Plant Biotechnology 26, 267–274 (2009)

Invited Review

Functional genomics of family 1 glycosyltransferases in *Arabidopsis*

Keiko Yonekura-Sakakibara*

RIKEN Plant Science Center, Yokohama, Kanagawa 230-0045, Japan
* E-mail: keikoys@psc.riken.jp Tel: +81-45-503-9491 Fax: +81-45-503-9489

Received April 16, 2009; accepted May 1, 2009 (Edited by T. Aoki)

Abstract Glycosylation plays an important role for the stabilization, enhancement of water solubility and detoxification of natural products. It also contributes to the highly diverse nature of plant secondary metabolites. Family 1 glycosyltransferases, often referred to as UDP-glycosyltransferases (UGTs), catalyze the transfer of sugar moieties to a wide range of acceptor molecules including plant hormones, phenylpropanoids, flavonoids, betalains, coumarins, terpenoids, steroids and glucosinolates *in planta*. Completion of the sequencing of the *Arabidopsis* genome revealed the presence of 107 UGT genes and also provided novel approaches to determine the functions of UGTs using genomics, transcriptomics and metabolomics. This review describes recent developments in functional genomics efforts on UGTs for plant natural products.

Key words: *Arabidopsis*, flavonoids, glycosyltransferase, secondary metabolism, transcriptome coexpression analysis.

Glycosyltransferases catalyze the transfer of sugar moieties to a wide range of acceptor molecules, and are found in all living organisms (Li et al. 2001). The enzymes can be classified into at least 91 subfamilies based on sequence similarities, catalytic mechanisms and presence of conserved sequence motifs (http://www.cazy.org/fam/acc_GT.html). Family 1 glycosyltransferases are often referred to as UDP-glycosyltransferases (UGTs) and are involved in glycosylation of plant metabolites. UGTs belong to a multigene family that is characterized by the presence of a carboxyl-terminal consensus sequence with 44-amino acid residues (Mackenzie et al. 1997). This motif is found in all plant UGTs and has been called the plant secondary product glycosyltransferase box (Gachon et al. 2005b). To date, UGTs have been reported to utilize plant hormones, including auxins, brassinosteroids, cytokinins (CKs), abscisic acid (ABA) and salicylic acid, phenylpropanoids, flavonoids, betalains, coumarins, terpenoids, steroids and glucosinolates as sugar acceptors and UDP-sugars as sugar donors (Bowles et al. 2006). Glycosylation plays an important role in stabilization, enhancement of water solubility and deactivation/detoxification of natural products, leading to the regulation of metabolic homeostasis, detoxification of xenobiotics and the biosynthesis, storage and transport properties of secondary metabolites. Glycosylation, together with other modification processes such as hydroxylation, methylation and acylation, contributes to the high chemical diversity among plant secondary metabolites. For example, more than 6,000 flavonoid compounds generally occur as the glycosidic form *in planta* (Anderson and Markham 2006).

UGT genes involved in plant secondary metabolism have been identified from various plant species. Identification of UGT functions was initially performed using biochemical or classical genetic approaches, such as enzyme purification (Vogt and Jones 2000) or mutant isolation (Dooner and Nelson 1977). Following the rapid progress in plant molecular biology over the past two decades, molecular-based strategies are now being employed to isolate genes of interest. Differential display methods and homology-based screening of cDNA libraries with homologous genes or conserved domains have been utilized for identification of novel UGT genes (Martin et al. 2001; Ono et al. 2006; Yamazaki et al. 1999). In 2000, sequencing of the *Arabidopsis* genome was completed (The Arabidopsis Genome Initiative 2000), providing novel functional genomics approaches to characterization of UGTs. This review describes recent progress in the study of UGTs in plant secondary metabolism based on genomics, transcriptomics and metabolomics.

Abbreviations: ABA, abscisic acid; CK, cytokinin; CKX, cytokinin oxidase; iP, N^6-(Δ^2-isopentenyl)adenine; IPT, adenosine phosphate-isopentenyltransferase; LC-MS, liquid chromatography mass spectrometry; *LOG*, *LONELY GUY*; N-GT, *N*-glycosyltransferase; O-GT, *O*-glycosyltransferase; tZ, *trans*-zeatin; UGT, family 1 glycosyltransferase, UDP-glycosyltransferase
This article can be found at http://www.jspcmb.jp/

Functional identification of UGTs based on genomics

Analysis of the *Arabidopsis* genome sequence revealed the presence of 107 UGTs (Harborne and Williams 2000). UGT recombinant proteins of 105 of these genes were used in a comprehensive *in vitro* screen to identify CK UGTs (Hou et al. 2004) and ABA UGTs (Lim et al. 2005) (Table 1). Five UGTs were shown to have CK UGT activity (Hou et al. 2004). Eight UGTs showed UGT activity for (±)-ABA, but only UGT71B6 had enantioselective UGT activity towards the naturally occurring form, (+)-ABA (Lim et al. 2005). Thirty six UGT recombinant proteins were also subjected to screening to identify the UGT genes involved in sinapate metabolism and lignin synthesis (Lim et al. 2001).

The availability of the complete genome sequence also allowed us to undertake a comprehensive search based on sequence similarities and phylogenetic analyses; these provide a well-established index that can be used to infer the function of a given gene. Based on sequence similarity, UGT78D1 and UGT73C6 were chosen as candidates for flavonol 3-*O*-glycosyltransferase and flavonol 7-*O*-glycosyltransferase, respectively (Jones et al. 2003). Flavonol analyses of *ugt78d1* and *ugt73c6* mutants and an enzymatic assay using the recombinant proteins indicated that *UGT78D1* and *UGT73C6* encode flavonol 3-*O*-rhamnosyltransferase and flavonol 7-*O*-glucosyltransferase, respectively (Table 1).

Complete genome sequences are now available for various plant species, including rice (*Oryza sativa*) (International Rice Genome Sequencing Project 2005), grapevine (*Vitis vinifera*) (Jaillon et al. 2007) and *Medicago truncatula* (Medicago Genome sequencing project; http://www.medicago.org/genome/downloads/Mt2/) in addition to *Arabidopsis*. In rice, 212 UGTs have been reported in the Rice GT database (http://ricephylogenomics.ucdavis.edu/cellwalls/gt/tree.php) and 146 UGTs have been predicted by an InterProScan search (ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/README.html). Grapevine and *Medicago truncatula* have been estimated to possess 240 and at least 164 UGT genes, respectively (Achnine et al. 2005; Jaillon et al. 2007). The grapevine genome shows large expansions of several gene families involved in the biosynthesis of aromatic compounds, such as stilbene synthases and terpene synthases, compared to other plant genomes (Jaillon et al. 2007). These observations indicate that comparative genomics of the sequenced plant genomes, together with detailed analyses of target metabolites, will provide an invaluable approach for isolating UGT genes of interest.

Table 1. *Arabidopsis* family1 glycosyltransferases with established or putative functions

UGT	AGI	Function	Evidence[a]	Reference
71B6	At3g21780	having ABA glucosyltransferase activity	B	Priest et al. 2006
72B1	At4g01070	having a xenobiotic glycosyltransferase activity	B, PP	Brazier-Hicks and Edwards 2005
72E1	At3g50740	having monolignol 4-*O*-glucosyltransferase activity	B	Lim et al. 2001
72E2	At5g66690	monolignol 4-*O*-glucosyltransferase	B, PM	Lanot et al. 2006; Lim et al. 2001
72E3	At5g26310	having monolignol 4-*O*-glucosyltransferase activity	B	Lim et al. 2001
73B2	At4g34135	having flavonol 7-*O*-glucosyltransferase activity	B	Kim et al. 2006
73B3	At4g34131	UGT required for pathogen resistance	PP	Langlois-Meurinne et al. 2005
73B5	At2g15480	UGT required for pathogen resistance	PP	Langlois-Meurinne et al. 2005
73C1	At2g36750	having cytokinin *O*-glucosyltransferase activity	B	Hou et al. 2004
73C5	At2g36800	brassinosteroid *O*-glucosyltransferase	B, PM	Poppenberger et al. 2005
		having cytokinin *O*-glucosyltransferase activity	B	Hou et al. 2004
73C6	At2g36790	flavonol 7-*O*-glucosyltransferase	B, PM	Jones et al. 2003
74B1	At1g24100	thiohydroximate *S*-glucosyltransferase	B, PM	Grubb et al. 2004
74C1	At2g31790	putative UGT in aliphatic glucosinolate biosynthesis	B	Gachon et al. 2005a
74F2	At2g43820	anthranilate glucosyltransferase	B, PM	Quiel and Bender 2003
75B1	At1g05560	callose synthase-associated glucosyltransferase	B	Hong et al. 2001
		p-aminobenzoate acylglucosyltransferase	B, PM	Eudes et al. 2008
75C1	At4g14090	anthocyanin 5-*O*-glucosyltransferase	PM	Tohge et al. 2005
76C1	At5g05870	having cytokinin *N*-glucosyltransferase activity	B	Hou et al. 2004
76C2	At5g05860	having cytokinin *N*-glucosyltransferase activity	B	Hou et al. 2004
78D1	At1g30530	flavonol 3-*O*-rhamnosyltransferase	B, PM	Jones et al. 2003
78D2	At5g17050	flavonoid 3-*O*-glucosyltransferase	B, PM	Tohge et al. 2005
78D3	At5g17030	flavonol 3-*O*-arabinosyltransferase	B, PM	Yonekura-Sakakibara et al. 2008
84A1	At4g15480	having hydroxycinnamate glucosyltransferase activity	B	Lim et al. 2001
84A2	At3g21560	sinapic acid glucosyltransferase	B, PM	Lim et al. 2001; Sinlapadech et al. 2007
84A3	At4g15490	having hydroxycinnamate glucosyltransferase activity	B	Lim et al. 2001
84B1	At2g23260	having IAA glucosyltransferase activity	B	Jackson et al. 2001
85A1	At1g22400	having cytokinin *O*-glucosyltransferase activity	B	Hou et al. 2004
89C1	At1g06000	flavonol 7-*O*-rhamnosyltransferase	B, PM	Yonekura-Sakakibara et al. 2007

[a]B, biochemical assays; PM, metabolite profiling *in planta*; PP, plant phenotype.

Functional identification of UGTs based on transcriptomics and/or metabolomics

Sequencing of the *Arabidopsis* genome stimulated the rapid and striking development of bioresources such as DNA resources and plant materials, web-based resources for the efficient utilization of the huge amount of biological data, and the development of so-called bioinformatics (Rhee et al. 2006). Transcriptome resources such as the gene expression and gene expression correlation databases obtained using DNA microarrays, have proved of value for selecting candidate genes. Various tools have been developed for secondary analysis of the public databases of transcriptome data

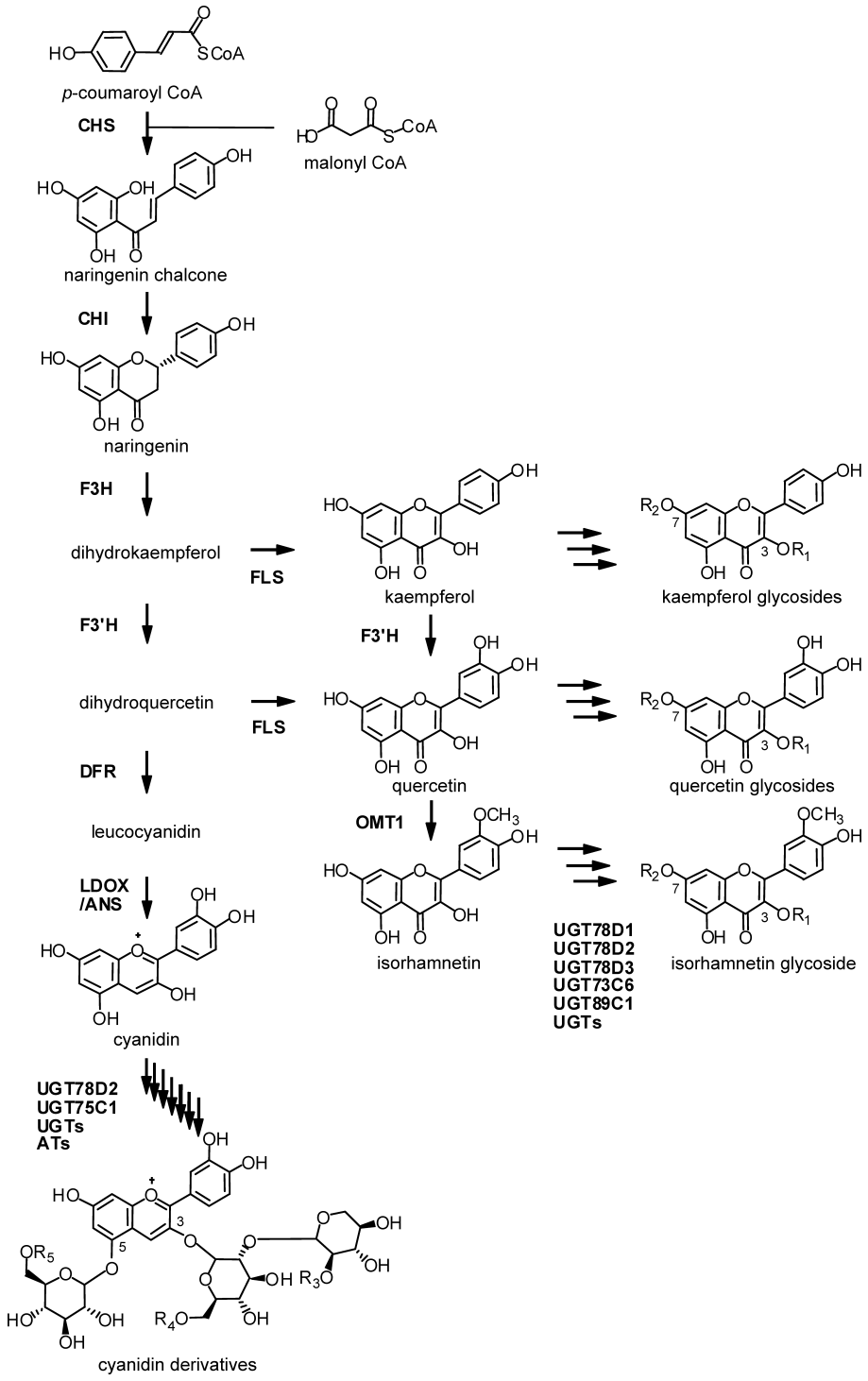

Figure 1. Flavonoid biosynthetic pathways in *Arabidopsis*. ANS, anthocyanidin synthase; AT, acyltransferase; CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydroflavonol reductase; FLS, flavonol synthase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; LDOX, leucoanthocyanidin dioxygenase; OMT1, flavonol 3′-*O*-methyltransferase; UGT, family 1 glycosyltransferase.

derived from microarray databases: Genevestigator (Zimmermann et al. 2004), CSB.DB (Steinhauser et al. 2004), Arabidopsis Co-expression Tool (Manfield et al. 2006), BAR (Toufighi et al. 2005), ATTED-II (Obayashi et al. 2007) and Cluster Cutting (Goda et al. 2008).

Transcriptome-based approaches, including transcriptome coexpression analysis (Saito et al. 2008), are powerful means for identifying genes involved in secondary metabolism. Genes operating within a particular metabolic pathway are often regulated by the same regulatory factors. This enables their cooperative expression in specific organs and/or under defined environmental conditions. Transcriptome coexpression analyses are based on the assumption that coordinately regulated genes function in the same metabolic pathway.

Flavonoids, including anthocyanins, flavonols and flavones, are among the most intensely studied secondary metabolites. The regulatory and biosynthetic pathways involved in flavonoid metabolism have been characterized in a variety of plant species and the corresponding genes have been isolated from various plants (Figure 1) (Anderson and Markham 2006; Tanaka and Fillipa 2006). Novel flavonoid UGTs were identified by a transcriptome coexpression analysis based on a public coexpression database, ATTED-II, using known flavonoid biosynthetic genes and all *Arabidopsis UGTs*. Of the 107 candidates, five *UGT*s correlated with known flavonoid biosynthetic genes, with the highest correlation being obtained for *UGT89C1* (Figure 2) (Yonekura-Sakakibara et al. 2007). In *ugt89c1* knockout mutants, no C-7 rhamnosylated flavonols were detected. Recombinant UGT89C1 protein showed flavonol 7-*O*-rhamnosyltransferase activity. These data demonstrate that *UGT89C1* encodes a flavonol 7-*O*-rhamnosyltransferase (Table 1) (Yonekura-Sakakibara et al. 2007). Of the remaining four candidates, two were known flavonoid UGTs, flavonoid 3-*O*-glucosyltransferase and anthocyanin 5-*O*-glucosyltransferase (Tohge et al. 2005), illustrating the power of transcriptome coexpression analysis to identify genes with specified metabolic properties. A similar strategy was also successful for identifying genes of relevance to the glucosinolate biosynthetic pathway. Hierarchical clustering of gene expression profiles revealed that *UGT74B1* and *UGT74C1* correlated with aromatic- and aliphatic-glucosinolate biosynthetic genes, respectively (Table 1) (Gachon et al. 2005a). A biochemical analysis using a recombinant protein and a glucosinolate analysis of *ugt74b1* loss-of-function mutants demonstrated that UGT74B1 is a UDP-glucose: thiohydroximate *S*-transferase in the aromatic-glucosinolate biosynthetic pathway (Grubb et al. 2004).

An integrated approach combining metabolomic and transcriptomic analyses has also been utilized for identification of gene function. Comprehensive analyses

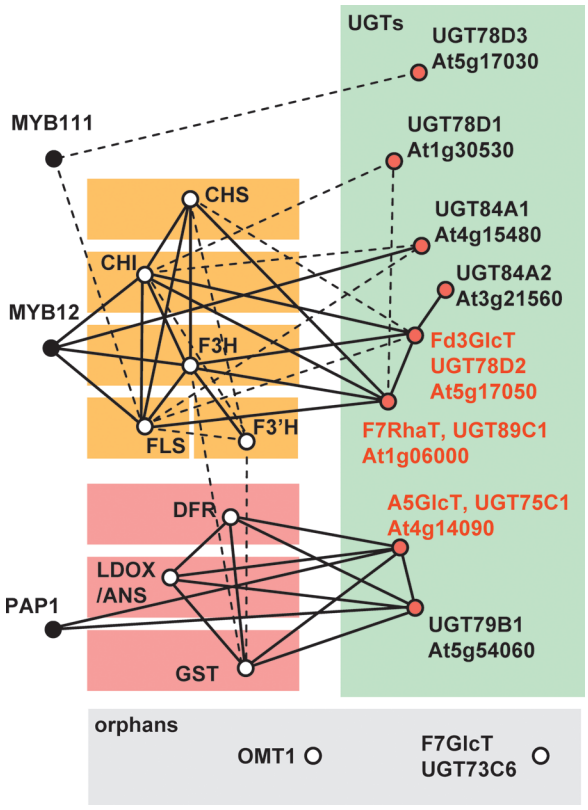

Figure 2. Correlation between genes of flavonoid biosynthetic pathways and UGTs. Coexpression analyses were carried out as described in Yonekura-Sakakibara et al. (2007). Black, white and red circles indicate genes encoding transcription factors, biosynthetic enzymes and UGTs, respectively. Positive correlations (correlation coefficient, $r > 0.6$ in all datasets, ver. 1 and $r > 0.525$ in all datasets, ver. 3) are indicated by solid connecting and dotted lines, respectively. Genes in the orange and pink backgrounds belong to general flavonoid and anthocyanin pathways, respectively. A5GlcT, anthocyanin 5-*O*-glucosyltransferase; ANS, anthocyanidin synthase; CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydroflavonol reductase; FLS, flavonol synthase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F7RhaT, flavonol 7-*O*-rhamnosyltransferase; Fd3GlcT, flavonoid 3-*O*-glucosyltransferase; GST, glutathione *S*-transferase; LDOX, leucoanthocyanidin dioxygenase; OMT1, flavonol 3′-*O*-methyltransferase.

of the metabolites and transcripts in *Arabidopsis* overexpressing *PAP1*, a MYB transcription factor for anthocyanin synthesis, led to the identification of novel flavonoid UGT genes (Tohge et al. 2005). Metabolome analyses by liquid chromatography mass spectrometry (LC-MS) and Fourier transform-MS revealed specific accumulation of cyanidin derivatives in *PAP1* overexpressing plants. Microarray analysis revealed 38 genes induced by *PAP1* overexpression including those encoding UGT, acyltransferase and glutathione *S*-transferase genes as well as known anthocyanin biosynthetic genes. Genes encoding flavonoid 3-*O*-glucosyltransferase (UGT78D2) and anthocyanin 5-*O*-glucosyltranferase (UGT75C1) were identified by flavonoid analysis of T-DNA inserted mutants and by

biochemical analysis using recombinant proteins (Table 1).

Recently, a comprehensive approach was used to investigate flavonoid metabolism (Yonekura-Sakakibara et al. 2008). The peaks derived flavonoids were estimated by a comparison of flavonol LC-MS patterns in wild type plants and in flavonoid deficient mutant, *tt4*, plants. The flavonol structures responsible for the peaks were deduced by MS/MS analysis and LC-MS profiling of other mutants lacking flavonoid biosynthetic genes such as flavonoid 3′-hydroxylase and flavonoid glycosyltransferases. As a result, 32 flavonol derivatives were detected in *Arabidopsis*. By using a transcriptome coexpression network analysis with the 24 known flavonoid biosynthetic genes as guides, candidate genes involved in the flavonoid pathway were delimited from the whole set of *Arabidopsis* genes. From these candidates, a gene encoding a novel flavonoid arabinosyltransferase (UGT78D3) was identified using *Arabidopsis* knockout mutants and *in vitro* enzymatic assays (Table 1). An UDP-rhamnose synthase gene, *RHM1*, was also highly correlated with flavonoid pathway genes. The flavonol profiles of the knockout mutants of *RHM1* showed a decrease of flavonol 3-*O*-rhamnosides and an increase of flavonol 3-*O*-glucosides. These results imply that UDP-rhamnose produced by RHM1 is used for flavonol rhamnosylation (Yonekura-Sakakibara et al. 2008).

Pitfalls of transcriptome coexpression analyses

Transcriptome coexpression analyses are based on the assumptions that genes in the same pathway are coordinately regulated and show similar expression profiles. However, as described below, these assumptions can be misleading under certain circumstances and users of these analyses need to be aware of the potential pitfalls.

One possible problem is that correlation coefficients may be underestimated when the metabolites of interest accumulate in specific regions, or in limited cells, or at lower concentrations compared to other metabolites in the same pathway. For example, trace amounts of flavonol 3-*O*-arabinosides are only detected in flowers whereas other flavonol 3-*O*-glycosides (3-*O*-glucosides and 3-*O*-rhamnosides) are widely distributed throughout the whole plant. By transcriptome coexpression analysis, flavonol biosynthetic genes accounted for 1.4% (2/142) of the genes showing a correlation coefficient greater than 0.4 with flavonol 3-*O*-arabinosyltransferase (UGT78D3) and 16% (13/80) with flavonoid 3-*O*-glucosyltranseferase (UGT78D2). Based on Mutual Rank by ATTED-II (Obayashi et al. 2009), flavonol biosynthetic genes accounted for 3.2% (1/31) and 19.2% (10/52), respectively, of the top 50 genes coexpressed with *UGT78D3* and *UGT78D2*.

Table 2. Crosshybridized pairs of *Arabidopsis* family1 glycosyltransferases

UGT	AGI		UGT	AGI
71D1	At2g29730	–	71D2	At2g29710
72D1	At2g18570	–	72D2P	At2g18560
73B2	At4g34135	–	73B3	At4g34131
73C3	At2g36780	–	73C4	At2g36770
73C5	At2g36800	–	73C6	At2g36790
76E8P	At5g37950	–	76E9	At5g38010
79B2	At4g27560	–	79B3	At4g27570

Some enzymes are involved in two or more pathways. These enzymes may be regulated by a common set of transcription factors in one pathway but separately regulated by different sets of transcription factors in others. In such a case, correlations with genes in the different pathways may be significantly lower. For example, OMT1 (At5g54160) is a caffeic acid *O*-methyltransferase/ flavonol 3′-*O*-methyltransferase that is active in both lignin and flavonol pathways, however, the correlation coefficients of *OMT1* for genes in the lignin pathway are much higher than for those in the flavonol pathway.

Cross hybridization is a critical issue in transcriptome coexpression analyses. Analysis of 117 UGTs, including 10 pseudogenes, indicated that 81 UGTs (69.2%) were located in tandem with or close to other UGTs. Of these genes, 7 sets (14 genes including 2 pseudogenes) cross hybridized to other UGT genes (Table 2). For example, *UGT73C6* (At2g36790) cross hybridized with the adjacent *UGT73C5* (At2g36800). UGT73C6 is a flavonol 7-*O*-glucosyltransferase (Jones et al. 2003) while UGT73C5 is a brassinosteroid glucosyltransferase (Poppenberger et al. 2005). None of the flavonoid biosynthetic genes had correlation coefficients greater than 0.4 against UGT73C6 probably due to cross hybridization with these two genes. Interestingly, *UGT73C6* (At2g36790) and *UGT76E12* (At3g46660) are unusual among UGTs in having potential natural antisense genes, At2g36792 and At3g46658, respectively. The antisense gene might cause the lack of a significant correlation between *UGT73C6* and flavonoid biosynthetic genes. By contrast, genes involved in metabolism frequently possess a redundant function. In the anthocyanin biosynthetic pathway, At1g03495 and At1g03940, which encode acyltransferases and are involved in the same acylation step of anthocyanin, cross hybridized with each other (Luo et al. 2007).

Application of transcriptome coexpression analysis to pathways with complex regulation

Transcriptome coexpression analysis has been

established as a powerful tool for the functional identification of genes involved in secondary metabolism. However, is it possible to use this approach to characterize other metabolic pathways, such as plant hormone biosynthesis, that have a complex pattern of regulation? The CK pathway was investigated to determine whether it was feasible to use this approach for genes with complex regulation.

The major CKs in plants are adenine derivatives that carry an isoprene-derived side chain at the N^6 terminus (Sakakibara 2006). As shown in Figure 3, the prenyl moiety of dimethylallyl diphosphate is conjugated with adenosine phosphates (AMP, ADP or ATP) by adenosine phosphate-isopentenyltransferases (IPTs), resulting in N^6-(Δ^2-isopentenyl)adenosine 5′-mono-, di- and/or triphosphate (Kakimoto 2001; Takei et al. 2001a). Hydroxylation of these N^6-(Δ^2-isopentenyl)adenine (iP)-nucleotide CKs by CYP735A leads to the formation of trans-zeatin (tZ)-nucleotides (Takei et al. 2004). CK oxidases (CKXs) are involved in the degradation of active form CKs (iP and tZ). CK *O*-glycosyltransferase (O-GT)/ *N*-glycosyltransferase (N-GT) is also involved in deactivation of active form CKs (Mok and Mok 2001).

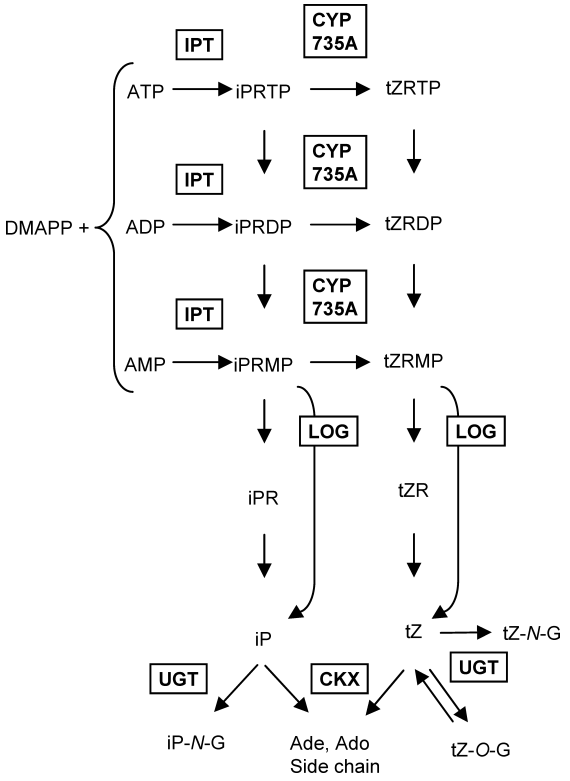

Figure 3. The biosynthetic pathways of major cytokinin species in *Arabidopsis*. Ade, adenine; Ado, adenosine; CKX, cytokinin oxidases; CYP, cytochrome P450 monooxygenase; DMAPP, dimethylallyl diphosphate; iP, N^6-(Δ^2-isopentenyl)adenine; iP-*N*-G, iP *N*-glucoside; iPR, iP riboside; iPRDP, iPR 5′-diphosphate; iPRMP, iPR 5′-monophosphate; iPRTP, iPR 5′-triphosphate; IPT, adenosine phosphate-isopentenyltransferase; LOG, cytokinin riboside 5′-monophosphate phosphoribohydrolase; tZ, trans-zeatin; tZ-*N*-G, tZ *N*-glucoside; tZR, tZ riboside; tZRDP, tZR 5′-diphosphate; tZ-*O*-G, tZ *O*-glucoside; tZRMP, tZR 5′-monophosphate; tZRTP, tZR 5′-triphosphate.

A number of genes have been reported to be involved in CK metabolism: 7 *IPTs* (Kakimoto 2001; Takei et al. 2001a), 2 *CYP735As* (Takei et al. 2004), 7 *CKXs* (Schmulling et al. 2003) and 5 *UGTs* (Hou et al. 2004) that have CK O-GT/N-GT activity. However, no significant correlations were observed by transcriptome coexpresssion analysis (all datasets, ver. 3, $r>0.46$) using ATTED-II with any of these 21 genes.

IPT and CYP735A enzymes catalyze adjacent enzymatic steps in the CK biosynthesis pathway. However, iP-type CKs, the enzymatic products of *IPT*s, and tZ-type CKs, the products of CYP735As, are supposed to have different physiological roles *in planta*. Supporting this idea, the manner of regulation of these enzymes is different. Some *IPT*s are down-regulated by CK treatment (Miyawaki et al. 2004), whereas *CYP735As* are up-regulated (Takei et al. 2004). Moreover, *IPT*s respond to auxin treatment in a completely opposite fashion to *CYP735As* (Miyawaki et al. 2004; Takei et al. 2004). Recently, the gene encoding a direct CK activation enzyme, *LONELY GUY* (*LOG*), was identified in rice (Kurakawa et al. 2007). A Blast search suggested the occurrence of 9 *LOG* orthologs in *Arabidopsis*. The metabolic step catalyzed by UGTs follows that catalyzed by LOGs. However, very low correlations were obtained between *Arabidopsis* LOG candidates and known UGTs having CK O-GT/N-GT activity ($r<0.27$, all datasets, ver. 3). Compared to those for flavonoid metabolism ($r<0.914$, all datasets, ver.3), the correlation coefficients between LOGs and UGTs were quite low, suggesting no correlation of gene expression between these two gene groups. In the CK biosynthetic pathway, CKs, including intermediates (riboside, ribotide, free and glycosides of iP-type and tZ-type CKs) have their own physiological roles and are regulated strictly in terms of accumulation level and molecular species. Furthermore, CKs have been found in the xylem sap of several plant species (Kudoyarova et al. 2007; Kuroha et al. 2002; Morris et al. 2001; Takei et al. 2001b; Yong et al. 2000); the CK species present in the phloem sap and xylem sap differ (Hirose et al. 2008). These data suggest that CK metabolism, including biosynthesis, degradation and transportation, is complex. Therefore, it will be difficult to determine the whole pathway using the public transcriptome databases based on microarray experiments using different conditions and developmental stages.

Can other transcriptomic strategies be applied to a system involving such a finely-tuned metabolism? To eliminate problems derived from the transportation of compounds or biased distribution of target compounds between plant tissues, specific or single cell separation techniques may be helpful. In medicinal and aromatic plants, transcriptomics-based approaches with trichome-

specific transcripts have proved successful in the identification of novel gene functions as some flavonoids and terpenes predominantly accumulate in glandular trichomes (Nagel et al. 2008; Rios-Estepa et al. 2008; Wang et al. 2008).

An Integration of transcriptomics and metabolomics in distinct cells and organs together with the anatomical techniques such as laser microdissection, specific cell separation indexed by GFP and single cell metabolomics may be also useful in such distinctly regulated pathways.

Conclusions

The completion of genome sequencing and the subsequent development of other "-omics", such as transcriptomics, proteomics, metabolomics and bioinformatics, have advanced the development of functional genomics approaches such as transcriptome coexpression analysis, especially in *Arabidopsis*, the first genome-sequenced plant. However, the physiological roles of most UGTs still remain uncertain even in *Arabidopsis*. Of 107 *Arabidopsis* UGT genes, the functions of only about 10% have been established *in planta* (Table 1). Biochemical analyses with recombinant UGT proteins can also be used to suggest functions (Table 1). However, in general, UGT functions *in planta* remain to be established because of the broad substrate specificity of these enzymes for sugar acceptors. Functional genomics approaches based on transcriptomics and metabolomics will provide robust evidence for their functions *in planta*. In addition, the development of new types of analytical equipment, such as next generation ultra-high speed sequencers, will permit comparative genomics-based approaches. The development of anatomical techniques should be also effective in the research into tightly regulated pathways.

Acknowledgements

I would like to thank Prof. K. Saito and Dr. H. Sakakibara for critical reading of the manuscript.

References

Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, et al. (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J* 41: 875–887

Anderson ØM, Markham KR (eds) (2006) *Flavonoids: chemistry, biochemistry, and applications.* CRC Press, Boca Raton

Bowles D, Lim EK, Poppenberger B, Vaistij FE (2006) Glycosyltransferases of lipophilic small molecules. *Annu Rev Plant Biol* 57: 567–597

Dooner HK, Nelson OE (1977) Controlling element-induced alterations in UDPglucose: flavonoid glucosyltransferase, the enzyme specified by the *bronze* locus in maize. *Proc Natl Acad Sci USA* 74: 5623–5627

Gachon CM, Langlois-Meurinne M, Henry Y, Saindrenan P (2005a) Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Mol Biol* 58: 229–245

Gachon CM, Langlois-Meurinne M, Saindrenan P (2005b) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci* 10: 542–549

Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, et al. (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J* 55: 526–542

Grubb CD, Zipp BJ, Ludwig-Muller J, Masuno MN, Molinski TF, et al. (2004) Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J* 40: 893–908

Harborne JB, Williams CA (2000) Advances in flavonoid research since 1992. *Phytochemistry* 55: 481–504

Hirose N, Takei K, Kuroha T, Kamada-Nobusada T, Hayashi H, et al. (2008) Regulation of cytokinin biosynthesis, compartmentalization and translocation. *J Exp Bot* 59: 75–83

Hou B, Lim EK, Higgins GS, Bowles DJ (2004) *N*-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. *J Biol Chem* 279: 47822–47832

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436: 793–800

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467

Jones P, Messner B, Nakajima J, Schaffner AR, Saito K (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. *J Biol Chem* 278: 43910–43918

Kakimoto T (2001) Identification of plant cytokinin biosynthetic enzymes as dimethylallyl diphosphate: ATP/ADP isopentenyltransferases. *Plant Cell Physiol* 42: 677–685

Kudoyarova GR, Vysotskaya LB, Cherkozyanova A, Dodd IC (2007) Effect of partial rootzone drying on the concentration of zeatin-type cytokinins in tomato (*Solanum lycopersicum* L.) xylem sap and leaves. *J Exp Bot* 58: 161–168

Kurakawa T, Ueda N, Maekawa M, Kobayashi K, Kojima M, et al. (2007) Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* 445: 652–655

Kuroha T, Kato H, Asami T, Yoshida S, Kamada H, et al. (2002) A trans-zeatin riboside in root xylem sap negatively regulates adventitious root formation on cucumber hypocotyls. *J Exp Bot* 53: 2193–2200

Li Y, Baldauf S, Lim EK, Bowles DJ (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J Biol Chem* 276: 4338–4343

Lim EK, Doucet CJ, Hou B, Jackson RG, Abrams SR, et al. (2005) Resolution of (+)-abscisic acid using an Arabidopsis glycosyltransferase. *Tetrahedron: Asym* 16: 143–147

Lim EK, Li Y, Parr A, Jackson R, Ashford DA, et al. (2001) Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in *Arabidopsis*. *J Biol Chem* 276: 4344–4349

Luo J, Nishiyama Y, Fuell C, Taguchi G, Elliott K, et al. (2007) Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *Plant J* 50: 678–695

Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, et al. (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* 7: 255–269

Manfield IW, Jen CH, Pinney JW, Michalopoulos I, Bradford JR, et al. (2006) *Arabidopsis* Co-expression Tool (ACT): web server tools for microarray-based gene expression analysis. *Nucl Acids Res* 34: W504–509

Martin RC, Mok MC, Habben JE, Mok DW (2001) A maize cytokinin gene encoding an *O*-glucosyltransferase specific to *cis*-zeatin. *Proc Natl Acad Sci USA* 98: 5922–5926

Miyawaki K, Matsumoto-Kitano M, Kakimoto T (2004) Expression of cytokinin biosynthetic isopentenyltransferase genes in *Arabidopsis*: tissue specificity and regulation by auxin, cytokinin, and nitrate. *Plant J* 37: 128–138

Mok DW, Mok MC (2001) Cytokinin Metabolism and Action. *Annu Rev Plant Physiol Plant Mol Biol* 52: 89–118

Morris SE, Turnbull CG, Murfet IC, Beveridge CA (2001) Mutational analysis of branching in pea. Evidence that Rms1 and Rms5 regulate the same novel signal. *Plant Physiol* 126: 1205–1213

Nagel J, Culley LK, Lu Y, Liu E, Matthews PD, et al. (2008) EST analysis of hop glandular trichomes identifies an O-methyltransferase that catalyzes the biosynthesis of xanthohumol. *Plant Cell* 20: 186–200

Obayashi T, Hayashi S, Saeki M, Ohta H, Kinoshita K (2009) ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucl Acids Res* 37: D987–D991

Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, et al. (2007) ATTED-II: a database of co-expressed genes and *cis* elements for identifying co-regulated gene groups in *Arabidopsis*. *Nucl Acids Res* 35: D863–D869

Ono E, Fukuchi-Mizutani M, Nakamura N, Fukui Y, Yonekura-Sakakibara K, et al. (2006) Yellow flowers generated by expression of the aurone biosynthetic pathway. *Proc Natl Acad Sci USA* 103: 11075–11080

Poppenberger B, Fujioka S, Soeno K, George GL, Vaistij FE, et al. (2005) The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids. *Proc Natl Acad Sci USA* 102: 15253–15258

Rhee SY, Dickerson J, Xu D (2006) Bioinformatics and its applications in plant biology. *Annu Rev Plant Biol* 57: 335–360

Rios-Estepa R, Turner GW, Lee JM, Croteau RB, Lange BM (2008) A systems biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint. *Proc Natl Acad Sci USA* 105: 2818–2823

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics—'majority report by precogs'. *Trends Plant Sci* 13: 36–43

Sakakibara H (2006) Cytokinins: activity, biosynthesis, and translocation. *Annu Rev Plant Biol* 57: 431–449

Schmulling T, Werner T, Riefler M, Krupkova E, Bartrina y Manns I (2003) Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, *Arabidopsis* and other species. *J Plant Res* 116: 241–252

Steinhauser D, Usadel B, Luedemann A, Thimm O, Kopka J (2004) CSB.DB: a comprehensive systems-biology database. *Bioinformatics* 20: 3647–3651

Takei K, Sakakibara H, Sugiyama T (2001a) Identification of genes encoding adenylate isopentenyltransferase, a cytokinin biosynthesis enzyme, in *Arabidopsis thaliana*. *J Biol Chem* 276: 26405–26410

Takei K, Sakakibara H, Taniguchi M, Sugiyama T (2001b) Nitrogen-dependent accumulation of cytokinins in root and the translocation to leaf: implication of cytokinin species that induces gene expression of maize response regulator. *Plant Cell Physiol* 42: 85–93

Takei K, Yamaya T, Sakakibara H (2004) *Arabidopsis CYP735A1* and *CYP735A2* encode cytokinin hydroxylases that catalyze the biosynthesis of *trans*-Zeatin. *J Biol Chem* 279: 41866–41872

Tanaka Y, Fillipa B (2006) Flower color, In Ainsworth C (ed) *Flowering and Its Manipulation.* Blackwell Publishers, London, pp 201–239

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, 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

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *Plant J* 43: 153–163

Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci* 5: 380–386

Wang G, Tian L, Aziz N, Broun P, Dai X, et al. (2008) Terpene biosynthesis in glandular trichomes of hop. *Plant Physiol* 148: 1254–1266

Yamazaki M, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, et al. (1999) Molecular cloning and biochemical characterization of a novel anthocyanin 5-*O*-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. *J Biol Chem* 274: 7405–7411

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, et al. (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *Plant Cell* 20: 2160–2176

Yonekura-Sakakibara K, Tohge T, Niida R, Saito K (2007) Identification of a flavonol 7-*O*-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. *J Biol Chem* 282: 14932–14941

Yong JW, Wong SC, Letham DS, Hocart CH, Farquhar GD (2000) Effects of elevated [CO(2)] and nitrogen nutrition on cytokinins in the xylem sap and leaves of cotton. *Plant Physiol* 124: 767–780

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol* 136: 2621–2632