A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity

Toshiyuki Waki¹, Ryo Mameda¹, Takuya Nakano¹, Sayumi Yamada¹, Miho Terashita¹, Keisuke Ito¹, Natsuki Tenma¹, Yanbing Li¹, Naoto Fujino¹, Kaichi Uno¹, Satoshi Yamashita², Yuichi Aoki³, Konstantin Denessiouk⁴, Yosuke Kawai⁵, Satoko Sugawara⁶, Kazuki Saito⁶, Keiko Yonekura-Sakakibara⁶, Yasumasa Morita⁷, Atsushi Hoshino⁸, Seiji Takahashi¹ & Toru Nakayama¹

Land plants produce diverse flavonoids for growth, survival, and reproduction. Chalcone synthase is the first committed enzyme of the flavonoid biosynthetic pathway and catalyzes the production of 2′,4,4′,6′-tetrahydroxychalcone (THC). However, it also produces other polyketides, including p-coumaroyltriacetic acid lactone (CTAL), because of the derailment of the chalcone-producing pathway. This promiscuity of CHS catalysis adversely affects the efficiency of flavonoid biosynthesis, although it is also believed to have led to the evolution of stilbene synthase and p-coumaroyltriacetic acid synthase. In this study, we establish that chalcone isomerase-like proteins (CHILs), which are encoded by genes that are ubiquitous in land plant genomes, bind to CHS to enhance THC production and decrease CTAL formation, thereby rectifying the promiscuous CHS catalysis. This CHIL function has been confirmed in diverse land plant species, and represents a conserved strategy facilitating the efficient influx of substrates from the phenylpropanoid pathway to the flavonoid pathway.
Flavonoids are an important class of specialized metabolites, consisting of more than 6900 different structures, that are produced by land plants (Embryophytes). After plants colonized the terrestrial environment 450 million years ago, flavonoids are believed to have played indispensable roles in plant survival, which include antioxidant and UV-screening functions as well as regulatory roles mediating the oxidative stress-induced activation of signaling cascades. In extant plants, flavonoids are also involved in the mechanisms underlying plant reproduction and defense against biotic and abiotic stresses. Flavonoids are derived from \( p \)-coumaroyl-CoA, which is a general precursor for a vast array of phenylpropanoids with diverse functions, such as cuticle biopolymers (in all land plants), lignins (in vascular plants), lignans, and hydroxycinnamoyl esters. The flavonoid biosynthetic pathway branches off from the general phenylpropanoid pathway, and chalcone synthase (CHS) is the first committed enzyme of this pathway. Chalcone synthase is a plant-specific type III polyketide synthase (PKS) that catalyzes the production of 2',4',4',6'-tetrahydroxychalcone (THC) from one \( p \)-coumaroyl-CoA and three malonyl-CoA molecules. The resulting THC subsequently undergoes a stereospecific isomerization catalyzed by chalcone isomerase (CHI) to produce (2S)-naringenin (a flavanone), which is metabolized to other classes of flavonoids in a plant lineage-specific manner. Additionally, CHS, CHI, and other flavonoid enzymes further down the pathway are reportedly organized to form a metabolon (i.e., a fragile, highly organized super-molecular enzyme complex) using P450 protein(s) as a nucleus for the enzyme association.

In many cases, enzymes involved in plant specialized metabolism coincidentally catalyze reactions other than those for which they evolved. This nature of specialized metabolic enzymes, referred to as "catalytic promiscuity," provides an important basis for the adaptive evolution of plant specialized metabolism. Like many other enzymes involved in plant specialized metabolism, CHS is a promiscuous enzyme, which, at least in vitro, coincidentally catalyzes the formation of other polyketides, including \( p \)-coumaroyltriacetic acid lactone (CTAL), due to the derailment of the chalcone-producing pathway. This promiscuity of CHS catalysis likely decreases the efficiency of flavonoid biosynthesis, but it is also believed to have served as the basis for the evolution of other plant-specific PKSs, including stilbene synthase (STS) and \( p \)-coumaroyltriacetic acid synthase (CTAS).

---

**Fig. 1 Chalcone synthase (CHS)-catalyzed formation of THC and other polyketides.** The CHS-catalyzed process (steps a through f) is presented in a gray background. One of the CHS substrates, \( p \)-coumaroyl-CoA, is derived from L-phenylalanine via the general phenylpropanoid pathway and serves as a precursor of lignin, cuticle biopolymers, and other phenylpropanoids in plants. A diffusible intermediate of the CHS-catalyzed reaction may serve as a substrate of chalcone reductase in deoxyflavonoid biosynthesis. Additionally, THC serves as a precursor of aurones, whereas naringenin serves as a precursor of other flavonoids. CL, 4-coumaric acid:CoA ligase.
In 2014, Morita et al. identified a gene in Japanese morning glory (Ipomoea nil) for which a loss-of-function mutation results in pale-colored flowers because of a decrease in the abundance of anthocyanins (a class of flavonoids responsible for cyan flower colors) and flavonols (a class of flavonoids involved in co-pigmentation)\(^{16}\). This gene, designated as **ENHANCER OF FLAVONOID PRODUCTION (EFP)**, encodes a type IV CHI-like protein (CHIL) that lacks CHI activity\(^{17}\). Moreover, CHIL may have evolved from a fatty acid-binding protein (FABP) when plants colonized land\(^{17}\). Although the CHIL-mediated enhancement of flavonoid production appears to occur in many plant species\(^{16}\), the underlying mechanism remains to be clarified. Recent studies indicated that CHIL interacts with CHI\(^{18}\) and CHS\(^{19}\), and may serve as an activator of these enzymes in *Arabidopsis thaliana*\(^{18}\) and some other plant species\(^{19}\).

In this study, we prove that CHIL is a component of flavonoid metabolons and binds to CHS to serve as a rectifier—rather than an activator—of promiscuous CHS: it enhances CHS-catalyzed THC production and diminishes CTAL formation without enhancing total polyketide production, thereby conserving the synthetic activity of CHS and CHIL. This role of CHIL as a rectifier for CHS is widely conserved among land plants. On the basis of the present findings, we discuss the differentiation of CHS-related enzymes (STS and CTAS) from CHS in regard to their ability to interact with CHIL. We propose that the role of CHS in THC production has been maintained throughout the evolution of land plants by the continued ability of CHS to interact with CHIL. However, some CHS homologs have lost their ability to interact with CHIL, resulting in neofunctionalizations to produce enzymes with differential product specificities, such as CTAS and STS.

**Results**

**CHIL is a component of flavonoid metabolons.** To clarify the role of CHIL in the enhancement of flavonoid production, we examined the possibility that CHIL is a flavonoid metabolon component. First, we used yeast two-hybrid systems to comprehensively analyze the binary protein–protein interactions between CHIL and flavonoid biosynthetic enzymes in plants confirmed to contain flavonoid metabolons, namely snapdragon (*Antirrhinum majus* L. (Lamiaceae))\(^{8,10}\) and soybean (*Glycine max* (Fabales))\(^{8,10}\) (Supplementary Fig. 1). We observed that in both plant species, CHIL interacts with CHS and cytochromes P450 \[i.e., 2-hydroxyisoflavanone synthase (IFS) in soybean and flavone synthase II (FNSII) in snapdragon\] (Fig. 2). These results were confirmed in planta by means of bimolecular fluorescence complementation (BiFC) experiments (Supplementary Fig. 2), suggesting that CHIL is a flavonoid metabolon component in these plants. To further confirm this, co-precipitation experiments involving the snapdragon flavonoid metabolon (Supplementary Fig. 1a)\(^{9}\) were conducted with the His\(_8\)-tagged snapdragon CHIL and CHS (His\(_8\)-AmCHIL and His\(_8\)-AmCHS, respectively) as bait proteins (see Methods for details). The results revealed that 150 and 192 proteins were able to specifically bind to His\(_8\)-AmCHIL and His\(_8\)-AmCHS (Supplementary Fig. 3), and AmCHS (Am04g40840.P01) and AmCHIL (Am07g21400.P01), respectively, were identified in the bound proteins. Specific binding of AmCHS to AmCHIL in the extract of snapdragon petal cells was further corroborated by the results of immunoblot analysis that followed co-precipitation (Supplementary Fig. 4). These results provided additional evidence that CHIL is a flavonoid metabolon component.

**Conservation of CHIL–CHS interactions among land plants.** Genes encoding CHIL and CHS (but not IFS and FNSII) are ubiquitous in the genomes of land plants. Thus, we analyzed the physical interactions between CHIL and CHS in phylogenetically distant land plants (bryophyte to angiosperms) (Table 1) with yeast two-hybrid systems and BiFC experiments. Consequently, CHIL and CHS were observed to interact with each other in all plant systems examined with both assay systems (Fig. 3a, c; see also Supplementary Fig. 5), strongly suggesting that this physical interaction is conserved throughout land plants. In contrast, CHI did not bind to CHS (Fig. 3a, Supplementary Fig. 5), with the exception of the enzymes from snapdragon\(^{17}\) and *A. thaliana*\(^{17}\). Moreover, CHIL did not interact with the STS of grapevine (*Vitis vinifera*) (VvSTS\(^{14}\)) or the CTAS of hydrangea (*Hydrangea macrophylla*) (HmCTAS\(^{15}\)) (Fig. 3a, c, Supplementary Fig. 5), both of which show close phylogenetic relationships with CHS. To examine the stoichiometric characteristics of CHIL binding to CHS, a mixture containing His\(_6\)-tagged *A. majus* CHIL and CHS \[i.e., His\(_6\)-AmCHIL (25 kDa, 10 µM) and His\(_6\)-AmCHS (monomer molecular mass \(45\) kDa, 8 µM)\] was combined with bis(sulfosuccinimidyl) suberate disodium salt (BS3, a bi-functional cross linker; 4 mM) at pH 7.5 for 5 min, followed by SDS-PAGE analysis. Protein band with a molecular mass of 70 kDa was detected only for the BS3-reacted protein mixture (Supplementary Fig. 6). This band was immuno-reactive to anti-CHMS and anti-CHIL antibodies and most consistently designated as an AmCHIL monomer cross-linked to an AmCHS monomer. These results were consistent with the binding of one CHIL monomer per CHS monomer. Apparent \(K_D\) values of the CHIL/CHS complexes of various plant species were calculated based on biolayer interferometry (Supplementary Table 1), and ranged from 1 nM (for morning glory) to 126 nM (for gingko (*Ginkgo biloba*)). Thus, the interactions between CHIL and CHS were slightly stronger than the interactions between the proteins of many other metabolons, with \(K_D\) values of 0.03–4 µM\(^{21,22}\).

We then examined whether the binding of CHIL to CHS takes place in a species-specific manner. The binding of heterologous CHILs cCHILs of morning glory (InCHIL) and bryophytes (PpCHIL and MpCHIL of *Physcomitrella patens* and *Marchantia polymorpha*, respectively) to *Arabidopsis CHS* (AtCHS) was assayed using a yeast two-hybrid system (Fig. 3b). The results showed that InCHIL bound weakly to AtCHS compared with AtCHIL, whereas PpCHIL and MpCHIL were unable to bind to AtCHS. To further confirm the species specificity of CHIL–CHS interactions in planta, AtCHIL, PpCHIL, and MpCHIL were respectively expressed in the CHIL-knockout mutant of *A. thaliana* (chil-3) under the control of the AtCHIL promoter (Fig. 4a). The seed coat of the wild-type *A. thaliana* (Col-0) was brown, whereas that of the chil-3 mutant was paler in color primarily owing to decreased contents in the seed coat of proanthocyanidin (PA) in the mutant (Fig. 4b). The chil-3 line expressing AtCHIL (chil-3/AtCHIL) consistently produced brown-colored seeds, which were indistinguishable from those of the wild type (Col-0) (Fig. 4b), whereas the chil-3 line expressing InCHIL (chil-3/InCHIL) displayed a slightly paler seed-color phenotype. Seed-color phenotypes of the chil-3 lines expressing a bryophyte CHIL (chil-3/PpCHIL-A, chil-3/PpCHIL-B, and chil-3/MpCHIL) were indistinguishable from those of the chil-3 mutant and negative controls (chil-3/35SGUS, chil-3/AtCHI, and chil-3/MpCHIL) (Fig. 4b). Furthermore, soluble and insoluble PAs in the seeds of the transgenic chil-3 lines were extracted and determined spectrophotometrically. Consistently, the contents of both types of PAs in the seeds of chil-3/AtCHIL were fully restored to those of the wild type (Fig. 4c, d). In this connection, the chil mutants expressing AtCHIL under the control of the *Cauliflower mosaic virus* 35S promoter showed insufficient recovery of PA contents\(^{18}\), suggesting that the use of the original
**Fig. 2 Interactions between CHIL and flavonoid enzymes in snapdragon and soybean as determined in yeast assay systems.** (*a*) Interactions between CHIL and cytochromes P450 in snapdragon (i–iii) and soybean (iv–vi) were analyzed with the split-ubiquitin yeast two-hybrid systems. Regarding the protein names, the initials of the scientific name of plant species are followed by the abbreviated enzyme/protein names, with homologs indicated with a hyphen. For example, GmCHS-1 refers to isozyme 1 of a soybean (*Glycine max*) CHS. (i–iii) Growth of yeast cells co-expressing X–Cub-LexA-VP16 with N$_{AD^G}$-AmCHIL, where X refers to (i) nothing, (ii) flavone synthase II (AmFNSII), and (iii) flavonoid 3′-hydroxylase (AmF3′H). (iv–vi) Growth of yeast cells co-expressing X–Cub-LexA-VP16 with N$_{AD^G}$-GmCHIL, where X refers to (iv) nothing, (v) isozyme 1 of 2-hydroxyisoflavonoid enzymes in snapdragon and soybean were analyzed with the split-ubiquitin yeast two-hybrid system. The growth of yeast cells co-expressing CHIL and flavonoid enzymes as fusions with the activation domain (AD, AD–Y column) and the DNA-binding domain (BD, BD–X column) of the yeast GAL4 transcription factor is presented. (i) BD-AmCHIL was co-expressed with AD–Y, where Y refers to CHS (AmCHS), CHI (AmCHI), dihydroflavonol 4-reductase (AmDFR), or flavone 3′-hydroxylase (AmF3′H). “None” refers to yeast cells co-expressing AD (without a fused protein) and BD-AmCHIL. (ii) AD-AmCHIL was co-expressed with BD–X, where X refers to AmCHS, AmCHI, AmDFR, or AmF3′H. “None” refers to yeast cells co-expressing BD (without a fused protein) and AD-AmCHIL. (iii) BD-GmCHIL was co-expressed with AD–Y, where Y refers to CHS isozymes (GmCHS-1 and GmCHS-7), chalcone reductase (CHR) isozymes (GmCHR-1 and GmCHR-5), CHI isozymes (GmCHI-1A, GmCHI-1B, and GmCHI-2), GmCHIL, or a 2-hydroxyisoflavonoid dehydratase isozyme (GmHID-1). “None” refers to yeast cells co-expressing AD (without a fused protein) and BD-GmCHIL. (iv) BD-GmCHIL was co-expressed with BD–X, where X refers to GmCHS-1, GmCHS-7, GmCHR-1, GmCHR-5, GmCHI-1A, or GmCHI-2. “None” refers to yeast cells co-expressing BD (without a fused protein) and AD-GmCHIL. Abbreviated growth media names are as follows: −WL, SD agar medium lacking tryptophan and leucine; −WLH, −WL medium lacking histidine; −WLHAd, −WLH medium lacking adenine.

The content of insoluble PA in the seeds of *chil-3/InCHIL* was higher than those of the *chil* mutants (*chil-1* and *chil-3*) and negative controls, although it was lower than that of the wild type (Fig. 4d). The content of soluble PA in the seeds of *chil-3/InCHIL* was only slightly higher than those of the *chil* mutants (Fig. 4c). No significant change in the contents of soluble and insoluble PAs was observed in *chil-3/PpCHIL-A, chil-3/PpCHIL-B,* and *chil-3/MpCHIL.* These results suggested that the binding of CHIL to CHS takes place in a species-specific manner and that heterologous CHIL could not fully implement its in planta function in phylogenetically distant species.

**Promiscuous CHS specificity and effect of CHIL binding.** When AmCHS was assayed with p-coumaroyl-CoA and malonyl-CoA as substrates under standard assay conditions (see Methods),
Table 1 Plant sources of CHILs and flavonoid enzymes used in this study.

| Taxonomic group | Scientific name | Common name |
|-----------------|-----------------|-------------|
| Bryophyte       | Physcomitrella patens | Spreading earth moss |
| Lycophyte       | Selaginella moellendorffii | Selaginella |
| Gymnosperm      | Ginkgo biloba | Ginkgo |
| Angiosperm      |                |             |
| Monocot Poales  |                |             |
| Dicot Poales    |                |             |
| Vitis           |                |             |
| Fabales         |                |             |
| Brassicales     |                |             |
| Comales         |                |             |
| Solanaceae      |                |             |
| Lamiales        |                |             |
| Ginkgo          |                |             |
| Physcomitrella  |                |             |
| Selaginella     |                |             |
| Spreading earth moss |            |             |
| Selaginella     |                |             |
| Ginkgo biloba   |                |             |
| Torenia majus   |                |             |
| Torenia hybrid  |                |             |

Effects of CHIL on kinetics and in vivo specificity of CHS. We analyzed the effects of CHIL on the kinetic parameters of the CHS-catalyzed formation of THC in the presence and absence of excess CHIL in representative plant systems. The $k_{cat}$ value of THC production was 2- to 15-times higher in the presence of CHIL than in its absence for all examined plant species (Tables 2 and 3). The $K_m$ and $K^*$ values (see Methods) for p-coumaryl-CoA and malonyl-CoA, respectively, were also higher in the presence of CHIL than in its absence (Tables 2 and 3).

We subsequently examined the effects of CHIL on the product specificity of in vivo chalcone synthesis in Escherichia coli. First, E. coli cells were genetically engineered to produce soybean 4-coumaric acid:CoA ligase (Gm4CL-3; corresponding to Cl in Fig. 1) and an isozyme of soybean CHS (GmCHS-1). When the recombinant cells (strain CHIL−) were incubated with 1 mM p-coumaric acid at 30 °C for 3 h, the cells produced 327 nM CLC, with an $E$ value of +66.5% (Fig. 6), which was significantly higher than the $E$ value (+7%) obtained via in vitro enzymatic production (Fig. 5e). To clarify the effects of GmCHIL on the product specificity of in vivo chalcone synthesis, GmCHIL was co-expressed with Gm4CL-3 and GmCHS-1 in strain CHIL− (see above). The results of immunoblot analyses revealed that the molar ratio (GmCHIL: GmCHS-1, on monomer basis) of the expressed proteins in the resulting cells (strain CHIL+) was 1.5:1.0 (Supplementary Fig. 8). The CHIL+ cells were incubated with 1 mM p-coumaric acid at 30 °C for 3 h. The HPLC results revealed that the $E$ value obtained with this system was +91%, indicating that the product specificity of the in vivo CHS-catalyzed reaction increased significantly (Fig. 6). The total amount of polyketides (CLC + CTAL) produced in strain CHIL+ was essentially unchanged from that obtained in strain CHIL−.

Discussion

Enzymes involved in specialized metabolism may exhibit promiscuous catalytic properties, which are believed to serve as the basis for the functional evolution (neofunctionalization) of enzymes. Despite the pivotal role of CHS as the first committed enzyme in the flavonoid pathway, CHS is a promiscuous enzyme irrespective of its plant origins. In vitro assay results indicated that CHS coincidently produces a large amount of CTAL as a byproduct in addition to THC, with $E$ values ranging from −67 to +33% depending on the plant species (Fig. 5e). The present study establishes that CHIL binds to CHS with a $K_0$ value between 10−9 and 10−7 M, thereby increasing the $E$ value of CHS catalysis to facilitate efficient THC production (Figs. 5 and 6). A previous study demonstrated that CHIL activates the CHS-mediated production of THC, but we showed that it only weakly induces the PKS-catalyzed production of polyketides (THC, CTAL, and BNY). Thus, in essence, CHIL is considered to be a rectifier—rather than an activator—of promiscuous CHS catalysis. Of note, during the review of this paper, a research group independently reported the ability of a fern CHIL to alter product specificity of CHS, consistent with our results. The CHIL genes are ubiquitous in the genomes of land plants and expressed along with CHS genes in two prominent product peaks (P1 and P2) and one very minor peak (P3) were observed during the LC-MS analysis (Supplementary Fig. 7a, upper panel). A subsequent tandem MS/MS analysis identified P1 and P2 as CTAL and naringenin (an isomer of THC) that forms via a non-enzymatic isomerization, respectively (Supplementary Fig. 7b). The P3 peak most likely represented bis-norflavonoylglycerol (BNY), according to its molecular mass and absorption spectra. Our assay results revealed that the product ratios of the CHS-catalyzed reactions varied with the enzyme source (Fig. 5a–e), with the formation of BNY being negligible in all cases (<1% (mol/mol) of all products). Thus, CHS is generally a promiscuous enzyme in terms of product specificity, irrespective of plant sources.

We then examined the effects of AmCHIL (1.0 μM) on the product specificity of AmCHS (0.1 μM) under standard assay conditions. In the presence of the added AmCHIL, the amount of THC and naringenin [collectively termed chalconoids (CLC)] formed in the reaction mixture increased 1.7-fold, with a substantial decrease in CTAL formation (Fig. 5a). The BNY-producing activity of AmCHS, which was negligible, was unaffected by the presence of CHIL. The total amount of CHS products (CLC, CTAL, and BNY) in the presence of CHIL was 116% of the amount in its absence, indicating that AmCHIL only slightly enhanced the total polyketide-producing activity of AmCHS. This was also the case for the CHS−CHIL systems of all other examined plant species (Fig. 5e). Thus, in essence, the presence of CHIL increases the $E$ value of the CHS-catalyzed reaction to enhance the product specificity for THC production. When a constant amount of CHS (0.1 μM) was assayed in the presence of varying amounts of CHIL, a 5- to 10-fold molar excess (on a monomer basis) of CHIL over CHS was observed to substantially alter the product specificity of CHS (Fig. 5f, g for snapdragon and P. patens systems). Moreover, the CHILs did not affect the $E$ values of CHS-catalyzed reactions (Fig. 5a, c). Furthermore, CHIL did not influence the product specificities of STS and CTAS (Fig. 5c, d).
dive,
land plants, CHIL likely represents an auxiliary subunit of CHS that is essential for CHS to efficiently fulfill its role (chalcone synthesis) as the first committed enzyme of the flavonoid pathway. This, in turn, provides the possibility that among the CHS-related homologs, CHSs may be defined as members that interact with CHIL (see below).

The results of kinetic studies provided an important insight into the mechanistic aspects of the CHIL-mediated increase in the $E$ value of CHS-catalyzed reactions in flavonoid metabolons. The $k_{\text{cat}}$ values of THC production in the presence of CHIL were 2- to 15-fold greater than those in its absence in all examined plant species (Tables 2 and 3). The $K_m$ values for both substrates for THC production were also greater in the presence of CHIL than in its absence in all examined plant species (Tables 2 and 3). Thus, the $k_{\text{cat}}/K_m$ values of CHS enzymes were similar irrespective of the presence of CHIL. However, even though two enzyme systems have similar $k_{\text{cat}}/K_m$ values, the ratio of the rates of these two systems may vary with the ratio of substrate concentration to $K_m$.

The $k_{\text{cat}}/K_m$ value only represents an estimation of the catalytic effectiveness of enzymes when substrate concentrations are near zero, which is unrealistic when the proposed involvement of CHS and CHIL in flavonoid metabolons is considered. In metabolons, a group of enzymes and metabolites in the metabolic pathway are believed to be concentrated in a small cellular region (i.e., micro-compartmentalization of cellular metabolism). Under such circumstances, the “effective” concentrations of $p$-coumaroyl-CoA and malonyl-CoA for CHS catalysis may be high enough relative to the $K_m$ values, and $k_{\text{cat}}$ may reflect the catalytic competency of the enzymes in metabolons relatively accurately.

Therefore, the observed increase in the $E$ value of CHS-catalyzed reactions due to CHIL may be, at least partly, explained in terms of the CHIL-mediated enhancement of the $k_{\text{cat}}$ of THC synthesis because the $k_{\text{cat}}$ of THC synthesis may be related to the rate of step $d$ (Fig. 1). CHIL increases the $E$ value of CHS catalysis via the
Fig. 5 Effects of CHIL and related proteins on the product specificities of CHS, CTAS, and STS. a Concentrations of products (blue bars, naringenin; orange bars, CTAL) formed after reactions catalyzed by 0.1 μM AmCHS in the presence and absence of 1 μM AmCHIL or AmCHI. b Concentrations of products (blue bars, naringenin; orange bars, CTAL) formed after reactions catalyzed by 0.1 μM PpCHS in the presence and absence of 1 μM PpCHIL. c Left, Concentrations of products formed after reactions catalyzed by 0.1 μM HmCHS in the presence and absence of 1 μM HmCHIL or HmCHI. Right, Concentrations of products formed after reactions catalyzed by 0.1 μM HmCTAS in the presence and absence of 1 μM HmCHIL or HmCHI. Blue bars, naringenin; orange bars, CTAL. d Left, Concentrations of products formed after reactions catalyzed by 0.1 μM VvCHS in the presence and absence of 1 μM VvCHIL. Right, Concentrations of products formed after reactions catalyzed by VvSTS in the presence and absence of VvCHIL. Blue bars, naringenin; orange bars, CTAL; and pink bars, resveratrol. e E values of reactions catalyzed by CHS (0.1 μM) in the presence (light-orange bars) and absence (gray bars) of homologous CHIL (1 μM) in various land plants. For (a–e), the numbers above panels indicate the relative percentages of the total amount of polyketide products, with the amount obtained in the absence of added CHIL or CHI set at 100%. f Effects of AmCHIL concentrations on the E values of the reaction catalyzed by 0.1 μM AmCHS. g Effects of PpCHIL concentrations on the E values of the reaction catalyzed by 0.1 μM PpCHS. For (a–g), enzymatic reactions were completed essentially under standard assay conditions (see Methods). Data are presented as the average of three independent determinations (± standard deviation). The underlying data for (a–g) are provided in the Source Data file.
specific rate enhancement of step d, resulting in the increased channeling of 3 toward THC production at the expense of CTAL formation (step e) (Fig. 1).

We revealed that CHIL is a non-catalytic component of flavonoid metabolons (Supplementary Fig. 1), suggesting that the importance of metabolon formation in plant specialized metabolism includes its role to rectify promiscuity of metabolic enzymes via macromolecular interactions to decrease undesirable minor activities and/or specificities. This potential consequence of macromolecular interactions in metabolons is based on the finding that the promiscuous product specificity of *Rhododendron dauricum* orcinol synthase, which is a type III PKS involved in orsellinic acid biosynthesis, may be rectified by CHIL (Fig. 1). Thus, the rectification of catalytic promiscuity via macromolecular interactions might not be limited to the CHIL−CHS system and may be generally important for metabolons involved in plant specialized metabolism. This strategy may enable the maintenance of the native promiscuity of enzyme catalysis for the evolution of functions, and facilitate the efficient synthesis of specialized metabolites based on the rectified specificity in extant plants.

The suppression of CHIL reportedly results in a 3- to 9-fold decrease in flavonoid content in petunia and torenia petals\(^1\). The in vitro effects of CHIL on the product specificity of CHS.

**Fig. 6 Effects of CHIL on in vivo specificity of CHS.** E values of CHS-catalyzed reactions during the production of CLC from p-coumaric acid in *E. coli* cells harboring pET15b-Gm4CL-3-GmCHS-1 (CHIL−) or pET15b-Gm4CL-3-GmCHS-1-GmCHIL (CHIL+) are shown. Data are presented as the average of three independent determinations of three biological replicates (±standard error). *P < 0.01 (Student’s t-test). The underlying data are provided in the Source Data file.

### Table 2 Kinetic parameters for p-coumaroyl-CoA of THC production catalyzed by CHS of some land plants in the presence (+) and absence (−) of CHIL.

| Enzyme     | CHIL | Kinetics | k_cat (min\(^{-1}\)) | K_m (µM) | k_cat/K_m (min\(^{-1}\)M\(^{-1}\)) |
|------------|------|----------|----------------------|---------|-----------------------------------|
| PpCHS      | +    | c        | [28.5 (±0.6)]\(^c\)  | not determined\(^c\) | 0.057\(^c\)                      |
|            | −    | H        | 1.69 (±0.17)         | 19.0 (±4.3) | 0.089                             |
| SmCHS      | +    | H        | 7.54 (±0.14)         | 2.58 (±0.16) | 2.92                              |
|            | −    | H        | 3.94 (±0.03)         | 3.15 (±0.77) | 1.25                              |
| GbCHS      | +    | H        | 44.5 (±6.2)          | 31.1 (±8.3)  | 1.43                              |
|            | −    | H        | 6.79 (±0.26)         | 4.29 (±0.73)  | 1.58                              |
| OsCHS-1    | +    | H        | 19.4 (±3.3)          | 76.3 (±19.0) | 0.25                              |
|            | −    | H        | 1.58 (±0.06)         | 7.70 (±0.93)  | 0.21                              |
| AmCHS      | +    | H        | 76.3 (±6.0)          | 49.7 (±6.4)   | 1.54                              |
|            | −    | H        | 5.23 (±0.17)         | 4.73 (±0.58)  | 1.11                              |

\(^{a}\)000 µM malonyl-CoA was used as the extender substrate.

\(^{b}\)H, hyperbolic. Hyperbolic vs [S] plots were obtained and fit to the Michaelis-Menten equation. For details, see Methods.

\(^{c}\)A linear relationship between initial velocity and substrate concentrations [S] was obtained in the range of [S] examined (up to 50 µM), suggesting that the K_m value of PpCHS for this substrate should be significantly greater than 50 µM because the enzyme-catalyzed reactions proceed with first-order kinetics under the conditions of [S] ≪ K_m. Therefore, only the K_m value was determined from slope of the initial velocity vs [S] plots. The k_cat value was calculated using the initial velocity at [S] = 50 µM.

### Table 3 Kinetic parameters for malonyl-CoA of THC production catalyzed by CHS of some land plants in the presence (+) and absence (−) of CHIL.

| Enzyme     | CHIL | Kinetics | k_cat (min\(^{-1}\)) | K_m (µM) | k_cat/K_m (min\(^{-1}\)M\(^{-1}\)) | Hill coeff. |
|------------|------|----------|----------------------|---------|-----------------------------------|-------------|
| PpCHS      | +    | s        | 31.5 (±0.6)          | 9.14 (±0.36) | 3.45                             | 2.96 (±0.32) |
|            | −    | s        | 2.60 (±0.12)         | 3.23 (±0.22) | 0.80                             | 3.25 (±0.85) |
| SmCHS      | +    | s        | 23.5 (±0.4)          | 3.86 (±0.14) | 6.09                             | 2.11 (±0.14) |
|            | −    | s        | 5.62 (±0.21)         | 1.15 (±0.09) | 4.90                             | 2.71 (±0.52) |
| GbCHS      | +    | s        | 28.9 (±1.7)          | 8.97 (±0.80) | 3.22                             | 2.46 (±0.45) |
|            | −    | s        | 8.01 (±0.36)         | 3.60 (±0.22) | 2.22                             | 3.03 (±0.59) |
| OsCHS-1    | +    | s        | 7.03 (±0.36)         | 7.60 (±0.58) | 0.93                             | 2.31 (±0.34) |
|            | −    | s        | 2.41 (±0.11)         | 3.35 (±0.16) | 0.72                             | 3.00 (±0.51) |
| AmCHS      | +    | s        | 35.4 (±1.4)          | 12.2 (±1.6)  | 2.90                             | 1.73 (±0.31) |
|            | −    | s        | 8.47 (±0.14)         | 2.57 (±0.10) | 3.30                             | 3.77 (±0.73) |

\(^{a}\)50 µM p-coumaroyl-CoA was used as the starter substrate.

\(^{b}\)S, sigmoidal. Sigmoidal vs [S] plots were obtained and fit to the Hill’s equation. For details, see Methods.
clarified in this study may, at least partly, account for such an in vivo effect of CHIL. However, CHIL appears to have another important role in enhancing flavonoid production in a plant species-specific manner. Specifically, the results of the present protein–protein interaction analyses proved that in snapdragon and soybean, CHIL physically interacts with CHS as well as cytochromes P450 (i.e., FNSII in snapdragon and IFS in soybean; see Supplementary Fig. 1), which also bind to CHI in these plant species19-21. Additionally, in A. thaliana, which lacks FNSII and IFS in its genome, CHIL was observed to interact with CHI18. It is generally accepted that cytochromes P450 serve as the nuclei of metabolism formation in plant specialized metabolism28. Thus, these observations imply there is another important CHIL function related to metabolism formation. Specifically, it recruits enzymes involved in the early flavonoid biosynthesis stage (CHS and CHI) to a flavonoid biosynthesis site in plant cells to establish an efficient supply of naringenin, which is a common precursor for the biosynthesis of flavonoids further down the pathway (Fig. 1). Therefore, both of these roles (i.e., rectification of catalytic promiscuity and recruiting enzymes to a flavonoid biosynthesis site) might be important for CHIL to fully enhance production in planta, with the latter occurring in a species-specific manner.

An evolutionary connection between fatty acid metabolism and flavonoid biosynthesis has been proposed based on the structural and mechanistic similarities between the enzymes involved in the biosynthesis of these compounds17,29. Consistent with this, CHI and CHILs are phylogenetically related to FABP. More specifically, CHILs are members of the CHI-fold protein family, which includes the following four phylogenetically separate groups of proteins: CHI enzymes (types I and II), FABPs (type III), and CHILs (type IV). In planta, FABP localizes to plastids, where it is involved in de novo fatty acid biosynthesis. A phylogenetic analysis suggests that FABPs may be the oldest members of the CHI-fold protein family and CHIL and CHI may have evolved from FABP when plants colonized the terrestrial environment17,29. In this context, it would be noteworthy that the diffusible linear tetraketide-CoA intermediate [3 (R = CoA), see Fig. 1] formed during CHS catalysis is structurally and biosynthetically related to fatty acids, and computational docking studies predicted that CHIL potentially binds to the intermediate 3 in an energetically favorable manner (Supplementary Fig. 9). This is consistent with the known abilities of other types of proteins in the family to bind to fatty acids (for FABPs) and cyclic tetraketides (for CHILs)17,29.

This prediction based on an evolutionary consideration should be addressed in conjunction with the mechanistic aspects of the role of CHIL to rectify CHS promiscuity in future studies.

Phylogenetic analysis suggested that CHS also evolved when plants colonized the terrestrial environment, similar to CHIL. Unlike the CHI-fold protein family, however, CHS constitutes a multigene family that consists of species-specific copies of CHS-related genes (Supplementary Fig. 10). Notably, four clades were resolved that each consist of multiple CHS genes from a single species (M. polymorpha, S. falax, P. patens, and soybean) (Supplementary Fig. 10a, c). This finding suggests that recent gene duplications have occurred within these species or a recent ancestor. It is also likely that these CHS-related genes evolved via a birth-and-death process, in which new copies of CHS genes are produced by repeated gene duplication and some duplicate genes are maintained in the genome while others are deleted, neo-functionalized by mutations, or rendered nonfunctional by deleterious mutations. We included the STS and CTAS sequences in the phylogenetic analysis to infer their evolutionary origin. The analysis indicated that STS (PsSTS and VvSTS) and CTAS (HmCTAS) were differentiated from CHS genes of seed plants (gymnosperms and angiosperms) in a lineage-specific manner (Supplementary Fig. 10b, d), thus functional differentiation of STS and CTAS from CHS arose during seed plant evolution. Our functional analysis of CHIL predicted that ancestral CHS had the ability to interact with CHIL to facilitate its biochemical role of THC synthesis. STS and CTAS did not interact with CHIL. Thus, during the proposed evolutionary process of CHS-related genes, some CHS homologs may have lost their ability to interact with CHIL, allowing neofunctionalization to produce enzymes that are capable of exclusively producing p-coumaroylitraceic acid (e.g., HmCTAS) and stilbenes (e.g., VvSTS). This is consistent with the notion that the ability of CHS to produce THC was maintained with continued ability to interact with CHIL throughout the course of land plant evolution.

In conclusion, CHIL is a rectifier of promiscuous CHS that allows CHS to precisely fulfill its physiological function—the synthesis of THC, which is the key precursor of a diverse array of flavonoids indispensable for the adaptation of land plants to the terrestrial environment. Given that this role of CHIL has been maintained during the evolution of land plants, it must have been of vital importance to survival of land plants in the terrestrial environment.

**Methods**

**Mass spectrometry, E. coli strains, and A. thaliana lines.** The plasmids used in this study were obtained from the following sources: pGEM-T Easy vector (Promega, Madison, WI, USA); pGBKT7 and pGADT7 (each used for the GAL4-based yeast two-hybrid assays) and pET15b and pCold I (each used for the heterologous production of His6-tagged proteins in E. coli cells) (Takara Bio, Otsu, Japan); pCOLA-Duet-1 (used for the simultaneous heterologous production of His6-tagged and S-tagged proteins with cholate cells) (Merck, Darmstadt, Germany); pGEX-KT-1 (used for the heterologous production of glutathione S-transferase-tagged proteins in E. coli cells) (GE Healthcare, Tokyo, Japan, Tokyo, Japan); pDOE-0530, ER:rh1 (each used for Biotin experiments), and pGWBI (The Arabidopsis Biological Resource Center, Columbus, OH, USA). Escherichia coli strains BL21 and BL21 (DE3) were obtained from Promega and strain Rosetta2 (DE3) was obtained from Merck KGaA. Arabidopsis thaliana accession Columbia-0 (Col-0; Lele Seeds, http://www.arabidopsis.com) was used as the wild type in this study. The T-DNA insertion mutants of Arabidopsis, lines SALK_096531 (chi-11)18 and GABI_189B03 (CS745226, chi-3) for CHIL were obtained from the Arabidopsis Biological Resource Center. T-DNA insertion lines were screened by PCR using specific primers (CHIL-F and CHIL-R for CHIL and TF and TR for T-DNA; see Supplementary Table 2). PCR products were sequenced to determine the exact insertion points.

**Chemicals.** The THC and p-coumaroyl-CoA were purchased from TransMIT (Gießen, Germany). Malonyl-CoA, naringenin, trans-p-coumaric acid, resveratrol, ampicillin, imidazole, and isopropyl 1-P-D-thiogalactoside were purchased from Nacalai Tesque (Kyoto, Japan). Acetonitrile and trifluoroacetic acid were obtained from Sigma-Aldrich Japan (Tokyo, Japan). Bis(sulfosuccinimidyl) disodium salt was obtained from Dojindo Laboratories (Kumamoto, Japan), whereas Ni2+-coated magnetic beads (His Mag Seraphose Ni) and Ni2+-coated Seraphose beads (His SpinTrap) were purchased from GE Healthcare Japan. All chemicals were of analytical grade except for acetonitrile and trifluoroacetic acid, which were of HPLC grade. [2-14C]Malonyl-CoA (140 MBq/mmol) was purchased from PerkinElmer Life Sciences (Boston, MA, USA).

**Synthesis of cDNAs encoding flavonoid enzymes and proteins.** The cDNAs encoding flavonoid enzymes and CHILs analyzed in this study are summarized in Supplementary Table 4. The cDNAs encoding flavonoid enzymes of snapdragon, torenia, and soybean [i.e., CHS, chalcone reductase (CHR, soybean), CHI, 2-hydroxyisoflavone dehydratase (HID, soybean), FNSII (snapdragon and torenia), IFS (soybean), flavanone 3-hydroxylase (F3H, snapdragon), and dihydroflavonol 4-reductase (DFR, snapdragon)] were obtained as previously described10. The CHIL cDNAs of snapdragon22, torenia22, soybean23,24, A. thaliana18, rice (Oryza sativa, OsCHIL-A and OsCHIL-B)25, and morning glory26 were obtained as described in the listed references.

The CHIL genes of grapevine (VvCHIL), Selaginella moellendorfii (SmCHIL), and P. patens (PpCHIL-A and PpCHIL-B) were identified via a BLAST search of the Phytosyme (version 12.1) database (https://phytosome.jgi.doe.gov/pz/portal). The CHIL sequences were used as a query sequence (Supplementary Table 4). The CHIL and CHI genes of hydrangea (HmCHIL and HmCHI) were identified via a BLAST search of the Hardwood Genomics Project database
Yeast two-hybrid assays. The Matchmaker Gold Yeast Two-Hybrid System (Clontec, Mountain View, CA, USA) was used to complete GAL4-based yeast two-hybrid assays as previously described. Briefly, the open reading frames encoding proteins to be tested for interaction were amplified by PCR to add SfiI restriction sites for ligation with the pGBK7 and PDAD7 vectors (Clontec). Each amplified DNA was purified, digested with SfiI, and inserted into the SfiI site of the pGβKT7 or pGAD7 vector. The S. cerevisiae strain Y2H Gold (Clontec) was transformed with each pair of the plasmids in accordance with the manufacturer’s guidelines.

RT-PCR. Total RNA was extracted using the RNeasy Total RNA Isolation Kit (ThermoFisher Scientific) in accordance with the manufacturer’s instructions. After DNase I treatment using the TURBO DNA-Free Kit (ThermoFisher Scientific), 500 ng total RNA was used for cDNA synthesis using the SuperScript ™ III First-Strand Synthesis System (Invitrogen). RT-PCR was performed using Takara ExScript reverse transcriptase (Takara, Japan) with a 2 μl cDNA sample by 35 cycles of amplification (98 °C for 10 s, 55 °C for 30 s, and 72 °C for 1 min), followed by a final extension step (72 °C for 5 min). The primers AtCHI RT_F and AtCHI-1 RT_R for AtCHI, AtGS5270-359F and AtCHI1 RT_R for AtCHI, Ipom. EFP_RT_F and Ipom. EFP_RT_RT for Ipom. EF1A53 RT_F and EF1A65 RT_F, and Ipom. EF1A53 RT_F and EF1A65 RT_R for Ipom. EF1A53 RT, Ipom. EFP_RT_F and poly. EFP_RT_RT for Ipom. EF1A53 RT, and poly. CHI_RT_F and poly. CHI_RT_RT for Ipom. EF1A53 RT were used for analyses (Supplementary Table 2).

PA analysis. Extraction and acid hydrolysis of PA were performed in triplicate as described previously. Mature seeds (10 mg) were homogenized in 0.75 ml of 70% (v/v) acetonitrile containing 5.26 mM Na3S2O3 in a mixer mill (Qagen Retsch MM400 Tissuelyser; Qagen, Tokyo, Japan) at 20 Hz for 1 min, followed by sonication for 20 min. After centrifugation at 15,000 × g for 5 min, the supernatant was evaporated and resuspended in 1 ml of a 2:1.03 (v/v) mixture of HCl/butanol 70% acetic. The absorbance of the solutions before and after hydrolysis at 95 °C for 60 min was measured at 545 nm and the difference was treated as the soluble PA fraction. The pellet after extraction with 70% acetic was also evaporated, suspended in the mixture of HCl/butanol70% acetic, and hydrolyzed, and was treated as the insoluble PA fraction.

Tandem mass spectrometry analyses. The amino acid sequences of the peptides obtained during the co-precipitation experiments (Supplementary Fig. 3) were determined with an EASY-nLC1000 system equipped with the Q Exactive Plus apparatus (ThermoFisher Scientific). The protein data were extracted and searched against the SwissProt protein database (v12.1) using the Mascot search engine. The results of the Mascot search were extracted as .mgf files and reviewed by the Targeted Reaction (Supplementary Fig. 7) were identified with the Q Exactive Plus apparatus.

Expression and purification of CHS, CHIL, and other enzymes. The complex coding sequences for CHS- and CHIL-related genes (i.e., CHS, CTAS, STS, CHI, and CHIL) were amplified by PCR, with the pGEM-T Easy- (Takara, Japan), or pBluescript-based constructs as templates. Restriction enzyme sites were added during the PCR amplification for the subsequent subcloning into the pCold I vector. The resulting fragments were subcloned into the pCold I vector to express a fusion protein with an N-terminal His tag, Escherichia coli BL21 cells were transformed with the resulting plasmids. The heterologous expressions of these cDNAs and the purification of the recombinant proteins were completed as previously described. The purified proteins were analyzed by SDS-PAGE and visualized with Coomassie Brilliant Blue R250. The concentrations of the recombinant proteins were determined based on the absorption coefficients at 280 nm, which were calculated based on the amino acid sequences.

Enzyme assays. The standard assay mixture consisted of 100 mM HEPES-NaOH buffer, pH 7.5, 30 μM p-coumaryl-CoA, 100 μM malonyl-CoA, CHS (typically 0.1 μM), and CHIL (typically 1.0 μM) in a final volume of 50 μl. The mixture without CHS was pre-incubated at 30 °C for 5 min, and the reaction was started by adding CHS. After a 60-min incubation at 30 °C, the reaction was stopped by adding a 50-μl aliquot of a mixture comprising acetonitrile and water (2,3, v/v) and 4% (v/v) trifluoroacetic acid. Flavonoids and related products formed in the reaction mixture were analyzed by reversed-phase HPLC with a J'sphere ODS M80 column (4.6 × 150 mm; YMC, Kyoto, Japan) under previously described conditions. Chromatograms were obtained by monitoring the absorbance at 290 nm. In some cases, flavonoids and related products were also analyzed by LC-MS with the following Shimadzu LCsolution system: column, CAPCELL CORE C18 (2.1 × 100 mm; Shisuido, Tokyo, Japan); flow rate, 0.2 ml/min; solvent A, 0.05% (by volume)
formic acid in H2O: solvent B, 0.05% (by volume) formic acid in acetonitrile. After an injection (5 μl) onto the column that was equilibrated with 20% B (by volume), the column was initially developed isocratically with 20% B for 2 min, followed by a linear gradient from 20 to 35% B in 18 min. The column was then washed isocratically with 35% B for 5 min, followed by a linear gradient from 35 to 90% B in 1 min. The column was washed isocratically with 20% B for 10 min before the next injection to ensure the column was sufficiently re-equilibrated. Chromatograms were recorded by mass spectrometry (SIM, negative ion mode).

To quantify the CTAL formed, the standard assay mixture was supplemented with [2-14C]malonyl-CoA (148 MBq/mmol). The CTAL and naringenin were separated by HPLC as described above and collected, after which the radioactivities of the CTAL and naringenin fractions were determined with the LSE500 Multi-Purpose Scintillation Counter (Beckman Coulter, Brea, CA, USA). The radioactivity of the naringenin fraction was correlated with absorbance peaks integrals of known amounts of naringenin, whose specific radioactivity was equal to that of CTAL.

**Enzyme kinetics.** Assays of the initial velocity (v) of the CHS-catalyzed reactions were completed with a standard assay system (see above) under steady-state conditions and with various substrate concentrations. The apparent Km and Vmax values and their standard errors for p-coumaroyl-CoA in the presence of a fixed concentration of malonyl-CoA were determined by fitting the initial velocity data to the Michaelis–Menten equation:

\[
v = \frac{V_{\text{max}}[S]}{K_m + [S]}
\]

with a nonlinear regression method of the SigmaPlot 12 program (Hulinks, Tokyo, Japan), where Vmax, Km, and [S] denote maximum velocity, the Michaelis constant, and the substrate concentration, respectively. When CHS assays were completed with malonyl-CoA as the varying substrate with a fixed concentration of p-coumaroyl-CoA, sigmoidal v vs [S] plots were obtained and fit to Hill’s equation:

\[
v = \frac{V_{\text{max}}[S]^n}{K^n + [S]^n}
\]

with a nonlinear regression method of the SigmaPlot 12 program, where K’ is related to Km and represents the effect of the substrate occupancy at one site on the substrate affinity at other sites, and n corresponds to Hill’s coefficient.

**Biolayer interferometry.** Biolayer interferometry measurements were recorded with a BLItz instrument (FortBio, Fremont, CA, USA). Biosensors were soaked in BLItz assay buffer (100 mM HEPES-NaOH, pH 7.5) for 10 min, and His6-tagged CHS was immobilized on anti-His6 sensors. Biolayer interferometry assays consisted of the following three steps, all performed in BLItz assay buffer: step 1, initial baseline (0–30 s); step 2, loading of glutathione S-transferase-tagged CHIL (30–150 s); and step 3, wash (150–270 s). Control values, which were obtained from empty sensors (i.e., no protein), were subtracted from experimental values before the data were processed. To calculate the equilibrium dissociation constant (Kd) and the association (ka) and dissociation (kd) rate constants, sensograms were fit to a 1:1 binding model with BLItz Pro (version 1.2.12).

**Production of chalcone in E. coli cells.** The GmNcLC-3 cDNA (DDBJ/EMBL/GenBank accession number, AF002258) was amplified by PCR from a cDNA library prepared from the lateral roots of soybean seedlings (cv. Enrei) as described previously and cloned into NdeI/XhoI-digested pET15b according to the BLItz method to obtain pET15b-GmNcLC-3. The GmCHS-1 cDNA was amplified by PCR and cloned into MCS2 of NdeI-digested pCOLDI-1-digested GmCHS-1 DNA. The DNA segment for the T7 promoter-driven expression of GmCHS-1 in the resulting plasmid was then introduced into EAgl-digested pET15b-GmNcLC-3 to obtain pET15b-GmNcLC-3-GmCHS-1. The GmCHIL cDNA was amplified by PCR and cloned into MCS1 of pET15b-GmNcLC-3-GmCHS-1 to obtain pET15b-GmNcLC-3-GmCHS-1-GmCHIL. The DNA segment for the T7 promoter-driven co-expression of GmCHS-1 and GmCHIL in the resulting plasmid was then introduced into EAgl-digested pET15b-GmNcLC-3-GmCHS-1-GmCHIL to obtain pET15b-GmNcLC-3-GmCHS-1-GmCHIL. The pET15b-GmNcLC-3-GmCHS-1-GmCHIL recombinant plasmids were respectively used to transform E. coli strain Rosetta2 (DE3) to obtain the strains CHIL– and CHIL+, respectively. The transformed cells were grown on Luria-Bertani (LB) medium containing 50 μg/ml kanamycin and 34 μg/ml chloramphenicol.

The transformed cells were grown in 20 ml LB medium containing 50 μg/ml ampicillin and 50 μg/ml chloramphenicol at 37 °C with shaking until the optical density at 600 nm (OD600) reached 0.4–0.5. Isopropyl 1-β-D-thiogalactoside was then added to the culture at a final concentration of 1.0 mM, after which the cells were incubated at 18 °C for 20 h. Cells were harvested by centrifugation, washed twice with 20 ml medium, and suspended in M9 medium for an OD600 of 0.5 (final volume, 200 μl). p-Coumaric acid was added to the cell suspension for a final concentration of 1 mM and the mixture was incubated at 30 °C for 3 h and then centrifuged. The target products (LCC, CTAL, and BNY) in the supernatant and precipitate fractions were extracted with 2:1 (v/v) mixture of methanol: chloroform and isolated by LC/MS as described above.

To quantify the expressed GmCHS-1 and GmCHIL proteins, in which GmCHS-1 was His6-tagged while GmCHIL was not, the crude extracts of the engineered E. coli cells, as well as known amounts of GmCHS-1 and GmCHIL proteins, were subjected to SDS-PAGE followed by western blotting analyses using anti-Gm4CL-3 (to determine GmCHS-1) and anti-CHIL IgG (to determine GmCHIL). The immunoblots were quantified using ImageJ software (National Institutes of Health, Bethesda, MD, USA).

**Preparation of antibodies.** Anti-AMCSH antibodies were prepared as described previously. To obtain anti-CHIL antibodies, the peptides corresponding to amino acid residues 121–125 of AMCSH, which were predicted to serve as epitopes, were used to immunize female Japanese white rabbits. The initial injection was administered in complete Freund’s Adjuvant with all subsequent immunizations in incomplete Freund’s Adjuvant. Immunoglobulin G (IgG) was purified from the collected serum as described previously. Monoclonal anti-His6 antibody was obtained from Wako, Osaka, Japan (Code: 011-23091, Lot. LKQ2757).

**Phylogenetic analysis.** The amino acid sequences of CHS and their homologs were used as queries to search a nucleotide sequence database of protein-coding genes from nine plant species (A. thaliana, G. max, V. vinifera, O. sativa, S. moellendorfii, S. falax, M. polymorpha, P. patens, and C. reinhardtii) using the tblast program. Sequences were retrieved from the Phytozone (version 12.1) database. Pimax solvent conditions and CHS amino sequences were obtained from the original publications. Matched sequences with e-value > 0.01 or match length < 350 bp were eliminated. After merging the query sequences with the matched sequences, a multiple alignment of the amino acid sequences was generated using MUSCLE with the default settings. Phylogenetic analyses were performed using the neighbor-joining (NJ) method and maximum likelihood (ML) method as implemented in SEA- VIEW version 4.7 and RaxML-NL version 0.9.6, respectively. Poisson corrected evolutionary distance was used for NJ tree reconstruction. An LG amino acid substitution matrix with gamma model rate heterogeneity and empirical amino acid frequencies was used for the ML analysis. Confidence of inferred phylogenetic trees were assessed by bootstrap method with 100 replicates.

**Modeling studies.** The AtCHIL crystal structure (PDB ID, 4DOK) was obtained from the Protein Data Bank (http://www.rcsb.org/pdb/) as the receptor molecule. The known structure of p-coumaroyl-lketide-CoA (PubChem CID: 70679933) was used to determine the structure of p-coumaroyl-tetraetide-CoA. Discovery Studio and Schrödinger (Schrödinger, LCC, New York, NY, USA) suites were used for modeling, structure creation, and validation. MAESTROSiteMap was used for identifying the binding sites on the surface of the AtCHIL structure, whereas GLIDE and GOLD were used to analyze docking. For all docked complexes, we calculated the Prime/Molecular Mechanics Generalized Born Surface Area function to estimate the Gibbs energy of binding (ΔG of binding) (Prime/ MM-GBSA) and the Piecewise Linear Potential (chemPLP) function to estimate the accuracy of the docking, with the hydrogen bonding term and multiple linear potentials used to model van der Waals and repulsive terms.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** All data needed to evaluate the conclusions in the manuscript are presented herein and/or as Supplementary Information. Additional data related to this study may be requested from the authors. The sequences for all genes described in this manuscript are available in the GenBank/EMBL/DDBJ databases and other databases shown in Supplementary Table 4. The source data underlying Figs. 4–6 as well as the source data of the gels and immunoblots in Supplementary Figs. 4, 6b, and 8 are provided in the Source Data file.

Received: 20 August 2019; Accepted: 17 January 2020; Published online: 13 February 2020

**References.**

1. Arita, M. & Suwa, K. Search extension transforms Wiki into a relational system: a case for flavonoid metabolite database. BioData Min. 1, 7 (2008).

2. Mouradov, A. & Spangenberg, G. Flavonoids: a metabolic network mediating plant adaptation to their environment. Front. Plant Sci. 5, 620 (2014).

3. Gould, K. S. & Lister, C. in Flavonoid Chemistry, Biochemistry and Applications (eds Andersen, Ø. & Markham, K.) (CRC Press, Taylor & Francis, 2006).

**ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-14558-9**
13. Moghe, G. D. & Last, R. L. Something old, Something new: Conserved
11. Nakayama, T., Takahashi, S. & Waki, T. Formation of
10. Mameda, R., Waki, T., Kawai, Y., Takahashi, S. & Nakayama, T. Involvement
9. Fujino, N. et al. Physical interactions among enzymes of the

6. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics,
5. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III

16. Morita, Y. et al. A chalcone isomerase-like protein enhances
31. Nelson, B. K., Cai, X. & Nebenfuhr, A. A multicolored set of
30. Gookin, T. E. & Assmann, S. M. Significance of crowding in living cells: leaving the test-tube behind.
29. Kaltenbach, M. et al. Evolution of chalcone isomerase from a noncatalytic
28. Thomasson, K. A. Glycolytic enzyme interactions with yeast and skeletal
27. Knecht, J. J. Exp. Bot. 60, 1384 (2009).
26. Tohge, T., Matsui, K., Ohme-Takagi, M., Yamazaki, M. & Saito, K. Enhanced radical scavenging activity of genetically modified Arabidopsis seeds. Biotechnol. Lett. 27, 297–303 (2005).
25. Kitamura, S. et al. Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of Arabidopsis thaliana flavonoid accumulation mutants. Plant J. 62, 549–559 (2010).
24. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption-coefficient of a protein. Protein Sci. 4, 2411–2423 (2015).
23. Zhang, Y., Werling, U. & Edelmann, W. Seamless ligation cloning extract (SLICE) cloning method. Methods Mol. Biol. 1116, 235–244 (2014).
22. Schwegendiek, A., Pfeffer, K. & Kindl, H. Pine stilbene synthesis cDNA, a tool for probing environmental stress. EBS Lett. 301, 41–44 (1992).
21. Fliegmann, J., Schröder, G., Schanz, S., Britsch, I. & Schroeder, J. Molecular analysis of chalcone and dihydroxyflavonol synthase from Scots pine (Pinus sylvestris), and differential regulation of these and related enzyme activities in stressed Plant. Mol. Biol. 18, 489–503 (1992).
20. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–427 (1987).
19. Felsenstein, J. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. Syst. Biol. 22, 240–249 (1973).
18. Kozakov, A. M. et al. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455 (2019).
17. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320 (2008).
16. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791 (1985).
15. Sytèmes D. BIOVIA Discovery Studio Modeling Environment, Release 2017 (Dassault Systèmes, 2016).
14. Abdel, R., Young, T., Fardin, R., Berne, B. J. & Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. J. Am. Chem. Soc. 130, 2817–2831 (2008).
13. Friesner, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47, 1739–1749 (2004).
12. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. Proteins 52, 609–623 (2003).
11. Schrödinger Release 2018-4: Prime S (LLC, New York, NY, 2018).
10. Korb, O., Stutzle, T. & Exner, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. J. Chem. Inf. Model. 49, 84–96 (2009).
9. Fujino, N. et al. cDNA cloning and characterization of chalcone isomerase-fold proteins from snapdragon (Antirrhinum majus L.) flowers. Plant Physiol. Biochem. 31, 105–113 (1993).
8. Shi, C. H. et al. Functional characterization of key structural genes in rice flavonoid biosynthesis. Planta 228, 1043–1054 (2008).
7. Pang, Y. et al. Characterization and expression of chalcone synthase gene from Ginkgo biloba. Plant Sci. 168, 1525–1531 (2005).
6. Jiang, C., Schömmner, C., Kim, S. Y. & Suh, D. Y. Cloning and characterization of chalcone synthase from the moss, Physcomitrella patens. Physicochemistry 67, 2531–2540 (2006).
5. Li, M. et al. Genome structure and evolution of Antirrhinum majus L. Nat. Plants 5, 174–183 (2019).
4. Yonekura-Sakakibara, K., Tohge, T., Niida, R. & Saito, K. Identification of a flavone 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. J. Biol. Chem. 282, 14932–14941 (2007).
3. Tohge, T., Matsui, K., Ohme-Takagi, M., Yamazaki, M. & Saito, K. Enhanced radical scavenging activity of genetically modified Arabidopsis seeds. Biotechnol. Lett. 27, 297–303 (2005).
2. Fuchs, S. et al. Arabidopsis structural genomics. Curr. Opin. Plant Biol. 8, 395–405 (2005).
1. Yoshida, S., Saka, K. & Sato, K. Identiﬁcation of an Arabidopsis homolog of Drosophila cinnabar. Biochem. J. 381, 519–524 (2004).

Acknowledgements
We are grateful to Prof. Richard L Smith, Graduate School of Environmental Studies, Tohoku University, for his invaluable advice. We thank Edanz Group (www.edanze-diting.com/ia) for editing a draft of this manuscript. This study was supported in part by a JSPS KAKENHI grant (18H03938) and the NIBB Collaborative Research Program (14-388, 14-352, 16-319, 17-332, and 18-337).

Author contributions
T.W., K.S., K.Y.-S., Y.M.H., A.H., S.T. and T. Nakayama conceived and designed experiments; T.W., R.M., T. Nakano, S. Yamada, S. Yamashita, M.T., K.I., N.T., Y.L., N.F. K.U., Y.A., S.S. and K.Y.-S. performed the experiments; T.W., K.Y.-S. and A.H. contributed plant materials, reagents, and analytical tools; K.D. performed the computational docking studies; Y.K., T.W. and T. Nakayama performed the phylogenetic analyses; T.W. and T. Nakayama wrote the paper.

Competing interests
The authors declare no competing interests.
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-14558-9.

Correspondence and requests for materials should be addressed to T.N.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020