Plant Biotechnology 28, 311–315 (2011)
DOI: 10.5511/plantbiotechnology.11.0111b

Gene Note

Identification of genes in *Arabidopsis thaliana* with homology to a novel acyl-glucose dependent glucosyltransferase of carnations

Taira Miyahara, Yuki Matsuba, Yoshihiro Ozeki*, Nobuhiro Sasaki

Department of Biotechnology, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588, Japan
* E-mail: ozeky@me.com Tel & Fax: +81-42-388-7239

Received November 5, 2010; accepted January 11, 2011 (Edited by K. Yazaki)

Abstract Recently, we characterized an acyl-glucose : anthocyanin 5-*O*-glucosyltransferase (DcAA5GT) in *Dianthus caryophyllus* that has novel characteristics. We also showed that a homologous enzyme, acyl-glucose:anthocyanin 7-*O*-glucosyltransferase (DgAA7GT), is present in *Delphinium grandiflorum.* These enzymes are both classified as members of glycoside hydrolase family 1 (GH1). Here, we searched the *Arabidopsis thaliana* database for genes with homology to *DcAA5GT* and *DgAA7GT* and identified 11 *β-glucosidases (BGLU1–11)* of unknown function. In a co-expression analysis, using genes for enzymes involved in anthocyanin modification as baits, we found that expression of *AtBGLU6* and *AtBGLU9* appeared to be correlated to that of the baits. Real-time RT-PCR showed that *AtBGLU1, 6, 7* and *9* were expressed at their highest levels in anthocyanin-inducing medium, although *AtBGLU6* was a pseudogene in which there was a stop codon on the middle of its cDNA sequence. These results suggest that AtBGLU1, 7 and 9 might be candidate glucosyltransferases for anthocyanin/flavonoid modification or biosynthesis.

Key words: Glucosyltransferase, glycoside hydrolase family1.

Glycoside hydrolases (GHs) are categorized as enzymes that catalyze the hydrolysis of the glycosidic bond between two or more carbohydrates or a carbohydrate and a non-carbohydrate moiety. They are classified into various families based on their sequence similarities. Protein folding similarities and conserved amino acid residues may be responsible for the homologous functions of enzymes within each family. At present, the GH classification in the Carbohydrate-Active enZYmes Database (http://www.cazy.org/) lists 118 families, and β-glucosidases are found in families 1, 3, 5, 9, and 30. Those in the GH family 1 (GH1) are enzymes that catalyze the hydrolysis of the β-glycosidic bond at the non-reducing end of their substrates. The GH1 enzymes belong to the GH-A clan and have an $(\alpha/\beta)_8$ barrel super secondary structure. They also have both the catalytic acid/base and the nucleophile consensus sequences. Both consensus sequences contain glutamate residues (Opassiri et al. 2006). In a recent β-glycosidase mutagenesis study, new enzymes were developed by replacing the nucleophile glutamate residues with different non-electric charge residues of other β-glycosidases (Hancock et al. 2005). These new enzymes were found to show a partial decrease to complete loss of hydrolysis activity and an increase in reverse hydrolysis activity, thereby generating catalysts for the synthesis of new glycosidic linkages. These mutant glycosidases have been termed 'glycosynthases' (Mayer et al. 2000). Glycosynthases are expected to have considerable potential as tools in the production of pharmaceuticals by enzymatic synthesis. Synthesis of oligosaccharides by glycosynthases is more advantageous than chemical synthesis because of the stereoselectivity of the sugar acceptors. However, the sugar donors for glycosynthases need to be carefully selected to maintain high glycosyltransferase activity (Shaikh and Withers 2008).

In a previous study, we were able to detect the activity of acyl-glucose:anthocyanin 5-*O*-glucosyltransferase (DcAA5GT) in *Dianthus caryophyllus* and of a homologue in *Delphinium grandiflorum* (DgAA7GT), and we also isolated the relevant cDNA in each species (Matsuba et al. 2010). Although DcAA5GT and DgAA7GT have glucosyltransferase activity, they are nevertheless classified as GH1 β-glucosidases. In addition, DcAA5GT and DgAA7GT do not use UDP-glucose as a donor. GH1 family covers a wide range of β-glycoside hydrolases with different sugar specificity and thus includes β-glycosidases of different EC numbers, such as β-glucosides (EC 3. 2. 1. 21), β-galactosides (EC 3. 2. 1. 23), β-mannosidases (EC 3. 2. 1. 25), β-glucuronidases (EC 3. 2. 1. 31), β-fucosidases (EC 3. 2. 1. 38), disaccharidases (EC 3. 2. 1. 149). This

Abbreviations: AA5(7)GT, Acyl-glucose : anthocyanin 5(7)-*O*-glucosyltransferase; BGLU, β-glucosidase; GH, glycoside hydrolase

This article can be found at http://www.jspcmb.jp/
Published online April 1, 2011

in turn implies that GH1-related glycosyltransferases might also include enzymes with different sugar donor specificities. In general, β-glycosidases catalyze the hydrolysis of β-glycosidic bonds between a monosaccharide and a carbohydrate. Recently, our previous data showed that the enzymes belonging to β-glycosidases catalyze the glycosyl transfer reaction (Matsuba et al. 2010). GH1 enzymes are retaining β-glycosidases, where the hydrolysis and glycosyl transfer reactions of β-glycosides proceed with a retention of anomeric configuration. The activities of DcAA5GT and DgAA7GT therefore represent the glucosyl transfer activities of GH1 β-glucosidases. To search for proteins homologous to DcAA5GT and DgAA7GT in *Arabidopsis*, we screened the GenBank/DDBJ database using BLASTN. Candidate genes were also identified using cDNA and amino acid sequences from The Arabidopsis Information Research (TAIR) database. ClustalW was used for protein sequence alignments and phylogenetic analysis was carried out using GENETYX-MAC Ver.14 (Figure 1). BLASTN identified *Arabidopsis thaliana β-glucosidase (AtBGLU) 1–47* as candidate homologous genes. The amino acid sequences encoded by *AtBGLU1* (At1g45191), *AtBGLU2* (At5g16580), *AtBGLU3* (At4g22100), *AtBGLU4* (At1g60090), *AtBGLU5* (At1g60260), *AtBGLU6* (At1g60270), *AtBGLU7* (At3g62740), *AtBGLU8* (At3g62750), *AtBGLU9* (At4g27820), *AtBGLU10* (At4g27830) and *AtBGLU11* (At1g02850) formed a cluster with those of *DcAA5GT* and *DgAA7GT* in the phylogenetic tree (Figure 1). This suggested that *AtBGLU1–11* were strong candidates for being glucosyltransferases similar to DcAA5GT or DgAA7GT. However, an earlier study categorized *AtBGLU1*, *2, 5, 6* as pseudogenes because of their lack of key catalytic motifs (Xu et al. 2004). Both of *AA5*/*7GT* and their homologous gene candidates *AtBGLU1–11*, and other β-glucosidases have key catalytic motifs except for pseudogenes, but their activity is distinguished by glucosyltransferase or glucosylhydrolase. It is considered that the glutamate residues of key catalytic motifs are not important for distinguishing whether their activities are glucosyltransferase or glucosylhydrolase (Matsuba et al. 2010). This result implies that the survey of the candidate genes which have glucosyltransferase activity

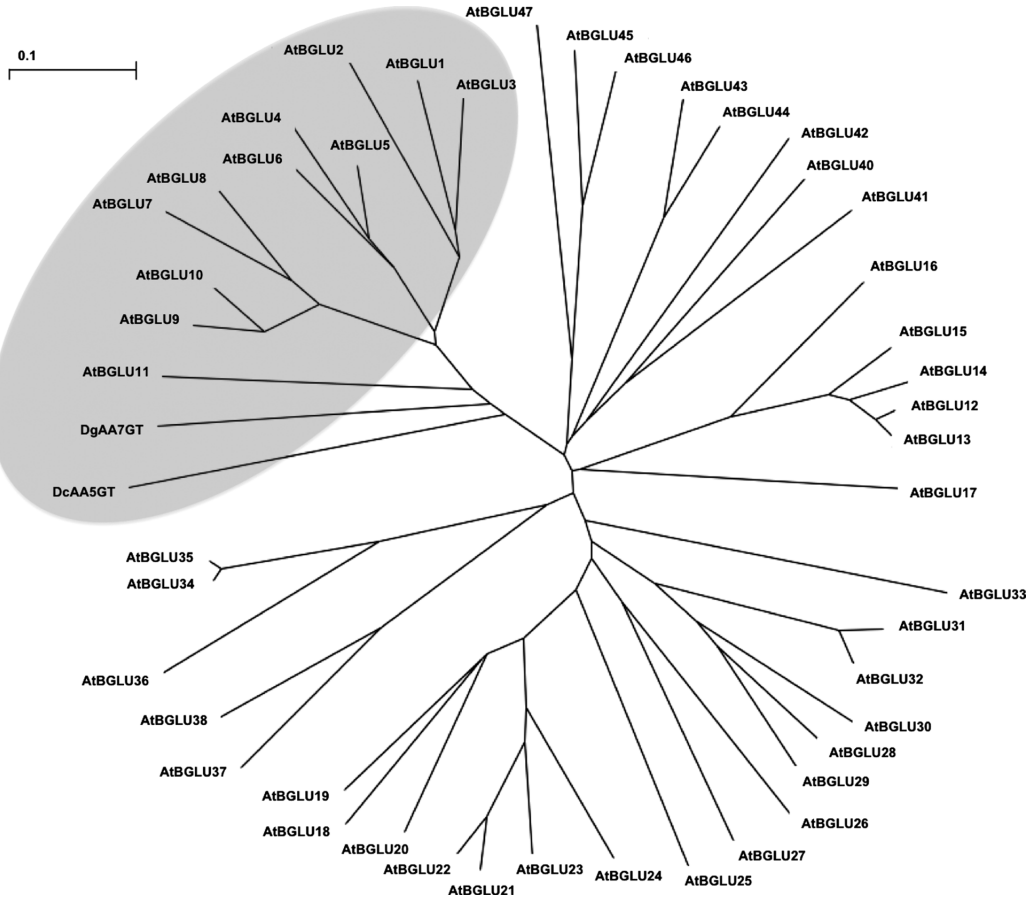

Figure 1. Phylogenetic relationships of the amino acid sequences of genes from *Arabidopsis thaliana* showing homology to DcAA5GT and DgAA7GT. The sequences were aligned using the ClustalW program at the DNA Data Bank of Japan (DDBJ), and the phylogenetic tree was generated with the internal neighbor-joining algorithm using default parameters, and viewed using GENETYX-Tree of GENETYX-MAC Ver.14. Bar=0.1 amino acid substitutions/site.

is insufficient only for homology research of their primary sequences.

Therefore, we investigated not only the primary sequences of *AA5*/*7GT* homologous genes in *Arabidopsis* but also performed a co-expression analysis of the *Arabidopsis thaliana trans*-factor and *cis*-element prediction database (ATTED-II) using a co-expression gene search algorithm as described by Sugano et al. (2010). We selected as baits, genes that encode enzymes involved in anthocyanin modification (Yonekura-Sakakibara 2009), namely At5g17050 (*UGT78D2, 3GT*) (Kubo et al. 2007), At4g14090 (*UGT75C1, 5GT*) (Thoge et al. 2005), and At3g29590 (*AT5MAT*) (D'Auria et al. 2007). The analysis indicated that *AtBGLU6* and *AtBGLU9* ranked third and 224th, respectively, in the top 300 co-expression genes retrieved in the search. *AtBGLU1–5, 7, 8, 10* and *11* did not rank in the top 300 (Table 1). However, *AtBGLU6* has been reported to be a pseudogene because it has a stop codon in the AtBGLU6 open reading frame (Xu et al. 2004). *DcAA5GT* has a

Table 1. Co-expression analysis of At5g17050, At4g14090 and At3g29590

RANK	AGI	FUNCTION
1	AT5G17050	UGT78D1 (UDP-GLUCOSYL TRANSFERASE 78D2)
2	AT5G17220	ATGSTF12 (ARABIDOPSIS THALIANA GLUTATHIONE S-TRANSFERASE PHI 12)
3	AT1G60270	BGLU6 (BETA GLUCOSIDASE 6)
4	AT3G21560	UGT84A2
5	265091_s_at	
6	253879_s_at	
7	AT4G04750	CARBOHYDRATE TRANSMEMBRANE TRANSPORTER
8	AT5G54060	UF3GT (UDP-GLUCOSE: FLAVONOID 3-O-GLUCOSYLTRANSFESASE)
9	AT3G51240	F3H (FLAVANONE 3-HYDROXYLASE)
10	AT4G14090	UDP-GLUCORONOSYL/ UDP-GLUCOSYL TRANSFERASE FAMILY PROTEIN)
11	AT5G13930	TT4 (TRANSPARENT TESTA 4)
12	267256_s_at	
13	254283_s_at	
14	AT5G42800	DFR (DIHYDRIFFLAVONOL 4-REDUCTASE)
15	AT4G23990	ATCSLG3
16	AT1G30530	UGT78D1 (UDP-GLUCOSYL TRANSFERASE 78D1)
17	AT1G06180	ATMYB13 (MYB DOMAIN PROTEIN 13)
18	AT3G22840	ELIP1 (EARLY LIGHT-INDUCABLE PROTEIN)
19	AT1G62540	FMO GS-OX2 (FLAVAIN-MONOOXYGENASE GLUCOSINOLATE S-OXYGENASE 2)
20	AT4G26950	UNKNOWN PROTEIN
21	AT4G13410	ATCSLA15
22	AT3G54990	SMZ (SCHILAFMUTZE)
23	267315_s_at	
24	AT5G01260	GLYCOSIDE HYDROLASE STARCH-BINDING DOMAIN-CONTAINING PROTEIN
25	AT2G34660	ATMRP2 (ARABIDOPSIS THALIANA MULTIDRUG RESISTENCE-ASSOCIATED PROTEIN2)
26	AT3G29590	AT5MAT
27	AT5G05270	CHALCONE-FLAVANONE ISOMERASE FAMILY PROTEIN
28	AT5G07990	TT7 (TRANSPARENT TESTA 7)
29	AT2G19860	HXK2 (HEXOKINASE 2)
30	AT3G13620	AMINO ACID PERMEASE FAMILY PROTEIN
224	AT4G27820	BGLU9 (BETA GLUCOSIDASE 9)

The ATTED-II database was screened using the "Search section" of CoexSearch with the setting 'Coexpression measure: MR' and 'top 300 genes among 22,263 genes'. The table lists the top 30 genes. Genes highlighted in light grey are known to be involved in anthocyanin modification (At5g17050: *UGT78D2, 3GT*, At4g14090: *UGT75C1, 5GT*, At3g29590: *AT5MAT*), those in dark grey are *AtBGLUs*.

different pattern of expression to other carnation genes involved in anthocyanin/flavonoid biosynthesis in the petals (Matsuba et al. 2010). It is thus possible that in *Arabidopsis,* expression of some *AA5/7GT* homologues might involve co-expression and be differently regulated from those of other genes of the anthocyanin/flavonoid biosynthetic pathways.

We used RT-PCR and real-time RT-PCR to investigate the expression patterns and functions of *AtBGLU1–11.* Sterilized *A. thaliana* (ecotype Colombia-0) seeds were germinated and grown for up to a month on solid Murashige and Skoog (MS) medium containing 3% (w/v) sucrose, and then the plants were transferred to soil. All plants were grown at 22°C under 16 h light/8 h dark. The following tissues were selected for analysis of *AtBGLU1–11* expression: roots, rosette leaves, leaves, flowers and siliques. Anthocyanin synthesis was induced in seedlings germinated on medium containing 6% (w/v) sucrose (Solfanelli et al. 2006). Seedlings grown on medium containing 1% (w/v) sucrose were used as a control. Osmotic stress was applied by growing seedlings on medium containing both 0.2 M sorbitol and 1% sucrose. All seedlings were harvested 14 days after germination. Total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The cDNAs were synthesized and real-time RT-PCR was performed as previously reported (Miyahara et al. 2010). RT-PCR was performed using Go Taq Hot Start DNA polymerase (Promega, Tokyo, Japan) with an Applied Biosystems 2720 Thermal Cycler (Life Technologies Japan, Tokyo, Japan). The cDNAs were amplified using the appropriate *AtBGLU1–11* primers (Supplemental Table 1). The elongation factor 1α (EF-1α) gene was used as the control for normalizing *AtBGLU1–11* expression levels. RT-PCR of *AtBGLU1–11* was performed by a preliminary denaturation step at 95°C for 2 min, and then 35 cycles at 95°C for 30 s, 55°C for 30 s and 72°C for 13 s.

Using real-time RT-PCR, we found that expression of *AtBGLU4*, *AtBGLU5* and *AtBGLU8* was below the level of detection. *AtBGLU1, 6, 7,* and *9* showed their highest levels of expression in seedlings grown in the anthocyanin-inducing medium (Table 2). *AtBGLU10* and *11* were highly expressed in plants grown on the 6% sucrose and 0.2 M sorbitol media. *AtBGLU2* and *3* showed the highest expression in leaves. All of the *AtBGLU* genes had low levels of expression in flowers and siliques. The real-time RT-PCR analysis showed that *AtBGLU6* and *AtBGLU9* had low levels of expression, while *AtBGLU1* had a specific increase in anthocyanin-inducing medium. These results suggest that AtBGLU7 and 9, as well as AtBGLU1 and 6, might possibly be involved in anthocyanin biosynthesis or modification by glucosyltransferase. AtBGLU10 and 11 appear to be involved in stress reactions. Although *AtBGLU1* and *6* has been reported to be a pseudogene (Xu et al. 2004), expression of *AtBGLU6* was induced in anthocyanin-inducing medium (Figure 2) and rose to the highest rank among other *AtBGLUs* in the co-expression analysis of genes for enzymes involved in anthocyanin modification

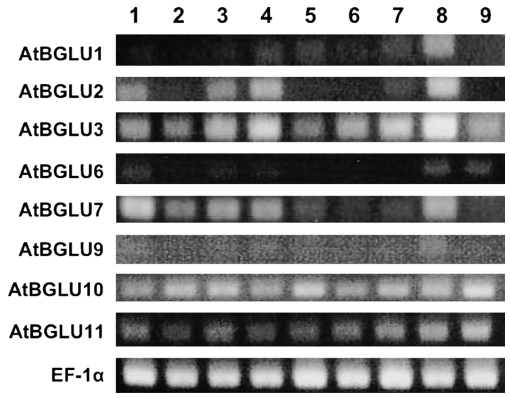

Figure 2. RT-PCR analysis of *AtBGLU1–11*. The tissues samples are numbered as follows: 1, roots; 2, stems; 3, rosette leaves; 4, leaves; 5, flowers; 6, siliques; 7, seedlings from 1% sucrose medium; 8, seedlings from 6% sucrose medium (anthocyanin-inducing medium); 9, seedlings from 0.2 M sorbitol medium (stress medium).

Table 2. Real-time RT-PCR analysis of *AtBGLU1–11* expression levels

	Roots	Stems	Rosette Leaves	Leaves	Flowers	Siliques	1% Sucrose	6% Sucrose	0.2 M Sorbitol
AtBGLU1	1.62 (0.72)	1.00 (0.00)	3.22 (1.14)	31.76 (16.13)	1.93 (0.40)	1.05 (0.16)	3.45 (0.54)	101.58 (33.6)	2.47 (0.88)
AtBGLU2	14.15 (3.23)	4.34 (1.00)	54.39 (7.29)	216.73 (34.63)	2.10 (0.39)	1.23 (0.22)	8.32 (3.09)	54.65 (12.03)	2.11 (0.78)
AtBGLU3	13.66 (4.84)	5.00 (1.84)	22.28 (6.98)	137.27 (29.48)	—	10.26 (3.12)	2.36 (0.032)	60.51 (4.26)	7.68 (1.59)
AtBGLU4	—	—	—	—	—	—	—	—	—
AtBGLU5	—	—	—	—	—	—	—	—	—
AtBGLU6	4.01 (0.80)	1.23 (0.24)	14.42 (5.58)	33.77 (1.57)	1.20 (0.19)	1.34 (0.41)	2.56 (0.30)	58.44 (7.73)	10.18 (3.24)
AtBGLU7	11.92 (0.71)	12.57 (2.77)	20.18 (2.58)	32.27 (4.29)	1.12 (0.012)	1.18 (0.40)	3.18 (0.41)	52.48 (4.82)	1.00 (0.00)
AtBGLU8	—	—	—	—	—	—	—	—	—
AtBGLU9	1.00 (0.00)	—	14.06 (4.90)	27.82 (6.83)	2.57 (0.64)	—	15.00 (5.21)	41.40 (10.83)	1.74 (0.54)
AtBGLU10	1.00 (0.00)	6.36 (0.96)	9.91 (3.55)	17.12 (4.52)	3.69 (0.51)	2.34 (1.11)	3.57 (0.72)	34.68 (8.21)	37.62 (10.01)
AtBGLU11	13.61 (0.10)	9.10 (0.94)	37.07 (6.77)	39.65 (2.81)	4.06 (0.72)	1.00 (0.00)	3.08 (0.63)	58.92 (12.49)	117.66 (18.61)

cDNAs were synthesized from mRNAs prepared from tissues and seedlings as described in the legend to Figure 1. The values presented are means (± SE in parentheses). n=3. Expression was below detectable levels in all tissues examined for *AtBGLU4*, *AtBGLU5* and *AtBGLU8,* in flowers for *AtBGLU3*, and in stems and siliques for *AtBGLU9*.

(Table 1). In order to confirm whether *AtBGLU1* and *6* were pseudogenes, we isolated their complete full length cDNAs and analyzed their sequences; the results showed that *AtBGLU1* cDNA had the complete open reading frame, but *AtBGLU6* cDNA had a stop codon on the middle of its sequence. Other two reported pseudogenes, *AtBGLU2* and *5*, were unable to isolate complete full length cDNAs, so that we abandoned further sequence analysis. These results suggest that *AtBGLU1* was not a pseudogene and it might be active, and the promoter of *AtBGLU6*, which had a stop codon in the middle of its open reading frame, was active and some expression of the gene occurred. Possibly, *AtBGLU6* has only recently become pseudogenic in character due to a mutation in its open reading frame. A similar scenario might apply to *AtBGLU2*; expression of *AtBGLU2* was detectable in our analysis (Figure 2, Table 2), although they have been described as pseudogenic in a previous report (Xu et al. 2004).

In order to clarify the functions of the *AtBGLU1–11* genes that have homology with *DcAA5GT* and *DgAA7GT*, it will be necessary to perform a metabolomics analysis in *AtBGLU* mutants. We have already initiated a program to develop *AtBGLU1–11* knockout mutants by T-DNA insertion and to prepare overexpression transformants.

Acknowledgements

We are grateful to Ms. Patricia McGahan, Tokyo University of Agriculture and Technology, for helpful advice and discussion. This research was supported by grants from the Research and Development Program for New Bio-industry Initiatives to Y.O. and Research Fellowships of Japan Society for the Promotion of Science for Young Scientists to Y.M.

References

D'Auria JC, Reicheit M, Luck K, Svatos A, Gershenzon J (2007) Identification and characterization of the BAHD acyltransferase malonyl CoA: anthocyanidin 5-*O*-glucoside-6″-*O*-malonyltransferase (At5MAT) in *Arabidopsis thaliana*. *FEBS Lett* 581: 872–878

Hancock SM, Corbett K, Fordham-Skelton AP, Gatehouse JA, Davis BG (2005) Developing promiscuous glycosidases for glycoside synthesis: residues W433 and E432 in Sulfolobus solfataricus beta-glycosidase are important glucoside- and galactoside-specificity determinants. *Chembiochem* 6: 866–875

Kubo H, Nawa N, Lupsea SA (2007) Anthocyaninless1 gene of Arabidopsis thaliana encodes a UDP-glucose:flavonoid-3-*O*-glucosyltransferase. *J Plant Res* 120: 445–449

Matsuba Y, Sasaki N, Tera M, Okamura M, Abe Y, Okamoto E, Nakamura H, Funabashi H, Takatsu M, Saito M, et al. (2010) A Novel Glucosylation Reaction on Anthocyanins Catalyzed by Acyl-Glucose Dependent Glucosyltransferase in the Petals of Carnation and Delphinium. *Plant Cell* 22: 3374–3389

Mayer C, Zechel DL, Reid SP, Warren RA, Withers SG (2000) The E358S mutant of Agrobacterium sp. beta-glucosidase is a greatly improved glycosynthase. *FEBS Lett* 466: 40–44

Miyahara T, Satoh S, Maeda K, Kimura S, Sasaki N, Ozeki Y (2010) Isolation of *Daucus carota* ethylene insensitive3-like (DcEIL) involved in stress-inducible *DcMYB1* expression in suspension-cultured carrot cells. *Plant Biotechnol* 27: 91–97

Opassiri R, Pomthong B, Onkoksoong T, Akiyama T, Esen A, Ketudat Cairns JR (2006) Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 beta-glucosidase. *BMC Plant Biol* 6: 33

Shaikh FA, Withers SG (2008) Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochem Cell Biol* 86: 169–177

Solfanelli C, Poggi A, Loreti E, Alpi A, Perata P (2006) Sucrose-specific induction of the anthocyanin biosynthetic pathway in Arabidopsis. *Plant Physiol* 140: 637–646

Sugano SS, Shimada T, Imai Y, Okawa K, Tamai A, Mori M, Hara-Nishimura I (2010) Stomagen positively regulates stomatal density in Arabidopsis. *Nature* 463: 241–244

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, et al. (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J* 42: 218–235

Xu Z, Escamilla-Treviño L, Zeng L, Lalgondar M, Bevan D, Winkel B, Mohamed A, Cheng CL, Shih MC, Poulton J, et al. (2004) Functional genomic analysis of Arabidopsis thaliana glycoside hydrolase family 1. *Plant Mol Biol* 55: 343–367

Yonekura-Sakakibara K (2009) Functional genomics of family 1 glycosyltransferases in Arabidopsis. *Plant Biotechnol* 26: 267–274