Identification of a Flavonol 7-O-Rhamnosyltransferase Gene Determining Flavonoid Pattern in Arabidopsis by Transcriptome Coexpression Analysis and Reverse Genetics*

Received for publication, December 15, 2006, and in revised form, February 2, 2007. Published, JBC Papers in Press, February 21, 2007, DOI 10.1074/jbc.M611498200

Keiko Yonekura-Sakakibara1, Takayuki Tohge1, Rie Niida1, and Kazuki Saito1,2

From the 1RIKEN Plant Science Center, Suehiro-cho 1-7-22, Tsurumi-ku, Yokohama 230-0045, Japan and the 2Graduate School of Pharmaceutical Science, Chiba University, Yayoi-cho 1-33, Inage-ku, Chiba 263-8522, Japan

Glycosylation plays a major role in the remarkable chemical diversity of flavonoids in plants including Arabidopsis thaliana. The wide diversity encoded by the large family-1 glycosyltransferase (UGT) gene family makes it difficult to determine the biochemical function of each gene solely from its primary sequence. Here we used transcriptome coexpression analysis combined with a reverse genetics approach to identify a gene that is prominent in determining the flavonoid composition of Arabidopsis. Using transcriptome coexpression analysis accessible on the ATTED-II public data base, the expression pattern of a UGT gene, UGT89C1, was found to be highly correlated with known flavonoid biosynthetic genes. No 7 rhamnosylated flavonols were detected in either of two T-DNA ugt89c1 mutants. This specific metabolite deficiency in the mutants was complemented by stable transformation with the genomic fragment containing intact UGT89C1. Glutathione S-transferase-fused recombinant UGT89C1 protein converted kaempferol 3-O-glucoside to kaempferol 3-O-glucosyl-rhamnoside-7-O-rhamnoside and recognized 3-O-glycosylated flavonols and UDP-rhamnose as substrates, but not flavonol aglycones, 3-O-glycosylated anthocyanins or other UDP-sugars. These results show that UGT89C1 is a flavonol 7-O-rhamnosyltransferase. The abundance of UGT89C1 transcripts in floral buds was consistent with the flavonoid accumulation of 7-O rhamnosylated flavonols in Arabidopsis organs. Our present study demonstrates that the integration of transcriptome coexpression analysis with a reverse genetic approach is a versatile tool for understanding a multigene family of a metabolic pathway in Arabidopsis.

Plants have the ability to produce a large variety of secondary metabolites, such as terpenoids, phenylpropanoids, flavonoids, and alkaloids, which together account for over 200,000 compounds (1). Even in Arabidopsis, which has a small genome when compared with other flowering plants, more than 170 secondary metabolites have been identified (2). Flavonoids, including the anthocyanins, flavonols and flavones, are among the most intensely studied secondary product groups, with over 6,000 known (3). Many of them play important roles as flower and fruit pigments, UV-B protectants, signaling molecules between plants and microbes, and regulators of auxin transport (4, 5). The flavonoids are also thought to have antioxidant, anti-allergic, and anti-inflammatory effects, thus contributing to human health (6, 7). Genes for the core biosynthetic steps leading to anthocyanins and flavonols have been identified, and knowledge of the molecular and biochemical properties of these enzymes and their regulatory factors has significantly progressed in the last few years (reviewed in Refs. 8 and 9). However, the flavonoid modification steps are yet to be completely understood. Although structural variations within the flavonoid skeleton are relatively limited, sequential modifications of the flavonoid core structure by glycosylation, acylation, and methylation result in increasing chemical diversity.

Completion of the Arabidopsis and rice genome sequencing projects as well as the contribution of identified nucleotide sequences from several other model plant species have allowed us to annotate a finite number of genes involved in primary or secondary metabolism by homology. However, most enzymes involved in the modification of secondary metabolites are encoded by multigene families (2, 10). For example, 107 UGT3 genes have been identified by homology in Arabidopsis (2), but only four, encoding flavonoid 3-O-glycosyltransferase (3GlCT), flavonol 3-O-rhamnosyltransferase (3RhaT), flavonol 7-O-glucosyltransferase (7GlcT), and anthocyanin 5-O-glycosyltransferase (5GlcT) have yet been characterized (11, 12). Previous reports (13) indicated the presence of eight flavonols and 11 anthocyanins whose structures were determined in Arabidopsis. Glycosylation patterns of the flavonoid structures suggest

* This study was supported in part by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and Uehara Memorial Foundation, Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 These authors contributed equally to this work.

2 To whom correspondence should be addressed. Tel.: 81-45-503-9488; Fax 81-45-503-9489; E-mail: ksaito@faculty.chiba-u.jp.

3 The abbreviations used are: UGT, family-1 glycosyltransferase; ANS, anthocyanidin synthase; CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydroflavonol reductase; ESI, electrospray ionization; MS, mass spectrometry; Q-TOF, quadrupole time-of-flight; RT, reverse transcription; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; GST, glutathione S-transferase; PDA, photodiode array; 3GlCT, 3-O-glycosyltransferase; 5GlCT, 5-O-glycosyltransferase; 7GlcT, 7-O-glucosyltransferase; 3RhaT, 3-O-rhamnosyltransferase; 7RhaT, 7-O-rhamnosyltransferase; Fd3GlCT, flavonol 3GlCT.
that five unknown UGT genes are involved in anthocyanin and/or flavonol glycosylation in *Arabidopsis* in addition to the previously identified four UGTs (flavonoid 3GlcT, flavonol 3RhaT, anthocyanin 5GlcT and flavonol 7GlcT). A majority of *Arabidopsis* flavonols (i.e. seven of eight flavonols whose structures have been determined) are 7-O-rhamnosylated (13–16). It is apparent that a flavonol 7-rhamnosyltransferase (7RhaT) is a prominent enzyme that is largely responsible for the determination of flavonoid patterns in *Arabidopsis*. The flavonol 7-O-rhamnosides are not limited to *Arabidopsis* but occur in a number of plant species (17). Of the 1,331 known flavonol O-glycosides, 140 are 7-O-rhamnosides (10.5%) (17). Of the 1,661 flavonol O-glycosides registered in the KNApSAcK data base, A Comprehensive Species-Metabolite Relationship Data-base (18), flavonol 7-rhamnosides account for 11.6%.  

However, no gene coding for 7RhaT has been identified despite its importance in flavonoid biosynthesis in *Arabidopsis* and other plant species. The huge diversity of UGTs makes it difficult to define their biochemical functions solely by their primary structures as a simple phylogenetic relationship based upon primary structure can give ambiguous and often misleading predictions of biochemical function. Therefore, a clearer understanding of the flavonoid modifying enzymes may be gained by developing of novel criteria by genome-based information.

Recently, the integration of transcriptome and metabolome analyses has emerged as a promising technology for functional genomics (12, 19, 20). In addition, publicly available *Arabidop-sis* transcriptome databases have been markedly improved and expanded. Notably, The *Arabidopsis* Information Resource (TAIR) (21) and the Nottingham *Arabidopsis* Stock Centre Arrays (NASCArrays) (22) have become available as *Arabidopsis*-specific repositories for microarray data. Furthermore, a number of secondary databases based on microarray data are accessible on-line: ATTED-II, a data base for co-regulated gene relationships in *Arabidopsis thaliana* to estimate gene functions (23); PRIme (Platform for RIKEN Metabolomics), a data base for integrated analysis of transcriptomics and metabolomics; a co-response data base in the comprehensive system-biology data base (CSDB); and Genevestigator, an *A. thaliana* microarray data base and analysis toolbox. Utilization of these public databases may greatly help to identify genes of interest for particular biochemical functions.

In the present study, we identified at least one flavonoid UGT gene from among 107 candidates using metabolomics, transcriptomics, and reverse genetics as a first step to complete the model flavonoid pathway in *Arabidopsis*. By coexpression analysis of transcriptome data sets of TAIR, we found that the expression of a UGT gene, *UGT89C1*, is highly correlated with the expression of known flavonoid biosynthetic genes. Flavonoid profiling in *ugt89c1* knock-out mutants and a recombinant protein assay confirmed that *UGT89C1* catalyzes rhamnosylation at the C-7 position of flavonols. Our results indicate that a comprehensive strategy combining coexpression analysis, metabolic profiling, reverse genetics, and biochemistry can be a versatile tool for the functional identification of genes that belong to a multigene family and to complete the model of a particular metabolic pathway in *Arabidopsis*.

**EXPERIMENTAL PROCEDURES**

**Plant Materials**—*A. thaliana* plants (ecotype Columbia-0) were grown on germination medium (24) at 22 °C under 16-h/8-h light and dark cycles. The light intensity was 40 μmol of photons m⁻² s⁻¹. T-DNA insertion lines for *UGT89C1* were obtained from the SALK Institute and were screened by PCR using specific primers for T-DNA and *UGT89C1*, *UGT89C1f*, *UGT89C1r*, *Lba1*, and *Rba1* (supplemental Table 1). PCR products were sequenced to determine the exact insertion points.

**Chemicals**—Chemicals of the highest grade commercially available were used unless specifically noted. Anthocya-nin and flavonol standards were purchased from Extrasyn-these and Analyticon. Kaempferol 3-O-α-L-rhamnose and kaempferol 3,7-O-bis-α-L-rhamnose were kindly provided by Professor H. Takayama, Chiba University, Japan. UDP-rhamnose synthesized as described previously (25) was purchased from Funakoshi. The resultant product consisted of UDP-rhamnose, UMP, UDP, and the condensation product of UDP-diphenylphosphate with UMP. The purity of UDP-rhamnose was determined by liquid chromatography-mass spectrometry (MS) as described previously (11).

**Coexpression Analyses**—Coexpression analyses were carried out using a Coexpression Gene Search algorithm on the RIKEN PRIME web site. The Coexpression Gene Search program is a web-based application designed to identify correlated genes from gene expression data produced using Affymetrix GeneChip technology by the AtGenExpress consortium (RIKEN Plant Science Center and the Max-Planck Institute for Molecular Plant Physiology) deposited in TAIR. Data for version 1 (all data sets, version 1, tissue and development, stress treatments, and hormone treatments) were used as a matrix for the analyses. Correlation coefficients have already been released in ATTED-II (23). To minimize the effects of experimental artifacts, data were renormalized, and Pearson’s correlation coefficient between genes was weighted in ATTED-II. The Pajek program was used for output.

**General Molecular Procedures**—The molecular procedures used were as described previously (26) unless otherwise specified. A phylogenetic tree was generated from a multiple alignment by the CLUSTALW program at the DNA Data Bank of Japan web site using the neighbor joining method of the TREE-VIEW program (27). Reverse transcription (RT)-PCR was performed as described previously (28) with primers *ugt89c1-1f* and *ugt89c1-1r* for *ugt89c1-1*, *ugt89c1-2f* and *ugt89c1-2r* for *ugt89c1-2*, and *TUBf* and *TUBr* for tubulin (GenBankTM Accession number AK117431) (supplemental Table 1).

**Plant Transformation**—A 3-kb genomic fragment covering the 1,544 bp of the promoter region, the entire *UGT89C1* coding region, and 187 bp of the 3’ non-coding region was amplified by PCR with primers *UGT89C1-GWf* and *UGT89C1-GWr* and used as a genomic clone for the complementation test. The Gateway system was used for construction of the binary vector for *Arabidopsis* transformation. The PCR product was cloned into the pENTR™/D-TOPO vector (Invitrogen Japan) as an

---

4. T. Tokimatsu and M. Arita, personal communication.
Arabidopsis Flavonol 7-O-Rhamnosyltransferase

entry vector and sequenced to confirm the absence of PCR errors. pGWY1 was used as a destination vector, and the LR reaction for the binary vector, pKYS320, was catalyzed by the Gateway LR clonase enzyme mix (Invitrogen). pKYS320 was transformed into Agrobacterium tumefaciens GV3101(pMP90) by the freeze and thaw method (29), and Arabidopsis plants were transformed by the floral dip method (30). Transgenic T2 plants were selected on 1/2 Murashige-Skoog medium containing 25 mg liter^{-1} hygromycin B and 50 mg liter^{-1} carbenicillin disodium salt. mRNA accumulation in T2 plants was checked by RT-PCR as described above.

Expression and Purification of Recombinant UGT89C1 Protein—A full-length cDNA clone of UGT89C1 (pda08132) was obtained from the RIKEN BioResource Center Arabidopsis full-length cDNA collection (31, 32). The full-length UGT89C1 was amplified by PCR using the primers UGT89C1-BamHI and UGT89C1-PstI (supplemental Table 1), and the amplified fragment was cloned into pCR-Blunt II-TOPO vector (Invitrogen) to confirm the nucleotide sequence and into pET-15b (Novagen, San Diego, CA). Escherichia coli strain BL21star(DE3) was used as a host for expression. Transformed cells were cultivated at 37 °C until A_{600} reached 0.5. After the addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 mM, cells were cultured at 20 °C for 4 h. The cells were collected, and the protein was purified as a glutathione S-transferase (GST) fusion according to the manufacturer’s instructions.

Enzyme Assays—The standard enzyme assay reaction mixture (final volume, 100 μl) consisted of 50 mM potassium phosphate, pH 7.0, 150 μM flavonoid substrate, and 500 μM UDP-sugar. The mixture was preincubated at 30 °C for 2 min, and the reaction was started by the addition of enzyme. Reactions were stopped after 0, 4, 8, or 12 min of incubation at 30 °C by the addition of 100 μl of ice-cold 1.0% (v/v) trifluoroacetic acid, and the supernatant was recovered by centrifugation at 12,000 × g for 3 min. Flavonoids in the resultant solution were analyzed as described below.

Flavonoid Analysis by UPLC/TM/PDA/ESI-Q-TOF/MS—Frozen Arabidopsis leaves were homogenized in extraction solvent (methanol: H₂O = 4:1) with 5 μl of solvent/mg of fresh weight in a mixer mill (MM300; Retsch GmbH & Co. KG) for 3 min at 30 Hz. After centrifugation at 12,000 × g for 10 min, cell debris was discarded, and supernatants were recentrifuged. The resultant supernatants were immediately analyzed with a Waters Acquity UPLC system (Waters Corp.) fitted with a Q-TOF Premier mass spectrometer (Micromass MS Technologies). A 2-μl sample was applied to an ACQUITY UPLC BEH C₁₈ column (Φ2.1 × 100 mm, 1.7 μm, Waters) at a flow rate of 0.5 ml/min with linear gradients of solvent A (0.1% formic acid in H₂O) and solvent B (0.1% formic acid in methanol) set according to the following profile: 0 min, 95% solvent A +5% solvent B; 9 min, 60% solvent A + 40% solvent B; 11 min, 100% solvent B; 13 min, 95% solvent A +5% solvent B. The column temperature was 35 °C. Photodiode array (PDA) was used for detection of UV-visible absorption in the range of 210–500 nm. Electrospray ionization (ESI) with positive mode was used. The TOF mass analyzer was used for detection of flavonoid glycosides [M + H]⁺ and fragment ion peak in a positive ion mode scanning with the following setting: desolvation temperature was 450 °C at a nitrogen gas flow rate of 600 liters/h, capillary spray 3.2 kV, source temperature 150 °C, and cone voltage 35 V.

Identification of the peaks in the plant extracts was based on comparisons of retention times, UV-visible absorption spectra, and mass fragmentation patterns by tandem MS analysis of the flavonoid standards. Other flavonoids with no standard compounds were annotated by comparison with the reported data in the UV-visible absorption spectra, elution time, m/z values, and MS² fragmentation patterns.

Quantitative Real-time PCR—RNA Extraction and cDNA synthesis were performed as described previously (28). The developmental stage of each organ used for analysis was as follows: leaves, 18 days after sowing; stems, 25 days after sowing; roots, 8 days after sowing; siliques, 11 days after flowering. Accumulation levels of the UGT89C1 transcripts were analyzed by a real-time PCR method, with an ABI PRISM 7500 real-time PCR system (Applied Biosystems) monitoring the amplification with the SYBR-Green I dye (Applied Biosys-
RESULTS

Coexpression Analysis to Predict the Genes Encoding Flavonoid-related Glycosyltransferases—To narrow the number of potential UGT candidates involved in flavonoid modification, we used coexpression analyses within the ATTED-II data base in a coexpression gene search program on the RIKEN PRIME web site. First, we verified the reliability of the data base by searching correlation coefficients among the known genes encoding flavonoid biosynthetic enzymes. The linkages between genes that had a higher correlation coefficient (r>0.6) in all GeneChip data are presented in Fig. 1. Genes involved in flavonoid biosynthesis were divided into two groups by this coexpression analysis. The first group contains chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonol 3′-hydroxylase (F3′H), and flavonol synthase (FLS), and the second group contains dihydroflavonol reductase (DFR), anthocyanidin synthase (ANS), and GST(TT19). The members of each resultant group are consistent with their classification as “early” (CHS, CHI, F3H, and FLS) or “late” genes (DFR and ANS) of anthocyanin biosynthesis (33, 34). In addition, this result suggested that F3′H is an early gene and GST is a late gene from the viewpoint of coexpression patterns. Coexpression analysis also indicated that two MYB transcriptional regulators, MYB12 for flavonol (35) and PAP1 for anthocyanin (36), form coexpression networks with structural genes for biosyntheses of flavonol and anthocyanin, respectively. PAP1 showed no correlation with DFR, ANS, or GST by these data sets. However, in the stress treatment sub-data set, PAP1 correlated with DFR (r = 0.636), ANS (r = 0.587), and GST (r = 0.660). MYB12 also correlates with CHS (r = 0.677) and F3′H (r = 0.706) by the sub-data set from the tissue and development context. These results suggest that correlation analysis can be useful for finding functional clues about uncharacterized genes in the flavonoid biosynthesis pathway.

To identify novel UGT genes likely to be involved in the flavonoid biosynthetic pathway, we analyzed the correlation networks of 107 UGT genes (37) using the known structural and regulatory genes (i.e. CHS, CHI, F3H, F3′H, FLS, DFR, ANS, GST, PAP1, and MYB12) as “bait” (Fig. 1). Significantly positive correlations were observed between the bait genes involved in flavonoid biosynthesis and five UGT genes: At4g14090 (UGT75C1, anthocyanidin 5GlcT) (12), At5g17050 (UGT78D2, flavonoid 3′-GlcT) (12), At5g54060 (UGT79B1), At4g15480 (UGT84A1) and At1g06000 (UGT89C1). At3g21560 (UGT84A2) was correlated with At5g17050 (UGT78D2). Although flavonoid 3GlcT (Fd3GlcT) catalyzes 3-O-glicosylation of both flavonols and anthocyanidins, it fell into the early genes cluster, implying that the early genes are involved both in biosynthesis of flavonol and in the early steps of anthocyanin production. Anthocyanidin 5GlcT fell within the late genes cluster. These results suggest that three UGTs (UGT84A1, UGT84A2, and UGT89C1) are involved in flavonol and/or anthocyanidin glycosylation and that UGT79B1 may be involved only in anthocyanin glycosylation. UGT89C1 was
Arabidopsis Flavonol 7-O-Rhamnosyltransferase

strongly correlated with the flavonol biosynthetic genes CHS, CHI, F3H, and FLS and Fd3GlcT. We thus gave priority to the characterization of UGT89C1, among the four unknown UGTs forming coexpression networks in the flavonoid pathway.

UGT89C1 Belongs to the Subfamily of UGTs Catalyzing the 7-O-Glycosylation of Flavonoids—The flavonoid UGT phylogenetic tree indicates that UGT89C1 belongs to the subfamily in which some UGTs catalyze glycosylation at the flavonoid C-7 position (cluster III) (Fig. 2). The amino acid sequence identity of UGT89C1 with the known Arabidopsis flavonoid UGTs, flavonol 7GlcT (UGT73C6), anthocyanin 5GlcT (UGT75C1), flavonoid 3GlcT (UGT78D2), and flavonol 3RhaT (UGT78D1) is 30, 26, 24, and 24%, respectively. When compared with the percentage of identity between Fd3GlcT and flavonol 3RhaT (72%), UGT89C1 had much lower identity with flavonol 7GlcT (30%). Furthermore, UGT89C1 has a 50-amino-acid gap when aligned with other plant UGTs involved in secondary metabolism (data not shown).

Flavonols, but not anthocyanins, that are glycosylated with glucose or rhamnose at their C-7 position have been identified in Arabidopsis (12, 15), implying the presence of 7GlcT and 7RhaT, and the gene encoding flavonol 7GlcT has already been identified (11). It is likely, then, that UGT89C1 encodes a novel flavonol 7RhaT.

Analysis of T-DNA Mutants of UGT89C1—To confirm the physiological function of UGT89C1 in Arabidopsis, two independent T-DNA insertion lines, SALK_068559 and SALK_071113, were designated as ugt89c1-1 and ugt89c1-2, respectively. The T-DNA of both mutants was inserted in the exon of UGT89C1 in both mutants but between positions +1034 to +1042 and +78 to +141 base pairs, respectively, from the putative transcription start site proposed from the full-length cDNA sequence (GenBank accession number AY093133) (Fig. 3A). No transcripts of UGT89C1 were detected by RT-PCR in homozygotes of either line (Fig. 3B), and there were no obvious phenotypic abnormalities in the mutant plants (data not shown).

The flavonoid profiles of wild-type, ugt89c1-1, and ugt89c1-2 plants were analyzed by UPLC/PDA/ESI-Q-TOF/MS (Fig. 4), focusing on flavonol composition because UGT89C1 was presumed to catalyze 7-O-glycosylation of flavonols, and no anthocyanin 7-O-glycosides have been detected in Arabidopsis, even by overexpression of PAP1. Three major kaempferol 7-O-rhamnosides, f1, kaempferol 3-O-rhamnoside-7-O-rhamnoside, f2, kaempferol 3-O-glucoside-7-O-rhamnoside, and f3, kaempferol 3-O-rhamnosyl(1→2)glucoside-7-O-rhamnoside, essentially disappeared, and new kaempferol derivative peaks were detected in the leaves and roots of the homozygous ugt89c1-1 line. Likewise, the levels of the corresponding quer- cetin 7-O-rhamnosides, i.e. f6 and f8, were also drastically reduced (supplemental Fig. 1). The flavonol profile of ugt89c1-2 was essentially identical (data not shown).

To demonstrate that the metabolic types deficient in 7-O-rhamnosylated flavonol are correctly ascribed to the UGT89C1 mutation, both ugt89c1-1 and ugt89c1-2 plants were transformed with a genomic clone of UGT89C1. Eight independent transgenic lines that accumulate UGT89C1 transcripts had the same flavonoid profile as wild-type plants (Fig. 4B). All of these

in vivo data indicate that UGT89C1 encodes UDP-rhamnose: flavonol 7-O-rhamnosyltransferase.

In Vitro Characterization of Recombinant UGT89C1—Recombinant UGT89C1 protein was expressed in E. coli as a GST fusion and purified. The GST-UGT89C1 fusion catalyzed the conversion of kaempferol 3-O-glucoside to a single product, kaempferol 3-O-glucoside-7-O-rhamnoside (Fig. 5), as confirmed by UPLC retention time and MS2 ionization when compared with the standard compound. GST alone as a negative control catalyzed no conversion to the 7-O-rhamnoside. Thus, UGT89C1 can be defined as a flavonol 7RhaT.

The specificity of UGT89C1 for a rhamnose acceptor was examined using the flavonols kaempferol, quercetin, isorhamnetin, and their glycosides, an anthocyanidin (cyanidin) and the anthocyanins cyanidin 3-O-glucoside, cyanidin 3-O-rhamnoside, and cyanidin 3-O-rhamnosyl(1→6)glucoside (Table 1). UGT89C1 had significant rhamnosyltransferase activity with the flavonol 3-O-monoglycosides and 3-O-diglycosides but less or no activity on the flavonol aglycones, cyanidin, or the anthocya-
nins. A single product peak was detected in each of the reaction mixtures with UGT89C1, flavonol glycosides, and UDP-sugars. Kaempferol 3-O-glucoside-7-O-rhamnoside and quercetin 3-O-rhamnoside-7-O-rhamnoside were used as standard compounds for peak identification.

The sugar donor specificity of UGT89C1 was examined with UDP-glucose, UDP-rhamnose, UDP-galactose, and UDP-glucuronic acid as donors and kaempferol 3-O-glucoside as acceptor (Table 1). No UGT activity was detected for UDP-glucose, UDP-galactose, or UDP-glucuronic acid, indicating that UGT89C1 is highly specific for UDP-rhamnose.
Expression of \textit{UGT89C1} and Flavonol Accumulation in \textit{Arabidopsis}—The accumulation of \textit{UGT89C1} transcripts in plant organs was measured by real-time PCR (Fig. 6A). Transcripts of \textit{UGT89C1} were detected in all tissues tested but were particularly abundant in floral buds and at very low levels in roots and siliques. This accumulation pattern was consistent with the distribution of \textit{UGT89C1}-MPSS signatures in the \textit{Arabidopsis} MPSS data base, in which \textit{UGT89C1} expression was strong in floral buds but weak in leaves, roots, and siliques.

The accumulation levels of three kaempferol glycosides (f1–f3) and three quercetin glycosides (f5, f6, and f8) in leaves, floral buds, and flowers were determined by UPLC/PDA/ESI-Q-TOF/MS. Concentrations of these six flavonols in buds and flowers ranged from ~10- to 70-fold higher than in leaves (Fig. 6B). These organ-specific concentrations are highly correlated with the pattern of \textit{UGT89C1} transcript accumulation as the final metabolic product would be expected to accumulate in a subsequent developmental stage. The kaempferol glycosides were the predominant flavonols in leaves, making up 97% of total flavonoids. However, 25% of total flavonoids in floral buds and flowers were quercetin glycosides.

\textbf{DISCUSSION}

\textbf{Evaluation of Coexpression Analysis for Functional Genomics and Other Candidate Genes Involved in Flavonoid Biosynthesis}—Flavonol 7\textit{RhaT} contributes significantly to the flavonoid composition of \textit{Arabidopsis} (13). Of the eight known flavonol glycosides in \textit{Arabidopsis}, seven are flavonol 7-\textit{O}-rhamnosides (13). In the present study, we could efficiently target flavonoid-related \textit{UGT}s among 107 candidates by coexpression analyses and were able to identify a flavonol 7\textit{RhaT} gene by a combination of reverse genetics and the enzymatics of recombinant proteins.

Results of our coexpression analyses are in remarkable agreement with the sum of what is known about flavonoid biosynthesis, such as classification of genes as either early or late. The results also confirmed what is known about tissue-specific timing of expression of the flavonoid biosynthetic genes. \textit{UGT89C1} was also significantly correlated with the putative nucleoside diphosphate-rhamnose synthase gene, \textit{At1g78570} (r = 0.641, all data sets; 0.765, tissue and development; 0.763, stress treatments). Besides \textit{UGT89C1}, \textit{At1g78570} also correlated with another \textit{UGT}, \textit{At1g30530} (flavonoid 3\textit{RhaT}, r = 0.654 from the stress matrix). These data suggest that \textit{At1g78570} may function as a UDP-rhamnose synthase, and the expression of \textit{UGT89C1} is well coordinated with the biosynthetic genes of UDP-rhamnose and the flavonols. There was no correlation between flavonol 7GlcT and the flavonoid 3\textit{RhaT} genes.
vonoid biosynthetic genes among the known flavonoid UGTs. Presumably this is due to the unique distribution of flavonol 7GlcT transcripts that are accumulated in leaves and flowers but not in roots, siliques, or stems, despite the constitutive expression of other flavonoid biosynthetic genes (11). This observation is also supported by “Digital in situ” in The Arabidopsis Gene Expression Data base (38), which indicates that flavonol 7GlcT expression is completely different from other flavonoid biosynthetic genes, even within the same organ (supplemental Fig. 2). The relationship among biosynthetic genes and transcription factors is also not so obvious with all data sets of the transcriptome. MYB12 and PAP1 link to flavonoid biosynthetic genes, but the regulators TTG1, GL3, and EGL3 do not. The lack of linkage by TTG1, GL3, and EGL3 may be due to their unique transcript distribution for root hair formation and mucilage production except for flavonoid biosynthesis and/or their lower expression levels when compared with structural genes and MYB12/PAP1. Correlation coefficients in a limited category or under highly specified conditions may be a much more effective index than any of the existing data sets for finding a particular gene, or set of genes, but it is difficult to find a gene that is expressed only under very limited conditions using coexpression analysis of many data sets because the limited number of correlations reduces its correlation coefficient to close to the background noise level. Metabolite target analysis and microarray analysis using a single cell from each of the various organs, or preferably, specific tissues within an organ, will be required to eliminate “signal noise” from genes that overlap the pathway of interest. However, identification of UGT89C1 is a clear example of a proof-of-concept for filling in the blanks in a metabolic scheme.

The Complete Flavonol Glycosylation Pathway in Arabidopsis—Integration of UGT89C1 substrate preference with previous reports allows us to draw the complete flavonol glycosylation pathway in Arabidopsis (Fig. 7). The substrate preference of UGT89C1 suggests that flavonol 7-O-rhamnosylation occurs after 3-O-glycosylation. The exact pathway to kaempferol 3-O-

### TABLE I

| Sugar acceptor          | Relative activity | %     |
|-------------------------|-------------------|-------|
| Kae 3-O-glucoside       | 7 ± 3             |
| Kae 3-O-rhamnosyl(1→6)glucoside | 100 ± 4          |
| Quercetin (Que) 3-O-glucoside | 74 ± 2            |
| Quercetin (Que) 3-O-galactoside | 67 ± 2            |
| Quercetin (Que) 3-O-rhamnosyl(1→6)glucoside | 31 ± 1            |
| Isorhamnetin (Iso) 3-O-glucoside | 73 ± 1            |
| Iso 3-O-glucosyl(1→6)glucoside | 76 ± 2            |
| Iso 3-O-rhamnosyl(1→6)glucoside | 68 ± 2            |
| Cyanidin (Cya) 3-O-glucoside | ND                 |
| Cya 3-O-rhamnosyl(1→6)glucoside | ND                 |
| Cya 3-O-rhamnosyl(1→6)glucoside | ND                 |
| Cya 3-O-rhamnosyl(1→6)glucoside | ND                 |
| Sugar donor             |                   |
| UDP-rhamnose            | 100 ± 4           |
| UDP-glucose             | ND                 |
| UDP-galactose           | ND                 |
| UDP-glucuronic acid     | ND                 |

* The enzymatic products are identified based on comparisons with the standards.

---

**FIGURE 5. UPLC analyses of the reaction products of UGT89C1 recombinant protein.** A, elution profiles of the standards (kaempferol 3-O-glucoside and kaempferol 3-O-glucosyl-7-O-rhamnoside), reaction products of GST protein (Control), or GST-fused UGT89C1 protein (UGT89C1) are shown. **kae-3glu**, kaempferol 3-O-glucoside; **kae-3glu-7rha**, kaempferol 3-O-glucosyl-7-O-rhamnoside. B, mass spectrum of the product catalyzed by UGT89C1 is identical to that of the standard, kaempferol 3-O-glucosyl-7-O-rhamnoside.
indicate that UGT89C1 can utilize kaempferol 3-O-rhamnosyl(1→6)glucoside and quercetin 3-O-rhamnosyl(1→6)glucoside as substrates about equally with flavonol 3-O-glucoside, but we could not test its the substrate specificity for kaempferol 3-O-rhamnosyl(1→2)glucoside because this compound is not available. A genetic study in Ipomea nil and Petunia hybrida demonstrated that anthocyanins are further glycosylated at the C-5 position after the completion of anthocyanidin 3-O-glucosyl(1→2)glucoside and anthocyanidin 3-O-glucosyl(1→6)glucoside modifications, respectively (39–41). Anthocyanidin 3-O-glucoside-2′-O-glucosyltransferase from L. nil recognizes anthocyanidin 3-O-glucoside but not anthocyanidin 3,5-O-diglucoside as a substrate (41). Anthocyanidin 3-O-glu- coside-2′′-O-glucuronosyltransferase from Bellis perennis also had negligible activity toward anthocyanidin 3,5-O-diglucoside (42).

These data may suggest two possibilities. One is that there are metabolic grids leading to the triglycosylated flavonols. The other is that triglycosylated flavonols are synthesized from flavonol 3-O-glucosides only through flavonol 3-O-rhamnosyl(1→2)glucoside or 3-O-glucosyl(1→6)glucoside, as is the case with anthocyanins.

Flavonol profiles in the roots of wild-type plants are more complicated than in leaves. In ugt89c1 mutants, unknown flavonol derivative peaks (U1, U3, and U4) were detected only in leaves. These results may be due to a different distribution pattern of enzymes in the two organs that can modify flavonoids. For example, flavonol 7-O-rhamnosides were missing, and unknown flavonol derivatives (but not flavonol 3-O-glycosides) were accumulated in leaves of ugt89c1 mutants, suggesting that flavonol 3-O-glycosides are unstable and/or rapidly modified further in planta.

Structural Consideration of UGTs in Terms of Substrate Specificity—The functional identification of UGT89C1 as a flavonol 7RhaT allowed us to compare the sequence of flavonoid 3GlcT, flavonol 3RhaT, flavonol 7GlcT, and flavonol 7RhaT from A. thaliana (At3GlcT, At3RhaT, At7GlcT, and At7RhaT). The three-dimensional structure of grape flavonoid 3GlcT was very recently determined and suggests the presence of several key residues that interact with UDP-sugar and the flavonoid backbone (43). The amino acid residues Gln-375, Asp-374, and Thr-141 are proposed to interact with hydroxyl groups at the C-2 and C-3, C-3 and C-4, and C-6 positions of the glucose moiety of UDP-glucose, respectively. The residue that corresponds to Gln-375 of the grape 3GlcT is present in At3GlcT and At7GlcT, but the position is Asn in At3RhaT and His in At7RhaT (supplemental Fig. 3). Asp-374 in the grape 3GlcT is conserved in all Arabidopsis flavonoid UGTs, although the configuration of hydroxyl groups at the C-3 and C-4 positions of UDP-sugars are different. Thr-141 varies in Arabidopsis UGTs (Thr in At3GlcT, Ala in At3RhaT, Gly in At7GlcT, and Pro in At7RhaT). It was reported that the His residue corresponding to Gln-375 in grape 3GlcT is conserved among galactosyltransferases and may be involved in recognition of a hydroxyl group configuration at the C-4 position (44). Although the configuration at C-4 is the same in galactose and rhamnose, the His residue is found in At7RhaT, but not in At3RhaT. Thus, the residues involved in sugar donor specificity cannot be ascribed to a single amino acid residue.

Arabidopsis Flavonol 7-O-Rhamnosyltransferase
Evolution of UGTs—The phylogenetic tree implies that flavonoid UGTs diverged from an ancestral gene into isogenes specific for the position of sugar attachment (i.e., divergence into 3-O-glycosyltransferase, 5-O-glycosyltransferase, 7-O-glycosyltransferase, etc.) and later gained the ability to use various UDP-sugars, such as UDP-glucose, UDP-rhamnose, and UDP-galactose. Furthermore, at least in Arabidopsis, UGT enzymes involved in C-3 glycosylation acquired the ability to use UDP-sugars after plant species divergence. However, C-7 glycosyltransferases apparently diverged before speciation. The sequence identity between At3GlcT and At3RhaT (72%) is much higher than between At7GlcT and At7RhaT (30%). It was reported that plant UGTs are divided into two major groups, those containing the conserved intron A and those that primarily contain no introns (45). The former group consists of UGT71-UGT73, UGT79, and UGT88-92, and the latter consists of UGT74-UGT78 and UGT82-UGT87. This divergence suggests that flavonoid 3-O-glycosyltransferase (UGT78) and 7-O-glycosyltransferase (UGT73 and UGT89) evolved independently and/or at different rates.

Coexpression analysis indicated that UGT89C1 and UGT78D2 belong to the early biosynthetic genes and UGT75C1 belongs to the late biosynthetic genes, although UGT89C1 has the ability to glucosylate the C-3 position of both flavonols and anthocyanidins. This trait may be due to the predominant and earlier production of flavonols under normal growth conditions or may be reflective of the evolutionary past of UGT. Chalcones and flavanones were first synthesized over 500 million years ago; flavonols followed later, and finally, anthocyanins appeared about 120 million years ago (46, 47). This time line indicates that early genes probably appeared first, and late genes emerged afterward. UGT78D2 could have evolved contemporaneously with flavonol and could have later adapted to glycosylate anthocyanidins. In general, the classification of UGTs, at least in terms of substrates, may be reflected in the time point of secondary metabolite appearance.

Improved analytical instruments and techniques will lead to the discovery of other secondary compounds and a subsequent unraveling of a number of structural and biosynthetic mysteries.
in the near future. Integration of metabolomics, including target analysis, and transcriptomics may reveal complete metabolic pathways, including intermediates, and may also suggest the evolutionary history, and in turn, the regulation system for each identified gene, protein, or metabolic pathway.

Acknowledgments—We thank M. Shibata and E. Inoue for technical support. We also thank Professor H. Takayama, Chiba University, and Dr. T. Nakagawa, Shimane University, for providing flavonoid standards and pGWB1 vector, respectively. We are grateful to Dr. T. Obayashi for expert advice on ATTED-II, to the AtGenExpress consortium, and to the researchers who donate their microarray data to the consortium.

REFERENCES

1. Dixon, R. A., and Strack, D. (2003) Phytochemistry 62, 815–816
2. D’Auria, J. C., and Gershenson, J. (2005) Curr. Opin. Plant Biol. 8, 308–316
3. Harborne, J. B., and Williams, C. A. (2000) Phytochemistry 55, 481–504
4. Dooner, H. K., Robbins, T. P., and Jorgensen, R. A. (1991) Annu. Rev. Genet. 25, 173–199
5. Dixon, R. A., and Paiva, N. L. (1995) Plant Cell 7, 1085–1097
6. Scalbert, A., Johnson, I. T., and Saltmarsh, M. (2005) Annu. Rev. Nutr. 215S–217S
7. Ross, J. A., and Kasum, C. M. (2002)
8. Winkel-Shirley, B. (2001)
9. Goossens, A., Hakkinen, S. T., Nuutila, A. M., Soderlund, H., Zabeau, M., and de Sutter, V. (2005) Plant Physiol. Biochem. 43910–43918
10. Bowles, D., Lim, E. K., Poppenberger, B., and Vaistij, F. E. (2006)
11. Jones, P., Messner, B., Nakajima, J., Schaffner, A. R., and Saito, K. (2003) J. Biol. Chem. 278, 43910–43918
12. Tohge, T., Yonekura-Sakakibara, K., Kojima, M., Yamaya, T., and Sakakibara, H. (2007) Pure Appl. Chem. 79, 811–823
13. Veit, M., and Pauli, G. F. (1999) J. Nat. Prod. (Lloydia) 62, 1301–1303
14. Routaboul, J. M., Kerhoas, L., Debeaujon, I., Pourcel, L., Caboche, M., Einhorn, J., and Lepiniec, L. (2006) Planta 224, 96–107
15. Williams, C. A. (2006) in Flavonoids: Chemistry, Biochemistry and Applications (Anderson, O. M., and Markham, K. R., eds) pp. 749–856, Taylor & Francis/CRC, Boca Raton, FL
16. Shibata, D., Saito, K., and Ohta, H. (2007) Plant Physiol. 143, 1437–1445
17. Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005) Plant Physiol. 138, 1083–1096
18. Bowles, D. (2002) Biochem. Soc. Trans. 30, 301–306
19. Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., and Cornish, E. C. (1994) Plant J. 5, 81–92
20. Kroon, J., Souer, E., de Graaff, A., Xue, Y., Mol, J., and Koes, R. (1994) Plant J. 5, 69–80
21. Rhee, Y. S., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L. A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D. C., Wu, Y., Xu, I., Yoo, D., Yoon, J., and Zhang, P. (2003) Nucleic Acids Res. 31, 224–228
22. Craigon, D. I., James, N., Oykere, J., Higgins, J., Jotham, J., and May, S. (2004) Nucleic Acids Res. 32, D575–577
23. Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., Shibata, D., Saito, K., and Ohta, H. (2003) Nucleic Acids Res., 35, D863–D869
24. Vevekens, D., Van Montagu, M., and Van Lijsebettens, M. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 5536–5540
25. Barber, G. A., and Behrmann, E. J. (1991) Arch. Biochem. Biophys. 288, 239–242
26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning a Laboratory Manual Second Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
27. Page, R. (1996) Comput. Appl. Biosci. 12, 357–358
28. Yonekura-Sakakibara, K., Nishiyama, Y., Yamaya, T., and Sakakibara, H. (2004) Plant Physiol. 134, 1654–1661
29. Clough, S. J., and Bent, A. F. (1998) Plant J. 16, 735–743
30. Seki, M., Carninci, P., Nishiyama, Y., Hayashizaki, Y., and Shinozaki, K. (1998) Plant J. 15, 707–720
31. Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Yagishita, A., and Shinozaki, K. (2002) Science 296, 141–145
32. Pelletier, M. K., and Shirley, B. W. (1996) Plant Physiol. 111, 339–345
33. Pelletier, M. K., Murrell, J. R., and Shirley, B. W. (1997) Plant Physiol. 113, 1437–1445
34. Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005) Plant Physiol. 138, 1083–1096
35. Koes, R., Verweij, W., and Quattrocchio, F. (2005) Trends Plant Sci. 10, 236–242
36. Bowles, D. (2002) Biochem. Soc. Trans. 30, 301–306
37. Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., and Cornish, E. C. (1994) Plant J. 5, 81–92
38. Kroon, J., Souer, E., de Graaff, A., Xue, Y., Mol, J., and Koes, R. (1994) Plant J. 5, 69–80
39. Morita, Y., Hoshino, A., Kikuchi, Y., Okuhara, H., Ono, E., Tanaka, Y., Fukui, Y., Saito, N., Nitasaka, E., Noguchi, H., and Iida, S. (2005) Plant J. 42, 353–363
40. Sawada, S., Suzuki, H., Ichimaida, F., Yamaguchi, M. A., Iwashita, T., Fukui, Y., Hemmi, H., Nishino, T., and Nakayama, T. (2005) J. Biol. Chem. 280, 899–906
41. Offen, W., Martinez-Fleites, C., Yang, M., Kiatt-Lim, E., Davis, B. G., Tilting, C. A., Ford, C. M., Bowles, D. J., and Davies, G. J. (2006) EMBO J. 25, 1396–1405
42. Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004) Arch. Biochem. Biophys. 429, 198–203
43. Paquette, S., Moller, B. L., and Bak, S. (2003) Phytochemistry 62, 399–413
44. Stafford, H. A. (1991) Plant Physiol. 96, 680–685
45. Quattrocchio, F., Wing, J. F., Leppen, H., Mol, J., and Koes, R. E. (1993) Plant Cell 5, 1497–1512