

Complete biosynthetic pathway of furochromones in *Saposhnikovia divaricata* and its evolutionary mechanism in Apiaceae plants

Received: 22 July 2024

Accepted: 18 March 2025

Published online: 01 April 2025



Jian-lin Zou^{1,4}, Hong-ye Li^{1,4}, Bao Nie^{2,4}, Zi-long Wang¹, Chun-xue Zhao¹, Yun-gang Tian¹, Li-qun Lin², Wei-zhe Xu³, Zhuang-wei Hou², Wen-kai Sun², Xiao-xu Han², Meng Zhang¹, Hao-tian Wang¹, Qing-yan Li³, Li Wang²✉ & Min Ye¹✉

Furochromones are specific bioactive secondary metabolites of many Apiaceae plants. Their biosynthesis remains largely unexplored. In this work, we dissect the complete biosynthetic pathway of major furochromones in the medicinal plant *Saposhnikovia divaricata* by characterizing prenyltransferase, peucenin cyclase, methyltransferase, hydroxylase, and glycosyltransferases. De novo biosynthesis of prim-*O*-glucosylcimifugin and 5-*O*-methylvisamminoside is realized in *Nicotiana benthamiana* leaves. Through comparative genomic and transcriptomic analyses, we further find that proximal duplication and high expression of a pentaketide chromone synthase gene *SdPCS*, together with the presence of a lineage-specific peucenin cyclase gene *SdPC*, lead to the predominant accumulation of furochromones in the roots of *S. divaricata* among surveyed Apiaceae plants. This study paves the way for metabolic engineering production of furochromones, and sheds light into evolutionary mechanisms of furochromone biosynthesis among Apiaceae plants.

Furochromones are an important class of bioactive natural products. They demonstrate anti-inflammatory^{1,2}, hepatoprotective³, and antiviral⁴ activities. For instance, khellin was used as a smooth muscle relaxant and cardioprotective drug in Europe⁵. While chromones are widely present in plants, furochromones have only been reported in a few families including Apiaceae, Ranunculaceae, and Leguminosae⁶. In Apiaceae, furochromones are the major bioactive compounds of *Saposhnikovia divaricata*⁷, *Ammi visnaga*⁸, and *Cnidium monnieri*⁹. Particularly, *S. divaricata* contains abundant prim-*O*-glucosylcimifugin (POG) and 5-*O*-methylvisamminoside

(5-*O*-MVG), and their total content could be above 0.24% of dry weight¹⁰. These two compounds may contribute to the bioactivities of *S. divaricata* for the treatment of respiratory virus infection¹¹, type I allergy¹², colitis¹³, and aging-impaired endogenous tendon regeneration¹⁴.

The structures of POG and 5-*O*-MVG feature in the substitution of an isoprenyl group at C-6, which forms a fused dihydrofuran ring^{15,16} (Supplementary Fig. 1). The biosynthesis of simple chromones has been extensively studied. The chromone skeletons are generated by polyketide synthases, such as PECPS from *Aquilaria sinensis* and AaPCS from

¹State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China.

²Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China. ³Civil Aviation Medicine Center, Civil Aviation Administration of China, A-1 Gaojing, Beijing 100123, China. ⁴These authors contributed equally: Jian-lin Zou, Hong-ye Li, Bao Nie.

✉ e-mail: wangli03@caas.cn; yemin@bjmu.edu.cn

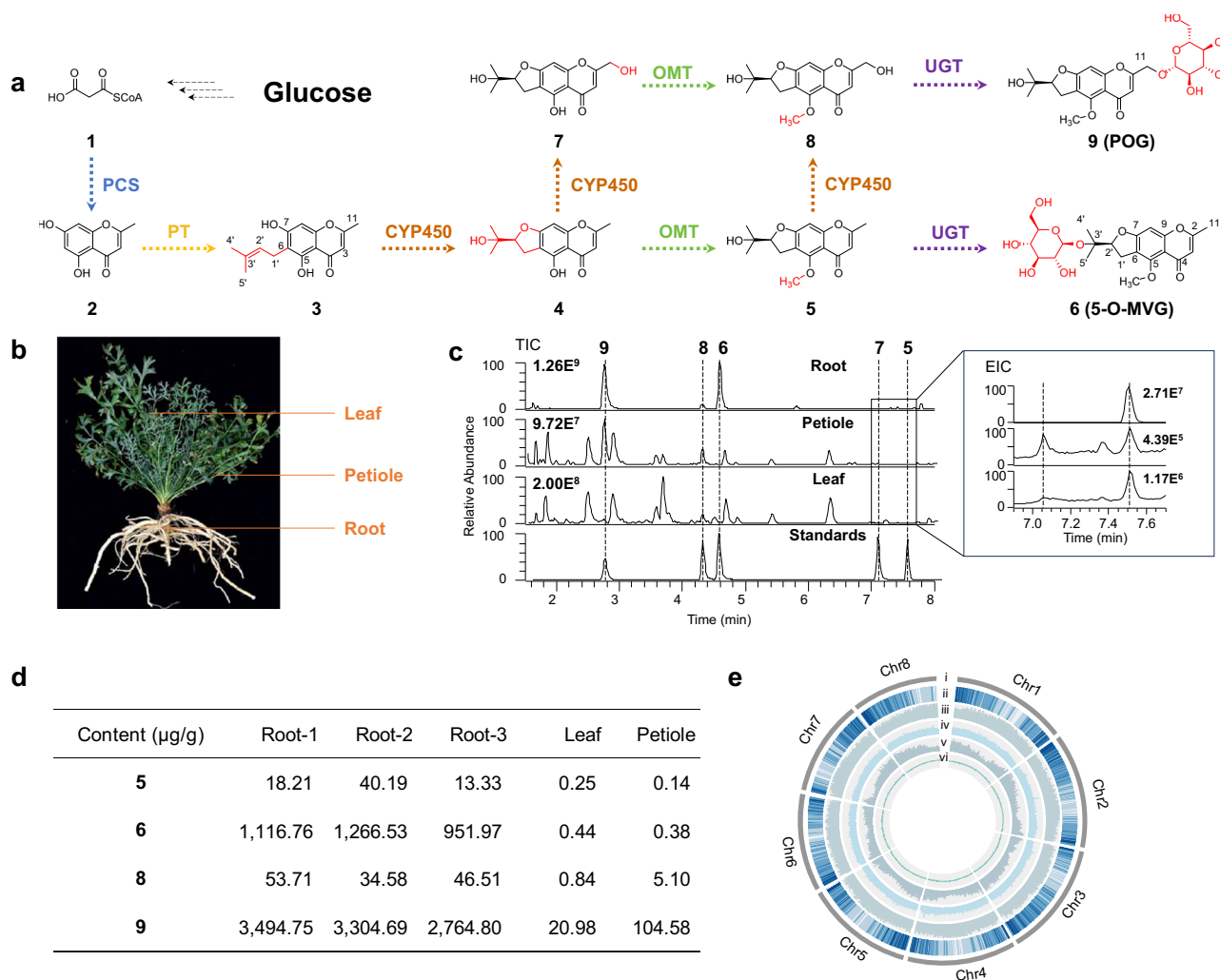


Fig. 1 | A proposed biosynthetic pathway of furochromones and genomic statistics in *S. divaricata*. **a** The proposed biosynthetic pathway and catalytic enzymes. PT prenyltransferase, PCS pentaketide chromone synthase, CYP450 cytochrome P450 enzyme, OMT *O*-methyltransferase, UGT uridine diphosphate-dependent glycosyltransferase. **1**, malonyl-CoA; **2**, noreugenin; **3**, peucenin; **4**, visamminol; **5**, 5-*O*-methylvisamminol; **6**, 5-*O*-methylvisamminoside; **7**, norcimifugin; **8**, cimifugin; **9**, prim-*O*-glucosylcimifugin. **b** Image of the sampled *S. divaricata*.

c, Total ion currents (TICs) and extracted ion chromatograms (EICs) of the root, petiole, and leaf of *S. divaricata* by LC/MS analysis. EIC mass range: *m/z* 291.11–291.12 + 293.09–293.10. **d** Contents of **5**, **6**, **8** and **9** in different organs, calculated on the basis of dry weight (*n* = 3, three biologically independent samples were tested). **e** Genomic statistics of *S. divaricata*, showing eight chromosomes (Chr1–Chr8). i, pseudochromosomes; ii, gene density; iii, Gypsy LTR density; iv, Copia LTR density; v, Helitron density; vi, GC content.

Aloe arborescens^{17,18}. However, the biosynthesis of furochromones remains largely unexplored. In the early 1970s, researchers fed sodium [¹⁴C] acetate to the shoots of *Ammi visnaga*, and revealed that peucenin and visamminol were biosynthetic intermediates of furochromones¹⁹. For the biosynthesis of POG or 5-*O*-MVG, a prenyltransferase (PT) is responsible to introduce an isoprenyl group to C-6 of the chromone skeleton²⁰. Thus far, very few enzymes have been reported to catalyze cyclization of an isoprenyl group to form a dihydrofuran ring. Although CYP76F112 from *Ficus carica*, PpDC and PpOC from *Peucedanum praeruptorum*, NiDC and NiOC from *Notopterygium incisum*, as well as AsDC and AsOC from *Angelica sinensis* have been reported to catalyze similar reactions to produce furocoumarins^{21–24}, no enzymes have been testified to generate furochromones. On the other hand, glycosyl substitutions at hydroxyl groups linking to the quaternary C-3' or the secondary C-11 are rare for natural products, and these reactions are usually catalyzed by uridine diphosphate-dependent glycosyltransferases (UGTs)²⁵. Moreover, both POG and 5-*O*-MVG contain a methoxyl group at C-5, and the methylation reaction was proposed to be catalyzed by an *O*-methyltransferase (OMT)²⁶. Although a big family of OMTs have been reported from plants, few OMTs could catalyze methylation at the less

active 5-OH. Limited examples include the isoflavone 5-*O*-methyltransferase from *Lupinus luteus*²⁷ and CdFOMT5 from *Citrus depressa*²⁸. For POG, a cytochrome P450 (CYP450) enzyme may introduce the extra primary hydroxyl group at C-11²⁹. Based on the above analysis, we hypothesized the biosynthetic pathway of 5-*O*-MVG (**6**) and POG (**9**) (Fig. 1a). While the enzyme categories catalyzing each step seem obvious, the specific enzymes with expected functions are still illusive.

In this work, we dissected the biosynthetic pathway of POG and 5-*O*-MVG in *S. divaricata*. The functions of seven enzymes were characterized, including SdPCS, SdPT, SdPC, SdCH, SdOMT, SdUGT1, and SdUGT2. Utilizing these gene elements, we realized the complete biosynthesis of POG and 5-*O*-MVG in *Nicotiana benthamiana* leaves. Moreover, we unravelled the genetic mechanisms for high abundances of POG and 5-*O*-MVG in *S. divaricata* among Apiaceae plants.

Results and discussion

Proposed biosynthetic pathway of furochromones in *Saposhnikovia divaricata* and gene mining

First, we analyzed the chemical constituents of three organs of *S. divaricata* (leaf, petiole, and root, Fig. 1b–c) by liquid chromatography

coupled with mass spectrometry (LC/MS). At least five furochromones (**5–9**) could be detected, which supported the validity of our proposed biosynthetic pathway. Subsequently, the contents of major compounds **5**, **6**, **8** and **9** in five tissue samples (roots at three developmental stages, petiole, and leaf) were quantitatively determined (Supplementary Figs. 2–6). The results indicated the roots contained more abundant furochromones, particularly the glycosides **6** and **9**, than the petiole and leaf samples (Fig. 1d).

In order to obtain a complete list of candidate genes involved in the biosynthesis of POG and 5-O-MVG, we sequenced, assembled, and annotated a chromosome-level genome of *S. divaricata*. Based on 43.72 Gb PacBio CCS long reads, we assembled the genome to 1.95 Gb (Supplementary Table 1), which was consistent with the estimate by flow cytometry (1.94 ± 0.02 Gb) (Supplementary Fig. 7) and the published assembly³⁰. The assembly contig N50 was 2.22 Mb and the Benchmarking Universal Single-Copy Ortholog (BUSCO) score was 96.1%, indicating good genome continuity and completeness (Supplementary Tables 2–3). By Hi-C technology, 94.27% contigs were anchored onto eight chromosomes (Fig. 1e, Supplementary Fig. 8 and Supplementary Table 4). Multiple-tissue RNA-Seq data (Supplementary Table 5), ab initio prediction, and homolog protein evidences were combined for genome annotation, which led to the identification of 38,704 high-confidence protein-coding genes and 65,734 transcripts. Finally, a total of 1,751,401 repetitive elements were annotated, accounting for 76.78% of the genome (Supplementary Table 6). With the high-quality genome and multiple-tissue RNA-Seq data, we quantified the gene expression abundance (fragments per kilobase of exon model per million mapped fragments, FPKM) of the five tissue samples mentioned above. Subsequently, we screened candidate genes according to genome annotation or local blastn search, and selected genes whose expression levels were correlated with the contents of downstream secondary metabolites in different organs for functional characterization.

Biosynthesis of the furochromone skeleton

The first step of the biosynthetic pathway converts malonyl-CoA (**1**) to noreugenin (**2**). The pentaketide chromone synthase AaPCS from *Aloe arborescens* is the only reported enzyme to catalyze this type of reaction¹⁸. Thus, we conducted a local blastn search using AaPCS as a query in the transcriptome of *S. divaricata*, and ten candidate genes with *e* values $< 10^{-21}$ were discovered. The expression level (FPKM) of one gene, *SdPCS*, was highly correlated with the furochromones contents with Pearson correlation coefficient (PCC) > 0.95 (Supplementary Table 7). It was sub-cloned into the pET28a(+) vector for protein expression in *E. coli* BL21 (DE3) cells. The function was characterized by enzyme catalysis reactions with **1** as substrate. According to high-performance liquid chromatography (HPLC) and LC/MS analyses, *SdPCS* generated a new peak, which was identified as **2** by comparing with a reference standard. *SdPCS* showed the highest catalytic efficiency at 37°C in $\text{Na}_2\text{HPO}_4\text{--NaH}_2\text{PO}_4$ buffer (pH 6.0). The k_{cat}/K_m value of *SdPCS* generating **2** was $22.84 \text{ M}^{-1}\text{s}^{-1}$ (Supplementary Fig. 9). From the genome of *S. divaricata*, we further discovered and characterized *SdPCS2* with the same function (Fig. 2a), though it was initially neglected due to low expression level (Supplementary Data 1). *SdPCS2* shows high amino acid sequence identity (91.33%) and short distance in the genome (135,334 bp) with *SdPCS* (Supplementary Figs. 10–11).

To discover the prenyltransferase (PT) converting **2** to peucenin (**3**), we obtained one candidate gene *SdPT* (PCC > 0.95 , Supplementary Table 8) among the 20 annotated PT genes. *SdPT* was sub-cloned to pESC-Leu vector and expressed in yeast WAT11 cells³¹. When the yeast microsomes were incubated with **2**, DMAPP and MgCl_2 , HPLC analysis showed a new product, which exhibited an $[\text{M} + \text{H}]^+$ ion at m/z 261.11 in LC/MS analysis. The MS/MS spectrum showed an abundant $[\text{M}-56 + \text{H}]^+$ fragment at m/z 205.05, indicating a prenyl substitution at C-6 or C-8³² (Fig. 2b). *SdPT* showed the highest catalytic efficiency at 45°C in $\text{Na}_2\text{CO}_3\text{--NaHCO}_3$ buffer (pH 10.0), with a K_m value of 29.08 μM

(Supplementary Fig. 12). Then we purified 0.8 mg of the product from scaled-up enzymatic reactions with the most suitable reaction condition. The $^1\text{H-NMR}$ spectrum showed two methylene signals at δ_{H} 3.17 (m, H-1'), one olefinic signal at δ_{H} 5.16 (t, $J = 6.0$ Hz, H-2'), and two methyl signals at δ_{H} 1.61 (H-4') and 1.71 (H-5'), indicating the presence of an isoprenyl group. The HMBC cross peaks from H-1' to C-5 (δ_{C} 158.1), C-6 (δ_{C} 111.1), and C-7 (δ_{C} 164.8) indicated the isoprenyl group was located at C-6 (Supplementary Figs. 13–16). Thus, the product was identified as peucenin (**3**) (Supplementary Table 9). *SdPT* represented the first prenyltransferase utilizing chromones as substrate. *SdPT2*, with an amino acid sequence identity of 85.79% (Supplementary Fig. 17), exhibited the same catalytic function as *SdPT*.

Few enzymes are known to catalyze the oxidative cyclization of isoprenyl groups, except for several CYP450 enzymes involved in the biosynthesis of furocoumarins^{22–24}. Since these enzymes belong to the CYP736 family, we screened candidates from the same family in *S. divaricata*, and chose four candidate genes whose expression levels were highly correlated with the furochromones contents (PCC > 0.90 , Supplementary Table 10). By incubating the microsomes of *SdPC* recombinant yeast WAT11 cells with **3** and NADPH, a new product was yielded. LC/MS analysis showed an $[\text{M} + \text{H}]^+$ ion at m/z 277, which could be fragmented into m/z 259 and m/z 205. Its structure was proposed to be visamminol (**4**). As no reference standard was available, we prepared **4** through hydrolysis of visamminol 3'-O-glucoside catalyzed by β -glucosidase (Supplementary Fig. 18), and confirmed its structure by NMR analysis. The $^1\text{H-NMR}$ spectrum showed two methyl signals at δ_{H} 1.13 (s, H-4') and δ_{H} 1.14 (s, H-5'), a tertiary proton signal at δ_{H} 4.71 (t, $J = 8.6$ Hz, H-2'), and a methylene signal at δ_{H} 3.02 (d, $J = 8.6$ Hz, H-1'), indicating the presence of a furan ring. The HMBC cross peaks from H-2' (δ_{H} 4.71) to C-1' (δ_{C} 26.6), C-7 (δ_{C} 166.4), and C-6 (δ_{H} 109.5) indicated the furan ring was conjugated with the benzene ring (Supplementary Figs. 19–22, Supplementary Table 9). HPLC and LC/MS analyses indicated the product had the same retention time and mass spectra with **4** (Fig. 2c). As the oxidative cyclization of isoprenyl phenolic compounds by chemical synthesis requires strong oxidizers like *m*-chloroperbenzoic acid³³, *SdPC* represents an efficient enzyme catalyst for this reaction.

Post-modification steps for the biosynthesis of furochromones

C-11 of compounds **7–9** is hydroxylated, indicating the presence of a CYP450 enzyme. However, very few enzymes have been reported to catalyze a similar reaction, thus no suitable templates are available for gene blast search. By analyzing the transcriptome data, we selected 12 candidate CYP genes, whose expression levels were highly correlated with the total contents of **8** and **9** (PCC > 0.95 , Supplementary Table 11). These genes were expressed in yeast WAT11 cells, and the microsomes were incubated with NADPH and **4** or **5** (Tris-HCl buffer, 50 mM) for functional characterization. LC/MS analysis indicated that *SdCH* could convert **4** and **5** (5-O-methylvisamminol) into **7** (norcimifugin) and **8** (cimifugin), respectively (Fig. 3a, Supplementary Fig. 23).

Likewise, we discovered the 5-O-methyltransferase *SdOMT* which converted **4** and **7** into **5** and **8**, respectively (PCC > 0.90 , Supplementary Table 12). Its function was characterized by enzymatic reaction and LC/MS analysis (Fig. 3b, Supplementary Fig. 24).

Glycosylation is the final step in the biosynthetic pathway. A total of 8 UGT genes with FPKM > 10 in the roots were chosen as candidate genes, and were cloned and expressed in *E. coli* BL21(DE3) (Supplementary Table 13). The functions were characterized by enzymatic catalysis with UDP-Glc (UDPG) as sugar donor, and **5** or **8** as sugar acceptor. *SdUGT1* (UGT93BA1) could catalyze the glucosylation of 3'-OH of **5** (tertiary alcohol) and 11-OH of **8** (primary alcohol) to produce **6** (5-O-methylvisamminoside, 5-O-MVG) and **9** (prim-O-glucosylcimifugin, POG), respectively. The products could lose 162 Da in the MS/MS spectra, and their structures were identified by comparing with reference standards (Fig. 3c–d). Moreover, we discovered *SdUGT2*

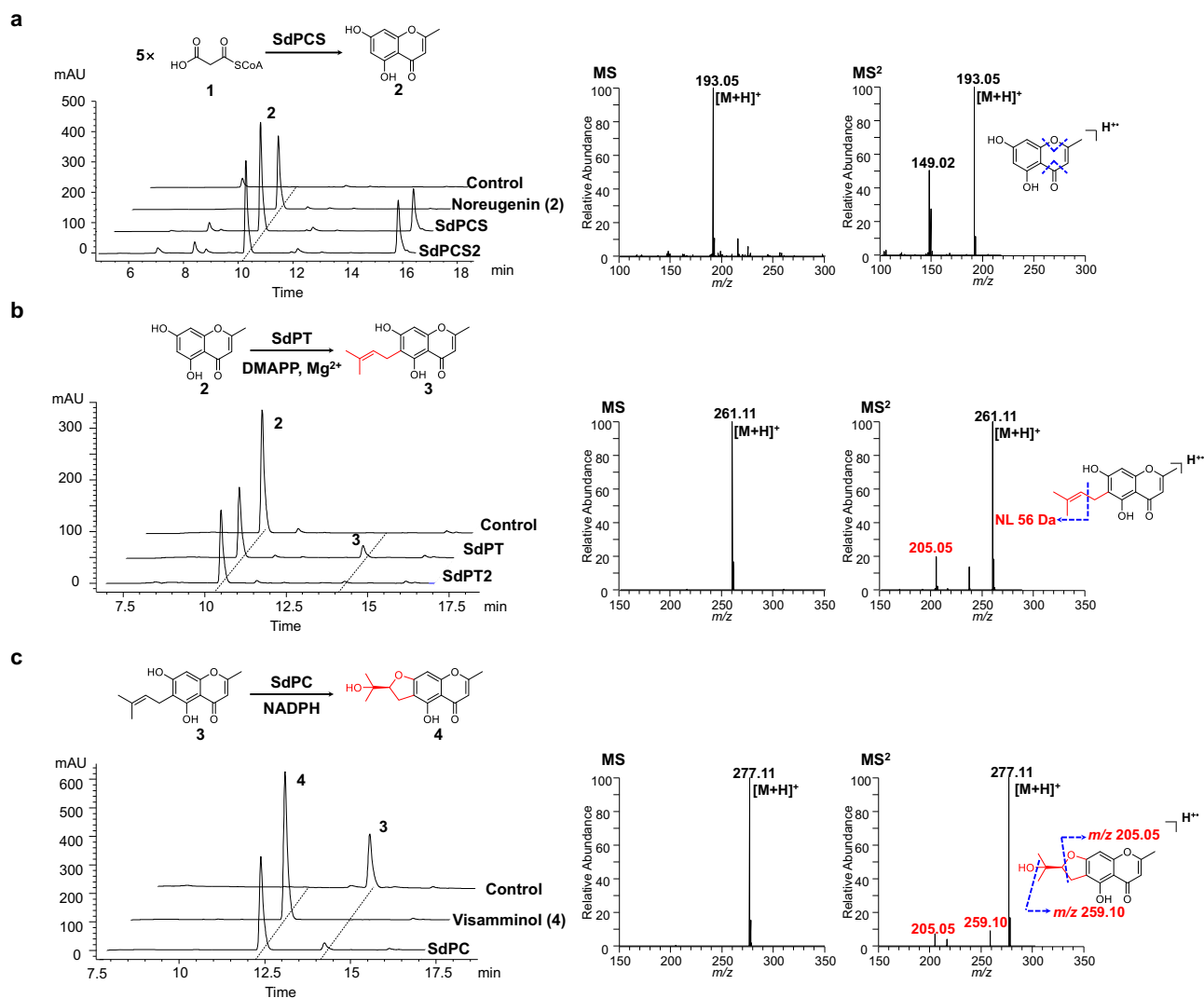


Fig. 2 | Biosynthesis of furochromone skeleton. Functional characterization of SdPCS (a), SdPT (b), and SdPC (c). Shown are HPLC/UV chromatograms of enzyme catalysis reactions ($\lambda = 280$ nm), together with (+)-ESI-MS and MS/MS spectra of the

products. Control, reaction mixtures incubated with boiled enzymes or microsomes.

(UGT93BB1), which exhibited a high amino acid sequence identity (54.93%) with SdUGT1 and showed the same catalytic activities (Supplementary Fig. 25). Interestingly, SdUGT1 and SdUGT2 only catalyzed 11-*O*-, but not 3'-*O*-glycosylation of **8**. Moreover, they showed 31 and 2.8-fold higher catalytic efficiency (k_{cat}/K_m value) with **8** than with **5** as substrate (Fig. 3e, Supplementary Figs. 26–27).

To elucidate mechanisms for the preference towards 11-OH, we acquired the crystal structure of SdUGT2 in complex with UDP through X-ray diffraction (PDB ID: 8ZNK, 1.88 Å) (Fig. 3f, Supplementary Fig. 28, Supplementary Table 14). The structure of SdUGT2 showed a typical GT-B fold with two Rossmann-like $\beta/\alpha/\beta$ domains. The N-terminal domain (NTD, residues 1–261 and 454–480) and the C-terminal domain (CTD, residues 262–453) are primarily responsible for sugar acceptor and sugar donor binding, respectively. Then we simulated the SdUGT2/UDP-Glc complex structure according to the GgCGT/UDP-Glc structure, and docked **8** into the structure in two potential binding modes through AutoDock 4.2 software^{34,35}. Alanine scanning of residues around the binding pocket led to remarkably decreased activities for most mutants, indicating reliability of the docking results (Supplementary Fig. 29). In both binding modes, His32 is close to the glycosylation sites (11-OH or 3'-OH) with a distance below 3.1 Å. Thus, the hydroxyl groups could be easily deprotonated to initiate the glycosylation reaction. However, **8**

could form more hydrogen bonds and hydrophobic interactions in the 11-*O*-glycosylation mode than in the 3'-*O*-glycosylation mode, as predicted by PLIP 2021³⁶. Thus, SdUGT2 preferred to catalyze 11-*O*-glycosylation of **8**. On the other hand, docking of **9** into SdUGT2 showed the distance between 3'-OH and His32 was too far for glycosylation reaction. This result was consistent with the absence of furochromone 3',11-di-*O*-glycosides in *S. divaricata*⁷ (Supplementary Fig. 30). We also simulated the structure of SdUGT1 using AlphaFold2³⁷, and docked **8** and UDP-Glc into the structure in the same way as described above. Comparing the binding modes of **8** in SdUGT1 and SdUGT2, we noticed that **8** was more stable in SdUGT1 as a result of extra hydrogen bonds and π -stacking interactions (Supplementary Fig. 31). This is probably the reason for the higher catalytic efficiency of SdUGT1 than SdUGT2.

Thus far, we have identified seven enzymes from *S. divaricata* catalyzing biosynthesis of the two major furochromones **6** and **9**. These genes are located at different chromosomes. Specifically, *SdCH* and *SdUGT1* are located at Chr1, *SdPCS* and *SdPC* at Chr2, *SdPT* at Chr3, *SdOMT* at Chr6, and *SdUGT2* at Chr8 (Fig. 4a). To our knowledge, this is the first report to unravel the complete biosynthetic pathway of furochromones. The expression levels of identified genes, except for *SdUGT1* and *SdUGT2*, are highly correlated with the distribution of major furochromones among different organs of *S. divaricata*.

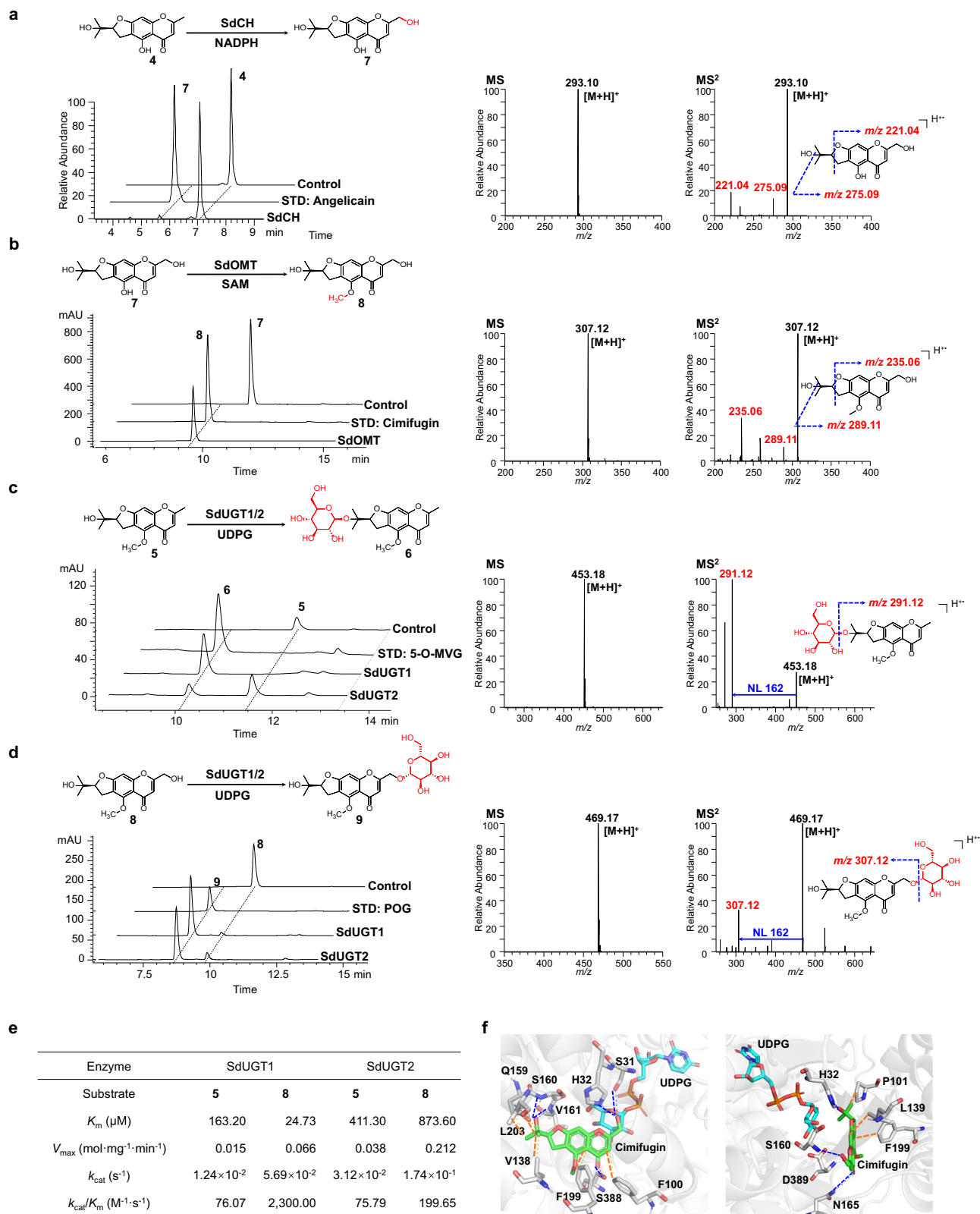


Fig. 3 | Post-modification reactions for the biosynthesis of furochromones. Functional characterization of SdCH (**a**), SdOMT (**b**), and SdUGT1/2 (**c**, **d**). Shown are HPLC/UV chromatograms of the enzyme catalysis reactions ($\lambda = 280$ nm), together with (+)ESI-MS and MS/MS spectra of the products. **e** Kinetic parameters of SdUGT1 and SdUGT2. **f** Simulated binding modes of **8** in the crystal

structure of SdUGT2 when catalyzing 11-O-glycosylation (left) and 3'-O-glycosylation (right). Hydrogen bonds and hydrophobic interactions were labeled with blue and orange dashes, respectively. STD, reference standard. Control, reaction mixtures incubated with boiled enzymes or microsomes.

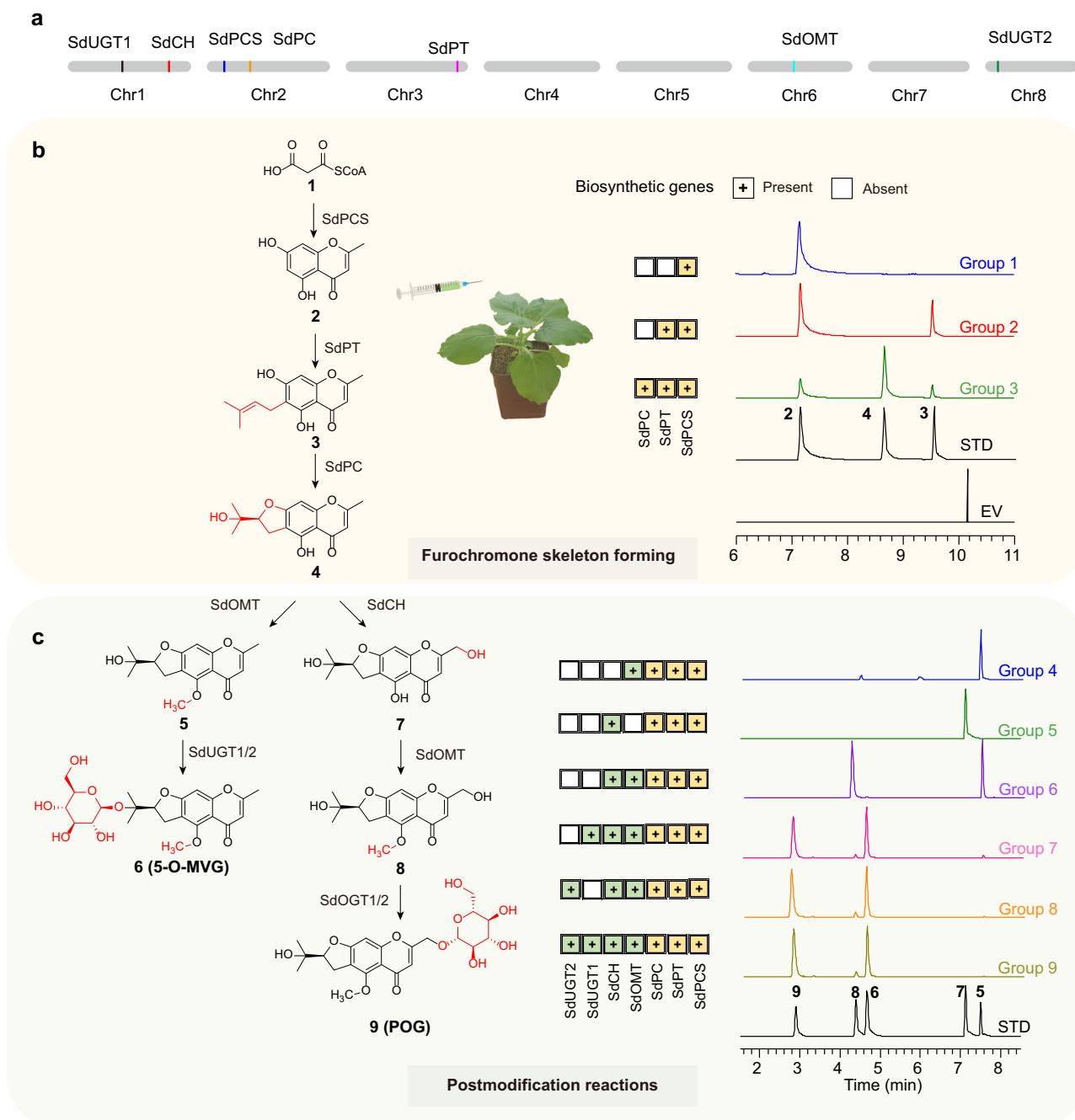


Fig. 4 | The dissected biosynthetic pathway of furochromones and their de novo biosynthesis in *N. benthamiana* leaves. **a** Genomic location of biosynthetic genes in *S. divaricata*. Catalytic functions of biosynthetic genes responsible for formation (**b**) and post-modification (**c**) of the furochromone skeleton. Extracted

ion chromatograms (EICs) of biosynthetic products in LC/MS analysis are shown. STD, reference standards. EV, agrobacterium-mediated transient expression using a vector without any biosynthetic genes.

De novo biosynthesis of *Saposhnikovia* furochromones in *Nicotiana benthamiana* leaves

POG and 5-O-MVG are important bioactive compounds in *S. divaricata*. Their extraction and purification are time and labor-consuming. It is imperative to engineer the biosynthetic pathway in chassis organisms. In this work, we realized de novo biosynthesis of furochromones in *Nicotiana benthamiana* leaves. Transient expression of the seven genes in *N. benthamiana* leaves revealed that all genes showed the expected catalytic activities (Fig. 4b-c). When all the seven genes were infiltrated into *N. benthamiana* leaves, **6** and **9** could be detected. Given the low catalytic efficiency of SdPT, we increased the OD₆₀₀ value of SdPT strain to 0.40, and **6** and **9** were generated at a yield of

17.48 µg/g and 3.82 µg/g (dry weight, Supplementary Fig. 32), respectively, where the yields were calculated with six independent biological replicates.

The distribution of chromones and their biosynthetic genes in Apiaceae plants

To gain deep insights into the evolution of biosynthetic pathway of furochromones in Apiaceae, we incorporated another seven Apiaceous species (*Coriandrum sativum*, *Apium graveolens*, *Angelica sinensis*, *Ligusticum chuanxiong*, *Daucus carota*, *Bupleurum chinense*, *Centella asiatica*) into metabolic, comparative genomic and transcriptomic analyses (Supplementary Data 2–4, Supplementary

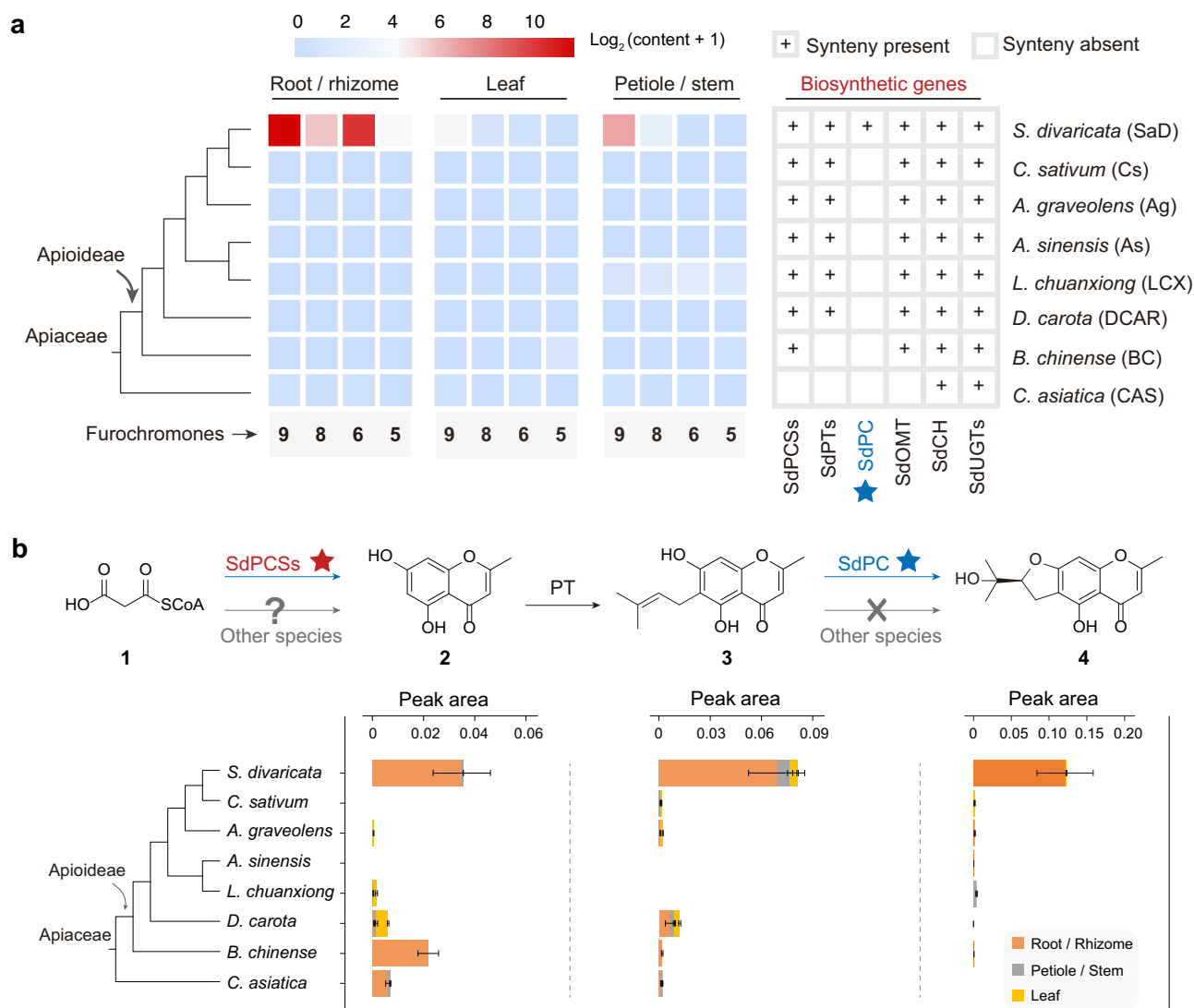


Fig. 5 | The distribution of chromones and their biosynthetic genes in Apiaceae plants. a Contents of typical furochromones in various organs of Apiaceae plants, and syntenic gene analysis. **b** Contents of simple chromones in various organs of

Apiaceae plants and the hypothesized causes ($n = 3$, three biologically independent samples were tested, and data are presented as mean values \pm SD).

Table 15). These species represented different evolutionary lineages, including the subfamilies Mackinlayoideae and Apioideae (including the tribe Bupleureae, Apieae, *Sinodielsia* Clade, Selineae). We first determined and compared the contents of four typical furochromones (5, 6, 8, and 9) among the eight species (Supplementary Figs. 33–53). Unexpectedly, the furochromones did not show a stepwise accumulation along the phylogeny backbone but exhibited a drastic enrichment in *S. divaricata*. The contents of furochromones in the other species were generally low (Fig. 5a, Supplementary Table 16). This result implied substantial differences in furochromone biosynthesis between *S. divaricata* and the other Apiaceous plants.

To investigate the evolutionary shift in furochromone biosynthesis from early diverged Apiaceous lineage to *S. divaricata*, we constructed a maximum likelihood (ML) phylogeny of Apiaceae species based on 398 strict single-copy orthologous genes (Fig. 5a). It revealed that *S. divaricata* belonged to the latest diverged clade including *C. sativum*. Then, the conserved syntenic gene blocks containing each furochromone-biosynthetic gene was identified in each Apiaceous species (Supplementary Figs. 54–64). The syntenic and homologous genes of SdUGTs and SdCH were detected in all Apiaceous species, while PCSs, OMTs and PTs were limited to Apioideae, indicating a

stepwise assembly of the furochromone biosynthetic pathway in Apiaceae (Fig. 5a). Remarkably, the last piece of the puzzle, SdPC, was detected only in *S. divaricata* (Fig. 5a). This clue motivated us to speculate that most Apiaceous species except *S. divaricata* may not contain any functional PC, thus leading to low furochromone content.

Additionally, we found that simple chromones (2 and 3) showed consistent distribution with furochromones (4, 5, 6, 8 and 9) in Apiaceae. They were also drastically enriched in *S. divaricata* (Fig. 5b, Supplementary Figs. 65–88). As 2 and 3 were generated before the catalysis of PC, their absence in other Apiaceous species was more likely caused by blocking of the initial step (PCS catalysis) (Fig. 5b). Moreover, most potential Apiaceous PTs involved in furochromone biosynthesis showed moderate or high expression abundances (Supplementary Table 17). Thus, except for PC, we mainly focused on the upstream PCS, the housekeeping gene in the chromone biosynthetic pathway.

Proximal duplication of *SdPCS* promotes furochromone accumulation in *S. divaricata*

We retrieved all potential PKS III genes in *S. divaricata* and the other seven Apiaceae species, and constructed an ML tree. A strongly

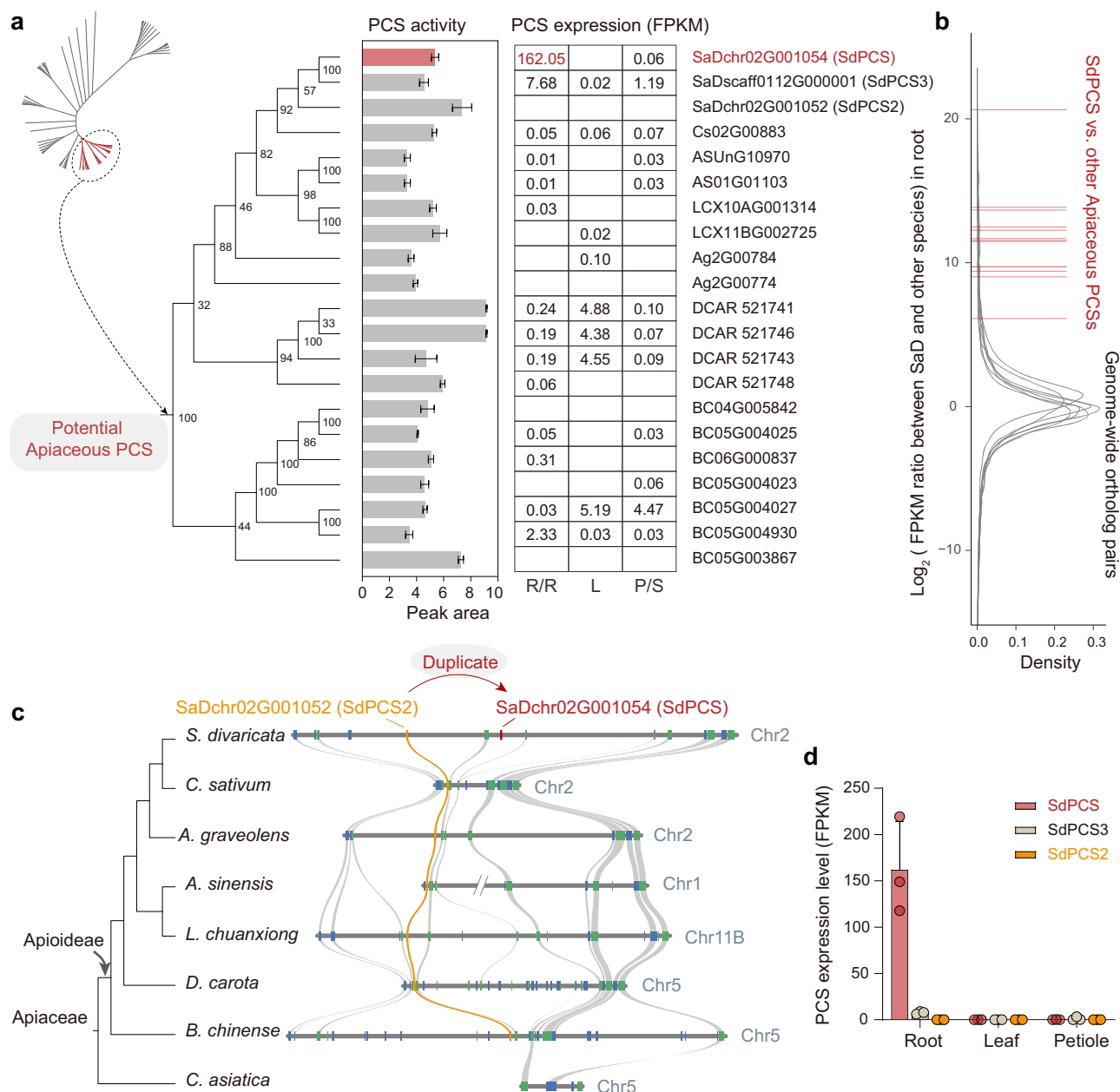


Fig. 6 | High expression of *SdPCS* promotes the accumulation of fur-chromones in the roots of *S. divaricata*. **a** Phylogenetic relationships, enzymatic activities, and expression abundances of Apioceous PCSs. The PCS enzyme activity was quantified by HPLC/UV peak area of generated noreugenin (2) ($\lambda = 280$ nm, $n = 3$, three biologically independent samples were tested, and data are presented as mean values \pm SD). R/R, L, and P/S represent Root/Rhizome, Leaf, and Petiole/Stem, respectively. **b** Comparison of the FPKM between *SdPCS* and other potential

Apioceous PCSs in the genome-wide context. Each red line represents a \log_2 (FPKM ratio) between *SdPCS* and a potential Apioceous PCS. Each grey density plot indicates the \log_2 (FPKM ratio) distribution of genome-wide orthologous genes of one Apioceae species. **c** Syntenic regions containing Apioceous PCSs. **d** Expression levels of the three PCS copies in *S. divaricata* ($n = 3$, three biologically independent samples were tested, data are presented as mean values \pm SD).

supporting clade (bootstrap support value (BS) = 100) containing *SdPCS* and 20 potential Apioceous PCSs was identified (Fig. 6a). Most genes in this clade were in the same syntenic region, implying they shared the same ancestor (Supplementary Fig. 63). No PCS was detected in *C. asiatica*, one of the most basal species in Apioceae, indicating that PCS may first emerge in the Apioideae subfamily. Then we expressed and characterized all the 20 genes, and compared their functions by enzymatic assays (Fig. 2a, Supplementary Figs. 89–94). Most of these enzymes showed similar catalytic abilities by converting 1 to 2 (Fig. 6a). However, the expression level of *SdPCS* in the root of *S. divaricata* was remarkably higher than the other homologous PCS genes (FPKM value, 162.05 vs 0–5.19) (Fig. 6a).

We are aware that direct inter-species comparison of FPKM might lead to misinterpretation since the utilization of FPKM value is usually limited to intra-species level. Instead, we compared the FPKM ratio between *SdPCS* and other potential Apioceous PCSs against 9470–27,214 pairs of orthologous genes as genomic background. The \log_2 (FPKM ratio) value of >95% genome-wide ortholog pairs between *S. divaricata* and other Apioceous species in root/rhizome is <5.00, with a mean near zero (–0.005). This result indicates that FPKM values of the investigated species are basically comparable. It is noteworthy that the \log_2 (FPKM ratio) between *SdPCS* and other Apioceous PCSs in root/rhizome, ranging from 6.12 to 20.63 (mean = 14.15), was higher than 95% of genome-wide orthologous gene pairs (Fig. 6b). Thus, we

deduced the expression abundance of *SdPCS* was significantly higher than the other Apiaceous *PCS*s. As the initial step is usually the rate-limiting step in biosynthetic pathway³⁸, the exceptionally high expression of *SdPCS* may contribute to the furochromone accumulation in *S. divaricata*.

Moreover, we traced the origin of *SdPCS* (SaDchr02G001054), and found it might originate from proximal duplication of the nearby *SdPCS2* (SaDchr02G001052) (Fig. 6c). The *PCS* ML tree revealed the three *PCS* copies in *S. divaricata* clustered in the same clade and *SdPCS2* diverged earlier, indicating *SdPCS* was originated from *S. divaricata* specific duplication event rather than directly inherited from ancestor species (Fig. 6a). Sequence analysis further confirmed this deduction. *SdPCS2* showed the highest coding sequence (CDS) identity with other syntenic Apiaceous *PCS*s among the three *SdPCS*s (Supplementary Fig. 95). Although the chromosomal position of *SdPCS3* (not anchored to any chromosome) is unknown, the lowest CDS identity with other Apiaceous *PCS*s indicated it was younger than *SdPCS* and *SdPCS2* (Supplementary Fig. 95). Therefore, *SdPCS2* is most likely the direct progenitor of *SdPCS* (Fig. 6c). Notably, we found that *SdPCS2* was nearly not expressed in any tissue (Fig. 6d), and its syntenic genes in other Apiaceous species were almost not expressed, either (Supplementary Table 18). Thus, the proximal duplication and high expression of *SdPCS* profoundly contributed to the biosynthesis of furochromones in the roots of *S. divaricata*.

The absence of functional *PC* gene leads to low furochromone contents in most Apiaceous plants

As shown in Fig. 7a, we did not detect syntenic genes of *SdPC* in the other Apiaceous plants. We retrieved all Apiaceous *CYP736s* and constructed an ML tree. The reported *PpDC* was also included²². A robust clade (BS=100) containing *SdPC* and 18 other potential *PC*s was identified (Fig. 7b, Supplementary Fig. 64). Although DCAR_313054 and As05G02751 were not included in this clade, they were syntenic with several genes including SaDchr05G002066 and LCX7BG003145 (Supplementary Fig. 96). The 11 potential *PC*s from *A. sinensis* lost one exon and was likely to lose the cyclization activity, thus they were not further analyzed. Finally, we cloned and characterized the other 10 potential *PC*s. Strikingly, none of them except for *SdPC* was effective in producing visamminol (4) (Fig. 7c). This observation confirmed our hypothesis that the absence of *PC* genes may lead to the low contents of furochromones in most Apiaceous species. However, we cannot exclude the possibility that these or other potential *PC*s might weakly catalyze the reactions in vivo, as trace furochromones were detected in all Apiaceous plants.

Since *PpDC* participated in the generation of furocoumarins in *Peucedanum praeruptorum*²², we tested the catalytic activities of homologous *PC*s using demethylisuberosin as substrate. *SdPC4*, *AsDC*, *LcPC2* and *DcPC* showed cyclization activities (Supplementary Fig. 97). However, *SdPC* could not catalyze this reaction despite its high sequence similarity with *SdPC4* (Supplementary Fig. 98). Interestingly, these four genes were located at the same syntenic block, which did not include *SdPC* (Supplementary Fig. 96). Thus, *SdPC* is a homologous enzyme with novel function, and its evolutionary origin warrants further investigation in the future.

In conclusion, this work dissected the complete biosynthetic pathway of prim-*O*-glucosylcimifugin and 5-*O*-methylvisamminoside, the major bioactive furochromone glucosides in *S. divaricata*. The functions of seven biosynthetic enzymes were characterized by enzymatic catalysis reactions, and the biosynthetic pathway was verified by de novo biosynthesis of major furochromones in *Nicotiana benthamiana*. Moreover, we explored the evolutionary mechanisms of furochromones biosynthesis in Apiaceae plants. Through comparative metabolic, genomic, and transcriptomic analyses of eight plant species, we found that proximal duplication and high expression of a pentaketide chromone synthase gene *SdPCS*, as well as the presence of

a lineage-specific peucenin cyclase gene *SdPC*, contribute to the abundant and specific accumulation of furochromones in the roots of *S. divaricata*. This work provides critical insights into the biosynthesis of furochromones and serves as a platform for their metabolic engineering production.

Methods

Materials and reagents

The sources of fresh plants of *Saposhnikovia divaricata*, *Centella asiatica*, *Bupleurum chinense*, *Daucus carota*, *Angelica sinensis*, *Apium graveolens*, and *Coriandrum sativum* are given in Supplementary Data 2. We sampled leaves, petioles, and roots of *Saposhnikovia divaricata* for both metabolic analyses and RNA-Seq, with the roots revealing three different growth levels (Supplementary Table 15). For each tissue/stage, three replicates were sampled.

The chemical reference standards and sugar donors used in this study were purchased from YuanYe Biotechnology Co., Ltd. (Shanghai, China). Methanol and acetonitrile (Thermo Fisher Scientific, USA) were of HPLC grade. The conversion rates were determined by HPLC/UV analysis on an Agilent HPLC 1260 instrument. Samples were separated on a Zorbax SB-C18 column (4.6×250 mm, 5 μm, Agilent, USA). The column temperature was 30 °C. To calculate the conversion rates, peak areas of both substrate and product were integrated by Chromeleon® at a certain wavelength. LC/MS analysis was performed on a Q-Exactive quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific, USA).

Genome sequencing, assembly, and annotation

For the PacBio library construction, 15 μg genomic DNA from the leaves of *S. divaricata* was fragmented into approximately 15 kb using g-TUBEs (Covaris, USA). After removing short fragments and single-strand overhangs, the retained fragments were converted into the proprietary SMRTbell library with the PacBio DNA Template Preparation Kit (Pacific Biosciences, CA, USA). Single Molecule Real Time (SMRT) sequencing was performed on a PacBio Sequel II sequencing platform. For Hi-C library construction, chromatin was first fixed in place with formaldehyde in the nucleus and then extracted. The extracted chromatin was digested with DpnII. The 5' overhangs of resulting fragments were then filled in with biotinylated nucleotides, and free blunt ends were ligated. After ligation, the DNA was purified from protein and treated following the Illumina Next Generation manufacturer's instructions. The libraries were subsequently sequenced on Illumina HiSeq X, producing 166.99 Gb 2 × 150 bp paired-end reads. The raw data of PacBio subreads was filtered to HiFi reads by PBccs (v6.4.0) (<https://github.com/PacificBiosciences/ccs>), and subsequently assembled with Hifiasm (v0.16.0)³⁹, which generated a pair of haplotype-resolved assemblies, hap1.p_ctg (1.95 Gb) and hap2.p_ctg (1.93 Gb). We selected the slightly larger hap1.p_ctg for subsequent BUSCO assessment, scaffolding, and annotation. The initial assembled contigs were anchored to chromosomes by 3D-DNA pipeline (v201008)⁴⁰ and further manually adjusted to produce a chromosome-level genome. BUSCO (v5.4.3)⁴¹ was used for benchmarking the genome with the “embryophyte_odb10” database.

EDTA (v2.0.0)⁴² was used to de novo identify, annotate, and classify the repetitive elements in the genome of *S. divaricata*. Prior to protein-coding gene annotation, the annotated repetitive elements in the genome were soft masked with bedtools (v2.28.0)⁴³. RNA-Seq raw reads of *S. divaricata* were filtered with fastx-toolkits (v0.0.14) (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and then assembled through Hisat2 (v2.2.1)⁴⁴ and Stringtie (v2.2.0)⁴⁵. The raw assembly of transcripts was further validated by PASA (v2.5.1)⁴⁶, which were then incorporated into the MAKER (v3.01.03)⁴⁷ pipeline to automatically identify protein-coding genes. Finally, the gene models identified by MAKER were updated by PASA (v2.5.1)⁴⁶. Some genes were recognized and annotated by manual examination. Their names were different

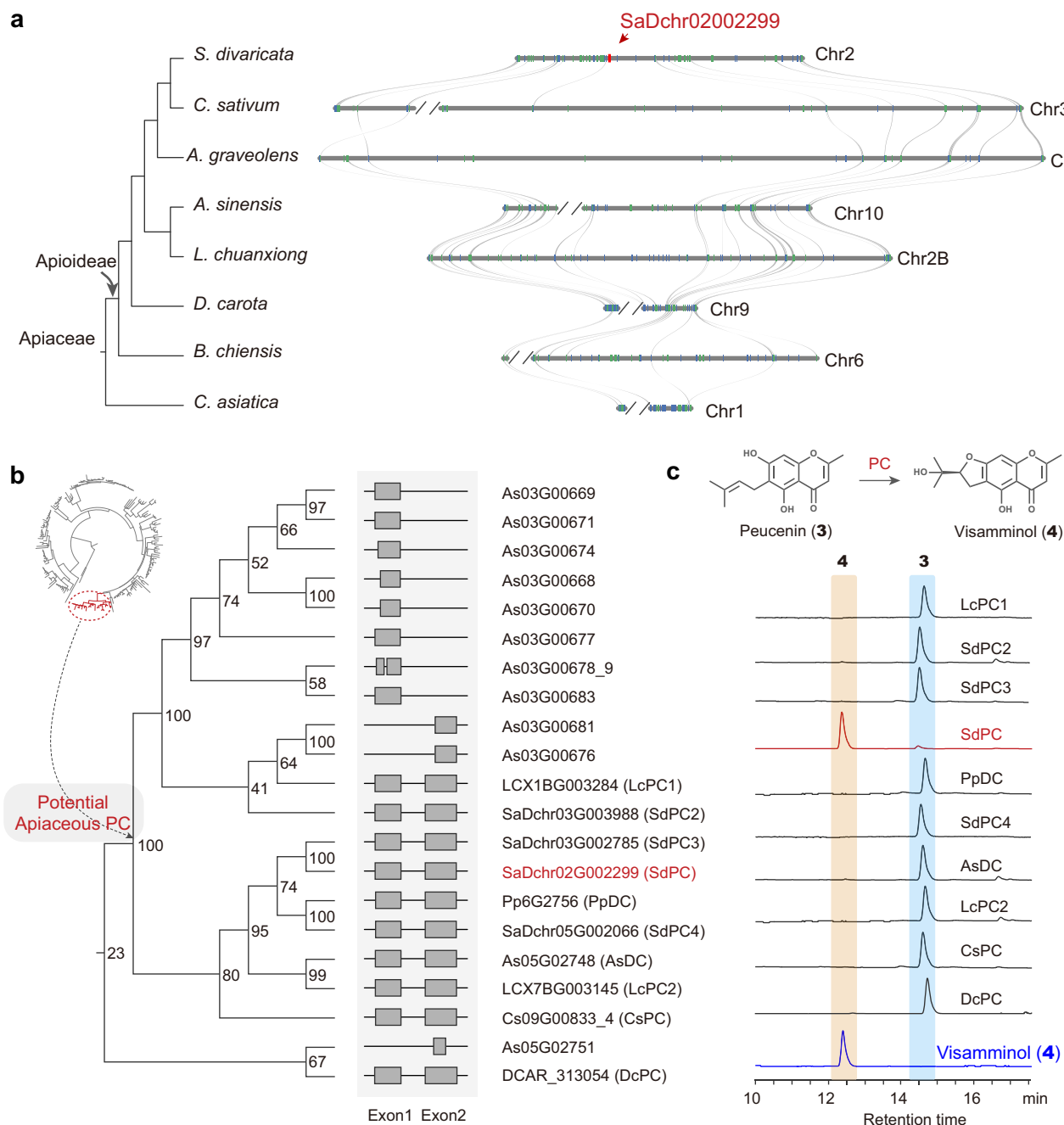


Fig. 7 | The absence of functional PC leads to low furochromone contents in most Apiaceous plants. a Syntenic regions containing *SdPC*. The syntenic gene pairs are connected by grey lines. **b** Phylogenetic relationship and gene structure of

potential Apiaceous PCs. As05G02748 (AsDC) is the same as As05G00644 in the initial annotation version. **c** HPLC/UV chromatograms showing the in vitro enzymatic activity of potential Apiaceous PCs ($\lambda = 280$ nm).

from the other genes that were annotated and named using software. To reveal the relative locations of a manual-annotated gene, we named it by its nearby upstream gene name plus a suffix of downstream_s1. For example, *SaDchr03g003979_downstream_s1* (*SdPT2*) is a manually annotated gene located downstream region of the *SaDchr03g003979*. Function annotations of the protein-coding genes were carried out by BLASTP searches against entries in both NCBI non-redundant protein (NR) (<https://www.ncbi.nlm.nih.gov/>) and Swiss-Prot (<https://www.uniprot.org/>) databases. The prediction of conserved domains for the genes was performed by InterProScan (v5.11-51.0)⁴⁸. The annotations of the GO terms (<http://geneontology.org/>) and KEGG pathways (<https://www.genome.jp/kegg/>) for the genes were annotated with eggNOG-mapper (v2.1.10-0)⁴⁹.

Total RNA isolation, RNA-Seq, and gene expression quantification

The total RNA was extracted with the TranZol™ kit (Transgen Biotech, China) following the manufacturer's instructions, and was used to synthesize the first-stranded complementary DNA (cDNA) with TransScript one-step genomic DNA (gDNA) removal and cDNA synthesis SuperMix (Transgen Biotech, China). The transcriptome data of different tissues of *S. divaricata* were sequenced at Novogene Co., Ltd. (Beijing, China).

The raw RNA-seq reads were filtered in fastp⁵⁰ with default parameters and then mapped to the reference genome of *S. divaricata* by Hisat2 (v2.2.1)⁴⁴. The counts of reads mapping to exons of each gene were calculated by featureCounts⁵¹. The FPKM value of each gene was calculated in R.

Genome-wide mining for furochromone biosynthetic genes

HMMER3 (v3.3.2)⁵² was used to identify 47PKSs, PTs, CYPs, OMTs and UGTs with an *e*-value of $1e^{-6}$. The HMMER profiles PF02797 and PF00195 were used for PKS III search. PF01040 and PF00067 were utilized to identify CYPs and PTs. PF00891 and PF08100 were employed to search OMTs, PF00201 was applied for UGT identification. The possible pseudogenes (length of predicted CDS < 200 amino acids) were discarded. Gene structures of all candidate genes were manually adjusted with IGV-GSaman (<https://gitee.com/CJchen/IGV-SRNA>).

Pearson correlation coefficients and the *P* values between contents of furochromones and the FPKM of genes among different tissues of *S. divaricata* were calculated with the *corr.test* function in R package psych (<https://rdocumentation.org/packages/psych/versions/2.3.3>). The correlation between the expression abundance of each gene and furochromone content was analyzed based on identical sample composition, involving 15 pairs of strict-matched samples. Those unexpressed genes were not incorporated in correlation analysis.

Phylogenetic and microsynteny analyses

ML phylogeny was constructed based on 398 strict single-copy orthologous genes identified by OrthoFinder (v2.5.4)⁵³ to clarify the phylogenetic relationship among the eight Apiaceous species. The protein sequences were aligned by MUSCLE (v5.1.linux64)⁵⁴ and subsequently concatenated by Phylosuite (v1.2.2)⁵⁵. ModelTest-NG (v0.1.7)⁵⁶ was used to detect the best-fit amino acid substitution model, based on which RAXML-NG (v1.1.0)⁵⁷ was employed to construct the ML phylogeny with 1,000 bootstrap analyses. The construction of phylogeny of biosynthetic genes follows the same method above.

The microsyntenic analyses generally followed the methods of Griesmann et al. (2018) and Yang et al. (2023). All *vs. all* blastp (*E*-value = $1e^{-5}$) was conducted for the protein sequences among eight Apiaceous genomes with BLAST (v2.13.0+)⁵⁸. The output protein identity matrix was loaded in JCVI (v1.2.7) to produce collinear gene blocks. Subsequently, we identified the syntenic region containing the furochromone biosynthetic genes (± 100 kb) in each species using the genome of *S. divaricata* as the reference. Because the syntenic retention varied between different species pairs, we compared the syntenic gene pairs for all species pairs and retained those gene pairs demonstrating consistent syntenic relationships. To eliminate the bias induced by mistaken annotation, we manually checked the corrected gene structure in syntenic region and re-organized the microsyntenic gene pairs.

Molecular cloning

The full-length candidate genes were amplified from cDNA with TransStart FastPfu DNA Polymerase (Transgen, China). Candidate genes for PCS, OMT and UGT were recombined in the pET-28a (+) vector (Invitrogen, USA) at BamHI site. Candidate genes for PT, PC and CH were cloned into pESC-Leu vector at BamHI (Invitrogen, USA). Sequences of the primers used in this study are listed in Supplementary Table 19.

Expression of candidate biosynthetic genes

The recombinant plasmids for PCS, OMTs and OGTs were introduced into *E. coli* BL21 (DE3) (Transgen Biotech, China) for heterologous expression. The *E. coli* cells were grown in 500 mL Luria-Bertani medium (JS0666, JSENB, China) containing kanamycin (50 μ g/mL) at 37 °C. After OD₆₀₀ reached 0.4–0.6, the cells were induced with 0.1 mM IPTG at 18 °C. After 18–24 h, the cell pellets were harvested by centrifugation (5632 \times g, 3 min at 4 °C), and then resuspended in 15 mL lysis buffer (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 30 mM imidazole, pH 8.0). Then cells were disrupted by sonication on ice, and the cell debris was removed by centrifugation at 5632 \times g for 50 min at 4 °C. The supernatant was collected and loaded onto a pre-equilibrated

column (His Trap™ HP, 5 mL, GE Healthcare), and eluted with different concentrations of elution buffer (50 mM NaH₂PO₄, pH 8.0, 300 mM NaCl, 30/300 mM imidazole)⁵⁹. The purified protein solution was added with approximately 0.5 mL glycerol (25%) and stored at –80 °C.

The recombinant plasmids for PT, PC and CH were introduced into yeast strain *Saccharomyces cerevisiae* WAT11 for heterologous expression. The yeast cells were grown in synthetic dropin medium without leucine (SD-Leu). Liquid cultures of the recombinant strains were set up by picking a single colony and growing in 50 mL of SD-Leu medium containing 20 g/L glucose at 28 °C overnight. The cells were collected by centrifugation (1000 g, 2 min) and resuspended in 25 mL of SD-Leu medium containing 20 g/L galactose to induce target protein expression for 24–48 hours at 28 °C. The microsomes of yeast cells were prepared as reported³¹.

Enzyme activity assay

The purified proteins and prepared microsomes were used for functional characterization by *in vitro* enzymatic reactions. The reactions were conducted in 100 μ L Tris-HCl buffer (50 mM, pH 8.0) containing 50 μ g purified enzymes or 20 μ L microsomes. The incubation mixtures include substrates (0.1 mM, malonyl-CoA for PCSs, noreugenin for PTs, peucenin for PCs, visamminol for OMTs and CHs, 5-*O*-methylvisamminol for CHs and SdUGT1/2, norcimifugin for OMTs, and cimifugin for SdUGT1/2), and donors/cofactors (0.5 mM, dimethylallyl pyrophosphate (DMAPP) and MgCl₂ for PTs, nicotinamide adenine dinucleotide phosphate (NADPH) for PCs and CHs, S-adenosylmethionine (SAM) and dithiothreitol (DTT) for OMTs, and uridine diphosphate glucose (UDPG) for SdUGT1/2). The reactions continued in a shaking incubator for 2 hours (37 °C for OMTs and UGTs, 30 °C for PCSs, PTs, PCs and CHs). For PCSs, reactions were terminated by adding 10 μ L 20% HCl followed by extraction with 300 μ L ethyl acetate and redissolution in 100 μ L MeOH. The other reactions were terminated by adding 100 μ L ice-cold MeOH. The mixtures were then centrifuged at 21,130 \times g for 20 min. The supernatants were analyzed by HPLC and LC/MS.

Samples were separated on a Zorbax SB-C18 column (4.6 \times 250 mm, 5 μ m, Agilent, USA). The HPLC methods are shown in Supplementary Table 20. LC/MS analysis was performed on a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer equipped with a heated ESI source (Thermo Fisher Scientific, USA). The MS parameters were as follows: sheath gas pressure 45 arb, aux gas pressure 10 arb, discharge voltage 4.5 kV, capillary temperature 350 °C. MS1 resolution was set as 70,000 FWHM, AGC target 1⁺E6, maximum injection time 50 ms, and scan range *m/z* 100–1000. MS2 resolution was set as 17,500 FWHM, AGC target 1⁺E5, maximum injection time 100 ms, NCE 35.

Biochemical properties assay

To optimize the pH value, different reaction buffers with pH from 4.0–6.0 (citric acid-sodium citrate buffer), 6.0–8.0 (Na₂HPO₄-NaH₂PO₄ buffer), 7.0–9.0 (Tris-HCl buffer), and 9.0–11.0 (Na₂CO₃-NaHCO₃ buffer) were tested. To optimize the reaction temperature, the reactions were incubated at 4, 18, 30, 37, 45, or 60 °C. All enzymatic reactions (100 μ L reaction mixtures, the same as those used for enzyme activity assay) were conducted in three parallel experiments (*n* = 3). The reactions were terminated and centrifuged at 21,130 \times g for 20 min for HPLC analysis as described above. The conversion rates in percentage were calculated from peak areas of products and substrates in HPLC/UV chromatograms (Agilent 1260, USA) (The peak area of product divided by the total peak area of product and substrate). The catalytic efficiency of SdPCS was evaluated by the peak area of target product (2).

Determination of kinetic parameters

Reactions were conducted in a final volume of 50 μ L with 50 mM reaction buffer, suitable concentration of protein or microsome,

1 mol/L of saturated donors/cofactors (UDPG for SdUGT1/2, DMAPP and MgCl₂ for SdPT), and different concentrations of substrate (**5** or **8** for SdUGT1/2, **2** for SdPT, **1** for SdPCS) (Supplementary Table 21, Supplementary Fig. 9d, Supplementary Fig. 11d). The reactions were quenched (adding 70 μ L pre-cooled methanol for SdUGT1/2 and SdPT, adding 5 μ L 20% HCl followed by extraction with 300 μ L ethyl acetate and redissolution in 100 μ L MeOH for SdPCS) after incubating at the optimal temperature for a certain reaction time (15 min for SdUGT1/2, 60 min for SdPT and SdPCS), and then centrifuged at 21,130 \times g for 20 min. The supernatants were used for HPLC analysis. All experiments were performed in triplicate. The conversion rates were calculated as described above. To determine the yields of **2** produced by SdPCS, we acquired its regression equation by testing calibration standard solutions (7.78 μ M reference standard **2** diluted by 2, 4, 8, 16, 32, 64, 128 and 256 folds, Supplementary Fig. 99), and calculated the conversion rates of each reaction according to HPLC/UV analysis. The kinetic parameters were calculated with Michaelis-Menten plot fitted by Graphpad Prism 8.0⁶⁰.

Scaled-up enzymatic reactions

To prepare the prenylated product, the reaction mixtures contained 100 μ L buffer (50 mM Tris-HCl, pH 8.0), 0.2 mM noreugenin, 1.0 mM DMAPP, 2.0 mM MgCl₂, and 20 μ L microsome. A total of 1200 parallel tube reactions were conducted. The reactions were performed at 30 °C overnight and terminated by extraction with 4-fold volume of ethyl acetate. The organic solvent was removed under reduced pressure. The residue was dissolved in 1.5 mL of methanol. The products were then purified by reversed-phase semi-preparative HPLC. The structures were characterized by HRMS and extensive 1D and 2D NMR analyses.

To prepare the hydrolyzed product of visamminol-3'-O-glucoside, the reaction mixture contained 20 mL buffer (50 mM NaH₂PO₄-Na₂HPO₄, pH 6.0), 0.5 mM visamminol-3'-O-glucoside, and 200 mg β -glucosidase (Solarbio, Beijing, China). A total of 5 parallel tubes were used. The reactions were performed at 45 °C for 4 hours and terminated by extraction with 4-fold volume of ethyl acetate. The extract was treated as described above.

Crystallization and structural determination

The full-length cDNA of *SdUGT2* was cloned into pET-28a (+) vector. The S-tag of pET28a was removed. A TrxA-tag and 6 \times His-tag followed by thrombin site were added before the N-terminus of the target protein to facilitate purification. The TrxA-His-thrombin-SdUGT2 protein was expressed in *E. coli* (DE3) strain and purified by Ni-affinity chromatography (GE Healthcare). After purification, the recombinant protein was digested by thrombin to remove tag (4 °C, 8 h). The sample was mixed with Ni-NTA affinity beads for the second time to purify the protein. The flow-through was concentrated and then applied to size-exclusion chromatography on a Superdex™ 200 increase 10/300 GL prepac column (GE Healthcare) for further purification. The elution buffer was 20 mM Tris-HCl (pH 7.5) and 50 mM NaCl. Fractions containing SdUGT2 were collected and concentrated to 20 mg/mL, flash-frozen on liquid nitrogen, and then stored in a -80 °C freezer. The purified protein was incubated with 6 mM UDP for 2 h. The crystals of SdUGT2 were obtained after 5 days at 16 °C in hanging drops containing 1 μ L of protein solution and 1 μ L of reservoir solution (0.2 M lithium sulfate monohydrate, 0.1 M Bis-Tris pH 5.25, 28% w/v polyethylene glycol 3,350) (Supplementary Fig. 28). The crystals were flash-frozen in the reservoir solution supplemented with 25% (v/v) glycerol.

The diffraction data of SdUGT2 crystal were collected at beamlines BL19U1 and BL02U1 Shanghai Synchrotron Radiation Facility (SSRF). The data were processed with XDS. The structures were solved by molecular replacement with Phaser. Crystallographic refinement was performed repeatedly with Phenix and COOT. The

refined structures were validated by Phenix and the PDB validation server (<https://validate.rcsb-1.wwpdb.org/>). The final refined structures were deposited in the Protein Data Bank. The diffraction data and structure refinement statistics are shown in Supplementary Table 14.

Molecular docking

Since all the reported UGT structures are highly conserved for the UDP-sugar binding domain, we simulated the SdUGT2/UDPG sugar complex structures by superimposing the UDP parts of UDPG to reported structures. With reference to the docking parameters of UGT71AP2, the axis of grid box for SdUGT1/2 is $x = -31.157$, $y = -21.476$, and $z = -11.443$ ⁶¹. Then we performed Auto-Dock analysis by Lamarckian Genetic Algorithm with default parameters for 250,000 evaluations in 100 cycles, and the other parameters followed the default settings. We selected conformations for further structural analysis according to the binding energies and possibilities for glycosylation reactions to happen. For SdUGT2 in complex with UDPG and **8**, we selected two conformations representing 11-O-glucosylation and 3'-O-glycosylation, respectively, among a set of 46 conformations with the lowest binding energy. For SdUGT2 in complex with UDPG and **9**, we selected one conformation from a set of two conformations with the second lowest binding energy, since no conformation with lower binding energy was suitable for further 3'-O-glycosylation (Supplementary Fig. 100).

De novo biosynthesis of furochromones in *N. benthamiana*

The full-length DNA sequences of SdPCS, SdPT, SdPC, SdOMT, SdCH and SdUGT1/2 were amplified with primers given in Supplementary Table 19. The PCR products were sub-cloned into pDonr207 vectors with the Gateway BP Clonase II Enzyme Mix and then cloned into pEAQ-HT-DEST1 vector with the Gateway LR Clonase II Enzyme Mix according to the manufacturer's instructions⁶². The recombinant pEAQ-HT-DEST1-target gene vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 by chemical conversion method. Single colonies were inoculated at 28 °C and subsequently shaken in LB culture medium (50 μ g/mL kanamycin and 50 μ g/mL rifampicin) until OD₆₀₀ = 0.6. After centrifugation, bacteria were re-suspended in MMA buffer to OD₆₀₀ = 0.2 for each strain. Different strains were mixed for transformation. The infection solution was infiltrated into leaves of 5–6 weeks old *N. benthamiana*. After 7 days, the samples were harvested and freeze-dried. The secondary metabolites were extracted by methanol and analyzed by LC/MS. The contents of compounds **5**, **6**, **8** and **9** were quantified by regression equations. Reference standards **5**, **6**, **8** and **9** were respectively dissolved in DMSO to make solutions of 1 mg/mL, which were 1:1 mixed to obtain the mixed stock solution. The stock solution was serially diluted with methanol containing 4 μ g/mL bergenin as internal standard to obtain calibration standard solutions (diluted by 2, 4, 8, 16, 32, 64, 128, 256, 512, 1,024, 2,048, 4,096, 8,192, 16,384, 32,768, 65,536 and 131,072 folds, respectively). The regression equations of **5**, **6**, **8** and **9** were listed in Supplementary Figs. 101–104. The LC/MS method parameters are listed in Supplementary Table 20. The data were analyzed using Xcalibur™ 4.3 software. The yields of **6** and **9** in each group were the average contents of six independent biological replicates.

Metabolite quantification

The secondary metabolites of different Apiaceae plants were extracted by methanol, and analyzed by LC/MS following the methods mentioned above.

Statistics and reproducibility

No data were excluded from the analyses. The experiments were not randomized.

Reporting summary

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

Data availability

Data supporting the findings of this study are available in the article, supplementary materials, or public database. The gene sequence data generated in this study have been deposited in the NCBI database under the accession numbers listed in Supplementary Table 22. The crystal structure in this study has been deposited in the RCSB PDB database under the accession number: SdUGT2 (8ZNK, <https://www.rcsb.org/structure/8ZNK>). The assembled genome and annotation files of *S. divaricata* are available at figshare [<https://doi.org/10.6084/m9.figshare.25904887.v1>]. The raw sequence data for the PacBio HiFi reads, Hi-C reads and RNA-Seq reads of *Saposhnikovia divaricata* generated in this study have been deposited in the Genome Sequence Archive at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation under BioProject number [PRJCA036214](https://www.ncbi.nlm.nih.gov/bioproject/PRJCA036214). The sources of genome data and RNA-seq data of other Apiaceous plants are listed in Supplementary Data 3, 4. Source data are provided with this paper.

References

- Sun, Y. et al. New chromones from the roots of *Saposhnikovia divaricata* (Turcz.) Schischk with anti-inflammatory activity. *Bioorg. Chem.* **134**, 106447 (2023).
- Abu-Hashem, A. A. & Youssef, M. M. Synthesis of new visnagen and khellin furochromone pyrimidine derivatives and their anti-inflammatory and analgesic activity. *Molecules* **16**, 1956–1972 (2011).
- Zhang, T. et al. Multi-omics reveals that 5-O-methylvisammioside prevention acute liver injury in mice by regulating the TNF/MAPK/NF- κ B/arachidonic acid pathway. *Phytomedicine* **128**, 155550 (2024).
- Abdelhazef, O. M., Abedelatif, N. A. & Badria, F. A. DNA binding, antiviral activities and cytotoxicity of new furochromone and benzofuran derivatives. *Arch. Pharm. Res.* **34**, 1623–1632 (2011).
- Edwards, A. M. & Howell, J. B. The chromones: history, chemistry and clinical development. A tribute to the work of Dr R. E. C. Altounyan. *Clin. Exp. Allergy* **30**, 756–774 (2000).
- Amen, Y., Elsbaey, M., Othman, A., Sallam, M. & Shimizu, K. Naturally occurring chromone glycosides: sources, bioactivities, and spectroscopic features. *Molecules* **26**, 7646 (2021).
- Kreiner, J., Pang, E., Lenon, G. B. & Yang, A. W. H. *Saposhnikovia divaricata*: a phytochemical, pharmacological, and pharmacokinetic review. *Chin. J. Nat. Med.* **15**, 255–264 (2017).
- Khalil, N., Bishr, M., Desouky, S. & Salama, O. Ammi visnaga L., a potential medicinal plant: a review. *Molecules* **25**, 301 (2020).
- Sun, Y., Yang, A. W. H. & Lenon, G. B. Phytochemistry, ethnopharmacology, pharmacokinetics and toxicology of *Cnidium monnieri* (L.) Cusson. *Int. J. Mol. Sci.* **21**, 1006 (2020).
- Commission, C. P. *Pharmacopoeia of the People's Republic of China*. Vol. 1, p156 (China Medical Science and Technology Press, 2020).
- Shang, Z. P. et al. Bioactive compounds of Jingfang Granules against SARS-CoV-2 virus proteases 3CL^{pro} and PL^{pro}. *J. Peking. Univ. (Health Sci.)* **54**, 907–919 (2022).
- Liang, X. et al. Effects and potential mechanisms of *Saposhnikovia divaricata* (Turcz.) Schischk. on type I allergy and pseudoallergic reactions in vitro and in vivo. *J. Ethnopharmacol.* **318**, 116942 (2024).
- Li, X., Qi, H., Zhang, X., Liang, H. & Zeng, N. Jing-Fang *n*-butanol extract and its isolated JFNE-C inhibit ferroptosis and inflammation in LPS induced RAW264.7 macrophages via STAT3/p53/SLC7A11 signaling pathway. *J. Ethnopharmacol.* **316**, 116689 (2023).
- Wang, Y. et al. Prim-O-glucosylcimifugin ameliorates aging-impaired endogenous tendon regeneration by rejuvenating senescent tendon stem/progenitor cells. *Bone Res.* **11**, 784–802 (2023).
- Erst, A. S., Petrova, N. V., Kaidash, O. A., Wang, W. & Kostikova, V. A. The genus *Eranthis*: prospects of research on its phytochemistry, pharmacology, and biotechnology. *Plants (Basel)* **12**, 3795 (2023).
- Urbagarova, B. M. et al. Chromones and coumarins from *Saposhnikovia divaricata* (Turcz.) Schischk. growing in Buryatia and Mongolia and their cytotoxicity. *J. Ethnopharmacol.* **261**, 112517 (2020).
- Wang, X. H. et al. Identification of a diarylpentanoid-producing polyketide synthase revealing an unusual biosynthetic pathway of 2-(2-phenylethyl) chromones in agarwood. *Nat. Commun.* **13**, 348 (2022).
- Abe, I. et al. Structure-based engineering of a plant type III polyketide synthase: formation of an unnatural nonaketide naphthopyrone. *J. Am. Chem. Soc.* **129**, 5976–5980 (2007).
- Harrison, P. G., Bailey, B. K. & Steck, W. Biosynthesis of furanochromones. *Can. J. Biochem.* **49**, 964–970 (1971).
- de Bruijn, W. J. C., Levisson, M., Beekwilder, J., van Berkel, W. J. H. & Vincken, J. P. Plant aromatic prenyltransferases: tools for microbial cell factories. *Trends Biotechnol.* **38**, 917–934 (2020).
- Villard, C. et al. A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution. *N. Phytol.* **231**, 1923–1939 (2021).
- Zhao, Y. et al. Two types of coumarins-specific enzymes complete the last missing steps in pyran- and furanocoumarins biosynthesis. *Acta Pharm. Sin. B* **14**, 869–880 (2024).
- Li, Q. et al. The chromosome-scale assembly of the *Notopterygium incisum* genome provides insight into the structural diversity of coumarins. *Acta Pharm. Sin. B* **14**, 3760–3773 (2024).
- Wang, K. et al. Three types of enzymes complete the furanocoumarins core skeleton biosynthesis in *Angelica sinensis*. *Phytochemistry* **222**, 114102 (2024).
- Tiwari, P., Sangwan, R. S. & Sangwan, N. S. Plant secondary metabolism linked glycosyltransferases: an update on expanding knowledge and scopes. *Biotechnol. Adv.* **34**, 714–739 (2016).
- Liu, Y., Fernie, A. R. & Tohge, T. Diversification of chemical structures of methoxylated flavonoids and genes encoding flavonoid-O-methyltransferases. *Plants (Basel)* **11**, 564 (2022).
- Khouri, H. E., Tahara, S. & Ibrahim, R. K. Partial-purification, characterization, and kinetic-analysis of isoflavone 5-O-methyltransferase from yellow lupin roots. *Arch. Biochem. Biophys.* **262**, 592–598 (1988).
- Itoh, N., Iwata, C. & Toda, H. Molecular cloning and characterization of a flavonoid-O-methyltransferase with broad substrate specificity and regioselectivity from *Citrus depressa*. *BMC Plant Biol.* **16**, 180 (2016).
- Isin, E. M. & Guengerich, F. P. Complex reactions catalyzed by cytochrome P450 enzymes. *Biochim. Biophys. Acta* **1770**, 314–329 (2007).
- Wang, Z. H. et al. Genomic, transcriptomic, and metabolomic analyses provide insights into the evolution and development of a medicinal plant *Saposhnikovia divaricata* (Apiaceae). *Hortic. Res.* **11**, uhae 105 (2024).
- Alagoz, Y., Mi, J., Balakrishna, A., Almarwaey, L. & Al-Babili, S. Characterizing cytochrome P450 enzymes involved in plant apocarotenoid metabolism by using an engineered yeast system. *Method Enzymol.* **671**, 527–552 (2022).
- Shang, Z. et al. Characterization of prenylated phenolics in *Glycyrrhiza uralensis* by offline two-dimensional liquid chromatography/mass spectrometry coupled with mass defect filter. *J. Pharm. Biomed. Anal.* **220**, 115009 (2022).
- Marumoto, S. & Miyazawa, M. Structure-activity relationships for naturally occurring coumarins as β -secretase inhibitor. *Bioorg. Med. Chem.* **20**, 784–788 (2012).

34. Zhang, M. et al. Functional characterization and structural basis of an efficient di-C-glycosyltransferase from *Glycyrrhiza glabra*. *J. Am. Chem. Soc.* **142**, 3506–3512 (2020).
35. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).
36. Adasme, M. F. et al. PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA. *Nucleic Acids Res* **49**, 530–534 (2021).
37. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
38. Zhou, X. et al. Phytoene synthase: the key rate-limiting enzyme of carotenoid biosynthesis in plants. *Front. Plant Sci.* **13**, 884720 (2022).
39. Cheng, H. et al. Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335 (2022).
40. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
41. Evgeny, Z. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **10**, 4647–4654 (2021).
42. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
43. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
44. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
45. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
46. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
47. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* **48**, 4.11.11–4.11.39 (2014).
48. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
49. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
50. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107 (2023).
51. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
52. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**, e121 (2013).
53. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
54. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
55. Zhang, D. et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **20**, 348–355 (2020).
56. Darriba, D. et al. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).
57. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
58. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
59. Labrou, N. E. Protein purification: an overview. *Methods Mol. Biol. (Clifton, N. J.)* **1129**, 3–10 (2014).
60. Hoare, S. R. J. Receptor binding kinetics equations: Derivation using the Laplace transform method. *J. Pharmacol. Toxicol. Methods* **89**, 26–38 (2018).
61. Wang, Z. et al. Functional characterization, structural basis, and protein engineering of a rare flavonoid 2'-O-glycosyltransferase from *Scutellaria baicalensis*. *Acta Pharm. Sin. B* **14**, 3746–3759 (2024).
62. 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 Key Research and Development Program of China (No. 2023YFA0914100 to M.Y., and No. 2023YFA0915800 to L.W.), Beijing Natural Science Foundation (No. QY23076 to J.L.Z., and 83001Y0439 to C.X.Z.), and National Natural Science Foundation of China (No. 81725023 to M.Y.). We thank Dr. Rong-shen Wang and Xi-ran Zhang of Ye Lab and Xun-meng Feng and Jiao-jiao Ji of Wang Lab for their technical assistance. We also thank the staff at BL19U1/BLO2U1 beamlines at SSRF of the National Facility for Protein Science in Shanghai (NFPS), Shanghai Advanced Research Institute, Chinese Academy of Sciences, for providing technical support in X-ray diffraction data collection and analysis. We use the element syringe from SciDraw website in Fig. 4b (<https://scidraw.io/about/>).

Author contributions

M.Y. and L.W. designed the research. M.Y., L.W., J.L.Z. and C.X.Z. acquired fundings for this study. B.N. contributed to the genome assembling and related bioinformatic analysis. J.L.Z. and H.Y.L. designed and performed the major experiments and data analysis. Z.L.W., Y.G.T., L.Q.L., W.Z.X., Z.W.H., W.K.S., X.X.H., M.Z., H.T.W. and Q.Y.L. assisted with experiments. M.Y., W.L., J.L.Z., H.Y.L. and B.N. wrote the manuscript. All authors have given approval to the final version of the manuscript.

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/s41467-025-58498-8>.

Correspondence and requests for materials should be addressed to Li Wang or Min Ye.

Peer review information *Nature Communications* thanks the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions 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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025