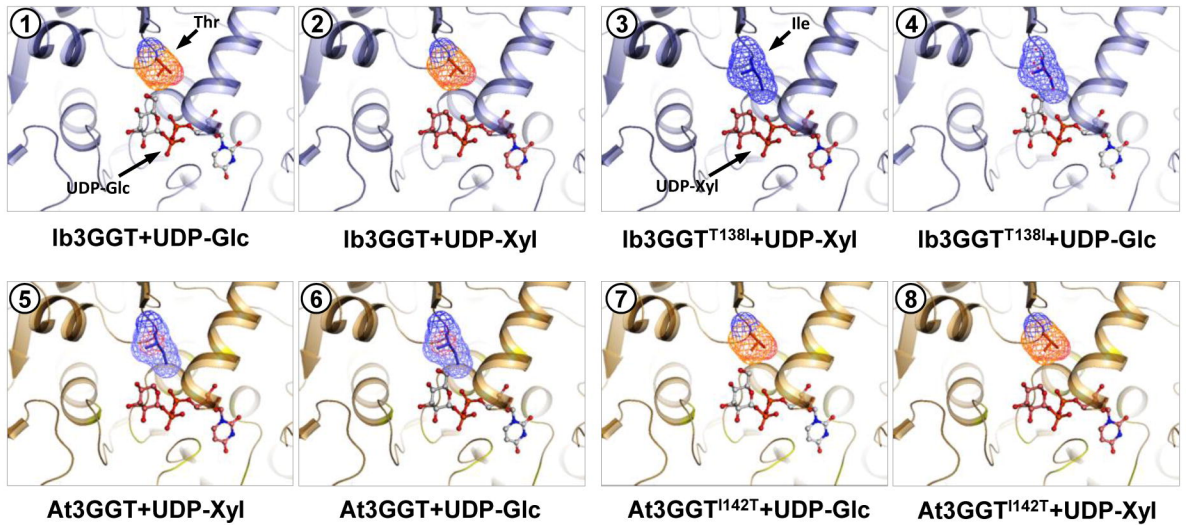




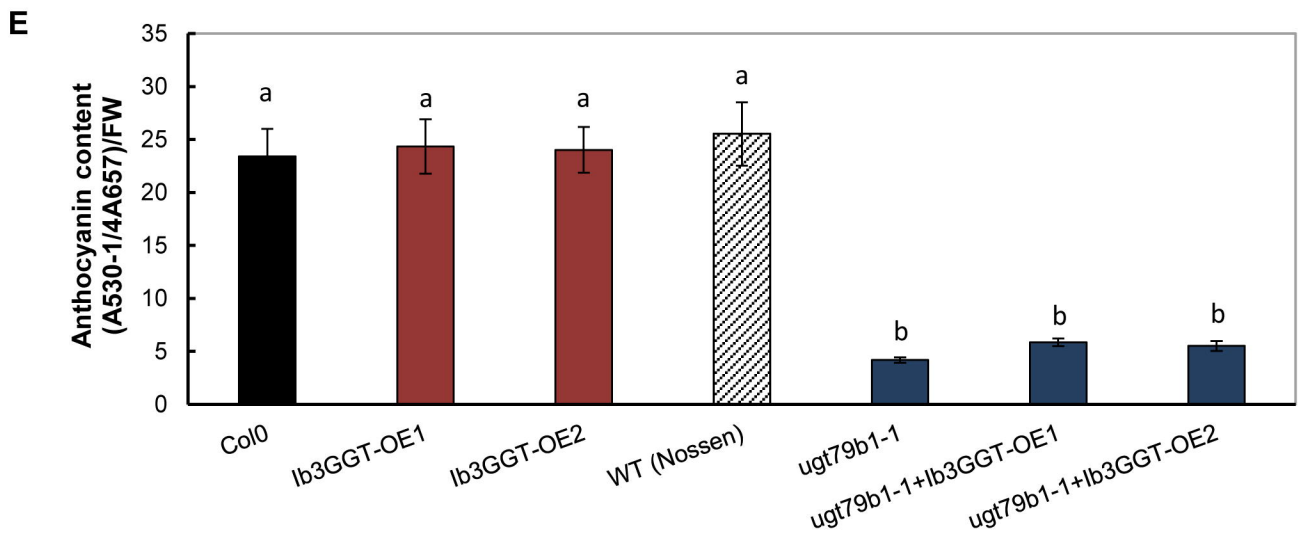
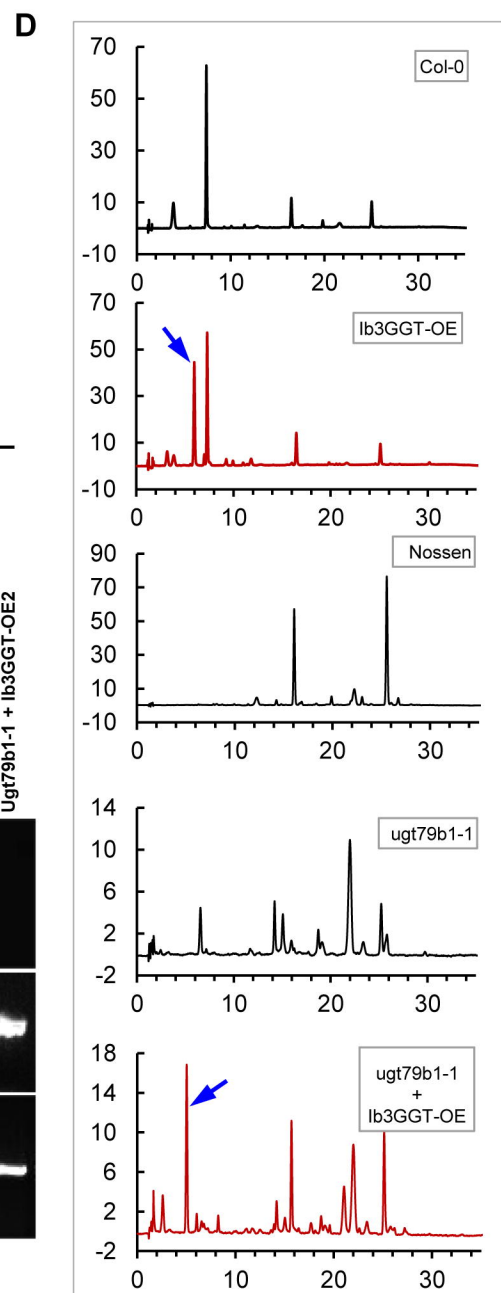
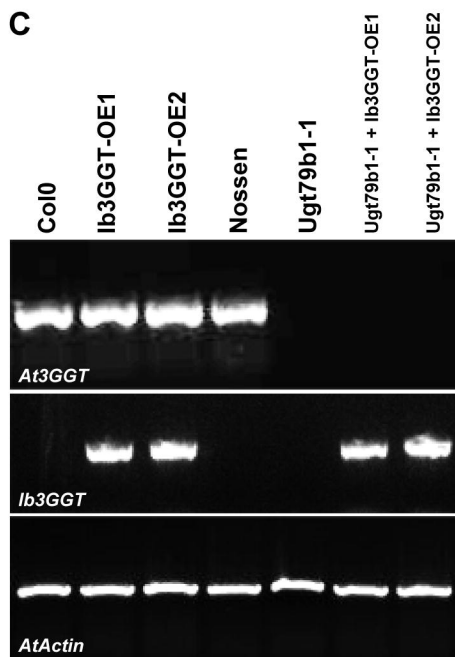
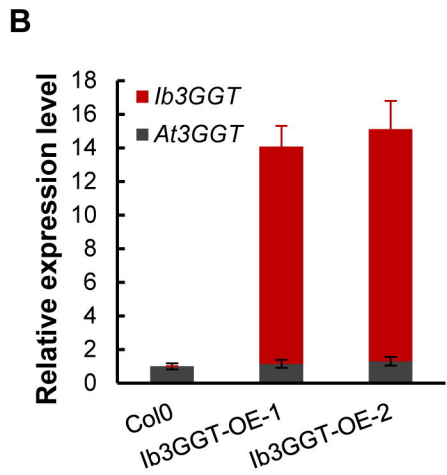
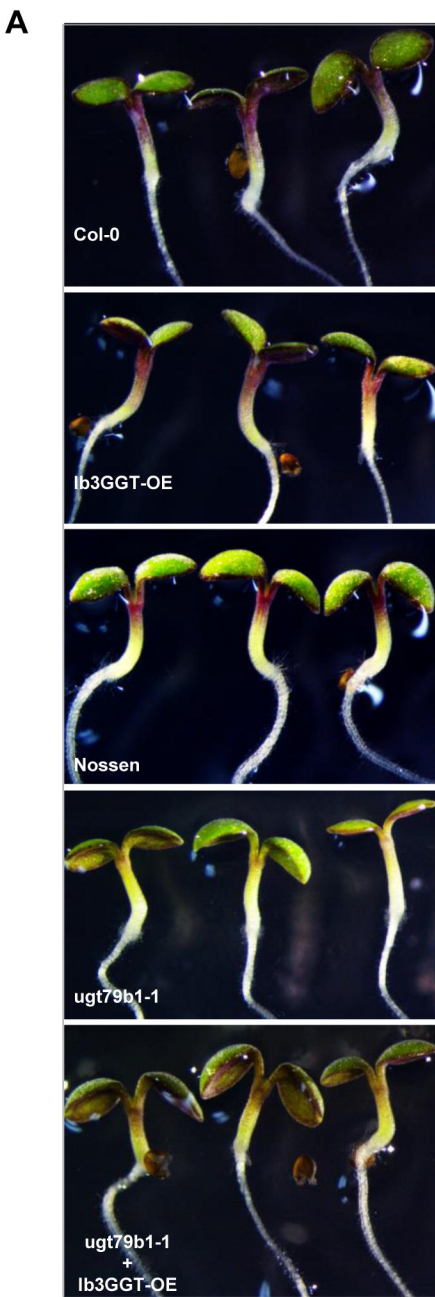
**A**



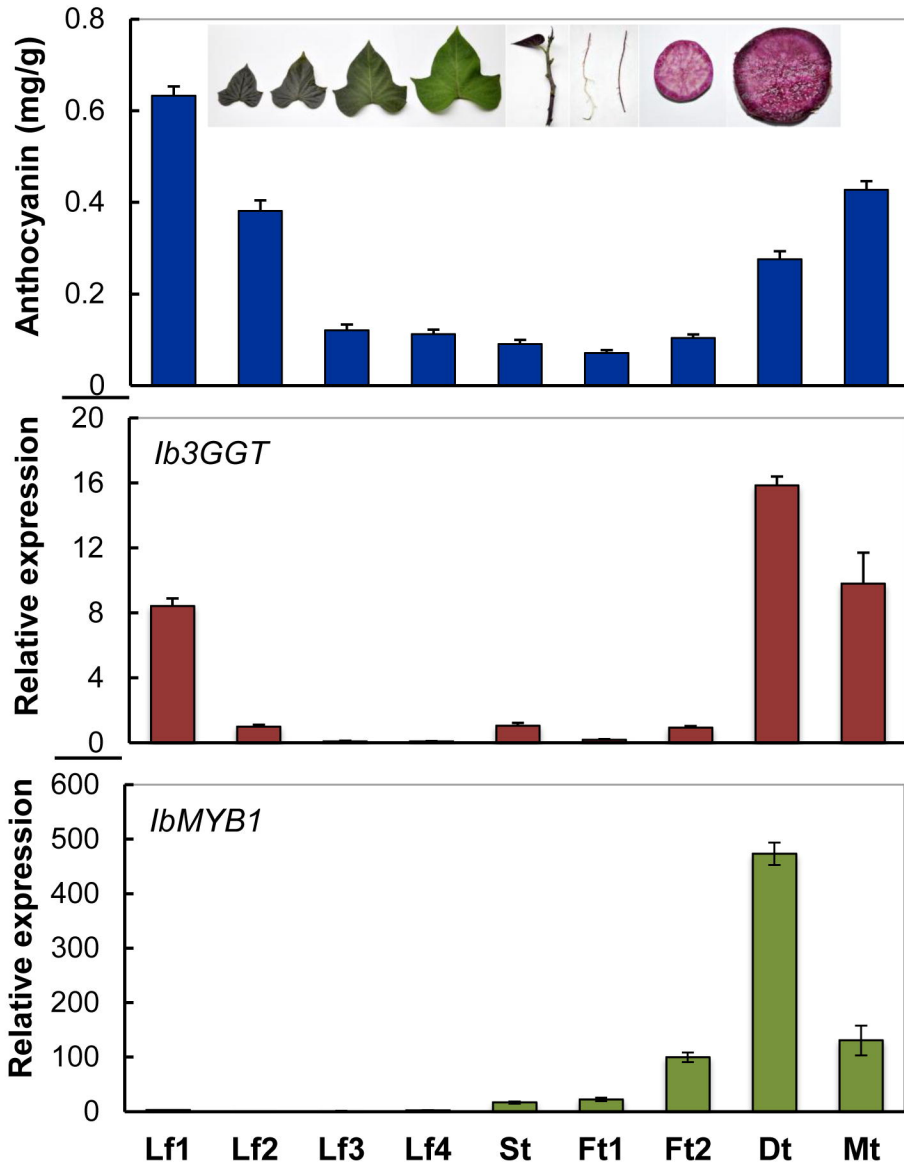
**B**



Reaction	Substrate	Enzyme	HPLC product
1	Cyanidin 3-glc + UDP-glc	Ib3GGT	Cyanidin 3-glc-2''-glc (100 ± 8.5)
2	Cyanidin 3-glc + UDP-xyl	Ib3GGT	Cyanidin 3-glc-2''-xyl (10.4 ± 1.2)
3	Cyanidin 3-glc + UDP-xyl	Ib3GGT <sup>T138I</sup>	Cyanidin 3-glc-2''-xyl (6.1 ± 0.79)
4	Cyanidin 3-glc + UDP-glc	Ib3GGT <sup>T138I</sup>	<b>No product</b>
5	Cyanidin 3-glc + UDP-xyl	At3GGT	Cyanidin 3-glc-2''-xyl (100 ± 15.0)
6	Cyanidin 3-glc + UDP-glc	At3GGT	<b>No product</b>
7	Cyanidin 3-glc + UDP-glc	At3GGT <sup>T142T</sup>	Cyanidin 3-glc-2''-glc (91.0 ± 12.5)
8	Cyanidin 3-glc + UDP-xyl	At3GGT <sup>T142T</sup>	Cyanidin 3-glc-2''-xyl (100 ± 18.3)



A

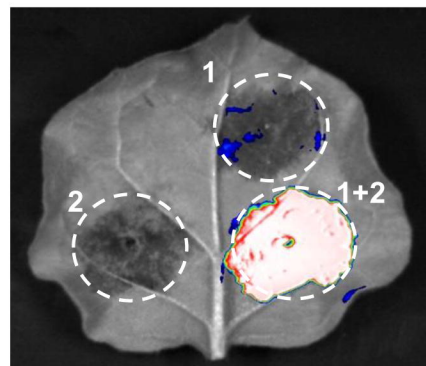


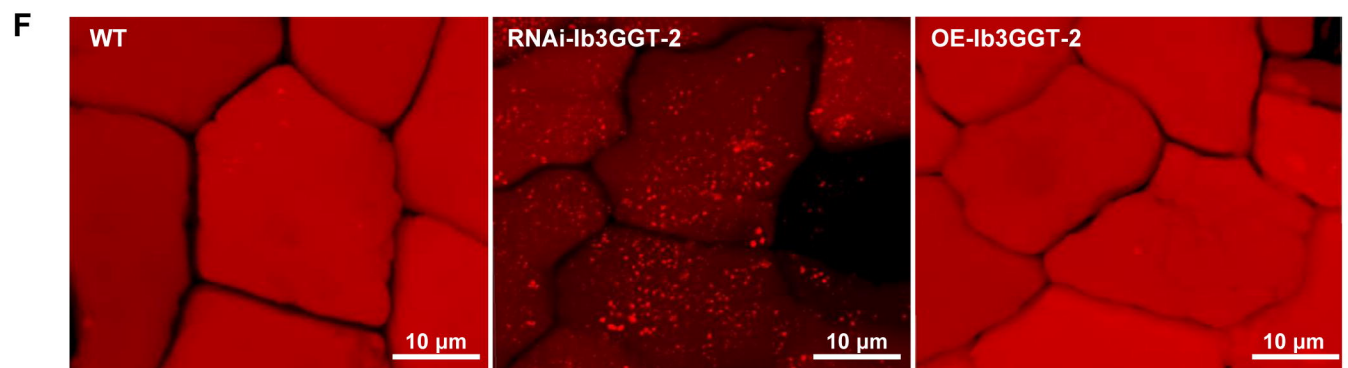
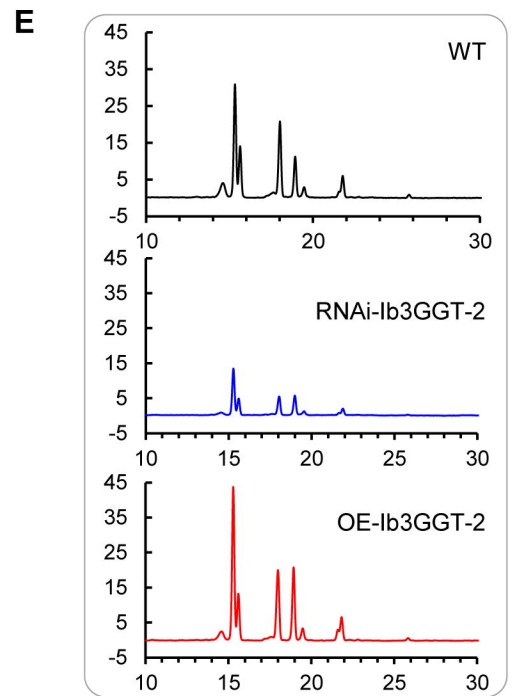
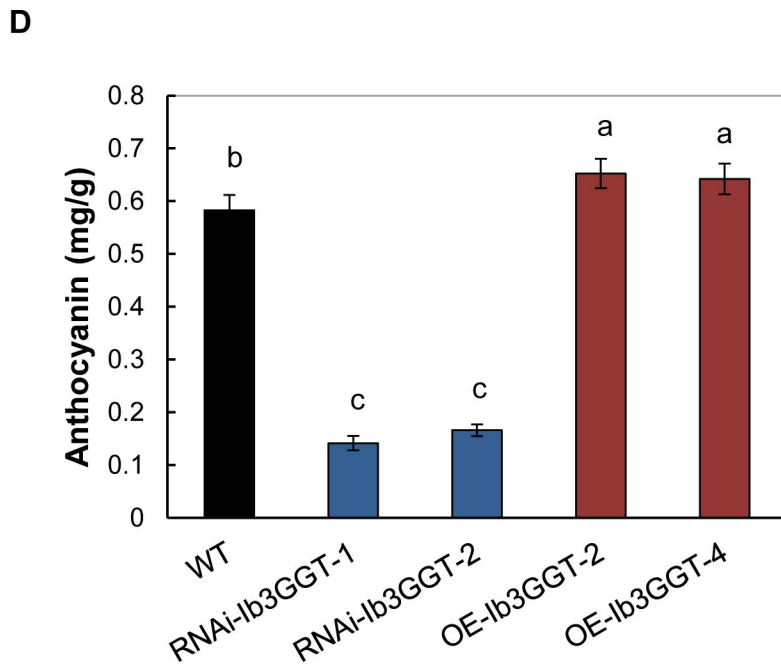
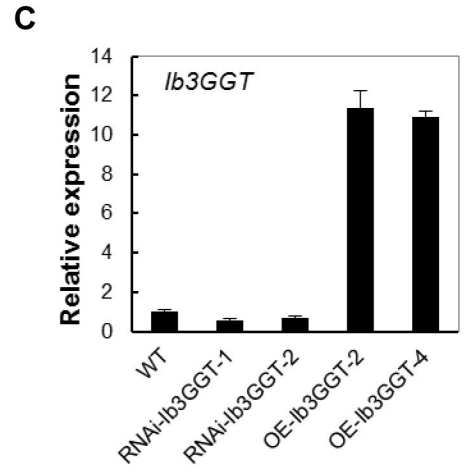
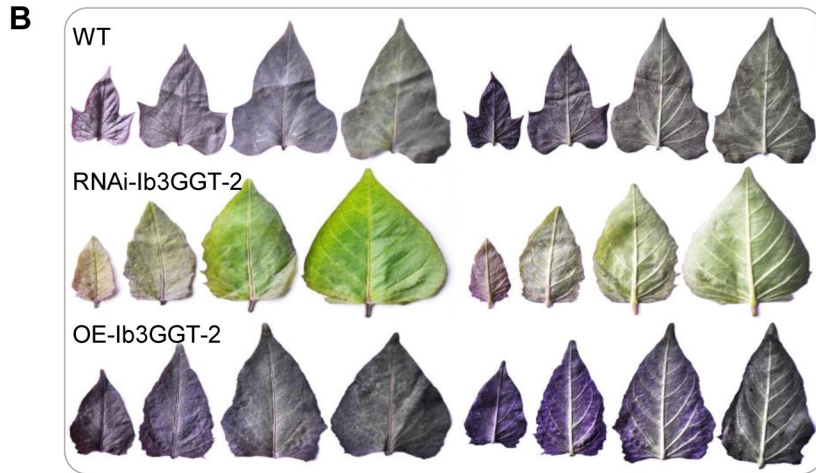
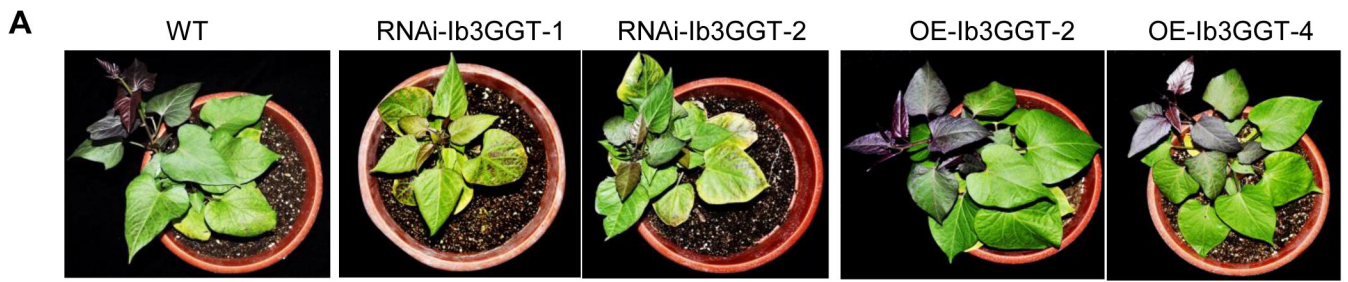
B

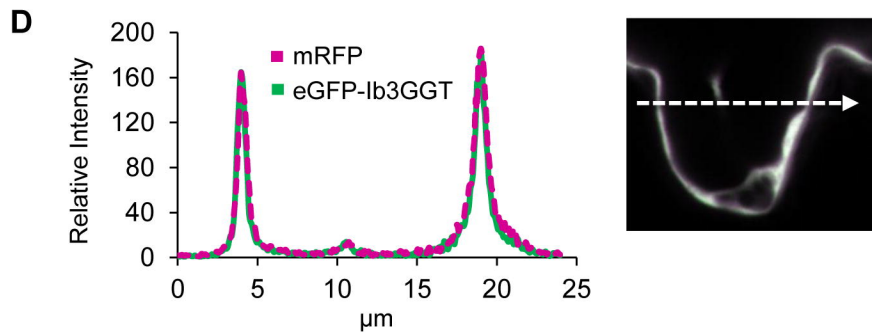
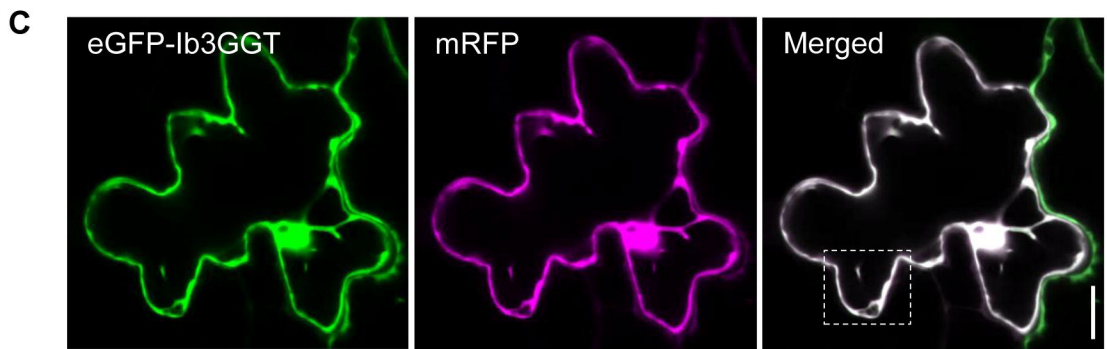
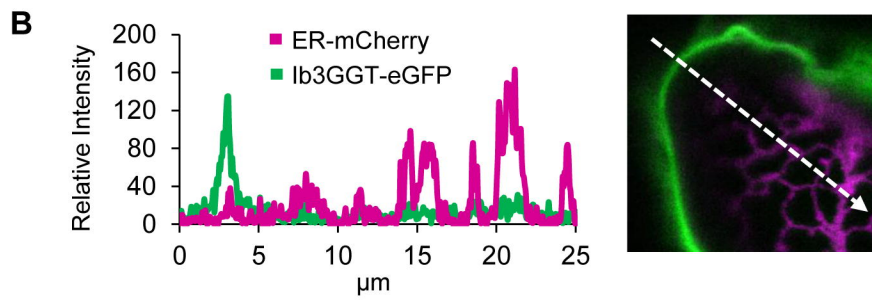
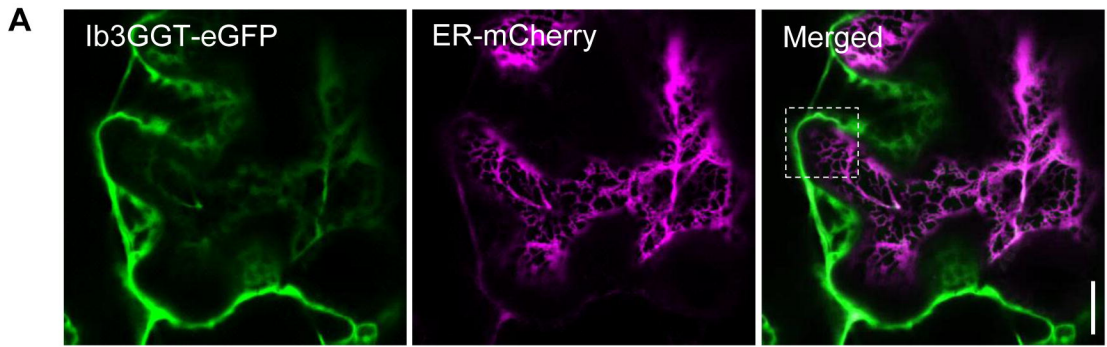
Reporter (1)



Effector (2)







Scale is 15  $\mu\text{m}$