

Point mutations that boost aromatic amino acid production and CO₂ assimilation in plants

Ryo Yokoyama^{1†}, Marcos V. V. de Oliveira^{1†}, Yuri Takeda-Kimura¹, Hirofumi Ishihara^{2‡}, Saleh Alseekh², Stéphanie Arrivault², Vandna Kukshal³, Joseph M. Jez³, Mark Stitt², Alisdair R. Fernie², Hiroshi A. Maeda^{1*}

Aromatic compounds having unusual stability provide high-value chemicals and considerable promise for carbon storage. Terrestrial plants can convert atmospheric CO₂ into diverse and abundant aromatic compounds. However, it is unclear how plants control the shikimate pathway that connects the photosynthetic carbon fixation with the biosynthesis of aromatic amino acids, the major precursors of plant aromatic natural products. This study identified *suppressor of tyra2* (*sota*) mutations that deregulate the first step in the plant shikimate pathway by alleviating multiple effector-mediated feedback regulation in *Arabidopsis thaliana*. The *sota* mutant plants showed hyper-accumulation of aromatic amino acids accompanied by up to a 30% increase in net CO₂ assimilation. The identified mutations can be used to enhance plant-based, sustainable conversion of atmospheric CO₂ to high-energy and high-value aromatic compounds.

Copyright © 2022
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

INTRODUCTION

Aromatic compounds have unusual stability because of their aromaticity (i.e., electron delocalization) and therefore can serve as a critical carbon sink of the global carbon cycle to reduce the atmospheric CO₂ concentration (1). Aromatics are also major platform precursors for pharmaceuticals, commodity chemicals, and materials, with global demand growing rapidly (2); however, the chemical conversion of CO₂ to aromatics is challenging, and fossil fuels remain the prime source of aromatics (3). Plants can directly convert atmospheric CO₂, using sunlight energy, into diverse and often abundant aromatic natural products that are primarily derived from aromatic amino acids (AAAs)—L-tyrosine, phenylalanine, and tryptophan (Tyr, Phe, and Trp, respectively). Plant AAA biosynthesis represents one of the most important biochemical pathways on Earth, because the AAA-derived phenolic polymer lignin accounts for up to 30% of vascular plant dry mass, second only to cellulose in abundance (4, 5). Plants also incorporate AAAs into plant hormones, antioxidants, cofactors, and specialized metabolites, many of which are widely used for pharmaceuticals, nutraceuticals, and industrial materials in our society (6).

The shikimate pathway connects photosynthetic carbon fixation, i.e., the Calvin-Benson-Bassham (CBB) cycle, with the biosynthesis of AAAs and aromatic phytochemicals (Fig. 1A). It is currently unclear how plants control this critical gateway pathway. Many studies showed that branch point enzymes in AAA biosynthesis are biochemically regulated in plants differently from their counterparts in microbes and that these differences may stem from the diverse biosynthetic use of AAAs in plants (7–9). To understand the shikimate pathway regulation, prior studies conducted genetic screenings of resistant plants to a shikimate pathway inhibitor, glyphosate, and toxic AAA analogs; however, these studies were either unsuccessful

or identified mutations in genes encoding 5-enolpyruvylshikimate-3-phosphate synthase, the glyphosate target, or branch point enzymes specific to AAA biosynthesis (10–14) rather than enzymes that regulate carbon flux through the entire shikimate pathway.

RESULTS

Suppressor of tyra2 identified dominant mutations targeting the entry step in the shikimate pathway

Here, we conducted genetic screening to isolate suppressors of the *Arabidopsis thaliana* *tyra2* knockout mutant that lacks one of two *TyrA* genes involved in Tyr biosynthesis (Fig. 1A) and that exhibits compromised growth and reticulated leaf phenotypes (Fig. 1B) (15). Roughly 10,000 *tyra2* seeds were mutagenized using ethyl methanesulfonate (EMS) and grown in eight separate pools (A to H). More than 10,000 M₂ seeds were harvested from each pool and screened for recovery of growth and/or the reticulated leaf phenotypes of *tyra2*. We isolated a total of 351 *suppressor of tyra2* (*sota*) mutants. Some lines (e.g., *sotaA4* and *sotaH1*) recovered both growth and reticulate phenotypes, whereas others (e.g., *sotaB4* and *sotaG1*) recovered only growth but remained reticulate (Fig. 1B and fig. S1) despite maintaining *tyra2* deficiency (fig. S2). When the overall profile of soluble metabolites was analyzed by gas chromatography–mass spectrometry (GC-MS) in 40 representative *sota* mutants, selected from different pools on the basis of a range of visible phenotypes, 21 lines showed elevated Tyr and Phe levels (Fig. 1C). As these Phe and Tyr levels highly positively correlated to each other ($R^2 = 0.929$) (Fig. 1D), these *sota* mutants likely affected the upstream synthesis of arogenate, the common substrate of Phe and Tyr, from the shikimate pathway (Fig. 1A). We designated them as “metabolic” *sota* mutants and focused our further study on them.

For genetic mapping, eight representative lines (*sotaA4*, *sotaA11*, *sotaB3*, *sotaB4*, *sotaF1*, *sotaG1*, *sotaH1*, and *sotaH9*) were backcrossed with the original *tyra2* mutant. Illumina whole-genome sequencing of *tyra2*-like and/or *sota*-like F₂ progenies identified high-frequency missense mutations from all eight lines in At4g39980, At4g33510, or At1g22410, which are the three loci encoding 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAHP synthase or DHS)

¹Department of Botany, University of Wisconsin-Madison, Madison, WI, USA.

²Max-Planck-Institut für Molekulare Pflanzenphysiologie, Potsdam-Golm, Germany.

³Department of Biology, Washington University in St. Louis, St. Louis, MO, USA.

*Corresponding author. Email: maeda2@wisc.edu

†These authors contributed equally to this work.

‡Present address: University of Helsinki, P.O. 65 (Viikinkaari 1), 00014 Helsinki, Finland.

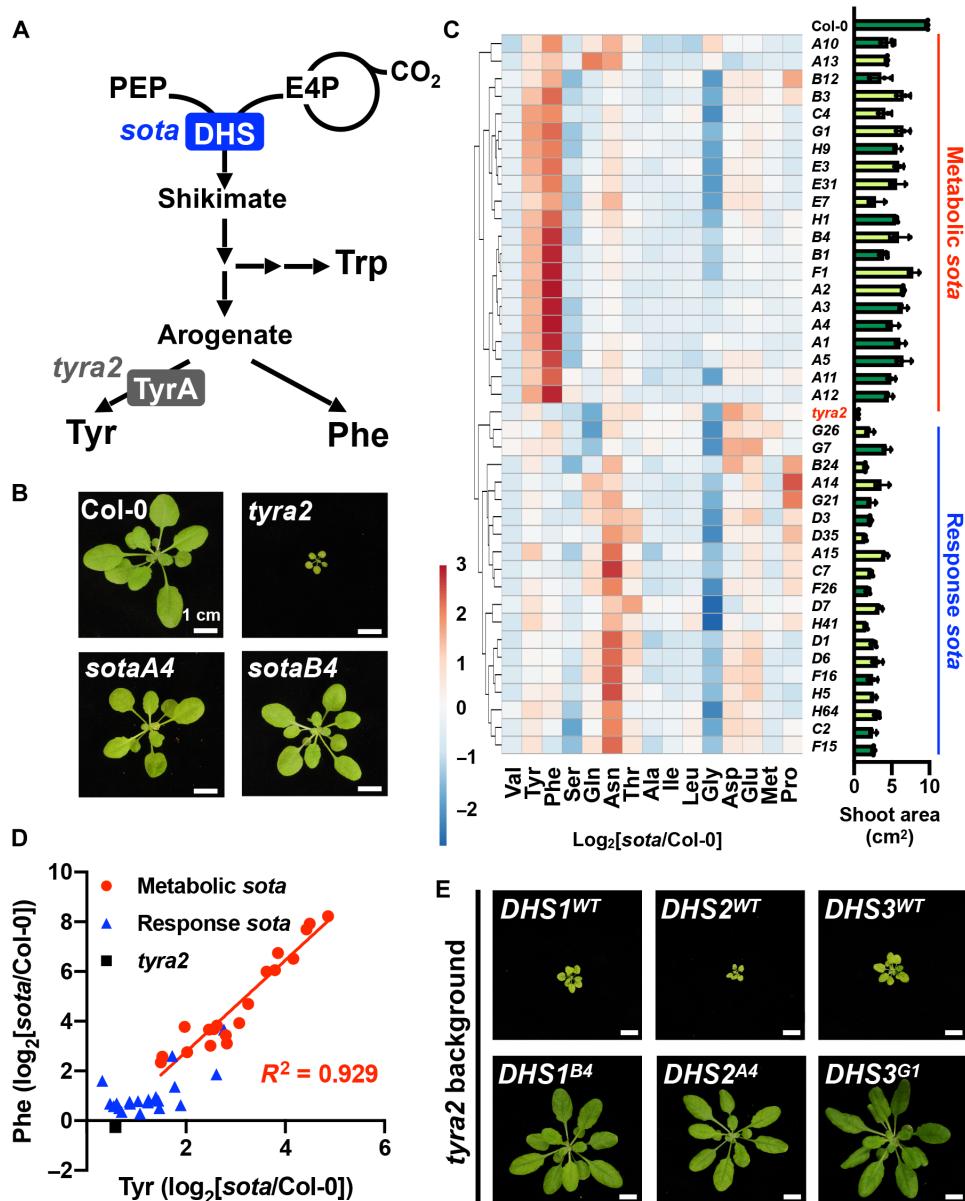


Fig. 1. Multiple suppressor of *tyra2* (*sota*) mutations rescued the *tyra2* growth inhibition and enhanced tyrosine (Tyr) and phenylalanine (Phe) accumulation. (A) A simplified diagram of the shikimate and AAA biosynthetic pathways. DHS, 3-deoxy-D-arabino-heptulonate 7-phosphate synthase; E4P, erythrose-4-phosphate; PEP, phosphoenolpyruvate; TyrA, TyrA arogenate dehydrogenase. (B) Plant pictures of 4-week-old *Col-0* wild-type (WT), *tyra2*, and two representative *sota* mutants of *Arabidopsis thaliana*. The remaining *sota* mutant plants are shown in fig. S1. (C) Soluble metabolite profiling and shoot area of the 3-week-old *Col-0*, *tyra2*, and *sota* mutants. Dark and light green bars represent that each *sota* mutant line showed *Col-0*-like fully mature green leaves and *tyra2*-like reticulated leaves, respectively. All the metabolic *sota* mutants exhibited significantly larger shoot area than *tyra2* [one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test, $P < 0.001$]. Data are means \pm SEM ($n = 4$ independent plant samples). (D) Relative amounts of Tyr and Phe against *Col-0* shown in (C) were plotted for metabolic *sota* (red circles), response *sota* (blue triangles), and *tyra2* (black square). (E) Plant pictures of representative complementation lines at T₂ generation that were generated by introducing either WT DHS (e.g., *DHS1^{WT}*) or *sota*-mutated DHS (e.g., *DHS1^{B4}*) genes, driven by the respective endogenous promoter, into the *Arabidopsis* *tyra2* background. Scale bars, 1 cm. The remaining lines are shown in fig. S6.

isoforms (fig. S3 and table S1). DHS catalyzes the first reaction in the shikimate pathway (Fig. 1A). The *DHS sota* mutations segregated with the *tyra2* suppression phenotypes as confirmed by derived cleaved amplified polymorphic sequence (dCAPS) marker genotyping in representative F₂ (fig. S4). Their F₂ population showed dominant or semidominant characteristics in terms of their growth

recovery and Tyr and Phe accumulation phenotypes (fig. S5). One exception is that the *sotaF1* homozygous, but not heterozygous, plants exhibited dwarfism likely because of its extreme accumulation of AAAs (fig. S5A). Transgenic expression of DHS genes with a *sota* mutation (e.g., *DHS1^{B4}*), but not the corresponding wild-type (WT) DHS genes (e.g., *DHS1^{WT}*), driven by their respective endogenous

promoters, in the *tyra2* mutant recovered its dwarf plant and reticulated leaf phenotypes (Fig. 1E and fig. S6A) and also led to elevated Phe and Tyr levels (fig. S6B), phenocopying the metabolic *sota* lines (Fig. 1). These results provide genetic evidence that these DHS *sota* mutations suppress the *tyra2* phenotype and enhance Tyr and Phe levels in a dominant fashion.

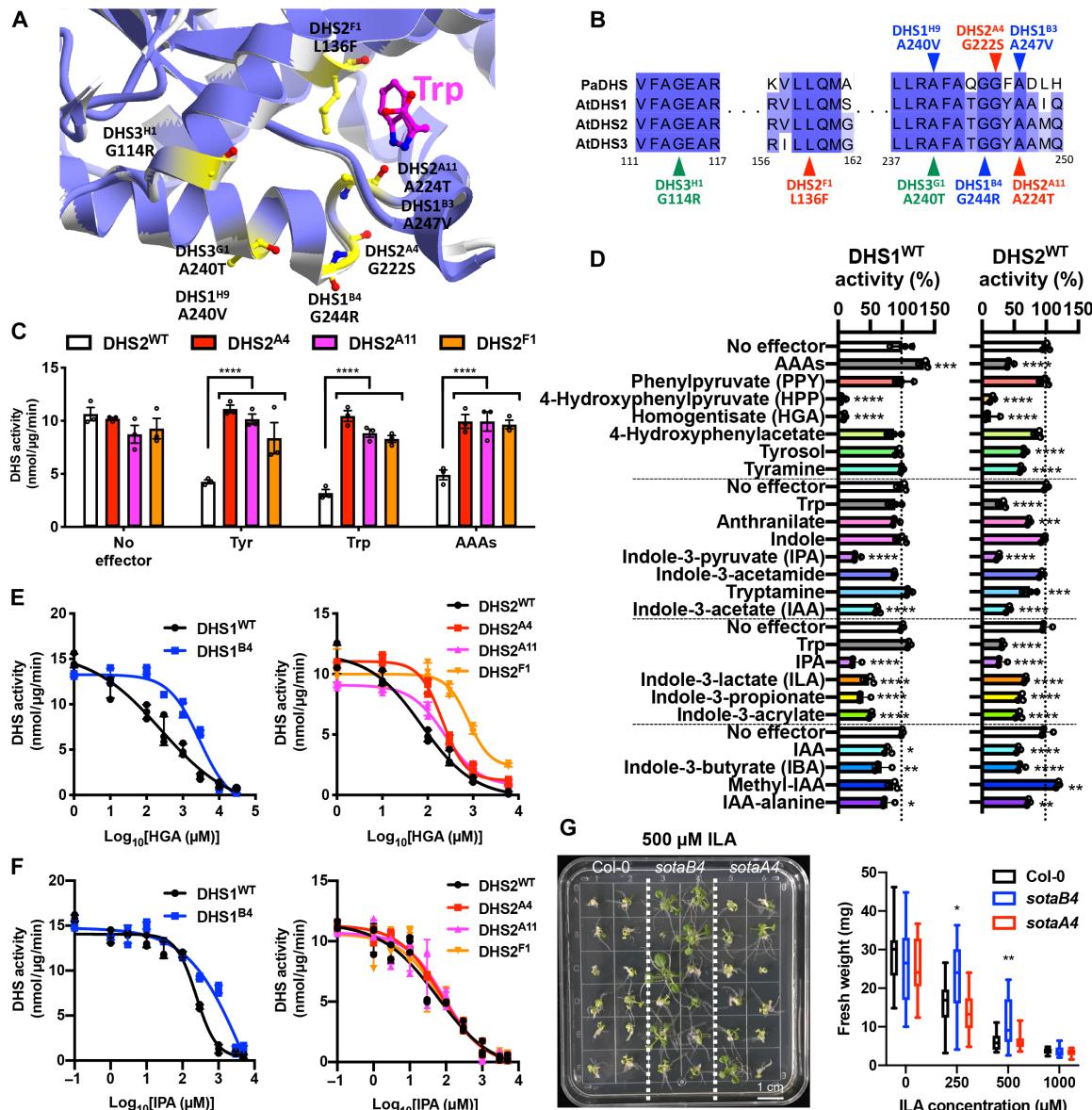


Fig. 2. The sota mutations biochemically deregulate the effector-mediated DHS negative feedback inhibition. (A) A structural model of *A. thaliana* DHS2 (AtDHS2, purple) generated from the *P. aeruginosa* DHS (PaDHS, white) with Trp (magenta) bound. Residues corresponding to the sota mutations mapped onto the AtDHS2 model are highlighted in yellow. The entire model is shown in fig. S7. (B) Selected regions of the amino acid sequence alignment of PaDHS and AtDHS enzymes, with the positions of the sota mutations indicated by blue, red, green arrows for AtDHS1, AtDHS2, and AtDHS3, respectively. The entire alignment is shown in fig. S8. (C) Enzymatic assay of DHS2 WT (DHS2^{WT}) and DHS2 with a sota mutation (DHS2^{A4}, DHS2^{A11}, and DHS2^{F1}) in the presence of Tyr, Trp, or mixture of all AAAs at 1 mM. ***P < 0.0001; significant differences by one-way ANOVA with Dunnett's multiple comparisons test against the corresponding DHS2^{WT} samples. Data are means \pm SEM (n = 3). (D) Screening of AAAs and AAA-derived metabolites as potential inhibitors of DHS1^{WT} and DHS2^{WT}. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, and ****P ≤ 0.0001 denote significant differences by one-way ANOVA with Dunnett's test against the corresponding "No effector" samples. Data are means \pm SEM (n = 3). The dotted horizontal lines separate four sets of independent experiments. (E and F) IC₅₀ curves of WT and sota mutant enzymes of DHS1 (left) and DHS2 (right) with varied concentrations of HGA (E) and indole-3-pyruvate (IPA) (F). Data are means \pm SEM (n = 3). (G) Plant picture (left) and fresh weight measurement (right) of 3-week-old Col-0, sotaB4, and sotaA4 mutants (Col-0 background) on the media containing ILA at 0, 250, 500, or 1000 μM. *P ≤ 0.05 and **P ≤ 0.01 denote significant differences by one-way ANOVA with Dunnett's test against the corresponding Col-0 samples (n = 12 to 16 independent plant samples).

protein structure (16). Introduction of the *sota* mutations to recombinant DHSs did not alter overall catalytic activity (Fig. 2C and fig. S9A). In addition, DHS transcript levels were unchanged in the *sota* mutants (fig. S10). We therefore hypothesized that the *sota* mutations might affect DHS enzyme regulation. We showed previously that Tyr and Trp inhibit *Arabidopsis* DHS2 but not the DHS1 or DHS3 isoforms (9). Chorismate and caffeate strongly inhibit all *Arabidopsis* DHS isoforms, whereas shikimate, prephenate, and arogenate slightly inhibit DHS2 (9). Here, we found that the DHS2 enzyme with the *sotaA4* mutation (DHS2^{A4}) is still inhibited by shikimate, prephenate, and arogenate as well as chorismate and caffeate, with similar median inhibitory concentration (IC₅₀) values to the corresponding DHS2 WT enzyme (DHS2^{WT}) (fig. S11 and table S2). Unlike DHS2^{WT}, the activity of the DHS2^{A4}, DHS2^{A11}, and DHS2^{F1} mutants was not inhibited by Tyr, Trp, or AAA mixtures at a concentration of up to 1 mM (Fig. 2C). Both structural docking simulation and differential scanning fluorimetry suggested that the DHS2^{A4} mutant enzyme still binds Tyr and Trp (fig. S12). These *sota* mutations on DHS2 therefore completely eliminate its sensitivity to Tyr and Trp, likely without altering their binding to the protein.

In contrast to DHS2, DHS1 is not inhibited by AAAs (9), and this was also the case for the DHS1^{B4} mutant enzyme (fig. S9A). We therefore hypothesized that the *sotaB4* mutation on DHS1 might eliminate inhibition by chorismate and caffeate. DHS1^{B4} was, however, still strongly inhibited by these effectors with comparable IC₅₀ values to that in DHS1^{WT} (fig. S9B and table S2). To explore how the *sotaB4* mutation affects DHS1 functionality, we further screened for additional aromatic compounds downstream of AAAs that might inhibit DHS1 and DHS2 enzymes (Fig. 2D and fig. S13A). Tyrosol and tyramine modestly inhibited DHS2 by ~30% at 1 mM (Fig. 2D). 4-Hydroxyphenylpyruvate (HPP) and homogentisate (HGA), but neither phenylpyruvate (PPY) nor 4-hydroxyphenylacetate, effectively inhibited all DHS^{WT} isoforms (Fig. 2D) with IC₅₀ of 75 to 250 μ M for HGA (Fig. 2E, fig. S13B, and table S2). Notably, nearly all of the *sota* mutants showed substantially higher IC₅₀ for HPP and HGA (up to 7- and 12-fold increase, respectively) than the corresponding WT enzymes. The one exception was DHS2^{A11}, the weakest *sota* allele, which showed no significant change in IC₅₀ with HPP and HGA (Fig. 2E, fig. S13B, and table S2). Thus, Tyr-derived compounds can also effectively inhibit the three DHS isoforms of *Arabidopsis*, and the *sota* mutations weaken this regulation.

Further screening found that Trp-derived indole-3-pyruvate (IPA), the immediate precursor of the plant hormone indole-3-acetate (IAA; auxin), and, to a lesser extent, IAA itself inhibit both DHS1 and DHS2 (Fig. 2D) with IC₅₀ of 241 and 58 μ M, respectively, for IPA (Fig. 2F and table S2). Indole-3-lactate (ILA), but not indole-3-acetamide, also reduced the activity of both DHS1 and DHS2, with ILA having a similar inhibitory effect as IPA (Fig. 2D, fig. S13B, and table S2). Anthranilate and indole, intermediates of Trp biosynthesis, as well as tryptamine did not affect the activity of DHS1, although DHS2 showed a slight reduction in the presence of anthranilate and tryptamine (Fig. 2D). The IPA-mediated inhibition was attenuated in the *sota* mutants of DHS1 and DHS3 (e.g., DHS1^{B4} and DHS3^{G1}), but not DHS2 (e.g., DHS2^{A4}), having 4 to 20 times higher IC₅₀ than corresponding WT (Fig. 2F, fig. S13B, and table S2). Similarly, DHS1^{B4} had a higher IC₅₀ than DHS1^{WT} for ILA but not for indole-3-propionate (fig. S13B and table S2). Although IPA impaired plant growth, independent of genotypes, even

at very low concentration (<10 μ M) likely due to its conversion to the plant hormone auxin, ILA feeding led to growth inhibition of *Arabidopsis* Col-0 WT plants, and the inhibition was significantly weakened in the *sotaB4* mutant plants (Fig. 2G). These *in vitro* and *in vivo* data together indicate that *Arabidopsis* DHS enzymes are inhibited by Trp-derived indolic compounds, an inhibition that is attenuated by the *sota* mutations of DHS1 and DHS3.

***sota* mutations deregulate the shikimate pathway and elevate AAAs**

To directly test whether the relaxed feedback regulation of DHS enzymes by the *sota* mutations indeed increase the shikimate pathway activity in plants, *Arabidopsis* Col-0 and the *sotaB4* and *sotaA4* mutant plants were fed with stable isotope-labeled ¹³CO₂ in the light for 6 hours from the beginning of the day. The following time course metabolite analyses showed that the ¹³C label was gradually incorporated into various metabolites (fig. S14). Compared to Col-0, the *sotaB4* and *sotaA4* mutants accumulated much higher levels of ¹³C-labeled shikimate and AAAs (Fig. 3A), but not other amino acids, with the exception of glycine that displayed slightly lower ¹³C incorporation in both *sota* lines than Col-0 (fig. S14). Similar results were obtained for 3-hour ¹³CO₂ labeling at the end of the day (fig. S15). These labeling studies are consistent with the GC-MS profiling of the overall metabolite pools of 21 metabolic *sota* mutants that revealed a large increase in AAAs and a slight reduction in glycine, but little change in other amino acids (Fig. 1C). These results indicate that the *sota* mutations specifically increased carbon flux through the shikimate pathway toward the biosynthesis of all three AAAs in planta.

To further assess the impacts of the *sota* mutations on AAA and AAA-derived metabolites, we conducted targeted metabolite profiling using GC-MS and liquid chromatography (LC)-MS. First, we generated the *sotaB4* and *sotaA4* mutants in the *Arabidopsis* Col-0 background by outcrossing to Col-0. Overall, these plants were indistinguishable from Col-0 in terms of their growth and seed yield (Table 1 and fig. S16). The comparison of AAA profiles of the *sota* mutants in the Col versus *tyra2* backgrounds revealed that the presence of the *tyra2* mutation increased Phe and Trp levels, resulting in elevated Phe/Tyr and Trp/Tyr ratios without altering the Phe/Trp ratio (fig. S17). However, the levels of all three AAAs remained elevated in the *sota* mutants in the Col-0 background (Fig. 3B), which were therefore used for the following analyses to eliminate the effects of the original *tyra2* mutation.

The levels of HGA and α - and γ -tocopherols derived from Tyr were, like Tyr, also elevated in both *sotaB4* and *sotaA4* mutants (Fig. 3B and table S3). In contrast, the levels of Trp-derived indole glucosinolates, such as indolyl-3-methyl glucosinolate (I3M), were not elevated in the *sota* lines (Fig. 3B and table S3). Similarly, the *sota* mutants and Col-0 had comparable levels of sinapate, sinapoylmalate, and flavonoids, including kaempferol-3-O-(2'-O-rhamnosyl) glucoside-7-O-rhamnoside (K3GR7R), which are phenylpropanoid compounds produced via Phe deamination catalyzed by Phe ammonia lyase (PAL) (Fig. 3B and table S3) (17, 18). ¹³C labeling of I3M, sinapoylmalate, and K3GR7R was also not increased within 6 hours of ¹³CO₂ labeling in the *sota* mutants compared to Col-0 (fig. S14). The overall lignin deposition, based on phloroglucinol staining and thioglycolic acid analyses, was unaltered in *sotaB4* and *sotaA4* mutants (fig. S18), unlike ectopic lignin accumulation previously observed in some *Arabidopsis* transgenics (19). After high

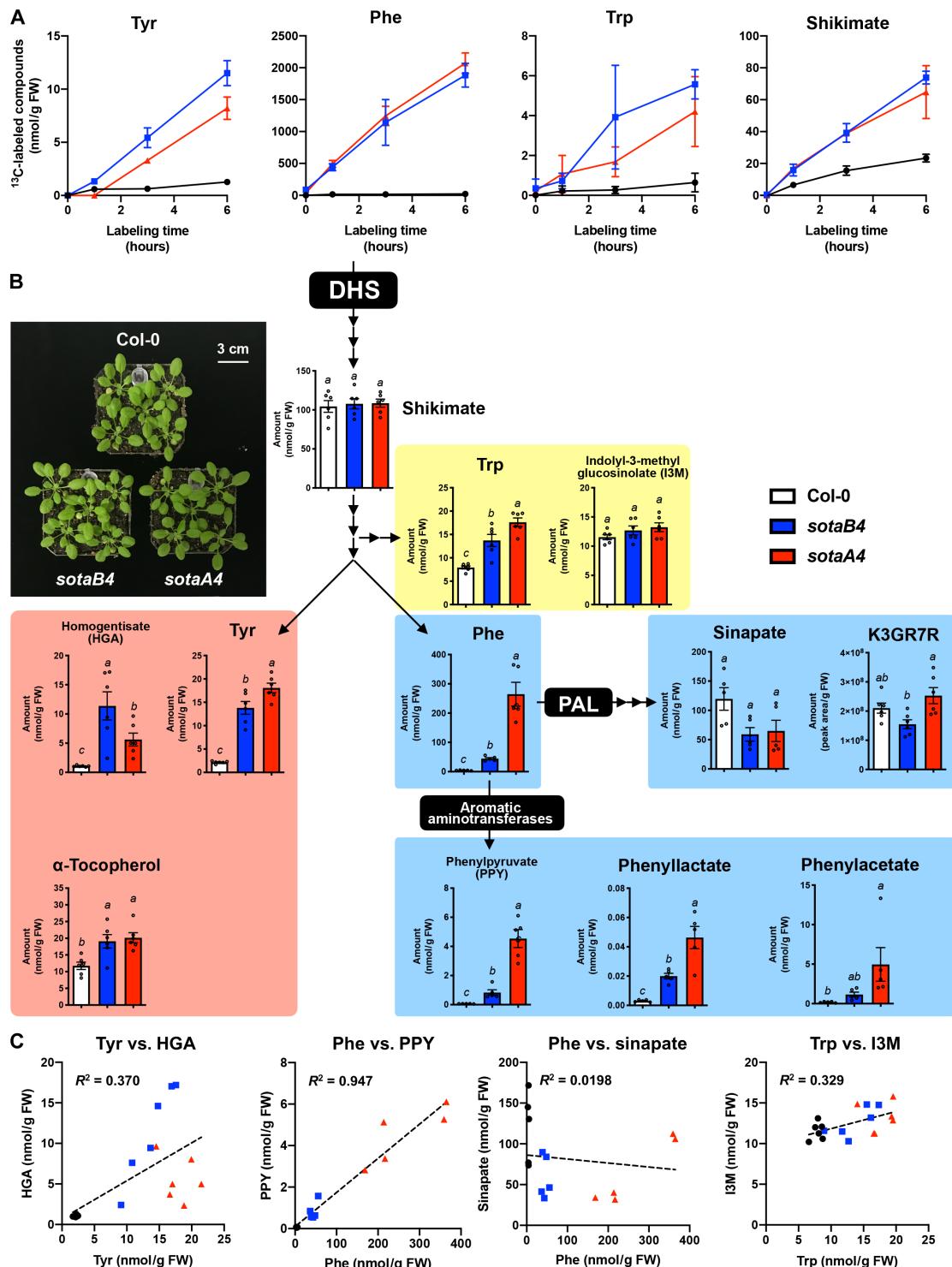


Fig. 3. Increased carbon flux elevates the levels of AAAs but not all AAA-derived compounds in the *sota* mutants. (A) $^{13}\text{CO}_2$ labeling experiment of Col-0, *sotaB4*, and *sotaA4* (tyra2 background), followed by quantification of ^{13}C -labeled Tyr and Phe by GC-MS and ^{13}C -labeled Trp and shikimate by liquid chromatography (LC)-MS. Data are means \pm SEM ($n = 3$ independent biological samples except for 0-hour time point having two replicates). (B) Targeted metabolomics analysis of AAAs and AAA-derived metabolites in 4-week-old Col-0, *sotaB4*, and *sotaA4* (Col-0 background) grown on soil (also see fig. S17 for data of the *sota* mutants in the *tyra2* background). Actual values are shown in table S3. Different letters indicate statistically significant differences among genotypes (one-way ANOVA with Tukey-Kramer test, $P < 0.05$). Data are means \pm SEM ($n = 5$ to 6 independent plant samples). K3GR7R, kaempferol-3-O-(2'-O-rhamnosyl)glucoside-7-O-rhamnoside; PAL, Phe ammonia lyase. (C) The correlations between the levels of AAAs and their representative derivatives shown in (B). The correlations of Phe versus phenyllactate or phenylacetate are shown in fig. S21.

Table 1. Parameters of growth, protein content, and photosynthesis. V_{cmax} , J_{max} , and R_{d} values, derived from the A - C_i curves in Fig. 4E, represent the maximum rate of Rubisco carboxylation activity, the potential rate of electron transport, and the rate of mitochondrial dark respiration, respectively. The initial slope and CO_2 compensation point (CCP) of the light response curves and A - C_i curves were determined using the first three and five points at low light and low C_i points, respectively (Fig. 4, D and E). Different letters (a and b) indicate statistically significant differences among genotypes (one-way ANOVA with Tukey-Kramer test, $P < 0.05$). Data are means \pm SEM ($n = 8$ independent plant samples for the growth and protein data and $n = 5$ to 6 for the photosynthetic parameters). FW, fresh weight; RbcL, Rubisco large subunit.

	Col-0	sotaB4	sotaA4
Individual shoot FW (g FW)	0.23 ± 0.0051^a	0.20 ± 0.010^a	0.24 ± 0.0090^a
Shoot area/shoot FW (cm 2 /g FW)	0.018 ± 0.000076^a	0.018 ± 0.00022^a	0.018 ± 0.00015^a
Total protein amount (mg/g FW)	9.11 ± 0.22^a	9.45 ± 0.19^a	8.86 ± 0.21^a
Relative RbcL amount	1.00 ± 0.018^a	0.97 ± 0.018^a	0.92 ± 0.027^a
V_{cmax} (μmol/m 2 per second)	18.57 ± 0.34^b	29.06 ± 0.89^a	27.46 ± 0.80^a
J_{max} (μmol/m 2 per second)	61.14 ± 0.61^b	83.51 ± 1.86^a	79.82 ± 1.71^a
R_{d} (μmol/m 2 per second)	1.80 ± 0.10^b	2.37 ± 0.10^a	2.67 ± 0.081^a
Initial slope	0.04 ± 0.0010^b	0.06 ± 0.0013^a	0.06 ± 0.0023^a
CCP (μmol/m 2 per second)	133.00 ± 3.81^a	121.96 ± 1.85^a	133.48 ± 1.18^a

light stress, which promotes the production of numerous AAA-derived compounds, Phe-derived phenylpropanoids, such as anthocyanins and K3GR7R, and Trp-derived compounds were elevated similarly between genotypes, despite all AAA levels being always higher in *sotaB4* and *sotaA4* than Col-0 (fig. S19 and table S4). AAA and shikimate levels were also elevated in shoots and roots of plate-grown *sotaB4* and *sotaA4* mutants, with the one exception of *sotaA4* roots (fig. S20 and table S5), possibly because of their isoform-specific functions (9). Again, the levels of these phenylpropanoids and Trp-derived metabolites were not significantly different between genotypes (fig. S20 and table S5). These results are consistent with the presence of multiple layers of tight regulations in the plant phenylpropanoid and indole metabolic network, which include both transcriptional and posttranscriptional regulations (18, 20).

Further careful comparisons of GC-MS traces between genotypes revealed that a few previously unidentified peaks appeared in both *sotaB4* and *sotaA4* mutants but not in Col-0 samples; on the basis of the National Institute of Standards and Technology library search and subsequent comparison to respective authentic standards, these peaks were identified as PPY, the keto acid of Phe produced by aromatic aminotransferases (21–23), as well as phenylacetate and phenyllactate, which are both likely derived from PPY (Fig. 3B) (24, 25). Notably, the levels of PPY and PPY derivatives, detected in *sota* mutants, positively correlated with the Phe level (Fig. 3C, fig. S21, and table S3). When the Col-0 WT plant was transformed with the *DHS1^{B4}* or *DHS2^{A4}* genes, but not their WT genes or empty vector control, the levels of AAAs were also elevated (fig. S22), as seen in *sota* mutants. Moreover, in these transgenic plants, the levels of Phe positively correlated with those of PPY and PPY-derived compounds without significant changes in the levels of phenylpropanoids, such as sinapate and K3GR7R (fig. S22 and table S6). Thus, the transgenic expression of a *sota*-mutated *DHS* gene, even in the presence of endogenous WT *DHS* genes, leads to elevated accumulation of all three AAAs and specific downstream products (e.g., HGA and PPY) in planta.

Deregulating the shikimate pathway enhances CO_2 assimilation

DHS uses two substrates, phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) that are directly supplied from glycolysis and the CBB cycle, respectively (Fig. 1A) (6, 26). We tested whether the markedly elevated AAA production in the *sota* mutants is supported by either starch or sugar storage pools by analyzing their levels during a day and night cycle. Compared to Col-0 WT, the *sotaB4* and *sotaA4* mutants had larger pools of Tyr and Phe and similar pools of Trp and shikimate at dawn. Throughout the day, the mutants increased the levels of Tyr, Phe, Trp, and shikimate up to 7.6-, 18-, 2.9-, and 2.4-fold higher levels than Col-0, respectively, by the end of the day. These large metabolite pools were then decreased during the night (Fig. 4A). Amounts of starch and soluble sugars, including sucrose and glucose, rose and declined during the day and night, respectively. Despite a trend to higher dusk starch levels, however, carbohydrates were not significantly different in *sotaB4* and *sotaA4* compared to Col-0 at all time points (Fig. 4, B and C).

To further test potential impacts of the *sota* mutations on photosynthetic carbon fixation, net CO_2 assimilation rates (A) in response to different light intensities were analyzed by measuring the gas exchange activity of Col-0, *sotaB4*, and *sotaA4* plants. Plants of both *sota* mutants exhibited significantly higher A levels at all light intensities at and above 100 microeinstein (μE), the growth light condition used in this study, and eventually reached a plateau to an approximately 30% higher assimilation than Col-0 (Fig. 4D). When A was analyzed under different intercellular CO_2 concentrations (C_i), the *sota* mutants exhibited up to 30% higher A than Col-0 especially at increased C_i (Fig. 4E). Although total protein and Rubisco contents were unaltered in the *sota* mutants (Table 1), the V_{cmax} values of both *sota* mutants were 50% higher than that of Col-0, suggesting that the carboxylation activity of Rubisco was elevated in the *sota* mutants. The CO_2 compensation point was comparable between genotypes, but R_{d} values, which represent dark respiration, were elevated in the *sota* mutants, which may further support

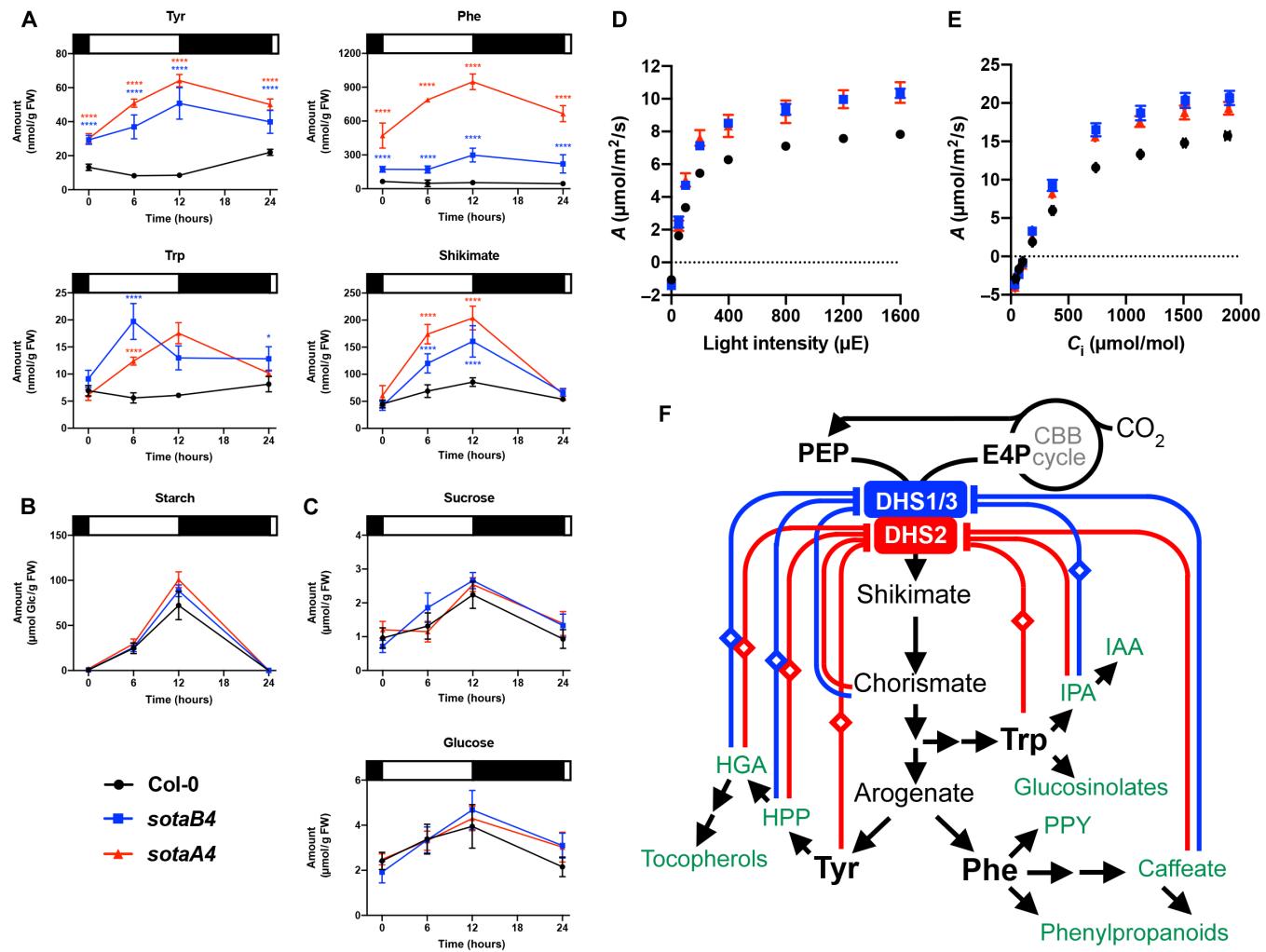


Fig. 4. Carbon fixation is accelerated to support high AAA production in the *sota* mutants. (A to C) The levels of AAA and shikimate (A), starch (B), and glucose and sucrose (C) of Col-0, *sotaB4*, and *sotaA4* (Col-0 background) harvested at the indicated time points under the 12-hour light /12-hour dark cycle (white/black bars above each graph). Starch is expressed as micromoles of glucose (Glc) equivalents per gram FW. * $P \leq 0.05$ and *** $P \leq 0.0001$; significant differences by one-way ANOVA with Dunnett's multiple comparisons test against the corresponding Col-0 samples. Data are means \pm SEM ($n = 4$ to 6 independent plant samples). (D and E) The response curves of CO_2 assimilation rate (A) to light intensity and CO_2 concentration in intercellular air spaces (C_i) of *Arabidopsis* Col-0 and the *sota* mutants (Col-0 background). Data are means \pm SEM ($n = 5$ to 6 independent plant samples). (F) The *sota* mutations eliminate or attenuate feedback regulation by certain effector molecules (open diamonds) of AtDHS1/3 and AtDHS2 (blue and red lines, respectively).

production of energy-intensive AAA biosynthesis (Table 1) (27). The enhanced CO_2 assimilation was similarly observed in the transgenic lines of the Col-0 background expressing the mutated *DHS1^{B4}* or *DHS2^{A4}* genes but not the WT *DHS* genes (fig. S23 and table S7). These results revealed that deregulation of the shikimate pathway by the *sota* mutations is accompanied by increased activity of carbon fixation.

DISCUSSION

The DHS-catalyzed reaction has been assumed to be important for the regulation of plant shikimate pathway on the basis of prior microbial studies (26, 28) and expression of deregulated microbial DHS expression in plants (29–31). Our study provides strong genetic evidence supporting this notion, as all eight metabolic *sota* mutations examined here mapped to the loci encoding DHSs, but

not to other shikimate pathway enzymes. Unlike microbial DHSs that are directly inhibited by the pathway product, AAAs, this study found that plant DHSs are subjected to highly complex feedback regulation mediated not only by AAAs but also by many AAA-derived compounds (Fig. 4F). The identified *sota* mutations relax DHS feedback inhibition without affecting the effector binding per se (fig. S12), similarly to that in a recently reported analogous mutation in *Mycobacterium tuberculosis* DHS (32). As no significant conformational change was observed in the protein structure (32), the molecular basis of how these mutations deregulate the feedback inhibition in microbial and plant DHS enzymes remains unknown. Although the *sota* mutations either abolished or attenuated DHS regulation by multiple effectors and we cannot pinpoint to a specific molecule, the degree of HGA and HPP inhibition (Fig. 2E and fig. S13B) inversely correlated with that of AAA accumulation among different DHS2 *sota* lines (figs. S5 and S6). Thus, plant DHS monitors the levels of multiple

downstream AAA-derived compounds and plays crucial roles in controlling the shikimate pathway and AAA production in plants. The dominant nature of the *sota* mutations (Fig. 1 and figs. S4 to S6 and S22) provides us ways to overcome the negative regulation of endogenous DHS enzymes by expressing an *Arabidopsis DHS sota* gene (e.g., *DHS1^{B4}*) in other plants in a specific tissue or developmental stage.

Elevated CO₂ assimilation observed in the *sota* mutants was notable and is likely important for efficient supply of E4P (Fig. 4F). This also agrees with prior reports that plant DHSs have high *K_m* for E4P (9) and that transketolase activity, which produce E4P in the CBB cycle, is important for AAA production in plants (33). Unlike *Arabidopsis* transgenics overexpressing CBB pathway enzymes that had elevated CO₂ assimilation and increased biomass (34), the *sota* mutations did not significantly alter plant biomass (Table 1 and fig. S16). Instead, the increased photosynthesis observed in the *sota* mutant plants (Fig. 4E, Table 1, and fig. S23) would provide additional energy to support the elevated activity of the highly energy-intensive shikimate pathway and AAA biosynthesis (27). Although the exact mechanism of the elevated CO₂ assimilation is currently unknown, a rapid use of E4P might alleviate negative regulations of the CBB cycle in the *sota* lines (35, 36). Notably, unlike the AAA imbalances and compromised growth caused by deregulation of a specific AAA biosynthetic branch (13, 15), the *sota* mutations had limited impacts on overall DHS activity in the absence of effectors (Fig. 2C and figs. S9 and S11) and overall plant growth (Table 1 and figs. S16 and S22). In addition, these *sota* mutations occur in well-conserved amino acid residues of DHSs among different plants, including important agricultural and bioenergy crops (e.g., maize and sorghum; figs. S24 and S25), and hence can be directly introduced in crops by gene editing (37). Thus, a series of the DHS point mutations identified in this study can provide powerful plant genetic tools to enhance the conversion of CO₂ into aromatic compounds for sustainable production of high-value chemicals while concomitantly reducing atmospheric CO₂.

MATERIALS AND METHODS

Plant materials

A. thaliana plants used in this study were grown under a 12-hour/12-hour 100-μE light/dark cycle with 85% air humidity in soil supplied with Hoagland solution or on the agarose-containing 0.5-strength Murashige and Skoog medium with 1% sucrose unless stated otherwise.

Screening of *sota*

The seeds of the *tyra2-1* transfer DNA insertion mutant (SALK_001756), previously characterized and confirmed as null homozygous with dwarf and reticulate phenotype (15), were used to conduct a forward genetic suppressor screening using EMS, following a method by Weigel and Glazebrook (38) with a few modifications. Briefly, ~10,000 *tyra2* homozygous seeds were mutagenized with 0.2% EMS (M0880, Sigma-Aldrich) for 15 hours in a 50-ml Falcon tube using a rocking platform. Seeds were rinsed with ultrapure water 10 times and soaked in the water for 1 hour. Subsequently, seeds were suspended in 400-ml 0.1% agarose and spread on eight different trays (~50 ml on each tray; the 1020 tray; CN-FLXHD, Greenhouse Megastore, Danville) containing germination soil mix (8269028, Sun-gro). Eight M₁ pools from different trays were named with alphabet letters (A to H), and each pool contained approximately 1000 M₁ plants. Mutagenesis efficiency was calculated as described previously (38), applying the Poisson distribution. Observation of siliques

from 50 M₁ plants identified 15 plants without aborted seeds, indicating that the mutagenesis was successful. M₂ screening was performed by germinating ~10,000 seeds from each M₁ pool on 10 trays containing the germination mix. A total of ~80,000 M₂ seeds were germinated in 80 trays. Phenotypes were evaluated at 4 to 5 weeks after germination. Col-0 and *tyra2-1* were germinated side by side with EMS mutants in each tray for comparison. Plants showing the *tyra2*-like dwarf and reticulate leaf phenotypes were removed, while ones showing any recovery on either one or both of the *tyra2* phenotypes were kept. The recovered plants were named as *sota* followed by a letter A to H for the isolated pool and then a number: For example, the *sotaB4* line was the fourth suppressor recovered from the pool B. Each M₂ *sota* line was allowed to self-fertilize, and the resulting M₃ seeds were collected for further experiments.

Whole-genome sequencing-based mapping of *sota* mutations

To identify the causal mutations leading to the suppression of the *tyra2* phenotypes and the accumulation of AAAs in the metabolic *sota* lines, the M₃ plants of *sotaA4*, *sotaA11*, and *sotaB4* were first backcrossed with *tyra2*, while the remaining lines were mapped later (see below). The F₁ population also showed the *tyra2* recovery phenotype, indicating that all three *sota* mutations had semidominant or dominant characteristics, with the F₁ plants of *sotaB4* being almost indistinguishable from its M₃ plants. As expected, roughly one-quarter of F₂ segregating populations showed the *tyra2*-like phenotypes (fig. S4). For genetic mapping, roughly 200 seedlings showing *tyra2*-like and *sota*-like phenotypes were separately harvested and pooled, and their genomic DNA was isolated using the DNeasy Plant Mini Kit (QIAGEN) according to the manufacturer's protocol. Subsequently, these six DNA samples were submitted to the University of Wisconsin Biotech Center for DNA library preparation, barcoding, and whole-genome sequencing by a 100-base pair single-end reading using the Illumina HiSeq 2500. To look for causal mutations that are present only in the *sota*-like F₂ population but not in the *tyra2*-like F₂ population, the sequencing data for both populations were analyzed using CLC Genomics Workbench 11.0.1 (QIAGEN). All single-nucleotide variants (SNVs) were first obtained by comparing each sequencing result to the TAIR10 reference genome. SNVs identified on the *tyra2*-like population were then subtracted from SNVs identified on the *sota*-like population. The remaining SNVs on the *sota*-like population were plotted on the basis of their frequency among obtained reads (y axis) and genome position (x axis) (fig. S3 and table S1). This frequency calculation allowed us to identify candidate causal mutations on gene loci encoding DHSs (table S1). Subsequently, five additional metabolic *sota* lines, namely, *sotaB3*, *sotaF1*, *sotaG1*, *sotaH1*, and *sotaH9*, were also backcrossed to the *tyra2* mutant and subjected to the genetic mapping using the next-generation sequencing as described above, but this time only sequencing the *sota*-like population of their F₂ population. Some of the identified mutations were further confirmed by the dCAPS analysis (fig. S4) and the complementation experiments (Fig. 1E and fig. S6) as described below.

dCAPS-based genotyping of the *sota* mutants

To test whether the DHS *sota* mutations, identified by the whole-genome sequencing, indeed segregated with the *sota*-like phenotype (i.e., suppression of *tyra2* phenotypes), the presence and absence of

each DHS *sota* mutation were examined in their F₂ populations by the dCAPS analysis. Primers for each *sota* SNV were designed using the bioinformatic tool dCAPS finder 2.0 (39), while complementary primers for each dCAPS primer were designed using primer3 v.0.4.0 (40), which are listed in tables S8 and S9. Polymerase chain reaction (PCR) was performed using EconoTaq PLUS green 2x master mix (Lucigen) in a 20- μ l reaction containing ~10 ng of genomic DNA and 0.5 μ M of each primer. After amplification, the PCR product was verified on 4% 1 \times tris-borate EDTA (TBE)-agarose gel electrophoresis, and 5 μ l of PCR product was digested by a respective restriction enzyme (Thermo Fisher Scientific; table S8) in a 20- μ l reaction. Digested fragments were separated by electrophoresis in 4 to 5% 1 \times TBE-agarose gel containing ethidium bromide. The GeneRuler Ultra Low Range DNA Ladder (Thermo Fisher Scientific) was used to verify the sizes of the digested fragments. In all eight *sota* lines, the corresponding DHS *sota* mutations were found only in F₂ individuals exhibiting the *tyra2* suppression phenotypes (fig. S4).

Generation of transgenic plants

We further examined whether the identified *sota* mutations are indeed responsible for the observed *sota* mutant phenotypes, including the *tyra2* suppression phenotypes and the elevated Tyr and Phe levels (Fig. 1, B and C). Considering the dominant (or semidominant) nature of the *sota* mutations, we transformed the original *tyra2-1* mutant with an individual DHS gene with and without an identified *sota* mutation to see whether we can recapitulate the *sota*-like phenotypes. The binary vectors containing the WT version of *DHS1*, *DHS2*, and *DHS3* complementary DNA (cDNA) as well as their own native promoter regions, which we used previously to rescue the corresponding *dhs* knockout mutants (9), were also used here as templates to conduct site-directed mutagenesis and to introduce different *sota* mutations. These vectors also contain a hygromycin resistance gene and the pFAST-R construct, a C-terminal red fluorescence protein (RFP) fusion protein driven by a seed-specific Oleosin1 (At4g25140) native promoter (41, 42). The mutagenesis PCR was carried out by mixing 1 ng of ribonuclease-treated plasmids as a template, 2 \times PrimeSTAR MAX DNA polymerase mix (R045A, Takara Bio, USA), and 0.5 μ M oligonucleotide primers (table S9), which were designed using the Takara web tool for mutagenesis (www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools). After 20 cycles of the PCR (98°C for 15 s, 58°C for 10 s, 72°C for 2 min, and final extension at 72°C for 5 min), the product was treated with FastDigest DpnI (Thermo Fisher Scientific), purified using the QIAquick PCR Purification Kit (QIAGEN), and introduced into ultracompetent *Escherichia coli* MC1061 cells (Lucigen). The final binary vector sequence was confirmed by the whole-plasmid sequencing (MGH DNA Core). For generation of *Arabidopsis* transgenic plants in the *tyra2-1* mutant background, the *tyra2-1* seeds were germinated on the germination mix and grown until flowering before being transformed by each construct using the floral dipping method (43). The transformed T₀ plants were allowed to complete their life cycle at the growth chamber, and dried T₁ seeds were harvested. The positive T₁ transformants were then selected using the RFP fluorescent marker expression by observing the seeds under the AxioZoom V16 (Zeiss) stereo fluorescent microscope with RFP settings as EX 572/25, BA590, and EM 629/62. Their T₂ seeds were further used to select lines that contain a single insertion of the transgene. Overall, eight individual T₂ plants

from each single insertion line were allowed to complete their life cycle, and their seeds were observed under a stereo RFP fluorescent microscope to identify homozygous T₃ seeds that were used for further analyses. Some T₂ homozygous plants, depending on the positional effects, could not complete their life cycle because of high accumulation of AAA, similarly to the *sotaF1* homozygous line. For these specific lines, T₂ heterogeneous plant populations were used for further analysis. Although the hygromycin resistance gene was also present, the selection by RFP seeds was more time efficient and less aggressive, allowing the germination of positive transformants directly on soil.

To generate the transgenic lines expressing the WT or *sota*-mutated DHS genes in the Col-0 background, the same constructs as used for the complementation test were transformed into Col-0 plants. One leaf of each 5-week-old T₂ plant of each line was first analyzed for photosynthetic measurement, and then other leaves were harvested from the same plant for metabolite analysis.

Enzyme preparation and enzymatic assay

To generate recombinant DHS proteins, the pET28a vectors carrying the *A. thaliana* DHS1 (AtDHS1), AtDHS2, or AtDHS3 WT sequence without the transit peptide toward plastid (amino acid residues 49 to 525, 34 to 507, and 52 to 527, respectively) were expressed in *E. coli* Rosetta-2 cells and purified using Ni-affinity chromatography, exactly as we conducted previously (9). For site-directed mutagenesis to generate DHS proteins with individual *sota* mutations, these pET28a plasmid templates were diluted by 500-fold, mixed with Phusion DNA polymerase (0.04 U/ μ l) (Thermo Fisher Scientific), 0.2 mM deoxynucleoside triphosphates (dNTPs), 1 \times Phusion reaction buffer (Thermo Fisher Scientific), and 0.5 μ M forward and reverse mutagenesis primers (table S9), followed by the PCR at 98°C for 30 s followed by 20 cycles of 10 s at 98°C, 20 s at 70°C, and 4.5 min at 72°C with a final extension at 72°C for 10 min. The PCR products were purified using the QIAquick Gel Extraction Kit (QIAGEN), treated with FastDigest DpnI (Thermo Fisher Scientific) to digest methylated plasmid template DNA for 20 min at 37°C and then transformed into *E. coli* cells. The mutagenized pET28a plasmids were sequenced to confirm that no errors were introduced during the mutagenesis process.

The DHS enzyme assays were conducted using the colorimetric method that we recently described (9). Briefly, the enzyme solution (7.7 μ l) containing 50 mM Hepes (pH 7.4) was preincubated with an effector molecule(s) at room temperature for 15 min. For assays using recombinant protein and enzyme fraction isolated from plant leaves, 0.01 to 0.1 μ g and approximately 50 μ g of proteins were used, respectively. After adding 0.5 μ l of 0.1 M dithiothreitol, the samples were further incubated at room temperature for 15 min. During these incubations, the substrate solution containing 50 mM Hepes (pH 7.4), 2 mM MnCl₂, 4 mM E4P, and 4 mM PEP at final concentration was preheated at 37°C. The enzyme reaction was started by adding 6.8 μ l of the substrate solution, then incubated at 37°C for 30 min, and terminated by adding 30 μ l of 0.6 M trichloroacetic acid. After a brief centrifugation, 5 μ l of 200 mM NaIO₄ (sodium meta-periodate) in 9 N H₃PO₄ was added to oxidize the enzymatic product and to incubate at 25°C for 20 min. To stop the oxidation reaction, 20 μ l of 0.75 M NaAsO₂ (sodium arsenite), which was dissolved in 0.5 M Na₂SO₄ and 0.05 M H₂SO₄, was added and immediately mixed. After 5 min of incubation at room temperature, one-third of the sample solution was transferred to a new tube to be mixed with 50 μ l of 40 mM thiobarbituric acid and incubated at 99°C for 15 min

in a thermal cycler. The mixture was added to 600 μ l of cyclohexanone in eight-strip solvent-resistant plastic tubes, mixed vigorously, and centrifuged at 4500g for 3 min to separate water- and cyclohexanone-based layers for the extraction off the developed pink chromophore. The absorbance of the pink supernatant was read at 549 nm with the microplate reader (Infinite 200 PRO, TECAN) to calculate DAHP production with the molar extinction coefficient at 549 nm ($\epsilon = 549$ nm) of 4.5×10^4 M $^{-1}$ cm $^{-1}$. Reaction mixtures with boiled enzymes were run in parallel and used as negative controls to estimate the background signal.

Structural modeling and differential scanning fluorimetry analysis

The three-dimensional structure of DHS2 WT was generated by homology modeling using the high-resolution structure 5uxm.pdb of type II DHS from *P. aeruginosa* as a template structure (16). DHS2 WT has more than 60% sequence identity with the template. Homology modeling was performed using Modeller 9.24 (44). The model with the lowest discrete optimized protein energy value was chosen for further validation. Modeled structure was validated by the inspection of phi/psi distributions of the Ramachandran plot obtained through PROCHECK (45), and the significance of consistency between template and models was evaluated using the ProSA server (46). In addition, the root mean square deviation (RMSD) was analyzed by Chimera (match-maker) (47) on superimposition of template (5uxm.pdb) with predicted structures to check the reliability of models. The model shows an RMSD of 0.207 Å to 5uxm.pdb Trp for 441 atom pairs. Trp binding site of the model was mapped on Chimera by superposition to Trp-bound 5uxm.pdb.

To examine the impacts of the *sota* mutation on the interaction between DHS and AAA effectors (fig. S12), the differential scanning fluorimetry analysis (48) was conducted using the recombinant DHS2^{WT} and DHS2^{A4} proteins. After diluting each recombinant protein solution to 0.1 μ g/ μ l, 15 μ l of the protein solution was mixed with 4 μ l of 25 times SYPRO orange fluorescence dye (Sigma-Aldrich) and 1 μ l of 20 μ M AAA ligand solution dissolved in 40% ethanol. The fluorescence signal was monitored during the stepwise increase in temperature (1°C/min from 25° to 95°C). The T_m was calculated by nonlinear regression analysis using the Boltzmann sigmoidal equation (48).

Soluble metabolite analyses

Approximately 50 to 80 mg of fully expanded mature leaves were pooled from multiple plants at the same developmental stages. For seedling analyses, approximately 50 mg of shoots and 10 to 20 mg of roots were pooled from more than five 10-day-old seedlings. After quickly measuring their fresh weight, obtained tissues were immediately frozen in liquid nitrogen and kept in -80°C until use. The frozen tissues were mixed in 800 μ l of extraction buffer containing (v/v) 2:1 of methanol and chloroform with isovitexin (0.5 μ g/ml) (MilliporeSigma), 100 μ M norvaline (Thermo Fisher Scientific), and Tocol (1.25 μ g/ml) (Matreya LLC), as internal standards for soluble metabolite analysis by LC-MS and GC-MS and tocopherol analysis by GC-MS, respectively. The mixtures were immediately homogenized for at least 3 min using the 1600 MiniG Tissue Homogenizer (SPEX SamplePrep) and 3-mm glass beads. After adding 600 μ l of H₂O and then 250 μ l of chloroform, polar phase containing amino acids and nonpolar phase containing tocopherols were separated by centrifugation and dried in new tubes for further analysis.

Metabolite analyses of amino acids and tocopherols using GC-MS were carried out after derivatization of the polar and nonpolar metabolites with *N*-*tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide with 1% *tert*-butyldimethylchlorosilane (Cerilliant) and *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (Restek), respectively, exactly as we previously described (15, 49). Metabolite identities are summarized in data S1.

For targeted metabolite analysis of Trp and AAA-derived compounds, reverse-phase LC-MS analysis with the Vanquish UHPLC system coupled with the Q Exactive Quadrupole-Orbitrap MS (Thermo Fisher Scientific) was conducted as previously described (9), with some modification. The metabolites were dissolved in 70 μ l of LC-MS-grade 80% methanol and separated using the mobile phases of 0.1% formic acid in LC-MS-grade water (solvent A) and 0.1% formic acid in LC-MS-grade acetonitrile (solvent B) at a flow rate of 0.4 ml/min and a column temperature of 40°C. The binary 25-min linear gradient with the following ratios of solvent B was used: 0 to 1 min, 1%; 1 to 10 min, 1 to 10%; 10 to 13 min, 10 to 30%; 13 to 14.5 min, 30 to 70%; 14.5 to 15.5 min, 70 to 99%; 15.5 to 21 min, 99%; 21 to 22.5 min, 99 to 10%; 22.5 to 23 min, 10 to 1%; and 23 to 25 min, 1%. The spectra were recorded using the full scan mode of negative ion detection, covering a mass range from mass/charge ratio (*m/z*) 100 to 1500. The resolution was set to 25,000, and the maximum scan time was set to 250 ms. The sheath gas was set to a value of 60, while the auxiliary gas was set to 35. The transfer capillary temperature was set to 150°C, while the heater temperature was adjusted to 300°C. The spray voltage was fixed at 3 kV, with a capillary voltage and a skimmer voltage of 25 and 15 V, respectively. The identity of amino acids and I3M peaks was confirmed by comparing their accurate masses and retention times with those of the corresponding authentic standards (data S1). The identity of the other compounds was confirmed by LC-tandem MS analysis as previously performed (data S1) (9). Quantification was based on the standard curves generated by injecting different concentrations of authentic chemical standards. The isovitexin peak of each sample was detected to normalize the sample-to-sample variation and to calculate the recovery rate by comparing with a blank sample corresponding to 800 μ l of the extraction buffer.

For quantification of some highly polar metabolites such as shikimate, we used hydrophilic interaction chromatography (HILIC) followed by compound detection with a Vanquish UHPLC (ultrahigh-performance LC) system coupled with the Q Exactive MS (Thermo Fisher Scientific). The same samples used for reverse-phase LC-MS analysis was injected onto a HPLC Poroshell 120 HILIC-Z column (150-mm by 2.1-mm inner diameter, 2.7- μ m particle size; Agilent) and eluted using mobile phases of 0.2% acetic acid in LC-MS-grade water containing 5 mM ammonium acetate (solvent A) and 0.2% acetic acid in LC-MS-grade acetonitrile containing 5 mM ammonium acetate (solvent B) with the following 22.5-min gradient at a flow rate of 0.45 ml/min and column temperature of 40°C. The binary linear gradient with the following ratios of solvent B was used: 0 to 1 min, 100%; 1 to 11 min, 100 to 89%; 11 to 15.75 min, 89 to 70%; 15.75 to 16.25 min, 70 to 20%; 16.25 to 18.5 min, 20%; 18.5 to 18.6 min, 20 to 100%; and 18.6 to 22.5 min, 100%. The spectra were recorded using the full-scan negative-ion mode, covering a mass range from *m/z* 70 to 1050. The resolution was set to 70,000, and the maximum scan time was set to 100 ms. The sheath gas was set to a value of 60, while the auxiliary gas was set to 35. The transfer capillary temperature was set to 150°C, while the heater temperature

was adjusted to 300°C. The spray voltage was fixed at 3 kV, with a capillary voltage and a skimmer voltage of 25 and 15 V, respectively. Retention times, MS spectra, and associated peak intensities were extracted from the raw files using the Xcalibur software (Thermo Fisher Scientific). The identities of metabolite peaks were confirmed by comparing their accurate masses and retention times with those of the corresponding authentic standards (data S1). Quantification was based on the standard curves generated by injecting different concentrations of authentic chemical standards. The isovitexin peak was also detected as an internal standard for the normalization and the recovery rate calculation as used in the reverse-phase LC-MS analysis above.

The IAA level was quantified as previously reported (50), with some modifications. Approximately 150 mg of 10-day-old *Arabidopsis* WT and the *sota* mutant seedlings grown on the agar plates were pooled and quickly frozen in a tube with three 3-mm glass beads. After grounding frozen tissues with the 1600 MiniG Tissue Homogenizer (SPEX SamplePrep), the sample was dissolved in 1 ml of ice-cold sodium phosphate buffer (100 mM; pH 7.0) containing 1% (w/v) diethyldithiocarbamic acid and 1 μ M isovitexin and shaken on an orbital shaker for 20 min at 4°C. After the centrifugation at 23,000g, 4°C for 20 min, the pH of the supernatant was adjusted to below 3.0 with 1 N hydrochloric acid. The IAA metabolite was obtained by solid-phase extraction using Oasis HLB columns (1 ml/30 mg; Waters), which were conditioned with 1 ml of methanol and then 1 ml of water, and equilibrated with 0.5 ml of sodium phosphate buffer (acidified with 1 N hydrochloric acid below 3). After the sample application, the column was washed with 2 ml of 5% methanol and then eluted with 2 ml of 80% methanol. The eluate was evaporated and stored at -20°C until LC-MS analysis. IAA was detected by the same reverse-phase LC-MS method as described above, with the following modifications. The metabolites were separated using the mobile phases of 0.1% formic acid in LC-MS-grade water (solvent A) and 0.1% formic acid in LC-MS-grade acetonitrile (solvent B) at a flow rate of 0.2 ml/min. The binary 25-min linear gradient with the following ratios of solvent B was used: 0 to 0.5 min, 10%; 0.5 to 10 min, 10 to 50%; 10 to 12.5 min, 50 to 60%; 12.5 to 14.5 min, 60 to 70%; 14.5 to 16 min, 70 to 99%; 16 to 21 min, 99%; 21 to 22.5 min, 99 to 10%; and 22.5 to 25 min, 10%. The separated metabolites were detected as described above in the reverse-phase LC-MS analysis, with a selective ion monitoring (SIM) mode. The identity of the IAA peak was confirmed by comparing its accurate mass and retention times with those of the corresponding authentic standard (data S1). Quantification was based on the standard curves generated by injecting different concentrations of authentic chemical standards. The isovitexin peak was also detected as an internal standard for the normalization and the recovery rate calculation.

For anthocyanin quantification, the polar phase isolated for amino acid analysis was diluted 10 times with water in a new tube. After adding 5 μ l of 5 N HCl for acidification, the absorption was measured at 530 and 657 nm with a microplate reader (Infinite 200 PRO, TECAN) to calculate anthocyanin contents with the formula $A_{530} - 0.25 \times A_{657}$ (51). For chlorophyll quantification, the nonpolar phase was dried down and then resuspended in 1 ml of 90% methanol. Several serial dilutions were prepared, and absorbance at 652 and 665 nm was measured using a microplate reader (Infinite 200 PRO, TECAN). The quantities of chlorophylls in each dilution were estimated by the following equations: Chl a = $16.72 \times A_{665} - 9.16 \times A_{652}$ and Chl b = $34.09 \times A_{652} - 15.28 \times A_{665}$ (52).

¹³CO₂ labeling experiments

The ¹³CO₂ labeling experiments were conducted following the previously published protocol (53, 54). Briefly, for the time course labeling experiment, Col-0 WT and the *sotaB4* and *sotaA4* mutants (in the *tyra2* background) were grown for 3 weeks under 12 hours of 150- μ E light and 12 hours of darkness. These plants were transferred to a 60-liter labeling chamber (75 cm in width, 40 cm in depth, and 20 cm in height; fig. S15A), 1 hour before the beginning of the light period (~7 a.m.), to which the air containing 450 to 460 parts per million of ¹³CO₂ was provided at approximately 5 liter/min. After the light was turned on at 8 a.m., the 1-, 3-, and 6-hour samples were harvested at 9 a.m., 11 a.m., and 2 p.m. At each time point, the entire shoots (above ground tissues) were harvested and immediately frozen with liquid nitrogen, in three biological replicates per genotype, where three or four individual plants were pooled together to make one replication. As a nonlabeled control, the samples for the 0-hour time point were harvested in duplicate right before the light period without any ¹³CO₂ labeling. Separately from the 6-hour time course experiment at the beginning of the day, Col-0 WT and the *sotaA4* mutant were also labeled with ¹³CO₂ for 3 hours toward the end of the day. Again, 3-week-old plants were placed in the labeling chamber at 4:45 p.m. After 3 hours of ¹³CO₂ labeling, plants were harvested as above at 7:45 p.m., just before the light was turned off.

The harvested shoot samples were ground-frozen to fine powders using the Retsch Ball Mill MM400, and soluble metabolites were extracted as described above, except ribitol, in addition to isovitexin, which was added as an internal standard for GC-MS analysis. Soluble metabolites were dried and derivatized by MSTFA and analyzed by GC-time-of-flight-MS as described previously (55). For quantification of shikimate and Trp, the dried samples were dissolved in 100 μ l of 80% MeOH and analyzed by the HILIC LC-MS and the reverse-phase LC-MS methods, respectively, as described above, with the following modified HILIC mobile phase gradient: 0 to 1 min, 100%; 1 to 1.5 min, 100 to 89%; 1.5 to 15.75 min, 89 to 70%; 15.75 to 16.25 min, 70 to 20%; 16.25 to 18.5 min, 20%; 18.5 to 18.6 min, 20 to 100%; and 18.6 to 22.5 min, 100%. To increase the sensitivity of peak detections, especially for ¹³C-labeled fragments, the MS compound detection was performed by a SIM mode. Metabolite identity was summarized in data S1.

The peak integration and labeling calculation were carried out as described previously (54). Briefly, the peak areas of nonlabeled and labeled ions (isotopomers) in different samples were integrated using the Xcalibur software (Thermo Fisher Scientific). The obtained data were corrected for natural abundance by comparing to unlabeled control samples using the CORRECTOR software as described previously (54). The amounts of ¹³C-labeled metabolites (nmol/mg of fresh weight) were calculated by multiplying the total metabolite pool sizes (nmol/mg of fresh weight) with the percent of ¹³C-labeled over total metabolite (the sum of both ¹²C- and ¹³C-labeled metabolites).

Quantification of starch, sugar, protein, and lignin contents

Quantification of starch and sugar contents was conducted as previously described (56), with some modifications. Thirty to 50 mg of 4-week-old fully mature leaves were harvested for each biological sample at indicated time points and frozen in a tube with three 3-mm glass beads. Soluble sugars were extracted twice by boiling the sample in 700 μ l of 80% ethanol at 80°C for 45 min until the leaves became bleached. The ethanol extract was evaporated and

dissolved in 200 μ l of distilled water. The sucrose and glucose levels were determined using the Total Sugar Assay Kit (Megazyme) according to the manufacturer's instruction. For starch analysis, the bleached leaves' tissues were air-dried and then ground in 1 ml of 100 mM sodium acetate buffer (pH 5.0) containing 5 mM CaCl_2 . The solubilized starch was enzymatically hydrolyzed into glucose by incubating with 10 μ l of α -amylase (3 U/ μ l; Megazyme) at 100°C for 15 min. After cooling to room temperature, the mixture was further incubated with 10 μ l of amyloglucosidase (3 U/ μ l; Megazyme) at 50°C for 50 min. The glucose concentration was determined using the Total Starch Assay Kit (Megazyme) according to the manufacturer's instruction and expressed as micromole glucose equivalent/g fresh weight (FW).

For determination of total protein content, frozen leaf tissues harvested from 4-week-old *Arabidopsis* plants were ground in liquid nitrogen and dissolved in 500 μ l of ice-cold isolation buffer containing 20 mM Hepes (pH 7.4) and 2.5 mM EDTA to determine the protein concentration by the Bradford assay (57). For analyzing the protein amount of Rubisco large subunit (RbcL), the same samples were applied to 4 to 20% Mini-PROTEAN TGX Stain-Free Protein Gels (Bio-Rad) to visualize and quantify the RbcL bands.

To determine the lignin deposition, 4-week-old leaves and roots were first fixed in formaldehyde/acetic acid/ethanol/water at a ratio of 5:5:45:45 (v/v) and decolorized with ethanol/acetic acid at a ratio of 6:1 (v/v). Phloroglucinol staining was conducted as previously described (58). Briefly, tissues were incubated in a mixture of one volume of 37% HCl (v/v) and two volumes of 3% phloroglucinol in ethanol (w/v) for 10 min and observed under bright-field lighting with an Olympus SZX12 stereoscope at the Newcomb Imaging Center at the Department of Botany, University of Wisconsin-Madison. For quantifying lignin content, 4-week-old leaves (whole aerial parts) and matured inflorescence stems were harvested and freeze-dried. Three individual plant samples were obtained for each genotype. The tissues were homogeneously pulverized with a tissue homogenizer (1600 MiniG, Spex SamplePrep). The homogenate was then extracted sequentially with distilled water, methanol, and hexane and then freeze-dried to give cell wall residues (CWRs). Thioglycolic acid lignin analysis was performed as described previously (59). The relative lignin content was expressed as absorbance of thioglycolic acid lignin at 280 nm (A_{280}) per weight of CWRs (mg).

Gas exchange measurement

The rate of net CO_2 assimilation was measured using an LI-6400XT photosynthesis system equipped with the 6400-40 leaf chamber (LI-COR). *Arabidopsis* plants were grown in the growth chamber under the condition of a 12-hour/12-hour 100- μE light/dark cycle with 85% air humidity for 4 weeks after germination, and fully expanded nonshaded leaves were used for the measurement. Because leaves did not fully fill the cuvette area, the leaf area inside the cuvette was photographed and quantified by ImageJ to normalize each assimilation rate. The temperature was kept at 25°C for all measurements. For analysis of the light response curve, the CO_2 concentration in the airstream was maintained at 400 $\mu\text{mol/mol}$. For analysis of the A - C_i curve, the light intensity was saturated at 1500 μE . After acclimating the leaves at the C_i level of 400 $\mu\text{mol/mol}$ to achieve a steady-state rate of assimilation, the C_i level of the response curve was set at 400, 185, 70, 35, 740, 1100, 1500, and 1900 $\mu\text{mol/mol}$, and measurements were taken when assimilation reached a steady-state rate. To determine the V_{cmax} , J_{max} , and R_d values, each A - C_i curve was fitted to the Farquhar-von Caemmerer-Berry model by the

"plantecophys" R package (60, 61). The initial slope and CO_2 compensation point of the light response curves and A - C_i curves were determined using the first three and five points at low light and low C_i points, respectively, as previously calculated (62).

Quantitative PCR expression analysis

To test the effects of the *sota* mutations on the *DHS* gene expression, the transcript levels of *DHS1*, *DHS2*, and *DHS3* were analyzed by reverse transcription quantitative PCR (RT-qPCR). Approximately 20 to 30 mg of fully expanded mature leaves were pooled from multiple 4-week-old plants grown on soils, immediately frozen in liquid nitrogen in a tube with three 3-mm glass beads, and ground using the 1600 MiniG Tissue Homogenizer (SPEX SamplePrep). Total RNA was isolated as previously described (63), treated with deoxyribonuclease I (Thermo Fisher Scientific), and reverse-transcribed to synthesize cDNA with M-MuLV reverse transcriptase and random hexamer primers (Promega) according to the manufacturer's protocol. RT-qPCR was conducted by the Stratagene Mx3000P (Agilent Technologies) using the GoTaq qPCR Master Mix (Promega) and target gene-specific primers listed in table S9. Four biological replicates with two technical RT-qPCR replicates were conducted. Expression of the *UBQ9* gene was used to normalize the sample-to-sample variations between different cDNA preparations. Relative expression levels among different genotypes were analyzed for each *DHS* gene using the $2^{-\Delta\Delta\text{Ct}}$ method.

Amino acid sequence alignment

DHS orthologs were first identified by BlastP searches using the amino acid sequence of AtDHS1 as a query against Phytozome 13 (64). *Nicotiana benthamiana* DHSs were searched from the *N. benthamiana* draft genome sequence v1.0.1 (65). The sequence sets used for figs. S8 and S24 as well as the calculation for the conservation of the residues among the plant kingdom (fig. S25) were available as FASTA files in data S2 to S4, respectively. Sequence alignment of figs. S8 and S24 was conducted with the MUSCLE algorithm and then visualized with Jalview (66), highlighting the residues by different depths of purple according to the percentage of the residues in each column that agree with the consensus sequence (>80, >60, >40, and <40%). Alignment of fig. S25 was generated with the MUSCLE algorithm and then visualized with Excel, coloring the residues that agree with the consensus amino acid at the *sota* mutation sites by yellow.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abo3416>

REFERENCES AND NOTES

- U.S. Department of Energy, Accelerating breakthrough innovation in carbon capture, utilization, and storage (2017); www.energy.gov/fe/downloads/accelerating-breakthrough-innovation-carbon-capture-utilization-and-storage.
- Global Aromatic Market: Information by type (benzene, toluene, O-xylene, P-xylene and others), by application (solvent, additive), by end-use industry (paint & coating, adhesive, pharmaceuticals, chemicals and others), region (North America, Europe, Asia Pacific, Latin America and Middle East & Africa)—Forecast till 2025 (Market Research Future, 2020); www.marketresearchfuture.com/reports/aromatics-market-930.
- T. Li, T. Shoinkhorova, J. Gascon, J. Ruiz-Martinez, Aromatics production via methanol-mediated transformation routes. *ACS Catal.* **11**, 7780–7819 (2021).
- W. Boerjan, J. Ralph, M. Baucher, Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546 (2003).
- A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan, C. E. Wyman, Lignin valorization: Improving lignin processing in the biorefinery. *Science* **344**, 1246843 (2014).

6. H. Maeda, N. Dudareva, The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.* **63**, 73–105 (2012).
7. C. S. Westfall, A. Xu, J. M. Jez, Structural evolution of differential amino acid effector regulation in plant chorismate mutases. *J. Biol. Chem.* **289**, 28619–28628 (2014).
8. C. A. Schenck, S. Chen, D. L. Siehl, H. A. Maeda, Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nat. Chem. Biol.* **11**, 52–57 (2015).
9. R. Yokoyama, M. V. V. de Oliveira, B. Kleven, H. A. Maeda, The entry reaction of the plant shikimate pathway is subjected to highly complex metabolite-mediated regulation. *Plant Cell* **33**, 671–696 (2021).
10. G. Jander, S. R. Baerson, J. A. Hudak, K. A. Gonzalez, K. J. Gruijts, R. L. Last, Ethylmethanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiol.* **131**, 139–146 (2003).
11. J. E. Brotherton, M. R. Jeschke, P. J. Tranell, J. M. Widholm, Identification of *Arabidopsis thaliana* variants with differential glyphosate responses. *J. Plant Physiol.* **164**, 1337–1345 (2007).
12. J. Li, R. L. Last, The *Arabidopsis thaliana* *trp5* mutant has a feedback-resistant anthranilate synthase and elevated soluble tryptophan. *Plant Physiol.* **110**, 51–59 (1996).
13. T. Huang, T. Tohge, A. Lytovchenko, A. R. Fernie, G. Jander, Pleiotropic physiological consequences of feedback-insensitive phenylalanine biosynthesis in *Arabidopsis thaliana*. *Plant J.* **63**, 823–835 (2010).
14. L. Pollegioni, E. Schonbrunn, D. Siehl, Molecular basis of glyphosate resistance: Different approaches through protein engineering. *FEBS J.* **278**, 2753–2766 (2011).
15. M. V. V. de Oliveira, X. Jin, X. Chen, D. Griffith, S. Batchu, H. A. Maeda, Imbalance of tyrosine by modulating TyrA arogenate dehydrogenases impacts growth and development of *Arabidopsis thaliana*. *Plant J.* **97**, 901–922 (2019).
16. O. W. Sterritt, S. A. Kessans, G. B. Jameson, E. J. Parker, A pseudoisostructural type II DAH7PS enzyme from *Pseudomonas aeruginosa*: Alternative evolutionary strategies to control shikimate pathway flux. *Biochemistry* **57**, 2667–2678 (2018).
17. T. Vogt, Phenylpropanoid biosynthesis. *Mol. Plant* **3**, 2–20 (2010).
18. X. Zhang, C.-J. Liu, Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol. Plant* **8**, 17–27 (2015).
19. L. J. Newman, D. E. Perazza, L. Juda, M. M. Campbell, Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the *det3* mutant phenotype. *Plant J.* **37**, 239–250 (2004).
20. C. Dubos, R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, L. Lepiniec, MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573–581 (2010).
21. H. Yoo, J. R. Widholm, Y. Qian, H. Maeda, B. R. Cooper, A. S. Jannasch, I. Gonda, E. Lewinsohn, D. Rhodes, N. Dudareva, An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nat. Commun.* **4**, 2833 (2013).
22. M. Wang, K. Toda, H. A. Maeda, Biochemical properties and subcellular localization of tyrosine aminotransferases in *Arabidopsis thaliana*. *Phytochemistry* **132**, 16–25 (2016).
23. M. Wang, K. Toda, A. Block, H. A. Maeda, TAT1 and TAT2 tyrosine aminotransferases have both distinct and shared functions in tyrosine metabolism and degradation in *Arabidopsis thaliana*. *J. Biol. Chem.* **294**, 3563–3576 (2019).
24. X. Wang, Y. Hou, L. Liu, J. Li, G. Du, J. Chen, M. Wang, A new approach for efficient synthesis of phenylactic acid from L-phenylalanine: Pathway design and cofactor engineering. *J. Food Biochem.* **42**, e12584 (2018).
25. M. J. Valera, E. Boido, J. C. Ramos, E. Manta, R. Radi, E. DellaCassa, F. Carrau, The mandelate pathway, an alternative to the phenylalanine ammonia lyase pathway for the synthesis of benzenoids in Ascomycete yeasts. *Appl. Environ. Microbiol.* **86**, e00701-20 (2020).
26. R. Bentley, The shikimate pathway—A metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* **25**, 307–384 (1990).
27. A. Arnold, Z. Nikoloski, Bottom-up metabolic reconstruction of *Arabidopsis* and its application to determining the metabolic costs of enzyme production. *Plant Physiol.* **165**, 1380–1391 (2014).
28. W. Jiao, E. J. Lang, Y. Bai, Y. Fan, E. J. Parker, Diverse allosteric componentry and mechanisms control entry into aromatic metabolite biosynthesis. *Curr. Opin. Struct. Biol.* **65**, 159–167 (2020).
29. V. Tzin, S. Malitsky, M. M. B. Zvi, M. Bedair, L. Sumner, A. Aharoni, G. Galili, Expression of a bacterial feedback-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of the shikimate pathway in *Arabidopsis* elucidates potential metabolic bottlenecks between primary and secondary metabolism. *New Phytol.* **194**, 430–439 (2012).
30. V. Tzin, I. Rogachev, S. Meir, M. Moyal Ben Zvi, T. Masci, A. Vainstein, A. Aharoni, G. Galili, Tomato fruits expressing a bacterial feedback-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of the shikimate pathway possess enhanced levels of multiple specialized metabolites and upgraded aroma. *J. Exp. Bot.* **64**, 4441–4452 (2013).
31. M. Oliva, A. Guy, G. Galili, E. Dor, R. Schweitzer, R. Amir, Y. Hacham, Enhanced production of aromatic amino acids in tobacco plants leads to increased phenylpropanoid metabolites and tolerance to stresses. *Front. Plant Sci.* **11**, 604349 (2020).
32. W. Jiao, Y. Fan, N. J. Blackmore, E. J. Parker, A single amino acid substitution uncouples catalysis and allostery in an essential biosynthetic enzyme in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **295**, 6252–6262 (2020).
33. S. Henkes, U. Sonnewald, R. Badur, R. Flachmann, M. Stitt, A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *Plant Cell* **13**, 535–551 (2001).
34. A. J. Simkin, P. E. Lopez-Calcagno, P. A. Davey, L. R. Headland, T. Lawson, S. Timm, H. Bauwe, C. A. Raines, Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphosphate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO₂ assimilation, vegetative biomass and seed yield in *Arabidopsis*. *Plant Biotechnol. J.* **15**, 805–816 (2017).
35. A. Gardemann, D. Schimkat, H. W. Heldt, Control of CO₂ fixation regulation of stromal fructose 1,6-bisphosphatase in spinach by pH and Mg²⁺ concentration. *Planta* **168**, 536–545 (1986).
36. M. A. J. Parry, A. J. Keys, P. J. Madgwick, A. E. Carmo-Silva, P. J. Andralojc, Rubisco regulation: A role for inhibitors. *J. Exp. Bot.* **59**, 1569–1580 (2008).
37. K. A. Molla, S. Sretenovic, K. C. Bansal, Y. Qi, Precise plant genome editing using base editors and prime editors. *Nat. Plants* **7**, 1166–1187 (2021).
38. D. Weigel, J. Glazebrook, EMS mutagenesis of *Arabidopsis* seed. *Cold Spring Harb. Protoc.* **2006**, pdb.prot4621 (2006).
39. M. M. Neff, E. Turk, M. Kalishman, Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* **18**, 613–615 (2002).
40. A. Untergasser, I. Cutucache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, S. G. Rozen, Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
41. T. L. Shimada, T. Shimada, I. Hara-Nishimura, A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. *Plant J.* **61**, 519–528 (2010).
42. C. Engler, M. Youles, R. Gruetzner, T.-M. Ehnert, S. Werner, J. D. G. Jones, N. J. Patron, S. Marillonnet, A golden gate modular cloning toolbox for plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
43. S. J. Clough, A. F. Bent, Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
44. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.37 (2016).
45. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
46. M. Wiederstein, M. J. Sippl, ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–W410 (2007).
47. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
48. F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).
49. M. Wang, S. Lopez-Nievez, I. L. Goldman, H. A. Maeda, Limited tyrosine utilization explains lower betalain contents in yellow than in red table beet genotypes. *J. Agric. Food Chem.* **65**, 4305–4313 (2017).
50. O. Novák, E. Héniková, I. Sairanen, M. Kowalczyk, T. Pospišil, K. Ljung, Tissue-specific profiling of the *Arabidopsis thaliana* auxin metabolome. *Plant J.* **72**, 523–536 (2012).
51. A. L. Mancinelli, Interaction between light quality and light quantity in the photoregulation of anthocyanin production. *Plant Physiol.* **92**, 1191–1195 (1990).
52. A. R. Wellburn, The spectral determination of chlorophylls *a* and *b*, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. *J. Plant Physiol.* **144**, 307–313 (1994).
53. M. Szecowka, R. Heise, T. Tohge, A. Nunes-Nesi, D. Vosloh, J. Huege, R. Feil, J. Lunn, Z. Nikoloski, M. Stitt, A. R. Fernie, S. Arrivault, Metabolic fluxes in an illuminated *Arabidopsis* rosette. *Plant Cell* **25**, 694–714 (2013).
54. R. Heise, S. Arrivault, M. Szecowka, T. Tohge, A. Nunes-Nesi, M. Stitt, Z. Nikoloski, A. R. Fernie, Flux profiling of photosynthetic carbon metabolism in intact plants. *Nat. Protoc.* **9**, 1803–1824 (2014).
55. J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, A. R. Fernie, Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387–396 (2006).
56. H. Maeda, W. Song, T. L. Sage, D. DellaPenna, Tocopherols play a crucial role in low-temperature adaptation and phloem loading in *Arabidopsis*. *Plant Cell* **18**, 2710–2732 (2006).

57. M. M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

58. P. Pradhan Mitra, D. Loqué, Histochemical staining of *Arabidopsis thaliana* secondary cell wall elements. *J. Vis. Exp.*, (2014).

59. S. Suzuki, Y. Suzuki, N. Yamamoto, T. Hattori, M. Sakamoto, T. Umezawa, High-throughput determination of thioglycolic acid lignin from rice. *Plant Biotechnol.* **26**, 337–340 (2009).

60. G. D. Farquhar, S. von Caemmerer, J. A. Berry, A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta* **149**, 78–90 (1980).

61. R. A. Duursma, Plantcophys—An R package for analysing and modelling leaf gas exchange data. *PLOS ONE* **10**, e0143346 (2015).

62. J. Kromdijk, K. Glowacka, S. P. Long, Photosynthetic efficiency and mesophyll conductance are unaffected in *Arabidopsis thaliana* aquaporin knock-out lines. *J. Exp. Bot.* **71**, 318–329 (2020).

63. L. Oñate-Sánchez, J. Vicente-Carbajosa, DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC. Res. Notes* **1**, 93 (2008).

64. D. M. Goodstein, S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D. S. Rokhsar, Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).

65. A. Bombléry, H. G. Rosli, J. Vrebalov, P. Moffett, L. A. Mueller, G. B. Martin, A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol. Plant Microbe Interact.* **25**, 1523–1530 (2012).

66. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

67. J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmüller, P. Dörmann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie, D. Steinhäuser, GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics* **21**, 1635–1638 (2005).

68. N. Schauer, D. Steinhäuser, S. Strelkov, D. Schomburg, G. Allison, T. Moritz, K. Lundgren, U. Roessner-Tunali, M. G. Forbes, L. Willmitzer, A. R. Fernie, J. Kopka, GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* **579**, 1332–1337 (2005).

Acknowledgments

Funding: This work was supported by U.S. National Science Foundation grant MCB-1818040 (to H.A.M. and J.M.J.), the Alexander von Humboldt Research Fellowship (to H.A.M.), the postdoctoral fellowship of the Japan Society for the Promotion of Science (to R.Y.), and the postdoctoral fellowship of the Uehara Memorial Foundation (to R.Y.). **Author contributions:** Conceptualization: R.Y., M.V.V.d.O., and H.A.M. Investigation: R.Y., M.V.V.d.O., Y.T.-K., H.I., S.A.I., S.Ar., V.K., and H.A.M. Analysis: R.Y., M.V.V.d.O., Y.T.-K., H.I., S.A.I., S.Ar., V.K., J.M.J., M.S., A.R.F., and H.A.M. Funding acquisition: J.M.J. and H.A.M. Project administration: H.A.M. Supervision: J.M.J., M.S., A.R.F., and H.A.M. Writing—Original draft: R.Y. and H.A.M. Writing—Review and editing: R.Y., M.V.V.d.O., Y.T.-K., H.I., S.A.I., S.Ar., V.K., J.M.J., M.S., A.R.F., and H.A.M. **Competing interests:** H.A.M., M.V.V.d.O., and R.Y. have a pending patent application related to the *sota* mutations identified in this study. The authors declare that they have no other competing interests. **Data and materials availability:** The *sota* mutants identified and characterized in this study are a part of the pending patent application (Wisconsin Alumni Research Foundation, filed on 7 December 2021, serial number: 63/286811) and will be made available upon request with a simple material transfer agreement under the terms of the Uniform Biological Material Transfer Agreement (UBMTA). All SNVs found from the *sota* genome sequencing are listed in data S5. The coding sequences of *sota*-mutated DHS genes have been deposited to NCBI under the consecutive accession numbers of ON314238 to ON314244. The original metabolite profiling data of the main results presented in this study have been deposited in the MetaboLights database under the identifier of MTBLS4694.

Submitted 27 January 2022

Accepted 19 April 2022

Published 8 June 2022

10.1126/sciadv.abo3416