

SYNTHETIC BIOLOGY

Discovery of a previously unknown biosynthetic capacity of naringenin chalcone synthase by heterologous expression of a tomato gene cluster in yeast

Deze Kong^{1*}, Sijin Li^{2*}, Christina D. Smolke^{1,3†}

Chalcone synthase (CHS) canonically catalyzes carbon-carbon bond formation through iterative decarboxylative Claisen condensation. Here, we characterize a previously unidentified biosynthetic capability of SlCHS to catalyze nitrogen-carbon bond formation, leading to the production of a hydroxycinnamic acid amide (HCAA) compound. By expressing a putative tomato (*Solanum lycopersicum*) gene cluster in yeast (*Saccharomyces cerevisiae*), we elucidate the activity of a pathway consisting of a carboxyl methyltransferase (SIMT2), which methylates the yeast primary metabolite 3-hydroxyanthranilic acid (3-HAA) to form a methyl ester, and a SlCHS, which catalyzes the condensation of 3-HAA methyl ester and *p*-coumaroyl-coenzyme A (CoA) through formation of an amide bond. We demonstrate that this aminoacylation activity could be a common secondary activity in plant CHSs by validating the activity in vitro with variants from *S. lycopersicum* and *Arabidopsis thaliana*. Our work demonstrates yeast as a platform for characterizing putative plant gene clusters with the potential for compound structure and enzymatic activity discovery.

INTRODUCTION

Plant specialized metabolism is a rich source of structurally and functionally diverse small molecules, also known as plant natural products. These specialized metabolites play important roles in plant communication and defense and have been widely applied as phytochemicals, antibiotics, antivirals, nutraceuticals, and cosmetics (1, 2). Recent developments in synthetic biology and metabolic engineering have enabled the assembly and expression of plant genes in heterologous hosts as a sustainable and efficient alternative for production of complex chemicals, including plant natural products and their synthetic derivatives (3, 4). However, the broader potential of these engineering efforts is challenged partially due to our limited knowledge of plant biosynthetic pathways and associated enzyme activities.

The elucidation of plant specialized metabolic pathways has been challenging, particularly in comparison to the elucidation of natural product pathways in microbes. In part, this has been due to the differences in the genomic organization of these pathways, where the genes encoding the biosynthetic pathway in plants are generally dispersed across the plant genome, whereas, in contrast, those in microbes tend to be tightly clustered in operons. However, recent work has revealed that certain genes constituting a number of plant natural product pathways are colocalized in the genome in operon-like structures. These plant biosynthetic gene clusters range from ~35 to several hundred kilobases (i.e., 3 to more than 10 genes) in size (5) and comprise genes that are physically colocalized and potentially coregulated. These gene clusters encode species-specific and/or specialized biochemical pathways modifying metabolites from primary metabolism, contributing to the vast chemical space present in the plant kingdom (6). Characterization of putative gene cluster activities and their resulting products assisted by genome mining and analytical chemistry may thus provide an abundant

source for the discovery of enzyme activities and compound structures (7, 8).

Gene cluster prediction in plants has been challenging because plant genomes are larger than those of bacteria and fungi, and plant genes are sparsely distributed along the genome, separated by a substantial amount of intergenic, noncoding sequences (7). A general approach for identifying plant gene clusters involves defining a “cluster core” by searching for backbone-generating enzymes—e.g., non-ribosomal peptide synthetase (NRPS), polyketide synthase (PKS), hybrid NRPS-PKS, and terpene synthase—from genome sequences and then expanding the cluster components based on catalytic domain analysis, physical colocalization, gene coexpression, and/or shared regulatory patterns (7, 8). Recently developed cluster-mining algorithms such as PhytoClust (9), PlantSMASH (10), and PlantClusterFinder (11) have demonstrated automated detection of hundreds to thousands of putative gene clusters from various plant genomes.

Despite the increasing number of putative plant biosynthetic gene clusters arising from computational prediction tools, characterizing the potential functionality of these clusters and associated enzymes in their host organisms has been limiting. In particular, in planta pathway characterization can be hindered by cryptic pathway gene expression, low concentrations of targeted compounds embedded in complex mixtures, and difficulties in genetically manipulating the native host for cluster activation (7). Facilitated by well-developed tools for genetic manipulation and pathway expression, baker's yeast (*Saccharomyces cerevisiae*) has proven to be a powerful platform for expression of heterologous gene clusters. Previous research has used yeast to characterize the biosynthetic activities of several gene clusters from various plant species, including triterpene biosynthetic clusters from *Arabidopsis thaliana* (12), a 10-gene noscapine-producing cluster from poppy (*Papaver somniferum*) (13), partial pathway genes for vinblastine and vincristine biosynthesis from Madagascan periwinkle (*Catharanthus roseus*) (14), cucurbitacin from cucurbit (*Cucurbitaceae*) (15), and a cyanogenic glycoside biosynthetic cluster from sorghum (*Sorghum bicolor*) (16). In these earlier studies, the previously identified plant gene clusters were heterologously expressed in yeast to validate the production of the compounds as expected from their plant hosts.

Copyright © 2020
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).

¹Department of Bioengineering, 443 Via Ortega, MC 4245, Stanford University, Stanford, CA 94305, USA. ²Robert Frederick Smith School of Chemical and Biomolecular Engineering, 230A Olin Hall, Cornell University, Ithaca, NY 14853, USA. ³Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: csmolke@stanford.edu

In this work, we use yeast as a plant natural product discovery platform to characterize the biosynthetic potential of a putative tomato gene cluster predicted from PlantClusterFinder (11), the activity of which has not been reported previously. By coexpressing the cluster genes with an early-step flavonoid pathway gene in yeast, we identified two previously unknown compounds in the yeast culture when fed *p*-coumaric acid, specifically 3-hydroxyanthranilic acid (3-HAA) methyl ester (1) and a hydroxycinnamic acid amide (HCAA) compound, dihydro-coumaroyl anthranilate amide (2) (Fig. 1A). Further analysis confirmed that a methyltransferase (SMT2) catalyzes the conversion of 3-HAA—a native yeast metabolite involved in tryptophan metabolism—to (1), and a naringenin chalcone synthase (SICH5) catalyzes the condensation of (1) and *p*-dihydro-coumaroyl-coenzyme A (CoA), reduced from *p*-coumaroyl-CoA by a yeast endogenous enoyl-CoA reductase (ECR), leading to production of (2). Knocking out the native ECR in yeast restored the production of an oxidized form of (2), coumaroyl anthranilate amide (3). Our characterization results reveal a previously uncharacterized amide synthesis activity for SICH5. In vivo site-directed mutagenesis results suggest that SICH5 uses the same active site for synthesis of (3) and for canonical synthesis of naringenin chalcone. Our work demonstrates the potential of yeast as a characterization tool for computationally aided discovery of compound structures and enzymatic activities from plant genomes.

RESULTS

Reconstitution of a putative five-gene tomato cluster in yeast leads to synthesis of an HCAA compound

Our study investigated the biosynthetic potential of a tomato-derived putative gene cluster that was predicted to produce hydroxylated naringenin chalcone and/or methyl esters of hydroxylated naringenin chalcone, natural compounds that are found in tomato but without an elucidated pathway for biosynthesis (11). The putative tomato gene cluster predicted from PlantClusterFinder [referred to as C584_4 (11)] consists of a CHS (*SICH5*, SOLYC09G091510), a putative cytochrome P450 (*SICYP*, SOLYC09G091570), and three methyltransferases (*SMT1/2/3*; SOLYC09G091530, SOLYC09G091540, and SOLYC09G091550). *SICH5* is a well-studied type III PKS, which is known to sequentially condense one *p*-coumaroyl-CoA and three malonyl-CoA molecules to make naringenin chalcone, the first committed intermediate in the biosynthesis of flavonoids and anthocyanins (17). Among the three methyltransferases, *SMT3* was previously characterized as a putative salicylic acid methyltransferase potentially regulating tomato hormone emission (18). To our knowledge, no studies have been reported characterizing *SMT1*, *SMT2*, and *SICYP* from the cluster.

We examined the biosynthetic capacity of the predicted tomato gene cluster in yeast. Yeast expression cassettes for complementary DNAs encoding the five genes identified in the cluster (*SICH5*, *SICYP*, and *SMT1/2/3*) were designed and assembled into a yeast artificial chromosome and transformed into a wild-type yeast strain (CEN.PK2), resulting in yeast strain CSY1210. Two additional enzymes supporting the putative pathway enzymes were expressed in CSY1210 from low-copy plasmids: (i) a yeast codon-optimized 4-coumarate-CoA ligase from tomato (*Sl4CL*), a precursor-producing gene from the flavonoid pathway, and (ii) an *Arabidopsis* NADPH-cytochrome P450 reductase (*AtATR1*), a reductase partner to support the activity of the putative cytochrome P450 (*SICYP*). We cultured CSY1210 trans-

formed with the additional plasmids and a control strain (transformed with the plasmids but not harboring the reconstructed tomato cluster) in synthetic dropout media supplemented with 100 μ M *p*-coumaric acid (the substrate for *Sl4CL*) for 72 hours at 25°C and analyzed the yeast media. The metabolites produced by the strain harboring the reconstructed tomato cluster were identified using an untargeted metabolomics analysis by qToF-MS (quadrupole time-of-flight hybrid mass spectrometry) (with a mass accuracy at 50 parts per million).

We observed two differential peaks representing compounds only produced in the strain harboring the reconstructed tomato cluster, one at mass/charge ratio (*m/z*) 168.0655 ($[M + H]^+$) (1) and the other at 316.1179 ($[M + H]^+$) (2) (fig. S1, A and B). To validate production of the two compounds in yeast, we analyzed the yeast culture media for production of (1) and (2) on liquid chromatography-tandem MS (LC-MS/MS). A product ion scan with a precursor ion set at 168.0 *m/z* showed two peaks at retention times of 4.291 and 5.872 min, respectively, and a product ion scan with a precursor ion set at 316.1 *m/z* showed a single peak at 5.872 min (Fig. 1B). On the basis of retention times and fragmentation patterns of (1) and (2) from qToF-MS analysis (fig. S1, A and B), we hypothesized that the peak at 4.291 min corresponds to (1) and that the peak at 5.872 min (for both precursor ion settings) corresponds to (2).

Identification of the minimal essential tomato gene set, leading to synthesis of a previously unknown HCAA compound in yeast

We next identified the genes from the predicted tomato cluster and supporting flavonoid pathway (i.e., *Sl4CL* and *AtATR1*) that participated in the production of (1) and (2) in yeast. We first examined whether the methyltransferases individually participated in the biosynthesis of (1) and (2). To enable stable expression of the gene cassettes, *Sl4CL*, *SICH5*, and *SMT1/2/3* were chromosomally integrated into the wild-type yeast strain (CEN.PK2) such that each engineered strain harbors *Sl4CL*, *SICH5*, and one of the methyltransferases—leading to construction of CSY1301 (*SMT1*), CSY1302 (*SMT2*), and CSY1303 (*SMT3*). As a control, we eliminated *SICYP* (and *AtATR1*) from the integration to isolate their functions in compound synthesis. We cultured the strains in synthetic complete media supplemented with 100 μ M *p*-coumaric acid for 72 hours at 30°C and analyzed the yeast culture media for production of (1) and (2). A product ion scan on LC-MS/MS with precursor ion set at 168.0 showed two peaks for *SMT1* and *SMT2* transformants at 4.324 and 5.864 min, respectively (Fig. 1C). A product ion scan with a precursor ion set at 316.1 showed a single peak at 5.864 min for *SMT1* and *SMT2* transformants (Fig. 1C). As previously hypothesized, the peak at 5.864 min detected at 168 *m/z* may be a molecular fragment of (2). Production of (1) and (2) in the absence of *SICYP* (and *AtATR1*) indicates that *SICYP* and *AtATR1* are not involved in the production of the compounds. From the data, we observed production of (1) and (2) in both CSY1301 and CSY1302, and the product ion detected in CSY1302 was 14-fold greater than that in CSY1301 (Fig. 1D). The results indicate that *SMT1* and *SMT2* participate individually in the production of (1) and (2) and that *SMT2* leads to ~21-fold higher level of (1) and ~14-fold higher level of (2) than *SMT1*. Since the activities of *SMT1* and *SMT2* appear to be redundant in the context of characterizing the production of (1) and (2), we focused on the activity of *SMT2* for subsequent characterizations. Together, the results of methyltransferase characterizations revealed that (1) and

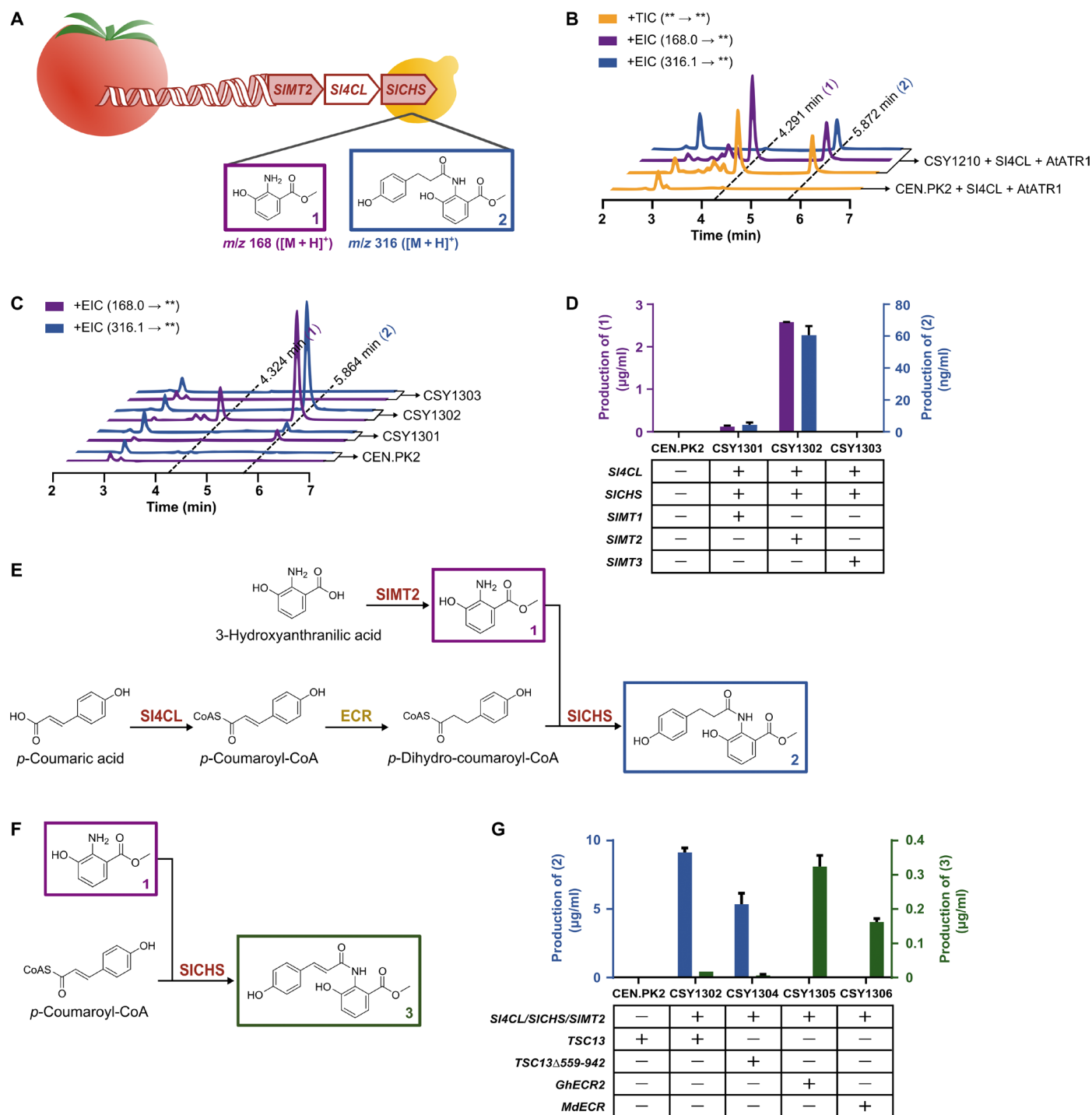


Fig. 1. Discovery of compounds and enzymatic activities via heterologous expression of a putative plant gene cluster in yeast. (A) Discovery of two previously unidentified compound structures by heterologous expression of genes from tomato cluster in yeast. Gene color: red, putative gene cluster; white, plant flavonoid pathway. (B) Validation of (1) and (2) production in yeast. CEN.PK2, wild-type yeast strain; CSY1210, strain expressing *SICH5*, *SICYP*, and *SIMT1/2/3*. (C) Characterization of (1) and (2) production with individual tomato methyltransferases in yeast. *SICH5* and *SICYP* are coexpressed with *SIMT1* (CSY1301), *SIMT2* (CSY1302), or *SIMT3* (CSY1303). (D) Summary of compound production with *SIMT1/2/3*. (E) Proposed pathway for biosynthesis of (1) and (2) in yeast. Enzyme color: red, tomato; yellow, yeast. (F) Proposed activity of *SICH5* in *TSC13* knockout strains. (G) Summary of compound production by *TSC13* knockout strains. TIC, total ion chromatogram; EIC, extracted ion chromatograms; ** indicates a thorough MS scan from *m/z* 10 to 168.0 or 316.1. "+"/"—" indicates the presence/absence of a gene or a gene fragment. Data show the mean of two biologically independent replicates, with error bar indicating SD. Compound color: purple, (1) methyl 3-hydroxyanthranilic acid; blue, (2) dihydro-coumaroyl anthranilate amide; green, (3) coumaroyl anthranilate amide. Enzyme abbreviations: *SIMT2*, methyltransferase 2; *SI4CL*, 4-coumarate-CoA ligase; *SICH5*, naringenin chalcone synthase; *ATR1*, NADPH-cytochrome P450 reductase 1; *ECR*, enoyl-CoA reductase.

(2) can be produced from a minimal set of genes consisting of *Sl4CL*, *SlCHS*, and *SlMT2*.

Validation of the proposed biosynthetic scheme of the HCAA compound in yeast

We next elucidated a biosynthetic scheme for the synthesis of (1) and (2) in yeast. Low-copy plasmids encoding the expression of *Sl4CL*, *SlCHS*, and *SlMT2* were cotransformed in different combinations into yeast, and the production of (1) and (2) were monitored in the presence and absence of fed *p*-coumaric acid after 72 hours of growth at 30°C (table S1). We first coexpressed the three genes with or without fed *p*-coumaric acid (groups 1 and 2). We then coexpressed all pairs of genes, e.g., *SlCHS* and *SlMT2*, *SlMT2* and *Sl4CL*, and *SlCHS* and *Sl4CL* with fed *p*-coumaric acid (groups 3 to 5). Last, we expressed each single gene in the absence of fed *p*-coumaric acid (groups 6 to 8). We observed that (i) the removal of fed *p*-coumaric acid eliminates the production of (2) (groups 1 and 2), (ii) the removal of the expression of *Sl4CL* or *SlCHS* eliminates the production of (2) (groups 3 and 4), (iii) the removal of the expression of *SlMT2* eliminates the production of both (1) and (2) (group 5), and (iv) the single expression of *SlMT2* without fed *p*-coumaric acid leads to production of (1) (groups 6 to 8). The observations (i) and (ii) indicate that *p*-coumaric acid is a precursor for the production of (2), and both *Sl4CL* and *SlCHS* are required for the production of (2). The observations (iii) and (iv) indicate that *SlMT2* is responsible for the production of (1), which is independent of fed *p*-coumaric acid, and that (1) is likely a substrate for the production of (2).

On the basis of the production patterns of (1) and (2) under different enzyme combinations, we proposed the sequencing of intermediates along the reconstructed pathway in yeast. *Sl4CL* is known to catalyze the conversion of *p*-coumaric acid to *p*-coumaroyl-CoA (19), and we observed that *p*-coumaric acid is an essential precursor for the production of (2) through the reconstructed pathway; thus, we hypothesized that *p*-coumaroyl-CoA is likely an intermediate of the pathway. A previous study reported that a group of methyltransferases from the salicylic acid benzoic acid theobromine (SABATH) enzyme family in maize is able to catalyze conversion of anthranilic acid to methyl anthranilate, a volatile methyl ester with potential function in plant defense (20). We hypothesized that *SlMT2* may use an anthranilate analog from yeast native metabolism (as the pathway precursor) and catalyze its conversion to a methyl ester (as a pathway intermediate). By searching anthranilate-related yeast native metabolites in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, we identified 3-HAA, a primary metabolite involved in tryptophan metabolism, as a putative substrate for the *SlMT2* methyltransferase and proposed the compound structure for the methyl ester (1) (Fig. 1A). We confirmed the compound structure of (1) with its chemical standard by retention time and tandem mass (MS/MS) spectrum (fig. S1C).

The data further support that (2) is the final product of the reconstructed pathway in yeast. Specifically, (2) may result from the condensation of the two identified intermediates, 3-HAA methyl ester (1) and *p*-coumaroyl-CoA, through the formation of an amide bond potentially catalyzed by *SlCHS*. However, direct condensation of the two intermediates would lead to a final *m/z* of 314.1023 ($[M + H]^+$), whereas the final *m/z* we observed for (2) from yeast culture was *m/z* 316.1179 ($[M + H]^+$). A native yeast ECR, encoded by *TSC13*, has been reported to reduce *p*-coumaroyl-CoA to *p*-dihydro-

coumaroyl-CoA (21). We hypothesized that native *Tsc13p* activity in yeast may reduce *p*-coumaroyl-CoA to *p*-dihydro-coumaroyl-CoA and that *SlCHS* catalyzes the condensation of *p*-dihydro-coumaroyl-CoA with (1), leading to production of (2) (Fig. 1E).

To validate our hypothesis, we used CSY1302 (which harbors chromosomally integrated *SlCHS*, *SlMT2*, and *ySl4CL*) to engineer *TSC13* knockout strains. As deletion of *TSC13* inhibited cellular growth due to its essential role in fatty acid synthesis (22), we partially disrupted *Tsc13p* activity by inserting three consecutive stop codons at two-thirds the length of *TSC13* coding sequence, resulting in strain CSY1304. The insertion of stop codons in *TSC13* may lead to low activity through a low frequency of stop-codon readthrough, enabling very low expression of *Tsc13p*. Stop-codon readthrough has been reported in yeast, where readthrough efficiencies can be as high as 8% (23) and be induced by stress conditions (24). We also replaced *TSC13* with heterologous ECR variants from *Gossypium hirsutum* (*GhECR2*) and *Malus domestica* (*MdECR*) that were reported to have low activity on *p*-coumaroyl-CoA (21), resulting in CSY1305 (*TSC13* $\Delta::GhECR2$) and CSY1306 (*TSC13* $\Delta::MdECR$), respectively. We cultured CSY1304 to CSY1306 in synthetic complete media supplemented with 100 μ M *p*-coumaric acid and 100 μ M 3-HAA methyl ester for 72 hours at 30°C and analyzed the yeast culture media for production of (1) and (2) on LC-MS/MS by multiple reaction monitoring (MRM) detection. Partial disruption of the native yeast ECR *Tsc13p* (CSY1304) resulted in a 40% reduction in production of (2), while replacement of *Tsc13p* with heterologous ECR variants (CSY1305 and CSY1306) resulted in the absence of production of (2) and the presence of a previously unknown compound (3), with an expected *m/z* of 314.1 ($[M + H]^+$) corresponding to the oxidized form of (2) (Fig. 1, F and G). The compound identities of (2) and (3) were validated by comparing the retention times and MS/MS spectrums to those of the chemical standards (fig. S1, D and E). The results suggest that the yeast native enzyme participated in the tomato cluster activity and produced a derivative product (2); we eliminated this interference by knocking out the yeast native gene *TSC13*, thereby restoring the true product (3) resulting from the minimal gene cluster (*Sl4CL*, *SlMT2*, and *SlCHS*).

Tomato methyltransferase shows activity on fed 3-HAA in yeast

On the basis of our in vivo functional characterization of *SlMT2*, the methyltransferase recognizes yeast native 3-HAA as a substrate. According to the KEGG pathway database, 3-HAA is involved in central metabolism, i.e., tryptophan metabolism, and the metabolite is also present in tomato. Since no previous studies have been reported on the functional roles of *SlMT1* and *SlMT2*, we investigated the activities of the methyltransferases on hydroxycinnamic acids, amines, and anthranilic acids by feeding these substrates to yeast engineered to express these methyltransferases. Among the three methyltransferases predicted in the tomato cluster, *SlMT3* has been reported to catalyze the methylation of salicylic acid (19). *SlMT1* and *SlMT2* showed high protein sequence similarity to *SlMT3* (78.12 and 81.42%, respectively), indicating that they may similarly exhibit activity on salicylic acid. In addition, the three methyltransferases were initially predicted as tailoring enzymes to modify *p*-coumaric acid and other moieties of hydroxycinnamic acids, contributing to the production of hydroxylated naringenin chalcone and/or methyl esters of hydroxylated naringenin chalcone in tomato flavonoid metabolism (11).

We tested the activity of *SIMT1/2/3* toward a variety of candidate substrates in yeast, including hydroxycinnamic acids (cinnamic, *p*-coumaric, caffeic, and salicylic acids), trace amines (tyramine, tryptamine, octopamine, dopamine, and serotonin), and anthranilic acid analogs (3-HAA and *p*-aminobenzoic acid). Low-copy plasmids encoding the expression of *SIMT1/2/3* or inactive *ccdB* (negative control) were transformed into the wild-type yeast strain (CEN. PK2). The transformed yeast strains expressing one of the methyltransferases (or negative control protein) were cultured in synthetic dropout media fed with 100 μ M of each substrate candidate for 72 hours at 30°C. The resulting yeast media was analyzed on qToF-MS for total ion scan, and the methylation products were evaluated by analyzing differential peaks detected from the transformants compared to the negative control. A methylation product is counted if the m/z ($[M + H]^+$) of a differential peak (between the sample and the negative control) qualifies a putative methylated product catalyzed from the substrate. Among all the potential substrates tested, *SIMT1* and *SIMT2* exhibited detectable activities toward 3-HAA, *p*-coumaric acid, and *p*-aminobenzoic acid (a primary metabolite that shares similar functional groups with 3-HAA), and *SIMT3* exhibited detectable activity only toward 3-HAA. The highest level of the methylation product was observed when supplying 3-HAA to *SIMT2* (Fig. 2). Among the three methyltransferases, *SIMT3* showed the lowest production of the methylation product from 3-HAA, and the methylation products catalyzed from *p*-coumaric acid and *p*-aminobenzoic acid were not detected in our assay. None of *SIMT1/2/3* showed detectable activity toward salicylic acid in the context of the yeast-based feeding assay. We hypothesized that either salicylic acid was not efficiently transported into yeast cells due to previously reported antagonism between salicylic acid and D-glucose (25) or the volatile salicylate methyl ester product may have evaporated. Our results indicate that all three methyltransferases (*SIMT1/2/3*) showed the highest activity toward 3-HAA (among the fed substrates tested) and that *SIMT2* led to the highest production of 3-HAA methyl ester in the yeast-based feeding assay.

Tomato CHS exhibits a previously uncharacterized activity for amide synthesis

Our *in vivo* characterization results of the minimal gene cluster (*SI4CL*, *SIMT2*, and *SICH5*) indicate that *SICH5* can potentially catalyze the condensation of *p*-coumaroyl-CoA and 3-HAA methyl ester, leading to the formation of a nitrogen-carbon (amide) bond. To our knowledge, this study is the first report of amide formation by CHS, which canonically catalyzes Claisen condensation (carbon-carbon bond formation) (26).

We further examined the amide bond catalytic activity of *SICH5* by expressing *SICH5* recombinantly in *Escherichia coli*, purifying the enzyme, and characterizing its activities via *in vitro* enzymatic as-

says. *SICH5* activity was examined with both its canonical substrates (malonyl-CoA and *p*-coumaroyl-CoA) and the substrates identified in the context of the minimal tomato gene cluster (3-HAA methyl ester and *p*-coumaroyl-CoA). The reactions were performed by incubating 4 μ g of purified enzyme with 200 μ M malonyl-CoA or 3-HAA methyl ester and 200 μ M *p*-coumaroyl-CoA for 4 hours and analyzed on LC-MS/MS by MRM detection. For *SICH5* canonical activity characterization, we observed spontaneous conversion of naringenin chalcone to naringenin under the *in vitro* reaction conditions, and we confirmed the production of naringenin by comparing the resultant peak with an authentic standard of naringenin (fig. S2A). We observed the production of (3) when 3-HAA methyl ester was added to the reaction mixture by comparing the peaks with a chemically synthesized standard of (3). The chemical standard of (3) yielded a single peak when dissolved in water (retention time, 6.872 min) but resulted in a secondary peak (retention time of 7.484 min) when dissolved in acidic methanol (fig. S2B). The secondary peak was also detected in acidic methanol-quenched *in vitro* reaction mixtures, from which the detection of (3) is expected. A previous study compared nonenzymatic and chalcone isomerase-catalyzed conversion of chalcone to flavanone and the pH dependence of this reaction (27). We hypothesized that the secondary peak could result from an isomerized form of (3), similar to the isomerization process of converting naringenin chalcone to naringenin, possibly formed during the *in vitro* reaction. Together, these results validate that *SICH5* is capable of amide formation.

We next examined whether the amide synthesis interferes with the canonical activity. We performed an *in vitro* reaction with *SICH5* under similar conditions but incubated equimolar amounts (200 μ M) of 3-HAA methyl ester and malonyl-CoA with 200 μ M *p*-coumaroyl-CoA. Analysis of the reaction products showed an 85% decrease in production of (3) (Fig. 3, reactions 2 and 3) and 6% decrease in production of naringenin (Fig. 3, reactions 1 and 3). The results suggest that 3-HAA methyl ester is likely competing with malonyl-CoA for a *p*-coumaroyl starter molecule at the *SICH5* active site, indicating that *SICH5* could use the same active site for amide formation as for Claisen condensation.

We next investigated whether *SICH5* exhibited a substrate specificity toward 3-HAA methyl ester for amide synthesis. We incubated *SICH5* with 200 μ M anthranilic acid analog and 200 μ M *p*-coumaroyl-CoA with similar *in vitro* reaction conditions, and the reaction mixture was analyzed on LC-MS/MS by product ion scan with a precursor ions set to match the m/z of expected condensation products. We tested numerous anthranilic acid analogs in this assay, including 3-HAA methyl ester, 2-amino-3/4/5-methoxybenzoic acid, 3-HAA, 2-amino-5-hydroxybenzoic acid, 3-hydroxybenzoic methyl ester, and anthranilic acid. Analysis of the m/z ($[M + H]^+$) of the expected product for each substrate indicated product peaks with 3-HAA methyl ester, 2-amino-5-methoxybenzoic acid, and 3-hydroxybenzoic methyl

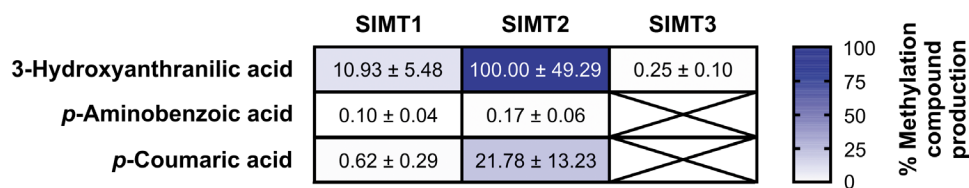


Fig. 2. Relative production of methylation products by *SIMT1/2/3* in yeast. Relative production of methylation products was calculated as a percentage of the highest production by *SIMT2* from substrate 3-HAA: 100% corresponds to the concentration of 3-HAA methyl ester (146 μ M) catalyzed from yeast endogenous 3-HAA and 100 μ M 3-HAA fed to yeast culture medium. Compounds not detected were crossed out. Data show the mean and SD of three biologically independent replicates.

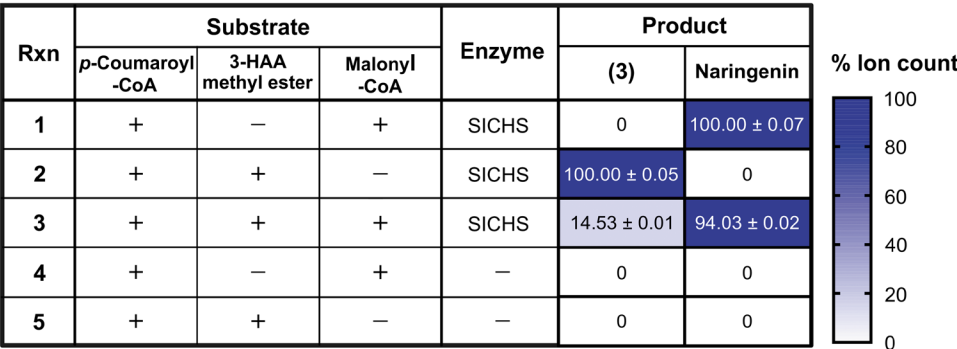


Fig. 3. In vitro characterization of SlCHS activity. +/– indicates the presence/absence of 200 μM *p*-coumaroyl-CoA, 200 μM 3-HAA methyl ester, 200 μM malonyl-CoA, or 4 μg of purified SlCHS protein. MRM (314.1 → 147.0) and MRM (273.0 → 152.8) detect the production of coumaroyl anthranilate amide (3) and naringenin, respectively. The ion counts are normalized by the highest ion count across reaction (rxn) 1 to 5 by each column; SD shows the percentage error among two independent replicates. Enzyme abbreviation: SlCHS, naringenin chalcone synthase.

ester, among which 3-HAA methyl ester yielded more than 15-fold and 49-fold higher product ion detected than those of 2-amino-5-methoxybenzoic acid and 3-hydroxybenzoic methyl ester, respectively (fig. S2C). In contrast, no amide product was observed when 3-HAA and anthranilic acid, which share a very similar molecular structure with 3-HAA methyl ester, were included in the reaction mixture. A trace amount of a possible ester product was observed when 3-hydroxybenzoic methyl ester was included as a substrate. The observed substrate preferences of SlCHS on the panel of anthranilic acid analogs tested indicate that methylation on the carboxyl group of the anthranilate may facilitate substrate access to the SlCHS active site and that SlCHS exhibits a high substrate preference toward 3-HAA methyl ester.

Last, we examined whether the observed amide synthesis activity was specific to the CHS variant from tomato (SlCHS). Specifically, we performed in vitro reaction assays with the CHS variant from *Arabidopsis* (AtCHS). AtCHS was recombinantly expressed in *E. coli* and purified, and its activities on malonyl-CoA and 3-HAA methyl ester were analyzed under the same assay conditions as were used for SlCHS. AtCHS exhibits identical patterns of catalytic activity and substrate preferences as SlCHS in vitro, i.e., highest production of amide with 3-HAA methyl ester, trace amounts of amide production with 2-amino-5-methoxybenzoic acid, and ester production with 3-hydroxybenzoic methyl ester (fig. S2, D and E). Together, the results indicate that the amide synthesis activity observed in SlCHS is not unique to this variant and could be a common secondary function in plant CHS enzymes.

In vivo site-directed mutagenesis confirms the utilization of CHS active site for both naringenin chalcone and amide synthesis

Type III PKSs are characterized by a conserved cysteine-histidine-asparagine catalytic triad, which corresponds to C164-H303-N336 in SlCHS. For canonical synthesis of naringenin chalcone, C164 and H303 form an imidazolium ion pair, which initiates a nucleophilic attack on the thioester carbonyl of *p*-coumaroyl-CoA that completes acyl transfer onto C164 (28). H303 and N336 coordinate the orientation of the incoming malonyl-CoA moieties during the process of iterative decarboxylation and condensation of the extender malonyl-CoA molecules in formation of the polyketide intermediate. In addition, F215 is an important gatekeeper residue that is reported to separate the CoA-binding tunnel from the active site cavity and

help with folding and internal orientation of the tetraketide intermediate (28–30). On the basis of our in vitro assay results, we hypothesized that SlCHS is likely to use the same active site for amide synthesis as for naringenin chalcone synthesis. We therefore investigated the catalytic mechanism of amide bond formation by examining the roles of these active site residues that are important for SlCHS canonical activity.

We first evaluated which residues could potentially interact with 3-HAA methyl ester and use the substrate for amide formation. We built a homology model for SlCHS using Phyre2 (31) and simulated the docking of 3-HAA methyl ester to the homology model structure using AutoDock Vina (32). The simulation shows that 3-HAA methyl ester favorably docks at the SlCHS active site, potentially interacting with H303, N336, and G305 by hydrogen bonding (Fig. 4A, fig. S3A). As a comparison, we simulated the docking of the canonical substrate malonyl-CoA to the SlCHS active site (fig. S3B), which shows that the substrate 3-HAA methyl ester is much smaller in size (molecular weight, 153 versus 854) than the canonical substrate and therefore can readily dock at the active site cavity.

On the basis of the results of the docking simulation, we first investigated the roles of C164, H303, N336 (canonical catalytic triad residues), and G305 on amide synthesis. We created a SlCHS knockout strain (CSY1307) by replacing the full sequence of SlCHS with three consecutive stop codons in CSY1305 (which harbors chromosomally integrated Sl4CL, SlMT2, SlCHS, and TSC13Δ::GhECR2). Low-copy plasmids encoding SlCHS point mutants (C164A, C164S, H303A, N336A, and G305A) were constructed and transformed into CSY1307. Transformed CSY1307 strains harboring individual SlCHS mutants were cultured in synthetic dropout media supplemented with 100 μM *p*-coumaric acid and 100 μM 3-HAA methyl ester for 72 hours at 30°C. Yeast culture media was analyzed for production of naringenin chalcone and (3) on LC-MS/MS by MRM detection. C164A, C164S, and H303A mutants completely eliminated both the canonical activity and the amide synthesis activity (Fig. 4B). The N336A mutant completely abolished naringenin chalcone production but resulted in an increase in the production of (3) compared to the wild-type variant, whereas the G305A mutant abolished canonical activity but exhibited only trace amounts of amide formation. The results indicate that C164 and H303 are essential for both canonical and amide synthesis, which is expected as these two residues are responsible for the loading of *p*-coumaroyl-CoA. The C164S mutant confirms the importance of the thiol group of cysteine

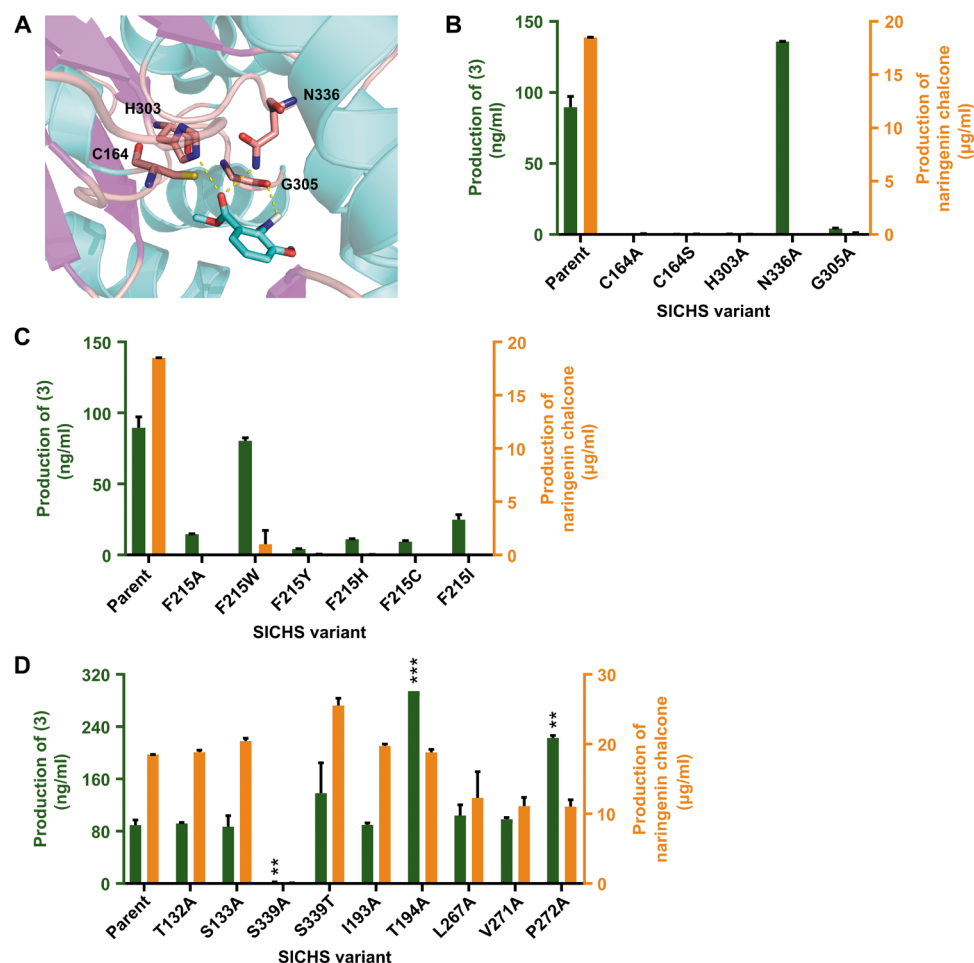


Fig. 4. Investigation of SICHS amide catalytic mechanism by active site modeling and in vivo site-directed mutagenesis. (A) Docking of (1) to SICHS active site. Dotted line, hydrogen bond interaction. (B to D) Production of (3) and naringenin chalcone in yeast by SICHS for C164, H303, N336, and G305 mutants (B); F215 mutants (C); and distal [~ 10 Å within docking site of (1)] residue mutants (D). Data show the mean of two biologically independent replicates with error bar indicating the SD. Unpaired two-tailed *t* test was performed between each variant and the parent for production of (3): ***P* < 0.01 and ****P* < 0.001 (D). Compound name: (1), methyl 3-hydroxyanthranilic acid; (3), coumaroyl anthranilate amide. Enzyme abbreviation: SICHS, naringenin chalcone synthase.

for forming the imidazolium ion pair with H303 to activate acyl transfer through nucleophilic attack during loading of *p*-coumaroyl-CoA onto C164. Although N336 is essential for canonical activity for binding of extender malonyl-CoA, it does not contribute to binding of 3-HAA methyl ester to the active site. This result is further supported by an uninterrupted docking of 3-HAA methyl ester to the active site of a N336A mutant homology model using AutoDock Vina (fig. S3C). The increase in production of (3) observed from the N336A mutant relative to the parent enzyme is likely due to a lack of competition between 3-HAA methyl ester and malonyl-CoA for the *p*-coumaroyl starter moiety at the active site of the N336A mutant. Last, the removal of amide and canonical activities observed in the G305A mutant suggests that G305 potentially performs a stabilizing role in anchoring 3-HAA methyl ester (as predicted by the docking simulation) and malonyl-CoA during their respective condensation reactions.

We next examined potential effects of F215 on amide formation (Fig. 4C). We tested different mutants of the residue to conserve either the ring structure (F215W, F215Y, and F215H) or spatial occupancy (F215I) of the residue side chain. Low-copy plasmids encoding

SICHS mutants (F215A, F215W, F215Y, F215H, F215C, and F215I) were each transformed into CSY1307. The transformed CSY1307 strains were cultured under identical conditions, and production of naringenin chalcone and (3) was analyzed on LC-MS/MS by MRM detection. All F215 mutants except F215W completely abolished the canonical activity, where F215W maintained only 5% naringenin chalcone production as compared to the wild-type variant (Fig. 4C). The results support the previously proposed role of F215 in orienting malonyl-CoA and polyketide intermediates at the active site (29, 30). We also observed that all mutants except F215W led to $\geq 70\%$ reduction in production of (3), while F215W maintained 90% production of (3) compared to the wild-type variant (Fig. 4C). The results suggest that the ring structure of residue 215 in wild-type and the F215W mutant may assist in orienting 3-HAA methyl ester at the active site to facilitate amide formation. However, the ring structure itself in the residue is not sufficient for 3-HAA methyl ester binding since decreased production of (3) was observed in F215Y and F215H (which conserved the ring structure); instead, spatial occupancy (F215I) by the residue may also contribute to substrate selection. Furthermore, reduced production of (3) observed in the F215Y and

F215H mutants could result from a poorly oriented residue side chain shielding the active site, thus preventing the access of 3-HAA methyl ester to C164-bound *p*-coumaroyl moiety. We also scanned for the production of pyrone derivatives bis-noryangonin (BNY) and 4-coumaroyltriacetic acid lactone (CTAL), the former a triketide and the latter a tetraketide early-released derailment by-product (29, 33), by F215 mutants in yeast culture media. We observed proportional levels of CTAL production compared to that of naringenin chalcone and no detectable levels of BNY production (fig. S4A). The results suggest that inhibited production of (3) by F215 mutants is unlikely due to pyrone by-product accumulation at the SlCHS active site. In summary, the results indicate that although F215 likely performs a specific structural role in orienting malonyl-CoA during extension of the polyketide intermediate in canonical activity, its function is less specific for selecting 3-HAA methyl ester as a substrate.

Last, we investigated the potential effects of nonspecific binding by 3-HAA methyl ester to SlCHS protein. We mutated nine residues (T132A, S133A, S339A, S339T, I193A, T194A, L267A, V271A, and P272A) within ~10 Å of the 3-HAA methyl ester docking site and analyzed the effects of these mutations on production of (3) in yeast (Fig. 4D and fig. S3D). CSY1307 strains transformed with the mutants encoded on low-copy plasmids were cultured under identical conditions, and production of naringenin chalcone and (3) was analyzed on LC-MS/MS by MRM detection. The results showed that most of the nine tested residues did not show statistically significant effects on production of (3), except for S339A, T194A, and P272A (Fig. 4D). S339A completely abolished SlCHS activity, and the two distal residue mutants (T194A and P272A) significantly improved SlCHS activity for production of (3). Since S339 is located at a loop structure near the SlCHS active site, the mutation may have interrupted the correct folding of the active site cavity and therefore disrupted both naringenin chalcone and amide synthesis. Removal of the two distal residues (T194A and P272A) may have altered the entrance geometry of the active site cavity, which facilitated the access of 3-HAA methyl ester to the active site and therefore increased production of (3). Similarly, fluctuations in the production of naringenin chalcone observed among the mutants could be caused by an altered geometry around the active site, which affected the access of *p*-coumaroyl-CoA or malonyl-CoA to the active site.

In vitro kinetic characterization shows competitive inhibition of CHS canonical activity by 3-HAA methyl ester

The results of the site-directed mutagenesis studies suggest that SlCHS uses the same active site for canonical and amide synthesis. We performed in vitro enzymatic assays to further investigate the kinetic properties of SlCHS on 3-HAA methyl ester. Kinetic assays were performed by incubating purified SlCHS with *p*-coumaroyl-CoA and varying concentrations of 3-HAA methyl ester (0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6, 3, 5, 10, and 15 mM). The reactions were stopped at different time points, and reaction products were analyzed on LC-MS/MS to derive the kinetic curve (Fig. 5A and fig. S5A). The kinetic data show that the amide synthesis has a K_m (Michaelis-Menten constant) of 3.06 mM and a V_{max} of 14.47 nM min⁻¹, resulting in a k_{cat} of 0.362 min⁻¹ and k_{cat}/K_m of 1.18×10^{-4} μM⁻¹ min⁻¹ under the in vitro reaction conditions (Fig. 5A). As a comparison, we performed in vitro enzymatic assays to characterize the kinetic properties of SlCHS canonical activity by incubating purified SlCHS with *p*-coumaroyl-CoA and varying concentrations of malonyl-CoA (0, 5, 50, 100, 200, 300, and 500 μM). The canonical

synthesis of naringenin chalcone has a K_m of 21.34 μM and V_{max} of 11.32 nM min⁻¹, resulting in a k_{cat} of 0.0943 min⁻¹ and k_{cat}/K_m of 4.42×10^{-3} μM⁻¹ min⁻¹ (fig. S5B). The results show a 143-fold difference between SlCHS's K_m for 3-HAA methyl ester and malonyl-CoA, indicating that the enzyme has a much higher affinity for malonyl-CoA than for 3-HAA methyl ester. The results also show a 37-fold higher catalytic efficiency (k_{cat}/K_m) of SlCHS for synthesis of naringenin chalcone than for that of amide. Together, the results indicate that amide synthesis is likely to be a less efficient secondary function of SlCHS.

We next examined the mechanism of 3-HAA methyl ester inhibition of SlCHS canonical activity. Kinetic assays were performed by incubating purified SlCHS with *p*-coumaroyl-CoA and varying concentrations of malonyl-CoA (5, 50, and 100 μM) as the substrate and 3-HAA methyl ester (0, 3, and 5 mM) as the inhibitor. The reactions were stopped at different time points, and reaction products were analyzed on LC-MS/MS to derive the kinetic curve for each inhibitor concentration (Fig. 5B). For the purpose of curve-fitting, only malonyl-CoA was considered as the substrate, since the reactions were performed under saturated concentrations of *p*-coumaroyl-CoA (200 μM). We first fit all data points (measured under 0, 3, and 5 mM inhibitor) to Eq. 1 (Fig. 5, B and C). By tuning the Hill coefficient, we observed that root mean square error (RMSE) is minimized for data points of 0 mM when $n = 1$, for data points of 3 mM when $n = 1.7$, and for data points of 5 mM when $n = 1.5$ (Fig. 5B and table S2A). The curve-fitting results suggest that the effects of cooperativity emerge only when inhibitors are present.

We then fit the data points taken under 3 and 5 mM inhibitors to competitive (Fig. 5, Eq. 2), uncompetitive (Fig. 5, Eq. 3), or mixed-type (Fig. 5, Eq. 4) inhibition modes to interpret inhibition coefficients (K_c for competitive inhibition and K_u for uncompetitive inhibition) by fixing the values for K_m and k_{cat} at those obtained at 0 mM inhibitor (Fig. 5, C to E). Here, we used the Hill coefficient n to represent the effect of cooperativity resulting from sequential binding of three molecules of malonyl-CoA to coumaroyl-bound enzyme complex. For the data points obtained under 3 and 5 mM inhibitors, we observed minimization of RMSE with the mixed-type inhibition model, and the best fits were obtained at $n = 1.7$ and 1.5 for 3 and 5 mM inhibitors, respectively (Fig. 5, D and E, and table S2D). For 3 mM inhibitor, $K_c = 0.377$ mM and $K_u = 1.01$ mM ($K_u/K_c = 2.67$). For 5 mM inhibitor, $K_c = 0.341$ mM and $K_u = 0.897$ mM ($K_u/K_c = 2.63$). Together, the results indicate that inhibition is dominated by competitive mode in both cases with a shift from competitive to uncompetitive mode as inhibitor concentration increases from 3 to 5 mM.

Last, we investigated the production of pyrone derivatives BNY and CTAL by SlCHS when inhibited by 3-HAA methyl ester. We scanned for BNY and CTAL production from reaction mixtures fed with 100 μM malonyl-CoA; 100 μM *p*-coumaroyl-CoA; and 0, 3, or 5 mM 3-HAA methyl ester inhibitor at the end of the kinetic assay time course. We detected proportional levels of CTAL production compared to that of naringenin and no detectable levels of BNY production (fig. S4, B and C). The results suggest that 3-HAA is unlikely to promote the release of derailment by-products due to early termination in extension and/or cyclization during polyketide synthesis.

DISCUSSION

We leveraged a yeast biosynthesis platform to characterize the activity of a computationally predicted biosynthetic gene cluster from

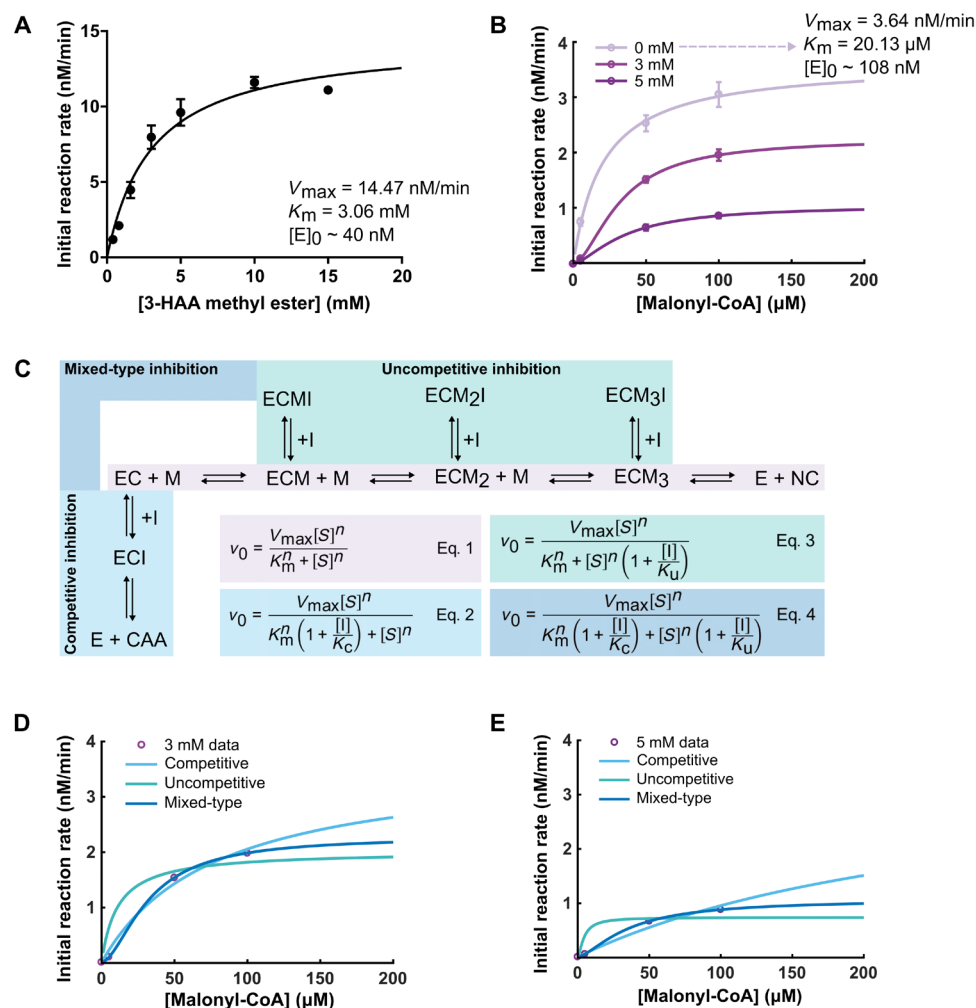


Fig. 5. SICHS in vitro kinetic assays for amide formation and inhibition mechanism to canonical activity. (A) Kinetic characterization of SICHS synthesis of coumaroyl anthranilate amide (3). (B) Kinetic characterization of SICHS synthesis of naringenin chalcone, inhibited with 0, 3, or 5 mM 3-HAA methyl ester. (C) Proposed inhibition mechanisms of 3-HAA methyl ester to SICHS canonical activity. E, enzyme (SICHS); EC, enzyme-coumaroyl complex; I, inhibitor (3-HAA methyl ester); ECI, enzyme-coumaroyl-inhibitor complex; CAA, coumaroyl anthranilate amide; M, malonyl-CoA; ECM, enzyme-diketide complex; ECM₂, enzyme-triketide complex; ECM₃, enzyme-tetraketide complex; NC, naringenin chalcone; ECMI, enzyme-diketide-inhibitor complex; ECM₂I, enzyme-triketide-inhibitor complex; ECM₃I, enzyme-tetraketide-inhibitor complex. Equation notations: v_0 , initial velocity; V_{max} , maximal velocity; K_m , Michaelis-Menten constant; S, substrate (i.e., malonyl-CoA); K_c , competitive inhibition coefficient; K_u , uncompetitive inhibition coefficient; n , Hill coefficient that simulates cooperativity effect by sequential binding of malonyl-CoA to the coumaroyl-bound enzyme complex. (D and E) Analysis on mode of inhibition by 3 mM (D) and 5 mM (E) 3-HAA methyl ester. Eq. 1, no inhibition; Eq. 2, competitive inhibition; Eq. 3, uncompetitive inhibition; Eq. 4, mixed-type inhibition. Data show the mean of two independent replicates, with error bar indicating the SD.

tomato, which led to the discovery of a previously undocumented HCAA compound and the potential of CHS for nitrogen-carbon bond synthesis. The HCAA compound is generated by the condensation of a hydroxycinnamic acid moiety and anthranilic acid moiety through formation of an amide bond. We showed that one of the substrates for HCAA production in yeast was 3-HAA methyl ester, which was converted from the native metabolite, 3-HAA, by each of the three methyltransferases in the predicted tomato gene cluster. Among the methyltransferases, SMIT2 exhibited the highest activity toward 3-HAA in yeast. Through systematic mutagenesis, in vivo activity screens, and in vitro substrate competition assays, we showed that SICHS uses the same active site for its canonical naringenin chalcone synthesis activity to catalyze the condensation of 3-HAA methyl ester and *p*-coumaroyl-CoA, leading to the production of coumaroyl an-

thranilate amide (3). To our knowledge, this is the first report of a type III PKS enzyme exhibiting amide bond formation activity. In vitro kinetic assays indicate that SICHS catalyzes the formation of (3) with a K_m of 3.06 mM for 3-HAA methyl ester.

To examine the catalytic mechanism of CHS for HCAA synthesis, we referred to mechanisms of other classes of enzymes that catalyze similar reactions. Specifically, the acyl-CoA *N*-acyltransferases are a category of benzylalcohol acetyl-, anthocyanin-O-hydroxy-cinnamoyl-, anthranilate-N-hydroxy-cinnamoyl/benzoyl-, deacetylindoline (BAHD) acyltransferases that catalyze the formation of HCAA in plants (34–41) and share a conserved HXXXDG domain, positioned near the center of the enzyme (38). A histidine residue in the HXXXDG motif deprotonates the oxygen or nitrogen atom on the corresponding acceptor substrate, thereby allowing a nucleophilic attack on the

carbonyl carbon of the CoA thioester and leading to the formation of a tetrahedral intermediate between the CoA thioester and acceptor substrate (39). The intermediate is reprotonated to release the free CoA and the acylated ester or amide. The aspartic acid residue in the conserved motif plays a structural rather than catalytic role by forming a salt bridge with a conserved arginine residue (39). Another family of enzymes, arylamine *N*-acetyltransferases (NATs), catalyzes a similar reaction that transfers an acetyl group from acetyl-CoA to the terminal nitrogen group of an arylamine substrate (42). The reaction is catalyzed by a cysteine-histidine-aspartic acid catalytic triad and is initiated by nucleophilic attack of the carbonyl group on acetyl-CoA by cysteine, activated by the histidine residue likely through formation of a thiolate-imidazolium ion pair (43, 44). The incoming arylamine attacks the carbonyl group bound to cysteine in forming a tetrahedral intermediate, with a general base deprotonating the amine group. Similarly to BAHD acyltransferases, it has been suggested that the deprotonation in NATs is assisted by the histidine residue in the catalytic triad (43). The aspartic acid residue was proposed to form a low-barrier hydrogen bond with the histidine residue to increase the basicity of the histidine for cysteine activation (43).

The catalytic mechanisms for BAHD acyltransferases and NATs suggest the potential roles of histidine at the SlCHS catalytic triad (C164-H303-N336) in (i) cysteine activation before nucleophilic attack of the carbonyl group of *p*-coumaroyl-CoA and (ii) deprotonating the incoming amine nucleophile in formation of a tetrahedral intermediate bound to cysteine. Previous studies on CHS catalytic mechanisms support (i) that H303 and C164 form a thiolate-imidazolium ion pair, which facilitates the nucleophilic attack of the thiolate anion on the thioester carbonyl of *p*-coumaroyl-CoA, resulting in transfer of the acyl moiety to C164 (28). Our *in vivo* mutagenesis data indicate that C164 and H303 are critical for canonical and amide synthesis. Therefore, it is likely that the mechanism for cysteine activation and acyl transfer is conserved for amide formation (fig. S6, A and B). In the next step, incoming 3-HAA methyl ester forms a covalent bond with the coumaroyl moiety bound to C164 by nucleophilic attack of the amine group on the carbonyl group of the coumaroyl moiety, leading to formation of a tetrahedral intermediate (fig. S6, C and D). The positively charged amide is then deprotonated by an unidentified general base (fig. S6, D and E), followed by release of the amide product (fig. S6F). H303 may play the role of the unidentified general base in deprotonating the incoming amine nucleophile as suggested for NATs (43); however, this process requires H303 to be regenerated (deprotonation of the imidazolium) after accepting a proton from a thiol group upon acyl transfer from *p*-coumaroyl-CoA to cysteine, the exact mechanism for which was not determined in this study.

Prior studies on CHS activity reported that mutations in an active site residue (F215) and acidification of *in vitro* reaction mixtures before extraction can lead to an increase in production of BNY and CTAL (29). In this work, we observed proportional levels of CTAL production compared to that of naringenin chalcone and no detectable levels of BNY production from CHS *in vitro* reaction mixtures. We also did not observe increases in BNY and CTAL from the F215 mutants expressed in yeast, in contrast to previously reported *in vitro* characterization of F215 mutants (29). The study reported the production of BNY from F215A and F215H mutants and CTAL from F215Y mutant, where BNY production was maximized at pH 7.0, and CTAL production was prominent within a pH range of 6.0

to 6.5 (29). The absence of detectable BNY and CTAL production by F215 mutants in our work may be due to differences in characterization conditions, i.e., yeast versus *in vitro*, and specifically may be due to the acidic pH ≤ 5.8 of yeast synthetic complete media. The observation also indicates that inhibited production of (3) observed with F215 mutants is not likely due to pyrone by-product accumulation at the CHS active site.

We observed that CHS exhibits catalytic promiscuity by catalyzing the synthesis of two different families of compounds: polyketide through its canonical activity and HCAA through the secondary activity characterized here. The syntheses of other HCAA compounds—e.g., *p*-coumaroyltyramine, *p*-coumaroyldopamine, and feruloyldopamine—by hydroxycinnamoyl-CoA:tyramine *N*-hydroxycinnamoyl transferase (THT), have been reported in tomato for defense against bacterial and fungal pathogens (45, 46). There is currently limited evidence to support that this secondary activity of CHS may be adapted by the plant host for HCAA synthesis, considering that the secondary activity shows ~40-fold lower efficiency (k_{cat}/K_m) compared to the canonical activity. However, this catalytic promiscuity may indicate a starting point for evolution of the enzyme to become an alternative route for HCAA compound production (47). For example, future work can compare the amine substrate specificity of both THT and CHS for HCAA synthesis, which may indicate an evolutionary advantage of CHS to catalyze hydroxycinnamoyl anthranilate-type HCAA if CHS shows higher activity toward anthranilic acid analogs than THT. Additional future work may focus on validating a role of the gene cluster in the native host by knocking out individual genes in tomato and performing metabolomics to search for metabolites that may be associated with the gene cluster. However, if the genes in the cluster are associated with a “cryptic pathway,” identification of a proper elicitor treatment would be required to induce the silent gene cluster and production of the target compound(s) in the host.

As more than 1000 putative plant gene clusters have now been predicted via computational tools (7, 9–11), future advances that further streamline high-throughput characterization workflows will be critical to characterizing activities encoded within these clusters. For example, future efforts may develop systematic criteria to prioritize gene clusters for yeast-based characterization and reliable high-throughput metabolite screening methods to accelerate the exploration of previously unidentified chemical space. Parallel genomic integration of multiple gene clusters can be facilitated by multiplexed CRISPR technology (48). Yeast harboring multiple gene clusters can then be screened for compound production using high-precision metabolomics, where improved computational workflows for untargeted metabolomics analysis can enable more efficient identification of novel low-abundance metabolites to distinguish robustly from background metabolite profiles. Thus, the integration of computational plant genome analysis, yeast-based heterologous pathway expression, and advances in analytics will allow for the streamlined characterization and discovery of biosynthetic routes that may be difficult to uncover in planta.

MATERIALS AND METHODS

Linear DNA template and plasmid construction

DNA sequences for heterologous biosynthetic enzymes were codon-optimized to improve expression in *S. cerevisiae* using GeneArt GeneOptimizer software (Thermo Fisher Scientific, Waltham, MA) and were synthesized as gene fragments (Twist Bioscience, San

Francisco, CA). For guide RNA (gRNA)/Cas9 plasmids, 20-base pair (bp) gRNAs targeting the genomic site were synthesized as primers (*TSC13* gRNA1: AACAGCTCAAATGTACGCAT; *TSC13* gRNA2: ATAACCTTAGCATTCCCAAAG; *SICH5* gRNA: TGTG-GTACATCATCAATCT), overlap polymerase chain reaction (PCR)-amplified with tRNA promoter/hepatitis delta virus (HDV) ribozyme PCR fragment (pCS3411), trans-activating CRISPR RNA (tracrRNA)/terminator PCR fragment (pCS3414), and cloned into a SpCas9 expression vector with G418 resistance (pCS3410) via Gibson assembly (49).

Plasmids for protein expression in *E. coli* were constructed by inserting DNA fragments encoding *At4CL*, *SICH5*, and *AtCHS* into pET28 vector via Gibson assembly, for which the PCR-amplified pET28 vector backbone and the protein inserts share a 40-base pair (bp) overhang at both ends of the linear DNA components. Plasmid encoding the parent *SICH5* protein in the site-directed mutagenesis study was constructed using Gibson assembly. The plasmid vector (pCS3305) was digested by restriction enzymes Xba I and Xho I, and the *SICH5* gene insert was amplified from a gene fragment.

Plasmids for single amino acid mutant variants were constructed either via Gibson assembly or blunt-end ligation. For the Gibson assembly method, primers encoding the single amino acid substitution were used to amplify the parent plasmid and the linear DNA product. The linear DNA product contained a 15-bp overlap between its 5' and 3' ends and was annealed by Gibson assembly. For blunt-end ligation method, a primer pair without overhang was used to amplify the parent plasmid, and the 5' primer encodes the single amino acid substitution. The linear DNA product is then incubated with T4 nucleotide kinase [New England Biolabs (NEB), Ipswich, MA] at 37°C for 30 min and subsequently with T4 DNA ligase (NEB, Ipswich, MA) at room temperature for 2 hours.

All the primers in this work were synthesized by the Stanford Protein and Nucleic Acid Facility (Stanford, CA). PCR amplifications were performed with Q5 High-Fidelity DNA polymerase (NEB, Ipswich, MA), and PCR products were purified using the DNA Clean and Concentrator Kit (Zymo Research, Irvine, CA). Plasmids generated in this work are listed in table S3.

Chemical substrates and standards

The chemical standard for methyl 3-hydroxy-2-(3-(4-hydroxyphenyl)propanamido)benzoate [dihydro-coumaroyl anthranilate amide (2)] and (E)-methyl 3-hydroxy-2-(3-(4-hydroxyphenyl)acrylamido)benzoate [coumaroyl anthranilate amide (3)] was purchased from Toronto Research Chemicals (Canada). Methyl 2-amino-3-hydroxybenzoate [3-HAA methyl ester (1)] was purchased from Apollo Scientific (UK). *p*-Coumaric acid, malonyl-CoA, 3-HAA, methyl 3-hydroxybenzoate, 2-amino-3-methoxybenzoic acid, 2-amino-4-methoxybenzoic acid, 2-amino-5-methoxybenzoic acid, 2-amino-5-hydroxybenzoic acid, and 2-anthranilic acid were purchased from Sigma-Aldrich (St. Louis, MO). Naringenin chalcone was purchased from Biosynth Carbosynth (USA). Naringenin was purchased from MedChemExpress (USA). *p*-coumaroyl-CoA standard was purchased from PlantMetaChem (Germany).

Yeast strain construction and transformation

Yeast strains used in this study are listed in table S3. All yeast strains are haploid, derived from CEN.PK2-1D (50) (MAT α URA3-52, TRP1-289, LEU2-3/112, HIS3 Δ 1, MAL2-8C, and SUC2), referred to as CEN.PK2. Genes in the predicted tomato cluster were codon-

optimized and assembled with corresponding promoter/terminator fragments and integrated into pYES1L (Life Technologies, Carlsbad, CA). To create the minimal pathway strain, the pathway genes (*SICH5*, *Sl4CL*, and *SlMT1/2/3*) were first cloned into pAG414-GDP1p/ADHt, pAG414-PGK1p/PHO5t, pAG414-PYK1p/MFA1t, or pAG414-TEF1p/CYC1t expression vector with Gibson assembly, and the linear DNA fragment for each pathway gene expression cassette with 30-bp overlap between each fragment was PCR-amplified from the pAG vectors, assembled, and integrated into YMR206W Δ ::locus with *SpHIS5* selection marker.

TSC13 and *SICH5* knockout strains were created by CRISPR-Cas9 genome editing method as previously described (51). The linear DNA repair templates were PCR-amplified and harbor a 30- to 45-bp overlap with the target genomic site. Two hundred nanograms of gRNA/SpCas9 plasmid and 500 ng of linear DNA template were cotransformed into yeast competent cells prepared from the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA), as described in the "Yeast strain construction and transformation" section. Colonies picked from G418 plate after 3 days were screened for metabolite production.

For yeast transformations, a single colony of the parent strain was inoculated in yeast peptone with 2% dextrose (YPD) media and incubated overnight at 30°C and 220 rpm. The saturated overnight culture was then diluted 50-fold in fresh YPD media and incubated for 4 to 6 hours. Cells (2.5 ml) were used per transformation. The cells were then harvested by centrifugation at 3500g for 4 min and prepared for transformation using the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA). For plasmid transformations, 50 ng of DNA was used per transformation. The transformed cells were plated directly onto synthetic dropout agar plates after 45-min incubation with EZ3 solution. For Cas9-based chromosomal integrations, 100 ng of the Cas9 plasmid (encodes G418 resistance) and 500 ng of the linear DNA fragments were used per transformation, and the transformed cells were subject to a 2-hour recovery at 30°C in YPD media after 45-min incubation with EZ3 solution. The cells were plated onto synthetic dropout plates supplemented with G418 (400 mg/liter) to select for colonies with successfully integrated constructs. The plate cultures were incubated 2 to 3 days before colonies were picked for metabolite production assays.

In vivo culture and assay conditions

To screen for metabolite production, two or three colonies were inoculated for each strain (or transformed strain) into 400 μ l of synthetic complete or dropout media with 2% dextrose in 2-ml 96-well plates, grown for 16 to 20 hours to saturation, diluted at a 1:8 ratio into fresh media with corresponding feeding conditions, and grown for 72 hours at 25° or 30°C, as indicated, before metabolite analysis of culture supernatant on LC-MS/qToF-MS.

LC-MS/MS analysis of metabolite production

For targeted metabolite production assays, 100 μ l of supernatant of yeast culture from 96-well plates was obtained by centrifugation at 4000g for 5 min. The sample was analyzed by an Agilent 1260 Infinity Binary high-performance LC (HPLC) paired with an Agilent 6420 Triple Quadrupole LC-MS, with a reversed-phase column (Agilent EclipsePlus C18, 2.1 \times 50 mm, 1.8 μ m), water with 0.1% formic acid as solvent A, and acetonitrile with 0.1% formic acid as solvent B, at a constant flow rate of 0.4 ml/min and an injection volume of 5 μ l. The following gradient was used for compound separation: 0 to 6 min,

3 to 50% B; 6 to 9 min, 50 to 97% B; 9 to 10 min, 97% B; 10 to 10.5 min, 97 to 3% B; 10.5 to 11 min, equilibration with 3% B. The liquid chromatogram eluent was directed to the MS for 1 to 10 min with electrospray ionization (ESI) source in positive mode, gas temperature at 350°C, gas flow rate at 10 liters/min, and nebulizer pressure at 50 psi. LC-MS data files were analyzed in Agilent MassHunter Workstation software. The liquid chromatograms and product ion scans were extracted either by specified precursor ion from total ion current or by MRM with ion transitions and related parameters specified in table S4. All the MRM transitions in this work were derived from product ion scan with specified precursor ion, and the most abundant product ion was chosen for MRM transition quantification. For each compound, production was quantified by integrating the peak area under the ion count curve. The ion counts were calibrated to a chemical standard curve and converted to measurements of titer (ng/ml or µg/ml) and molar concentration (nM) for in vivo and in vitro assays, respectively.

qToF-MS analysis of metabolite production

For untargeted metabolite production assays, 200 µl of yeast culture from 96-well plates was flash-frozen, lyophilized overnight, and dissolved in 100 µl of 75% methanol (with 25% water) with 0.1% formic acid. The sample was analyzed by the Agilent 1260 Infinity Binary HPLC paired with an Agilent 6545 Quadrupole Time-of-Flight LC-MS, with a reversed-phase column (Agilent EclipsePlus C18, 2.1 × 50 mm, 1.8 µm), water with 0.1% formic acid as solvent A, and acetonitrile with 0.1% formic acid as solvent B, at a constant flow rate of 0.6 ml/min and an injection volume of 1 µl. The following gradient was used for compound separation: 0 to 0.40 min, 5% B; 0.40 to 8.40 min, 5 to 95% B; 8.40 to 10.40 min, 95% B; 10.40 to 10.41 min, 95 to 5% B; 10.41 to 12.00 min, 5% B. The liquid chromatogram eluent was directed to the MS for 1 to 12 min with ESI source in positive mode, gas temperature at 250°C, gas flow rate at 12 liters/min, nebulizer pressure at 10 psig, Vcap at 3500 V, fragmentor at 100 V, skimmer at 50 V, octupole 1 RF Vpp at 750 V, and acquisition scan rate at 2.50 spectra/s.

Homology modeling and docking simulation

SlCHS homology model was built using Phyre2 (31) from amino acid sequence, with 85% identity with template c1cml chain A from Protein Data Bank. Docking simulation was performed by AutoDock Vina (32), and docking results were visualized using PyMOL. Geometry optimizations of substrate structures before docking simulations were conducted using Gaussian 16 (DFT, B3LYP, and LANL2DZ).

Protein expression and purification

Protein expression plasmids were transformed into *E. coli* BL21(DE3) cells. For each protein construct, single colony was inoculated into 5 ml of LB media with kanamycin (50 mg/liter) and incubated at 37°C and 220 rpm for 16 hours (overnight). Overnight culture (5 ml) was then inoculated into 1 liter of Luria-Bertani (LB) media with kanamycin (50 mg/liter) and incubated at 37°C and 200 rpm for around 5 hours until an optical density at 600 nm (OD₆₀₀) reached 0.6. The culture was then cooled to 18°C, induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside, and incubated for 16 hours at 200 rpm. The cells were harvested by centrifugation at 4000g for 15 min, and all the following steps were performed on ice with prechilled buffers and reagents. The cell pellet was first washed in 50 mM (pH 8.0) Tris buffer, resuspended in lysis buffer [10 mM imidazole, 50 mM sodium phosphate, and 300 mM

sodium chloride (pH 7.4)], and lysed by sonication. The cellular debris was removed from cell lysate by centrifugation at 16,000g and 4°C for 1 hour. The enzyme proteins were purified from the supernatant using Ni-NTA agarose affinity chromatography and eluted using a range of imidazole concentrations (40, 100, 150, 200, 250, and 450 mM) with the target protein most efficiently eluted at 200 mM imidazole. The purified proteins were then buffer-exchanged and concentrated to storage buffer [50 mM potassium phosphate, 100 mM NaCl, and 10% (v/v) glycerol (pH 7.5)]. The protein concentration was determined by NanoDrop and corrected by extinction coefficient. The final yield for all three proteins is ~2.2 mg/ml. Aliquots of the purified proteins were flash-frozen and stored at -80°C.

In vitro synthesis of *p*-coumaroyl-CoA

p-Coumaroyl-CoA was synthesized by a batch of in vitro reactions with purified protein (40 µg/ml) of At4CL, 400 µM *p*-coumaric acid, 400 µM CoA, 4 mM adenosine 5'-triphosphate, and 5 mM MgCl₂, added to a buffer with 50 mM potassium phosphate and 100 mM NaCl at pH 7.5. The reaction mixture was incubated at 37°C and 500 rpm for 4 hours. Aliquots of the reaction products were stored at -20°C.

In vitro enzymatic assays

For SlCHS and AtCHS in vitro activity validation, 4 µg of purified protein and 200 µM *p*-coumaroyl-CoA were incubated with 200 µM malonyl-CoA and/or 3-HAA methyl ester in a 50-µl reaction volume at 30°C and 450 rpm for 4 hours in the dark. The reaction volume was quenched in equal volume of acidic methanol (with 0.1% formic acid), the mixture was centrifuged at 32,000g for 10 min, and the supernatant was used for LC-MS analysis. For the specificity assay, 4 µg of purified protein and 200 µM *p*-coumaroyl-CoA were incubated with 200 µM 3-HAA, methyl 3-hydroxybenzoate, 2-amino-3-methoxybenzoic acid, 2-amino-4-methoxybenzoic acid, 2-amino-5-methoxybenzoic acid, 2-amino-5-hydroxybenzoic acid, or 2-anthranilic acid, with the same incubation and extraction protocol described above.

For amide synthesis kinetic assays, 680 or 40 nM purified SlCHS protein and 200 or 500 µM *p*-coumaroyl-CoA were incubated with 0, 1, 5, 10, 50, 100, and 200 µM or 0, 0.4, 0.8, 1.6, 3, 5, 10, and 15 mM 3-HAA methyl ester. For canonical activity kinetic assay, 120 nM purified SlCHS protein and 200 µM *p*-coumaroyl-CoA were incubated with 0, 5, 50, 100, 200, 300, and 500 µM malonyl-CoA. For each assay, duplicates were performed in 50-µl reaction volumes; incubated at 30°C and 450 rpm under dark condition; and quenched by adding equal volume of acidic methanol (with 0.1% formic acid) at 5, 10, 15, 20, and 25 min (for amide synthesis with low concentration range of 3-HAA methyl ester); at 6, 24, 30, and 36 min (for amide synthesis with high concentration range of 3-HAA methyl ester); or at 5, 10, 17, 24, and 31 min (for canonical activity). The samples were further diluted by adding 30 µl of water and filtered using 0.2 µm filter plates before measurements on LC-MS/MS.

For enzymatic inhibition assays, 108 nM purified SlCHS protein was incubated with 200 µM *p*-coumaroyl-CoA and 5, 50, or 100 µM malonyl-CoA and 0, 3, or 5 mM 3-HAA methyl ester. For each assay, duplicates were performed in 40-µl reaction volumes; incubated at 30°C, 450 rpm under dark condition; and quenched by adding equal volume of acidic methanol (with 0.1% formic acid) at 5, 10, 17, 24, and 31 min. The samples were further diluted by adding 30 µl of

water and filtered using 0.2 μ M filter plates before measurements on LC-MS/MS.

Statistical analysis

For untargeted metabolomic analysis, data were obtained from $n = 3$ biologically independent replicates. Biological independence refers to individual colonies of a yeast strain inoculated into separate culture volumes under the same feeding and growth conditions. qToF-MS data files were converted to mzXML files using MSConvert, and untargeted metabolomics differential analysis was performed using the xcms package in R (52). The differential peaks were then identified by sorting the “diffreport” generated from xcms differential analysis by “fold” parameter, with a filter set for a P value smaller than 0.01.

For metabolite production, each liquid chromatogram trace is representative of two biologically independent replicates. Ion count data show the mean of $n = 2$ or 3 biologically independent replicates, with error bar indicating the SD. Biological independence refers to individual colonies of a yeast strain inoculated into separate culture volumes under the same feeding and growth conditions. Statistical significance analysis was performed (for selected data) by unpaired two-tailed t test.

For in vitro kinetic assay, progress curve data show the mean of compound produced from $n = 2$ independent replicates performed simultaneously in separated reaction volumes, with error bar indicating the SD. For amide synthesis kinetic assays, initial reaction rates and error bars were calculated by fitting progress curves with a built-in linear regression tool in GraphPad Prism 7 for amide formation reactions. For canonical activity inhibition assay by 3-HAA methyl ester, progress curves were fitted using DynaFit (53) through an ordinary differential equation (ODE)-based system derived from the kinetic model specified in fig. S5E. Because of an initial lag phase in the progress curve, the reaction rates were obtained from the first derivative of the progress curve (calculated by DynaFit) and then fitted to the general equation $M(1 - \exp(-ax))$ in MATLAB 2017a, in which “ M ,” i.e., plateau of the rate function, represents the reaction rate at steady state, i.e., linear region of the progress curve. For kinetic curve, data show the slope or M obtained from progress curve data analysis, with error bar representing the relative error (%) of the slope (calculated by GraphPad Prism 7 linear regression tool) or relative RMS (%) for progress curve fitting (calculated by DynaFit). K_m and V_{max} for kinetic data were estimated using built-in Michaelis-Menten kinetic nonlinear regression tool in GraphPad Prism 7 (for amide synthesis) or MATLAB 2017b by fitting data with kinetic equations as specified in Fig. 5C (for canonical activity inhibition assay).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/44/eabd1143/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. A. G. Atanasov, B. Waltenberger, E.-M. Pferschy-Wenzig, T. Linder, C. Wawrosch, P. Uhrin, V. Temml, L. Wang, S. Schwaiger, E. H. Heiss, J. M. Rollinger, D. Schuster, J. M. Breuss, V. Bochkov, M. D. Mihovilovic, B. Kopp, R. Bauer, V. M. Dirsch, H. Stuppner, Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnol. Adv.* **33**, 1582–1614 (2015).
2. T. Aburjai, F. M. Natsheh, Plants used in cosmetics. *Phytother. Res.* **17**, 987–1000 (2003).
3. A. Cravens, J. Payne, C. D. Smolke, Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun.* **10**, 2142 (2019).
4. S. Li, Y. Li, C. D. Smolke, Strategies for microbial synthesis of high-value phytochemicals. *Nat. Chem.* **10**, 395–404 (2018).
5. H.-W. Nützmann, A. Huang, A. Osbourn, Plant metabolic clusters – from genetics to genomics. *New Phytol.* **211**, 771–789 (2016).
6. E. T. Wurtzel, T. M. Kutchan, Plant metabolism, the diverse chemistry set of the future. *Science* **353**, 1232–1236 (2016).
7. A. K. Chavali, S. Y. Rhee, Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief. Bioinform.* **19**, 1022–1034 (2018).
8. M. J. Smanski, H. Zhou, J. Claesen, B. Shen, M. A. Fischbach, C. A. Voigt, Synthetic biology to access and expand nature’s chemical diversity. *Nat. Rev. Microbiol.* **14**, 135–149 (2016).
9. N. Töpfer, L.-M. Fuchs, A. Aharoni, The PhytoClust tool for metabolic gene clusters discovery in plant genomes. *Nucleic Acids Res.* **45**, 7049–7063 (2017).
10. S. A. Kautsar, H. G. Suarez Duran, K. Blin, A. Osbourn, M. H. Medema, PlantSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* **45**, W55–W63 (2017).
11. P. Schlöpfer, P. Zhang, C. Wang, T. Kim, M. Banf, L. Chae, K. Dreher, A. K. Chavali, R. Nilo-Poyanco, T. Bernard, D. Kahn, S. Y. Rhee, Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol.* **173**, 2041–2059 (2017).
12. B. Field, A.-S. Fiston-Lavier, A. Kemen, K. Geisler, H. Quesneville, A. E. Osbourn, Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16116–16121 (2011).
13. Y. Li, C. D. Smolke, Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat. Commun.* **7**, 12137 (2016).
14. E. C. Tatsis, I. Carqueijeiro, T. Dugé De Bernonville, J. Franke, T.-T. T. Dang, A. Oudin, A. Lanoue, F. Lafontaine, A. K. Stavrinides, M. Clastre, V. Courdavault, S. E. O’Connor, A three enzyme system to generate the *Strychnos* alkaloid scaffold from a central biosynthetic intermediate. *Nat. Commun.* **8**, 316 (2017).
15. Y. Shang, Y. Ma, Y. Zhou, H. Zhang, L. Duan, H. Chen, J. Zeng, Q. Zhou, S. Wang, W. Gu, M. Liu, J. Ren, X. Gu, S. Zhang, Y. Wang, K. Yasukawa, H. J. Bouwmeester, X. Qi, Z. Zhang, W. J. Lucas, S. Huang, Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* **346**, 1084–1088 (2014).
16. B. J. Kotopka, C. D. Smolke, Production of the cyanogenic glycoside dhurrin in yeast. *Metab. Eng. Commun.* **9**, e00092 (2019).
17. J.-L. Ferrer, J. M. Jez, M. E. Bowman, R. A. Dixon, J. P. Noel, Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **6**, 775–784 (1999).
18. K. Aoki, K. Yano, A. Suzuki, S. Kawamura, N. Sakurai, K. Suda, A. Kurabayashi, T. Suzuki, T. Tsugane, M. Watanabe, K. Ooga, M. Torii, T. Narita, T. Shin-i, Y. Kohara, N. Yamamoto, H. Takahashi, Y. Watanabe, M. Egusa, M. Kodama, Y. Ichinose, M. Kikuchi, S. Fukushima, A. Okabe, T. Arie, Y. Sato, K. Yazawa, S. Satoh, T. Omura, H. Ezura, D. Shibata, Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics* **11**, 210 (2010).
19. A. Rodriguez, T. Strucko, S. G. Stahlhut, M. Kristensen, D. K. Svenssen, J. Forster, J. Nielsen, I. Borodina, Metabolic engineering of yeast for fermentative production of flavonoids. *Bioresour. Technol.* **245**, 1645–1654 (2017).
20. T. G. Köllner, C. Lenk, N. Zhao, I. Seidl-Adams, J. Gershenzon, F. Chen, J. Degenhardt, Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using S-adenosyl-L-methionine. *Plant Physiol.* **153**, 1795–1807 (2010).
21. B. J. Lehka, M. Eichenberger, W. E. Björn-Yoshimoto, K. G. Vanegas, N. Buijs, N. B. Jensen, J. D. Dyekjær, H. Jenssen, E. Simon, M. Naesby, Improving heterologous production of phenylpropanoids in *Saccharomyces cerevisiae* by tackling an unwanted side reaction of Tsc13, an endogenous double-bond reductase. *FEMS Yeast Res.* **17**, fox004 (2017).
22. S. D. Kohlwein, S. Eder, C.-S. Oh, C. E. Martin, K. Gable, D. Bacikova, T. Dunn, Tsc13p is required for fatty acid elongation and localizes to a novel structure at the nuclear-vacuolar interface in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **21**, 109–125 (2001).
23. I. Williams, J. Richardson, A. Starkey, I. Stansfield, Genome-wide prediction of stop codon readthrough during translation in the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **32**, 6605–6616 (2004).
24. M. V. Geraschenko, A. V. Lobanov, V. N. Gladyshev, Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17394–17399 (2012).
25. T. G. Scharff, M. Badr, R. J. Doyle, C. Zaleski, Antagonism between salicylic acid and glucose in *Saccharomyces cerevisiae*. *Toxicol. Appl. Pharmacol.* **60**, 485–491 (1981).
26. Y. Shimizu, H. Ogata, S. Goto, Type III polyketide synthases: Functional classification and phylogenomics. *Chembiochem* **18**, 50–65 (2017).
27. J. M. Jez, J. P. Noel, Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. *J. Biol. Chem.* **277**, 1361–1369 (2002).
28. I. Abe, H. Morita, Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat. Prod. Rep.* **27**, 809–838 (2010).

29. I. Abe, Y. Sano, Y. Takahashi, H. Noguchi, Site-directed mutagenesis of benzalacetone synthase: The role of Phe²¹⁵ in plant type III polyketide synthases. *J. Biol. Chem.* **278**, 25218–25226 (2003).
30. J. M. Jez, M. E. Bowman, J. P. Noel, Expanding the biosynthetic repertoire of plant type III polyketide synthases by altering starter molecule specificity. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5319–5324 (2002).
31. C. Modelos, Trabajo práctico N° 13. Varianzas en función de variable independiente categórica. *Nat. Protoc.* **10**, 845–858 (2016).
32. O. Trott, A. J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
33. T. Akiyama, M. Shibuya, H.-M. Liu, Y. Ebizuka, p-Coumaroyltriatic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. *thunbergii*. *Eur. J. Biochem.* **263**, 834–839 (1999).
34. H. Hohlfield, W. Schurmann, D. Scheel, D. Strack, Partial purification and characterization of hydroxycinnamoyl-coenzyme A: tyramine hydroxycinnamoyltransferase from cell suspension cultures of *Solanum tuberosum*. *Plant Physiol.* **107**, 545–552 (1995).
35. J. Negrel, F. Javelle, Purification, characterization and partial amino acid sequencing of hydroxycinnamoyl-CoA: tyramine N-(hydroxycinnamoyl)transferase from tobacco cell-suspension cultures. *Eur. J. Biochem.* **247**, 1127–1135 (1997).
36. K. Back, S. M. Jang, B.-C. Lee, A. Schmidt, D. Strack, K.-M. Kim, Cloning and characterization of a hydroxycinnamoyl-CoA: Tyramine N-(hydroxycinnamoyl) transferase induced in response to UV-C and wounding from *Capsicum annuum*. *Plant Cell Physiol.* **42**, 475–481 (2001).
37. Q. Yang, K. Reinhard, E. Schiltz, U. Matern, Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA: Anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. *Plant Mol. Biol.* **35**, 777–789 (1997).
38. J. C. D'Auria, Acyltransferases in plants: A good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331–340 (2006).
39. X. Ma, J. Koepke, S. Panjikar, G. Fritzsche, J. Stöckigt, Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *J. Biol. Chem.* **280**, 13576–13583 (2005).
40. P. J. Facchini, J. Hagel, K. G. Zulak, Hydroxycinnamic acid amide metabolism: Physiology and biochemistry. *Can. J. Bot.* **80**, 577–589 (2002).
41. T. Bontpart, V. Cheynier, A. Ageorges, N. Terrier, BAHD or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. *New Phytol.* **208**, 695–707 (2015).
42. J. Sandy, A. Mushtaq, S. J. Holton, P. Schartau, M. E. M. Noble, E. Sim, Investigation of the catalytic triad of arylamine N-acetyltransferases: Essential residues required for acetyl transfer to arylamines. *Biochem. J.* **390**, 115–123 (2005).
43. A. L. Sikora, B. A. Frankel, J. S. Blanchard, Kinetic and chemical mechanism of arylamine N-acetyltransferase from *Mycobacterium tuberculosis*. *Biochemistry* **47**, 10781–10789 (2008).
44. H. Wang, G. M. Vath, K. J. Gleason, P. E. Hanna, C. R. Wagner, Probing the mechanism of hamster arylamine N-acetyltransferase 2 acetylation by active site modification, site-directed mutagenesis, and pre-steady state and steady state kinetic studies. *Biochemistry* **43**, 8234–8246 (2004).
45. L. Zacarés, M. P. López-Gresa, J. Fayos, J. Primo, J. M. Bellés, V. Conejero, Induction of p-coumaroyldopamine and feruloyldopamine, two novel metabolites, in tomato by the bacterial pathogen *Pseudomonas syringae*. *Mol. Plant Microbe Interact.* **20**, 1439–1448 (2007).
46. E. von Roepenack-Lahaye, M.-A. Newman, S. Schornack, K. E. Hammond-Kosack, T. Lahaye, J. D. G. Jones, M. J. Daniels, J. M. Dow, p-Coumaroylnoradrenaline, a novel plant metabolite implicated in tomato defense against pathogens. *J. Biol. Chem.* **278**, 43373–43383 (2003).
47. S. D. Copley, Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* **7**, 265–272 (2003).
48. N. S. McCarty, A. E. Graham, L. Studená, R. Ledesma-Amaro, Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* **11**, 1281 (2020).
49. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison III, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
50. K.-D. Entian, P. Kötter, 25 yeast genetic strain and plasmid collections. *Methods Microbiol.* **36**, 629–666 (2007).
51. O. W. Ryan, J. M. Skerker, M. J. Maurer, X. Li, J. C. Tsai, S. Poddar, M. E. Lee, W. DeLoache, J. E. Dueber, A. P. Arkin, J. H. D. Cate, Selection of chromosomal DNA libraries using a multiplex CRISPR system. *eLife* **3**, e03703 (2014).
52. C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
53. P. Kuzmič, Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV proteinase. *Anal. Biochem.* **237**, 260–273 (1996).

Acknowledgments: We thank A. Cravens for the providing the Cas9/single-guide RNA plasmids (pCS3410, 3411, and 3414) for yeast genomic editing, J. Payne for performing the geometry optimizations of substrate structures for docking simulations, T. Valentini and J. Payne for training in protein purification and valuable discussions on protocol design for in vitro experiments, J. E. Jeon and X. Guan for assistance with tomato metabolomics analyses, and the Stanford ChEM-H Metabolic Chemistry Analysis Center and C. Fischer for instrument (qToF-MS) access and training. We thank E. Sattely, S. Y. Rhee, and C. Khosla for discussions and advice on experimental design. We thank T. Valentini, P. Srinivasan, and B. Kotopka for feedback in the preparation of this manuscript. **Funding:** This work was supported by the NIH U01GM110699 Genome to Natural Products Initiative and Chan-Zuckerberg Biohub Foundation. **Author contributions:** All authors designed the research, analyzed the data, and wrote the paper. D.K. and S.L. performed the research. S.L. performed untargeted metabolomics analysis and found the new compounds. D.K. and S.L. proposed and characterized the tomato cluster activity in yeast. D.K. performed and analyzed CHS in vivo site-directed mutagenesis studies and in vitro enzyme assays. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 1 June 2020
Accepted 14 September 2020
Published 30 October 2020
10.1126/sciadv.abd1143

Citation: D. Kong, S. Li, C. D. Smolke, Discovery of a previously unknown biosynthetic capacity of naringenin chalcone synthase by heterologous expression of a tomato gene cluster in yeast. *Sci. Adv.* **6**, eabd1143 (2020).