# Discovery of a novel amino acid racemase through exploration of natural variation in *Arabidopsis thaliana*

Renee C. Strauch[a,b], Elisabeth Svedin[c], Brian Dilkes[c], Clint Chapple[d], and Xu Li[a,b,1]

[a]Plants for Human Health Institute, North Carolina State University, Kannapolis, NC 28081; [b]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695; [c]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907; and [d]Department of Biochemistry, Purdue University, West Lafayette, IN 47907

Plants produce diverse low-molecular-weight compounds via specialized metabolism. Discovery of the pathways underlying production of these metabolites is an important challenge for harnessing the huge chemical diversity and catalytic potential in the plant kingdom for human uses, but this effort is often encumbered by the necessity to initially identify compounds of interest or purify a catalyst involved in their synthesis. As an alternative approach, we have performed untargeted metabolite profiling and genome-wide association analysis on 440 natural accessions of *Arabidopsis thaliana*. This approach allowed us to establish genetic linkages between metabolites and genes. Investigation of one of the metabolite–gene associations led to the identification of *N*-malonyl-D-allo-isoleucine, and the discovery of a novel amino acid racemase involved in its biosynthesis. This finding provides, to our knowledge, the first functional characterization of a eukaryotic member of a large and widely conserved phenazine biosynthesis protein PhzF-like protein family. Unlike most of known eukaryotic amino acid racemases, the newly discovered enzyme does not require pyridoxal 5′-phosphate for its activity. This study thus identifies a new D-amino acid racemase gene family and advances our knowledge of plant D-amino acid metabolism that is currently largely unexplored. It also demonstrates that exploitation of natural metabolic variation by integrating metabolomics with genome-wide association is a powerful approach for functional genomics study of specialized metabolism.

D-amino acid | racemase | genome-wide association | secondary metabolism | natural variation

Plants have the ability to create over 200,000 small compounds known as secondary or specialized metabolites (1). These chemically diverse compounds help mediate plant adaptation to their environment and play important roles in plant defense mechanisms, pigmentation, and development. In addition, many of these metabolites are desirable to humans as medicinal and nutritional compounds. Therefore, furthering our understanding of plant specialized metabolism will have profound impacts on various applications from crop improvement to human health.

To date, only a small fraction of the chemical and catalytic space in plant specialized metabolism has been explored. Even in the best-studied model plant *Arabidopsis thaliana*, there are still many uncharacterized metabolites, and the vast majority of genes encoding enzymes implied to be involved in specialized metabolism do not have known associations with any metabolites. Several studies of *Arabidopsis* natural accessions (individuals collected from wild populations) revealed considerable qualitative and quantitative variation in the accumulation of various compounds such as glucosinolates, terpenoids, and phenylpropanoids (2–4). This extensive metabolite variation can be attributed to genetic variation in genes encoding enzymes and regulatory factors of the pathways involved; quantitative trait locus (QTL) mapping has successfully uncovered several genes involved in the production of these metabolites (3–7). Liquid chromatography–mass spectrometry (LC-MS)–based untargeted metabolic profiling has further extended such analysis to unknown metabolites, finding genetic

contribution to the variation in at least three-fourths of detected mass peaks (8).

Here we describe an integrated transdisciplinary platform, combining metabolomics, genetics, and genomics, to exploit the biochemical and genetic diversity of natural accessions of the model plant *A. thaliana* to uncover associations between genes and metabolites. Using this platform, we linked a differentially accumulating metabolite, identified through chemical analysis as *N*-malonyl-D-*allo*-isoleucine (NMD-Ile), to a previously uncharacterized gene identified as an amino acid racemase through reverse genetics and biochemical analysis.

Amino acids exist in two forms; L-amino acids, the proteogenic form, and their enantiomorphs, D-amino acids. D-amino acids also play important structural and physiological roles in diverse life systems. In bacteria, D-amino acids confer cell-wall protease resistance and regulate cell-wall remodeling (9–11). D-amino acids were also found to be involved in signaling mechanisms in animal nervous systems and plant pollination (12, 13). Enzymes that catalyze the conversion of L-amino acids to D-amino acids are a class of isomerases known as amino acid racemases (14). Amino acid racemases are categorized into pyridoxal 5′ phosphate (PLP)-dependent and PLP-independent families. PLP-dependent racemases include AlaR, SerR, ArgR, and AspR and are found in both bacteria and eukaryotes, including mammals and plants. PLP-independent amino acid racemases include bacterial ProR, GluR, AspR, and diaminopimelate (DAP) epimerase.

## Significance

We describe how untargeted metabolic profiling and genome-wide association analysis was used in *Arabidopsis thaliana* to link natural products (secondary metabolites) with genes controlling their production. This powerful approach exposed metabolite–enzyme connections even without prior knowledge of the metabolite identity or the biochemical function of the associated enzyme. Further chemical and genetic analysis synergistically led to the discovery and characterization of a D-amino acid derivative, *N*-malonyl-D-allo-isoleucine, and a novel amino acid racemase responsible for its biosynthesis. Little is known about D-amino acid metabolism and its natural variation in plants. Additionally, this is the first functional characterization of a eukaryotic member of a large family of phenazine biosynthesis protein phzF-like proteins conserved across all the kingdoms.
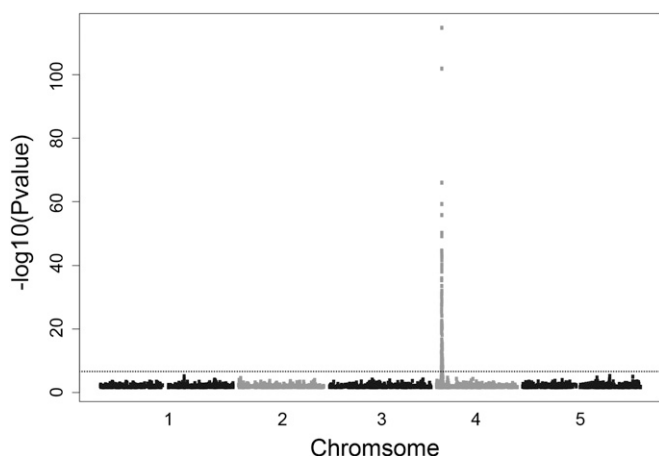
**Fig. 1.** Manhattan plot showing GWAS results for M172T666. The $-\log_{10}$ $P$ values for association with M172T666 for each SNP ($y$ axis) are plotted against chromosome positions ($x$ axis). The horizontal dashed line corresponds to a Bonferroni-corrected $\alpha < 0.05$.

Except for DAP epimerase, which is required for the biosynthesis of Lys in bacteria and plants, ProR in *Trypanosoma cruzi*, the protozoan parasite responsible for Chagas disease, has been the only known eukaryotic PLP-independent racemase (15). The *Arabidopsis* amino acid racemase that we describe in this paper is a new member of the PLP-independent family, making its identification a significant addition to this class of enzymes.

## Results

### Genome Wide Association Analysis Revealed a Strong QTL for an Unknown Metabolite.

A total of 440 natural *A. thaliana* accessions were chosen for this study due to their genetic diversity and the availability of their genome-wide single nucleotide polymorphism (SNP) data. LC-MS–based untargeted metabolite profiles of leaf tissue were collected for each accession. The metabolite phenotypes were subjected to linear mixed-model–based genome wide association (GWA) analysis using the available SNP data (16, 17). The strongest association revealed from this analysis was for a metabolite feature with a mass-to-charge ratio of 172 and a retention time of 666 s (M172T666) (Fig. 1). In LC-MS–based experiments, each metabolite often gives rise to multiple metabolite features. These features were grouped into a pseudospectrum of the metabolite during mass data analysis. Another two features, M130T666 and M216T665, were grouped with M172T666. All these features gave similar GWA results, showing a strong QTL at the top of chromosome 4 (*SI Appendix*, Table S1). For the convenience of discussion, we name this unknown metabolite by its molecular weight and will refer it as "M217" in the rest of the paper.

### Accumulation of M217 Is Controlled by DAAR1.

Loci near the best-associated SNP (4-1266038) were investigated as potential candidate genes controlling the accumulation of M217. Although there are several unknown and uncharacterized genes in the interval of interest, AT4G02850 (D-amino acid racemase1, *DAAR1*) and AT4G02860 (*DAAR2*) were deemed promising candidates for further study as both genes share sequence similarity with the PhzF gene involved in the biosynthesis of phenazines (18), a class of bacteria-specific metabolites, suggesting that they may encode enzymes for plant specialized metabolism (*SI Appendix*, Figs. S1 and S2). To investigate whether one of these two genes are involved in the accumulation of M217, T-DNA insertion lines for AT4G02850 and AT4G02860 (*daar1* and *daar2*) were obtained from the *Arabidopsis* Biological Resource Center (ABRC) (19). The expected insertions were confirmed by PCR amplification

of the expected insertion site (Fig. 2A). Homozygous T-DNA mutants were compared with wild-type plants, and disruption of neither gene resulted in obvious morphological defects. Metabolite profiles comparing levels of M217 showed that accumulation of M217 was severely decreased in *daar1* but unaffected in *daar2* compared with wild type (Fig. 2B). These data indicate that DAAR1 is responsible for the majority of M217 accumulation.

### Identification of *N*-Malonyl-D-*allo*-Isoleucine.

The chemical identity of M217 was needed to further elucidate biochemical basis for the genetic link found between this metabolite and *DAAR1*. By using a variety of chemical analysis techniques, including MS, MS/MS, NMR, and HPLC, we determined that M217 is NMD-Ile (Fig. 3).

The accurate mass and isotope profile information obtained from the MS data suggested a molecular formula of $C_9H_{15}NO_5$. More structural insights were obtained by NMR analysis on M217 purified from *Arabidopsis* leaf tissue using semipreparative HPLC. All assignments were done using $^1H$, $^{13}C$, heteronuclear single quantum coherence (HSQC), and correlation spectroscopy (COSY) spectra (Table 1). The COSY spectrum was especially useful in identifying the isoleucine moiety as there is a high degree of proton–proton coupling (J-coupling) in the R-group. Once the R-group was assigned, comparison with spectra of isoleucine standards confirmed the identity of the amino acid.

Identification of the malonyl derivatization was based on several lines of chemical evidence. Two carbonyl groups and a methylene ($CH_2$) group remained to be accounted for in the NMR spectrum after isoleucine was assigned. Although no additional direct structural information could be obtained from NMR, the lower-than-expected integration of C8 methylene ($CH_2$) in the $^1H$ spectrum was suggestive of a β-dicarbonyl group. Methylene protons in this group exchange with the solvent through a keto-enol tautomerization mechanism resulting in lower integration (20). In addition, MS/MS analysis suggested the presence of a very labile carboxylic acid group and an acetyl group; malonyl groups are known to degrade easily to acetyl groups by loss of a carboxylic acid, thus corroborating the MS/MS data.

The common presence of malonyl conjugates of D-amino acids in plants (21) led us to further investigate the stereo-configuration



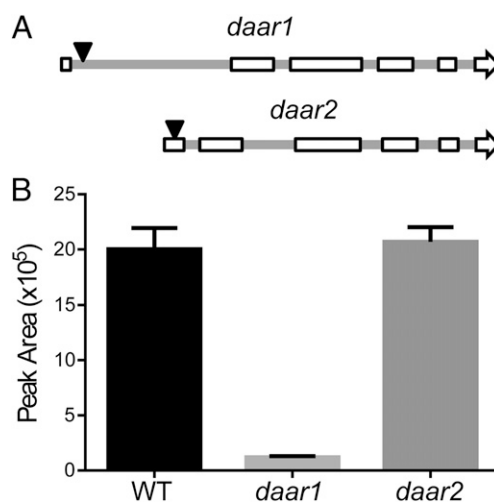**Fig. 2.** Accumulation of M217 in wild-type and T-DNA insertion mutants of *DAAR1* and *DAAR2*. (A) Schematic representation *DAAR1* and *DAAR2* genes with arrows marking the position of T-DNA insertions in *daar1* and *daar2*. Introns are gray lines, and white boxes represent exons. (B) Accumulation of M217 (mean peak areas) with SDs from four biological replicates of wild-type (WT) *daar1* and *daar2*.
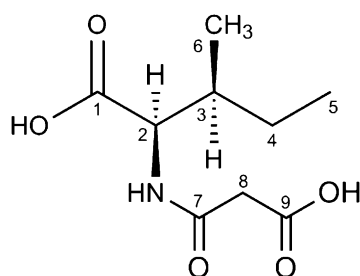
**Fig. 3.** Chemical structure of *N*-malonyl-D-*allo*-isoleucine. M217 was identified as *N*-malonyl-D-*allo*-isoleucine using NMR, MS/MS, and chiral derivatization.

of M217. Isoleucine contains two stereocenters at carbon 2 (α-carbon) and 3 (Fig. 3). Determination of the stereochemistry was performed using NMR and LC-MS. Comparison of $^1$H spectra of M217 and authentic L-Ile and L-*allo*-Ile indicated that carbon 3 is in the allo-configuration (*SI Appendix*, Fig. S3). To determine the stereoconfiguration of the α-carbon, the amino acid was released from M217 by base hydrolysis and subjected to derivatization with Marfey's reagent. This chiral derivatizing reagent causes intramolecular bonding of D-isomers resulting in longer retention on reverse-phase HPLC, allowing separation of D- and L-enantiomers that would otherwise coelute on a nonchiral column (22). Comparison with a derivatized authentic L-Ile standard revealed that the amino acid is the D-enantiomer.

**Enzymatic Characterization of DAAR1.** Sequence analysis revealed that DAAR1 belongs to the phenazine biosynthesis-like protein family (PF02567), which is closely related to the DAP epimerase (PF01678), proline racemase (PF05544), and PrpF (PF04303) families (pfam.xfam.org). The characterized members of these families all catalyze isomerization reactions, suggesting that DAAR1 is likely also an isomerase. Due to the D-configuration of the α-carbon of Ile in NMD-Ile and the shared conserved catalytic residues with other PLP-independent racemase enzymes, we tested the hypothesis that DAAR1 catalyzes the conversion of L-Ile to D-Ile.

The *DAAR1* gene was recombinantly expressed in *Escherichia coli* as a His-tagged protein and purified by affinity chromatography. Racemase activity of the purified protein was tested using L-Ile as substrate. Recombinant DAAR1 catalyzed the conversion of L-Ile to D-Ile whereas extracts from empty vector controls showed no racemase activity (Fig. 4*A*). Thus, DAAR1 is responsible for the observed racemization of L-Ile. Such activity was not detected for recombinant DAAR2 (*SI Appendix*, Fig. S4). We noted a rapid loss of activity at room temperature and when storing the enzyme at 4 °C for overnight, indicating enzyme instability. The affinity tag did not affect enzyme activity or stability and was not removed in subsequent experiments.

Further analysis of the recombinant DAAR1 activity revealed that this enzyme is a PLP-independent racemase. After dialysis against PLP-free buffer, DAAR1 showed no difference in activity in the presence or absence of PLP (*SI Appendix*, Fig. S5), suggesting a catalytic mechanism independent of PLP. PLP independence is also consistent with the sequence similarity of DAAR1 to other PLP-independent enzymes and the observation that DAAR1 required reducing conditions to function consistent with the Cys-residue active site and two-base mechanism used by PLP-independent racemases (15, 23). All these data suggest that DAAR1 is a PLP-independent racemase.

To determine enzyme kinetics for DAAR1, activity was assayed with 1–100 mM L-Ile (Fig. 4*B*). Apparent $K_m$ and $V_{max}$ values were calculated to be 17 ± 2.7 mM and 0.26 ± 0.01 mM·min$^{-1}$, respectively. The $k_{cat}$ was determined to be 662 min$^{-1}$. We further investigated DAAR1 substrate selectivity by testing racemization

of other branch-chain amino acids. Compared with Ile, DAAR1 showed 1 and 5% relative activity toward Leu and Val, respectively. DAAR1 showed no activity toward *N*-acetyl-L-Ile, a commercially available surrogate for *N*-malonyl-L-Ile, suggesting that DAAR1 does not use a *N*-derivatized substrate and that the malonylation of D-amino acids occurs downstream in the metabolic pathway.

## Discussion

Currently, we are not certain about the identities of the majority of specialized metabolites nor have we molecularly identified the genes encoding their biosynthetic pathways. As a result, the associations between the vast majority of metabolic diversity and organismal adaptation remain elusive. We have demonstrated a strategy to take anonymous "omic" observation of metabolites and standing variation and generate information regarding the metabolite identities, genes encoding biosynthetic enzymes, and alleles affecting variation. Our description of DAAR1 as a PLP-independent amino acid racemase provides evidence for the function of a large and unannotated gene family present in all major divisions of eukaryotes.

In contrast to primary metabolism where sequence similarity can serve as a reliable guide for gene function prediction, specialized metabolism has a high rate of gene divergence and undergoes rapid evolution; as a result, although useful for general classification, sequence analysis is often insufficient for correct prediction of the specific reaction catalyzed by individual enzymes (24). The relationship between genes, mRNA, and proteins is obvious because of the template-based nature of transcription and translation whereas the connection between enzymes and their substrates and products is harder to predict even with the availability of high-resolution crystallographic or NMR structural information. In this study, by exploiting within-species natural variation, we established genetic connection between an *Arabidopsis* gene and its substrate and the end product of the pathway in which they are involved. This metabolome–genome connection empowers us to synergistically integrate genetic and genomic information with chemical and biochemical knowledge for gene function and metabolite identification. Our work, as well as a recent study on rice using a similar approach (25), clearly demonstrates that GWA is a powerful discovery tool for functional genomics of specialized metabolism.

D-amino acids have been detected in many different organisms including bacteria, humans, and plants. It has long been known that D-amino acids such as D-Ala and D-Glu are structural components of bacterial peptidoglycan, which make the bacterial cell wall resistant to proteases (10). D-amino acid is also an important structural and functional component of diverse bioactive peptides that act as analgesics, antimicrobials, and bactericidal agents among many others (9). More recently, research has begun to reveal important physiological roles of D-amino acids in bacteria and mammals. Bacteria use D-amino acids as paracrine and autocrine effectors by secreting large amounts into their surroundings

**Table 1. NMR (D$_2$O, 600 MHz) data for *N*-malonyl-D-*allo*-isoleucine**

| No. | $^{13}$C | $^1$H | COSY |
|---|---|---|---|
| 1 | 170.6 | — | — |
| 2 | 57.8 | 4.26 (d) | H5 |
| 3 | 36.9 | 1.90 (m) | H2, H3, H6 |
| 4 | 26.1 | 1.18, 1.26 (m) | H1, H5 |
| 5 | 11.1 | 0.83 (t) | H3 |
| 6 | 14.1 | 0.81 (d) | H5 |
| 7 | 178.8 | — | — |
| 8 | 36.4 | 2.62 (s) | — |
| 9 | 170.6 | — | — |

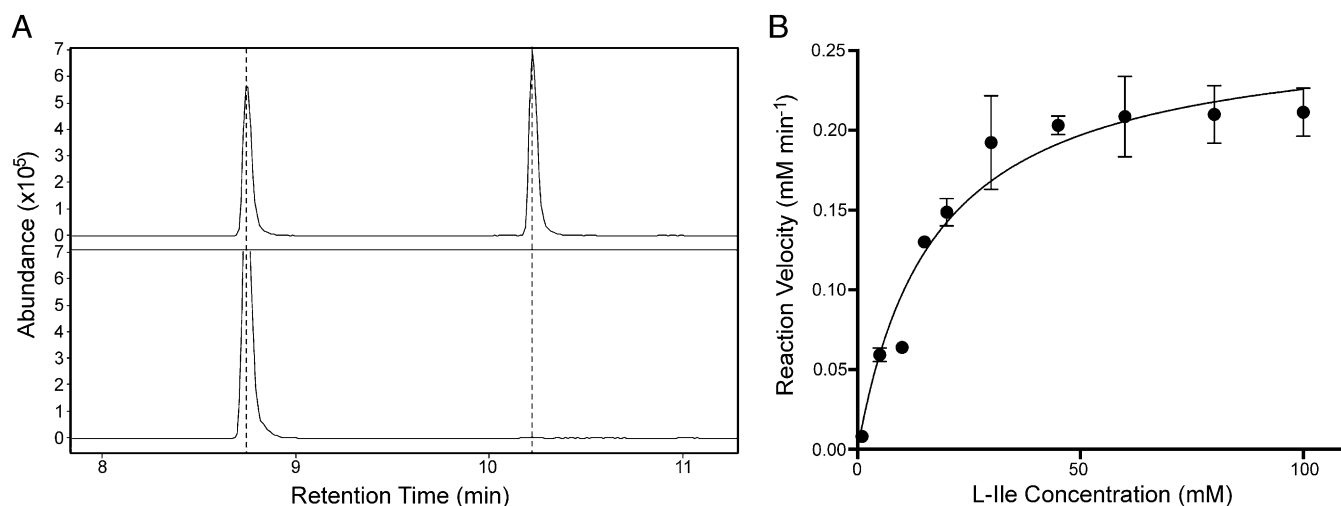d, doublet; m, multiplet; s, singlet; t, triplet.

**Fig. 4.** Activity and kinetic characterization of DAAR1. (*A*) Purified recombinant DAAR1 protein and the soluble crude fraction of an empty vector control were incubated with 10 mM L-Ile at room temperature for 8 h. The reactions were quenched and derivatized with Marfey's reagent (FDVA), an amine reactive agent that allows separation of optical isomers of amino acids by HPLC. DAAR1 catalyzed the conversion from L-Ile to D-*allo*-Ile (*Top*), whereas the empty vector control showed no isomerase activity (*Bottom*). (*B*) DAAR1 activity was assayed at L-Ile concentrations from 1 to 100 mM to determine its kinetic parameters.

to trigger population transition to stationary-phase growth and to regulate cell-wall architecture (11). In the mammalian nervous system, D-Asp is involved in development and acts as a neurotransmitter, and in the endocrine system it is an important player in hormone synthesis and release from the hypothalamus, pituitary glands, and gonads (12). D-Ser is a coregulator of NMDA receptors and an important component in excitatory neurotransmission in the cerebral cortex and hippocampus and plays a role in learning, memory, and behavior (13).

Although inroads have been made in uncovering the biological role of D-amino acids in mammals and bacteria, their function in plants remains mostly unknown despite having been investigated for decades. D-amino acids and their conjugates have been detected in diverse plant species; however, their physiological roles were often overlooked or discounted in the past because D-amino acids were deemed unnatural and exogenous application caused significant plant growth inhibition (26–28). Recently, investigation of different *Arabidopsis* accessions has revealed considerable natural variation in uptake and metabolism of D-amino acids (29, 30). Identification of plant genes encoding D-amino acid metabolism enzymes such as D-amino acid oxidases (31), D-amino acid transaminases (32), and particularly the D-Ser and D-Ala racemases (33–36) provides clear evidence for endogenous D-amino acid biosynthesis and metabolism, implying that D-amino acids may have important biological functions in plants. This suggestion was substantiated by the finding that D-Ser activates glutamate receptors and plays a role in pollen tube growth and morphogenesis (37). Despite this progress, the functions of many other D-amino acids found in plants and the genes responsible for their biosynthesis are currently unknown. In an early study, NMD-Ile was detected in barley shoots when fed with D-Ile (27). Our discovery of a novel type of racemase in *Arabidopsis* reveals the biosynthetic enzyme for this D-amino acid, adding new knowledge to D-amino acid metabolism in plants and allowing for the manipulation of the pathway to further investigate its biological significance.

To date, PLP-independent amino acid racemases have been found only in bacteria and in the protozoan parasite *T. cruzi* (15). The *Arabidopsis* DAAR1 required for *N*-malonyl-D-*allo*-isoleucine biosynthesis identified in this work represents the first member of this class of enzyme in multicellular eukaryotes. DAAR1 has similar enzymatic properties to other characterized amino acid racemases, both PLP-dependent and -independent. $K_m$ values

between 2.5 and 110 mM have been reported for amino acid racemases (33, 35, 36, 38–40); the $K_m$ of DAAR1, at 17 mM, is within the expected range. It is interesting to note that *Arabidopsis* DAP-epimerase, a related enzyme involved in L-Lys biosynthesis, has a 100-fold lower $K_m$ value (23). The low substrate affinity of amino acid racemases may be a typical feature needed for preventing the depletion of the L-amino acid pool essential for protein synthesis.

DAAR1 belongs to a large phenazine biosynthesis-like family that has nearly 3,500 members and is represented in more than 2,000 species from bacteria to eukaryotes, including humans (pfam.xfam.org). Several genes of this family were shown to have important biological functions. For example, knocking out the yeast gene family member YHI9 results in an increased sensitivity to the pore-forming antifungal drug nystatin, suggesting a possible role of this gene in membrane regulation and stability (41). Experiments on the human gene family member, MAWD-binding protein, indicate that it may have a role in suppressing tumorigenicity (42). The biochemical functions of YHI9 and MAWD-binding protein are unknown, as are all members of this large gene family, except for the bacterial phenazine biosynthetic gene PhzF (18). Our work on *Arabidopsis* DAAR1 demonstrates the enzymatic function of a eukaryotic family member, facilitating further investigation of molecular mechanisms of the biological roles of these pathways.

## Materials and Methods

**Chemicals.** All chemicals were obtained from Sigma-Aldrich.

**Plant Material and Growth Condition.** Seeds of *Arabidopsis* natural accession [Core360 set, stock no. CS76309, and 80 accessions described by Cao et al. (43), stock no. CS76427] and T-DNA mutant lines (SALK_059126C, *daar1*; and WISCDSLOX485-488C16/CS864801, *daar2*) were ordered from ABRC (https://abrc.osu.edu). Twenty of the Core360 set lines for which seeds were not available for ordering were kindly provided by Prof. David Salt, University of Aberdeen, United Kingdom.

A restricted randomization design was used in the natural accession experiment to minimize possible confounding environmental effects. Specifically, plants were grown in 36-well flats with each plant per well. Three Col-0 plants were planted in each flat at three fixed positions, serving as a control for variation between flats. For each of the 440 accessions, three plants were planted. The total of 1,320 plants were randomly allocated to 40 flats with the restriction that plants of the same accession were not planted in the same flat. These 40 flats were put on a single bench in a growth room at 22 °C and

50% humidity under long-day conditions (16 h light, 8 h dark). The second pair of rosette leaves was collected from 4-wk-old plants, weighed, flash-frozen in liquid nitrogen, and stored at −80 °C. Leaves from each plant were analyzed individually.

**Metabolite Extraction.** Tissues were extracted with 50% (vol/vol) methanol containing 2 mg/L tetrapeptide Ala-Leu-Ala-Leu at 60 °C for 30 min. The tissue-to-solvent ratio has been kept constant across all of the samples. When necessary, samples were 0.4-μm-filtered before analysis by LC-MS.

**LC-MS Instrument and Column.** Metabolite profiles were generated on a G6530A Q-TOF LC/MS equipped with an Eclipse Plus C18, 3- × 100-mm column from Agilent Technologies. The mobile phase used a gradient of acetonitrile with 0.1% formic acid and water with 0.1% formic acid. Data were acquired in negative mode.

**Semipreparative HPLC and Column.** The Semi-Prep HPLC system with a L-2450 Elite LaChrome diode array detector (Hitachi) was used. A custom-packed Zorbax Eclipse plus C18, 5 μm, i.d. 9.4- × 150-mm column was used from Agilent Technologies. The mobile phase used a gradient of acetonitrile with 0.1% formic acid and water with 0.1% formic acid.

**NMR.** NMR spectra were obtained on a Bruker 700 MHz NMR spectrometer equipped with AVANCE II console and cryogenically cooled QNP probe. $^1$H experiments were performed at a frequency of 700.13 MHz and $^{13}$C experiments at a frequency of 176.05 MHz. $^1$H experiments used a 30 ° pulse, number of points of 64 k, a sweepwidth of 20.6 ppm, delay of 1 s, and number of scans of 64). COSY experiments used 2,048 × 128 points, a sweepwidth of 13.3 ppm, delay of 1.5 s, and number of scans of 8. HSQC experiments used 1,024 × 256 points, an $^1$H sweepwidth of 13.3 ppm, an $^{13}$C sweepwidth of 165.6 ppm, delay of 1.5 s, and number of scans of 16.

**LC-MS Data Processing.** Agilent MassHunter raw data (.d directories) were converted to mzXML format using the msconvert tool in the ProteoWizard software (44). The mzXML files were grouped into different folders by accession and processed with the open-source software XCMS (45). Peak picking was performed using the centWave algorithm with the following parameters: ppm = 50, peakwidth = c (8, 50), snthresh = 10, and prefilter = c (3, 100). Peaks were aligned using "obiwarp" method and grouped with bw = 10, mzwid = 0.025, and minsamp = 3. The R package CAMERA (46) was used to identify isotope and adduct peaks. The parameters used for finding isotopes were the following: ppm = 10, mzabs = 0.01, intval = "into," and minfrac = 0.25. The following parameters were used to find adducts: ppm = 20, mzabs = 0.015, and multiplier = 4. A custom R script was written to select a representative peak from each pseudospectrum based on intensity. The final peak table was exported to a csv file for downstream analysis.

**GWA Analysis.** Each peak was used as a trait for GWA analysis. The 206,087 SNP genotypes of the accessions used in this study were extracted from the available data described previously (16). To reduce the chance of spurious association, those with a minor allele frequency less than 5% were excluded, leaving 192,616 SNPs for the analysis. The Efficient Mixed-Model Association eXpedited procedure (17) was used for testing the association between each SNP and the phenotype, as previously described (7).

**Genotyping of T-DNA Insertion Mutants.** Genomic DNA was isolated by grinding a small piece of leaf tissue (∼0.25 cm$^2$) in Shorty Buffer (5 mM EDTA, 80 mM LiCl, 0.2% SDS, 40 mM Tris, pH 9). The cell debris was pelleted by centrifugation, and an aliquot of supernatant was added to an equal volume of isopropanol to precipitate the DNA. The DNA was pelleted by centrifugation, and the supernatant was removed by decanting and air drying. The DNA was redissolved in TE Buffer (10 mM Tris, 1 mM EDTA, pH 9) and stored at −20 °C.

Genotyping of the T-DNA lines were performed using two pairs of primers. A T-DNA border primer (BP) and a primer downstream of the insertion site (RP) were used to detect the T-DNA allele. The presence of wild-type allele was probed using RP and a primer upstream of the insertion site. For SALK_059126C, the T-DNA border primer used was 5′-ATTTTGCCGATTTCG-GAAC-3′, the RP used was 5′-TGAGAGGCAAAGCCGTTACT-3′, and the LP used was 5′-CAAGTGGAAGCAAGCACTCA-3′. For WISCDSLOX485-488C16, the BP used was 5′-AACGTCCGCAATGTGTTATTAAGTTGTC-3′, the RP used was 5′-TGAAGCCACGTGTCATCTCT-3′, and the LP used was 5′-AGATGCT-CCTGTAACCGTCG-3′.

**Recombinant Protein Expression and Purification.** *Arabidopsis DAAR1* and *DAAR2* full-length cDNA clones (S68253 and U67563) were obtained from ABRC. The ORF of DAAR1 (∼1 kb) was amplified by PCR using primers 5′-GAACAGATTGGAGGTATGGCCATGCTCGTGAA-3′ and 5′-ATTCGGATCCTCT-AGTGGCAAAGAGTCGAAAGATTAAACTAAGATAGAG-3′. The pE-SUMOpro Kan plasmid (LifeSensors) was linearized with BsaI, and the expression construct was assembled using Gibson Assembly Cloning Kit (New England BioLabs), following the manufacturer-recommended protocol. The DAAR2 expression construct was made in the same way, using primers 5′-CCGCGAACAGATT-GGAGGTATGGGGAAGAAGAAAGGTGTC-3′ and 5′-GCTCGAATTCGGATCCTT-CAGACCAAGACATGGCCTT-3′. The constructs were transformed into the Rosetta 2 pLysS *E. coli* strain (EMD Millipore). The cultures were grown in Terrific Broth (47) with 50 mg/L kanamycin and 34 mg/L chloramphenicol, inoculated at 30 °C, and induced at 18 °C with 0.5 mM isopropyl β-D-1-thi-ogalactopyranoside. The cells were harvested by centrifugation, and the pellet was stored at −80 °C until use.

Cell pellets were lysed with B-PER protein extraction reagents containing DNaseI and EDTA-free protease inhibitor mixture. The soluble extract was collected by centrifugation and applied to Ni-NTA Resin (750 μL resin per 10 mL soluble extract) previously equilibrated with lysis buffer. After a 30-min incubation at 4 °C with occasional mixing, the resin was washed several times with Wash Buffer [50 mM Tris, 500 mM NaCl, 20 mM imidazole, 10% (vol/vol) glycerol, 10 mM β-mercaptoethanol, pH 8]. The resin-bound protein was eluted with Elution Buffer [50 mM Tris, 500 mM NaCl, 250 mM imidazole, 10% (vol/vol) glycerol, 10 mM β-mercaptoethanol, pH 8]. The elution fractions were combined and desalted into Reaction Buffer [50 mM Tris, 50 mM NaCl, 10% (vol/vol) glycerol, 1 mM EDTA, 10 mM β-mercapto-ethanol, pH 8] and digested with SUMO protease for 1 h at room temperature or stored in protein-stabilizing mixture at −20 °C.

Protein extraction and purification reagents including HisPur Ni-NTA Resin, Pierce Zeba 7-kDa, 10-mL desalting column, Pierce DNaseI (protein extraction grade), Pierce EDTA-free protease inhibitor mixture, Pierce protein-stabilizing mixture, and Pierce 20-kDa cutoff spin concentrator were all purchased from Thermo Scientific.

**Western Blot.** The identity of the expressed fusion protein was confirmed by Western blot. A SDS/PAGE gel was incubated in Transfer Buffer [25 mM Tris, 192 mM glycine with 10% (vol/vol) aqueous methanol, pH 8.3] for 15 min before being transferred to nitrocellulose membrane (100 V, 30 min). The membrane was washed for 10 min in TBS (25 mM Tris, 150 mM NaCl, pH 7.6) and blocked for 1 h in Blocking Buffer [25 mM Tris, 150 mM NaCl, 0.1% Tween, 5% (wt/vol) nonfat dry milk powder, pH 7.6]. The membrane was washed twice in TBST (25 mM Tris, 150 mM NaCl, 0.1% Tween, pH 7.6) before incubation with 1° Antibody Solution (Blocking Buffer with 1:10,000 dilution of Anti-HisTag antibody) for 1 h with gentle agitation. The membrane was washed four times in TBST before incubation with 2° Antibody Solution (TBST with 0.01% SDS and 1:15,000 dilution of donkey anti-mouse antibody conjugated with IR Dye 800CW). The membrane was washed four times in TBST and two times in TBS and visualized on a LI-COR Odyssey imager at 700 nm (protein ladder) and 800 nm (2° Antibody).

Anti-HisTag monoclonal mouse antibody (catalog #ab5000) was purchased from AbCam; donkey anti-mouse antibody with IR Dye 800CW (catalog #926–32212) from LI-COR; and nitrocellulose membrane, prestained 10- to 250-kDa Protein Ladder, and Power Pac HC from Bio-Rad.

**Amino Acid Racemase Activity Assay.** Enzyme activity was assayed in an optimized buffer containing 50 mM Tris (pH 8), 50 mM NaCl, 10% (vol/vol) glycerol, 1 mM EDTA, and 10 mM β-mercaptoethanol. L-Ile (10 mM) was incubated with enzyme at room temperature for 0.5–2 h, depending on the analysis, and the reaction was stopped by heat inactivation at 95 °C. The amino acids were derivatized with a modified Marfey's reagent, N-α-(2,4-dinitro-5-fluorophenyl)-L-valinamide (FDVA) (Sigma Aldrich), for LC-MS analysis. Ten-microliter samples treated with 10 μL 6% (vol/vol) triethyl amine and 10 μL 0.1 mg/mL FDVA dissolved in acetone were incubated at 50 °C for 1 h, and the reaction was quenched with 5% (vol/vol) acetic acid. The reaction was dried by speedvac and dissolved in 1:1 H$_2$O:MeOH for analysis by LC-MS.

Strauch et al.

1. Yonekura-Sakakibara K, Saito K (2009) Functional genomics for plant natural product biosynthesis. *Nat Prod Rep* 26(11):1466–1487.
2. Kliebenstein DJ, et al. (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiol* 126(2):811–825.
3. Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. *Plant J* 42(5):757–771.
4. Li X, Bergelson J, Chapple C (2010) The ARABIDOPSIS accession Pna-10 is a naturally occurring *sng1* deletion mutant. *Mol Plant* 3(1):91–100.
5. Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. *Plant Cell* 13(3):681–693.
6. Kroymann J, et al. (2001) A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiol* 127(3):1077–1088.
7. Li X, et al. (2014) Exploiting natural variation of secondary metabolism identifies a gene controlling the glycosylation diversity of dihydroxybenzoic acids in Arabidopsis thaliana. *Genetics* 198(3):1267–1276.
8. Keurentjes JJ, et al. (2006) The genetics of plant metabolism. *Nat Genet* 38(7):842–849.
9. Cava F, Lam H, de Pedro MA, Waldor MK (2011) Emerging knowledge of regulatory roles of D-amino acids in bacteria. *Cell Mol Life Sci* 68(5):817–831.
10. Höltje JV (1998) Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiol Mol Biol Rev* 62(1):181–203.
11. Lam H, et al. (2009) D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science* 325(5947):1552–1555.
12. D'Aniello S, Somorjai I, Garcia-Fernàndez J, Topo E, D'Aniello A (2011) D-aspartic acid is a novel endogenous neurotransmitter. *FASEB J* 25(3):1014–1027.
13. Wolosker H (2007) NMDA receptor regulation by D-serine: New findings and perspectives. *Mol Neurobiol* 36(2):152–164.
14. Conti P, et al. (2011) Drug discovery targeting amino acid racemases. *Chem Rev* 111(11):6919–6946.
15. Buschiazzo A, et al. (2006) Crystal structure, catalytic mechanism, and mitogenic properties of *Trypanosoma cruzi* proline racemase. *Proc Natl Acad Sci USA* 103(6):1705–1710.
16. Seren Ü, et al. (2012) GWAPP: A web application for genome-wide association mapping in Arabidopsis. *Plant Cell* 24(12):4793–4805.
17. Kang HM, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4):348–354.
18. Blankenfeldt W, et al. (2004) Structure and function of the phenazine biosynthetic protein PhzF from *Pseudomonas fluorescens*. *Proc Natl Acad Sci USA* 101(47):16431–16436.
19. Alonso JM, et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301(5633):653–657.
20. Nichols MA, Waner MJ (2010) Kinetic and mechanistic studies of the deuterium exchange in classical keto–enol tautomeric equilibrium reactions. *J Chem Educ* 87(9):952–955.
21. Zenk M, Scherf H (1963) D-tryptophan in höheren pflanzen. *Biochim Biophys Acta* 71:737–738.
22. Bhushan R, Kumar V, Tanwar S (2009) Chromatographic separation of enantiomers of non-protein $\alpha$-amino acids after derivatization with Marfey's reagent and its four variants. *Amino Acids* 36(3):571–579.
23. Pillai B, et al. (2009) Crystal structure of diaminopimelate epimerase from *Arabidopsis thaliana*, an amino acid racemase critical for L-lysine biosynthesis. *J Mol Biol* 385(2):580–594.
24. Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. *Trends Plant Sci* 5(10):439–445.
25. Chen W, et al. (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet* 46(7):714–721.
26. Robinson T (1976) D-amino acids in higher plants. *Life Sci* 19(8):1097–1102.
27. Rosa N, Neish AC (1968) Formation and occurrence of N-malonylphenylalanine and related compounds in plants. *Can J Biochem* 46(8):799–806.
28. Näsholm T, Kielland K, Ganeteg U (2009) Uptake of organic nitrogen by plants. *New Phytol* 182(1):31–48.
29. Gördes D, Kolukisaoglu Ü, Thurow K (2011) Uptake and conversion of D-amino acids in *Arabidopsis thaliana*. *Amino Acids* 40(2):553–563.
30. Gördes D, Koch G, Thurow K, Kolukisaoglu U (2013) Analyses of Arabidopsis ecotypes reveal metabolic diversity to convert D-amino acids. *Springerplus* 2(1):559.
31. Gholizadeh A, Kohnehrouz BB (2009) Molecular cloning and expression in *Escherichia coli* of an active fused *Zea mays* L. D-amino acid oxidase. *Biochemistry (Mosc)*. 74(2):137–144.
32. Funakoshi M, et al. (2008) Cloning and functional characterization of *Arabidopsis thaliana* D-amino acid aminotransferase: D-aspartate behavior during germination. *FEBS J* 275(6):1188–1200.
33. Ono K, Yanagida K, Oikawa T, Ogawa T, Soda K (2006) Alanine racemase of alfalfa seedlings (*Medicago sativa* L.): First evidence for the presence of an amino acid racemase in plants. *Phytochemistry* 67(9):856–860.
34. Fujitani Y, Horiuchi T, Ito K, Sugimoto M (2007) Serine racemases from barley, *Hordeum vulgare* L., and other plant species represent a distinct eukaryotic group: Gene cloning and recombinant protein characterization. *Phytochemistry* 68(11):1530–1536.
35. Fujitani Y, et al. (2006) Molecular and biochemical characterization of a serine racemase from *Arabidopsis thaliana*. *Phytochemistry* 67(7):668–674.
36. Gogami Y, Ito K, Kamitani Y, Matsushima Y, Oikawa T (2009) Occurrence of D-serine in rice and characterization of rice serine racemase. *Phytochemistry* 70(3):380–387.
37. Michard E, et al. (2011) Glutamate receptor-like genes form Ca2+ channels in pollen tubes and are regulated by pistil D-serine. *Science* 332(6028):434–437.
38. Mutaguchi Y, Ohmori T, Wakamatsu T, Doi K, Ohshima T (2013) Identification, purification, and characterization of a novel amino acid racemase, isoleucine 2-epimerase, from Lactobacillus species. *J Bacteriol* 195(22):5207–5215.
39. Dietrich D, van Belkum MJ, Vederas JC (2012) Characterization of DcsC, a PLP-independent racemase involved in the biosynthesis of D-cycloserine. *Org Biomol Chem* 10(11):2248–2254.
40. Goytia M, et al. (2007) Molecular and structural discrimination of proline racemase and hydroxyproline-2-epimerase from nosocomial and bacterial pathogens. *PLoS One* 2(9):e885.
41. Liger D, et al. (2005) Crystal structure of YHI9, the yeast member of the phenazine biosynthesis PhzF enzyme superfamily. *Proteins* 60(4):778–786.
42. Li DM, et al. (2013) MAWBP and MAWD inhibit proliferation and invasion in gastric cancer. *World J Gastroenterol* 19(18):2781–2792.
43. Cao J, et al. (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43(10):956–963.
44. Chambers MC, et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 30(10):918–920.
45. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78(3):779–787.
46. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S (2012) CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 84(1):283–289.
47. Tartoff KD, Hobbs CA (1987) Improved media for growing plasmid and cosmid clones. *Focus* 9(2):12.

PLANT BIOLOGY