Exploiting Natural Variation of Secondary Metabolism Identifies a Gene Controlling the Glycosylation Diversity of Dihydroxybenzoic Acids in *Arabidopsis thaliana*

Xu Li,^{*,1} Elisabeth Svedin,[†] Huaping Mo,^{*,§} Susanna Atwell,^{**,} Brian P. Dilkes,^{†,2} and Clint Chapple^{*,2}

^{*}Department of Biochemistry, [†]Department of Horticulture and Landscape Architecture, [‡]Purdue Interdepartmental NMR Facility,

[§]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana 47907, and

^{**}Department of Plant Sciences, University of California, Davis, California 95616

ABSTRACT Plant secondary metabolism is an active research area because of the unique and important roles the specialized metabolites have in the interaction of plants with their biotic and abiotic environment, the diversity and complexity of the compounds and their importance to human medicine. Thousands of natural accessions of *Arabidopsis thaliana* characterized with increasing genomic precision are available, providing new opportunities to explore the biochemical and genetic mechanisms affecting variation in secondary metabolism within this model species. In this study, we focused on four aromatic metabolites that were differentially accumulated among 96 *Arabidopsis* natural accessions as revealed by leaf metabolic profiling. Using UV, mass spectrometry, and NMR data, we identified these four compounds as different dihydroxybenzoic acid (DHBA) glycosides, namely 2,5-dihydroxybenzoic acid (gentisic acid) 5-*O*- β -D-glucoside, 2,3-dihydroxybenzoic acid 3-*O*- β -D-glucoside, 2,5-dihydroxybenzoic acid 5-*O*- β -D-xyloside, and 2,3-dihydroxybenzoic acid 3-*O*- β -D-xyloside. Quantitative trait locus (QTL) mapping using recombinant inbred lines generated from C24 and Col-0 revealed a major-effect QTL controlling the relative proportion of xylosides vs. glucosides. Association mapping identified markers linked to a gene encoding a UDP glycosyltransferase gene. Analysis of Transfer DNA (T-DNA) knockout lines verified that this gene is required for DHBA xylosylation *in planta* and recombinant protein was able to xylosylate DHBA *in vitro*. This study demonstrates that exploiting natural variation of secondary metabolism is a powerful approach for gene function discovery.

PLANTS produce >200,000 diverse low-molecular weight compounds, known as secondary or specialized metabolites (Dixon and Strack 2003; Yonekura-Sakakibara and Saito 2009). These metabolites are not essential to short-term survival but play important roles in many aspects of plant life, including growth regulation, defense against herbivores, UV protection, and other adaptations to the environment (Hartmann 2007). Apart from their roles in plant

adaptation, plant specialized metabolites are rich sources for industrial and medicinal materials such as dyes, flavors, and pharmaceuticals (Balandrin *et al.* 1985).

Studies of *Arabidopsis thaliana* have greatly advanced our understanding of the biochemical pathways and gene networks involved in a variety of specialized metabolism. For example, genetic analysis of artificially induced *Arabidopsis* mutants has led to the discovery of genes responsible for the biosynthesis of flavonoids, sinapate esters, and lignin (Shirley *et al.* 1995; Ruegger and Chapple 2001). Exploiting natural variation between different accessions in the accumulation of glucosinolates and terpenoids has led to the identification of genes involved in their biosynthesis and alleles regulating their variation that are under strong selection (Kliebenstein *et al.* 2001a,b; Kroymann *et al.* 2001; Chen *et al.* 2003; Tholl *et al.* 2005). Despite this progress, there are many more unknown than known compounds, even in this model species, and the pathways that generate them are still waiting to be

Copyright © 2014 by the Genetics Society of America

doi: 10.1534/genetics.114.168690

Manuscript received July 22, 2014; accepted for publication August 21, 2014; published Early Online August 29, 2014.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.168690/-DC1>.

¹Present address: Plants for Human Health Institute, North Carolina State University, Kannapolis, NC 28081.

²Corresponding authors: Department of Biochemistry, Purdue University, 175 South University St., West Lafayette, IN 47907. E-mail: chapple@purdue.edu; Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana 47907. E-mail: bdilkes@purdue.edu.

discovered (D'Auria and Gershenzon 2005). In recent years, mapping on genome-wide polymorphism information (Nordborg *et al.* 2005; Atwell *et al.* 2010; Chan *et al.* 2010; Li *et al.* 2010b; Cao *et al.* 2011) and experimental genetic material exploiting natural variation such as recombinant inbred lines (RILs) and near isogenic lines (NILs) (Torjek *et al.* 2006, 2008) has facilitated the discovery of the alleles responsible for metabolic variation and thus the genes responsible for specialized metabolism. This approach provides an opportunity to discover the molecular mechanisms underlying evolutionarily relevant changes in metabolism and link metabolites to adaptive differences conditioned by metabolism (Keurentjes *et al.* 2006).

To expand our knowledge of aromatic secondary metabolites, we surveyed soluble UV-absorbing metabolites in leaf extracts of a set of 96 *Arabidopsis* accessions, and found substantial variation in their accumulation. A previous study of one of those metabolic variations identified a gene involved in sinapate ester biosynthesis (Li *et al.* 2010a). Here we report the identification of four dihydroxybenzoic acid (DHBA) glycosides that are differentially accumulated among accessions, and the discovery of a gene encoding a UDP-xylose:DHBA xylosyltransferase that underlies this metabolic polymorphism.

Materials and Methods

Plant material and growth conditions

The 96 accessions of *Arabidopsis thaliana* (stock no. CS22660) and the two T-DNA knockout lines (SALK_081110C and SALK_147583C) for At5g03490 were obtained from the Arabidopsis Biological Resource Center (Alonso *et al.* 2003). F₂ plants were generated by crossing Col-0 and C24 and subsequently selfing of F₁. The C24 × Col-0 and Col-0 × C24 RILs and NILs were kindly provided by Rhonda Meyer at Max Planck Institute of Molecular Plant Physiology, Germany. All plants were grown at 22° under a 16-h light/8-h dark photoperiod.

HPLC profiling of leaf soluble metabolites

HPLC sample preparation and analysis were done according to a previously published procedure (Li *et al.* 2010a). Briefly, leaf tissue was harvested from 3-week-old plants and immediately extracted with 50% (v/v) methanol (1 ml per 100 mg tissue) at 65° for 30 min and injected onto a Shim-pack XR-ODS column (3.0 × 75 mm, 2.2 μm) (Shimadzu, Kyoto, Japan) using a binary gradient from 2% acetonitrile in 0.1% formic acid to 35% acetonitrile in 0.1% formic acid at a flow rate of 0.8 ml·min⁻¹. The 2,5-DHBA and 2,3-DHBA were used as standard for quantification of 2,5-DHBA glycosides and 2,3-DHBA glycosides, respectively.

Compound purification and identification

A methanol extract of ~500 g of leaf tissue was concentrated using a rotary evaporator and fractionated on a semi-prep C18 column using a binary gradient of water and acetonitrile. The collected fractions were dried and redissolved in 5 mM

Figure 1 Comparison of leaf metabolite profile of Col-0 and C24. HPLC chromatograms of leaf extract from Col-0 and C24 clearly show the differential accumulation of compounds 1 and 2 (indicated by arrow-heads) between these two accessions.

NH₄AC (pH 9), and then loaded onto a DE52 anion exchange column. The DHBA glycosides were eluted with 7.5 mM HCl. The fraction was freeze dried and redissolved in appropriate solvent for analysis. Base hydrolysis was performed in 1 M NaOH at room temperature overnight. Acid hydrolysis was done by incubating the compounds with 1 M HCl at 80° for 1 hr. Liquid Chromatography–Mass Spectrometry (LC-MS) experiments were performed with an Agilent accurate mass time-of-flight instrument in negative electrospray ionization mode using standard settings. For NMR analysis, 2,5-DHBAG and 2,3-DHBAG were dissolved with CD₃OD:D₂O (9:1, v/v). The 2,5-DHBAX was dissolved in D₂O and 2,3-DHBAX in CD₃OD. ¹H and ¹³C spectra were acquired on a Bruker Avance 500 equipped with a broadband probe operating at room temperature.

Recombinant protein expression

The *UGT89A2* gene was amplified from Col-0 genomic DNA with attB site containing primers: 5'-ggggacaagtgtgtacaaaaagcaggctcatgacggaagtgtattgttg-3' and 5'-ggggaccactttgtacaa gaaagctgggttagacttttcaaattcttgac-3'. The PCR product was cloned into pDONR221 (Invitrogen) and subsequently cloned into pDEST17 (Invitrogen) by standard Gateway cloning technology. The resulting expression vector was transformed into *Escherichia coli* BL21 (DE3) pLysS strain. The expression of the recombinant protein was induced with 0.4 mM IPTG at 16° overnight.

Xylosyltransferase activity assay

Total protein was extracted from *E. coli* with 50 mM MOPS-KOH buffer, pH 7.0. Protein concentration was measured by using a microvolume bicinchoninic acid assay on a NanoDrop spectrophotometer (Desjardins *et al.* 2009). Xylosyltransferase activity assay was carried out in 50 mM MOPS-KOH buffer, pH 7.0 containing 14 mM 2-mercaptoethanol, 5 mM UDP-xylose, and 1 mM 2,5-DHBA. After incubation at 30° for 30 min, the enzyme reaction was stopped by adding one-fifth of the total assay volume of 24% (w/v) trichloroacetic acid. After a brief centrifugation, the supernatant was

Figure 2 Compounds 1 and 2 are resistant to base hydrolysis but labile to acid hydrolysis. The two compounds were subjected to no treatment, base hydrolysis, or acid hydrolysis, and subsequently analyzed by LC-MS as described in *Materials and Methods*. The m/z value of the molecular ion for a compound is shown to the right.

analyzed on reverse-phase HPLC as described above except using a starting gradient of 5% acetonitrile.

QTL mapping

Phenotyping of DHBA glycoside accumulation was done by HPLC on 265 Col $\times$ C24 RILs and 31 NILs. The RIL and NIL populations and genotypes for the RILs and NILs were described previously (Torjek *et al.* 2006, 2008). The proportion of DHBA xyloside to total DHBA in the RIL population was nonnormal. To improve the trait distribution for mapping and eliminate false positive interactions, the traits were log transformed after applying pseudocounts to adjust trait values of zero (Burkart-Waco *et al.* 2012). QTL mapping was performed in R using the package R/QTL (Broman *et al.* 2003). The implementation of Multiple QTL Mapping (MQM) and 1000 permutations were used to calculate experiment-wide multiple test corrections (Arends *et al.* 2010).

Genome-wide association analysis

DHBA glycosides were measured in the 96 accession populations described by Nordborg *et al.* (2005). The 214,051 SNP genotypes derived from genotyping of these 96 accessions on the ATSNPTile1 array were previously described (Atwell *et al.* 2010). Alleles with a minor allele frequency of <10% were

excluded from analysis to reduce observation of spurious associations leaving 177,088 SNPs for the analysis.

The association between each SNP and the phenotype was calculated using the Efficient Mixed-Model Association eXpedited (EMMAX) procedure (Kang *et al.* 2008, 2010), which minimizes computational time for each SNP by estimating variance parameters and genetic relatedness once and then applies that across all association tests. The output of EMMAX includes the beta value (the main effect) and the P -value (significance) for each SNP. Manhattan plots were generated using an R script (Turner 2001) to visualize the EMMAX results. The Manhattan script requires a file that contains four columns: SNP (SNP ID), CHR (chromosome), BP (base pair position), and P (P -value). A custom Perl script was used to convert the EMMAX output into a file that is read by the Manhattan script.

Results

Identification of four DHBA glycosides from *Arabidopsis*

During a previous survey of sinapate ester profiles from 96 *Arabidopsis* accessions (Li *et al.* 2010a) we observed a number of other peaks for several UV-absorbent compounds with

Figure 3 Four DHBA glycosides identified in this study. (A) HPLC chromatograms of leaf extract from C24 and selected Col-0 $\times$ C24 F₂ plants, showing the retention times of the four DHBA glycosides. (B) UV spectra of the four DHBA glycosides. (C) Chemical structures of the four DHBA glycosides.

qualitative and quantitative differences. Two compounds were readily detected in half of the 96 accessions, including C24, but were below the limits of detection in the remainder, including Col-0 (Figure 1). These two compounds elute early from reverse-phase HPLC, indicating they are relatively hydrophilic. Because sugar conjugation is a common hydrophilic modification in plant secondary metabolism, we carried out acid and base hydrolysis to test if these compounds could be sugar esters or glycosides. As shown in Figure 2, these compounds remain intact after sodium hydroxide treatment, suggesting that they are unlikely to be esters. In contrast, treatment with hydrochloric acid at elevated temperature produced a UV-absorbing compound with a longer retention time, consistent with release of an aglycone moiety from a glycoside by acid hydrolysis. Further analysis revealed that the putative aglycones of these two compounds have the same retention time, UV spectrum, and mass as authentic 2,5-dihydroxybenzoic acid (2,5-DHBA;

gentisic acid) and 2,3-dihydroxybenzoic acid (2,3-DHBA), respectively. Further, the mass difference between the hydrolysis product and the intact compound is 162, which is consistent with the loss of a hexose. To further elucidate their structures, we purified the two putative glycosides from C24 leaves by reverse-phase HPLC and anion exchange chromatography. The ¹H and ¹³C NMR spectra of the two purified compounds matched the previously reported NMR spectra of 2,5-dihydroxybenzoic acid 5-*O*- β -D-glucoside (2,5-DHBAG) and 2,3-dihydroxybenzoic acid 3-*O*- β -D-glucoside (2,3-DHBAG), respectively (Figure 3C; [Supporting Information, Figure S1](#)) (Sakushima *et al.* 1995).

We then crossed C24 with Col-0 and analyzed the levels of 2,5-DHBAG and 2,3-DHBAG in 290 F₂ plants (see below for details). Some F₂ plants accumulated two additional compounds (3 and 4), which had similar UV spectrum to 2,5-DHBAG and 2,3-DHBAG, but with delayed retention times (Figure 3, A and B). Like 2,5-DHBAG and 2,3-DHBAG,

Figure 4 Accumulation patterns of DHBA glycosides in Col-0 × C24 F₂ population. The levels of different DHBA glycosides accumulated in 290 individual F₂ plants were plotted against each other. (A) The 2,3-DHBAG vs. 2,5-DHBAG; (B) 2,3-DHBAX vs. 2,5-DHBAX; (C) 2,3-DHBA glycosides (sum of 2,3-DHBAG and 2,3-DHBAX) vs. 2,5-DHBA glycosides (sum of 2,5-DHBAG and 2,5-DHBAX); (D) 2,5-DHBAX vs. 2,5-DHBAG; (E) 2,3-DHBAX vs. 2,3-DHBAG; (F) DHBAX (sum of 2,3-DHBAX and 2,5-DHBAX) vs. DHBAG (sum of 2,3-DHBAG and 2,5-DHBAG); and (G) distribution pattern of the relative abundance value of DHBAX in the F₂ population.

these two compounds were also resistant to base hydrolysis but susceptible to hot acid hydrolysis. LC-MS analysis revealed that acid hydrolysis of compounds 3 and 4 also gave rise to 2,5-DHBA and 2,3-DHBA, respectively. In addition, the mass difference between intact compounds and their hydrolysis products coincided with the loss of a pentose, suggesting that these two compounds are pentosides of 2,5-DHBA and 2,3-DHBA. To identify the compounds, they were purified and subjected to NMR analysis. This analysis confirmed that they are DHBA glycosides and revealed that the sugar moiety is β -D-xylose in both compounds 3 and 4. The ¹H and ¹³C NMR spectra of compounds 3 and 4 agreed with those previously described for 2,5-DHBA 5-O- β -D-xyloside

(2,5-DHBAX) and 2,3-DHBA 3-O- β -D-xyloside (2,3-DHBAX), respectively (Figure 3C; Table S1) (Fayos *et al.* 2006; Bartsch *et al.* 2010). We reexamined Col-0 by more sensitive methods and were able to detect 2,5-DHBAX using fluorescence detection or LC-MS, although the level of this compound was below the detection limit of the UV detector used in our initial HPLC profiling experiments (data not shown).

The accumulation of the four DHBA glycosides has a shared genetic basis

The four DHBA glycosides identified above are chemically related, suggesting that they may share common steps in their biosynthesis. To test this genetically and investigate

Figure 5 QTL analysis of molar percentage of DHBA xylosides from the Col-0 × C24 RIL population. The y-axis corresponds to the negative $\log_{10}$ of the *P*-value as determined by MQM mapping in the R/QTL package. The x-axis indicates the chromosome, and the positions of each marker are indicated by the upward tick marks on the axis line. Only one position, at the top of chromosome 5, exceeded a 1000-permutation-estimated threshold of 0.05.

how the biosynthesis of these four DHBA glycosides are related, we examined their accumulation in a 290-individual Col-0 × C24 F₂ population. The levels of 2,3-DHBAG and 2,5-DHBAG were positively correlated in this population. In some plants, we detected only 2,3-DHBAG but not 2,5-DHBAG and vice versa (Figure 4A) indicating some independence in their accumulation and the detection limits of our analysis. Similarly, the accumulation of 2,3-DHBAX and 2,5-DHBAX was correlated (Figure 4B). In contrast, the levels of 2,5-DHBAX and 2,5-DHBAG can be used to clearly categorize F₂ individuals into three groups: one predominately accumulates 2,5-DHBAX, another predominately accumulates 2,5-DHBAG, and the third group accumulates both at comparable levels (Figure 4D). A similar accumulation pattern was also observed for the glycosides of 2,3-DHBA (Figure 4E). The accumulation of these four DHBA glycosides were also compared by an aglycone group, that is, 2,3-DHBA glycosides (2,3-DHBAG and 2,3-DHBAX) vs. 2,5-DHBA glycosides (2,5-DHBAG and 2,5-DHBAX), or by glycosylation type, DHBA xylosides (2,5-DHBAX and 2,3-DHBAX) as compared to DHBA glucosides (2,5-DHBAG and 2,3-DHBAG) (Figure 4, C and F). These analyses clearly indicate that these four DHBA glycosides share some genetic control in their accumulation. Moreover, the relative abundance of DHBA xylosides (expressed as percentage of total DHBA glycosides) segregates into three clusters with a near 1:2:1 ratio in the Col-0 × C24 F₂ population (Figure 4G), consistent with a single Mendelian locus controlling a majority of the variation in the relative abundance of xylosides vs. glucosides of DHBAs.

QTL mapping of DHBA glycosylation to single locus resolution

To map the genetic variation responsible for the variation in the sugar moiety of the DHBA glycosides, we measured DHBA glycoside levels in 268 RILs generated from C24 and Col-0 (Torjek *et al.* 2006) and carried out QTL mapping analysis (Table S2). We calculated the molar percentage of DHBA xylosides (the total amount of 2,3- and 2,5-DHBAX divided by the total amount of the four DHBA glycosides) and performed a $\log_{10}$ transformation of the trait for QTL mapping analysis. Strikingly, a single QTL at the top of chromosome 5 accounted for 84% of the phenotypic variance and was the only QTL responsible for the anticorrelation of glucosides and xylosides (Figure 5). This was consistent with the F₂ population segregation patterns and indicated that a single Mendelian locus

may segregate within this QTL. The high penetrance of the trait suggested that molecular identification of the natural genetic variation in DHBA glycosylation was feasible.

We further analyzed the accumulation of DHBA glycosides in NILs. Lines derived from crosses between C24 and Col-0 containing introgressions of the QTL region (Torjek *et al.* 2008) were phenotyped for DHBA glycosides. This analysis confirmed that the genotype of the chromosome 5 QTL is sufficient to determine the relative abundance of DHBAG and DHBAX (Figure S2), corroborating the QTL mapping result. Using the NIL and RILs containing recombination junctions at the top of chromosome 5 were insufficient to localize the QTL to single gene resolution. The region identified by the NIL and RIL was a 450-kb interval containing 165 annotated genes (data not shown). One of these, At5g03490, encodes a putative UDP-glycosyltransferase (UGT), an enzyme predicted to catalyze a reaction that transfers a sugar from a UDP-sugar to an acceptor molecule.

To test whether this locus on chromosome 5 was responsible for the global variation of *A. thaliana* for DHBA glycosylation types observed, we performed a genome-wide association experiment. Leaf DHBA glycoside levels were determined for a 96-accession association panel that has previously been genotyped (Nordborg *et al.* 2005). This association panel is small and population structure is a confounding factor that required corrections (Kang *et al.* 2008). The ratio of the two sugar types of DHBA glycosides was strongly influenced by a locus at the top of chromosome 5 (Figure 6). This corresponds to a region contained within the QTL identified by linkage mapping (Figure 5). The SNP that had the greatest effect was of moderate frequency (SNP at Chr5 880586: *P*-value = 1.2×10^{-9} ; FDR-adjusted *P*-value < 10^{-6} ; Col-like allele frequency = 0.52; Table S3). Notably, multiple significant SNP residues within a few kilobases of At5g03490, the putative UGT gene (Table S3).

Identification of a 2,5-DHBA:UDP-xylose xylosyltransferase

If variation in At5g03490 is responsible for the QTL detected in the RILs and by genome-wide association, it is likely that the Col-0 allele encodes a DHBA:UDP-xylose xylosyltransferase. According to a recommended UGT superfamily nomenclature system (Mackenzie *et al.* 1997), At5g03490 encodes UGT89A2 (Ross *et al.* 2001; Paquette *et al.* 2009), which belongs to GT family 1 as defined in the widely used

Figure 6 Identification and localization of genetic variation responsible for variation in the xyloside proportion of DHBA glycosides. Association tests were carried out using a genome-wide collection of 177,088 SNPs and DHBA glycoside measurements from 96 accessions of *A. thaliana*. Each point represents an individual SNP. The y-axis is the negative log of the *P*-value for the test for association between the segregation of that SNP and the proportion of DHBA xylosides present in leaf tissue (see *Materials and Methods*). The x-axis corresponds to the position of each SNP in the genome. The five chromosomes of *A. thaliana* are indicated by alternating black and gray color for each. The red and blue horizontal lines correspond to *P*-values of 10^{-5} and 10^{-7} , respectively.

CAZy database (<http://www.cazy.org/>) (Lombard *et al.* 2014). To test the activity of UGT89A2, we expressed its gene from Col-0 in *E. coli* and tested the protein extract for xylosyltransferase activity toward 2,5-DHBA. In the presence of UDP-xylose, the protein extract of the UGT89A2-expressing *E. coli* converted 2,5-DHBA to 2,5-DHBAX, but that of the empty vector control did not (Figure 7). This indicates that Col-0 UGT89A2 has 2,5-DHBA:UDP-xylose xylosyltransferase activity.

To test if UGT89A2 is responsible for the accumulation of 2,5-DHBA xyloside *in planta*, we obtained two lines harboring independent T-DNA insertions within At5g03490 in Col-0 background. In both of the T-DNA lines, the open reading frame of *UGT89A2* was disrupted (Figure 8A), so it is likely both lines are knockout mutants. No visible phenotypes were observed for these mutants. We measured the DHBA glycoside accumulation in these mutants and compared it with that of wild-type Col-0. Whereas wild-type Col-0 accumulates 2,5-DHBAX, the two T-DNA insertion mutants lack this compound and instead accumulate 2,5-DHBAG at a comparable level (Figure 8B). These results demonstrated that UGT89A2 is required for the accumulation of 2,5-DHBAX in Col-0 plants, and in the absence of this enzyme, a cryptic UGT activity for glucosylation of 2,5-DHBA becomes predominant.

Discussion

Plants have evolved the ability to produce myriad chemicals to cope with adverse environments and ward off herbivores. It is estimated that 15–25% of the genes in a plant genome encode enzymes for specialized metabolism (Pichersky and

Figure 7 Col-0 UGT89A2 has a 2,5-DHBA xylosyltransferase activity. Crude protein extract from *E. coli* expressing Col-0 *UGT89A2* or harboring an empty pET30a vector was tested for 2,5-DHBA xylosyltransferase activities using a HPLC-based assay.

Gang 2000). By using a combination of approaches, including compound identification, QTL mapping, genome-wide association analysis, and protein activity assays, to study a biochemical polymorphism revealed from nontargeted metabolite profiling of *Arabidopsis* natural accessions, we have identified a *UGT* gene responsible for glycosylation of DHBAs. In a previous study that screens the *in vitro* activity of recombinant *Arabidopsis* UGTs expressed from *E. coli* against benzoate compounds, UGT89A2 was shown to be able to glucosylate 2,5-DHBA (Lim *et al.* 2002). However, its activity for xylosylation was not tested because only UDP-glucose was used as the sugar donor in that study. By metabolite profiling and identification, we were able to test the biologically relevant substrates for this enzyme and discovered the *in planta* function of UGT89A2 as a DHBA xylosyltransferase. This study demonstrates that exploiting natural variation in *Arabidopsis* is an efficient and powerful approach to elucidate gene function in secondary metabolism.

Phenotypic variation in metabolite accumulation has been shown to have relatively high heritability compared to complex traits, probably due to their simple genetic architecture (Baxter and Dilkes 2012; Korte and Farlow 2013). In this case, the F_2 population indicated that DHBA compounds were either xylosylated or glucosylated. The ratio of these two glycosylation types was the parameter that

Figure 8 Disruption of *UGT89A2* in Col-0 results in accumulation of 2,5-DHBAG instead of 2,5-DHBAX. (A) Schematic representation of the positions of the T-DNA insert in two independent *UGT89A2* knockout lines, SALK_147583C and SALK_081110C. (B) HPLC chromatograms of the leaf extract from Col-0 and the two T-DNA knockout lines.

best predicted the genotype. We identified a single locus underlying the molecular basis for the variation. QTL and association studies control for the effects of genetic background via the inclusion of genetic co-factors in the tests (Zeng 1993; Kang *et al.* 2008, 2010), but measurement imprecision and uncontrolled environmental effects still contribute to the error term and interfere with QTL detection and localization. In the case of these metabolites, the precision of the HPLC methods used to quantify DHBA glycosides in this study resulted in little technical inconsistency in the measurement of the compounds. Thus, appropriate phenotyping and crude metabolic modeling permitted extremely high sensitivity for the detection of variation at an enzyme locus. This has also been demonstrated by a recent genome-wide association study of human serum metabolites (Gieger *et al.* 2008). We predict that simple modeling steps for pathways, such as principle components analysis (Sebastian *et al.* 2002; Gilbert and Le Roy 2003) as well as ratios of

anticorrelated compounds such as we used here, will accelerate the discovery of genes responsible for metabolic variation in natural populations.

Moderate frequency polymorphisms will provide the best power to detect QTL, as maximum replication of the minor allele at each SNP will be achieved at a SNP allele frequency of 0.5. In our case, the accumulation of xylosylated DHBA was observed in ~50% of the 96 lines assayed and our lowest *P*-value was obtained at a locus with nearly 50–50 allele frequencies. This contrasts with other successful work in which dramatic differences in metabolite abundance were used to identify underlying polymorphisms using phenotypically extreme accessions, to map alleles that were not widespread in natural populations (Li *et al.* 2010a).

2,5-DHBA and 2,3-DHBA are structurally closely related to salicylic acid (2-hydroxybenzoic acid) (SA), an endogenous signaling molecule that plays an important role in plant defense against pathogens (Mettraux 2002). Feeding experiments using isotope-labeled SA demonstrated that 2,5-DHBA and 2,3-DHBA could be made from SA in leaves of various plants including *Arabidopsis* (Ibrahim and Towers 1959; Bartsch *et al.* 2010). Consistent with this observation, *Arabidopsis* isochorismate synthase mutant *sid2*, which is defective in SA biosynthesis, was found to be deficient in DHBA glycosides (Bartsch *et al.* 2010). More recently, a SA 3-hydroxylase gene was identified from *Arabidopsis*, providing direct evidence for the involvement of SA in DHBA biosynthesis (Zhang *et al.* 2013).

DHBA glycoside biosynthesis has been implicated in defense responses and leaf senescence regulation (Belles *et al.* 1999; Fayos *et al.* 2006; Bartsch *et al.* 2010; Zhang *et al.* 2013), but it is yet unclear whether these are due to modulation of SA activity or new activities conferred by DHBAs and/or their glycosides. Glycosylation is a common modification that contributes to the diversity of secondary metabolism. It has been shown that glycosylation has various effects on plant metabolites including increasing stability, controlling compartmentalization, and modifying biological activity (Jones and Vogt 2001). DHBA glycosides may serve as an inactive storage form, parallel to the role of SA β -glucoside for SA (Chen *et al.* 1995). Alternatively, the sugar moiety of DHBA glycosides may be important for their suggested signaling activity in plant defense (Belles *et al.* 1999; Bartsch *et al.* 2010). There are known cases where the sugar moiety of a glycoside has been shown to be important for the biological activity of the molecule. For example, an intact sugar moiety was found to be essential for the membrane disruption and antifungal activity of oat saponins, a group of triterpene glycosides (Armah *et al.* 1999). It has also been reported that the 3- β -glucoside of resveratrol, piceid (Waterhouse and Lamuela-Raventos 1994), has significant higher inhibitory effect than the aglycone on spore germination and penetration of *Venturia inaequalis*, the causal agent for apple scab (Schulze *et al.* 2005).

There is no clear association between geographic origin of the accessions and the sugar type of the DHBA glycosides

they accumulate. This biochemical phenotypic difference may be neutral to plant fitness but if xylosides and glucosides confer specificity to DHBAs in defense against different pathogens, frequency-dependent selection is possible (Chaboudez and Burdon 1995; Clay and Kover 1996). Consistent with a function for these molecules, DHBA xylosides accumulation can be induced in plants by some pathogens (Fayos *et al.* 2006). Furthermore, the polymorphisms at UGT89A2 comprise moderate-frequency extended haplotypes, which could be produced by frequency-dependent selection. Further research into the functions of these DHBA glycosides may shed light on the possible evolutionary forces behind the observed metabolite variation.

Acknowledgments

We are grateful to Rhonda Meyer for kindly providing the C24 x Col-0 and Col-0 x C24 RILs. We thank Bruce Cooper for carrying out the LC-MS analysis and Junko Maeda for technical assistance. This work was primarily funded by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the US Department of Energy through grant DE-FG02-07ER15905 to C.C.

Literature Cited

- Alonso, J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. M. Chen *et al.*, 2003 Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301: 653–657.
- Arends, D., P. Prins, R. C. Jansen, and K. W. Broman, 2010 R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* 26: 2990–2992.
- Armah, C., A. Mackie, C. Roy, K. Price, A. Osbourn *et al.*, 1999 The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. *Biophys. J.* 76: 281–290.
- Atwell, S., and Y. S. Huang, B. J. Vilhjálmsson, G. Willems, M. Horton *et al.*, 2010 Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.
- Balandrin, M. F., J. A. Klocke, E. S. Wurtele, and W. H. Bollinger, 1985 Natural plant chemicals: sources of industrial and medicinal materials. *Science* 228: 1154–1160.
- Bartsch, M., P. Bednarek, P. D. Vivanos, B. Schneider, E. von Roepenack-Lahaye *et al.*, 2010 Accumulation of isochlorismate-derived 2,3-dihydroxybenzoic 3-O-beta-D-xyloside in *Arabidopsis* resistance to pathogens and ageing of leaves. *J. Biol. Chem.* 285: 25654–25665.
- Baxter, I., and B. P. Dilkes, 2012 Elemental profiles reflect plant adaptations to the environment. *Science* 336: 1661–1663.
- Belles, J. M., R. Garro, J. Fayos, P. Navarro, J. Primo *et al.*, 1999 Gentisic acid as a pathogen-inducible signal, additional to salicylic acid for activation of plant defenses in tomato. *Mol. Plant Microbe Interact.* 12: 227–235.
- Broman, K. W., H. Wu, S. Sen, and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890.
- Burkart-Waco, D., C. Josefsson, B. Dilkes, N. Kozloff, O. Torjek *et al.*, 2012 Hybrid incompatibility in *Arabidopsis* is determined by a multiple-locus genetic network. *Plant Physiol.* 158: 801–812.
- Cao, J., K. Schneeberger, S. Ossowski, T. Günther, S. Bender *et al.*, 2011 Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43: 956–963.
- Chaboudez, P., and J. Burdon, 1995 Frequency-dependent selection in a wild plant-pathogen system. *Oecologia* 102: 490–493.
- Chan, E. K., H. C. Rowe, and D. J. Kliebenstein, 2010 Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185: 991–1007.
- Chen, F., D. Tholl, J. C. D'Auria, A. Farooq, E. Pichersky *et al.*, 2003 Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell* 15: 481–494.
- Chen, Z., J. Malamy, J. Henning, U. Conrath, P. Sanchez-Casas *et al.*, 1995 Induction, modification, and transduction of the salicylic acid signal in plant defense responses. *Proc. Natl. Acad. Sci. USA* 92: 4134–4137.
- Clay, K., and P. X. Kover, 1996 The red queen hypothesis and plant/pathogen interactions. *Annu. Rev. Phytopathol.* 34: 29–50.
- D'Auria, J. C., and J. Gershenzon, 2005 The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Curr. Opin. Plant Biol.* 8: 308–316.
- Desjardins, P., J. B. Hansen, and M. Allen, 2009 Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer. *J. Vis. Exp.* (33): doi: 10.3791/1610.
- Dixon, R. A., and D. Strack, 2003 Phytochemistry meets genome analysis, and beyond. *Phytochemistry* 62: 815–816.
- Fayos, J., J. M. Bellés, M. P. Lopez-Gresa, J. Primo, and V. Conejero, 2006 Induction of gentisic acid 5-O-beta-D-xylopyranoside in tomato and cucumber plants infected by different pathogens. *Phytochemistry* 67: 142–148.
- Gieger, C., L. Geistlinger, E. Altmaier, M. H. De Angelis, F. Kronenberg *et al.*, 2008 Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 4: e1000282.
- Gilbert, H., and P. Le Roy, 2003 Comparison of three multitrait methods for QTL detection. *Genet. Sel. Evol.* 35: 281–304.
- Hartmann, T., 2007 From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry* 68: 2831–2846.
- Ibrahim, R. K., and G. Towers, 1959 Conversion of salicylic acid to gentisic acid and o-pyrocatechuic acid, all labelled with carbon-14, in plants. *Nature* 184: 1803.
- Jones, P., and T. Vogt, 2001 Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta* 213: 164–174.
- Kang, H. M., N. a. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman *et al.*, 2008 Efficient control of population structure in model organism association mapping. *Genetics* 178: 1709–1723.
- Kang, H. M., J. H. Sul, S. K. Service, N. a. Zaitlen, S. Kong *et al.*, 2010 Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42: 348–354.
- Keurentjes, J. J., J. Fu, C. R. De Vos, A. Lommen, R. D. Hall *et al.*, 2006 The genetics of plant metabolism. *Nat. Genet.* 38: 842–849.
- Kliebenstein, D. J., V. M. Lambrix, M. Reichelt, J. Gershenzon, and T. Mitchell-Olds, 2001a Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell* 13: 681–693.
- Kliebenstein, D. J., J. Kroymann, P. Brown, A. Figuth, D. Pedersen *et al.*, 2001b Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* 126: 811–825.
- Korte, A., and A. Farlow, 2013 The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9: 29.
- Kroymann, J., S. Textor, J. G. Tokuhisa, K. L. Falk, S. Bartram *et al.*, 2001 A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiol.* 127: 1077–1088.
- Li, X., J. Bergelson, and C. Chapple, 2010a The *Arabidopsis* accession pna-10 is a naturally occurring sng1 deletion mutant. *Mol. Plant* 3: 91–100.

- Li, Y., Y. Huang, J. Bergelson, M. Nordborg, and J. O. Borevitz, 2010b Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 107: 21199–21204.
- Lim, E. K., C. J. Doucet, Y. Li, L. Elias, D. Worrall *et al.*, 2002 The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.* 277: 586–592.
- Lombard, V., H. G. Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat, 2014 The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42: D490–D495.
- Mackenzie, P. I., I. S. Owens, B. Burchell, K. Bock, A. Bairoch *et al.*, 1997 The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenet. Genomics* 7: 255–269.
- Metraux, J., 2002 Recent breakthroughs in the study of salicylic acid biosynthesis. *Trends Plant Sci.* 7: 332–334.
- Nordborg, M., T. T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 3: e196.
- Paquette, S. M., K. Jensen, and S. Bak, 2009 A web-based resource for the *Arabidopsis* P450, cytochromes *b*₅, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (<http://www.P450.kvl.dk>). *Phytochemistry* 70: 1940–1947.
- Pichersky, E., and D. R. Gang, 2000 Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 5: 439–445.
- Ross, J., Y. Li, E. Lim, and D. J. Bowles, 2001 Higher plant glycosyltransferases. *Genome Biol.* 2: 3004.1–3004.6.
- Ruegger, M., and C. Chapple, 2001 Mutations that reduce sinapoylmalate accumulation in *Arabidopsis thaliana* define loci with diverse roles in phenylpropanoid metabolism. *Genetics* 159: 1741–1749.
- Sakushima, A., M. Coşkun, and T. Maoka, 1995 Hydroxybenzoic acids from *boreava orientalis*. *Phytochemistry* 40: 257–261.
- Schulze, K., L. Schreiber, and I. Szankowski, 2005 Inhibiting effects of resveratrol and its glucoside piceid against *venturia inaequalis*, the causal agent of apple scab. *J. Agric. Food Chem.* 53: 356–362.
- Sebastian, R., M. Kearsey, and G. J. King, 2002 Identification of quantitative trait loci controlling developmental characteristics of brassica oleracea L. *Theor. Appl. Genet.* 104: 601–609.
- Shirley, B. W., W. L. Kubasek, G. Storz, E. Bruggemann, M. Koornneef *et al.*, 1995 Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* 8: 659–671.
- Tholl, D., F. Chen, J. Petri, J. Gershenzon, and E. Pichersky, 2005 Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. *Plant J.* 42: 757–771.
- Torjek, O., H. Witucka-Wall, R. C. Meyer, M. von Korff, B. Kusterer *et al.*, 2006 Segregation distortion in *Arabidopsis* C24/col-0 and col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theor. Appl. Genet.* 113: 1551–1561.
- Torjek, O., R. C. Meyer, M. Zehnsdorf, M. Teltow, G. Strompen *et al.*, 2008 Construction and analysis of 2 reciprocal *Arabidopsis* introgression line populations. *J. Hered.* 99: 396–406.
- Turner, S., 2001 Annotated Manhattan plots and QQ plots for GWAS using R, Revisited. Available at: <http://gettinggenetics-done.blogspot.com/2011/04/annotated-manhattan-plots-and-qq-plots.html>.
- Waterhouse, A. L., and R. M. Lamuela-Raventos, 1994 The occurrence of piceid, a stilbene glucoside, in grape berries. *Phytochemistry* 37: 571–573.
- Yonekura-Sakakibara, K., and K. Saito, 2009 Functional genomics for plant natural product biosynthesis. *Nat. Prod. Rep.* 26: 1466–1487.
- Zeng, Z. B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA* 90: 10972–10976.
- Zhang, K., R. Halitschke, C. Yin, C. J. Liu, and S. S. Gan, 2013 Salicylic acid 3-hydroxylase regulates *Arabidopsis* leaf longevity by mediating salicylic acid catabolism. *Proc. Natl. Acad. Sci. USA* 110: 14807–14812.

Communicating editor: C. D. Jones

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.168690/-/DC1>

Exploiting Natural Variation of Secondary Metabolism Identifies a Gene Controlling the Glycosylation Diversity of Dihydroxybenzoic Acids in *Arabidopsis thaliana*

Xu Li, Elisabeth Svedin, Huaping Mo, Susanna Atwell, Brian P. Dilkes, and Clint Chapple

^{13}C NMR spectrum of compound 1

^{13}C NMR spectrum of compound 2

^1H NMR spectrum of compound 1

^1H NMR spectrum of compound 2

2,5-DHBAG

2,3-DHBAG

Figure S1. ^{13}C and ^1H NMR spectra of 2,5-DHBAG and 2,3-DHBAG.

Figure S2. leaf soluble metabolites for two Col-0/C24 NILs. The M42 line has a C24 background with the top of chromosome 5 region from Col-0. The N69 line has a Col-0 background with the top of chromosome 5 region from C24.

Tables S1-S3

Available for download as Excel files at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.168690/-/DC1>

Table S1 ^{13}C and ^1H NMR spectral data of 2,5-DHBAX and 2,3-DHBAX.

Table S2 DHBA glycoside data of 268 Col-0/C24 RILs for QTL mapping.

Table S3 SNPs identified from genome wide association.