



Versatility in acyltransferase activity completes chicoric acid biosynthesis in purple coneflower

Rao Fu ^{1,5}, Pingyu Zhang^{1,5}, Ge Jin¹, Lianglei Wang¹, Shiqian Qi², Yang Cao³, Cathie Martin ⁴ & Yang Zhang ¹✉

Purple coneflower (*Echinacea purpurea* (L.) Moench) is a popular native North American herbal plant. Its major bioactive compound, chicoric acid, is reported to have various potential physiological functions, but little is known about its biosynthesis. Here, taking an activity-guided approach, we identify two cytosolic BAHD acyltransferases that form two intermediates, caftaric acid and chlorogenic acid. Surprisingly, a unique serine carboxypeptidase-like acyltransferase uses chlorogenic acid as its acyl donor and caftaric acid as its acyl acceptor to produce chicoric acid in vacuoles, which has evolved its acyl donor specificity from the better-known 1-O- β -D-glucose esters typical for this specific type of acyltransferase to chlorogenic acid. This unusual pathway seems unique to *Echinacea* species suggesting convergent evolution of chicoric acid biosynthesis. Using these identified acyltransferases, we have reconstituted chicoric acid biosynthesis in tobacco. Our results emphasize the flexibility of acyltransferases and their roles in the evolution of specialized metabolism in plants.

¹Key Laboratory of Bio-resource and Eco-environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610064, China.

²Department of Urology, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, and Collaborative Innovation Center for Biotherapy, Chengdu 610041, China. ³Center of Growth, Metabolism and Aging, College of Life Sciences, Sichuan University, Chengdu 610064, China. ⁴Department of Metabolic Biology and Biological Chemistry, John Innes Centre, Norwich NR4 7UH, UK. ⁵These authors contributed equally: Rao Fu, Pingyu Zhang.

✉email: yang.zhang@scu.edu.cn

The discovery of America by Columbus changed the world in many ways, particularly through the increased exchange of valuable natural resources. One example was the introduction of *Echinacea*, a genus of herbs native to North America, and used by Native Americans as traditional medicines to treat common colds and infections¹. Since the 1880s, *Echinacea* extracts have become some of the most popular commercial herbal preparations, and today over 1000 products containing extracts of *Echinacea* are listed in the Dietary Supplement Label Database (<https://dslod.nih.gov>). The total sales of products from *E. purpurea*, *E. angustifolia*, and *E. pallida*, reached US\$ 120.2 million on the US market alone in 2019 and grew by 90.9% in the first half of 2020, partially due to the COVID-19 pandemic². The main bioactives in *Echinacea* are caffeic acid derivatives (CADs)³. Chicoric acid is the principal CAD and provides an index for the quality of raw material and commercial preparations of *Echinacea*⁴. Although there isn't sufficient evidence to conclusively establish its functions, chicoric acid has been reported as a potential HIV-1 integrase inhibitor⁵ along with other bioactivities including antiviral, anti-inflammatory, glucose and lipid homeostatic, neuroprotective, anti-atherosclerosis, and anti-aging activities⁶. Despite the large number of studies in the potential health benefits of chicoric acid, its biosynthetic pathway remains unclear.

Enzymes that use caffeoyl CoA and meso-tartaric acid to generate mono-*O*-caffeoyl-meso-tartrate and add an additional caffeoyl group from caffeoyl CoA to form di-*O*-caffeoyl-meso-tartrate were reported in *Equisetum arvense* in 1996, but without identification of the encoding genes⁷. In *Echinacea*, caftaric acid and chicoric acid are L-isomers, which differ from those in *Equisetum arvense* suggesting a distinct biosynthetic pathway⁸. A hydroxycinnamoyl-CoA: tartaric acid hydroxycinnamoyl transferase (HTT) activity synthesizing L-caftaric acid has been reported in *Arachis glabrata*, but without gene information⁹. However, no enzyme activity has yet been described for L-chicoric acid biosynthesis. One of the best-studied CADs is chlorogenic acid. Hydroxycinnamoyl-Coenzyme A: quinate hydroxycinnamoyl transferase (HQT), a BAHD acyltransferase family member, is involved in the biosynthesis of chlorogenic acid in different plant species^{10–12}. Another important acyltransferase group is the serine carboxypeptidase-like (SCPL) family. The major differences between these two families are the acyl donor source and cellular compartment: BAHDs use acyl-CoA thioesters as acyl donors in the cytosol, whereas SCPLs use 1-*O*- β -glucose esters as their acyl donors in the vacuole¹³.

Here we report the elucidation of the chicoric acid biosynthesis pathway in purple coneflower. This pathway involves both BAHD and SCPL acyltransferases. Two BAHD enzymes, EpHTT and EpHQT are responsible for the biosynthesis of caftaric acid and chlorogenic acid in the cytosol, respectively. These two CADs serve as substrates for the generation of chicoric acid in the vacuole by a unique SCPL enzyme. Unlike other SCPLs, this SCPL prefers chlorogenic acid to 1-*O*-caffeoyl β -D-glucose as its acyl donor. Among the species we checked, this chicoric acid biosynthetic route is unique to *Echinacea* spp. and it is likely that chicoric acid biosynthesis in other species evolved by convergent catalytic mechanisms. Our results expand the understanding of acyltransferases and highlight the rapid evolution possible with these types of enzymes.

Results

EpBAHDs catalyze the synthesis of caftaric acid and chlorogenic acid. Previous studies indicated that purple coneflower (*Echinacea purpurea*) contains much higher levels of chicoric acid than other *Echinacea* species¹⁴. Metabolic profiling of purple

coneflower seedlings revealed the presence of chicoric acid throughout the plant, making it difficult to conduct tissue-specific coexpression analysis to identify biosynthetic genes (Fig. 1)^{15–17}. We generated hairy root cultures of purple coneflower and confirmed the presence of chicoric acid (Fig. 1). RNA-seq of hairy roots treated with methyl jasmonate to induce chicoric acid accumulation failed to provide good candidate genes due to the large number of differently expressed genes. As an alternative, we took a traditional biochemical approach and searched for caftaric acid synthase activity as caftaric acid is a likely intermediate (Fig. 1d)^{12,18}. Based on the structural similarity of caftaric acid to chlorogenic acid, we investigated whether hydroxycinnamoyl-CoA: tartaric acid hydroxycinnamoyl transferase (HTT) could use caffeoyl CoA as an acyl donor to produce caftaric acid as reported⁹. When tartaric acid and caffeoyl CoA were incubated with a crude protein extract from hairy roots of purple coneflower, we detected the production of caftaric acid, indicating the presence of HTT activity. In addition, the crude protein extract also catalyzed the generation of chlorogenic acid using quinic acid and caffeoyl CoA as substrates, confirming the activity of hydroxycinnamoyl-Coenzyme A: quinate hydroxycinnamoyl transferase (HQT).

We purified HTT activity by ammonium sulfate precipitation and ion-exchange chromatography and found one candidate protein annotated as HQT-like in the fraction with high HTT activity through comparison of peptide mass fingering data to hairy root RNA-seq data (Supplementary Fig. 1a–b). This protein belongs to the BAHD acyltransferase family, which utilize CoA thioesters as acyl donors: BAHD enzymes have two conserved amino acid motifs HXXXD and DFGWG^{11,13}. This HQT-like protein was closely related to the HCT (Hydroxycinnamoyl-Coenzyme A: quinate/shikimate hydroxycinnamoyl transferase) clade, with HHLVD and DFGWG motifs (Fig. 2a and Supplementary Fig. 2). In vitro enzyme assays using recombinant protein from *E. coli* confirmed that this enzyme can catalyze the biosynthesis of caftaric acid (Fig. 2b and Supplementary Fig. 3a and 4a–c), and we named it EpHTT. Similarly, in fractions with high HQT activity, we identified another BAHD acyltransferase with classic HTLSD and DFGWG motifs belonging to the HQT clade of BAHD acyltransferases (Fig. 2a, Supplementary Fig. 1a–b and 2), which could catalyze the formation of chlorogenic acid in vitro, and which we named EpHQT (Fig. 2c and Supplementary Fig. 3b and 4d–f). No putative signal peptides were found in the EpHTT or EpHQT sequences. Both EpHTT and EpHQT GFP-tagged proteins were cytosolically localized as reported for other HQTs^{19,20} and showed highest activities at pH 7 (Fig. 2h and Supplementary Fig. 5a–b). Although both enzymes showed promiscuity in their acyl donors, caffeoyl CoA was preferred by both (Supplementary Table 1–2). EpHTT differs from HTT activity from perennial peanut which prefers *p*-coumaroyl CoA as its acyl donor⁹. We verified the function of EpHTT and EpHQT in vivo using both overexpression and gene silencing in hairy roots. When the EpHTT gene was overexpressed or silenced, the caftaric acid levels in transgenic hairy roots increased or decreased correspondingly (Fig. 2d, e). Similar results were seen for the EpHQT gene, where the contents of chlorogenic acid in the transgenic hairy roots were correlated positively in the different lines to the expression levels of EpHQT (Fig. 2f, g). Our in vitro and in vivo results confirmed that EpHTT and EpHQT are responsible for caftaric acid and chlorogenic acid biosynthesis, respectively in purple coneflower.

Substrate specificities of EpBAHDs. Phylogenetic analysis indicated that EpHTT might have evolved from the hydroxycinnamoyl-coenzyme A: shikimate/quinic hydroxycinnamoyl transferase

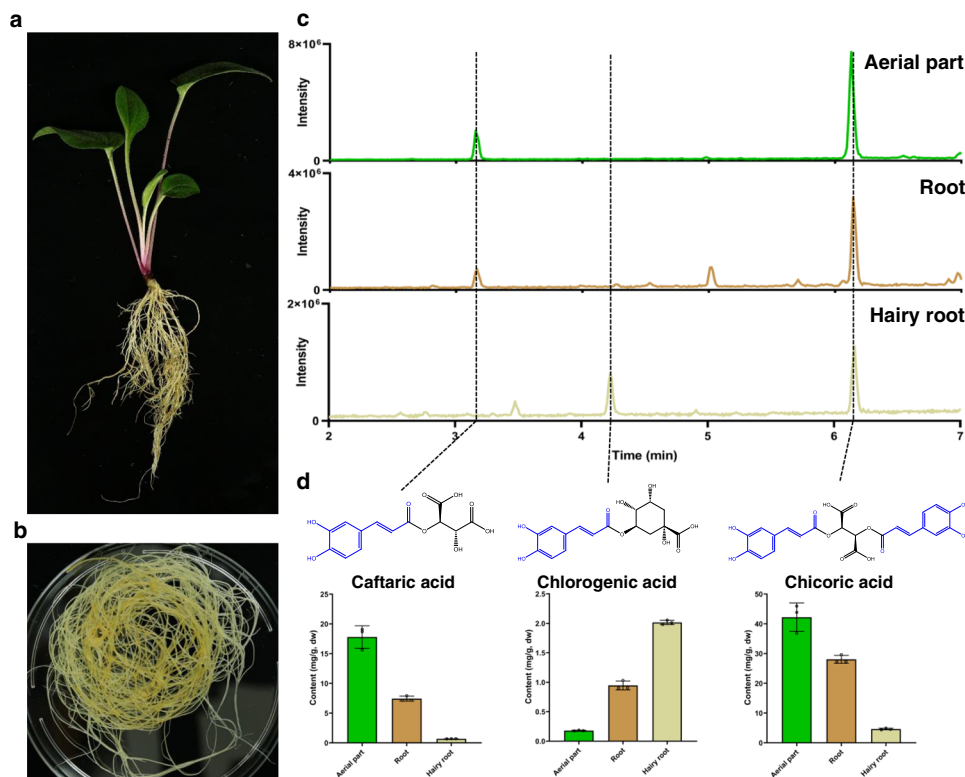


Fig. 1 Main caffeic acid derivatives in purple coneflower. **a** Two-month-old seedlings and **b** hairy root cultures of purple coneflower. **c** Total ion chromatograms (TIC) of aerial parts, roots, and hairy root cultures. **d** Main caffeic acid derivatives (CADs) and their abundance. Data are mean \pm s.d. ($n = 3$ biologically independent samples). Source data underlying Fig. 1d are provided as a Source Data file.

(HCT) clade of BAHD acyltransferases (Fig. 2a). Using AtHCT (AED95744.1) and NtHCT (CAE46932.1) to screen RNA-seq data, we also identified a candidate transcript encoding EpHCT. EpHCT belongs to the HCT clade with conserved HHAAD and DFGWG motifs (Fig. 2a and Supplementary Fig. 2)²¹. We confirmed EpHCT function by in vitro enzyme assays which could catalyze the production of *p*-coumaroyl shikimic acid using *p*-coumaroyl CoA and shikimic acid as substrates, and the generation of 5-*O*-caffeoyl shikimic acid using caffeoyl CoA and shikimic acid as substrates (Supplementary Fig. 6). All three BAHD enzymes, EpHTT, EpHQT and EpHCT, showed preferences amongst their potential acyl acceptors (Table 1). EpHTT could use only tartaric acid but not quinic acid nor shikimic acid, EpHQT and EpHCT could use both quinic acid and shikimic acid as acyl acceptors but with significant preferences by EpHQT for quinic acid and by EpHCT for shikimic acid as previously reported (Supplementary Fig. 7)^{12,21,22}. Molecular docking was undertaken based on the structural model for *p*-coumaroyl shikimic acid bound AtHCT (PDB: 5KJU)²³. All three EpBAHDs have similar scaffolds. Compared with EpHQT and EpHCT, EpHTT formed a smaller pocket, that could accommodate only tartaric acid, and not quinic acid and shikimic acid (Supplementary Fig. 8a). Differences between the active pockets of EpHQT and EpHCT appeared critical in determining their acyl acceptor preferences (Supplementary Fig. 8b). In summary, EpHTT has likely evolved from the HCT clade of enzymes, principally through specialization of its acyl acceptor recognition.

Chicoric acid is synthesized from caftaric acid and chlorogenic acid. As suggested by the formation of di-caffeoyl quinic acid in tomato and di-*O*-caffeoyl-*meso*-tartrate in *Equisetum arvense*^{7,24}, we tried to establish a similar in vitro chicoric acid biosynthesis assay. However, neither EpHTT nor crude protein extract could convert caftaric acid or caftaric acid with caffeoyl CoA to chicoric

acid (Supplementary Fig. 9a, b). We noticed that decreases in *EpHQT* expression also led to the decline of chicoric acid content in transgenic hairy roots (Supplementary Fig. 9c). Chlorogenic acid has been reported as an acyl donor for the biosynthesis of di-caffeoyl quinic acid by SLHQT, as well as caffeoylglucarate by a GDSL lipase-like acyltransferase in tomato^{24,25}. This suggested that chlorogenic acid might be used as acyl donor in chicoric acid biosynthesis. Consequently, we incubated caftaric acid and chlorogenic acid with a crude protein extract from purple coneflower and detected the formation of chicoric acid. We optimized this assay using acidic and low salt conditions (PBS pH 4.0 without NaCl or KCl) (Supplementary Fig. 9b). Together, these data suggested that chicoric acid can be synthesized from caftaric acid and chlorogenic acid by a chicoric acid synthase in purple coneflower.

A SCPL enzyme catalyzes the biosynthesis of chicoric acid. The activity of chicoric acid synthase was measured and fractions with the highest activities were used for peptide mass fingerprinting (Supplementary Fig. 1c, d). However, peptides of neither BAHDs nor GDSL lipase-like proteins were detected. Instead, we noticed peptides of a Serine carboxypeptidase-like (SCPL) protein were enriched in the fractions (Supplementary Fig. 1e). This protein belongs to the SCPL clade IA which has been reported to have acyltransferase activity (Fig. 3a)²⁶. In vitro assays indicated that this protein, when expressed and purified from yeast, could catalyze the trans-acylation of caftaric acid by chlorogenic acid to form chicoric acid and quinic acid (Fig. 3b and Supplementary Fig. 10). Therefore, we named this SCPL protein, chicoric acid synthase (EpCAS). Compared to BAHD acyltransferases, EpCAS required longer incubation times for accumulation of its chicoric acid product to become apparent, which is typical of other SCPLs^{27,28}. We determined the kinetic parameters of EpCAS

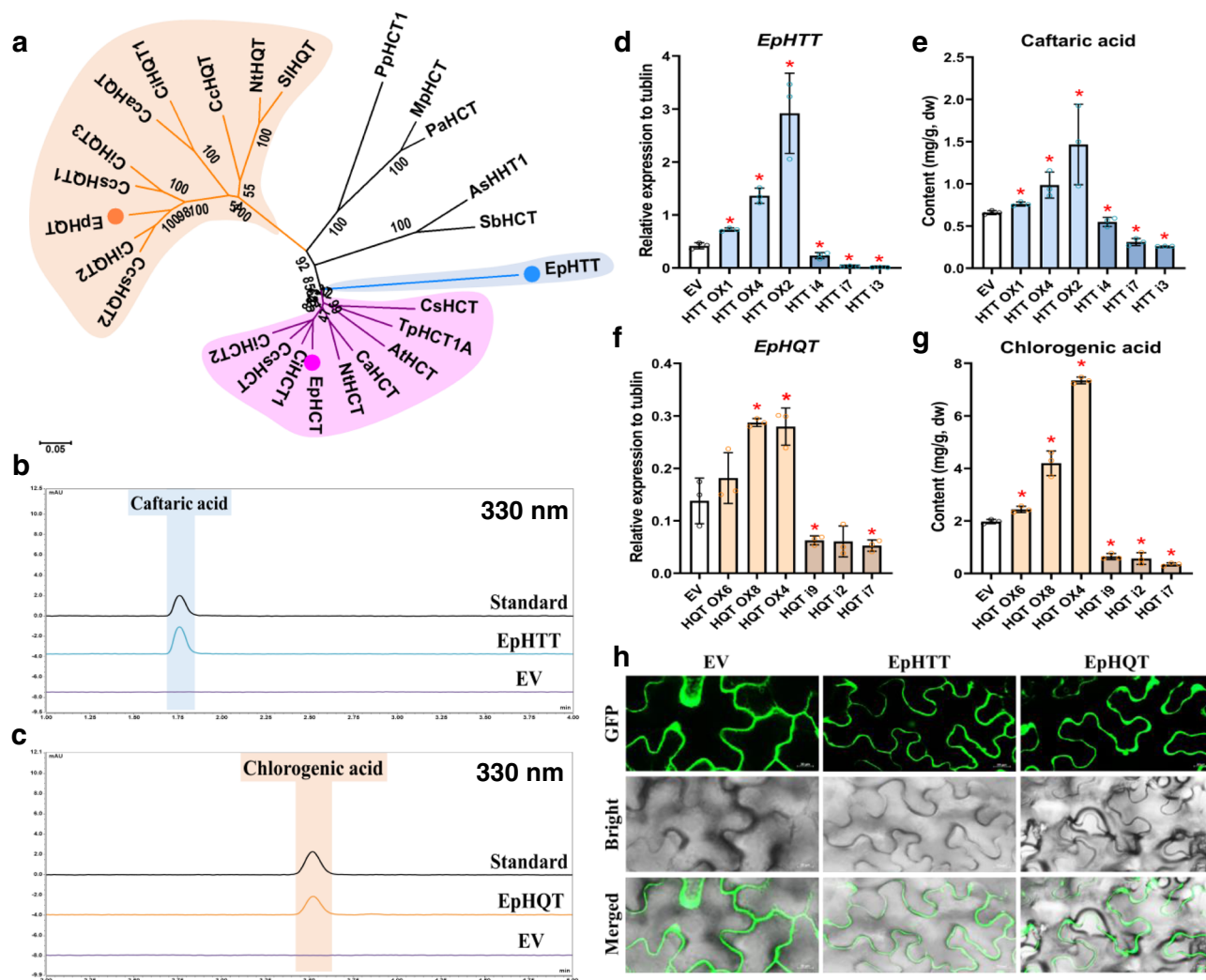


Fig. 2 Two cytosolic BAHD acyltransferases catalyze the synthesis of caftaric acid and chlorogenic acid. **a** Phylogenetic analysis of EpHHT, EpHQT, and EpHCT with functionally identified BAHDs. GeneBank ID of the proteins used in the tree are listed in methods. The HQT and HCT clades are highlighted in orange and pink, respectively, and EpHHT is highlighted in blue. UPLC detection of caftaric acid and chlorogenic acid production by EpHHT (**b**) and EpHQT (**c**), respectively. **d** Relative expression levels of *EpHHT* and **e** caftaric acid contents in transgenic overexpression (OX) and RNAi (i) hairy root lines. **f** Relative expression levels of *EpHQT* and **g** chlorogenic acid contents in transgenic hairy root lines. Data are mean \pm s.d. ($n = 3$ biologically independent samples). * indicates significant difference from empty vector (EV) control line ($P < 0.05$) analyzed by two-sided Student's *t*-test. * $P = 0.0013$ (HTT OX1), * $P = 0.0005$ (HTT OX4), * $P = 0.0047$ (HTT OX2), * $P = 0.0163$ (HTT i4), * $P = 0.0003$ (HTT i7), * $P = 0.0002$ (HTT i3) in **d**; * $P = 0.0072$ (HTT OX1), * $P = 0.0230$ (HTT OX4), * $P = 0.0430$ (HTT OX2), * $P = 0.0302$ (HTT i4), * $P = 0.0002$ (HTT i7), * $P = 0.000008$ (HTT i3) in **e**; * $P = 0.0042$ (HQT OX8), * $P = 0.0118$ (HQT OX4), * $P = 0.0420$ (HQT i9), * $P = 0.0298$ (HQT i7) in **f**; * $P = 0.0053$ (HQT OX6), * $P = 0.0013$ (HQT OX8), * $P = 0.0000004$ (HQT OX4), * $P = 0.0001$ (HQT i9), * $P = 0.0005$ (HQT i2), * $P = 0.000009$ (HQT i7) in **g**. **h** Cytosolic localization of EpHHT and EpHQT in *Nicotiana benthamiana* leaves. Scale bars show 20 μm . This experiment was repeated independently three times with similar results. Source data underlying d-g are provided as a Source Data file.

using *in vitro* enzyme assays and found that EpCAS showed quite low *K_{cat}* values for its caftaric acid and chlorogenic acid substrates (Supplementary Table 3). Similar to other SCPLs^{29,30}, EpCAS has a conserved vacuolar targeting peptide at its N-terminus (Met1-His24) (Supplementary Fig. 11). We confirmed that an EpCAS GFP-fusion protein was located in vacuoles, and the enzyme shows highest activity at pH 4 consistent with its vacuolar location (Fig. 3e and Supplementary Fig. 5c). We verified the role of EpCAS in chicoric acid biosynthesis *in vivo* using hairy roots. When *EpCAS* was overexpressed or silenced, the chicoric acid content was increased or decreased respectively, consistent with *EpCAS* expression levels which were accordingly upregulated or downregulated (Fig. 3c, d). These results showed that EpCAS catalyzes the synthesis of chicoric acid in purple coneflower.

When the EpCAS was aligned with other functionally characterized SCPL acyltransferases, we found it had the conserved Ser-His-Asp catalytic triad (Supplementary Fig. 12), which is necessary for hydrolysis of peptide bonds by serine proteinase and can be inhibited by phenylmethylsulfonyl fluoride (PMSF)^{31,32}. Indeed, the *in vitro* activity of EpCAS was inhibited by PMSF in a dose-dependent manner (Fig. 4a). The significance of the conserved Ser-His-Asp catalytic triad of EpCAS was also confirmed using site-directed mutagenesis. When the Ser180 was replaced by Ala, EpCAS lost acyltransferase activity completely. Mutations converting His430 and Asp374 to alanine also showed significantly lower activities (Fig. 4b). These results demonstrated that the catalytic triad is crucial for the acyl transfer activity³³. Molecular docking based on the structure of the yeast serine

Table 1 EpBAHD substrate preferences.

Enzyme	Acyl donors	Acyl acceptors		
		Tartaric acid	Quinic acid	Shikimic acid
EpHTT	Caffeoyl CoA	+	n.d.	n.d.
	<i>p</i> -Coumaroyl CoA	+	n.d.	n.d.
	Feruloyl CoA	+	n.d.	n.d.
EpHQT	Caffeoyl CoA	n.d.	+	+
	<i>p</i> -Coumaroyl CoA	n.d.	+	+
	Feruloyl CoA	n.d.	+	n.d.
EpHCT	Caffeoyl CoA	n.d.	+	+
	<i>p</i> -Coumaroyl CoA	n.d.	+	+
	Feruloyl CoA	n.d.	n.d.	n.d.

“+” indicates that the reaction can happen. n.d. means no detection of corresponding products under these experimental conditions.

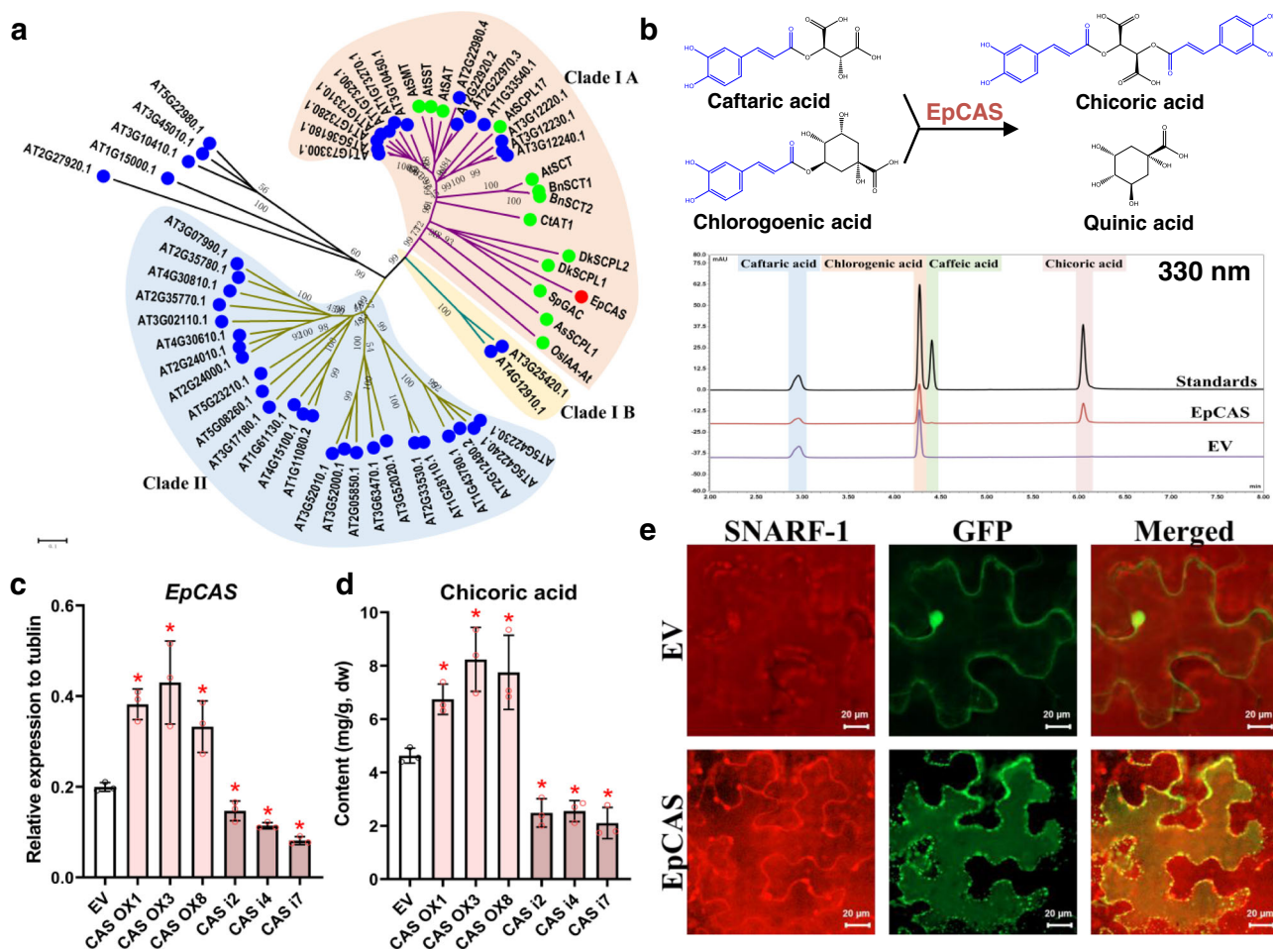


Fig. 3 A SCPL enzyme catalyzes the biosynthesis of chicoric acid. **a** Phylogenetic analysis of EpCAS together with *Arabidopsis thaliana* and several additional functionally characterized SCPLs. EpCAS is marked in red. SCPLs previously characterized functionally are marked in green. **b** UPLC detection of chicoric acid generated by recombinant EpCAS. **c** Relative expression levels of EpCAS and **d** chicoric acid contents in transgenic over expression (OX) and RNAi (i) hairy root lines. Data are mean±s.d. ($n = 3$ biologically independent samples). * indicates a significant difference from the empty vector (EV) control line ($P < 0.05$) analyzed by two-sided Student’s *t*-test. * $P = 0.0008$ (CAS OX1), * $P = 0.0123$ (CAS OX3), * $P = 0.0162$ (CAS OX8), * $P = 0.0187$ (CAS i2), * $P = 0.0002$ (CAS i4), * $P = 0.0001$ (CAS i7) in **c**; * $P = 0.0045$ (CAS OX1), * $P = 0.0071$ (CAS OX3), * $P = 0.0189$ (CAS OX8), * $P = 0.0034$ (CAS i2), * $P = 0.0017$ (CAS i4), * $P = 0.0025$ (CAS i7) in **d**. **e** EpCAS is vacuole-localized in *Nicotiana benthamiana* leaves. The vacuole was visualized by SNARF-1⁵⁵. Scale bars, 20 μm. This experiment was repeated independently three times with similar results. Source data underlying **c** and **d** are provided as a Source Data file.

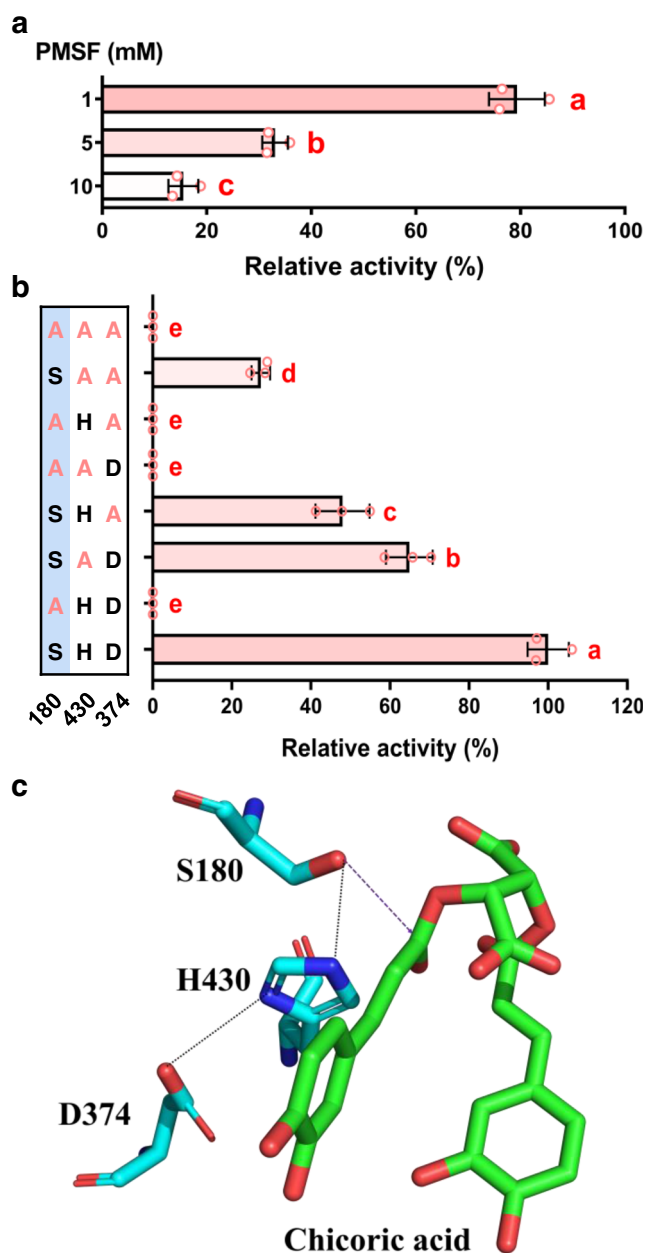


Fig. 4 The Ser-His-Asp catalytic triad is necessary for the acyltransferase activity of EpCAS. **a** EpCAS activity was inhibited by Phenylmethylsulfonyl fluoride (PMSF) in a dose-dependent manner. **b** Site-directed mutagenesis of Ser-His-Asp catalytic triad. Ser180 is necessary for the acyltransferase activity of EpCAS. Data are mean \pm s.d. ($n = 3$ independent experiments). Different letters in **a** and **b** indicate significantly different values at $P < 0.05$ analyzed by one-way ANOVA with Tukey's multiple comparisons test. **c** Molecular docking of Ser-His-Asp catalytic triad with chicoric acid. Source data underlying **a** and **b** are provided as a Source Data file.

carboxypeptidase (PDB: 1YSC) was used to investigate the catalytic mechanism of EpCAS³⁴. The catalytic triad of EpCAS formed the reactive part of the active site, especially Ser180 which could act as a nucleophilic residue and attack the carbonyl C of the chlorogenic acid to form an ester with the caffeic acid intermediate. The ester bond of this complex could be subsequently attacked by caftaric acid and result in the production of chicoric acid (Fig. 4c). Surprisingly, we also detected caffeic acid as a byproduct in the EpCAS reaction (Fig. 3b and Supplementary Fig. 10f-g). An SCPL (1-*O*-sinapoyl- β -D-glucose: L-malate

sinapoyl transferase) from *Arabidopsis* has also been reported to have a minor hydrolytic activity producing sinapic acid from 1-*O*-sinapoyl- β -D-glucose (pH 6.0)³⁵. When EpCAS was mixed with caftaric acid or chlorogenic acid alone (pH 4.0), it cleaved caffeic acid from both these substrates (Supplementary Fig. 13). Another SCPL from *Zea mays*, 1-*O*-(indole-3-acetyl)- β -D-glucose: *myo*-inositol indoleacetyl transferase has also been reported to have hydrolytic activity towards its product³⁶. EpCAS could also hydrolyze chicoric acid to release caftaric acid and caffeic acid, but this reaction was irreversible (Supplementary Fig. 13c-d). It seems that the Ser180 of the catalytic triad can attack the ester bond of these caffeic acid esters directly and release minor amounts of caffeic acid.

The substrate specificity of EpCAS. Since previous studies have indicated that SCPL acyltransferases use 1-*O*- β -D-glucose esters as their acyl donors^{13,37}, we investigated whether EpCAS could also use 1-*O*-caffeoyl- β -D-glucose as its acyl donor. When we incubated 1-*O*-caffeoyl- β -D-glucose (Supplementary Fig. 14) with caftaric acid, very little chicoric acid was detected (Supplementary Fig. 15a-b). Compared to chlorogenic acid, the relative activity of EpCAS in using 1-*O*-caffeoyl- β -D-glucose was reduced to $0.67 \pm 0.06\%$ (Supplementary Fig. 16a). The in vitro kinetics of EpCAS against 1-*O*-caffeoyl- β -D-glucose compared to chlorogenic acid showed that even though they had comparable K_m values, the K_{cat} and K_{cat}/K_m values were much lower for 1-*O*-caffeoyl- β -D-glucose (Supplementary Table 3). Docking results showed that the glucose group of 1-*O*-caffeoyl- β -D-glucose would sink into the inner pocket and hinder the entry of caftaric acid. Although quinic acid and glucose have similar structures, the extra carboxyl group of quinic acid would hinder the entrance of chlorogenic acid to the inner pocket of EpCAS. This difference suggested that quinic acid would be liberated more easily from chlorogenic acid than glucose from 1-*O*-caffeoyl- β -D-glucose, explaining the preference of EpCAS for chlorogenic acid as its acyl donor (Supplementary Fig. 15c). Interestingly, we found that compounds with structures similar to chlorogenic acid (neochlorogenic acid, cryptochlorogenic acid and 5-*O*-caffeoyl shikimic acid) could also be used by EpCAS but with much lower relative activities ($<1\%$) than with chlorogenic acid under identical conditions (Supplementary Fig. 16). These results demonstrated that EpCAS has evolved unique specificity for its acyl donor, distinct from that of other characterized SCPL enzymes in plants^{33,38}. Interestingly, although we could detect very low activity of purified EpCAS enzyme when using 1-*O*-caffeoyl- β -D-glucose as an acyl donor in vitro, crude protein extracts of purple coneflower could not use 1-*O*-caffeoyl- β -D-glucose as an acyl donor to synthesize caftaric acid, chlorogenic acid or chicoric acid (Supplementary Fig. 17). In addition, we failed to detect any 1-*O*-caffeoyl- β -D-glucose in extracts of purple coneflower (Supplementary Fig. 18). All these results demonstrated that in purple coneflower, EpCAS has evolved specificity for its acyl donor.

Chicoric acid biosynthesis is polyphyletic in the plant kingdom. So far, over sixty genera and species have been found to contain chicoric acid³⁹. To investigate the conservation of EpCAS's unique acyltransferase activity in other chicoric acid producing species, we selected several representative species including: another two *Echinacea* species (*Echinacea pallida* and *Echinacea angustifolia*) and two distantly related species (*Cichorium endivia* and *Lactuca sativa*) known to produce caftaric acid, chlorogenic acid and chicoric acid; as well as *Helianthus annuus* and *Chrysanthemum coronarium*, from two different intervening plant lineages which accumulate only chlorogenic acid (Fig. 5a-b and Supplementary Fig. 19). When we incubated

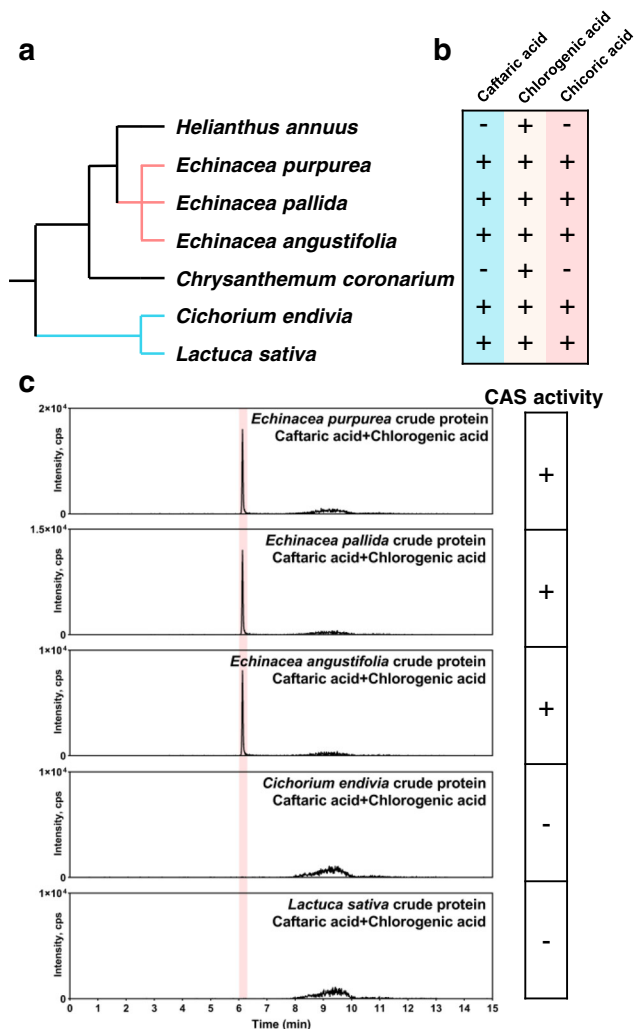


Fig. 5 Chicoric acid biosynthesis is polyphyletic between different species. **a** Phylogenetic relationships among several related species. Two *Echinacea* species (*Echinacea pallida* and *Echinacea angustifolia*) and two distantly related species (*Cichorium endivia* and *Lactuca sativa*) containing both substrates (caftaric acid and chlorogenic acid) and chicoric acid; as well as *Helianthus annuus* and *Chrysanthemum coronarium*, from two intervening plant lineages which contain only chlorogenic acid were selected for the study of chicoric acid biosynthetic routes. The tree was adopted from NCBI taxonomy (<http://lifemap-ncbi.univ-lyon1.fr/>). **b** The presence of target compounds in each species. **c** CAS activities from different chicoric acid containing plant crude protein extracts. Only *Echinacea* species can use caftaric acid as an acyl acceptor and chlorogenic acid as an acyl donor to produce chicoric acid.

crude protein extracts from the other chicoric acid-producing species with caftaric acid and chlorogenic acid, *E. pallida* and *E. angustifolia* crude protein extracts showed CAS activities. However, *Cichorium endivia* and *Lactuca sativa* failed to produce chicoric acid (Fig. 5c). Crude protein extracts from all samples exhibited HQT activities, but *Cichorium endivia* and *Lactuca sativa* crude protein extracts did not show HTT activities agreeing with their lack of CAS activity (Supplementary Fig. 20). In addition, we were unable to find any sequences encoding highly homologous proteins to EpHTT and EpCAS from the *Lactuca sativa* genome or any other public available databases by BLAST, while the BAHD proteins with the HTLSD motif (conserved in HQT-like proteins) were found. It seems likely that the pathway for chicoric acid biosynthesis catalyzed by HTT and CAS in

Echinacea species is unique. In other chicoric acid containing species, we conclude that chicoric acid is likely synthesized using different substrates and enzyme activities. Our results support the idea that chicoric acid biosynthesis is polyphyletic and has evolved in different plant lineages by convergence⁴⁰.

Reconstruction of chicoric acid biosynthesis in *N. benthamiana*. As a proof of concept, we attempted to reconstruct chicoric acid biosynthesis in planta. As *N. benthamiana* leaves produce abundant chlorogenic acid but no tartaric acid¹², different biosynthetic genes (*EpHTT* and *EpCAS*) and precursor (tartaric acid) combinations were investigated for their ability to produce chicoric acid. Coexpression of *EpHTT* and *EpCAS* and subsequent supplementation with tartaric acid generated chicoric acid verifying the functions of the identified genes and showing that they could be used for engineering chicoric acid production in plant chassis (Fig. 6). Future attempts to scale-up production of chicoric acid could be explored in chassis with a good source of tartaric acid combined with other metabolic engineering tools⁴¹.

Discussion

For millennia, humans have used plant specialized metabolites as herbal medicines. During the past two decades, especially in the postgenomic era, strategies such as co-expression and genome mining have been used extensively for metabolic pathway identification⁴². In this study, however, we describe steps in a metabolic pathway identified using traditional biochemistry. While RNA-seq, co-expression analysis and genome mining are becoming more popular approaches, discovery of genuinely novel activities may still be served best by strategies based on characterizing enzyme activities⁴².

We have shown that the biosynthesis of chicoric acid, the major bioactive compound of purple coneflower, involves a two-step process. In the cytosol, two BAHD acyltransferases, EpHTT and EpHQT catalyze the production of caftaric acid and chlorogenic acid intermediates, respectively. Both compounds are transported to the vacuole to form chicoric acid catalyzed by EpCAS (Fig. 7). In purple coneflower, chicoric acid has been reported to be stored in vacuoles⁴³, in line with our proposed biosynthetic pathway. Chicoric acid is also present in intercellular spaces and cell walls which may be due to release following cell damage. Like other identified SCPLs^{13,35,44}, EpCAS prefers acidic conditions and exhibits highest in vitro enzyme activity at pH 4, like SIHQT chlorogenate-chlorogenate transferase activity in vacuoles²⁴. EpCAS activity was relatively high (>80% of the maximum) over the pH range of 3.6 to 4.9, which should ensure the production of chicoric acid under vacuolar pH conditions. Based on our results, the biosynthesis of chicoric acid is likely to depend on transport of caftaric acid and chlorogenic acid into the vacuoles, although, to date, no transporters for hydroxycinnamic acid derivatives have been reported. Chlorogenic acid has been reported to be transported into vacuoles by vesicles in *Lonicera japonica* flowers⁴⁵.

Chicoric acid has been reported to have multiple physiological functions⁶, making it an important target for synthetic biology. Using *N. benthamiana*, we have successfully reconstituted the chicoric acid biosynthetic pathway. However, due to the absence of tartaric acid, *N. benthamiana* is not the ideal chassis for producing chicoric acid. Since the biosynthesis of tartaric acid has not been fully elucidated^{46,47}, further attempts to reconstruct chicoric acid biosynthesis should be focused on chassis with naturally high tartaric acid contents.

Our results emphasize the flexibility of specialized metabolism in plants, particularly the broad importance of transfer of acyl groups to form specialized bioactives. Unlike other SCPLs, in

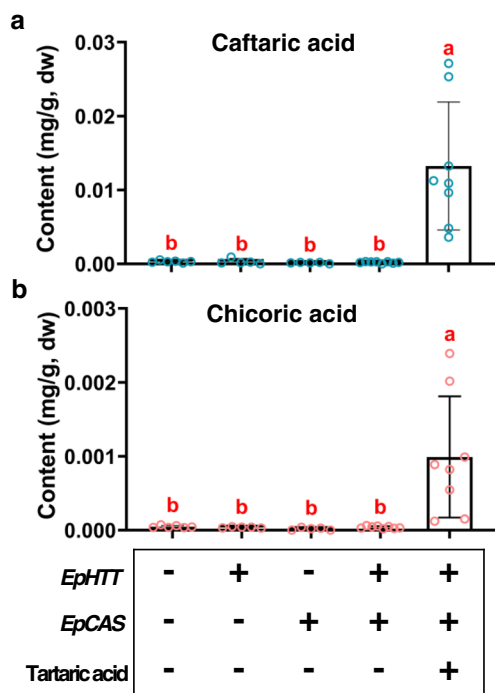


Fig. 6 Reconstruction of chicoric acid biosynthesis in *Nicotiana benthamiana* leaves. **a** Caftaric acid and **b** chicoric acid contents of *N. benthamiana* leaves infiltrated with different combinations of genes and precursor. Data are mean±s.d. ($n = 8, 5, 5, 6, 8$ biologically independent samples, respectively). Lowercase letters indicate significant differences from each other ($P < 0.05$) analyzed by one-way ANOVA with Tukey's multiple comparisons test. Source data are provided as a Source Data file.

purple coneflower, EpCAS utilizes chlorogenic acid rather than 1-*O*-caffeoyl β -glucose as its acyl donor^{13,37}. Among the species we tested, this pathway is unique, existing only *Echinacea* species and it is likely that chicoric acid biosynthesis is polyphyletic and has evolved by convergent mechanisms in other plant lineages. Chicoric acid is also vacuole-localized in *Cichorium endivia*⁴⁸ where it may be synthesized using caffeoyl CoA as an acyl donor as suggested for *Equisetum arvense*⁷, or produced by modification of other intermediates, such as *p*-coumaroylcaffeoyl tartaric acid. Further research is needed to clarify its biosynthesis in other species. Additionally, it will be interesting to identify the biological functions (if any) of chicoric acid in *Echinacea* and other species. This could help us better understand its evolution and regulation. In summary, our results expand our understanding of acyltransferases in specialized metabolism and highlight the rapid evolution possible with these types of enzymes.

Methods

Plant materials. Seeds of *Echinacea purpurea* (No. PI656830) were obtained from United States National Plant Germplasm System. Seeds were pretreated with 40 °C water for 6 h and then grown in sterilized soil in the greenhouse (23 ± 2 °C, 16 h light/8 h dark). Two-month-old seedlings were used for profiling of caffeic acid derivatives. Germination of sterile seedlings was conducted as follows: seeds were sterilized with 5% Plant Preservative Mixture (PPM) (Plant Cell Technology, USA) in 3× Murashige and Skoog (MS) basal salt medium for 10 h and washed with sterile water three times. Superfluous water was absorbed with sterile filter paper. The seeds were plated on MS solid medium in the dark (covered with aluminum foil paper) for 2 weeks at 23 °C before germination. Seedlings were transferred into an incubator (23 °C, 16 h light/8 h dark) and cultivated for one month before hairy root production. *Nicotiana benthamiana* was grown in sterilized soil in a greenhouse (23 ± 2 °C, 16 h light/8 h dark) for subcellular localization and metabolic engineering. *Helianthus annuus*, *Chrysanthemum coronarium*, *Cichorium endivia*, and *Lactuca sativa* were collected from a local supermarket.

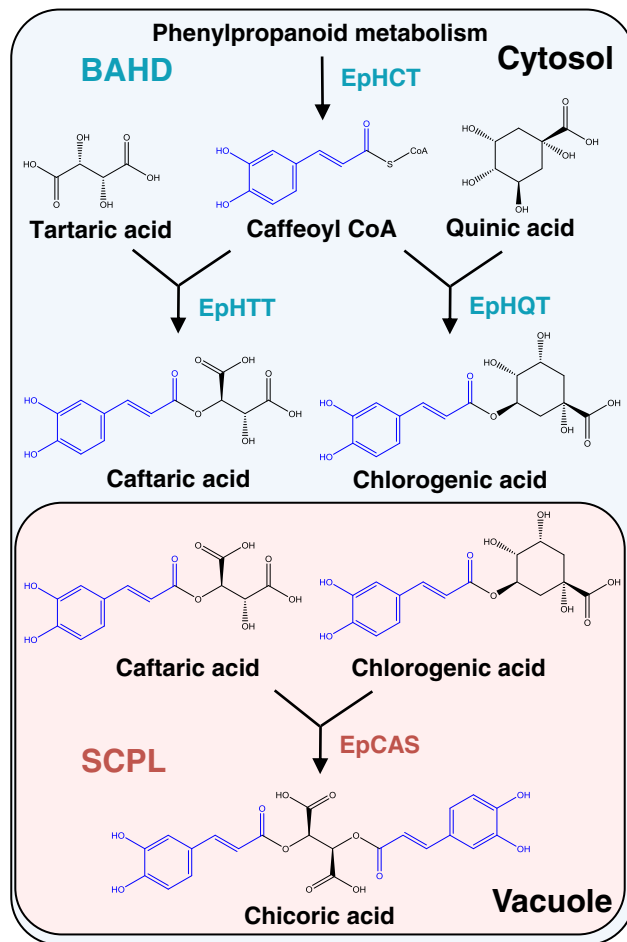


Fig. 7 Schematic representation of the proposed chicoric acid biosynthetic pathway in purple coneflower. Two types of acyltransferases distributed in distinct subcellular compartments are involved in the biosynthesis of chicoric acid. In the cytosol, the BAHDs including EpHTT and EpHQT use caffeoyl CoA from phenylpropanoid metabolism as an acyl donor to synthesize caftaric acid and chlorogenic acid, respectively. Both products are then transported into the vacuole. EpCAS, a specialized SCPL acyltransferase, uses chlorogenic acid as its acyl donor instead of 1-*O*-caffeoyl- β -D-glucose and transfers the caffeoyl group to caftaric acid to generate chicoric acid.

Extraction of caffeic acid derivatives. Plant materials were collected and frozen in liquid nitrogen immediately and stored at -80 °C. On the day of extraction, samples were lyophilized and ground into powder. Ten milligrams dry sample were weighed and extracted with 1 mL 70% methanol (v/v) in an ultrasonic bath (Qsonica 700, USA) with 50% amplitude at 4 °C, 20 s on 40 s off, for a total of 15 min. Extracts were centrifuged at $5,000 \times g$ for 10 min at 4 °C and the supernatants were collected. The residue was extracted again using the same conditions. The two supernatants were combined and diluted to 2.5 mg/mL (dry material vs. solvent) for analysis using an ultra-performance liquid chromatography-diode array detector (UPLC-DAD) for detection of caffeic acid derivatives (CADs) and 0.5 mg/mL for liquid chromatography-high-resolution mass spectrometry (LC-HRMS) analysis. The extracts were further centrifuged at $20,000 \times g$ for 10 min at 4 °C and the supernatants were collected and stored at -80 °C before analysis.

Qualitative analysis of CADs using LC-HRMS system. CADs were qualified based on the comparison of their retention times and mass spectra from samples and standards using LC-HRMS (Nexera UHPLC LC-30A and AB SCIEX qTOF X500R) coupled with a Turbol V™ source and SCIEX OS software (version 1.7). The LC analytical conditions were as follows: samples were separated using a Hypersil Gold C18 column (100×2.1 mm, 1.9 μ m; Thermo Fisher Scientific, USA) and the column temperature was set at 40 °C. The flow rate was 0.4 mL/min. The mobile phases were 0.1% formic acid (A) and acetonitrile (B), and the gradient was as follows: 2% B for 0.5 min; 2–20% B for 5.5 min; 20–95% B for 4 min; 95% B for 2 min followed by a decrease to 2% B for 0.1 min and re-equilibration of the

column for 2.9 min with 2% B. The injection volume was 1.0 μ L and the sampler temperature was set at 15 °C. Mass spectrometry was performed using an electrospray ionization (ESI) source operating with the information-dependent acquisition (IDA) model. The source parameters in negative polarity were as follows: ion source gas 1: 50 psi; ion source gas 2: 50 psi; curtain gas: 35 psi; CAD gas: 7; temperature: 450 °C; spray voltage: -4500 V; for TOF-MS, the mass range: 100–1000 Da; declustering potential (DP): -80 V; DP spread: 0; collision energy (CE): -10 V; CE spread: 0; accumulation time: 0.1 s; IDA criteria: small molecule; for TOF-MS/MS, the mass range: 50 <https://mpsd3.mpslimited.com/Digicore/DigiEditPage.aspx?FileName=421352456100491411330137.xml> 1000 Da; DP: -80 V; DP spread: 0; CE: -40 V; CE spread: 20 V; accumulation time: 0.05 s.

Quantitative analysis of CADs using UPLC-DAD system. CAD contents were determined quantitatively using a UPLC-DAD system. Three CAD standards including caftaric acid, chicoric acid and chlorogenic acid were used. Retention times and the UV-vis spectra of standards and samples were compared for confirmation. Quantification was performed on a DIONEX UltiMate 3000 UHPLC system with Chromeleon software (Thermo Fisher Scientific, USA). Sample separation was performed using a Hypersil Gold C18 column (100 \times 2.1 mm, 1.9 μ m; Thermo Fisher Scientific, USA), and the column temperature was set at 40 °C. The mobile phases were 0.1% formic acid water (A) and acetonitrile (B). The gradient was as follows: 2% B for 0.5 min; 2–14% B for 1 min; 14–18% B for 5.5 min; 18–80% B for 2 min; 80% B for 1 min, followed by re-equilibration of the column for 3 min with 2% B. The flow rate was set as 0.5 mL/min and the injection volume was 2.0 μ L. The sampler temperature was 10 °C. UV-vis absorption spectra were recorded on-line from 190 to 800 nm. Detection was conducted at 330 nm for quantitative purposes. Quantifications were carried out using external standards for calibration curves. The calibration curve coefficients for caftaric acid, chicoric acid, and chlorogenic acid were all above 0.9999. The results were expressed as mg of each CADs per 1 g dry material.

Generation of *Echinacea purpurea* hairy roots. The hairy roots were constructed as previously reported⁴⁹. Briefly, single colonies of pK7WG2R⁵⁰ plasmid containing *Agrobacterium rhizogenes* A4 strain were inoculated into 5 mL of tryptone yeast broth (with 10 mM CaCl₂, 50 mg/L spectinomycin, and 50 mg/L kanamycin). The cultures were incubated at 28 °C overnight with shaking (180 rpm). The cultures were centrifuged for 15 min at 3000 \times g at 4 °C and resuspended in MS liquid medium containing 200 μ M acetosyringone to make infection solution.

Leaf explants from one-month-old *Echinacea purpurea* plants grown on MS medium were scratched by a sterilized knife blade in cold, sterilized water. Explants were immersed in infection solution for 5 min and blotted dry on sterile filter paper. Infected explants were then cocultured on MS medium containing 200 μ M acetosyringone at 25 °C in the dark for 3 days before moving into MS medium supplemented with cefotaxime (250 mg/L). Hairy roots developed on cut ends of tissue were cultured in fresh 1/2 B5 solid medium (with 250 mg/L cefotaxime) at 25 °C in the dark. Hairy root cultures were transferred to fresh solid medium every 3 weeks and maintained as separate independent clones. Elongated root tips were transferred to flasks containing 30 mL of 1/2 B5 liquid medium (250 mg/L cefotaxime) to make liquid cultures, which were subcultured every week.

Methyl jasmonate treatment and RNA-sequencing. Empty vector-induced hairy roots were used for methyl jasmonate (MeJA) treatment. Hairy root cultures coming from independent progenitor roots were prepared and part of each independent culture received 100 μ M MeJA and the other part was used as the control. The hairy roots were collected after 24 h for RNA-sequencing and after 48 h for profiling of CADs⁵¹.

Total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, German). RNA-sequencing was performed using the BGI Illumina-HiSeq2000 platform. Library construction, sequencing, and analysis were carried out by the Beijing Genomic Institute (Shenzhen, China).

Crude protein extracts. The plant material was frozen in liquid nitrogen and ground into a fine powder. One gram of powdered material was extracted with 4 mL extraction buffer (phosphate buffer saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, and 2 mM KH₂PO₄) pH 7.0 containing 10 mM EDTA, 2% PVPP, 5 mM DTT, and 1 \times proteinase inhibitor cocktail (product No.: 4693132001, Roche) on ice for 2 h. The mixture was centrifuged at 4 °C at 5,000 \times g for 10 min. The supernatant was further centrifuged at 4 °C at 15,000 \times g for 10 min. The supernatant was filtered through a 0.22 μ m polyethersulfone (PES) membrane and dialyzed against PBS buffer (pH 7.0) overnight at 4 °C refrigeration to remove low molecular weight compounds. The protein concentration was determined using a BCA kit (Sangon Biotech, China).

Assay conditions for BAHD enzymes during protein purification. A 100 μ L reaction containing PBS buffer (pH 7.0), 20 μ L crude protein extract (0.52 μ g/ μ L), 120 μ M caffeoyl CoA, and 120 μ M tartaric acid or quinic acid was used for EpHTT and EpHQT assays, respectively. The mixture was incubated at 30 °C for 30 min and the reaction was stopped by adding 300 μ L methanol. The supernatant was collected after centrifugation at 20,000 \times g at 4 °C for 10 min and 2.0 μ L was injected

for UPLC-DAD analysis. The UPLC conditions were the same as for quantitative analysis of CADs except for the gradient which was isocratic 5% B for 10 min.

Purification of BAHD enzymes. First, ammonium sulfate precipitation was used to fractionate crude protein extracts. Following pre-tests, 0–30%, 30–40%, 40–60%, and >60% (NH₄)₂SO₄ fractions were collected. After the addition of (NH₄)₂SO₄ powder, the mixture was shaken at 4 °C for 1 h and then centrifuged at 20,000 \times g at 4 °C for 15 min. The supernatant was used for further precipitation and the sediment was resuspended in PBS (pH 7.0). (NH₄)₂SO₄ was removed by dialysis against PBS (pH 7.0) at 4 °C overnight. During this period, PBS buffer was replaced three times. The enzymes were purified further by FPLC (Fast Protein Liquid Chromatography, ÄKTA, GE Healthcare) coupled with HiTrap Q HP column (GE Healthcare) using a gradient of increasing concentration of NaCl from 0 to 1 M in PBS buffer (pH 7.0). The enzyme activities of fractions were measured, and those fractions with the greatest activity were loaded into SDS-PAGE gels and run for 5 min. When all the samples had entered the separation gel, each lane was cut out and processed for protein MS identification as detailed below.

Protein MS identification. Gels segments cut from SDS-PAGE gels were processed for protein MS identification. Peptides were digested by trypsin (Promega) and analyzed using nanoLC-QTOF-MS/MS (Eksper NanoLC 425 and AB SCIEX Triple TOF 5600)⁵². Peptides were separated using a SiELC ChromXP C18 column (75 μ m \times 15 cm, 3 μ m, 120 Å). LC flow rate was set to 0.3 μ L/min with phase A (2% acetonitrile, 98% water, 0.1% formic acid) and phase B (98% acetonitrile, 2% water, 0.1% formic acid). Gradients were: 5–7% B for 1 min, 7–22% B for 31 min, 32–38% B for 10 min, 32–55% B for 7 min, 55–80% B for 0.5 min, and hold at 80% B for 4.5 min, followed by re-equilibration with 5% B for 10 min. The MS method was as follows: TOF MS: CUR: 30; GS1: 12; GS2: 0; IHT: 150; ISVF: 2300; Mass Range: 300–1250 Da; Accumulation Time: 250 ms. IDA Criteria: 20; TOF MS/MS spectra acquired after each TOF MS scan; Rolling CE was applied. Mass Range: 100–1500 Da; Accumulation Time: 50 ms; CES: 5. Peptides were analyzed with ProteinPilot software and compared with the transcriptome database.

Expression of recombinant BAHD enzymes in *E. coli* and purification. Full-length CDS of EpHTT, EpHQT, and EpHCT were amplified from *Echinacea purpurea* hairy root cDNA using primers listed in Supplementary Table 4 and cloned into the pDEST17 plasmid with 6 \times His at the N-terminus⁵³. The recombinant plasmids were transformed into *E. coli* BL21(DE3) which was grown on LB solid medium (50 μ g/mL ampicillin). Overnight cultures were transferred to 200 mL LB liquid medium containing antibiotics until the OD₆₀₀ reached 0.5–1.0. Isopropyl β -D-1-thiogalactopyranoside was added to a final concentration of 0.4 mM, and the cultures were grown overnight at 16 °C with shaking at 200 rpm. Cells were harvested by centrifugation at 5,000 \times g for 5 min.

Cells were resuspended in lysis buffer (25 mM Tris-HCl 8.0, 150 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine). ATP (1 mM) and PMSF (1 mM) were also added. The cells were broken using a disruptor with 600 bar. After centrifugation at 20,000 \times g at 4 °C for 30 min, the supernatant was collected. Protein was purified using a Ni²⁺-NTA column (Qiagen) according to the manufacturer's instructions. Imidazole was removed using FPLC coupled with HiTrap Q HP column for purification of enzymes from crude extracts. The protein concentration was determined using a BCA kit (Sangon Biotech, China). Recombinant proteins were stored in PBS (pH7.0 with 10% glycerol) at -80 °C before use.

Recombinant BAHD activities and kinetics. According to our preliminary results, a UPLC based endpoint method was adopted for the determination of kinetics. Standard BAHD in vitro enzyme assays were performed in a 200 μ L reaction mix containing PBS buffer (at appropriate pH), acyl-CoA thioesters as the acyl donors, and organic acids as the acyl acceptors. Purified protein (500 ng) was added to start the reaction. The mixture was incubated at 30 °C for 30 min. Methanol (300 μ L) was added to stop the reaction. After centrifugation at 20,000 \times g at 4 °C for 10 min, the supernatant was collected and analyzed quantitatively using UPLC-DAD under the same conditions used for purification of BAHD activity. The enzyme reaction products were identified using LC-HRMS under the same conditions used for profiling of CADs. Enzyme activity was calculated as production of products and expressed as pkat (pmol/s). Kinetic parameters were determined from Michaelis-Menten plots using GraphPad Prism software (version 8.01). All in vitro enzyme assays were repeated at least three times.

For investigation of the effects of pH on BAHD activity at different pH values (from 2 to 9), PBS buffers were used in the enzyme assays. 100 μ M tartaric acid or quinic acid and 100 μ M caffeoyl CoA were used in the enzyme assays. The generation of products was quantified using UPLC.

For kinetic analyses, PBS pH 7.0 was selected as an optimum buffer. The kinetics for the acyl donors of EpHTT were determined using 0.5 mM tartaric acid and caffeoyl CoA (6.25–62.5 μ M), *p*-coumaroyl CoA (12.5–500 μ M), feruloyl CoA (25–200 μ M), respectively. When determining the kinetics for acyl acceptors for EpHTT, different combinations were used, including caffeoyl CoA (50 μ M) and tartaric acid (200–2400 μ M), *p*-coumaroyl CoA (100 μ M) and tartaric acid (500–4000 μ M), and feruloyl CoA (100 μ M) and tartaric acid (250–4000 μ M). Quinic acid was fixed at 1 mM and various acyl-CoA thioesters (caffeoyl CoA

(6.25–62.5 μM); *p*-coumaroyl CoA (25–800 μM) and feruloyl CoA (50–400 μM) were used for the study of the kinetics of EpHQT activity with different acyl donors. The kinetics of EpHQT's activity with different acyl donors were measured using the following combinations: caffeoyl CoA (50 μM) and quinic acid (200–2400 μM), *p*-coumaroyl CoA (100 μM) and quinic acid (250–2000 μM), and feruloyl CoA (100 μM) and quinic acid (1000–5000 μM). Products including *p*-coumaroylquinic acid, *p*-coumaroyltartaric acid, feruloylquinic acid, and feruloyltartaric acid were identified using LC-HRMS under the same conditions described above. Quantification was based on UPLC-DAD using caftaric acid and chlorogenic acid as references for tartaric acid esters and quinic acid esters, respectively.

For investigating the substrate preferences of BAHDs, acyl-CoA thioesters (100 μM) and organic acids (2 mM) including tartaric acid, quinic acid, and shikimic acid were used. The products were measured with LC-HRMS.

Assay conditions for SCPL enzyme during protein purification. Optimum conditions for SCPL activity were established by pre-tests. Assays were performed in a 200 μL in vitro enzyme reaction containing chlorogenic acid (250 μM), caftaric acid (250 μM), 10 mM Na_2HPO_4 , and 2 mM KH_2PO_4 (pH 4.0). A total of 20 μL crude protein extract (0.4 $\mu\text{g}/\mu\text{L}$) was added to start the reaction. After incubation at 30 °C for 1 h, the reaction was stopped by adding 300 μL methanol. The mixture was centrifuged at 4 °C at 20,000 $\times g$ for 10 min and the supernatant was used for UPLC analysis under the same conditions as described above except the gradients were as follows: 2% B for 0.5 min, 2–15% B for 3.5 min, stay 15% B for 3 min, then drop to 2% B in 0.1 min and re-equilibrated with 2% for 2.9 min.

Purification of chicoric acid synthase activity. Similar to the purification of BAHD enzymes, ammonium sulfate precipitation was used to separate chicoric acid synthase activity initially. Following pre-tests, 0–20%, 20–40%, 40–60%, and >60% $(\text{NH}_4)_2\text{SO}_4$ fractions were used for fractionation. The fractions were prepared and desalted as described for BAHD acyl transferases. The active fraction was separated further by two steps including the use of the same anion exchange column used for BAHD purification and sieve chromatography. Molecular sieve separation was performed on FPLC with Superdex 200 10/300 GL (GE Healthcare) using PBS buffer (pH 7.0). Finally, the active fraction was used for the identification of EpCAS using the protein MS identification method described above.

Recombinant SCPL enzyme expression in *S. cerevisiae* and purification. The full-length cDNA of EpCAS was cloned into the pESC-His expression vector⁵⁴ with a 6 \times His tag at the C-terminus using ClonExpress II One Step Cloning Kit (Vazyme). *Saccharomyces cerevisiae* BY4741 was used as the host yeast strain. Yeast transformation procedures were conducted using a lithium acetate method⁵⁵ and were incubated at 30 °C. Single colonies were grown on SD dropout media containing 2% dextrose without Histidine (Solarbio) and were picked and then used for purification by inoculation into 15 mL of SD dropout medium containing 2% dextrose. Cells were grown overnight at 30 °C on a rotary shaker set at 200 rpm. Cultures were then inoculated in 200 mL of SD dropout media containing 2% dextrose and grown to an OD_{600} of 0.4. All cells were transferred into 200 mL SD dropout medium containing 2% galactose. Yeast cells were grown at 30 °C on a rotary shaker for 36 h. The cells were centrifuged at 1500 $\times g$ for 5 min at 4 °C. The supernatant was discarded, and the cells were resuspended in 10 mL of sterile water. Cells were centrifuged at 1500 $\times g$ for 5 min at 4 °C. The supernatant was discarded, and the cell pellets were stored at -80 °C for later experiments.

Cells were resuspended in the same volume of Ni-NTA lysis buffer (50 mM Na_2HPO_4 , 300 mM NaCl and 10 mM imidazole, pH 8.0) and broken using acid-washed glass beads (425–600 μm , Sigma) by shaking in a grinding machine (60 Hz, 60 s for 5 times) and then centrifuged at 5,000 $\times g$ at 4 °C for 10 min. The supernatant was collected and further centrifuged at 20,000 $\times g$ at 4 °C for 10 min. After passage through a 0.22 μm PES membrane, the protein was purified using a Ni²⁺-NTA column (Qiagen) according to the manufacturer's instructions with some modifications. The column was washed using 20 mM imidazole and eluted using pH 4.5 buffer (50 mM Na_2HPO_4 , 300 mM NaCl). Protein concentration was determined using a BCA kit (Sangon Biotech, China) and recombinant protein was stored at -80 °C before use.

Mutagenesis and expression of EpCAS in *S. cerevisiae*. Overlapping gene fragments were separately amplified using EpCAS as a template with primers containing the desired single and combined amino acid substitutions⁵⁶. The fragments were simultaneously cloned into pESC-His vector for expression as described above.

Determination of recombinant EpCAS characteristics. Standard conditions for EpCAS assays were 100 μL reactions containing chlorogenic acid (200 μM), caftaric acid (100 μM), 10 mM Na_2HPO_4 , and 2 mM KH_2PO_4 (pH 4.0). Protein (1.41 μg) was added to start the reaction. After incubation at 30 °C for 12 h (based on pre-experiments), the reaction was stopped by adding 400 μL methanol. The mixture was centrifuged at 4 °C at 20,000 $\times g$ for 10 min and the supernatant was used for UPLC-DAD analysis under the same conditions used for the SCPL enzyme assay

during purification. Reactions were also analyzed using LC-HRMS by the same method as used for assaying BAHD activities.

For determination of the effects of pH, starting solution (10 mM Na_2HPO_4 and 2 mM KH_2PO_4) was adjusted into different pH (ranging from pH 1.6 to 10.7) by adding 1 M HCl or 1 M NaOH, and all other procedures were performed using standard conditions. For further identification of EpCAS as an SCPL protein, PMSF at different concentrations including 1, 5, and 10 mM was used to inhibit enzyme activity. For product and substrate degradation assays, 100 μM chicoric acid, 100 μM caftaric acid and 200 μM chlorogenic acid were used. To establish whether caftaric acid and caffeic acid can generate chicoric acid, 100 μM caftaric acid and 200 μM caffeic acid were combined. For kinetic analysis, caftaric acid ranging from 10 to 200 μM was mixed with 200 μM chlorogenic acid and chlorogenic acid ranging from 20 to 200 μM was mixed with 200 μM caftaric acid and assays were incubated with EpCAS for caftaric acid and chlorogenic acid analysis, respectively. For kinetic analysis against 1-*O*-caffeoyl β -D-glucose, caftaric acid 200 μM was mixed with 1-*O*-caffeoyl β -D-glucose ranging from 10 to 400 μM . For the characterization of activities with other acyl donors, 100 μM caftaric acid was mixed with 200 μM candidate donors including chlorogenic acid, 1-*O*-caffeoyl β -D-glucose, neochlorogenic acid, cryptochlorogenic acid and 5-*O*-caffeoylshikimic acid. The products were determined using LC-HRMS and UPLC-DAD. All in vitro enzyme assays were conducted at least three times.

Synthesis of 1-*O*-caffeoyl- β -D-glucose. 1-*O*-caffeoyl- β -D-glucose was synthesized according to published methods^{57,58}, and its identity was confirmed using LC-MS and NMR. UV/Vis: λ_{max} 330 nm. ¹H NMR (400 MHz, CD_3OD) δ : 7.66 (d, J = 16.0 Hz, 1H), 7.06 (s, 1H), 6.97 (d, J = 8.4 Hz, 1H), 6.79 (d, J = 8.0 Hz, 1H), 6.31 (d, J = 16.0 Hz, 1H), 5.58 (d, J = 7.6 Hz, 1H), 3.86 (d, J = 12.0 Hz, 1H), 3.72–3.68 (m, 1 H), 3.47–3.39 (m, 4H). HRMS (m/z): [M]⁻ calculated for $\text{C}_{15}\text{H}_{18}\text{O}_9$, 341.0878; found, 341.0881 (Supplementary Fig. 12).

Phylogenetic analysis. The sequences of BAHD (CcsHCT (AFL93686), CcsHQT1 (CAM84302) and CcsHQT2 (CAR92145) from *Cynara cardunculus* var. *scolymus*; CiHCT1 (ANN12608), CiHCT2 (ANN12609), CiHQT1 (ANN12610), CiHQT2 (ANN12611), and CiHQT3 (ANN12612) from *Cichorium intybus*; NtHCT (CAD47830) and NtHQT (CAE46932) from *Nicotiana tabacum*; CaHCT (CAJ40778) from *Coffea arabica*; CcHQT (ABO77957) from *Coffea canephora*; TpHCT1A (ACI16630) from *Trifolium pratense*; AtHCT (AED95744) from *Arabidopsis thaliana*; SbHCT (4KE4_A) from *Sorghum bicolor*; AsHHT1 (BAC78633) from *Avena sativa*; CcaHQT (ABK79690) from *Cynara cardunculus* var. *altilis*; SlHQT (CAE46933) from *Solanum lycopersicum*; MpHCT (AXN55971) from *Marchantia paleacea*; PaHCT (AXN55972) from *Plagioclasma appendiculatum*; CsHCT (AEJ88365) from *Cucumis sativus*; PpHCT1 (AMK38063) from *Physcomitrium patens* and SCPL (SpGAC (AAF64227) from *Solanum pennellii*; AtSMT (AAF78760), AtSCT (AAK52316), AtSAT (AEC07395), AtSST (AEC07397) and AtSCPL17 (AAS99709) from *Arabidopsis thaliana*; BnSCT1 (AAQ91191) and BnSCT2 (CAM91991) from *Brassica napus*; DkSCPL1 (BAF56655) and DkSCPL2 (BAH89272) from *Diospyros kaki*; CtAT1 (BAF99695) from *Clitoria ternatea*; AsSCPL1 (ACT21078) from *Avena strigosa*; OsIAA-At (EEE56946) from *Oryza sativa*) were obtained from NCBI and other public databases. Amino acid sequences were aligned and the Neighbor-Joining (NJ) trees were built using MEGA6 (Molecular Evolutionary Genetics Analysis version 6.0) with the following parameters: at least 1000 bootstrap replications, Poisson model, uniform rates and complete deletion⁵⁹.

Subcellular localization. Target genes were cloned into pSuper1300-GFP⁶⁰ using primers listed in Supplementary Table 4 to create in-frame, C-terminal GFP fusion proteins, and transformed into *Agrobacterium tumefaciens* strain GV3101. Transformed clones were grown on 2 mL LB selective medium (50 mg/L kanamycin, 50 mg/L gentamicin, and 50 mg/L rifampin) for 24 h at 28 °C with shaking at 180 rpm. The cultures were transferred to 20 mL LB medium containing antibiotics and grown until the OD_{600} reached 1.0. Cells were harvested and resuspended in infiltration buffer (10 mM MES pH 5.6, 10 mM MgCl_2 , and 200 μM acetosyringone) and incubated for an additional 2 h in the dark at room temperature. Strains containing target plasmids were infiltrated into the leaves of 4-week-old *Nicotiana benthamiana* plants. The plants were grown in the dark for 24 h (EpHTT and EpHQT) or 72 h (EpCAS) and prepared for microscopic observations. Leaves were cut into small pieces and placed on glass slides and observed with a ZEISS Cell Observer SD Spinning Disk Confocal Microscope. The vacuoles were stained by 10 μM SNARF-1 for 2 h before observation⁶¹. Detection parameters were as follows: GFP excitation at 488 nm and emission at 500–530 nm, SNARF-1 excitation at 514 nm and emission at 600–700 nm.

Molecular docking. The protein 3D structures of BAHD enzymes were modeled using I-TASSER (<https://zhanglab.cmb.med.umich.edu/I-TASSER/>)⁶² and CIS-RR⁶³. Docking experiments were conducted on CB-Dock⁶⁴ based on AutoDock Vina⁶⁵. EpCAS was modeled by SwissModel⁶⁶ using SERINE CARBOX-YPEPTIDASE (PDB:IYSC) as template. AutoDock 4.2 was used for docking experiments⁶⁷. Molecular graphics were rendered with PyMOL (version 2.3.2).

Overexpression and RNAi of BAHDs and SCPL in hairy root cultures. *EpHTT*, *EpHQT* and *EpCAS* genes were cloned into plasmid pK7WG2R with a 35 S promoter⁵⁰ using Gateway recombination according to the manufacturer's instructions for the overexpression of target genes in hairy roots. Nonhomologous sequences of *EpHTT*, *EpHQT* and *EpCAS* were amplified with the primers listed in Supplementary Table 4 and were cloned into binary hpRNA vector pK7WGIGW2R⁶⁸ using Gateway recombination for RNAi. The vectors were introduced into *Agrobacterium rhizogenes* A4 by electroporation. Successful clones were selected on tryptone yeast broth solid medium supplemented with spectinomycin (50 mg/L) and kanamycin (50 mg/L). All other procedures for generation and analysis of hairy roots were as described as above.

Quantitative reverse transcription polymerase chain reaction. Total RNA was extracted with an RNeasy Plant Mini Kit (Qiagen). First-strand cDNA was synthesized from 1 µg RNA using a PrimeScript™ RT reagent Kit with gDNA Eraser (Takara). qRT-PCR was performed using a Bio-Rad CFX384 and iTaq Universal One-Step RT-qPCR Kits (Bio-Rad) according to the manufacturer's instructions. Results were calculated using Bio-Rad CFX manager software. Tubulin was used as internal control. The relative expression of genes was calculated using the $\Delta\Delta C_t$ method. The primer pairs for qRT-PCR were designed by Primer3Plus (<http://www.primer3plus.com>) (Supplementary Table 4) and compared by BLAST analysis to the NCBI database for confirmation of primer specificity.

Reconstruction of chicoric acid biosynthesis in *N. benthamiana*. For transient overexpression, *EpHTT* and *EpCAS* were cloned into the destination vector pEAQ-HT-DEST3⁶⁹ using Gateway recombination and transformed into *Agrobacterium tumefaciens* strain GV3101 by electroporation. Single clones with each target construct were transferred into 2 mL LB liquid culture (kanamycin, rifampicin, and gentamicin, 50 mg/L), and the cultures were grown at 28 °C in a shaker at 200 rpm for 1 day. For subculturing, 0.5 mL were transferred into 10 mL LB liquid medium (kanamycin, rifampicin, and gentamicin, 50 mg/L) and cells were grown for 6–8 h until the OD₆₀₀ reached 0.5–0.8. Cells were collected and resuspended in infiltration buffer (10 mM MgCl₂, 10 mM MES pH 5.6, and 250 µM acetosyringone) to OD₆₀₀ = 1.0. The *Agrobacterium tumefaciens* strains with vectors carrying target genes were infiltrated into *Nicotiana benthamiana* leaves by syringe. Substrates (tartaric acid 1 mM, caftaric acid 500 µM) were injected into infiltrated leaves 5 days after inoculation. After 24 h, the leaves were collected for LC-HRMS analysis using the same conditions used for CAD profiling.

Statistical analysis. Unless specifically described, all the experiments in this paper were repeated at least three times and results from representative data sets are presented. GraphPad Prism (version 8.02) and Microsoft Excel (version Office 365) was used for the statistical analysis. The statistical evaluations used unpaired *t* tests and one-way analysis of variance (ANOVA) with multiple comparisons, followed by Tukey tests. The results were considered statistically significant at **P* < 0.05.

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

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. The sequences of the genes reported in this article have been deposited in NCBI GenBank: *EpHTT* (MT936803), *EpHQT* (MT936804), *EpHCT* (MT936805), and *EpCAS* (MT936806). The RNA-seq data from MeJA treatment and the control have been deposited in the SRA database of NCBI under accession numbers SRR8935731, SRR8935732, SRR8935733 for MeJA treatments and SRR8937034, SRR8937035, SRR8937036 for controls. The source data underlying Figs. 1d, 2d–g, 3c, 3d, 4a, 4b, and 6, Supplementary Figures 1, 3, 5, 9c, and 16a, as well as Supplementary Tables 1–3 are provided as a Source Data file.

Received: 12 November 2020; Accepted: 11 February 2021;

Published online: 10 March 2021

References

- Shah, S. A., Sander, S., White, C. M., Rinaldi, M. & Coleman, C. I. Evaluation of echinacea for the prevention and treatment of the common cold: a meta-analysis. *Lancet Infect. Dis.* **7**, 473–480 (2007).
- Smith, T., May, G., Eckl, V. & Reynolds, C. M. US sales of herbal supplements increase by 8.6% in 2019. *HerbalGram* **127**, 54–69 (2020).

- Barrett, B. Medicinal properties of Echinacea: a critical review. *Phytomedicine* **10**, 66–86 (2003).
- Perry, N. B., Burgess, E. J. & Glennie, V. L. Echinacea standardization: analytical methods for phenolic compounds and typical levels in medicinal species. *J. Agric. Food Chem.* **49**, 1702–1706 (2001).
- Robinson, W. E. Jr., Reinecke, M. G., Abdel-Malek, S., Jia, Q. & Chow, S. A. Inhibitors of HIV-1 replication [corrected; erratum to be published] that inhibit HIV integrase. *Proc. Natl. Acad. Sci. USA* **93**, 6326–6331 (1996).
- Peng, Y., Sun, Q. & Park, Y. The bioactive effects of chicoric acid as a functional food ingredient. *J. Med. Food* **22**, 645–652 (2019).
- Hohlfeld, M., Veit, M. & Strack, D. Hydroxycinnamoyltransferases involved in the accumulation of caffeic acid esters in gametophytes and sporophytes of *Equisetum arvense*. *Plant Physiol.* **111**, 1153–1159 (1996).
- Nakajima, K., Hashimoto, T. & Yamada, Y. Two tropinone reductases with different stereospecificities are short-chain dehydrogenases evolved from a common ancestor. *Proc. Natl. Acad. Sci. USA* **90**, 9591–9595 (1993).
- Sullivan, M. L. Perennial peanut (*Arachis glabrata* Benth.) leaves contain hydroxycinnamoyl-CoA:tartaric acid hydroxycinnamoyl transferase activity and accumulate hydroxycinnamoyl-tartaric acid esters. *Planta* **239**, 1091–1100 (2014).
- Clifford, M. N., Jaganath, I. B., Ludwig, I. A. & Crozier, A. Chlorogenic acids and the acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. *Nat. Prod. Rep.* **34**, 1391–1421 (2017).
- D'Auria, J. C. Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331–340 (2006).
- Niggeweg, R., Michael, A. J. & Martin, C. Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat. Biotechnol.* **22**, 746–754 (2004).
- Bontpart, T., Cheynier, V., Ageorges, A. & Terrier, N. BAHD or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. *N. Phytologist* **208**, 695–707 (2015).
- Pellati, F., Benvenuti, S., Magro, L., Melegari, M. & Soragni, F. Analysis of phenolic compounds and radical scavenging activity of Echinacea spp. *J. Pharm. Biomed. Anal.* **35**, 289–301 (2004).
- Lau, W. & Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **349**, 1224–1228 (2015).
- Jeon, J. E. et al. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* **180**, 176–187 e119 (2020).
- Caputi, L. et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* **360**, 1235–1239 (2018).
- Chen, X. et al. A pathogenesis-related 10 protein catalyzes the final step in the biosynthesis of theanine. *Nat. Chem. Biol.* **14**, 738–743 (2018).
- Legrand, G. et al. Identification and characterization of five BAHD acyltransferases involved in hydroxycinnamoyl ester metabolism in chicory. *Front Plant Sci.* **7**, 741 (2016).
- Cheevanunapakul, K., Khaksar, G., Panpetch, P., Boonjing, P. & Sirikantaramas, S. Identification and functional characterization of genes involved in the biosynthesis of caffeoylquinic acids in Sunflower (*Helianthus annuus* L.). *Front Plant Sci.* **10**, 968 (2019).
- Lallemand, L. A. et al. A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. *Plant Physiol.* **160**, 249–260 (2012).
- Sonnante, G. et al. Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. *Plant Physiol.* **153**, 1224–1238 (2010).
- Levsh, O. et al. Dynamic conformational states dictate selectivity toward the native substrate in a substrate-permissive acyltransferase. *Biochemistry* **55**, 6314–6326 (2016).
- Moglia, A. et al. Dual catalytic activity of hydroxycinnamoyl-coenzyme A quinate transferase from tomato allows it to moonlight in the synthesis of both mono- and dicaffeoylquinic acids. *Plant Physiol.* **166**, 1777–1787 (2014).
- Teutschbein, J. et al. Identification and localization of a lipase-like acyltransferase in phenylpropanoid metabolism of tomato (*Solanum lycopersicum*). *J. Biol. Chem.* **285**, 38374–38381 (2010).
- Fraser, C. M., Rider, L. W. & Chapple, C. An expression and bioinformatics analysis of the Arabidopsis serine carboxypeptidase-like gene family. *Plant Physiol.* **138**, 1136–1148 (2005).
- Lehfeldt, C. et al. Cloning of the SNG1 gene of Arabidopsis reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. *Plant Cell* **12**, 1295–1306 (2000).
- Mugford, S. T. et al. A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell* **21**, 2473 (2009).
- Hause, B., Meyer, K., Viitanen, P. V., Chapple, C. & Strack, D. Immunolocalization of 1-O-sinapoylglucose:malate sinapoyltransferase in Arabidopsis thaliana. *Planta* **215**, 26–32 (2002).

30. Mugford, S. T. et al. Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell* **25**, 1078–1092 (2013).
31. Shirley, A. M., McMichael, C. M. & Chapple, C. The *sng2* mutant of *Arabidopsis* is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. *Plant J.* **28**, 83–94 (2001).
32. van der Hoorn, R. A. Plant proteases: from phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* **59**, 191–223 (2008).
33. Stehle, F., Brandt, W., Milkowski, C. & Strack, D. Structure determinants and substrate recognition of serine carboxypeptidase-like acyltransferases from plant secondary metabolism. *FEBS Lett.* **580**, 6366–6374 (2006).
34. Endrizzzi, J. A., Breddam, K. & Remington, S. J. 2.8-A structure of yeast serine carboxypeptidase. *Biochemistry* **33**, 11106–11120 (1994).
35. Stehle, F., Brandt, W., Schmidt, J., Milkowski, C. & Strack, D. Activities of *Arabidopsis* sinapoylglucose: malate sinapoyltransferase shed light on functional diversification of serine carboxypeptidase-like acyltransferases. *Phytochemistry* **69**, 1826–1831 (2008).
36. Kowalczyk, S., Jakubowska, A., Zielińska, E. & Bandurski, R. S. Bifunctional indole-3-acetyl transferase catalyses synthesis and hydrolysis of indole-3-acetyl-myo-inositol in immature endosperm of *Zea mays*. *Physiol. Plant.* **119**, 165–174 (2003).
37. Ciarkowska, A., Ostrowski, M., Starzyńska, E. & Jakubowska, A. Plant SCPL acyltransferases: multiplicity of enzymes with various functions in secondary metabolism. *Phytochem. Rev.* **18**, 303–316 (2018).
38. Stehle F., et al. Snap-shot of serine carboxypeptidase-like acyltransferase evolution: the loss of conserved disulphide bridge is responsible for the completion of neo-functionalization. *J. Phylogenet. Evol. Biol.* **1**, 3 (2013).
39. Lee, J. & Scagel, C. F. Chicoric acid: chemistry, distribution, and production. *Front. Chem.* **1**, 40 (2013).
40. Pichersky, E. & Lewinsohn, E. Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* **62**, 549–566 (2011).
41. Fu, R., Martin, C. & Zhang, Y. Next-generation plant metabolic engineering, inspired by an ancient chinese irrigation system. *Mol. Plant* **11**, 47–57 (2018).
42. Jacobowitz, J. R. & Weng, J. K. Exploring uncharted territories of plant specialized metabolism in the postgenomic era. *Annu. Rev. Plant Biol.* **71**, 631–658 (2020).
43. Li, Z. et al. The synthesis and storage sites of phenolic compounds in the root and rhizome of *Echinacea purpurea*. *Am. J. Plant Sci.* **03**, 551–558 (2012).
44. Liu, Y. et al. Purification and characterization of a novel galloyltransferase involved in catechin galloylation in the tea plant (*Camellia sinensis*). *J. Biol. Chem.* **287**, 44406–44417 (2012).
45. Li, Y. et al. Correlation of the temporal and spatial expression patterns of HQT with the biosynthesis and accumulation of chlorogenic acid in *Lonicera japonica* flowers. *Hortic. Res.* **6**, 73 (2019).
46. DeBolt, S., Cook, D. R. & Ford, C. M. L-tartaric acid synthesis from vitamin C in higher plants. *Proc. Natl Acad. Sci. USA* **103**, 5608–5613 (2006).
47. Jia, Y. et al. An aldo-keto reductase with 2-keto-l-gulonate reductase activity functions in l-tartaric acid biosynthesis from vitamin C in *Vitis vinifera*. *J. Biol. Chem.* **294**, 15932–15946 (2019).
48. Goupy, P. M., Varoquaux, P. J. A., Nicolas, J. J. & Macheix, J. J. Identification and localization of hydroxycinnamoyl and flavonol derivatives from endive (*Cichorium endivia* L. cv. Geante Maraichere) leaves. *J. Agric Food Chem.* **38**, 2116–2121 (1990).
49. Liu, C. Z., Abbasi, B. H., Gao, M., Murch, S. J. & Saxena, P. K. Caffeic acid derivatives production by hairy root cultures of *Echinacea purpurea*. *J. Agric Food Chem.* **54**, 8456–8460 (2006).
50. Ding, Y. et al. Abscisic acid coordinates nod factor and cytokinin signaling during the regulation of nodulation in *Medicago truncatula*. *Plant Cell* **20**, 2681 (2008).
51. Goossens, A. et al. A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proc. Natl Acad. Sci. USA* **100**, 8595–8600 (2003).
52. Dzieciatkowska M., Hill R., Hansen K. C. GeLC-MS/MS analysis of complex protein mixtures. In: *Shotgun Proteomics: Methods and Protocols* (ed Martins-de-Souza D.). Springer New York (2014).
53. Ivanov, D. et al. A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase-like serine proteases*. *J. Biol. Chem.* **281**, 17246–17252 (2006).
54. Wang, Y. & Yu, O. Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells. *J. Biotechnol.* **157**, 258–260 (2012).
55. Gietz, D., St Jean, A., Woods, R. A. & Schiestl, R. H. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* **20**, 1425 (1992).
56. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
57. Zhu, Y. & Ralph, J. Stereoselective synthesis of 1-O- β -feruloyl and 1-O- β -sinapoyl glucopyranoses. *Tetrahedron Lett.* **52**, 3729–3731 (2011).
58. Zhang, S.-Q., Li, Z.-J., Wang, A.-B., Cai, M.-S. & Feng, R. Total synthesis of the phenylpropanoid glycoside, grayanoside A. *Carbohydr. Res.* **299**, 281–285 (1997).
59. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
60. Cui, Q. et al. Analysis of WRKY transcription factors and characterization of two *Botrytis cinerea*-responsive LrWRKY genes from *Lilium regale*. *Plant Physiol. Biochem.* **127**, 525–536 (2018).
61. Bak, A., Cheung, A. L., Yang, C., Whitham, S. A. & Casteel, C. L. A viral protease relocalizes in the presence of the vector to promote vector performance. *Nat. Commun.* **8**, 14493 (2017).
62. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7 (2014).
63. Cao, Y. et al. Improved side-chain modeling by coupling clash-detection guided iterative search with rotamer relaxation. *Bioinformatics* **27**, 785–790 (2010).
64. Liu Y., et al. CB-Dock: a web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacol. Sin.* **41**, 138–144 (2019).
65. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
66. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
67. Morris, G. M. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
68. Zhao, Q. et al. Two CYP82D enzymes function as flavone hydroxylases in the biosynthesis of root-specific 4'-deoxyflavones in *Scutellaria baicalensis*. *Mol. Plant* **11**, 135–148 (2018).
69. Sainsbury, F., Thuenemann, E. C. & Lomonosoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693 (2009).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31670352 and 31800258), the Institutional Research Fund of Sichuan University (2020SCUNL106), the Fundamental Research Funds for the Central Universities (SCU2020D003), the China Postdoctoral Science Foundation (2018M631080) and the Post-Doctor Research Project, Sichuan University. C.M. was supported by Institute Strategic Programs: Understanding and Exploiting Plant and Microbial Secondary 336 Metabolism (BB/J004596/1) and Molecules from Nature (BB/P012523/1) from the UK Biotechnology and Biological Sciences Research Council (BBSRC) to the John Innes Centre (JIC). We acknowledge the Mass Spectrometry Core Facility in College of Life Sciences, Sichuan University for the assistance in proteomics and metabolic analysis and Dr. Qing Zhao (Chenshan Botanical Garden, China) for advice on hairy root production and maintenance. We thank Prof. Xiaoya Chen (Institute of Plant Physiology and Ecology, Chinese Academy of Science), Prof. Jie Luo (Hainan University, China) and Prof. Chengchao Xu (Artemisinin Research Center, China) for critical comments on the manuscript.

Author contributions

Y.Z. and C.M. devised the initial experimental strategies and R.F. and Y.Z. designed the experiments with C.M.'s advice. R.F. and P.Z. performed the experiments with the assistance of G.J. and L.W. Both S.Q. and Y.C. assisted with docking assays. R.F., C.M., and Y.Z. wrote the paper with inputs from all the authors who approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21853-6>.

Correspondence and requests for materials should be addressed to Y.Z.

Peer review information *Nature Communications* thanks Michael Wink and other, anonymous, reviewers for their contributions to the peer review of this work.

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) 2021