Current Plant Biology 46 (2026) 100599

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

Phylogeny-guided discovery of a crocetin dialdehyde-producing carotenoid cleavage dioxygenase from *Paulownia tomentosa*

Lucía Morote [a,1], Eduardo Parreño [a,1], Elena Moreno Giménez [a], Verónica Aragonés [b], Alberto José López Jiménez [a,c], Ángela Rubio-Moraga [a], Oussama Ahrazem [a,c], Gianfranco Diretto [d], José-Antonio Daròs [b], Lourdes Gómez-Gómez [a,e,*]

[a] Instituto Botánico, Departamento de Ciencia y Tecnología Agroforestal y Genética, Universidad de Castilla-La Mancha, Campus Universitario s/n, Albacete 02071, Spain
[b] Instituto de Biología Molecular y Celular de Plantas (Consejo Superior de Investigaciones Científicas-Universitat Politècnica de València), Valencia 46022, Spain
[c] Escuela Técnica Superior de Ingeniería Agronómica y de Montes y Biotecnología, Departamento de Ciencia y Tecnología Agroforestal y Genética, Universidad de Castilla-La Mancha, Campus Universitario s/n, Albacete 02071, Spain
[d] Italian National Agency for New Technologies, Energy, and Sustainable Development (ENEA), Biotechnology laboratory, Casaccia Research Centre, Rome 00123, Italy
[e] Facultad de Farmacia, Universidad de Castilla-La Mancha, Campus Universitario s/n, Albacete 02071, Spain

ARTICLE INFO

Keywords:
Carotenoid cleavage dioxygenase (CCD)
β-carotene
Crocetin
Apocarotenoids
Paulownia tomentosa
Genome
Heterologous expression

ABSTRACT

Carotenoid cleavage dioxygenases (CCDs) mediate the oxidative cleavage of carotenoids, producing apocarotenoids with diverse biological functions and industrial relevance. Phylogenetic analysis of CCDs from *Paulownia tomentosa* identified CCD4–2 as a candidate enzyme potentially involved in crocetin biosynthesis. CCD4–2 clustered closely with functionally characterized CCD4 enzymes from *Buddleja* and *Verbascum* (73.7% and 70.0% identity, respectively), both of which cleave carotenoids at the C7-C8 and C7′-C8′ double bonds to produce crocetin dialdehyde. *In silico* predictions and subcellular localization studies, supported that *P. tomentosa* CCD4–2 is plastid-targeted and likely interacts with membrane-bound carotenoid pools. Functional assays demonstrated that CCD4–2 cleaves specifically β-carotene at C7-C8 and C7′-C8′ but does not act on linear or asymmetric carotenoids. Although β-carotene was abundant in *P. tomentosa* leaves and flowers, crocins were not detected, likely due to substrate competition and the high expression of other CCD enzymes, particularly CCD1 and CCD4–4, which may divert β-carotene toward alternative cleavage pathways or products. To verify the catalytic function in planta, CCD4–2 was expressed using a viral vector in *Nicotiana benthamiana*. In this low-competition background, crocins accumulation reached values close to 4 mg/g dry weight, confirming the cleavage activity of CCD4–2. These findings underscore the presence of catalytically valuable CCD enzymes within plant genomes and demonstrate how prior knowledge of CCD enzymatic activity and phylogenetic relationships can facilitate the identification of novel candidates for the biotechnological production of high-value apocarotenoids.

1. Introduction

Carotenoid biosynthesis in plants is governed by a network of biochemical pathways, many of which have been extensively characterized [1]. Carotenoid cleavage dioxygenases (CCDs) represent a major family of enzymes involved in carotenoid catabolism, thereby modulating carotenoid levels and influencing a wide range of downstream physiological and developmental processes. Some major subfamilies of CCD genes CCD1, CCD2, CCD4, CCD7, CCD8, and CCD10 have been identified. Among these, CCD7 and CCD8 are involved in the regulation of lateral shoot branching [2], CCD1 contributes to apocarotenoid degradation [3], CCD2 to crocetin biosynthesis in monocots [4–6], and the CCD10 subgroup incudes the zaxinone synthase enzyme, involved in growth and strigolactone biosynthesis in rice [7]. In comparison with these CCDs subfamilies, the CCD4 subfamily comprises enzymes involved in a broad range of processes, many associated with the

* Corresponding author at: Instituto Botánico, Departamento de Ciencia y Tecnología Agroforestal y Genética, Universidad de Castilla-La Mancha, Campus Universitario s/n, Albacete 02071, Spain.
E-mail address: marialourdes.gomez@uclm.es (L. Gómez-Gómez).
[1] Both authors contribute equally to this work.

https://doi.org/10.1016/j.cpb.2026.100599
Received 11 December 2025; Received in revised form 17 February 2026; Accepted 23 February 2026
Available online 28 February 2026

coloration dynamics during flower development or fruit maturation, and the biosynthesis of volatile aroma compounds such as β-ionone, pseudoionone, and geranylacetone [8–10]. In *Crocus*, *Osmanthus*, potato, sweet potato, *Cucumis*, *Chrysanthemum*, and apple, CCD4 catalyzes the oxidative cleavage of β-carotene at the C9-C10 and/or C9′-C10′ double bonds, yielding β-ionone [11–13]. In contrast, CCD4 enzymes from other species, such as *Citrus*, *Buddleja davidii*, *Gardenia jasminoides*, and *Nycthantes arbor-tristis*, exhibit alternative cleavage sites, C7-C8 and C7′-C8′, on carotenoids like zeaxanthin and β-cryptoxanthin, lycopene or β-carotene, producing apocarotenoids including β-citraurin, crocetin dialdehyde and 3-OH-cyclocitral [4,5,14,15–18].

Despite the known role of CCDs in many plant species, their functional significance in trees largely remain to be elucidated [19]. The genus *Paulownia*, belonging to the family *Paulowniaceae*, comprises nine recognized species [20], of which *Paulownia tomentosa*, *P. elongata,* and *P. fortunei* are the most extensively cultivated. Native to China and Southeast Asia, *Paulownia* species are widely appreciated for their rapid growth, ease propagation, and broad environmental adaptability. These trees have gained global attention for reforestation, biomass production, and hardwood timber industries due to their high wood yield and short rotation cycles [21].

Genome analyses of *P. fortunei* [22], revealed the notable expansion of the *CCD4* gene family, with 12 *CCD4* homologs identified, of which eight are tandemly duplicated on chromosomes 7 and 18 [23]. Among these 12 *Paulownia CCD4* genes, four homologous genes were expressed in flowers of *P. tomentosa*, and one gene, *CCD4–4*, was identified as recognizing and cleaving β-carotene at the C9-C10 and/or C9′-C10′ positions, producing β-ionone, playing a dual role in pollinators attraction through scent and visual signaling [23]. In this study, we cloned and characterized the *P. tomentosa CCD4–2* gene, which was expressed along the development of the flowers at relatively low levels, compared with *CCD4–4*, which showed higher levels of expression in the bud and at anthesis [23]. Interestingly, phylogenetic analysis of the CCD4–2 amino acid sequence revealed that clusters with CCD enzymes involved in crocetin biosynthesis in other species of the *Scrophulariaceae* family. The evaluation of CCD4–2 activity in bacteria and plants demonstrated its capacity of crocetin dialdehyde biosynthesis, although this apocarotenoid was not identified in the flowers of *P. tomentosa*. Finally, we identified the presence of *CCD4* genes in the genome of *P. tomentosa* and compare its identity and distribution with those previously identified in *P. fortunei*.

2. Materials and methods

2.1. Cloning, and expression of CCD4-2

Total RNA was isolated from *Paulownia* flowers and leaves obtained from the Jardin Botánico de Castilla-La Mancha (Albacete, Spain), using the Trizol Reagent (Thermo Fisher, Spain), followed by on-column DNaseI treatment and purification with the Direct-Zol RNA micro-prep kit (Zymo Research, USA). The obtained RNA was used for the SMARTer® RACE 5′/3′ Kit (Takara, Japan) following the manufacturer's specifications, and the specific oligonucleotides depicted in Table S1. Amplified *CCD4.2* fragments were purified using Gel Band Purification Kit (Promega, USA) and inserted in the pGEM-T vector (Promega, USA). After transformation, positive colonies were cultured for miniprep preparation (BioTools, Spain) and the obtained plasmids were sent to Macrogen Inc (South Korea) for sequencing. Full-length *CCD4–2* (GenBank ID: OZ253763.1) was amplified from 5 μL of cDNA by a High-Fidelity DNA Polymerase (Takara, Japan) according to the manufacturer's instructions and using specific primers (Table S1). Further, CCD4–2 was inserted into the vector pTHIO-dam1, to yield the plasmids pTHIO-CCD4–2. The integrity of the construction was verified by sequencing. Expression analyses of *CCD4–2* were performed using the specific oligonucleotides depicted in Table S1.

2.2. Phylogenetic and structural analyses analysis

Multiple amino acid sequence alignments and phylogenetic tree construction were done using the MEGA 11.0.10 program (https://megasoftware.net/). Phylogenetic analyses were done with the Maximum Parsimony method with 2500 bootstrap replicates.

Subcellular localization predictions were carried out using DeepLoc-1.0 (https://services.healthtech.dtu.dk/services/DeepLoc-1.0/). The 3D structure was predicted using the Phyre2 software (http://www.sbg.bio.ic.ac.uk/phyre2/), Chimera X (https://www.cgl.ucsf.edu/chimerax/), and COACH (https://seq2fun.dcmb.med.umich.edu/COACH/). Tunnels prediction was done using MOLE (https://webchem.ncbr.muni.cz/Platform/App/Mole).

2.3. In vivo enzyme assays in E.coli

BL21 *E. coli* cells were transformed with different PAC vectors purchased from Addgene (https://www.addgene.org/): PAC-ZETA for ζ-carotene biosynthesis, PAC-LYC for lycopene biosynthesis, PAC-BETA for β-carotene biosynthesis, PAC-ZEAX for zeaxanthin biosynthesis, and PAC-DELTA for δ-carotene biosynthesis. After the initial transformation and selection, the confirmed positive bacterial colonies were propagated to prepare chemically competent cells. These competent cells were then transformed with the pTHIO-CCD4–2 construct, and with the empty pTHIO vector, used as a negative control. Cells were grown, induced and extracted according to previous methodology [24,25].

2.4. Analytical methods for carotenoids and nonvolatile apocarotenoids

HPLC-DAD analysis was performed with an Agilent system (Agilent Technologies, Germany) equipped with a photodiode array detector (model 996). A C18-reversed phase column (Kromasil 100–5-C18 4.6 ×150 mm) was used for substrates and products analyses of activity assays in bacterial cells. The solvents used were A: methanol/water/tert-butylmethylether 60:20:2 (v/v/v), B: methanol/tert-butylmethylether 50:50 (v/v/v), and C: 100% tert-butylmethylether with water containing 0.1 g/L ammonium acetate. The flow rate was 1 mL/min with 95% A and 4% B and held for 3 min, followed by a ramp to 10% A and 90% C in 19 min, followed by a ramp to 100% C in 8 min, 10 min with 100% C, and return to initial conditions for 5 min. The extraction and analyses of crocins from infected leaf tissue was done as previously described [26]. While for carotenoid and chlorophylls analyses a C30-reversed phase column (YMC Europe, Schermbeck, Germany) was used and the methodology followed has been already described [26].

2.5. HPLC-HRMS analysis of apocarotenoids in Paulownia flowers

HPLC-HRMS of *P. tomentosa* leaves was performed as previously described [27]. Slight adjustments were made to the heated electrospray ionization (HESI) settings. Specifically, nitrogen was employed as both the sheath and auxiliary gas, set at 35 and 25 units, respectively. The vaporizer and capillary temperatures were each maintained at 280 °C. Additional parameters included a discharge current of 4.5 μA and an S-lens RF level of 50. Data acquisition was performed in both positive and negative ionization modes across a mass range of 110–1600 m/z, using the following settings: resolution of 70,000, one microscan, an AGC target of 1×10^6, and a maximum injection time of 50 ms. The sets for the identification of crocins, and their related compounds were conducted as previously described [28,29].

2.6. Analytical methodology for volatile capture and analysis

Volatile capture and analyses were done following a previous methodology paper [25]. In brief, following arabinose induction of bacterial cultures, volatile compounds were extracted after 24 h using headspace sorptive extraction. Samples were incubated under controlled

conditions, and extracted analytes were subsequently prepared for GC-MS analysis. Thermal desorption of stainless-steel tubes containing the stir bars was performed using a TurboMatrix Automated Thermal Desorber (PerkinElmer, Norwalk, CT, USA). GC-MS analyses were performed using a ThermoFisher Scientific DSQ II mass spectrometer coupled to a Trace GC gas chromatograph equipped with a 30 m Zebron ZB 5 column (5% phenyl–95% dimethylpolysiloxane, 0.25 mm I.D. and 0.25 μm film thickness; Phenomenex, Aschaffenburg, Germany). Compounds were identified by comparison of mass spectra with the NIST database (National Institute of Standards and Technology Mass Spectral Search Program Version 2.0) and by using synthetic standards.

2.7. Activity assays in N. benthamiana

Plasmid pGTEV-CCD4–2 was constructed by PCR amplification of cDNA of *P. tomentosa CCD4–2* (Table S1) and Gibson DNA assembly at the N-t end of the viral polyprotein. The pGTEVa contains a wild-type TEV cDNA (GenBank DQ986288, with 2 silent and neutral mutations G273A and A1119G) flanked by the cauliflower mosaic virus (CaMV) 35S promoter and terminator in a binary vector derived from pCLEAN-G181 [26]. The plasmid used as control (pGTEV-GFP) express the recombinant TEV-GFP clone was previously described [18].

Agrobacterium tumefaciens strain C58C1, harboring the helper plasmid pCLEAN-S48, was independently transformed via electroporation with plasmids pGTEV-CCD4–2, and pGTEV-GFP. Transformed cells were selected on agar plates containing 50 μg/mL kanamycin, 50 μg/mL rifampicin, and 7.5 μg/mL tetracycline. Liquid cultures from selected colonies were prepared and used to infiltrate leaves of one-month-old *N. benthamiana* plants. Following infiltration, plants were maintained in a growth chamber under controlled conditions (25°C, 16 h light/8 h dark photoperiod). Leaf tissue was harvested 14 days post-inoculation (dpi) and analyzed for crocin and carotenoid content.

2.8. Cellular localization

The C-terminal region of CCD4–2 was fused to the N-terminus of eGFP in the pBI-eGFP vector [30] using the In-Fusion® HD Cloning Plus CE kit (Clontech, www.clontech.com) and gene-specific primers (Table S1). The integrity of the construct was verified by sequencing prior to transformation into *A. tumefaciens* C58C1. Transient expression was carried out via agroinfiltration of *N. benthamiana* leaves. 5–7 days post-infiltration the agroinfiltrated leaves were analyzed using a confocal laser scanning microscope. Fluorescence signals were detected using 488 nm (argon) and 635 nm (diode) lasers for eGFP (green fluorescence, 488 nm excitation / 505–540 nm emission) and chlorophyll auto-fluorescence (red, 660–750 nm emission), respectively. Images were acquired in xyz scan mode at a resolution of 800 × 800 pixels using a 60 × objective lens (numerical aperture 1.35) with 3 × optical zoom.

2.9. Synteny analysis

Genomic regions showing homology to previously characterized CCD4 sequences were identified in *P. tomentosa* and *P. fortunei* using BLASTn. The retrieved regions were subsequently annotated with Augustus to predict gene models [31]. Syntenic relationships between the corresponding chromosomal segments (chromosomes 6 and 19 of *P. tomentosa* and chromosomes 19 and 18 of *P. fortunei*) were visualized using the clinker software [32], which connects homologous genes based on amino acid identity. Genome assemblies for *P. fortunei* (GCA_019321725.1) and *P. tomentosa* (GCA_965263725.1) were retrieved from NCBI and analyzed using SynMap2 on the CoGe platform (accessed 9 October 2025). Whole-genome pairwise comparisons were performed with SynMap2's default workflow, which computes nucleotide alignments and identifies collinear (syntenic) blocks via DAGChainer. Resulting macrosynteny dot plots and block lists were exported directly from SynMap2 for visualization and downstream interpretation.

2.10. Expression analysis

RNA-seq reads from leaves and flower transcriptomes were aligned to the *P. fortunei* reference genome using HISAT2 with default parameters [33]. The reads mapping to the previously identified *CCD4*-containing regions were quantified using Cufflinks [34], and expression levels were normalized and expressed as FPKM (Fragments Per Kilobase of transcript per Million mapped reads) in $\log_2$ scale. The resulting expression matrix was visualized as a heatmap to compare *CCD4* expression in leaves and flowers at anthesis.

3. Results

3.1. Phylogenetic analyses and cleavage activity in E.coli cells

We performed a phylogenetic analysis of *P. tomentosa* CCD4–2 with other CCD4s with known and characterized activities in different plant species. This analysis indicates that CCD4–2 belongs to a specific clade, together with three previously characterized CCDs, catalyzing the C7-C8 and C7′-C8′ cleavage of zeaxanthin and/or β-carotene (Fig. 1). BdCCD4.1 and BdCCD4.3, from *B. davidii* were previously shown to have activity on zeaxanthin [4], while *V. giganteum* VgCCD4.1 recognized zeaxanthin and β-carotene as substrates for crocetin biosynthesis [35]. To determine the cleavage activity of *P. tomentosa* CCD4–2, its full-length cDNA was obtained from flower tissues using sequence information obtained in a previous work [23]. CCD4–2 contains 565 amino acids and showed 98% identity with a *P. fortunei* protein (KAI3468520.1), 73.7% with *B. davidii* CCD4.1 (APU54674.1), and 70.0% with *V. giganteum* CCD4.1 (WQF66837.1). *In silico* analyses of CCD4–2 suggested a plastid subcellular localization. To confirm this predicted location, we performed transient expression analyses in leaves of *Nicotiana tabacum* with CCD4–2 C-terminally fused to the green fluorescent protein GFP (Fig. 2A). Confocal microscopy images showed the location of CCD4–2 in plastids, confirming the *in silico* prediction. Because of the lipophilic nature of the substrates used by the CCDs, we investigated the possible affinity of CCD4–2 to the membrane bilayer. To predict membrane interaction, we used Position of Proteins in Membrane (PPM 3.0) web server. The prediction resulted in numerical values assigned to each surface residue representing the likelihood of that residue being involved in a membrane interaction. A favorable transfer energy ($\Delta G_{transfer} = -9.7$ kcal/mol) was calculated and lipid contacts were assigned to hydrophobic residues 56–57,60,189,192–193,195–199,203, 281,330,333 present in the helical "dome" that can penetrate the lipid bilayer core up to a depth of 7.5 ± 0.4 Å (Fig. 2B). Next, we searched for the presence of substrate-enter tunnels in CCD4–2 structure using the MOLE 2.5 online program. Three main tunnels were predicted, which are surrounded by the membrane-binding residues (Fig. 2B). The position of the iron cofactor is shown in Fig. 2C.

Based on the close phylogenetic relationship between CCD4–2 and *B. davidii* CCD4.1 and *V. giganteum* CCD4.1, we hypothesized that CCD4–2 would exhibit the same C7-C8 and C7′-C8′ cleavage regioselectivity. To test this hypothesis, we expressed CCD4–2 in *E. coli* producing β-carotene and zeaxanthin and measured its activity. As predicted, CCD4–2 showed cleavage C7-C8 and C7′-C8′ activity (Fig. 3), but only β-carotene was recognized as substrate (Fig. 3A and B). In addition, the volatiles from the head space of the cultures of *E. coli* cells producing β-carotene and expressing CCD4–2 were collected using Stir-Bar Sorptive Extraction (SBSE) and analyzed by GC-MS. As shown in Fig. 3C, the GC-MS analysis revealed the formation of β-cyclocitral, confirming the cleavage of the C7-C8 and C7′-C8′double bonds of β-carotene.

Further, acyclic carotenoids as ζ-carotene and lycopene were tested as substrates (Fig. S2), but they were not recognized by CCD4–2. To delimit more the substrate preference of CCD4–2, the asymmetric carotenoid δ-carotene, was selected for further assays, but no products were detected in the activity assays (Fig. S2). Taken together, the results

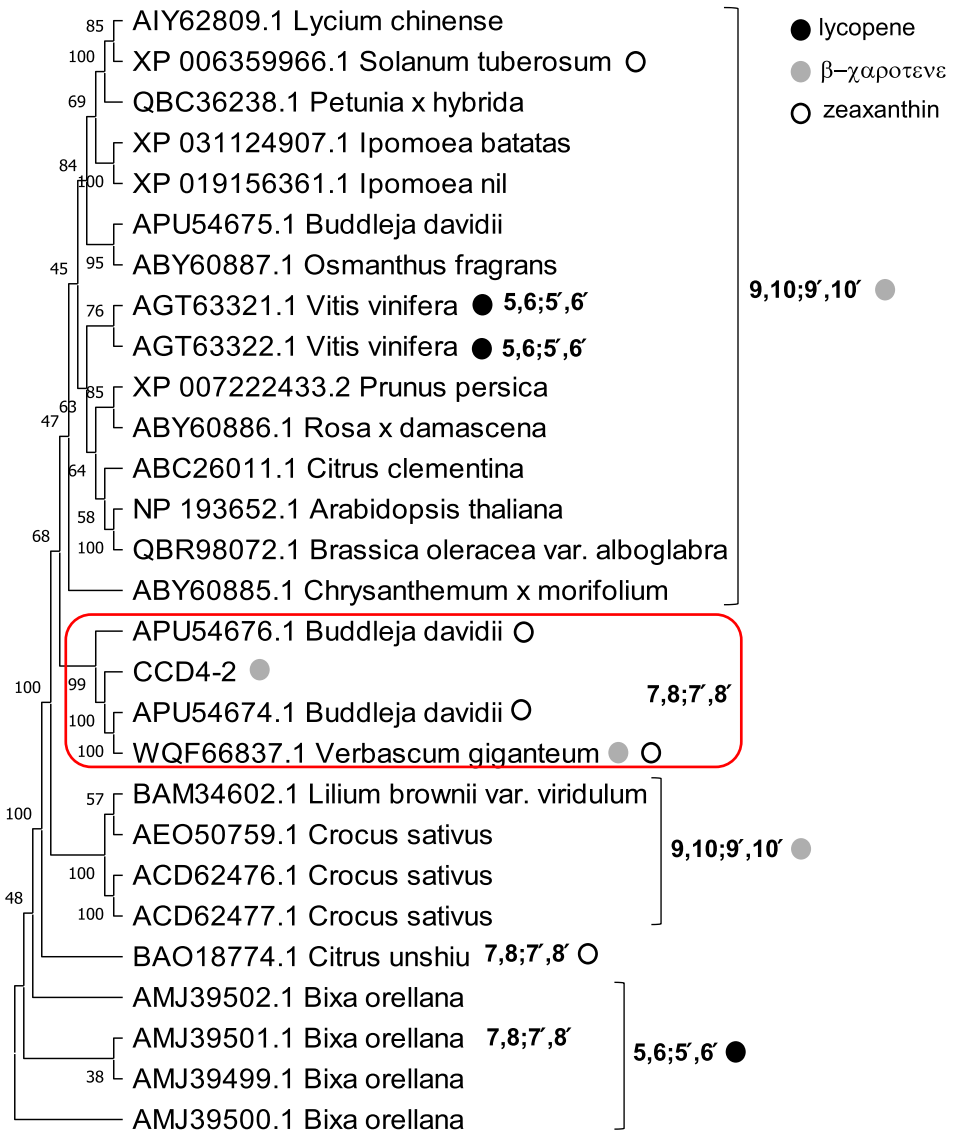

Fig. 1. Phylogenetic relationship of CCD4 enzymes from different plant species, with characterized cleavage activity. The phylogenetic tree was constructed by the Maximum Likelihood method using 2500 replications of bootstrap analysis. The number on the nodes corresponds to the percentage of bootstrap values. The sub-tree containing the CCD4–2 sequence is highlighted in red. The cleavage sites are indicated on the right side of the tree. Black, grey, and white dots represent the recognized substrates.

obtained showed a clear preference of CCDD4–2 for β-carotene among the different carotenoid substrates tested.

3.2. *Characterization of the enzymatic activity of CCD4-2 in heterologous plant systems*

To determine whether the observed activity in *E. coli* was also reproducible in a plant system, we transiently expressed CCD4–2 with a viral vector. Using the previously developed tobacco etch virus (TEV)-based expression system for CsCCD2L [26], we investigated whether, through the heterologous expression of CCD4–2 in *Nicotiana benthamiana,* crocin biosynthesis can be achieved. The cDNA encoding CCD4–2 was inserted at the N-t of the viral polyprotein to generate TEV-CCD4–2 (Fig. 4A). Further, an artificial NIaPro recognition sequence was incorporated at the 3′ terminus of the cloning site to facilitate proteolytic release of CCD4–2 from the viral polyprotein. Following our previous work, we used the standard −8/+ 3 NIaPro cleavage site that naturally separates NIb and CP in TEV, but including silent mutations to avoid recombination during viral vector replication. The resulting recombinant construct was then introduced into *N. benthamiana* plants via agroinoculation. As control, plants were agroinoculated with TEV-GFP (negative control). Inoculated plants (both with TEV-CCD4–2 and TEV-GFP) exhibit viral infection symptoms approximately 8 dpi. By 14 dpi, leaves from TEV-CCD4–2 infected plants displayed yellow spotting, suggestive of the presence of crocins (Fig. 4B). Chromatographic analysis of apolar extracts, which include carotenoids and chlorophylls revealed a general decrease in these pigments across all virus-infected samples carrying CCD4–2 (Fig. 4C and D). Polar extracts from collected infected leaves from TEV-CCD4–2 showed the presence of crocins, while these soluble apocarotenoids were not detected in the polar extracts from TEV-GFP infected leaves. The levels of crocins reached 3.89 ± 0.95 mg/g DW. Thin-layer chromatography (TLC) showed the presence of crocins in the analyzed extracts compared with the control plants (Fig. 4E). Overall, these results demonstrate the successful accumulation of crocins in *N. benthamiana* upon expression of CCD4–2, indicating the functional role of CCD4–2 in plants as a C7-C8 and C7′-C8′ cleavage dioxygenase. These results also suggested that this cleavage activity could be widespread in dicotyledonous plants,

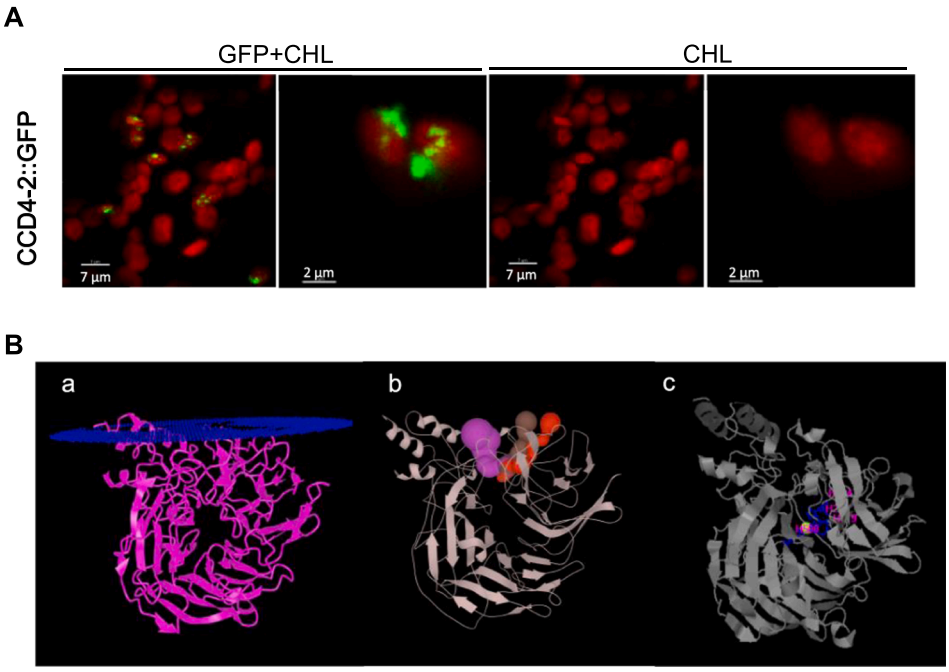

Fig. 2. Fig. 2. CCD4–2 intracellular location and structural analyses. A) Intracellular localization of CCD4–2::GFP in *Nicotiana benthamiana* leave cells using confocal microscopy. GFP is in green, plastid autofluorescence (CHL) is in red. All images are single confocal section. B) *In silico* structural analyses of CCD4–2. In a is shown the interaction of CCD4–2 with the membrane, based on the results obtained with PPM 3.0 (https://opm.phar.umich.edu/ppm_server3_cgopm) the membrane is represented as a blue grid. In b are shown predicted substrate-access tunnels (pink, brown and red) in CCD4–2 obtained with MOLE (https://mole.upol.cz/). In c are shown the histidine (H) residues involved in the interaction with the iron cofactor (in green).

although in just a few species, including *Buddleja*, gardenia, *Verbascum*, and *N. arbor-tristis* crocetin has been reported [35–38].

3.3. Absence of crocins in flowers and leaves extracts of P. tomentosa

We previously evaluated the carotenoid content of *P. tomentosa* flowers in different developmental stages [23], where crocetin was not detected. Polar extracts of flowers and leaves were obtained and analyzed for the presence of crocins (Fig. S3). The extract of flowers mainly showed the presence of phenolic compounds and flavonoids. Apigenin and quercetin have been reported before as main flavonoids in flowers [39]. In the polar extracts of leaves, flavonoids and phenolic compounds were also detected (Fig. S3). The leaves of paulownia contain several bioactive compounds as iridoids, phytosterols, flavonoids and triterpenoids [20]. In addition, we analyzed the apolar extracts from leaves to determine the presence of crocetin and other carotenoids by HPLC-DAD (Fig. S4). Crocetin was not detected, but lutein and β-carotene were the most abundant carotenoids.

3.4. The CCD4 family in the genome of P. tomentosa and synteny analyses

We finally analyzed the recent sequenced genome of *P. tomentosa* (daPauTome1.hap1.1) for chromosome location of *CCD4–2*, and the determination of its genomic structure and presence of related *CCD4* sequences. The ccd4–2 gene was located on chromosome 19 and contained a single intron of 99 bp (Fig. 5A). Blastn analyses on this chromosome using *CCD4–2* as bait, allowed the detection of four additional *CCD4*-like encoding genes, all containing a single intron, surrounding the CCD4–2 gene (Fig. 5A), expanding a total region of 52.938 bp. The identities at the nucleotide level between *CCD4–2* and the other *CCD4*-like genomic sequences range from 75% to 67%. Further, the whole genome was searched for related *CCD4* sequences, and five additional genes were identified, all of them located in chromosome 6 (Fig. 5A). In order to further understand the evolutionary origins and orthologous relationship of *CCD4* genes, a microsynteny analysis was performed between *P. Tomentosa* and *P. fortunei (assembly GCA_019321725.1)* (Fig. 5A). The analysis revealed a complex but well-conserved pattern of collinearity among *CCD4* gene family members in *P. fortunei* and P. *tomentosa* (Fig. 5A). A total of ten *CCD4* homologs (*CCD4.1-CCD4.10*) were identified across several chromosomal segments in *P. tomentosa*, whereas a slightly smaller number of syntenic copies were detected in *P. fortunei*, suggesting partial gene loss/gain following species divergence. The gray ribbons connecting homologous regions varied in color intensity, reflecting differences in sequence conservation. Darker ribbons indicated regions of higher nucleotide similarity, representing more conserved syntenic blocks, while lighter ribbons corresponded to more divergent or partially aligned genomic segments. Several *CCD4* genes (*CCD4.1, CCD4.3, CCD4.9,* and *CCD4.10*) were linked by multiple dark ribbons, indicating strong conservation and the presence of multiple homologous anchor genes in the corresponding genomic regions of *P. fortunei*. This pattern suggests that these *loci* are evolutionarily stable and may encode functionally important isoforms. In contrast, *CCD4.5*, *CCD4.7* and *CCD4.8* displayed fewer and lighter syntenic connections, implying greater sequence divergence or the loss of corresponding regions in *P. fortunei*, possibly due to post-speciation deletions, or lineage-specific pseudogenization. The strong and multiple connections between *CCD4.2*, *CCD4.4*, and *CCD4.5* highlight regions of segmental duplication, where duplicated blocks within *P. tomentosa* correspond to a single or fewer counterparts in *P. fortunei*. The consistent gene orientation and conserved gene order across most syntenic blocks underscore the evolutionary stability of these loci within the *Paulownia* lineage. *CCD4.1*, *CCD4.5*, *CCD4.9* and *CCD4.10* are connected to *P. fortunei* by multiple dark ribbons, indicating (1) highly conserved orthologous sequences and (2) the presence of multiple homologous anchor genes in corresponding regions of *P. fortunei*.

Further, the expression of all the *CCD4* genes from *P. tomentosa* was evaluated in flowers at anthesis and in leaves using transcriptome data. All the genes except for *CCD4–5* and *CCD4–7* were expressed in flowers, and in leaves only were detected *CCD4–3, CCD4–4* and *CCD4–5* (Fig. 5),

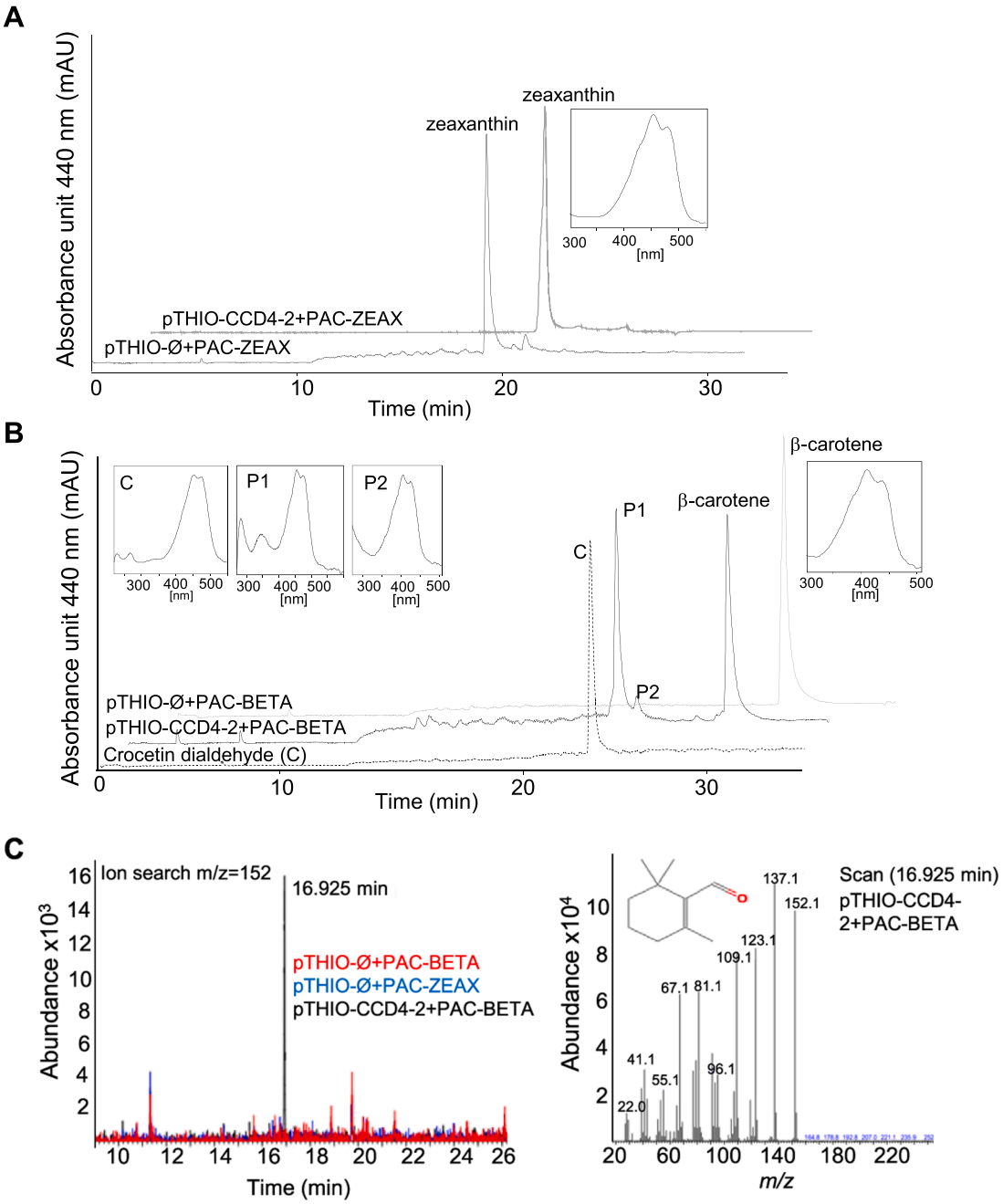

Fig. 3. Functional expression of CCD4–2 in zeaxanthin and β-carotene-producing *E. coli* cells. A) HPLC-DAD analysis CCD4–2 in bacterial cultures producing zeaxanthin. Inset is shown the zeaxanthin spectra. B) HPLC-DAD analysis of β-carotene cleaved by CCD4–4 in bacterial cultures. Inset are shown the spectra of the pigments extracted from each bacterial pellet. C) GC/MS analysis of products synthesized by CCD4–2 cleavage activity during bacterial growth, volatiles were collected by Stir-bar Sorptive Extraction (SBSE) from the gas space of *E. coli* cultures producing zeaxanthin and β-carotene and analyzed by GC-MS. CCD4–2 produced β-cyclocitral in cell cultures of *E. coli* producing β-carotene. β-cyclocitral was identified by comparing the detected molecule ion of 152 and detected fragments with the NIST library.

with variable expression levels.

4. Discussion

The present study investigates the phylogenetic relationships and *in vivo* activities of ccd4–2 of the tree *P. tomentosa*. CCD4–2 catalyzes C7-C8 and C7′-C8′ oxidative cleavage of β-carotene, resulting in the formation of products of biological and economic importance [40]. Several studies have shown that the CCD4 enzymes from plants catalyzed the cleavage at different double bonds positions from cyclic and acyclic carotenoid substrates [19]. The genome of *P. fortunei* showed the presence of 8 *CCD4* genes [23], while 10 has been found in the genome of *P. tomentosa*. The *CCD4–2* homolog from *P. fortunei* was in chromosome 18 and surrounded by additional 3 *CCD4*-like encoding genes [23]. In the genome of *P. tomentosa*, *CCD4–2* was located on chromosome 19, and surrounded by 4 *CCD4*-like encoding genes, including the initially characterized *CCD4–4* gene [23] suggesting tandem duplications in this area. Additional *CCD4* homologs were detected on the chromosome 6 of *P. tomentosa*, and on the chromosome 7 of *P. fortunei*. Interestingly, in *P. fortunei* the chromosome 7 contains 73.5% of the homologous genes present on chromosome 17, and 64.7% of the homologous genes from chromosome 18 [22]. *Paulownia* has experienced two rounds of whole

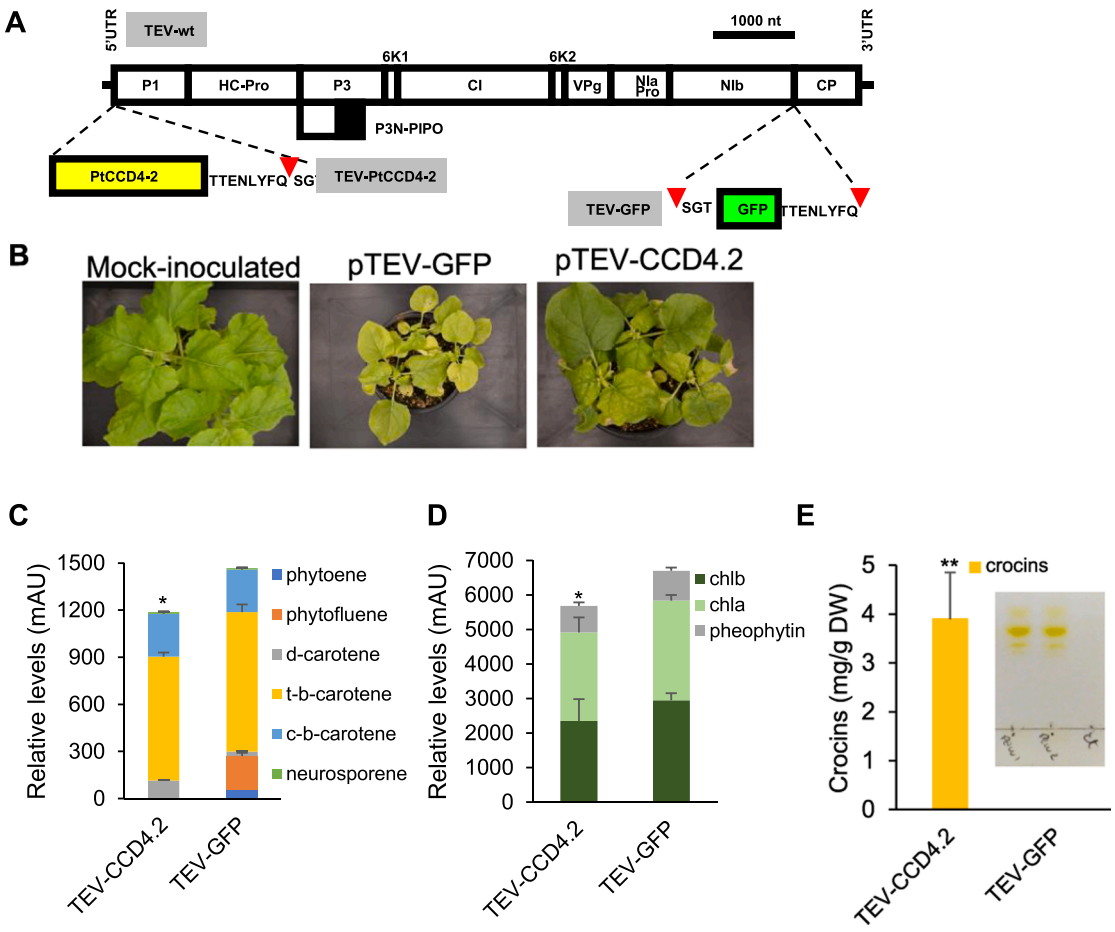

Fig. 4. Crocins accumulation in *Nicotiana benthamiana* following expression of CCD4–2 via a recombinant TEV vector. A) Schematic diagram of the TEV genome showing the insertion sites of GFP (green box, control vector TEV-GFP) or PtCCD4–2 (yellow box, vector TEV-CCD4–2). Artificial NIaPro cleavage sites used to release the recombinant proteins from the viral polyprotein in both cases are indicated, with the red arrow marking the cleavage site. TEV cistrons (P1, HC-Pro, P3, P3N-PIPO, 6K1, CI, 6K2, VPg, NIaPro, NIb, and CP) are shown as boxes; lines represent the 5′ and 3′ UTRs. Scale bar = 1000 nucleotides. B) Representative leaf images from *N. benthamiana* plants that were mock-inoculated, agroinoculated with TEV-GFP and TEV-CCD4–2, photographed at 14 days post-inoculation (dpi). Scale bars = 5 mm. C) Carotenoid composition and relative levels of TEV-CCD4–2 infected leaves and TEV-GFP infected leaves, used as control. D) Chlorophyll relative levels of TEV-CCD4–2 infected leaves and TEV-GFP infected leaves, used as control. E) Crocins content of TEV-CCD4–2 infected leaves and TEV-GFP infected leaves, used as control. Inset is shown the TLC analysis of the polar extracts containing crocins in TEV-CCD4–2 infected leaves. Significant difference (*P < 0.05, **P < 0.01).

genome duplication (WGD) [22]. The expansion and organization of the *CCD4* gene family in *Paulownia* reflect the combined effects of WGD, and local tandem duplication. The detection of *CCD4* homologs on chromosomes 6 and 7, including the strong syntenic relationships between chromosome 7 and chromosomes 17 and 18 in *P. fortunei*, and chromosome 6 with chromosomes 18 and 19 (Fig. S5) supports the contribution of ancient WGD in shaping the chromosomal landscape. Following WGD, diploidization processes typically lead to gene loss and genomic fractionation; however, the persistence of numerous *CCD4* paralogs suggests selective retention, likely due to their functional roles in carotenoid metabolism and potential contributions to adaptive phenotypes such as pigmentation and volatile formation. In addition to WGD-derived copies, the presence of *CCD4–2* surrounded by several *CCD4*-like genes in both species clearly indicates recent tandem duplication events, which further expanded the family in a lineage-specific manner. These tandem arrays likely allowed fine-tuned or tissue-specific regulation. Together, the interplay of ancient genome duplications, subsequent diploidization, and localized gene amplification has produced a complex *CCD4* repertoire in *Paulownia*, potentially underpinning biochemical and phenotypic diversification within the genus.

Phylogenetic analyses showed CCD4–2 closely related to CCD4 enzymes from *Buddleja* and *Verbascum* (73.7% and 70.0% identity, respectively) that are involved in crocetin biosynthesis through C7-C8 and C7′-C8′ double bonds cleavage, suggesting a similar activity for CCD4–2. As the *Verbascum* and *Buddleja* enzymes, CCD4–2 is localized in plastids, and *in silico* analysis suggested its interaction with membranes allowing substrate availability [41], interacting with membrane-bound pools of β-carotene. *Verbascum* and *Buddleja* enzymes showed activity on β-carotene and/or zeaxanthin and in both plants crocins accumulate in the flower tissues [35,42]. The activity of CCD4–2 was tested over different carotenoid substrates, being unable to recognize linear carotenoids with C7-C8 and/or C7′-C8′ double bonds, such as lycopene, this last one is substrate for the GjCCD4a enzyme from *Gardenia*, which also recognized zeaxanthin and β-carotene. The asymmetric carotenoid δ-carotene was not recognized, and from the symmetric cyclic carotenoids only β-carotene was cleaved by CCD4–2. This carotenoid has been detected in the flowers of *Paulownia*, where is also substrate of additional CCDs, including CCD1 and CCD4–4 [35]. In this study, β-carotene was identified as the predominant carotenoid in the leaves, alongside lutein. However, similar to the observations in flowers, crocins were not detected, suggesting several possible explanations. One scenario is that *Paulownia* may contain competing metabolic pathways, enzyme repressors, or rapid degradation mechanisms that prevent the accumulation or detection of crocins under natural conditions. Alternatively, other CCD enzymes co-expressed with CCD4–2 could be responsible for

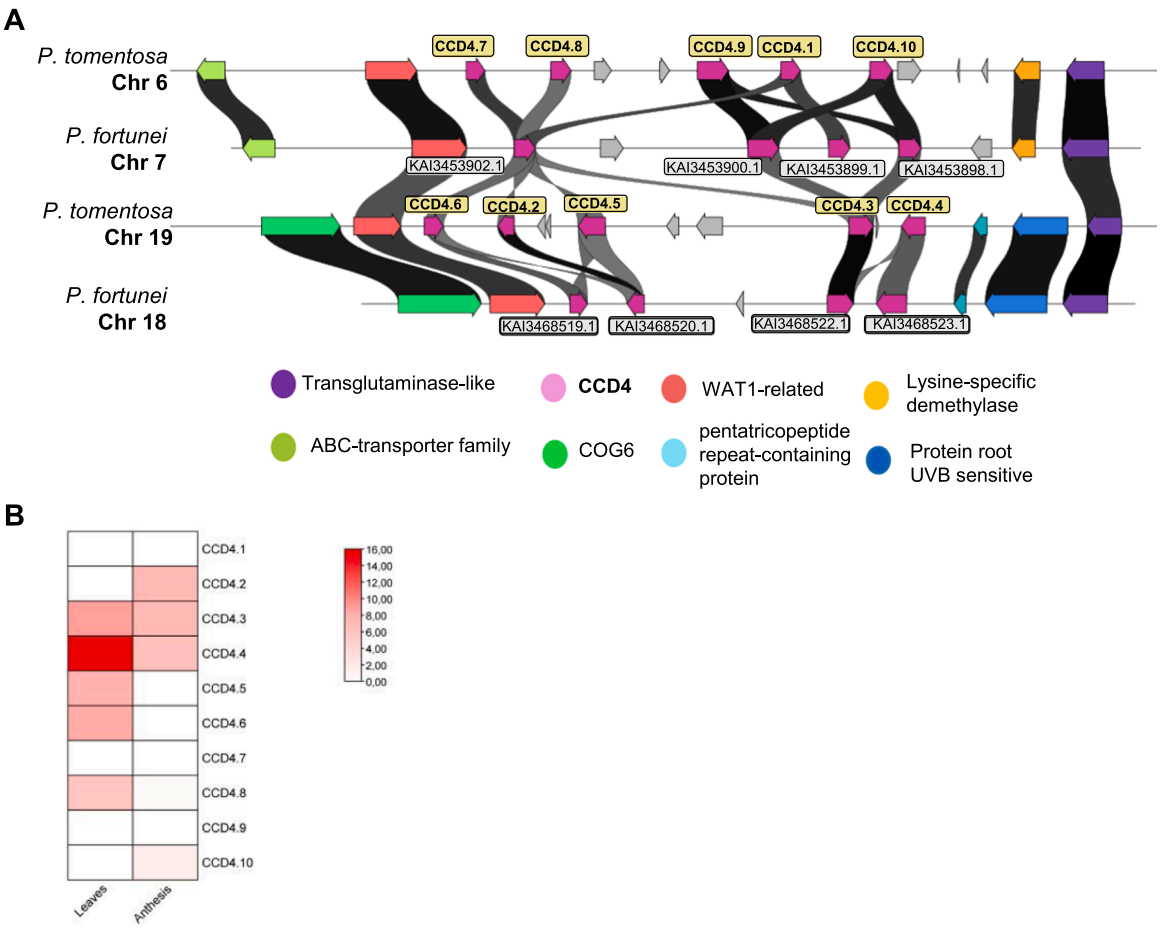

Fig. 5. Synteny and expression analysis of *CCD4* genes in *Paulownia tomentosa* and *Paulownia fortunei*. A) Two syntenic regions are shown: chromosomes 6 and 19 of *P. tomentosa* aligned with chromosomes 7 and 18 of *P. fortunei*. Colored blocks represent predicted genes (color legend on the right), with *CCD4* family members highlighted in yellow. Ribbons connect homologous genes sharing > 65% amino acid identity, and homologs annotated in *P. fortunei* are indicated in gray. B) Heatmap showing *CCD4* transcript levels (FPKM, $\log_2$ scale) in leaves and flowers at anthesis, illustrating tissue-specific differential expression patterns.

the cleavage of crocetin or the substrate β-carotene, thereby diverting the pathway away from crocin production [23]. One possibility is the presence of CCD1, which has been implicated in maintaining carotenoid homeostasis through the cleavage of apocarotenoids [3], and can be acting on crocetin. Alternatively, other members of the CCD4 family could be contributing to this process. At least CCD4–3, CCD4–4, CCD4–8 and CCD4–10 are expressed in flowers. This last possibility may explain why heterologous expression of CCD4–2 in *N. benthamiana*, a species encoding only two CCD4 enzymes in the genome, resulted in the accumulation of crocins. The reduced competition for substrates or lack of degradation of the product might facilitate the accumulation of crocins and their detection.

Previously, a viral expression system was developed for the production of crocins using the CsCCD2L enzyme from saffron [26]. This system has since been successfully adapted to express other crocins biosynthetic enzymes derived from *Verbascum* [35] and *Crocosmia x crocosmiiflora* [5], demonstrating its versatility. The accumulation of crocins at several milligrams per gram of DW (3.89 ± 0.95) indicates that CCD4–2 introduces a productive metabolic branch point exceeding or similar to the efficiencies previously reported using this expression system. The enzyme from *B. Davidii* BdCCD4.1 (0.75 ± 0.02 mg/g DW), that from saffron CsCCD2L (2.18 ± 0.33 mg/g DW) [26], from *V. giganteum* VgCCD4.1 (1.78 ± 0.11 mg/g DW) [35], montbretia (4.37 ± 0.07 mg/g DW) [5] and from *N. arbor-tristis* (2.32 ± 0.69 mg/g DW) [18]. This outcome suggests that the viral platform not only supports rapid expression but also provides a metabolic environment in which CCD4-type enzymes remain catalytically active and well supplied with carotenoid substrates. The magnitude of crocin accumulation strengthens the case that transient viral systems can serve as practical biotechnological tools for generating specialty apocarotenoids at levels meaningful for downstream purification or functional testing. The demonstrated efficiency with CCD4–2 suggests that additional apocarotenoid targets could be pursued using the same approach, offering a scalable route for producing compounds that are otherwise difficult to obtain in quantity from natural sources.

In summary, the results obtained in this work highlight two key points. First, CCD enzymes with known catalytic activity, established through phylogenetic and biochemical characterization, can be harnessed to drive apocarotenoid production in heterologous hosts. CCD4-2 demonstrates this principle, joining a growing list of CCD4 family members that act on β-carotene to produce crocins when expressed in *N. benthamiana*. Second, the plant genome evidently harbors a reservoir of catalytically valuable CCD enzymes. Phylogenetic inference, positioning CCD4-2 alongside crocin-synthesizing CCD4s, and functional assays together inform targeted screening and bioprospecting of plant CCD gene families. The prior identification of enzymatic activity in related CCDs thus becomes a powerful predictor in selecting novel candidates for metabolic engineering. This approach accelerates the discovery of new CCDs for biotechnological applications, particularly in producing high-value soluble apocarotenoids, such as crocins and picrocrocin. The convergence of enzyme characterization and metabolic engineering opens new avenues for sustainable, plant-based bioproduction of apocarotenoids with pharmaceutical and industrial relevance.

CRediT authorship contribution statement

Conceptualization: O.A. and L.G-G.; writing, reviewing, and editing: L.G-G., O.A., J-A.D and G-D, A.R-M.; plasmids construction, E.M.J, L.M. and V.A.; plant agroinfiltration, L.M., and V.A.; metabolite extraction and analyses, L.G-G., E.P., A.R-M, and L.M; A.J.L.J: bioinformatic analysis; activity assays, L.M., L.G-G, and E.P. All the authors read and approved the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgements

This work was supported by grants PID2023–146186OB-I00 and PID2023–146418OB-I00 from the MICIU/AEI/10.13039/501100011033 and ERDF, EU, and PROMETEO (CIPROM/2022/21) from Generalitat Valenciana. LGG is member of the Spanish Carotenoid Network (CaRed), grant RED2022–134577-T, funded by MCIN/AEI/10.13039/501100011033.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cpb.2026.100599.

Data availability

Data will be made available on request.

References

[1] G. Sandmann, Diversity and evolution of carotenoid biosynthesis from prokaryotes to plants, Adv. Exp. Med. Biol. 1261 (2021) 79–94.
[2] A. Alder, et al., The path from beta-carotene to carlactone, a strigolactone-like plant hormone, Science 335 (6074) (2012) 1348–1351.
[3] A. Ilg, P. Beyer, S. Al-Babili, Characterization of the rice carotenoid cleavage dioxygenase 1 reveals a novel route for geranial biosynthesis, FEBS J. 276 (3) (2009) 736–747.
[4] O. Ahrazem, et al., Evolutionarily distinct carotenoid cleavage dioxygenases are responsible for crocetin production in Buddleja davidii, J. Exp. Bot. 68 (16) (2017) 14.
[5] L. Morote, et al., Montbretia flowers as a source of bioactive crocins: biotechnology tools and delivery systems, Biotech. Rep. 46 (2025) e00891.
[6] Q. Fang, et al., Cloning and functional characterization of a carotenoid cleavage dioxygenase 2 gene in safranal and crocin biosynthesis from Freesia hybrida, Plant Physiol. Biochem. 154 (2020) 439–450.
[7] J.Y. Wang, et al., The apocarotenoid metabolite zaxinone regulates growth and strigolactone biosynthesis in rice, Nat. Commun. 10 (1) (2019) 810.
[8] P. Awasthi, et al., Transgene-free genome editing supports CCD4 role as a negative regulator of β-carotene in banana, J. Exp. Bot. (2022).
[9] I. Amaya, et al., Differential expression of CCD4(4B) drives natural variation in fruit carotenoid content in strawberry (Fragaria spp, Plant Biotechnol. J. 23 (3) (2025) 679–691.
[10] A. Ohmiya, et al., Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals, Plant Physiol. 142 (3) (2006) 1193–1201.
[11] M. Bruno, P. Beyer, S. Al-Babili, The potato carotenoid cleavage dioxygenase 4 catalyzes a single cleavage of beta-ionone ring-containing carotenes and non-epoxidated xanthophylls, Arch. Biochem. Biophys. (2015).
[12] A. Rubio, et al., Cytosolic and plastoglobule-targeted carotenoid dioxygenases from Crocus sativus are both involved in beta-ionone release, J. Biol. Chem. 283 (36) (2008) 24816–24825.
[13] F.C. Huang, et al., Substrate promiscuity of RdCCD1, a carotenoid cleavage oxygenase from Rosa damascena, Phytochem 70 (4) (2009) 457–464.
[14] M.J. Rodrigo, B. Alquezar, L. Zacarias, Cloning and characterization of two 9-cis-epoxycarotenoid dioxygenase genes, differentially regulated during fruit maturation and under stress conditions, from orange (Citrus sinensis L. Osbeck), J. Exp. Bot. 57 (3) (2006) 633–643.
[15] G. Ma, et al., Enzymatic formation of beta-citraurin from beta-cryptoxanthin and zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of Citrus fruit, Plant Physiol. 163 (2) (2013) 682–695.
[16] O. Ahrazem, et al., The carotenoid cleavage dioxygenase CCD2 catalysing the synthesis of crocetin in spring crocuses and saffron is a plastidial enzyme, N. Phytol. 209 (2) (2016) 13.
[17] A. Ji, et al., Transcriptome-guided mining of genes involved in crocin biosynthesis, Front Plant Sci. 8 (2017) 518.
[18] L. Morote, et al., Biotechnological production of crocetin and crocins using a carotenoid cleavage ddioxygenase (CCD4) from Nyctanthes arbor-tristis, Front Plant Sci. (2025).
[19] D.-C. Jonathan, et al., Carotenoid cleavage dioxygenases 4 from woodiness plant and their relationships with herbaceous plants, Ind. Crops Prod. 205 (2023) 117529.
[20] N. Sławińska, J. Zając, B. Olas, Paulownia organs as interesting new sources of bioactive compounds, Int J. Mol. Sci. 24 (2) (2023).
[21] M. Jakubowski, Cultivation potential and uses of Paulownia wood: a review, Forests 13 (2022) 668.
[22] Y. Cao, et al., Genomic insights into the fast growth of paulownias and the formation of Paulownia witches' broom, Mol. Plant 14 (10) (2021) 1668–1682.
[23] L. Morote, et al., A carotenoid cleavage dioxygenase 4 from Paulownia tomentosa determines visual and aroma signals in flowers, Plant Sci. 329 (2023) 111609.
[24] L. Gomez-Gomez, et al., Determination of in vitro and in vivo activities of plant carotenoid cleavage oxygenases, Methods Mol. Biol. 2083 (2020) 63–74.
[25] L. Morote, et al., In vitro dioxygenase activity characterization using headspace stir bar sorptive extraction (HSSE), Anal. Methods 16 (33) (2024) 5733–5740.
[26] M. Martí, et al., Efficient production of saffron crocins and picrocrocin in Nicotiana benthamiana using a virus-driven system, Metab. Eng. 61 (2020) 238–250.
[27] G. Diretto, et al., UGT709G1: a novel uridine diphosphate glycosyltransferase involved in the biosynthesis of picrocrocin, the precursor of safranal in saffron (Crocus sativus), N. Phytol. (2019).
[28] A.A. D'Archivio, et al., UHPLC analysis of Saffron (Crocus sativus L.): optimization of separation using chemometrics and detection of minor crocetin esters, Molecules 23 (8) (2018).
[29] O. Ahrazem, et al., Engineering high levels of saffron apocarotenoids in tomato, Hortic. Res 9 (2022) uhac074.
[30] D.Q. Shi, et al., SLOW WALKER1, essential for gametogenesis in Arabidopsis, encodes a WD40 protein involved in 18S ribosomal RNA biogenesis, Plant Cell 17 (8) (2005) 2340–2354.
[31] M. Stanke, et al., AUGUSTUS: ab initio prediction of alternative transcripts, Nucleic Acids Res. 34 (_2) (2006) W435–W439.
[32] C.L.M. Gilchrist, Y.H. Chooi, clinker & clustermap.js: automatic generation of gene cluster comparison figures, Bioinformatics 37 (16) (2021) 2473–2475.
[33] D. Kim, et al., Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (8) (2019) 907–915.
[34] C. Trapnell, et al., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (5) (2010) 511–515.
[35] L. Morote, et al., Verbascum species as a new source of saffron apocarotenoids and molecular tools for the biotechnological production of crocins and picrocrocin, Plant J. 118 (2024) 14.
[36] Y.H. Liao, P.J. Houghton, J.R. Hoult, Novel and known constituents from Buddleja species and their activity against leukocyte eicosanoid generation, J. Nat. Prod. 62 (9) (1999) 1241–1245.
[37] T.H. Tseng, et al., Crocetin protects against oxidative damage in rat primary hepatocytes, Cancer Lett. 97 (1) (1995) 61–67.
[38] E.G. Hill, A.P. Sirkar, A new colouring matter from Nyctanthes Arbor-tristis, J. Chem. Soc. Trans. 91 (1907) 4.
[39] N. Guo, X.Q. Zhai, G.Q. Fan, Chemical composition, health benefits and future prospects of Paulownia flowers: A review, Food Chem. 412 (2023) 135496.
[40] T. Abu-Izneid, et al., Nutritional and health beneficial properties of saffron (Crocus sativus L): a comprehensive review, Crit. Rev. Food Sci. Nutr. 62 (10) (2022) 2683–2706.
[41] X. Sui, P.D. Kiser, J.v. Lintig, K. Palczewski, Structural basis of carotenoid cleavage: from bacteria to mammals, Arch Biochem Biophys. 539 (2) (2013) 203–213, https://doi.org/10.1016/j.abb.2013.06.012. Epub 2013 Jul 1. PMID: 23827316; PMCID: PMC3818509.
[42] G. Diretto, et al., Identification and characterization of apocarotenoid modifiers and carotenogenic enzymes for biosynthesis of crocins in Buddleja davidii flowers, J. Exp. Bot. 72 (8) (2021) 3200–3218.