Journal of Experimental Botany, Vol. 71, No. 1 pp. 290–304, 2020
doi:10.1093/jxb/erz425 Advance Access Publication September 26, 2019
This paper is available online free of all access charges (see https://academic.oup.com/jxb/pages/openaccess for further details)

RESEARCH PAPER

Identification and evolutionary analysis of chalcone isomerase-fold proteins in ferns

Rong Ni, Ting-Ting Zhu, Xiao-Shuang Zhang, Piao-Yi Wang, Chun-Jing Sun, Ya-Nan Qiao, Hong-Xiang Lou and Ai-Xia Cheng*,

Key Laboratory of Chemical Biology of Natural Products, Ministry of Education, School of Pharmaceutical Sciences, Shandong University, Jinan, 250012, China

* Correspondence: aixiacheng@sdu.edu.cn

Received 30 July 2019; Editorial decision 10 September 2019; Accepted 12 September 2019

Editor: Howard Griffiths, University of Cambridge, UK

Abstract

The distribution of type I and II chalcone isomerases (CHIs) in plants is highly family specific. We have previously reported that ancient land plants, such as the liverworts and *Selaginella moellendorffii*, harbor type II CHIs. To better understand the function and evolution of CHI-fold proteins, transcriptomic data obtained from 52 pteridophyte species were subjected to sequence alignment and phylogenetic analysis. The residues determining type I/II CHI identity in the pteridophyte CHIs were identical to those of type I CHIs. The enzymatic characterization of a sample of 24 CHIs, representing all the key pteridophyte lineages, demonstrated that 19 of them were type I enzymes and that five exhibited some type II activity due to an amino acid mutation. Two pteridophyte chalcone synthases (CHSs) were also characterized, and a type IV CHI (CHIL) was demonstrated to interact physically with CHSs and CHI, and to increase CHS activity by decreasing derailment products, thus enhancing flavonoid production. These findings suggest that the emergence of type I CHIs may have coincided with the divergence of the pteridophytes. This study deepens our understanding of the molecular mechanism of CHIL as an enhancer in the flavonoid biosynthesis pathway.

Keywords: Chalcone isomerase, chalcone synthase, evolution, ferns, flavonoids, type IV CHI.

Introduction

Flavonoids represent a large group of polyphenolic secondary metabolites that are widely present in land plants and possess significant and diverse biotic and abiotic functions (Falcone Ferreyra *et al.*, 2012; Hassan and Mathesius, 2012); some of these compounds are also pharmacologically active (Kozlowska and Szostakwegierek, 2014). The biosynthesis of flavonoids starts with the condensation of one *p*-coumaroyl-CoA molecule with three malonyl-CoA molecules, a reaction catalyzed by the enzyme chalcone synthase (CHS; EC 2.3.1.74); this results in the formation of the compound naringenin chalcone (6'-hydroxychalcone). CHS is a homodimer, comprising two 40–45 kDa subunits; each monomer utilizes a Cys–His–Asn catalytic triad lying within a buried active site cavity (Ferrer *et al.*, 1999). The enzyme chalcone isomerase (CHI; EC 5.5.1.6) catalyzes the intramolecular and stereospecific cyclization of naringenin chalcone to form 5,7,4′-trihydroxyflavanone (naringenin). Although chalcones are readily isomerized into (2*RS*)-flavanones spontaneously, only the (2*S*)-flavanones are used as intermediates *in planta* for the synthesis of flavonoids. Four types of CHI are recognized according to their phylogenetic relationships and function (Ralston *et al.*, 2005). Type I enzymes exclusively isomerize naringenin chalcone to form (2*S*)-naringenin (Shimada *et al.*,

2003), whereas type II enzymes accept both naringenin chalcone and isoliquiritigenin (6'-deoxychalcone), yielding, respectively, (2*S*)-naringenin and liquiritigenin (5-hydroxyflavanone) (Fig. 1). It has been suggested that type II CHIs are unique to leguminous plants, but, according to our recent investigation, they are also produced by the ancient terrestrial plant species liverworts and *Selaginella* (Cheng *et al.*, 2018). In addition to the bona fide type I and II CHIs, additional CHI-fold proteins are also found in plants; these are termed type III and type IV CHIs (Ralston *et al.*, 2005; Ngaki *et al.*, 2012). Type III CHIs, which are found both in many land plant species and in green algae, do not exhibit any CHI activity. This type of CHI-fold proteins were shown to bind fatty acids *in vitro*, and have thus been proposed to be involved in fatty acid metabolism (Ngaki *et al.*, 2012). Type IV CHIs (CHILs) are completely devoid of CHI catalytic activity because of several substitutions among the catalytic residues and are thought to be confined to land plants (Ralston *et al.*, 2005; Ngaki *et al.*, 2012). A loss-of-function mutation in type IV CHIs led to the lower accumulation of flavonoid and anthocyanins in *Ipomoea nil* (Morita *et al.*, 2014), suggesting that they acted as enhancers of flavonoid production and contribute towards flower pigmentation (Morita *et al.*, 2014). The underlying mechanism of the type IV CHIs was suggested to be that they could physically interact with CHIs or CHSs (Jiang *et al.*, 2015; Ban *et al.*, 2018). The products of two CHIL-encoding genes harbored by the hop (*Humulus lupulus*) appear to be significant for the synthesis of demethylxanthohumol in the trichomes; while one protein enhances the catalytic efficiency of CHS_H1 and PT1L, the other stabilizes the open ring conformation of the product (Ban *et al.*, 2018). Phylogenetic- and sequence-based analyses have suggested that the type IV CHIs are only loosely related to the bona fide CHIs, which in turn probably evolved from a common ancestor thought to have been a type III CHI (Ngaki *et al.*, 2012).

The pteridophytes, which first appeared in the fossil record ~360 million years ago, are the second most diverse group of vascular plants (Duff and Nickrent, 1999; Pryer *et al.*, 2001).

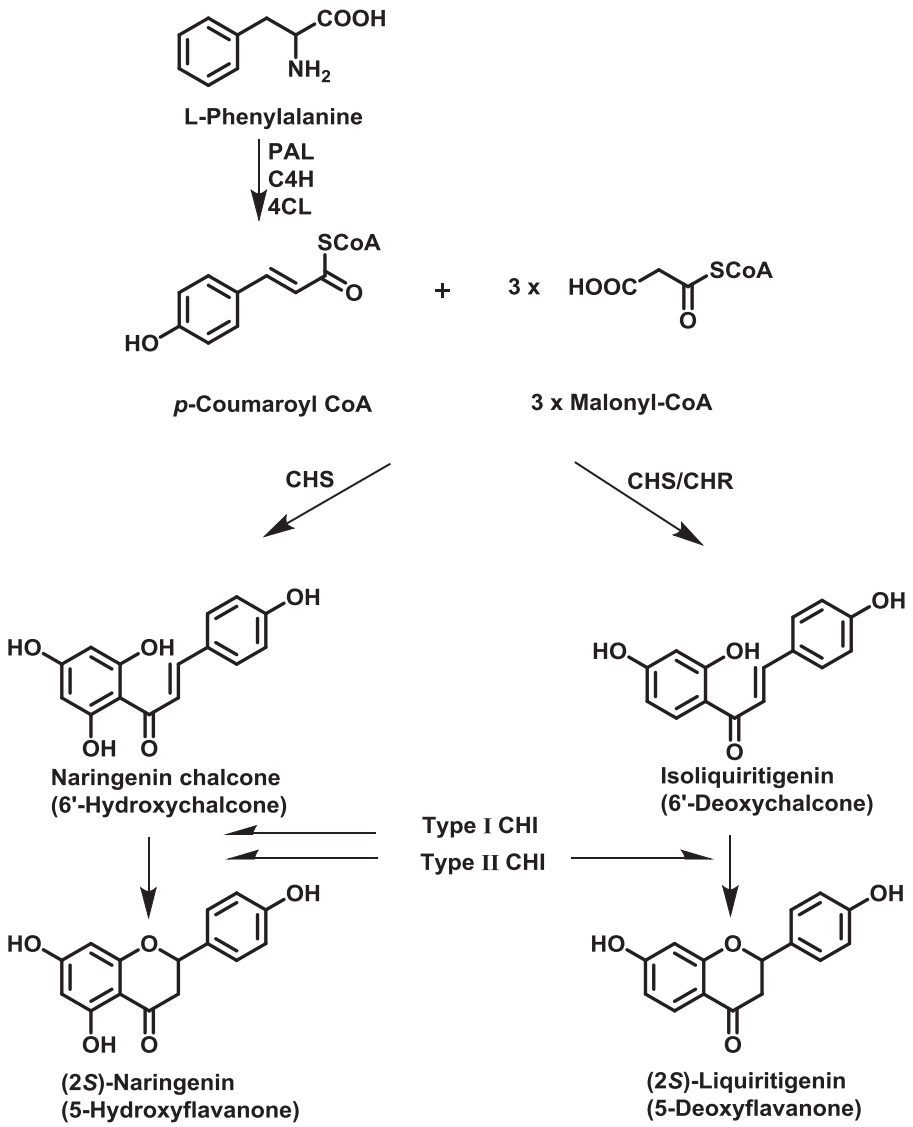

Fig. 1. A schema illustrating synthesis of flavonoids in plants. PAL, phenylalanine ammonia lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate:coenzyme A ligase; CHS, chalcone synthase, CHR, chalcone reductase; CHI, chalcone isomerase. Type I CHIs isomerize exclusively naringenin chalcone, whereas type II CHIs accept both isoliquiritigenin and naringenin chalcone.

The genus *Selaginella* is the sole surviving genus within the lycophyte Selaginellaceae family; it is held to represent the oldest lineage of vascular plants (Weng and Noel, 2013). A recent suggestion is that type II CHIs are characterized from liverworts and spike moss but type I CHIs are not present either in bryophytes or in *Selaginella*. In addition, in contrast to the liverwort CHIs, *Selaginella* CHI favors naringenin chalcone over isoliquiritigenin as its substrate, acting like type I CHI enzymes (Cheng *et al.*, 2018). Here, an attempt has been made to clarify the function and evolution of CHI-fold proteins in the pteridophytes, by exploiting a set of transcriptomic data collected from 52 pteridophyte species representing all of the important lineages. An analysis of their predicted peptide sequences has focused on determining the identity of the key residues responsible for type I or type II CHI activity. The functional characterization results indicated that while most of enzymes expressed type I activity, a few did show some evidence of type II functionality. In addition, evidence has been provided to show that type IV CHIs can promote the accumulation of flavonoids by physically interacting with CHS and CHI, and decreasing the derailment product in the CHS-catalyzed reaction.

Materials and methods

Chemicals and reagents

All available commercial chemicals used in this study were purchased from either Chengdu Must Bio-technology (Chengdu, China) or Sigma-Aldrich (St. Louis, MO, USA), except for naringenin chalcone, which was purchased from Nanjing GOREN Bio-technology (Nanjing, China). *p*-Coumaroyl-CoA was synthesized by a plant 4-coumarate-CoA ligase (4CL) (Gao *et al.*, 2015) following a published procedure (Beuerle and Pichersky, 2002). To synthesize *p*-coumaroyl-CoA, the reaction were performed at 37 °C for 12 h in 10 ml of 200 mM Tris–HCl (pH 7.5) containing 5 mM $MgCl_2$, 5 mM ATP, 300 μM CoA, 500 μM *p*-hydroxycinnamic acid, and 5 mg of recombinant 4CL protein. The reaction mixture was acidified with 1 ml of 6 N HCl and extracted three times with 5 ml of ethyl acetate to remove residual cinnamic acid. After evaporating the remaining organic solvent, 44% ammonium acetate was added to a final concentration of 4% (w/v). Following centrifugation at 5000 *g* for 10 min, the aqueous solution was loaded on a pre-conditioned 1000 mg SPE cartridge (Supelco, Bellefonte, PA, USA). The column was successively rinsed with 10 ml of 4% ammonium acetate solution to remove unreacted CoA and protein, and with 20 ml of distilled water to elute the CoA ester. The change in absorbance of the reaction mixture was monitored at a wavelength of 333 nm according to the reported absorption maxima for *p*-coumaroyl-CoA. After lyophilization, the sample was dissolved in distilled water, the absorbance at the maximum absorption wavelength was measured by a UV spectrophotometer, and the concentration of *p*-coumaroyl-CoA was calculated using Lange–Beer's law ($A=\varepsilon bc$; *p*-coumaroyl-CoA ε333 nm=21 000 M^{-1} cm^{-1}).

CHI and CHS identification and phylogenetic analysis

In a recent publication, transcriptome data were generated from 69 fern species (Shen *et al.*, 2018). In addition, we performed transcriptome sequencing of five additional fern species: *Stenoloma chusanum* (SRR8185333), *Lindsaea orbiculata* (SRR8185332), *Polypodiodes amoena* (SRR8185331), *Adiantum capillus-veneris* (SRR8185330), and *Cyclosorus parasiticus* (SRR8185329). To investigate the function and evolution of fern CHI-fold proteins, we searched these transcriptome sequence data. Fifty-six putative bona fide CHI-fold proteins and 45 putative *CHI-like* genes were recognized from the pteridophyte species from Polypodiales and the Early Leptosporangiate group (the amino acid sequences of each enzyme are summarized in Supplementary Datasets S1 and S2 At *JXB* online). In addition, two putative *CHS* genes were recognized by searching the transcriptome database of *L. orbiculata*, and these genes were given the designations *LoCHS1* and *LoCHS2*.

To construct a phylogenetic tree, the amino acid sequences were first aligned with CLUSTALW integrated in the MEGA v4.0 program (Tamura *et al.*, 2007). Then, the tree was constructed using the Neighbor–Joining method. The statistical reliability of the nodes of the tree was assessed by bootstrap analyses with 1000 replicates (Tamura *et al.*, 2007).

Sequence alignment and homology modeling

Multiple sequence alignment was performed using the DNAMAN v7 program (Lynnon Biosoft, Quebec, Canada), in which the deduced 56 putative CHI sequences were aligned with those of AtCHI (*Arabidopsis thaliana*, P41088), MsCHI (*Medicago sativa*, P28012), and SmCHI1 (*Selaginella moellendorffii*, EFJ22013.1). The LoCHIL and AcCHIL sequences were aligned with that of AtCHIL (*A. thaliana*, NP_850770), while the LoCHSs from *L. orbiculata* were aligned with MsCHS (*M. sativa*, AAA02824). Homology modeling of AcCHI1 and AcCHI1-F188V was carried out by the Swiss-model server (http://swissmodel.expasy.org) (Arnold *et al.*, 2006), based on the crystal structure of CHIs from *M. sativa* (PDB: 1EYQ and 1FM7) (Jez *et al.*, 2000; Jez and Noel, 2002).

Plant material and nucleic acid extraction

The pteridophyte species were cultured in Shanghai Chenshan Botanical Garden or in the greenhouse in Shandong University. Sporophylls and trophophylls were collected, immediately frozen in liquid nitrogen, and preserved in an ultralow temperature refrigerator at –80 °C before RNA extraction. RNA was isolated using the cetyltrimethylammonium bromide (CTAB) method (Gambino *et al.*, 2008) and used as a template for cDNA synthesis, based on a RevertAid™ First Strand cDNA Synthesis kit (MBI Fermentas, Waltham, MA, USA) primed with $(dT)_{18}$, according to the manufacturer's protocol. *Arabidopsis thaliana* were grown at 22 °C under a 16 h light/8 h dark photoperiod. *Nicotiana benthamiana* were grown at 24 °C under a 12 h light/12 h dark photoperiod.

Gene cloning, heterologous expression, and protein purification

The full-length sequences of 24 CHIs from 21 pteridophyte species, two representative *CHI-like* genes (*LoCHIL* and *AcCHIL*), and two CHS genes (*LoCHS1* and *LoCHS2*) were separately amplified from the synthesized cDNAs with the primer pairs listed in Supplementary Table S1. The amplified ORFs were digested with corresponding restriction endonucleases and ligated into the pET-32a vector (Novagen, Malaysia). After validation by sequencing, the recombinant plasmids as well as the empty pET-32a plasmid were transformed into *Escherichia coli* strain BL21 (DE3), respectively. The overnight cultures (2.0 ml) generated from a single transgenic colony were inoculated into 200 ml of LB medium and incubated at 37 °C with shaking (200 rpm) until reaching an OD_{600} of 0.5. Isopropyl-β-D-thiogalactoside (IPTG) was then added at a final concentration of 0.5 mM. The bacterial cultures were continually incubated for 18 h at 16 °C, then the cells were harvested with centrifugation (4000 *g*, 5 min) and homogenized by sonication on ice. An Ni-NTA Sefinose His-bind column (Bio Basic, Canada) was used to purify the recombinant proteins, following the manufacturer's protocol. Protein concentrations were determined using the Bradford reagent (Beyotime, Shanghai, China), with BSA employed as the standard. The purified proteins were separated by denaturing SDS–PAGE (12% acrylamide), and visualized by Coomassie blue R250 staining (Sun *et al.*, 2013).

CHI enzymatic assay and kinetic analysis

Enzyme assays based on isoliquiritigenin as a substrate were performed at 25 °C for 10 min in a 200 μl reaction containing 50 mM Tris–HCl (pH 7.5), 50 μM substrate, and 10 μg of purified recombinant proteins; control reactions contained a protein extract of *E. coli* BL21 (DE3) carrying an

empty pET-32a plasmid. The reactions were extracted twice in 200 µl of ethyl acetate. Because naringenin chalcone is unstable in buffer solution, the reaction was performed at room temperature, and the substrate was added last. After mixing for 1 min, an equal volume of ethyl acetate was added immediately, and the reactions were extracted twice in 200 µl of ethyl acetate. After removal of the solvent under vacuum, the residue was dissolved in 100 µl of methanol. The samples were separated through a reverse-phase C18 column (XDB-C18, 5 µm; Agilent, Santa Clara, CA, USA) using HPLC (1260 Infinity Binary LC system, Agilent) equipped with a multiwavelength diode array detector using the previously reported procedure (Cheng *et al.*, 2018). Enzyme kinetic assays were performed following the procedure reported previously (Cheng *et al.*, 2018).

Site-directed mutagenesis

The AcCHI1-F188V mutant was created using the Stratagene QuikChange site-directed mutagenesis method. The necessary pair of primers, AcCHI1-F188V-F/R, were designed using the Primer X online software (www.bioinformatics.org/primerx) (Supplementary Table S1). The resulting PCR product was digested with *Dpn*I (Thermo Scientific, USA) to cleave methylated DNA, and an aliquot of gel-purified restriction reaction product was transformed into *E. coli* DH5α. The mutated genes were confirmed by sequencing and expressed in *E. coli* BL21 (DE3) as mentioned above. Enzyme assays were performed using purified mutant recombinant proteins with the same substrates and under the same conditions as mentioned above. Catalytic efficiency was estimated from a standard calibration curve.

Chemical measurements in transgenic Arabidopsis lines

The cDNAs of LoCHI1 and AcCHI1 were amplified with the primer pairs listed in Supplementary Table S1 from the plasmids pET32a-LoCHI1 and pET32a-AcCHI1. The PCR products were used in a BP Clonase reaction for recombination into the pDONR207 vector. Positive clones were then used in an LR Clonase reaction for recombination into the pGWB5 destination vector (Curtis and Grossniklaus, 2003). The resulting pGWB5-LoCHI1 and pGWB5-AcCHI1 constructs were transformed into *Agrobacterium tumefaciens* strain GV3101 using the freeze/thaw method (Bade and Damm, 1995). To achieve gene overexpression *in planta*, the resulting expression cassettes were transformed into *tt5* by the floral dipping method (Clough and Bent, 1998). A similar exercise was carried out with respect to LoCHIL and AcCHIL. LoCHIL and AcCHIL were transformed into *chil* mutant Arabidopsis using the above-mentioned method. Approximately 10 individual transgenic lines for each gene were generated, and two lines with each construct were selected for further analysis based on preliminary HPLC profiling of accumulated flavonoids. The selected transgenic lines were germinated on agar plates containing half-strength Murashige and Skoog medium, 1% agar, and 2% sucrose under a 16 h light and 8 h dark cycle at 22 °C. To measure the anthocyanin content, we employed a spectroscopic method, as previously described (Zhang *et al.*, 2013; Cheng *et al.*, 2018). For proanthocyanidin (PA) staining, seeds of the *chil* mutant, transgenic Arabidopsis overexpression lines of *LoCHIL* and *AcCHIL*, and Col were stained with 0.1% DMACA [4-(dimethylamino)-cinnamaldehyde] reagent in methanol:6 N HCl (1:1, v/v) (Jiang *et al.*, 2015). For extraction of PAs, the dry seeds were ground in liquid nitrogen, and 100 mg batches were extracted with 800 µl of extraction solution (70% acetone:0.5% acetic acid) by vortexing followed by sonication at room temperature for 30 min. Following centrifugation at 2500 *g* for 10 min, the residues were extracted twice as above. The pooled supernatants were then extracted with chloroform, and the aqueous supernatant was extracted twice with chloroform and three times with hexane. Samples were lyophilized and resuspended in extraction solution to a final concentration of 3 mg of original sample per microliter. Samples were mixed briefly and transferred to another tube. In brief, aliquots of samples (2.5 µl) were mixed with 197.5 µl of DMACA reagent [0.2% (w/v) DMACA in methanol:3 N HCl (1:1)] in microplate wells; for blanks, the samples were replaced with 2.5 µl of methanol:3 N HCl. Samples and blanks were read within 15 min on a Wallac Victor 2 plate reader equipped with a 640 nm emission filter. Blanks were subtracted from samples. For measurement of insoluble PAs, the residues of the above tissue extractions were dried in air for 2 d, butanol-HCl reagent was added, and the mixture was sonicated at room temperature for 60 min, followed by centrifugation at 2500 *g* for 10 min. The supernatants were transferred to cuvettes for determination of absorption at 550 nm, and the samples were then boiled for 1 h. After cooling to room temperature, the A_{550} was recorded again, and the first value was subtracted from the second (Pang *et al.*, 2007).

Characterization of LoCHS enzyme activity with type IV CHI in vitro

Enzyme assays for the CHS reaction were performed at 35 °C for 30 min in 250 µl of 50 mM potassium phosphate buffer (pH 6.5) containing 50 µM *p*-coumaroyl-CoA, 100 µM malonyl-CoA, 20 µg of purified recombinant proteins encoded by CHS and CHIL or CHS together with empty pET-32a in an equal molar ratio. The reactions were terminated by adding 25 µl of 10% glacial acetic acid, then extracted twice with 250 µl of ethyl acetate and centrifuged at 5000 *g* for 10 min. After the solvent had been removed under vacuum, the residue was dissolved in 50 µl of 80% methanol. Analysis of enzymatic products was performed using an HPLC–MS Agilent 1100 system equipped with a multiwavelength diode array detector and an electron spray ionization (ESI) mass spectrometer. Samples were separated using a 5 µm reverse-phase XDB-C18 column (Agilent) with a flow rate of 0.8 ml min^{-1}. The eluents were water plus 0.1% glacial acetic acid (A) and methanol (B); a linear gradient of 70% A and 30% B to 20% A and 80% B was supplied over 30 min. Standard solutions of reference compounds were used for calibration.

Yeast two-hybrid assays

Yeast two-hybrid (Y2H) assays were based on the Matchmaker GAL4 two-hybrid system (Clontech, Mountain View, CA, USA). LoCHIL and LoCHI1 were amplified from the pET-32a vector and cloned into the pGBKT7 (BD) vector. LoCHS1, LoCHS2, and LoCHI1 were then subcloned into the pGADT7 (AD) vector (the primers used in this study are listed in Supplementary Table S1). The pGBKT7 and pGADT7 constructs were co-transformed into yeast strain AH109 using the LiAc method (Gietz and Woods, 2002). The presence of the transgenes was confirmed by growth on SD/-Leu/-Trp plates. To assess protein interactions, the transformed yeast was suspended in liquid SD/-Leu/-Trp medium and cultured to OD_{600}=1.0. A 10 µl aliquot of suspended yeast was spread on SD/-His/-Leu/-Trp medium. Interactions were observed after 3 d of incubation at 30 °C.

Luciferase complementation imaging assays

For luciferase complementation imaging (LCI) assays, the genes encoding LoCHIL, LoCHI1, LoCHS1, and LoCHS2 were cloned into the pCAMBIA1300-nLuc or pCAMBIA1300-cLuc vectors using *Kpn*I and *Sal*I, or *Kpn*I and *Bgl*II digestion (the primers used in this study are listed in Supplementary Table S1). The constructs were introduced into *A. tumefaciens* GV3101. Overnight cells were collected by centrifugation, washed with infiltration buffer containing 10 mM MES-KOH (pH 6.5) and 40 µM acetosyringone, and re-suspended in AS medium containing 10 mM $MgCl_2$ and 200 µM acetosyringone to an OD_{600} of 1.0–1.5. The suspensions were prepared by compounding at a 1:1:1 ratio the three *A. tumefaciens* strains carrying the nLuc fusion, the cLuc fusion, and the gene-silencing inhibitor p19, respectively. The *Agrobacterium* mixture was then infiltrated into different locations in the same leaves of *N. benthamiana* for transient expression (Ban *et al.*, 2018). Two to three days after infiltration, an exposure time of 2 min with 3×3 binning was used to capture the LUC images (Chen *et al.*, 2008). Each data point contained three replicates.

Results

Identification of CHI-fold proteins from pteridophyte species

Searching the transcriptome data generated from 74 fern species, 56 putative type I or II CHIs were recognized from 52

fern species. Phylogenetic analysis was performed based on the predicted polypeptide sequences of the putative bona fide type I or II CHI-fold proteins found in ferns, together with the sequences of previously recognized type II CHIs in liverworts and *S. moellendorffii* (Cheng *et al.*, 2018), which represent the main lineages of ancient land plants (Fig. 2). The outcome of the analysis implied that the pteridophyte proteins were not closely related to standard type II CHIs. A recent fern phylogeny, constructed by a coalescent-based method using nucleotide sequences with calculated divergence times, showed that 69 pteridophyte species have been taxonomically divided into three major groups, namely the Eusporangiates, the early Leptosporangiates, and the Polypodiales (Shen *et al.*, 2018). Species belonging to the Eusporangiates harbored no *CHI* homologs. The evolution of the *CHI* genes mirrors speciation within the pteridophytes as a whole. However, the *CHI* genes of various Polypodiales species, namely *A. capillus-veneris CHI1*, *Adiantum caudatum CHI1*, *Antrophyum callifolium CHI1*, *Haplopteris amboinensis CHI1*, and *Pteris vittata CHI1*, appeared to be intermediate between *CHI* genes encoded by *S. moellendorffii* and species belonging to the early Leptosporangiates.

The full length of the deduced polypeptide sequences of the 56 pteridophyte *CHI* genes aligned well with both bona fide type I and type II CHI sequences (Fig. 3). Their level of identity of amino acid sequence to alfalfa MsCHI (P28012) ranged from 19% to 29%, and to *A. thaliana* AtCHI (P41088) from 38% to 40%. Despite this relatively low level of sequence identity, the pteridophyte CHIs have retained a number of conserved residues, for example Arg36, Gly37, Leu38, Phe47, Thr48, Ile50, Leu101, Tyr106, Lys109, and Val110 (using the numbering scheme of CHI from *M. sativa*). The two residues of MsCHI (Thr190 and Met191) postulated to determine substrate preference (Forkmann and Dangelmayr, 1980) were represented by Thr and Ile in 53 of the pteridophyte proteins and by Ser and Ile in the remaining three (Fig. 3). The residues postulated to determine the substrate preference of higher plant type I CHIs are either Ser/Ile or Thr/Ile (Park *et al.*, 2018). In the pteridophytes, the indication is that the two critical residues differentiating type I from type II CHIs are either Thr/Ile or Ser/Ile rather than Thr/Met. Accordingly, the pteridophyte CHIs were considered most likely to be type I enzymes.

Catalytic activities and homology modeling of pteridophyte CHIs

A set of 24 representative pteridophyte *CHI* genes (from 21 species) were isolated and heterologously expressed in *E. coli* to characterize their enzymatic properties. The recombinant enzymes were provided with either 6'-hydroxychalcone (naringenin chalcone) or 6'-deoxychalcone (isoliquiritigenin) as substrate. Nineteen of the enzymes (LoCHI1, ScCHI1, PamCHI1, CpaCHI1, AlCHI1, AlCHI2, CbCHI1, HpCHI1, DaCHI1, DbCHI1, PbCHI1, MspCHI1, SpCHI1, SpCHI2, PniCHI1, PniCHI2, LmCHI1, ApaCHI1, and MstCHI1) were only able to accept naringenin chalcone and so were classified as type I CHIs (Fig. 4; Supplementary Figs S1, S2), while the other five (AcCHI1, AcaCHI1, AdcCHI1, HaCHI1, and PvCHI1) exhibited slight type II CHI activity (Fig. 4; Supplementary Figs S1, S2). As delineated in the two representative HPLC chromatograms shown in Fig. 4, AcCHI1 yielded 5-hydroxyflavanone (naringenin) and 5-deoxyflavanone (liquiritigenin) when provided with, respectively, naringenin chalcone and isoliquiritigenin (Fig. 4b, g). On the other hand, LoCHI1 produced naringenin when presented with naringenin chalcone, but was unable to catalyze the conversion of isoliquiritigenin to any product (Fig. 4c, h). Thus, AcCHI1 was deemed to exhibit type II CHI cyclization activity, since it could accept both isoliquiritigenin and naringenin chalcone as substrate, whereas LoCHI1 showed type I CHI activity and was only able to cyclize naringenin chalcone. Although the identity of the residues postulated to determine their substrate preference was consistent with their being type I CHIs (Supplementary Fig. S3), the five enzymes AcCHI1, AdcCHI1, AcaCHI1, HaCHI1, and PvCHI1 were all able to catalyze type II-specific substrates. To further elucidate the mechanism, protein homology modeling was performed for AcCHI1 using MsCHI (PDB code 1EYQ and 1FM7) as a template. The results indicated that the residue Phe188 in the substrate-binding pocket of AcCHI1 was conserved in these five type II CHIs; this residue corresponds to the Val amino acid in type I CHIs (Supplementary Figs S4, S5). When Phe188 was mutated to Val188 in AcCHI1, the altered version of the enzyme converted naringenin chalcone to naringenin, and was unable to catalyze the conversion of isoliquiritigenin (Fig. 5). Taken together, these results suggested that the amino acid Phe rather than the residues Thr and Met determined the type II CHI activity of this subgroup of enzymes.

Kinetic parameters of the type I/II CHIs

The kinetic characteristics of four representative CHIs (LoCHI1, ScCHI1, AcCHI1, and AcaCHI1) were identified using naringenin chalcone and isoliquiritigenin as substrates. These enzymes displayed comparable kinetic properties to those previously reported liverwort and *Selaginella* bona fide CHIs toward naringenin chalcone (Table 1) (Cheng *et al.*, 2018). However, when isoliquiritigenin was used as substrate, the catalytic efficiencies (K_{cat}/K_m values) of type II enzymes from pteridophyte (AcCHI1 and AcaCHI1) were slightly lower than *Selaginella* CHI (Cheng *et al.*, 2018). Moreover, the catalytic efficiencies of AcCHI1 and AcaCHI1 for isoliquiritigenin (3.95×10^5 and 4.14×10^5, respectively) were much lower than those for naringenin chalcone (9.58×10^7 and 9.41×10^7, respectively) (Table 1). These data demonstrate that pteridophyte type II CHIs showed similar catalytic efficiencies towards naringenin as type I CHI and exhibited lower activity against isoliquiritigenin than type II CHIs from the basal land plants liverworts and spike moss.

Complementation of the tt5 *mutant with pteridophyte bona fide CHIs*

The genes encoding both AcCHI1 and LoCHI1 were transformed separately into the *A. thaliana tt5* mutant to check their

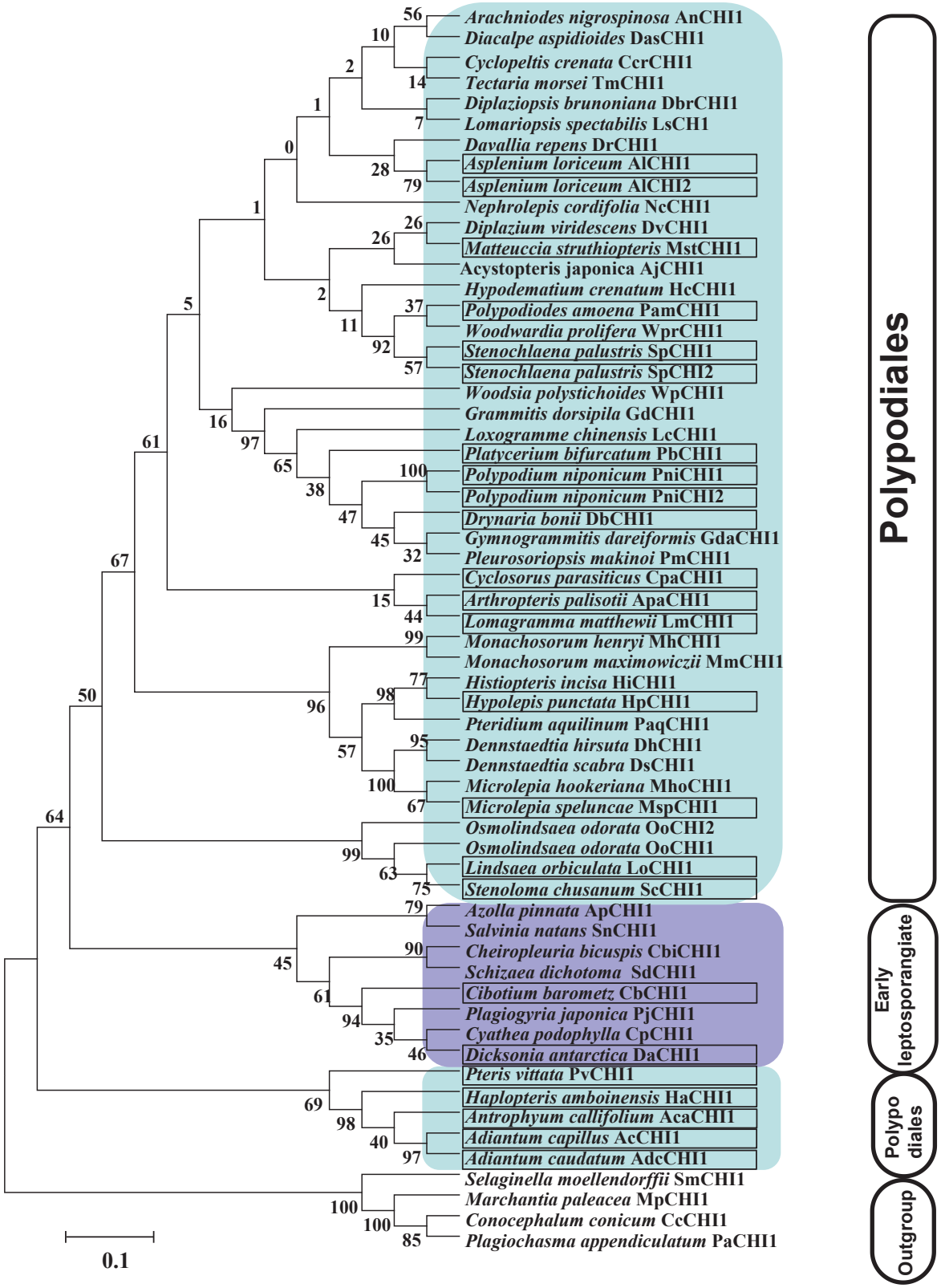

Fig. 2. The phylogeny of CHI-fold proteins. The tree was constructed using the Neighbor–Joining algorithm with the aid of 1000 bootstrap replicates. The length of each branch reflects the distance between nodes. The outgroup sequences used include *Selaginella moellendorffii* CHI1 (EFJ22013.1), *Conocephalum conicum* CHI1 (AOC83884.1), *Plagiochasma appendiculatum* CHI1 (AOC83883.1), and *Marchantia paleacea* CHI1 (AOC83881.1). All of the CHI polypeptide sequences are given in Supplementary Dataset S1. The genes isolated and functionally characterized in the present investigation are shown in boxes. (This figure is available in color at *JXB* online.)

ability to complement the absence of *AtTT5*. The transparent testa of mature seeds set by the *tt5* mutant is pale yellow (Focks *et al.*, 1999); in both of the transformants, this was altered to dark brown, rendering the appearance of the seed similar to that of seed set by wild-type plants (Fig. 6A). Both the anthocyanin and flavonol aglycone content of the transgenic plants

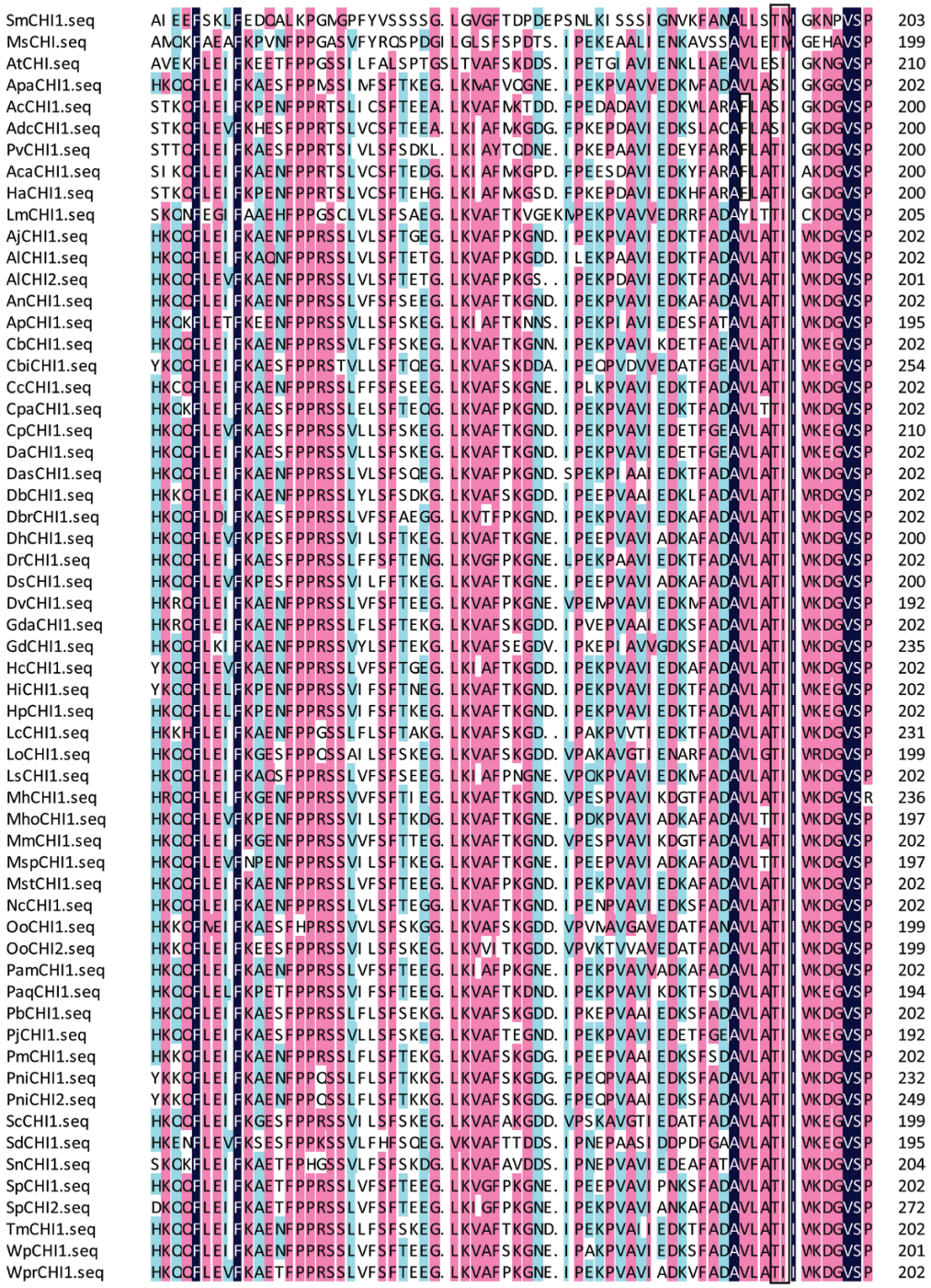

Fig. 3. Sequence alignment of 56 CHIs produced by 52 pteridophyte species with CHIs from other plants. SmCHI1 (*Selaginella moellendorffii*, EFJ22013.1), MsCHI (*Medicago sativa* CHI, P28012), and AtCHI (*Arabidopsis thaliana* CHI, P41088). Residues responsible for the preference between isoliquiritigenin and naringenin chalcone as substrate are shown in boxes. (This figure is available in color at *JXB* online.)

was higher than that of the *tt5* mutant (Fig. 6B–E). In particular, anthocyanin pigment accumulated at the junction between cotyledons and hypocotyl in the transgenic plants, but not in non-transgenic *tt5* mutant plants (Fig. 6B), and the total anthocyanin content of the transgenic seedlings was higher than that of the non-transgenic seedlings (Fig. 6C). The conclusion was that both AcCHI1 and LoCHI1 were able, at least partially, to rescue the *tt5* mutant phenotype, confirming that pteridophyte CHIs possess catalytic activity similar to that of higher plant CHIs.

Complementation of the chil *mutant by two genes encoding pteridophyte type IV CHIs*

The level of peptide identity between AtCHIL (NP_850770) and LoCHIL was 39.6%, and between AtCHIL and AcCHIL was

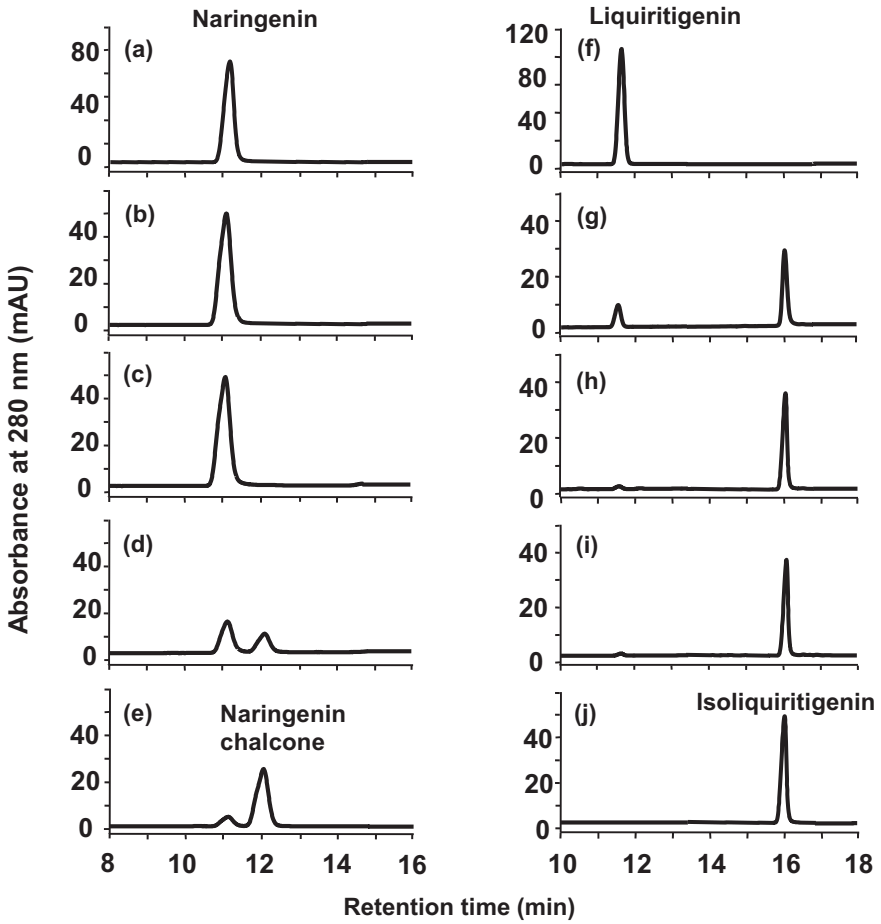

Fig. 4. HPLC analysis of the products of *in vitro* reactions including recombinant AcCHI1 and LoCHI1. The substrate provided was either naringenin chalcone or isoliquiritigenin. (a) Naringenin standard. (b–d) HPLC profiles of the reaction products of (b) AcCHI1, (c) LoCHI1, and (d) negative control (empty pET-32a vector) using naringenin chalcone as substrate. (e) Naringenin chalcone standard. (f) Liquiritigenin standard. (g–i) HPLC profiles of the reaction products of (g) AcCHI1, (h) LoCHI1, and (i) negative control (empty pET-32a vector) using isoliquiritigenin as substrate. (j) Isoliquiritigenin standard.

36.3%. Type IV CHIs lack the catalytic core of the CHI fold and feature substitutions at nearly all the critical sites, including the widely conserved alfalfa Asn113 (Supplementary Fig. S6). When recombinant AcCHIL and LoCHIL proteins were provided with either naringenin chalcone or isoliquiritigenin as substrate, no activity was detectable (Supplementary Fig. S7). When the genes encoding AcCHIL and LoCHIL were separately transformed into the *A. thaliana chil* mutant, DMACA staining suggested that the seeds of the transgenic plants accumulated more PAs than the mutant plants, although less than that accumulated in the wild-type seeds (Fig. 7A). The presence of either transgene increased the content of both extractable and non-extractable PA in the mature seeds (Fig. 7B), although, in both cases, the level remained slightly lower than that achieved in wild-type seeds. The transgenic plants also produced more flavonol aglycones than did the *chil* mutants (Fig. 7C, D). Thus, both AcCHIL and LoCHIL were able to partially rescue the *chil* phenotype.

Type IV CHI-fold proteins enhance the activity of CHSs

To test the interactions between CHIL and CHS, two CHS-encoding genes were isolated from *L. orbiculata*: these were designated *LoCHS1* and *LoCHS2*. The former gene encoded a 405 residue polypeptide and the latter a 439 residue polypeptide. A sequence alignment indicated that both of these polypeptides aligned well with higher plant CHSs; they also retained the conserved Cys–His–Asn catalytic triad and most of the active site residues identified in *M. sativa* CHS (AAA02824) (Supplementary Fig. S8) (Ferrer *et al.*, 1999; Focks *et al.*, 1999). The influence of LoCHIL on the enzymatic efficiency of both LoCHS1 and LoCHS2 was tested by assessing the level of enzymatic activity when the reactions were provided with *p*-coumaroyl-CoA and malonyl-CoA; the products were characterized using LC-MS (Fig. 8A, B). Previous studies have indicated that, during CHS reactions *in vitro*, 4-coumaroyltriacetic acid lactone (CTAL) and bisnoryangonin (BNY) are derailment products formed at the triketide or tetraketide stages, respectively (Jiang *et al.*, 2006; Ma *et al.*, 2009). Here, reactions based on LoCHS1 or LoCHS2 or LoCHS1/LoCHIL or LoCHS2/LoCHIL generated two products (p1 and p2). The latter product shared the same HPLC retention time as a naringenin standard. Its negative ESI-ion mass spectrum comprised a molecular ion $[M\text{-}H]^-$ at an m/z of 271, consistent with it being naringenin, which was produced by the spontaneous cyclization of naringenin chalcone. The p1 product also showed an $[M\text{-}H]^-$ ion at m/z 271 (Fig. 8B); the MS^2 spectrum

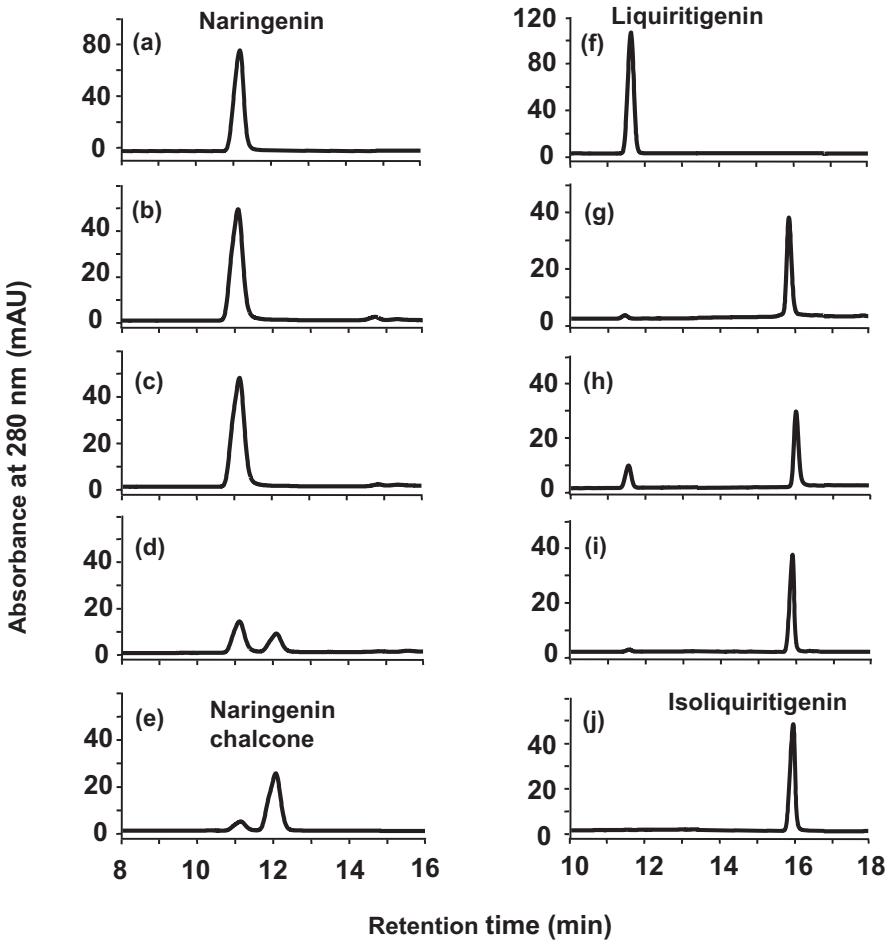

Fig. 5. HPLC analysis of the products of *in vitro* reactions comparing recombinant wild-type and a site-directed mutagenesis-altered form of AcCHI1. The substrate provided was either naringenin chalcone or isoliquiritigenin. (a) Naringenin standard. (b–d) HPLC profiles of the reaction products of (b) AcCHI1-F188V, (c) AcCHI1, and (d) negative control (empty pET-32a vector) using naringenin chalcone as substrate. (e) Naringenin chalcone standard. (f) Liquiritigenin standard. (g–i) HPLC profiles of the reaction products of (g) AcCHI1-F188V, (h) AcCHI1, and (i) negative control (empty pET-32a vector) using isoliquiritigenin as substrate. (j) Isoliquiritigenin standard.

Table 1. *Kinetic analysis of recombinant LoCHI1, ScCHI1, AcCHI1, and AcaCHI1 using naringenin chalcone and isoliquiritigenin as substrates*

Substrate	Enzyme	K_m (μM)	V_{max} (μmol mg^{-1} min^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (M^{-1} min^{-1})
Naringenin chalcone	LoCHI1	90.28±19.71	204.66±15.60	9221.75±702.8	10.2×10^7
	ScCHI1	96.48±20.3	213.74±16.21	9304.19±705.75	9.64×10^7
	AcCHI1	98.2±21.8	214.32±17.31	9406.33±759.56	9.58×10^7
	AcaCHI1	99.59±22.4	213.90±17.62	9369.0±771.76	9.41×10^7
Isoliquiritigenin	LoCHI1	–	–	–	–
	ScCHI1	–	–	–	–
	AcCHI1	61.4±9.29	0.553±0.05	24.27±2.08	3.95×10^5
	AcaCHI1	54.29±9.806	0.514±0.05	22.51±2.20	4.14×10^5

of this product was identified as that of CTAL by comparing it with the published spectroscopic data (Yu *et al*., 2018) and according to the reaction mechanism (Fig. 8B, C). Reactions run with the enzyme combination LoCHS1/LoCHIL produced 2.7-fold higher naringenin than did those based on only LoCHS1, while those using LoCHS2/LoCHIL produced 2.3-fold higher naringenin than did those based on only LoCHS2 (Fig. 8D). In contrast, when LoCHS1 or LoCHS2 interacted with LoCHIL, the content of CTAL decreased significantly. The levels of CTAL decreased greatly (down to 56% and 58%, respectively) in both LoCHS1/LoCHIL and LoCHS2/LoCHIL compared with CHS alone (Fig. 8D).

Type IV CHI-fold proteins physically interact with CHI and CHSs

A Y2H assay, designed to assess interactions between LoCHIL, LoCHSs, and LoCHI1, showed that strains co-transformed

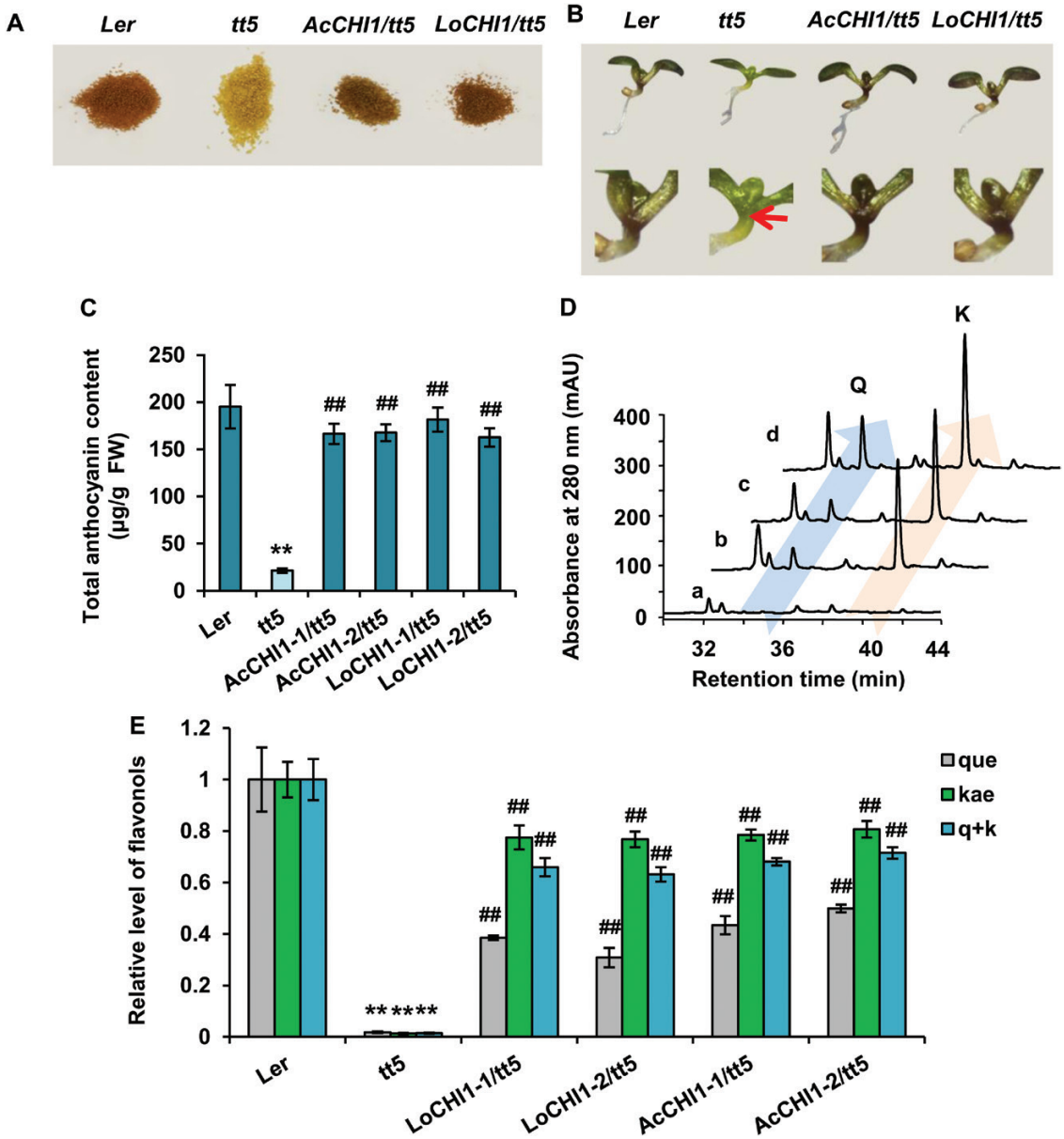

Fig. 6. Complementation of the *A. thaliana tt5* mutant phenotype by *AcCHI1* and *LoCHI1*. (A) Mature *A. thaliana* seeds harvested from wild-type plants (ecotype Ler), the *tt5* mutant, and transgenic lines constitutively expressing *AcCHI1* or *LoCHI1* in the *tt5* mutant background. (B) Seedlings 10 d after germination (DAG) of Ler, *tt5*, and the rescued lines overexpressing *AcCHI1* or *LoCHI1* showing that both types of transgenic plant accumulate anthocyanin in their hypocotyl, which does not occur in *tt5* mutant seedlings (indicated by an arrow). (C) The anthocyanin content of 10 DAG seedlings of Ler, *tt5*, *AcCHI1/tt5*, and *LoCHI1/tt5*. (D) Representative HPLC chromatograms of methanolic extracts of 10 DAG seedlings of (a) *tt5* mutant, (b) *AcCHI1/tt5* transgenic plants, (c) *LoCHI1/tt5* transgenic plants, and (d) the wild type (Ler). (E) Relative levels of quercetin (Q) and kaempferol (K) in 10 DAG seedlings. Levels of quercetin and kaempferol from wild-type Ler were set as a value of 1.0. Data are presented in the form mean ±SD (n=3); *P<0.05 and **P<0.01 versus the Ler group, #P<0.05 and ##P<0.01 versus the *tt5* group. (This figure is available in color at *JXB* online.)

with LoCHIL and LoCHS1 or LoCHIL and LoCHS2 were able to grow on a medium lacking Trp, Leu, and His. Moreover, when LoCHSs or LoCHIL were co-expressed with LoCHI1, yeast growth was also detected. However, no perceptible yeast growth was detected with the negative control (Fig. 9A). These results suggested the protein–protein interactions occurred between LoCHIL and LoCHS1 or LoCHS2, between LoCHIL and LoCHI1, and between LoCHI1 and LoCHS1 or LoCHS2.

The binary protein–protein interactions of the above combinations were further examined *in planta* using LCI assays. The results showed that strong luminescence signals could be detected when the pairs LoCHIL/LoCHS1, LoCHIL/LoCHI1, and LoCHI1/LoCHS1 were co-expressed in *N. benthamiana* leaves, irrespective of the location of the proteins in the N- or C-termini (Fig. 9B). However, for the LoCHS2/LoCHIL or LoCHS2/LoCHI1 pairs, a binary interaction was identified when LoCHS2 was located in the C-termini (Fig. 9B), but was not obviously detected with LoCHS2-nluc/LoCHIL-cluc or LoCHS2-nluc/LoCHI1-cluc. These results were probably due to the difference of the location of the N- and C-termini in the protein 3D structures of the partner proteins used in the LCI assays.

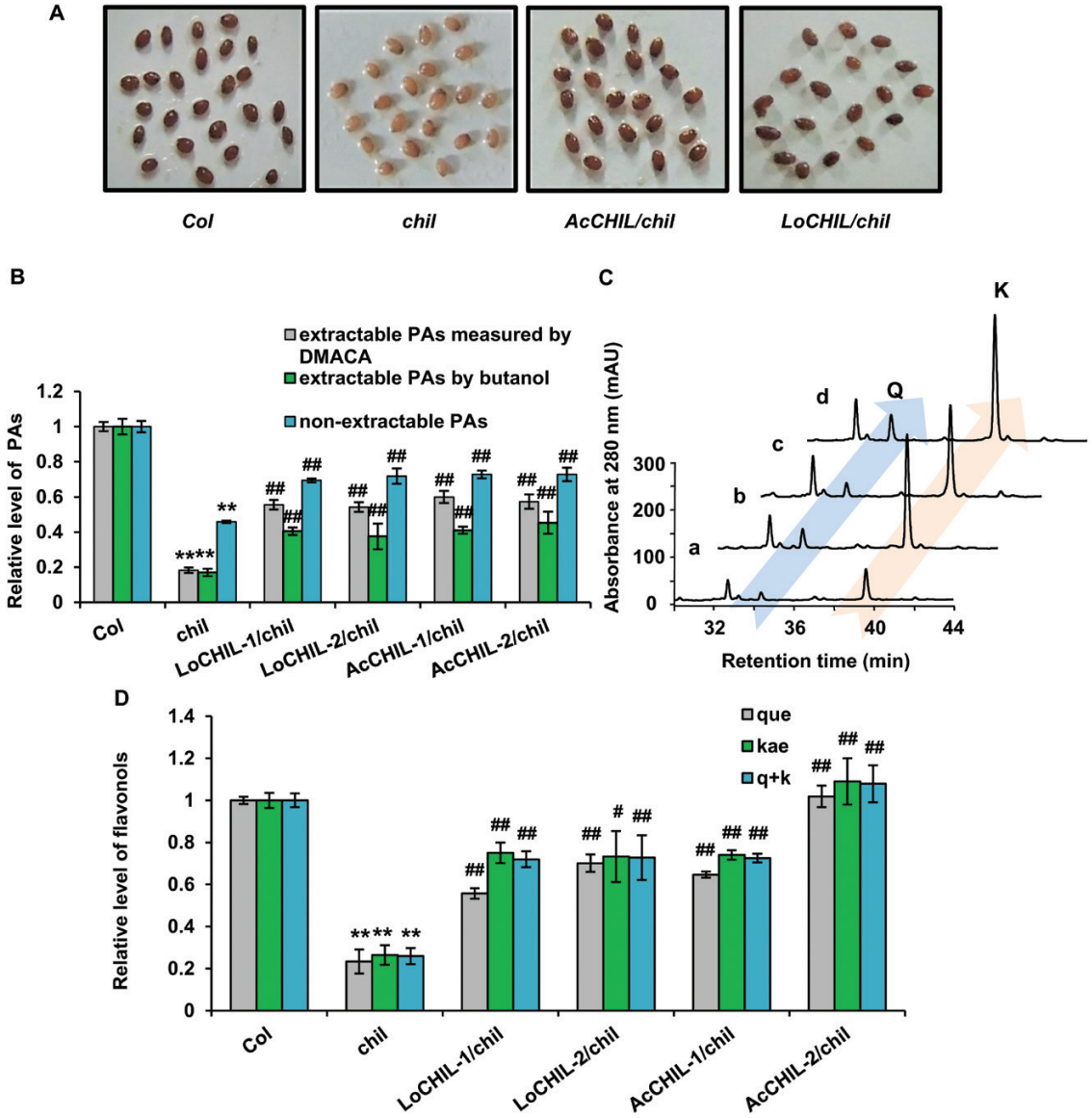

Fig. 7. The effect of constitutively expressing pteridophyte *CHIL* genes in an *A. thaliana chil* mutant background. (A) DMACA-stained seed set by a wild-type (ecotype Col) plant, the *chil* mutant, and transgenic lines constitutively expressing *AcCHIL* or *LoCHIL* in the *chil* mutant background. (B) Relative levels of proanthocyanidins in mature seeds. (C) HPLC chromatograms of methanolic extracts of 10 DAG seedlings of (a) *chil*, (b) *AcCHIL/chil* transgenic plants, (c) *LoCHIL/chil* transgenic plants, and (d) the wild type (Col). (D) Levels of quercetin (Q) and kaempferol (K) in 10 DAG seedlings. Levels of quercetin and kaempferol from wild-type Col were set as a value of 1.0. Data are presented in the form mean ±SD (n=3); *P<0.05 and **P<0.01 versus the Col group, #P<0.05 and ##P<0.01 versus the *chil* group. (This figure is available in color at *JXB* online.)

Discussion

Type I CHIs are present in the pteridophytes

CHIs catalyze the stereospecific isomerization of chalcones into their corresponding (2*S*)-flavanone. Bona fide CHIs are classified into two types: type I CHIs isomerize only naringenin chalcone to naringenin, while type II CHIs can accept either isoliquiritigenin or naringenin chalcone, forming liquiritigenin and naringenin, respectively. We discovered that the early land plants liverworts and *S. moellendorffii* do possess bona fide CHI activity, which could catalyze the stereospecific intramolecular cyclization of both naringenin chalcone and isoliquiritigenin to the corresponding flavanones, which resembles the function of the 'legume-specific' type II CHIs (Cheng *et al.*, 2018). Vascular plants arose ~410 million years ago and subsequently diverged into several lineages, of which only two have survived to the present: the euphyllophytes (pteridophytes and spermatophytes) and the lycophytes (Kenrick and Crane, 1997; Banks *et al.*, 2011). The currently extant lycophytes comprise the Lycopodiales (club mosses), the Isoetales (quillworts), and the Selaginellales (spike mosses) (Banks, 2009). *Selaginella* is the sole genus in the spike moss family (Selaginellaceae) of lycophytes, which belong to the more ancient tracheophyte lineage. The pteridophytes arose early in the Carboniferous period, ~360 million years ago (Pryer *et al.*, 2001). In the present investigation, we characterized bona fide CHIs produced by pteridophyte species lying at the major evolutionary nodes. The indications are that in pteridophyte CHIs, the residues postulated to determine the substrate preference of type II CHIs (i.e. Thr190 and Met191: see Forkmann and Dangelmayr, 1980) have been replaced by,

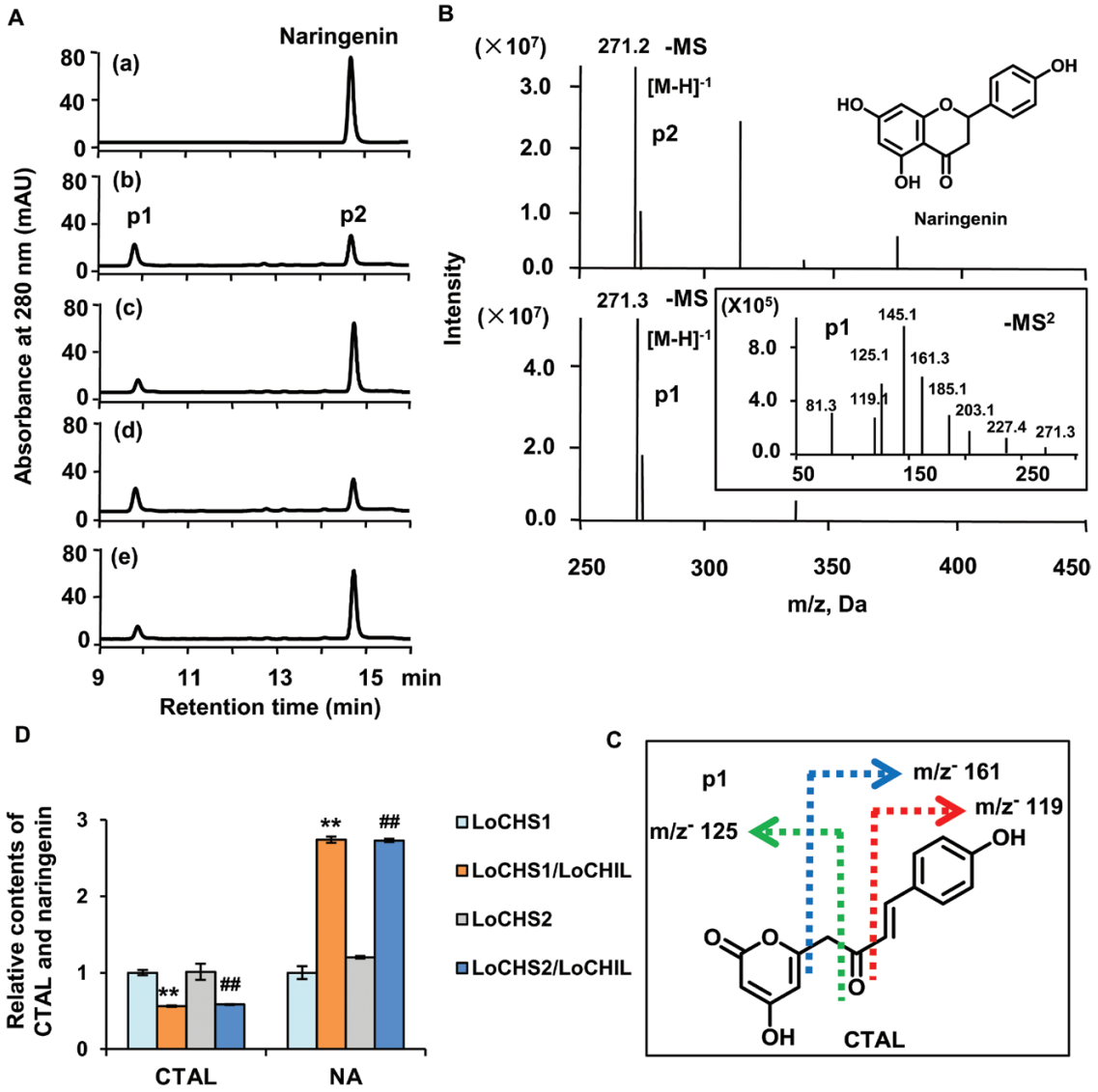

Fig. 8. Effects of LoCHIL on the production of naringenin and CTAL *in vitro*. (A) HPLC chromatograms showing the effect of LoCHIL on the enzymatic activity of LoCHS *in vitro*: (a) naringenin standard. (b–e) HPLC profiles of the reaction products of (b) LoCHS1, (c) LoCHS1/LoCHIL, (d) LoCHS2, and (e) LoCHS2/LoCHIL. (B) MS profiles of p1 (CTAL) and p2 (naringenin). (C) The putative splitting of p1 on MS^2. (D) Relative contents of CTAL and naringenin produced by the *in vitro* reactions. Data are presented in the form mean ±SD (n=3); $^{**}P<0.01$ versus the LoCHS1 group, $^{\#\#}P<0.01$ versus the LoCHS2 group. (This figure is available in color at *JXB* online.)

respectively, either Thr or Ser and Ile. Most of the pteridophyte CHIs exhibited type I CHI activity, although a low level of type II activity was displayed by a small number of entries. Sequence alignment and structural modeling suggested that these type II CHIs harbor a Val/Phe substitution in their active center, since a site-directed mutagenesis experiment was able to show that this substitution converted the activity of the CHI produced by *A. capillus-veneris*. Interestingly, the phylogenetic analysis indicated that these five type II CHIs were located at the base of the fern CHI clade (Fig. 2). In fact, in a previous phylogenetic analysis that reflects the fern evolutionary history, the species harboring these type II CHIs were included in the same branch as the ferns from Polypodiales (Shen *et al.*, 2018). The implication is that the evolutionary development of the CHIs in the pteridophytes differed somewhat from what has been postulated to have occurred in fern evolution. The key amino acid residues determining the type I and type II CHIs are Thr and Ile in most ferns, which probably represent an intermediate form between the Thr and Met of type II CHIs and the Ser and Ile of type I CHIs. The data have unambiguously demonstrated that type II CHIs appeared in the most ancient land plants, liverworts and spike moss, while in the pteridophytes, type I CHIs with a more constrained requirement for substrate tended to supplant the type II CHIs as evolution proceeded. *Selaginella* CHI may represent an evolutionary intermediate in the transition from type II to type I enzymes; it may well be an ancestor of the higher plant type I CHIs.

Type IV CHIs enhance CHS activity in the pteridophytes

In recent years, the genetic and biochemical roles of type IV CHIs (also called CHILs) have been extensively investigated in plants. Although no catalytic activity has been shown

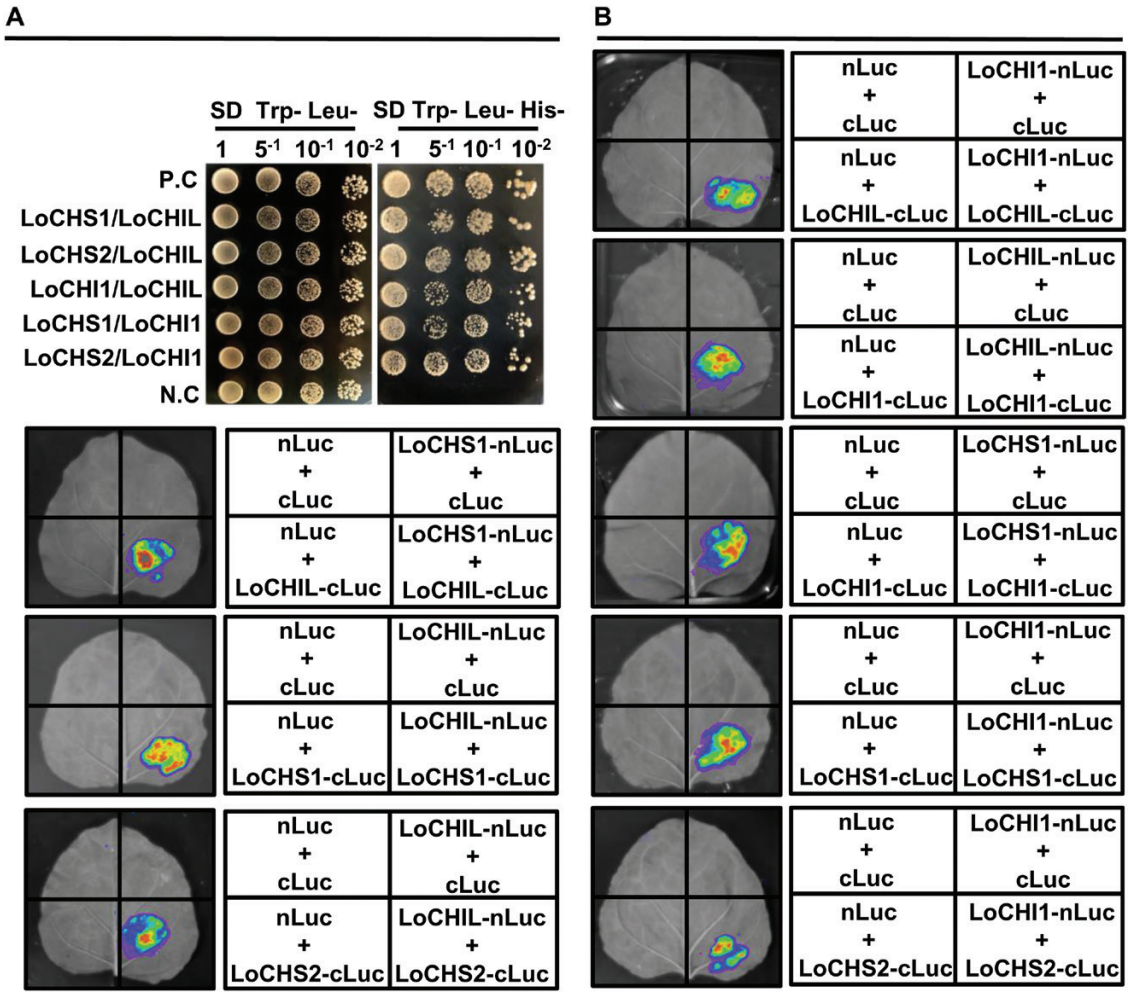

Fig. 9. Interaction between LoCHIL, LoCHSs, and LoCHI1. (A) Yeast two-hybrid experiment demonstrating the protein–protein interaction between LoCHIL, LoCHS1, LoCHS2, and LoCHI1. P.C. refers to the positive control; N.C. refers to the negative control. (B) Luciferase complementation imaging assay was used for analyzing the protein–protein interaction of LoCHIL, LoCHS1, LoCHS2, and LoCHI1. *Agrobacterium tumefaciens* carrying different plasmids were infiltrated into *Nicotiana benthamiana* leaves. Images were recorded at 72 h after infiltration. (This figure is available in color at *JXB* online.)

by recombinant CHILs (Ngaki *et al.*, 2012), their presence enhances flavonoid and anthocyanin production in both *A. thaliana* (Jiang *et al.*, 2015) and *Petunia hybrida* (Morita *et al.*, 2014). Previous investigation demonstrated that CHIL could interact with CHS and CHI to increase the production of flavonoids in plants (Jiang *et al.*, 2015; Ban *et al.*, 2018). In the current investigation, two functional CHSs with different catalytic efficiencies were characterized from *L. orbiculata*. CHS catalyzes the iterative condensation of one *p*-coumaroyl-CoA molecule and three malonyl-CoA molecules to form a tetraketide intermediate and then cyclizes the resulting tetraketide intermediate via an intramolecular C6–C1 Claisen condensation reaction to form naringenin chalcone (Abe and Morita, 2010). Naringenin chalcone was easily transformed to naringenin through spontaneous cyclization. The compound CTAL is a minor byproduct of the C5-oxygen-C1 lactonization from the tetraketide intermediate in the CHS-catalyzed reaction (Yamaguchi *et al.*, 1999). We demonstrated that LoCHIL was able to increase both LoCHS1 and LoCHS2 activity and naringenin formation by limiting the amount of CTAL produced (Fig. 8A), which indicated that CHIL may promote the proper intramolecular cyclization of the linear tetraketide intermediate produced by CHS-catalyzed reactions. Y2H and LCI assays were able to show that LoCHIL could interact physically with CHS and the bona fide type I CHI, LoCHI1 (Fig. 9). In addition, the interaction between LoCHSs and LoCHI1 was detected. Flavonoid synthesis enzymes are thought to co-localize in loosely associated macromolecular complexes (Saslowsky *et al.*, 2001). The interaction between CHS and CHI has been reported in several species including *Glycine max* (Dastmalchi *et al.*, 2016), *A. thaliana* (Saslowsky *et al.*, 2001; Crosby *et al.*, 2011), and *Antirrhinum majus* (Fujino *et al.*, 2018). In the CHI-catalyzed reaction, only the (*2S*)-naringenin isomer can be further converted to downstream products. The interaction and channeling between CHS and CHI was believed to be important to prevent the formation of a mixture of (*2S*)- and (*2R*)-naringenin (Jez *et al.*, 2000). Our results suggested that CHILs may ensure the proper folding and cyclization of polyketide intermediates, which avoids the formation of CTAL during the course of a CHS-catalyzed reaction. At the same time, they may also mediate the interaction between CHS and bona fide CHI to channel the substrate and product, thereby enhancing the production of flavonoids (Fig. 10). *CHIL* loss-of-function

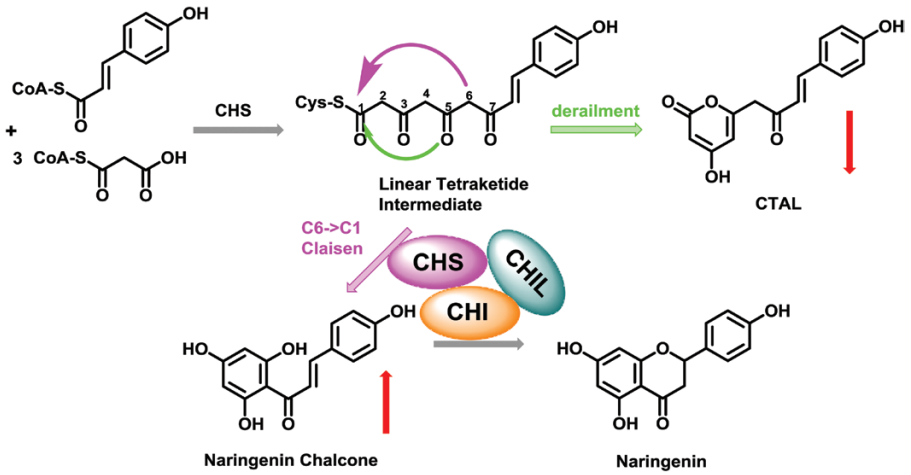

Fig. 10. Proposed model illustrating the interactions of CHIL with CHS or CHI during flavonoid synthesis. CHIL interacts with CHS to enhance the content of naringenin chalcone and to reduce the production of CTAL. (This figure is available in color at *JXB* online.)

mutations are characterized by a strong reduction in the level of both proanthocyanidin and flavonol deposited in the seed (Jiang *et al.*, 2015). The constitutive expression of *LoCHIL* in the *A. thaliana chil* mutant partially rescued the mutant phenotype and enhanced the accumulation of both proanthocyanidin and flavonol. Thus pteridophyte CHILs appear to share a similar function to their *A. thaliana* homologs, further demonstrating that the CHS-enhancing property of type IV CHI-fold proteins has been conserved over the course of plant evolution (Ban *et al.*, 2018).

In conclusion, the evidence supports the notion that type II CHIs evolved into type I CHIs at the time of the emergence of the euphyllophytes. Type IV CHIs may act as an anchor for both CHS and CHI proteins; their function is to promote the proper folding and cyclization of polyketide intermediates formed during a CHS-catalyzed reaction, thereby avoiding the formation of derailment products; in so doing, they act to support the accumulation of flavonoids.

Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. *In vitro* enzyme activity of the 22 pteridophyte CHIs, based on isoliquiritigenin as substrate.

Fig. S2. *In vitro* enzyme activity of the 22 pteridophyte CHIs, based on naringenin chalcone as substrate.

Fig. S3. Sequence alignment of CHI polypeptides from six pteridophytes with CHIs from other plants.

Fig. S4. Homology modeling of AcCHI1 and AcCHI1-F188V using MsCHI (PDB:1EYQ) as template.

Fig. S5. Homology modeling of AcCHI1 and AcCHI1-F188V using MsCHI (PDB:1FM7) as template.

Fig. S6. Sequence alignment of CHIL from two pteridophyte species with AtCHIL (NP_850770).

Fig. S7. *In vitro* enzymatic assays of the pteridophyte recombinant proteins AcCHIL and LoCHIL.

Fig. S8. Sequence alignment of CHSs from *Lindsaea orbiculata* with MsCHS (AAA02824).

Table S1. The primers used in the study

Table S2. The GenBank accession numbers of the cloned CHI-fold proteins and CHS proteins.

Dataset S1. The amino acid sequences of 56 CHIs from 52 pteridophyte species.

Dataset S2. The amino acid sequences of 45 CHILs from 45 pteridophyte species.

Acknowledgements

We thank Professor Yongzhen Pang from the Institute of Botany of the Chinese Academy of Sciences for providing the *chli* mutant seeds, and *tt5* mutant seeds, and Professor Yue-Hong Yan from Shanghai Chenshan Botanical Garden in China for providing the transcriptomic sequence data of ferns and the fern plants. This work was supported by the National Natural Science Foundation of China (nos 31770330 and 31370330) and Science & Technology Development Plan Project of Shandong (no. 2016GSF121032).

References

Abe I, Morita H. 2010. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. Natural Product Reports **27**, 809–838.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22**, 195–201.

Bade JB, Damm B. 1995. Introduction II *Agrobacterium rhizogenes*, a natural transformation system. In: Potrykus I, Spangenberg G, eds. Gene transfer to plants. Springer Lab Manual. Berlin, Heidelberg: Springer, 30–38.

Ban Z, Qin H, Mitchell AJ, Liu B, Zhang F, Weng JK, Dixon RA, Wang G. 2018. Noncatalytic chalcone isomerase-fold proteins in *Humulus lupulus* are auxiliary components in prenylated flavonoid biosynthesis. Proceedings of the National Academy of Sciences, USA **115**, 5223–5232.

Banks JA. 2009. Selaginella and 400 million years of separation. Annual Review of Plant Biology **60**, 223–238.

Banks JA, Nishiyama T, Hasebe M, *et al*. 2011. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science **332**, 960–963.

Beuerle T, Pichersky E. 2002. Enzymatic synthesis and purification of aromatic coenzyme A esters. Analytical Biochemistry **302**, 305–312.

Chen H, Zou Y, Shang Y, Lin H, Wang Y, Cai R, Tang X, Zhou JM. 2008. Firefly luciferase complementation imaging assay for protein–protein interactions in plants. Plant Physiology **146**, 368–376.

Cheng AX, Zhang X, Han XJ, Zhang YY, Gao S, Liu CJ, Lou HX. 2018. Identification of chalcone isomerase in the basal land plants reveals an

ancient evolution of enzymatic cyclization activity for synthesis of flavonoids. New Phytologist **217**, 909–924.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. The Plant Journal **16**, 735–743.

Crosby KC, Pietraszewska-Bogiel A, Gadella TW Jr, Winkel BS. 2011. Förster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes. FEBS Letters **585**, 2193–2198.

Curtis MD, Grossniklaus U. 2003. A gateway cloning vector set for high-throughput functional analysis of genes *in planta*. Plant Physiology **133**, 462–469.

Dastmalchi M, Bernards MA, Dhaubhadel S. 2016. Twin anchors of the soybean isoflavonoid metabolon: evidence for tethering of the complex to the endoplasmic reticulum by IFS and C4H. The Plant Journal **85**, 689–706.

Duff RJ, Nickrent DL. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. American Journal of Botany **86**, 372–386.

Falcone Ferreyra ML, Casas MI, Questa JI, Herrera AL, Deblasio S, Wang J, Jackson D, Grotewold E, Casati P. 2012. Evolution and expression of tandem duplicated maize flavonol synthase genes. Frontiers in Plant Science **3**, 101.

Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP. 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nature Structural and Molecular Biology **6**, 775–784.

Focks N, Sagasser M, Weisshaar B, Benning C. 1999. Characterization of *tt15*, a novel *transparent testa* mutant of *Arabidopsis thaliana* (L.) Heynh. Planta **208**, 352–357.

Forkmann G, Dangelmayr B. 1980. Genetic control of chalcone isomerase activity in flowers of *Dianthus caryophyllus*. Biochemical Genetics **18**, 519–527.

Fujino N, Tenma N, Waki T, *et al*. 2018. Physical interactions among flavonoid enzymes in snapdragon and torenia reveal the diversity in the flavonoid metabolon organization of different plant species. The Plant Journal **94**, 372–392.

Gambino G, Perrone I, Gribaudo I. 2008. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. Phytochemical Analysis: **19**, 520–525.

Gao S, Yu HN, Xu RX, Cheng AX, Lou HX. 2015. Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum*. Phytochemistry **111**, 48–58.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods in Enzymology **350**, 87–96.

Hassan S, Mathesius U. 2012. The role of flavonoids in root–rhizosphere signalling: opportunities and challenges for improving plant–microbe interactions. Journal of Experimental Botany **63**, 3429–3444.

Jez JM, Bowman ME, Dixon RA, Noel JP. 2000. Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. Nature Structural and Molecular Biology **7**, 786–791.

Jez JM, Noel JP. 2002. Reaction mechanism of chalcone isomerase. pH dependence, diffusion control, and product binding differences. Journal of Biological Chemistry **277**, 1361–1369.

Jiang C, Schommer CK, Kim SY, Suh DY. 2006. Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. Phytochemistry **67**, 2531–2540.

Jiang W, Yin Q, Wu R, Zheng G, Liu J, Dixon RA, Pang Y. 2015. Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. Journal of Experimental Botany **66**, 7165–7179.

Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. Nature **389**, 33–39.

Kozłowska A, Szostak-Wegierek D. 2014. Flavonoids—food sources and health benefits. Roczniki Panstwowego Zakladu Higieny **65**, 79–85.

Ma LQ, Pang XB, Shen HY, Pu GB, Wang HH, Lei CY, Wang H, Li GF, Liu BY, Ye HC. 2009. A novel type III polyketide synthase encoded by a three-intron gene from *Polygonum cuspidatum*. Planta **229**, 457–469.

Morita Y, Takagi K, Fukuchi-Mizutani M, Ishiguro K, Tanaka Y, Nitasaka E, Nakayama M, Saito N, Kagami T, Hoshino A. 2014. A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. The Plant Journal **78**, 294–304.

Ngaki MN, Louie GV, Philippe RN, Gerard M, Florence P, Bowman ME, Ling L, Elise L, Eve Syrkin W, Noel JP. 2012. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. Nature **485**, 530–533.

Pang Y, Peel GJ, Wright E, Wang Z, Dixon RA. 2007. Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. Plant Physiology **145**, 601–615.

Park SH, Lee CW, Cho SM, Lee H, Park H, Lee J, Lee JH. 2018. Crystal structure and enzymatic properties of chalcone isomerase from the Antarctic vascular plant *Deschampsia antarctica* Desv. PLoS One **13**, e0192415.

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf PG, Hunt JS, Sipes SD. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature **409**, 618–622.

Ralston L, Subramanian S, Matsuno M, Yu O. 2005. Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. Plant Physiology **137**, 1375–1388.

Saslowsky D, Winkel-Shirley B. 2001. Localization of flavonoid enzymes in Arabidopsis roots. The Plant Journal **27**, 37–48.

Shen H, Jin D, Shu JP, *et al*. 2018. Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. Gigascience **7**, 1–11.

Shimada N, Aoki T, Sato S, Nakamura Y, Tabata S, Ayabe S. 2003. A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*. Plant Physiology **131**, 941–951.

Sun Y, Wu Y, Zhao Y, Han X, Lou H, Cheng A. 2013. Molecular cloning and biochemical characterization of two cinnamyl alcohol dehydrogenases from a liverwort *Plagiochasma appendiculatum*. Plant Physiology and Biochemistry **70**, 133–141.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24**, 1596–1599.

Weng JK, Noel JP. 2013. Chemodiversity in *Selaginella*: a reference system for parallel and convergent metabolic evolution in terrestrial plants. Frontiers in Plant Science **4**, 119.

Yamaguchi T, Kurosaki F, Suh DY, Sankawa U, Nishioka M, Akiyama T, Shibuya M, Ebizuka Y. 1999. Cross-reaction of chalcone synthase and stilbene synthase overexpressed in *Escherichia coli*. FEBS Letters **460**, 457–461.

Yu HN, Liu XY, Gao S, Sun B, Zheng HB, Ji M, Cheng AX, Lou HX. 2018. Structural and biochemical characterization of the plant type III polyketide synthases of the liverwort *Marchantia paleacea*. Plant Physiology and Biochemistry **125**, 95–105.

Zhang X, Gou M, Liu CJ. 2013. *Arabidopsis* Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. The Plant Cell **25**, 4994–5010.