nature communications

Article

https://doi.org/10.1038/s41467-025-59774-3

Elucidation of the biosynthetic pathway of hydroxysafflor yellow A

Received: 9 September 2024

Accepted: 29 April 2025

Published online: 14 May 2025

Check for updates

Zi-Long Wang[1,6], Hao-Tian Wang[1,6], Guowei Chang[2], Guo Ye[1], Meng Zhang[1], Jiang Chen[3] & Min Ye[1,4,5]

Hydroxysafflor yellow A (HSYA) is a clinical investigational new drug for the treatment of acute ischemic stroke. It has a unique quinochalcone di-*C*-glycoside structure and is exclusively found in the flowers of safflower (*Carthamus tinctorius*). To date, little is known about the biosynthesis of HSYA. In this work, we characterize four key biosynthetic enzymes from *C. tinctorius*: CtF6H (6-hydroxylation of naringenin to produce carthamidin), CtCHI1 (isomerization between carthamidin and isocarthamidin), CtCGT (flavonoid di-*C*-glycosyltransferase), and Ct2OGD1 (2-oxoglutarate-dependent dioxygenase). Notably, Ct2OGD1 coordinates with CtCGT to convert carthamidin or isocarthamidin to HSYA. Functions of these genes are confirmed by VIGS (virus-induced gene silencing) in *C. tinctorius*, de novo biosynthesis of HSYA in *Nicotiana benthamiana*, semi-synthesis in yeast, and in vitro enzyme assays. We further find that the simultaneous presence and high expression of the above four key genes, together with the absence of F2H (flavanone 2-hydroxylase) genes, are essential for the biosynthesis of HSYA, and thus interpret mechanisms for the unique presence of HSYA in safflower. This work elucidates the biosynthetic pathway of HSYA and provides a foundation for the green and efficient production of this valuable medicinal natural product.

C arthamus tinctorius L. (safflower) is a popular medicinal plant worldwide. Its flowers have been used as the traditional Chinese herbal medicine Hong-Hua to treat cardiovascular and cerebrovascular diseases for a long history[1]. One predominant bioactive compound is hydroxysafflor yellow A (HSYA), a unique quinochalcone di-*C*-glycoside. HSYA exhibits a variety of biological activities. According to experimental investigations, HSYA could alleviate atherosclerosis, cerebral ischemia and reperfusion injury, myocardial ischemia, vascular injury, alcohol-induced liver injury, and diabetes[2–4]. Safflor Yellow Injection contains 85% HSYA (*w*/*w*), and has been used as a prescription drug to treat coronary heart disease and angina pectoris in China. Recently, a phase III clinical trial of HSYA Injection for the treatment of acute ischemic stroke has been completed, and its new drug application (NDA) has been submitted to the National Medical Products Administration of China (https://english.nmpa.gov.cn/). HSYA is a widely concerned natural product with promising clinical value.

HSYA was first reported in 1981, and its unique chemical structure was not correctly identified until 2013[5]. Thus far, only a few quinochalcone *C*-glycosides have been reported in nature, and all of them are from the flowers of *C. tinctorius*[6,7]. The total chemical synthesis of HSYA has attracted a lot of organic chemists, but still remains a great challenge[8,9]. It appears to be even more challenging to elucidate the biosynthetic pathway of HSYA. No information is available on how the

[1]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China. [2]Key Laboratory of Seed Innovation, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China. [3]State Key Laboratory of Southwestern Chinese Medicine Resources, College of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China. [4]Peking University-Yunnan Baiyao International Medical Research Center, 38 Xueyuan Road, Beijing 100191, China. [5]Southwest United Graduate School, Kunming 650092, China. [6]These authors contributed equally: Zi-Long Wang, Hao-Tian Wang.
e-mail: yemin@bjmu.edu.cn

quinochalcone structure is formed, and how the glycosyl group is attached to the hydroxylated and sp^3-hybridized carbon atom. Although a number of novel *C*-glycosyltransferases have been reported in the past several years, all of them could only attach a glycosyl moiety to an sp^2 carbon of a benzene ring[10]. A common upstream chalcone synthase responsible for naringenin chalcone biosynthesis has been identified from *C. tinctorius*, and it is currently the only reported enzyme probably involved in the biosynthesis of HSYA[11]. Recently, the chromosome-scale genome of *C. tinctorius* has been reported, which provides a platform to investigate the biosynthesis of HSYA[12,13].

In this work, we characterized from *C. tinctorius* four key genes involved in the biosynthesis of HSYA, namely *CtF6H* (flavanone 6-hydroxylase, cytochrome P450), *CtCGT* (flavonoid di-*C*-glycosyltransferase, UGT), *Ct2OGD1* (2-oxoglutarate-dependent dioxygenase, 2OGD), and *CtCHI1* (chalcone-flavanone isomerase, CHI). The functions of these genes were identified through de novo biosynthesis of HSYA in *Nicotiana benthamiana*, in vitro enzymatic assays, and yeast expression systems. The unique presence of HSYA in safflower was also interpreted by bioinformatics analysis.

Results

Bioinformatics analysis

To discover candidate biosynthetic genes, we determined HSYA in different parts of safflower by liquid chromatography coupled with mass spectrometry (LC/MS) (Fig. 1a). The results indicated that HSYA was only present in the flowers. According to its chemical structure, HSYA may be derived from a chalcone, which is usually biosynthesized by a chalcone synthase (CHS) from coumaroyl-CoA and malonyl-CoA. Subsequently, HSYA could be generated from the chalcone through hydroxylation, *C*-glycosylation, and dearomatization. A proposed biosynthetic pathway is shown in Fig. 1b. Then, we obtained the transcriptome data of the budding flower, blooming flower, calyx, and leaf, respectively. We further discovered eight transcripts of chalcone synthases, which belonged to *CtCHS1*, *CtCHS2*, and *CtCHS3*[14]. Through comparative analysis, we found that most of the transcripts of *CtCHS1* and *CtCHS2* genes showed high expression levels in the budding flowers but not the blooming flowers (Fig. 1c and Supplementary Table 1), indicating the expression of CHS decreased remarkably as the flowers mature.

Based on our proposed biosynthetic pathway, we deduce that UDP-dependent glycosyltransferases (UGTs) and cytochrome P450 enzymes may participate in the biosynthesis of HSYA. We recruited 306 transcripts of UGT and 616 transcripts of P450 genes from the transcriptome using the features of PF00201 and PF00067 in the Pfam database, respectively[15]. The total expression levels of *CtCHS1* and *CtCHS2* were used as the bait for co-expression analysis. For *CtCHS1*, only a few genes showed a strong correlation ($r \geq 0.8$, Pearson correlation coefficient) (Supplementary Table 2). In contrast, *CtCHS2* exhibited a high correlation with 22 UGT and 24 P450 genes, which were tentatively considered as candidate genes. We further analyzed their expression levels in different tissues (Supplementary Tables 3, 4). Two candidate UGTs (*CtAH10T0218300.1* and *MSTRG.178.6*) and one P450 (*CtAH11T0228700.1*) showed high expression levels in the budding flowers (Fig. 1d). Phylogenetic analysis indicated MSTRG.178.6 was clustered with 2-hydroxyflavanone *C*-glycosyltransferases (CGTs) (Supplementary Fig. 1), and CtAH11T0228700.1 was clustered with flavonoid hydroxylases (Supplementary Fig. 2)[16]. Moreover, MSTRG.178.6 contains the conserved DPF motif characteristic of type I CGTs[10], which is composed of aspartic acid, proline, and phenylalanine.

Functional characterization of CGT and P450 enzymes

To identify the in vitro function of CGT and P450, we cloned and expressed *MSTRG.178.6* and *CtAH11T0228700.1* in *E. coli* and WAT11 yeast, respectively (Supplementary Data 1). MSTRG.178.6 (CtCGT) was purified using Ni affinity chromatography (Supplementary Fig. 3). It could accept phloretin (**1**) and 2-hydroxynaringenin (**2**) but not naringenin or naringenin chalcone as substrate to generate mono-*C*-glycosides and di-*C*-glycosides (Fig. 2a and Supplementary Figs. 4–6). The structures of **1a**, **1b**, and **2a** were identified as nothofagin (mono-*C*-glycosylated phloretin), 3′,5′-di-*C*-glycosylphloretin, and 2-hydroxynaringenin mono-*C*-glycoside by comparison with reference standards, respectively. Product **2b** could spontaneously undergo dehydration to form vicenin-2[17,18], which allowed its identification as 6,8-di-*C*-glycosylated 2-hydroxynaringenin (Supplementary Fig. 7). CtCGT exhibited its maximum activity at pH 9.0 (50 mM Tris-HCl) and 45 °C (Supplementary Fig. 8). The reaction was independent of divalent metal ions. The apparent K_m values for **1** and **1a** were 1.86 μM and 284 μM using the "Michaelis-Menten" method, with saturated UDP-Glc as sugar donor (Supplementary Fig. 9). CtCGT showed similar functions as previously reported CGTs of the UGT708 family[17,18], and was named UGT708U8 by the UGT Nomenclature Committee. Next, we extracted the microsomes containing CtAH11T0228700.1 (CtF6H, CYP706S4) from WAT11 yeast. With the addition of NADPH, CtF6H could catalyze 6-hydroxylation of apigenin (**3**) to generate scutellarein (**3a**) (Fig. 2a and Supplementary Fig. 10). CtF6H exhibited its maximum activity at 4 °C, and the catalytic activity decreased significantly at increased temperatures (Supplementary Fig. 11). The above results proved that CtCGT and CtF6H exhibited *C*-glycosylation and hydroxylation activities, respectively.

To investigate whether *CtCGT* and *CtF6H* are involved in the biosynthesis of HSYA in vivo, we employed virus-induced gene silencing (VIGS) technology[19]. Two-week-old seedlings were treated by *Agrobacterium*-mediated infiltration to silence a 300–400 bp DNA fragment at the N-terminal of these genes (Fig. 2b). Around two months later at the budding period, the flowers were analyzed by LC/MS (Fig. 2c). The contents of HSYA decreased by 29.6% and 30.8%, respectively, in the VIGS-*CtCGT* and VIGS-*CtF6H* groups, when compared with the empty vector (EV) control group (Fig. 2d and Supplementary Fig. 12). Meanwhile, the expression levels of *CtCGT* and *CtF6H* were 60% and 42.9% lower than EV in VIGS-*CtCGT* and VIGS-*CtF6H* groups (Fig. 2e), respectively. It was noteworthy that the expressions of *CtF6H* and *CtCGT* also decreased in VIGS-*CtCGT* and VIGS-*CtF6H* groups, respectively, similar to our previous report[17]. The above results proved that *CtCGT* and *CtF6H* participated in the biosynthesis of HSYA in *C. tinctorius*.

Ct2OGD1, together with CtF6H and CtCGT realized the biosynthesis of HSYA in *Nicotiana benthamiana*

Nicotiana benthamiana is an efficient platform to validate plant gene functions[20–22]. We introduced five upstream genes (*AtPAL*, *AtC4H*, *At4CL*, *AtCHS*, and *AtCHI*) into *N. benthamiana* to synthesize the flavonoid skeleton[23], along with two genes (*pgm*, *galU*) for UDP-Glc biosynthesis, co-expressed with *CtCGT* and *CtF6H* (Supplementary Table 5). However, no product was detected. There should be additional unknown genes participating in the biosynthesis of HSYA. We extracted all high-expression genes (TPM ≥ 10) in the budding flower, and constructed a library of 9307 transcripts. Using *CtCGT* and *CtF6H* as the "bait", we conducted co-expression analysis and obtained 1227 transcripts (Fig. 3a, $r \geq 0.8$, Pearson correlation coefficient). Among them, 34 transcripts were annotated as isomerization, oxidation or reduction functions according to Pfam, NR, and Swiss-Prot databases (Supplementary Data 2). In addition, we obtained 29 other transcripts annotated with similar functions (Supplementary Data 3), which showed low correlations with *CtCGT* or *CtF6H* but exhibited high expression in the budding flower.

The above mentioned 63 candidate genes were randomly distributed into eight groups and co-infiltrated with *CtCGT*, *CtF6H*, flavonoid skeleton synthetic module, and UDP-Glc synthetic module into

https://doi.org/10.1038/s41467-025-59774-3

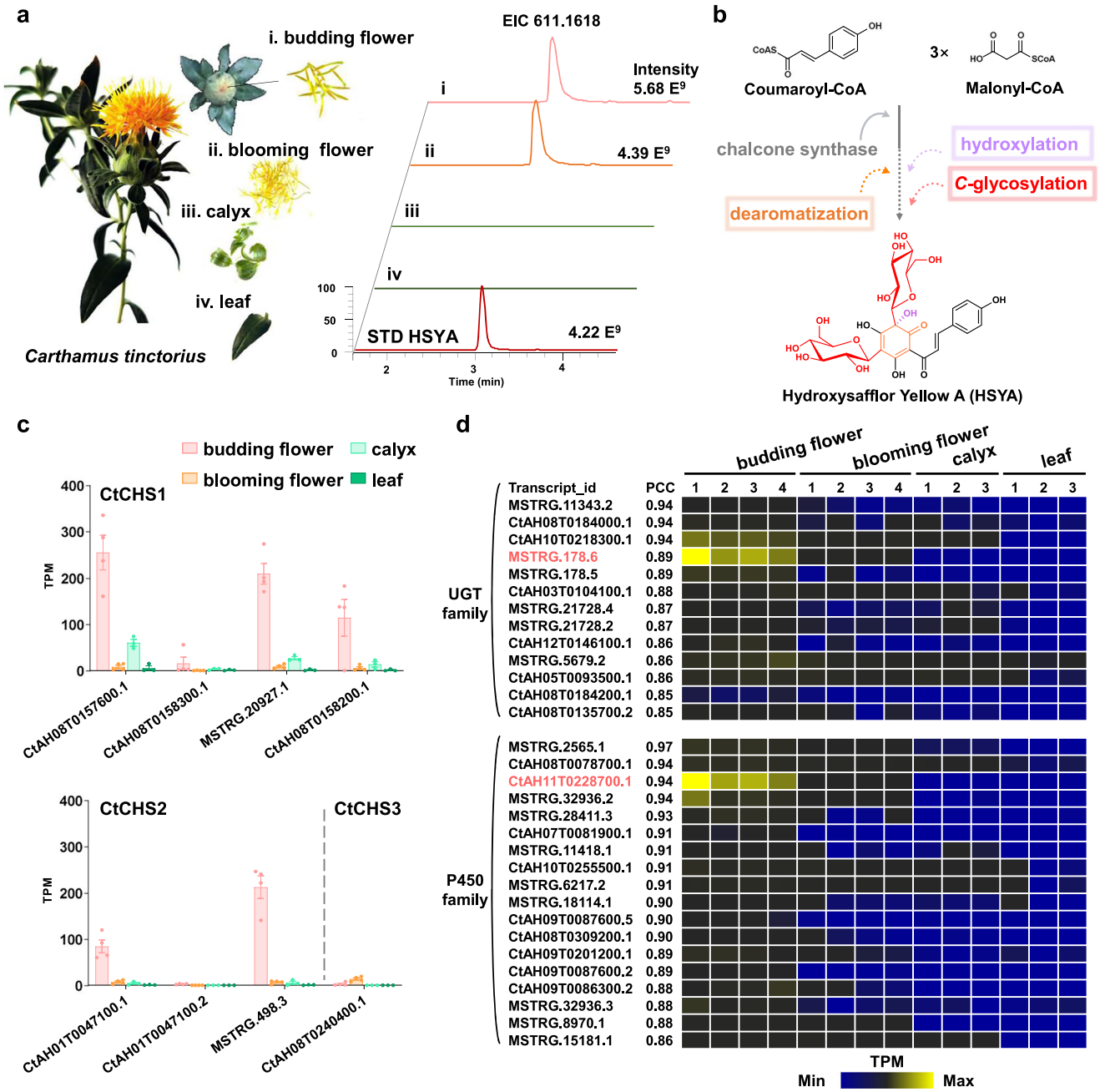

Fig. 1 | Bioinformatic analysis of candidate genes involved in the biosynthesis of HSYA. a LC/MS analysis of the budding flower, blooming flower, calyx, and leaf of *Carthamus tinctorius*, showing extracted ion chromatograms (EICs, *m/z* 611.1618) of HSYA. The peak intensity is shown on the EIC plot. **b** A proposed biosynthetic pathway of HSYA. **c** Expression levels of *CtCHS1*, *CtCHS2*, and *CtCHS3* in the transcriptomes of different parts of *C. tinctorius*. For budding flower and blooming flower, $n = 4$, four biologically independent samples were tested; For calyx and leaf, $n = 3$, three biologically independent samples were tested; The data are presented as mean values ± SEM. TPM, transcripts per million. For the raw data, see Supplementary Table 1. **d** Expression levels of candidate UGTs and P450s in the transcriptomes of different parts of *C. tinctorius*. Three to four replicates were used. PCC Pearson correlation coefficient. For the raw data, see Supplementary Tables 3, 4.

N. benthamiana, respectively (Supplementary Table 6). For each group, leaf disks from the infiltrated parts were collected 7 days post-infiltration, and the samples were extracted and analyzed by LC/MS (Fig. 3b and Supplementary Fig. 13). Among the 8 groups, only group 4 generated one product exhibited an $[M\text{-}H]^-$ ion at *m/z* 611, which could yield a diagnostic fragment for *C*-glycosides at *m/z* 491 ($[M\text{-}H\text{-}120]^-$) in the MS/MS spectrum[24]. By comparing with a reference standard, this product was identified as HSYA (Supplementary Fig. 14). Then, we screened each of the 8 genes in group 4, and found that only CtAH01T0057600.1 (Ct_3) could produce HSYA (Supplementary Fig. 15).

CtAH01T0057600.1 was annotated as a flavanone 3-hydroxylase of the 2-oxoglutarate-dependent dioxygenase (2OGD) family, and was clustered with FNSI and F3H in phylogenetic analysis (Supplementary Fig. 16). It was named as Ct2OGD1. Its role in safflower was further evaluated by VIGS. In the VIGS-*Ct2OGD1* group, the expression level of *Ct2OGD1* and contents of HSYA were 34.7% and 26.6% lower than EV group (Fig. 2d, e). Moreover, the expression levels of *CtCGT* and *CtF6H* decreased by 37.6% and 55.4%, respectively. The above results proved that *Ct2OGD1* participated in the biosynthesis of HSYA in safflower.

https://doi.org/10.1038/s41467-025-59774-3

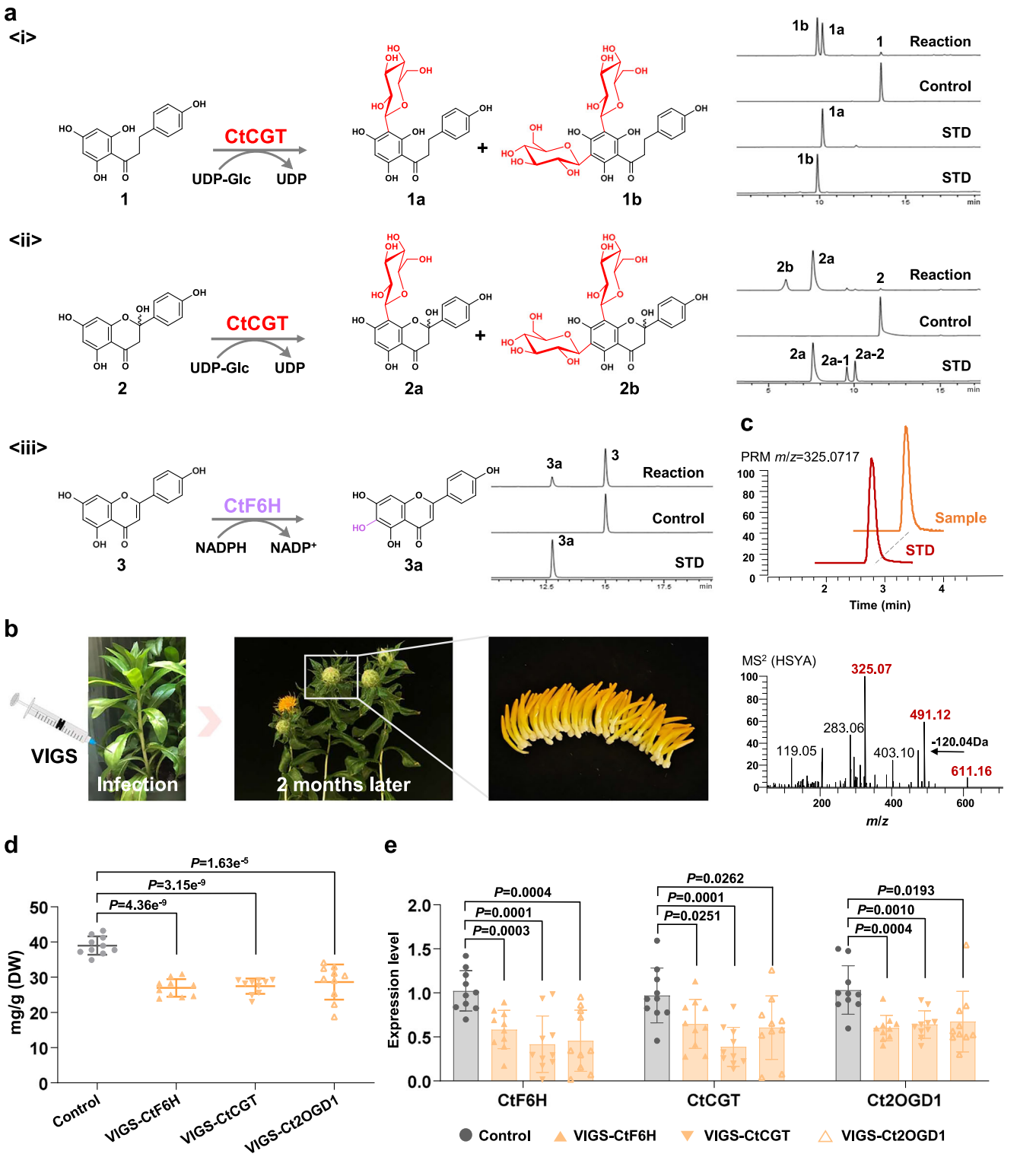

Fig. 2 | Functional characterization of CtCGT and CtF6H. a Catalytic functions of CtCGT and CtF6H in vitro. HPLC chromatograms of enzyme catalytic products of CtCGT using **1** (<i>) or **2** (<ii>) as sugar acceptor and UDP-Glc as sugar donor, and of CtF6H using **3** (<iii>) as acceptor and NADPH as donor. The ultraviolet absorption wavelengths for **1**/**2** and **3** were 300 and 340 nm, respectively. **1a** (nothofagin), **1b** (3′,5′-di-*C*-glycosylphloretin), **2a** (2-hydroxynaringenin mono-*C*-glycoside), **2a-1** (isovitexin), **2a-2** (vitexin), **3a** (scutellarein). **b** The workflow for VIGS experiments in *C. tinctorius*. **c** LC/MS analysis of VIGS samples, showing parallel reaction monitoring (PRM) chromatograms of HSYA, and (-)-ESI-MS/MS spectrum of $[M\text{-}H]^-$ ion at *m/z* 611. **d** The contents of HSYA in *C. tinctorius* upon VIGS treatment ($n = 10$, ten biologically independent samples were tested; The data were presented as mean values ± SD; Statistical significance was analyzed using a two-tailed *t*-test.). **e** Expression levels of *CtF6H*, *CtCGT*, and *Ct2OGD1* in the EV and VIGS groups ($n = 10$, ten biologically independent samples were tested; The data were presented as mean values ± SD; Statistical significance was analyzed using a two-tailed *t*-test.). The source data underlying Fig. 2d, e are provided in a Source Data file.

https://doi.org/10.1038/s41467-025-59774-3

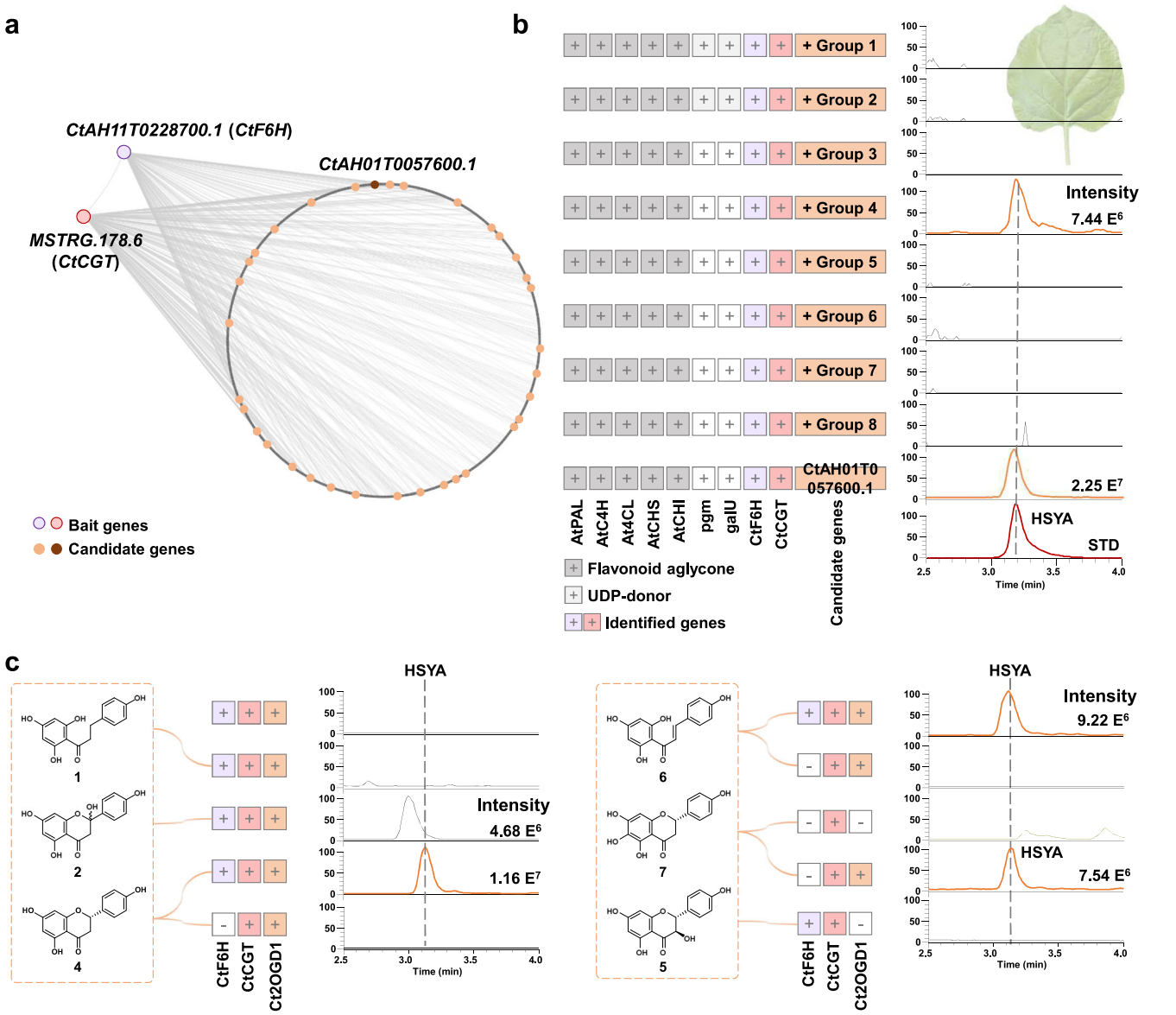

Fig. 3 | De novo biosynthesis of HSYA in *Nicotiana benthamiana*. **a** Co-expression analysis of *C. tinctorius* transcriptome using *CtCGT* and *CtF6H* as bait genes. The co-expression result was visualized by Cytoscope. **b** Screening of candidate genes involved in the biosynthetic pathway of HSYA using *N. benthamiana* as chassis, showing extracted ion chromatograms (EIC: *m/z* 611.1618) of HSYA. The group information is available in Supplementary Table 6. **c** Screening of key precursors in HSYA biosynthesis, showing extracted ion chromatograms (EIC: *m/z* 611.1618) of HSYA. The peak intensity is shown on the EIC plot.

Naringenin and carthamidin act as key precursors in the biosynthesis of HSYA

To further explore the biosynthetic precursor of HSYA, we co-infiltrated the combination of *CtCGT*, *CtF6H*, and *Ct2OGD1* into the leaves of *N. benthamiana*. Three days later, we separately infiltrated potential precursors **4** and **6**, along with the common precursors (**1** and **2**) for flavonoid *C*-glycoside biosynthesis, into the leaves. Four days post-infiltration, leaf disks were collected for LC/MS analysis (Fig. 3c and Supplementary Fig. 17). When all three genes were present, only the addition of naringenin (**4**) or naringenin chalcone (**6**) could produce HSYA. We further found that naringenin chalcone was not stable in *N. benthamiana* or in vitro reaction buffer solution (PBS, pH=7.4), and could be easily transformed to naringenin (Supplementary Figs. 18, 19), while the opposite reaction was unlikely to occur. Thus, we deduce that naringenin is the key biosynthetic precursor for HSYA.

Next, we used naringenin as a substrate to test the functions of CtCGT, CtF6H, and Ct2OGD1 using enzymatic reactions. Ct2OGD1 and CtF6H could catalyze the hydroxylation of naringenin to generate 3-hydroxynaringenin (dihydrokaempferol, **5**) and 6-hydroxynaringenin (carthamidin, **7**), respectively (Supplementary Figs. 20–23). As CtCGT could not catalyze the *C*-glycosylation of naringenin (Supplementary Fig. 6), we infiltrated *CtCGT*/*CtF6H*/dihydrokaempferol (**5**) and *CtCGT*/*Ct2OGD1*/carthamidin (**7**) into *N. benthamiana*, respectively. HSYA could be synthesized only in the carthamidin group (Fig. 3c and Supplementary Fig. 17). This result suggested that Ct2OGD1 did not catalyze 3-hydroxylation, but instead functioned together with CtCGT to catalyze a novel last-step reaction leading to the biosynthesis of HSYA. While Ct2OGD1 had been reported as a flavanone 3-hydroxylase (CtF3H)[25], it was renamed in this work for easy understanding.

CtCHI improves the biosynthetic yield of HSYA in vitro

To verify the function of CtCGT and Ct2OGD1 in vitro, we co-incubated carthamidin, UDP-Glc, Fe^{2+}, ascorbic acid, α-ketoglutarate (α-KG), and purified CtCGT and Ct2OGD1, and observed the generation of HSYA,

https://doi.org/10.1038/s41467-025-59774-3

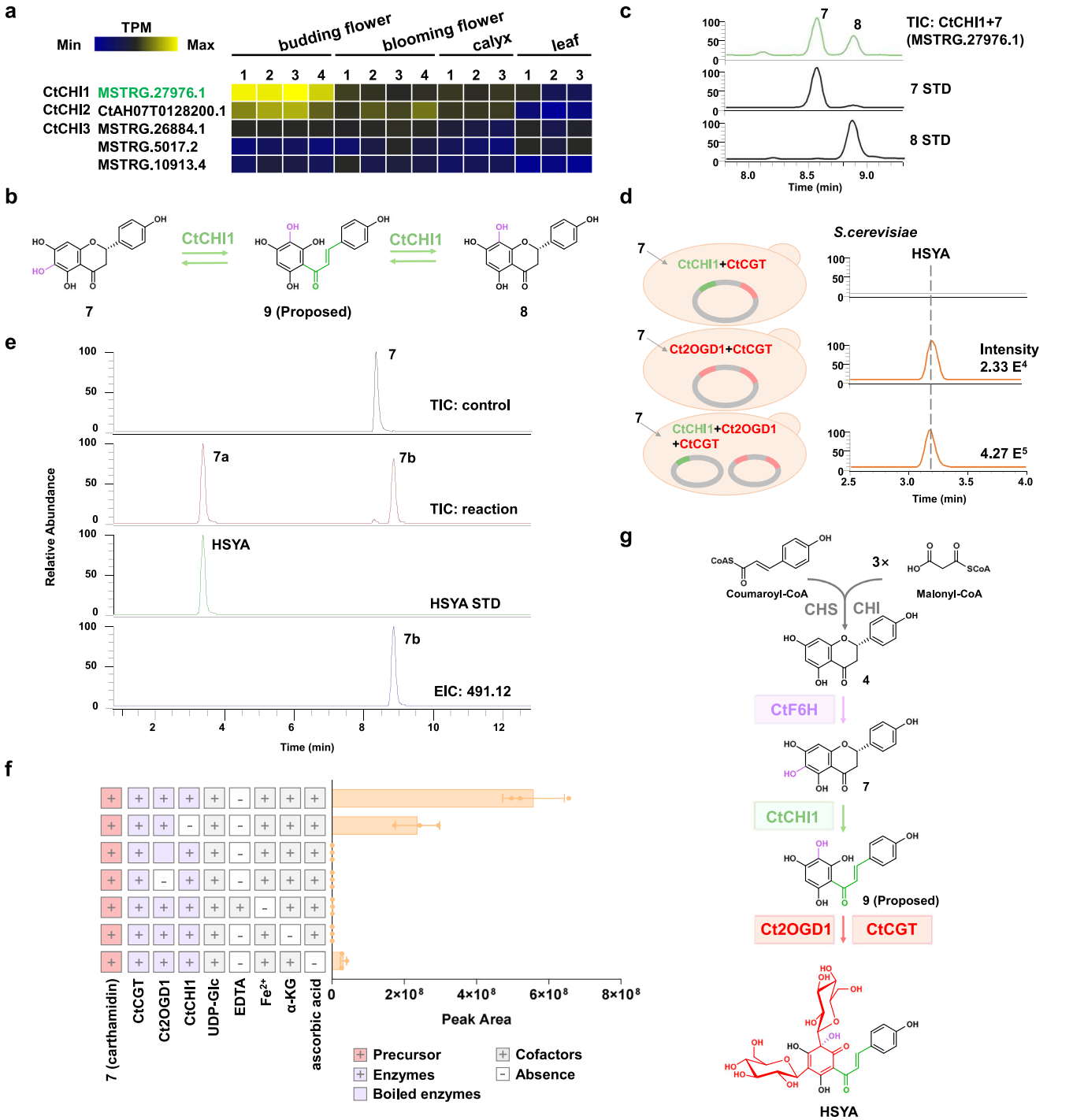

Fig. 4 | Biosynthesis of HSYA in vitro. a Expression levels of five *CtCHI* genes in the transcriptomes of different parts of *Carthamus tinctorius*. **b** The isomerization catalyzed by chalcone-flavanone isomerase. **c** The total ion current (TIC) of the CtCHI1 catalyzed reaction using **7** as substrate. **d** Functional characterization of CtCHI1, CtCGT, and Ct2OGD1 in yeast, and EIC chromatograms (*m/z* 611.1618). **e** The TIC of CtCGT-Ct2OGD1-CtCHI1catalyzed reaction using **7** as substrate. **f** Effects of enzymes and cofactors on the in vitro biosynthesis of HSYA. To verify the role of Fe^{2+}, EDTA was added to the reaction system to chelate residual Fe^{2+} in the buffer. $n = 3$, three biologically independent samples were tested; The data are presented as mean values ± SD. **g** The proposed biosynthetic pathway of HSYA. The source data underlying Fig. 4f is provided in a Source Data file.

though at a low yield. Interestingly, feeding with isocarthamidin also led to the synthesis of HSYA in *N. benthamiana* (Supplementary Fig. 17). As carthamidin and isocarthamidin could convert to each other via a chalcone intermediate, we speculate on the presence of a chalcone-flavanone isomerase (CHI) in the biosynthetic pathway of HSYA. We discovered five potential CHI genes in the transcriptome of safflower (Fig. 4a and Supplementary Table 7). Three genes were successfully cloned and expressed in *E. coli*, including *MSTRG.27976.1* (*CtCHI1*), *CtAH07T0128200.1* (*CtCHI2*), and *MSTRG.26884.1* (*CtCHI3*). Among them, only CtCHI1 could convert carthamidin (**7**) to isocarthamidin (**8**), and carthamidin chalcone might be an intermediate, though it could not be detected (Fig. 4b, c and Supplementary Figs. 24, 25). Notably, *CtCHI1* showed similar expression patterns with *CtF6H*, *CtCGT*, and *Ct2OGD1* (Fig. 4a).

To investigate the role of *CtCHI1* in the biosynthetic pathway of HSYA, we first tested its expression levels in different VIGS groups. The expression levels of *CtCHI1* decreased by 39.2%, 39.8%, and 17.3% in *CtF6H*-, *Ct2OGD1*-, and *CtCGT*-VIGS groups, respectively (Supplementary Fig. 26). Then, we constructed the pESC-Leu-*CtCHI1-CtCGT*, pESC-Leu-*Ct2OGD1-CtCGT*, and pESC-His-*CtCHI1* recombinant plasmids and expressed them individually in the WAT11 yeast (Fig. 4d). When fed with carthamidin, the yield of HSYA was approximately 20-fold higher than that in the non-*CtCHI1* group.

Subsequently, we established an in vitro multi-enzyme catalysis system, which consists of purified CtCGT, Ct2OGD1, and CtCHI1. Carthamidin could be efficiently converted to two products (**7a** and **7b**). Product **7a** was identified as HSYA by comparing with a reference standard (Fig. 4e and Supplementary Figs. 27, 28). Product **7b** was tentatively identified as a mono-*O*-glycoside byproduct according to the diagnostic fragment ion at *m/z* 286 ($[M\text{-}H\text{-}162]^-$ (Supplementary Figs. 27–29). HSYA could also be generated when isocarthamidin was used as substrate (Supplementary Fig. 30). The reaction conditions were then optimized using carthamidin as substrate. The reaction showed high catalytic activity at 30 °C in citrate buffer at pH 6, or in PBS buffer at pH 7 or 8 (Supplementary Fig. 31). When Ct2OGD1 was absent or when boiled Ct2OGD1 was used, no HSYA could be generated. The presence of CtCHI1 improved the yield of HSYA by 2.4 times (Fig. 4f). Generally, Fe^{2+} and α-KG serve as key factors for 2OGD-type enzymes, while ascorbic acid helps prevent the oxidation of Fe^{2+} during the reaction process. Herein, we proved that Fe^{2+} and α-KG were indispensable cofactors for the CtCGT-Ct2OGD1-CtCHI1 catalyzed reaction, and ascorbic acid could improve the yield. The above results further demonstrated the role of CtCHI1 in the biosynthesis of HSYA, as well as the coordinating activities of Ct2OGD1 and CtCGT to catalyze di-*C*-glycosylation and dearomatization reactions for HSYA biosynthesis (Fig. 4g).

Subcellular locations and interactions of biosynthetic enzymes

To investigate the subcellular locations of the above identified four enzymes, we fused the green fluorescent protein (GFP) gene with the biosynthetic genes individually at their C-terminal. As shown in Fig. 5a, CtCGT, Ct2OGD1, and CtCHI1 are all localized in the cytoplasm and cell nucleus. In contrast, CtF6H is located at the endoplasmic reticulum (ER), as it displayed a similar distribution pattern to that of the ER-marker.

To test the interactions among Ct2OGD1, CtCGT, and CtF6H, we conducted split firefly luciferase complementation (SFLC) imaging assays in *Nicotiana benthamiana* leaves[26,27]. As shown in Fig. 5b, the two components of nLUC and cLUC in the three groups were able to come into spatial proximity, thereby restoring the firefly luciferase function and producing fluorescence. Likewise, we further used the bimolecular fluorescence complementation (BiFC) assay to confirm the interactions. In the above three groups, the two proteins could interact with each other, leading to the reconstitution of yellow fluorescence (Fig. 5c and Supplementary Fig. 32). Moreover, we used microscale thermophoresis (MST) to investigate potential in vitro protein-protein interactions. The results demonstrated that CtCGT directly bound to Ct2OGD1 with a dissociation constant (K_d) of 541.4 nM (Fig. 5d).

Based on the above results, we depicted the biosynthetic process of HSYA in plant cells (Fig. 5e). Naringenin acts as the key precursor for HSYA biosynthesis. Firstly, it is hydroxylated to form carthamidin in the ER. Then, carthamidin is transferred to the cytoplasm, where it is converted to carthamidin chalcone. Finally, Ct2OGD1 and CtCGT complete the last step to generate HSYA.

Mechanisms for the unique distribution of HSYA in safflower

While HSYA is uniquely present in safflower, similar key enzymes, including CGT, 2OGD, and F6H, have been reported in many other plants, and CHI is generally present in plant secondary metabolism. We tried to replace *CtCGT*, *CtF6H*, and *Ct2OGD1* with known analogous genes *GuCGTa*[17], *SbF6H*[28], and *AtF3H*, respectively, and infiltrated them into *Nicotiana benthamiana*. Interestingly, all the groups were able to produce HSYA when fed with carthamidin (Fig. 6a, Supplementary Table 5, and Supplementary Fig. 33). To interpret why HSYA is only present in safflower, we analyzed the transcriptome data of 1,341 plant species of the ONEKP database (Supplementary Fig. 34a)[29]. CtCGT, CtF6H, and Ct2OGD1 were used as query sequences for BLASTp search, and the top 250 genes were obtained for each query sequence. A total of 523 species contained at least one candidate gene (Supplementary Fig. 34b). However, only 29 species contained all three key genes (Supplementary Table 8).

On the other hand, we found that when *CtCGT*, *CtF6H*, and *Ct2OGD1* were infiltrated into *N. benthamiana*, and 2-hydroxynaringenin (**2**) was added as a precursor, it could produce common *C*-glycosides but not HSYA (Supplementary Fig. 35). This is a general biosynthetic pathway of flavone *C*-glycosides in higher plants[10,17]. Naringenin could be converted to 2-hydroxynaringenin by flavanone 2-hydroxylases (F2H), which are popularly present in plants[28]. Interestingly, no typical F2H genes were discovered in the transcriptome of *C. tinctorius*. Thus, we speculate that F2H is a competing enzyme which could inhibit the production of HSYA. Once an F2H is present, the flavanone precursor would produce the popular flavonoid *C*-glycosides, instead of HSYA (Fig. 6b).

Among the above screened 29 plant species, ten species contain at least one *F2H*. The other 19 species contain homologs of the aforementioned three key genes, but no F2H gene (Fig. 6c). Then, we constructed homology trees of the 2OGD, CGT, and P450 from the 19 species, respectively. All the 2OGD genes exhibited high similarity with Ct2OGD1 (Fig. 6d and Supplementary Fig. 36). The CGT genes from 4 species (*C. pumila*, *Q. shumardii*, *M. cerifera*, and *H. prostrata*) showed low similarity with *CtCGT* (Supplementary Fig. 37), and the P450 genes from only three species (*F. cronquestii*, *F. sonorensis*, and *S. marianum*) showed high similarity with CtF6H (Fig. 6d and Supplementary Fig. 38). We further analyzed the chemical constituents of the popular medicinal plant *S. mariamum*, and detected neither 6-hydroxylated flavonoids nor HSYA[30]. Moreover, the homologous genes of *CtCGT*, *Ct2OGD1* and *CtF6H* did not show high expression levels in the same tissue or at the same growing period (Supplementary Fig. 39). On the other hand, we obtained a "white flower" natural variant of safflower, which contains all the three key genes, but could not produce HSYA. While the expression levels of *CtCHI1*, *CtCGT*, and *Ct2OGD1* were similar to those of the "red flower", *CtF6H* showed extremely low expression in the "white flower" (Fig. 6e). These results indicate that simultaneous high expression of these key genes is critical for the biosynthesis of HSYA.

Discussion

HSYA is a clinical investigational new drug for the treatment of acute ischemic stroke. The unique quinochalcone di-*C*-glycoside structure of HSYA renders its biosynthetic study very challenging. In this work, we analyzed chemical constituents and transcriptomes of different parts of *C. tinctorius* (Supplementary Fig. 40), and from hundreds of candidates, we discovered four key genes (*CtF6H*, *CtCGT*, *Ct2OGD1*, and *CtCHI1*) responsible for the biosynthesis of HSYA. By testing a number of substrates, we determined naringenin as a critical biosynthetic precursor of HSYA. Among the four key enzymes, CtF6H could convert naringenin to 6-hydroxynaringenin (carthamidin), and CtCHI1 catalyzes carthamidin to generate carthamidin chalcone. Ct2OGD1 could coordinate with CtCGT to catalyze di-*C*-glycosylation and dearomatization reactions of carthamidin chalcone to synthesize HSYA. By infiltrating the above genes, together with upstream genes responsible for the biosynthesis of naringenin, we realized de novo biosynthesis of HSYA in *Nicotiana benthamiana*. The functions of these key genes were further confirmed by VIGS experiments in *C. tinctorius*, semi-synthesis in yeast,

https://doi.org/10.1038/s41467-025-59774-3

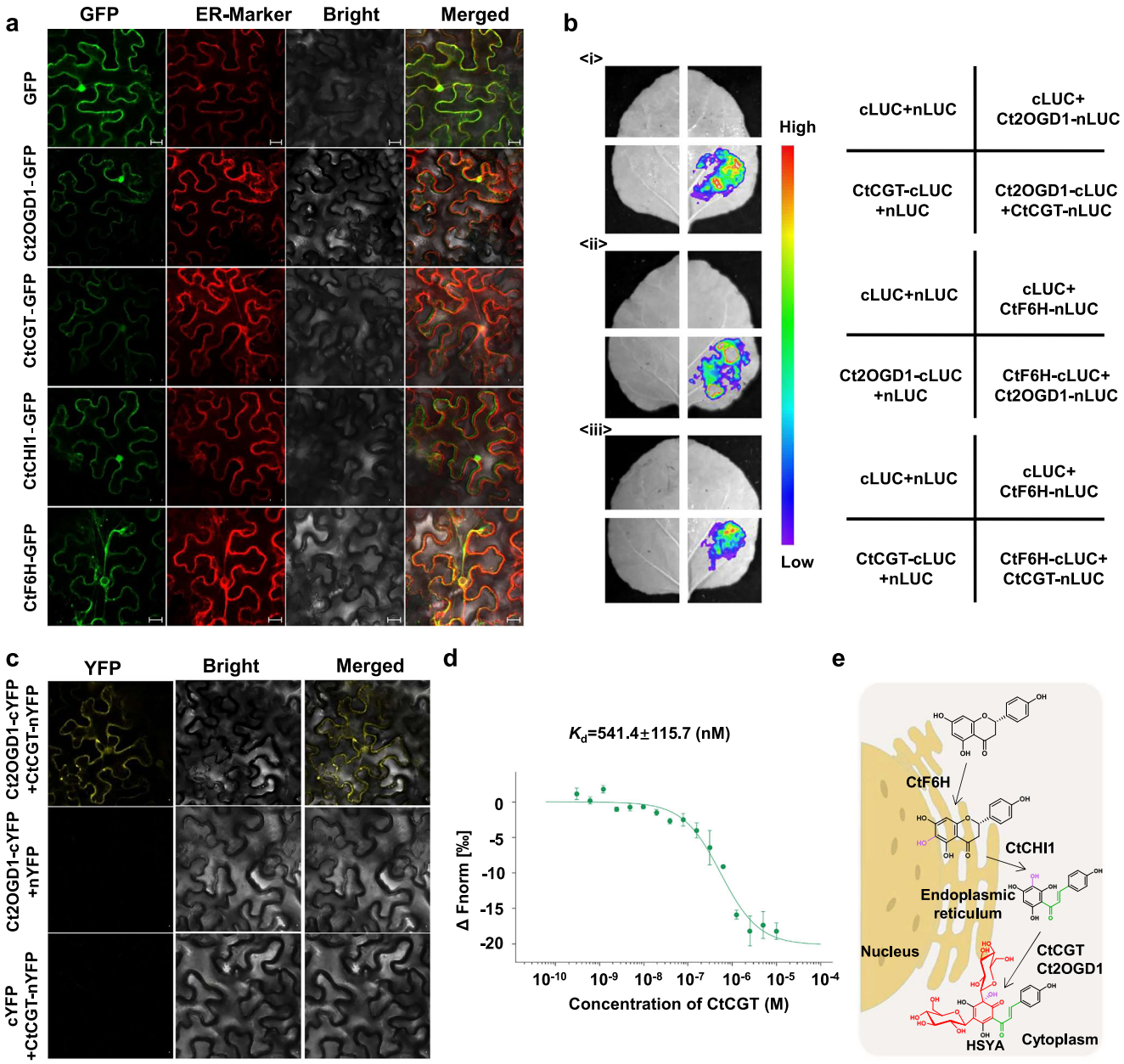

Fig. 5 | Subcellular location and interaction of HSYA biosynthetic enzymes. **a** Subcellular localization of Ct2OGD1, CtCGT, CtCHI1, and CtF6H in *N. benthamiana* leaf epidermal cells. GFP fusions with Ct2OGD1, CtCGT, CtCHI1, and CtF6H under control of the CaMV 35S promoter were transiently expressed in *N. benthamiana* leaf epidermal cells. HDEL: DsRed (red) is an endoplasmic reticulum (ER) marker. Scale bars are 20 μm. $n = 3$, three biologically independent samples were tested. **b** Split firefly luciferase complementation assay in *N. benthamiana* leaves. <**i**> The interaction between Ct2OGD1 and CtCGT. <**ii**> The interaction between CtF6H and Ct2OGD1. <**iii**> The interaction between CtF6H and CtCGT. $n = 3$, three biologically independent samples were tested. **c** Ct2OGD1 physically interacts with CtCGT using BiFC assays. Ct2OGD1 was cloned in-frame with the C-YFP vector with the respective deletion, and CtCGT was cloned in-frame with the N-YFP vector. Scale bars are 20 μm. $n = 3$, three biologically independent samples were tested. **d** The interaction between Ct2OGD1 and CtCGT using MST. Signal to noise ratio, 17.5. $n = 3$, three biologically independent samples were tested; The data were presented as mean values ± SEM. **e** Illustration of the proposed biosynthetic process of HSYA in plant cells. The source data underlying Fig. 5d is provided in a Source Data file.

and enzyme catalysis. To our knowledge, this work represents the first report to elucidate the biosynthetic pathway of HSYA.

It is interesting that Ct2OGD1 coordinates with CtCGT to complete the final step of HSYA biosynthesis. This is not the only case where two enzymes function together to catalyze one reaction. For instance, the combination of CsCYP88A51 and CsMOI2 catalyzes melianol to generate *apo*-melianol[31], and CAL1 coordinates with CAL2 to participate in the biosynthesis of huperzine A[32]. In this work, we reveal that Ct2OGD1 and CtCGT are both located in the cytoplasm and cell nucleus, and their interactions were confirmed by SFLC, BiFC, and MST experiments. There could be two possible functional modes for these two enzymes. First, the two proteins may form a complex, which directly catalyzes the final reaction step. We simulated the Ct2OGD1-CtCGT complex structure (Supplementary Fig. 41). The β-strand at the C-terminal of Ct2OGD1 could insert into the structure of CtCGT, and form a large parallel β-sheet consisting of six β-strands, thereby stabilizing the protein. Alternatively, the two enzymes may function separately, with the complex stabilizing the proteins. In the future, more studies are needed to elucidate the detailed mechanisms of how CtCGT and Ct2OGD1 coordinate to catalyze the dearomatization

https://doi.org/10.1038/s41467-025-59774-3

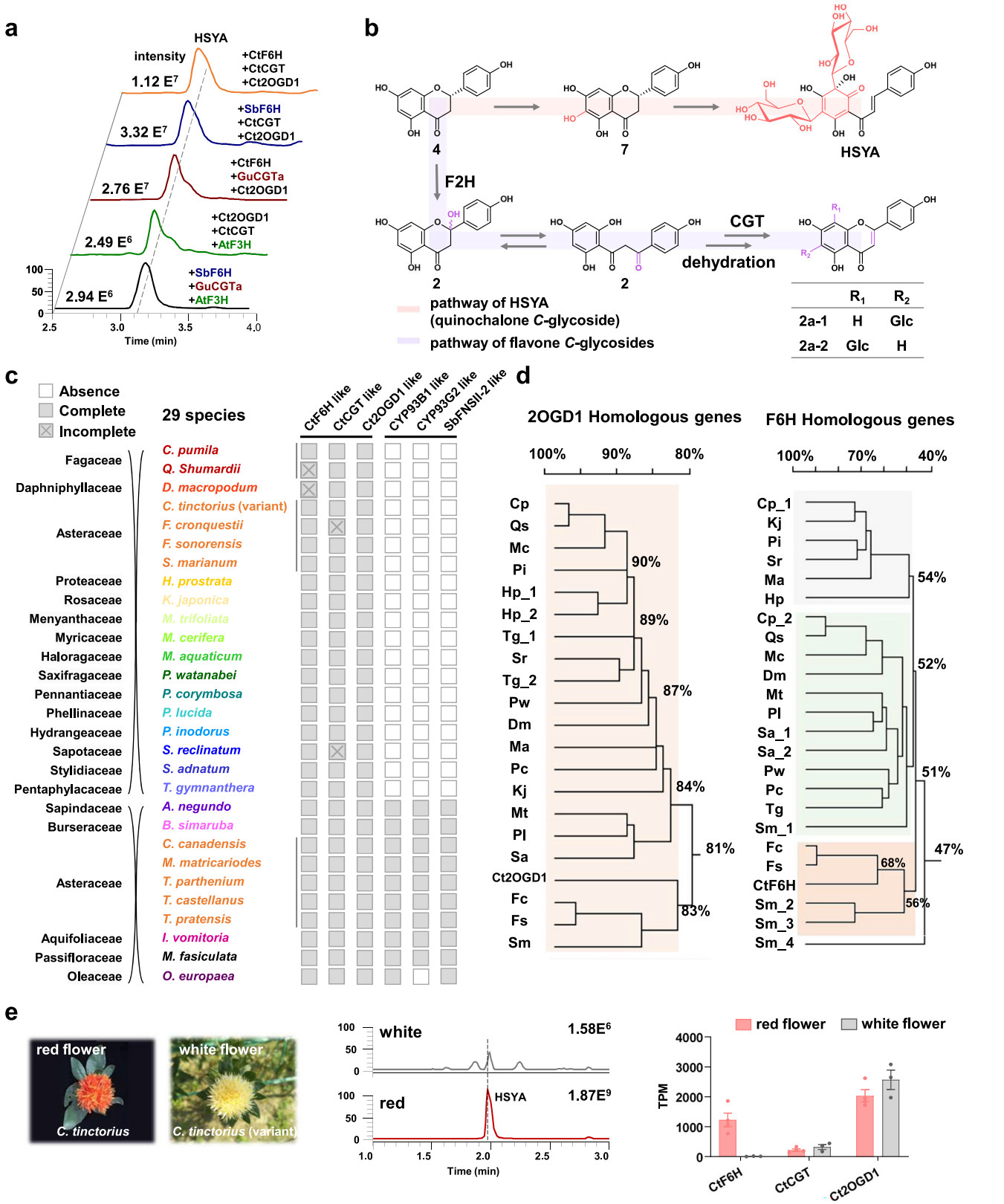

Fig. 6 | The distribution of HSYA-related genes in plants. a LC/MS analysis of *Nicotiana benthamiana* leaves infiltrated with different combinations of genes, showing extracted ion chromatograms (EIC: *m/z* 611.1618) of HSYA. The peak intensity is shown on the EIC plot. **b** The competing biosynthesis pathways of HSYA (quinochalcone *C*-glycoside) and flavone *C*-glycosides. **c**, **d** Homology analysis of Ct2OGD1 and CtF6H homologous genes from 19 plant species of the ONEKP database (https://db.cngb.org/onekp/). Some genes were not complete in four species (*Q. Shumardii*, *D. macropodum*, *F. cronquestii*, and *S. reclinatum*). **e** LC/MS analysis and expression level analysis of *C. tinctorius* (red flower) and its variant (white flower). The peak with a retention time similar to that of HSYA in the white flowers was not identified as HSYA, according to mass spectral analysis. $n = 3$, three biologically independent samples were tested; The data were presented as mean values ± SEM. The source data underlying Fig. 6e are provided in a Source Data file.

reaction, and how the rare *C*-glycosylation of an sp^3 carbon atom takes place.

This work also interprets why HSYA is uniquely present in *C. tinctorius*. Analysis of the transcriptomes of 1341 plant species indicates the simultaneous presence and high expression of the these key enzymes is essentially required for the biosynthesis of HSYA. Moreover, F2H enzymes should be absent to avoid the production of common flavonoid *C*-glycosides.

In conclusion, this work elucidates the unique biosynthetic pathway of HSYA, and interprets why HSYA is only present in safflower. Our results serve as a critical platform to further unravel detailed mechanisms of this pathway, and demonstrate the potential of synthetic biology to prepare HSYA as a promising investigational new drug.

Methods

Plant materials and chemicals

The plants of *Carthamus tinctorius* L. were grown in a greenhouse at 22 °C, 14-h light/10-h dark. The budding flower, blooming flower, calyx (around 3 months), and leaf (around 3 months) were collected for total RNA extraction and transcriptome sequencing. Chemicals used in this study are shown in Supplementary Table 9.

Total RNA isolation and transcriptome sequencing

The total RNA was extracted using the RNAprep Pure Plant Plus Kit for polysaccharides & polyphenolics-rich (Tiangen Biotech, China) following the manufacturer's instructions. cDNA was synthesized using TransScript one-step genomic DNA (gDNA) removal and cDNA synthesis SuperMix (Transgen Biotech, China). The transcriptome data of different parts of *C. tinctorius* were acquired using the Illumina sequencing platform by Majorbio Bioinformatics Technology Co., Ltd (Shanghai, China). Transcription level analyses were performed using the online platform of Majorbio Cloud Platform (www.majorbio.com).

Bioinformatics

Co-expression analysis was conducted using R Studio. For UGT and P450 analysis, a library with 306 UGT transcripts and 616 P450 transcripts was constructed. The total expression levels of *CtCHS1* and *CtCHS2* were used as the "bait", respectively. For the final gene screening, all high-expression genes (TPM ≥10) in budding flowers were extracted. A library consisting of 9307 transcripts was constructed. *CtCGT* and *CtF6H* were used as the "bait". The co-expressed genes were further filtered according to Pfam (https://www.ebi.ac.uk/interpro), NR (non-redundant protein sequences, https://www.ncbi.nlm.nih.gov/), and UniProtKB/Swiss-Prot (https://www.uniprot.org/) database annotation and Pearson's correlation coefficient ($r \geq 0.8$). The co-expression network was visualized by Cytoscape. Homologous gene screening of *CtCGT*, *CtF6H*, and *Ct2OGD1* was conducted using ONEKP. BLASTp was used as a BLAST search with default parameters. The top 250 homologous genes were obtained for further analyses.

Phylogenetic and homology tree analysis

Molecular phylogenetic analysis was conducted using MEGA6 software with the maximum likelihood method (Supplementary Figs. 1, 2, 16 and Supplementary Data. 4–6). The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the taxa analyzed. Homology tree was conducted by DNAMAN (Supplementary Figs. 36–38).

Molecular cloning and expression

The full length of ORF was amplified from cDNA by using 2×Hieff Canace® Gold PCR Master Mix (Yeasen, China) (Supplementary Data 1). *CtCGT*, *Ct2OGD1*, and *CtCHI1-3* were cloned into pET-28a(+) vector at the BamHI site using MultiF Seamless Assembly Mix (ABclonal, China). *CtF6H* was cloned into the pESC-leu vector at the BamHI site using seamless assembly. For the yeast feeding experiment, *CtCHI1* and *CtCGT* were inserted into the SpeI and BamHI sites of the same pESC-Leu vector, respectively. *Ct2OGD1* and *CtCGT* were inserted into the SpeI and BamHI sites of the pESC-Leu vector, respectively. *CtCHI1* was inserted into the BamHI sites of the pESC-His vector. The recombinant plasmids were transformed into *E. coli* BL21(DE3) and WAT11 yeast (an engineered *Saccharomyces cerevisiae* strain constructed by replacing the endogenous CPR gene in yeast with the CPR1 gene from *Arabidopsis thaliana*), respectively.

For the *E. coli* expression system, single colonies were incubated in LB media (50 μg/mL Kanamycin) at 37 °C in a shaking incubator at 200 rpm. When the OD_{600} value reached 0.6–0.8, protein expression was induced with 0.1 mM IPTG for 20 h at 18 °C. The total proteins were extracted by sonication on ice. The recombinant proteins were purified using a nickel-affinity column. After SDS-PAGE analysis, the purified protein was concentrated and desalted by a 30-kDa ultrafiltration tube (Merck Millipore) with storage buffer (20 mM Tris, 200 mM NaCl, 20% (*v*/*v*) glycerol, pH 7.4).

For the yeast expression system, single colonies were incubated in SD-Leu medium at 28 °C in a shaking incubator at 200 rpm. The cultures were centrifuged (1500×*g*, 5 min, 4 °C), then the pellet was collected, washed with sterile water, and resuspended (OD_{600} = 1.0) in fresh SG-Leu medium for gene expression. After 24 h, the cells were harvested, rinsed in TEK buffer (50 mM Tris, pH 7.4, 1 mM EDTA, 100 mM KCl), and resuspended in TESB buffer (50 mM Tris pH 7.4, 1 mM EDTA, 0.6 M sorbitol). The samples were homogenized with acid-washed glass beads (425–600 mm, Sigma-Aldrich) eight times in a vortex mixer (each cycle: 30 s vortexing, 30 s cooling on ice). After centrifugation at 21,130×*g* for 10 min, the supernatant was collected. For every 500 μL supernatant should be added 1 mL TESB precipitation solution (TESB, 0.225 M NaCl, 0.15 g/mL PEG-4000). After 15 min, the microsome was obtained by centrifugation at 21,130×*g* for 10 min. Then the microsome was resuspended in TEG (50 mM Tris, pH 7.4, 1 mM EDTA, 20% glycerol). For the yeast feeding experiment, the yeast was initially cultured in SD-Leu or SD-Leu/His medium, followed by collection of the pellet, washing with sterile water, and resuspension (OD_{600} = 1.0) in fresh SG-Leu or SG-Leu/His medium. Then, 0.05 mmol/L carthamidin was added to the yeast 10 h after protein expression. One day later, an aliquot of the culture was extracted with an equal volume of methanol for LC/MS analysis.

Enzyme activity assay

For CGT enzyme assay, the reactions were carried out in 100-μL systems containing 50 mM PBS (pH 7.4), 0.1 mM substrate, 0.5 mM UDP-Glc, and 10 μg of purified CtCGT at 37 °C for 1 h. For the P450 enzyme assay, the reactions were carried out in 100-μL systems containing 50 mM PBS (pH 7.4), 0.1 mM substrate, 0.5 mM NADPH, and 100 μg of microsomes containing CtF6H at 18 °C for 8 h. For the 2OGD enzyme assay, the reactions were carried out in 100-μL systems containing 50 mM PBS (pH 7.4), 0.1 mM substrate, 1 mM Fe^{2+}, 1 mM ascorbic acid, 1 mM α-ketoglutaric acid (α-KG), and 10 μg of purified Ct2OGD1 at 37 °C for 1 h. For the CHI enzyme assay, the reactions were carried out in 100-μL systems containing 50 mM citrate buffer (pH 6), 0.1 mM substrate, 0.5 mM DTT, and 20 μg of CtCHI at 30 °C for 15 min. The reactions were terminated with 200 μL ice-cold methanol and centrifuged at 21,130×*g* for 20 min. For multi-enzyme catalysis assay, the reactions were carried out in 50-μL systems containing 50 mM citrate buffer (pH 6.0), 0.1 mM substrate, 0.5 mM UDP-Glc, 5 mM DTT, 0.6 mM Fe^{2+}, 0.6 mM α-KG, 0.6 mM ascorbic acid, and 40 μg of purified CtCGT, Ct2OGD1 and CtCHI1 at 30 °C for 3 h. The reactions were terminated with 100 μL ice-cold methanol and centrifuged at 21,130×*g* for 20 min.

LC/MS analysis

For HPLC analysis, an Agilent 1260 instrument (Agilent Technologies, Waldbronn, Germany) was used. Samples were separated on an

Agilent ZORBAX SB-C18 column (4.6 × 250 mm, 5 µm). The mobile phase consisted of methanol (A) and water containing 0.1% formic acid (*v*/*v*, B). The analytes were eluted using a linear gradient program: 0 min, 20% A; 20 min, 100% A; 25 min, 100% A; 26 min, 20% A; 31 min, 20% A. The flow rate was 1 mL/min. The column temperature was 30 °C. For UHPLC/MS analysis, a Thermo UHPLC instrument was coupled with a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer through a heated ESI source (Thermo Fisher Scientific, USA). Samples were separated on an Acquity UPLC HSS T3 column (2.1 × 100 mm, 1.8 µm). Methanol (A) and water containing formic acid (0.1%, *v*/*v*) (B) were used as the mobile phase. The gradient elution program was as follows: 0 min, 20% A; 4 min, 35% A; 7 min, 40% A; 11.5 min, 100% A; 13 min, 100% A. The flow rate was 0.3 mL min^{-1}. The column temperature was 50 °C. 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. MS^1 resolution was set as 70,000 FWHM, AGC target 1*E6, maximum injection time 50 ms, and scan range *m*/*z* 100–1500. MS^2 resolution was set as 17,500 FWHM, AGC target 1*E5, maximum injection time 100 ms, NCE 35. The mass spectra were recorded in the negative ion mode. The mass data were analyzed by Xcalibur 4.1 software (Thermo Fisher).

VIGS

DNA fragments of 300 to 500 base pairs of the target genes were amplified by PCR using gene-specific primers and then cloned into the pTRV2 vector using BamHI. After transfer to *Agrobacterium tumefaciens* GV3101, the cells were grown in LB culture medium (50 µg/mL Kanamycin, 50 µg/mL Rifampicin, and 20 µg/mL Gentamycin) overnight at 28 °C and 200 rpm, collected, and resuspended in MMA buffer (10 mM $MgCl_2$, 10 mM MES, and 150 mM acetosyringone, pH 5.8) and adjusted to optical density (OD_{600}) of around 1.0. The strains containing TRV1 (empty) or TRV2 (harboring the target gene fragment) vectors were mixed at a ratio of 1:1, and infiltrated into around 2–3-week seedlings via a syringe. Two months later, the budding flowers were collected and lyophilized. The dried samples were pulverized into powder, and an aliquot of 5 mg was mixed with 0.5 mL of 50% (*v*/*v*) methanol. The mixture was ultrasonicated for 15 min and then centrifuged at 21,130×*g* for 20 min. The samples were then diluted 40- to 80-fold by 50% (*v*/*v*) methanol. A 2-µL aliquot of the sample was injected for UHPLC/MS analysis. The PRM mode was used for quantitative analysis. Reference standard was accurately weighed and dissolved in methanol to prepare stock solution (2 mg/mL). The stock solution was diluted with methanol to obtain a series of working solutions (1, 5, 10, 50, 100, 500, 1000, 5000, 10,000 ng/mL). A 2-µL aliquot was injected for analysis.

Gene functional characterization in *Nicotiana benthamiana*

The genes for *Nicotiana benthamiana* expression were cloned into the modified pEAQ-HT-DEST1 vector at the XhoI site using seamless assembly. The recombinant plasmids were transformed into *Agrobacterium tumefaciens* strain GV2260. Single colonies of the transformed *Agrobacterium tumefaciens* strain GV2260 mentioned above were inoculated at 28 °C with shaking in LB culture medium (50 µg/mL Kanamycin, 50 µg/mL Rifampicin) until OD_{600} = 0.6. After centrifugation, bacteria were resuspended in MMA buffer. Different strains were mixed at a final OD_{600} of 0.2 for each strain before transformation. The infection solution was infiltrated into the leaves of 5–6 week-old *Nicotiana benthamiana*. For de novo biosynthesis, the samples were harvested and freeze-dried after infiltration for 7 days. For substrate-feeding experiments, *Nicotiana benthamiana* leaves were infiltrated with 0.2 mM substrates (dissolved in MMA buffer) on the 3rd day after strain infiltration. After incubation for 4 days, the leaves were harvested, flash frozen, and extracted for UHPLC/MS analysis.

Subcellular localization and confocal microscopy analysis

The full-length CDS (coding sequence) of CtF6H, CtCGT, Ct2OGD1, and CtCHI1 were fused with green fluorescent protein (GFP) in the vector pCAMBIA1300 (Supplementary Data 1). *Agrobacterium* strain GV3101 containing recombinant plasmids, GFP positive control, and HDEL (ER-marker) were infiltrated into *N. benthamiana* leaves and then cultured at 25 °C for 48 h. The GFP fluorescence was detected at excitation wavelengths of 488 and 543 nm, respectively, using a confocal laser scanning microscope (Carl Zeiss, LSM980).

Split firefly luciferase complementation (SFLC) assay

The full-length genes were amplified, and then inserted into the BamHI and SaLI sites of pCAMBIA1300-cLUC or pCAMBIA1300-nLUC vectors to generate fusion constructs (Supplementary Data 1). Different combinations of the constructs were co-transfected into *N. benthamiana* leaf epidermal cells by *Agrobacterium*-mediated infiltration (GV3101). After 48 h of incubation in darkness, the injected leaves were sprayed with 1 mM luciferin (Promega, E1605), and then the LUC signal was captured by an CCD imaging apparatus (Berthold, LB985).

Bimolecular fluorescence complementation (BiFC) assay

The full-length CDSs were inserted into the BamHI and XhoI sites of *YFPN* and *YFPC*, respectively (Supplementary Data 1). The resulting constructs were co-transfected into tobacco leaf epidermal cells by *Agrobacterium*-mediated infiltration. The transformed *N. benthamiana* leaves were incubated at 25 °C for 48–72 h. Yellow fluorescence signals were visualized using a confocal microscope (Zeiss LSM980).

Microscale thermophoresis analysis

The binding affinity of CtCGT with Ct2OGD1 was measured by microscale thermophoresis (MST). The proteins of CtCGT and Ct2OGD1 were purified by Ni-NTA affinity chromatography and size exclusion chromatography, and were centrifuged and concentrated to 20 mg/mL. Then they were diluted with 10 mM PBS (containing 0.05% Tween-20, UDPG 2 mM, Fe^{2+} 2 mM, α-KG 2 mM, ascorbic acid 2 mM). Ct2OGD1 was labeled with the *N*-hydroxysuccinimide (NHS) (NanoTemper Technologies). Serially diluted CtCGT using the above buffer solution, with concentrations of 20 µM to 0.61 nM, were mixed with labeled Ct2OGD1 at room temperature and then loaded into Monolith standard-treated capillaries. Binding was measured by monitoring the samples at an excitation power of 2% on a Monolith NT.115 instrument (NanoTemper Technologies). The K_d value was determined using the MO. Affinity Analysis software (NanoTemper Technologies).

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 a public database. The gene sequence data generated in this study have been deposited in the NCBI database under the following accession numbers: *CtF6H* [PQ040214, https://www.ncbi.nlm.nih.gov/nuccore/PQ040214], *CtCGT* [PQ040212, https://www.ncbi.nlm.nih.gov/nuccore/PQ040212], *Ct2OGD1* [PQ040215, https://www.ncbi.nlm.nih.gov/nuccore/PQ040215], *CtCHI1* [PV014870, https://www.ncbi.nlm.nih.gov/nuccore/PV014870]. The raw reads from the RNA-sequencing profiling analysis of *Carthamus tinctorius* have been deposited in the NCBI Sequence Read Archive (SRA) database under the

BioProject accession PRJNA1135894. The primers, candidate genes information, and sequences used for phylogenetic tree construction are given in Supplementary Data 1–6. Source data are provided with this paper.

References

1. Zhou, X., Tang, L., Xu, Y., Zhou, G. & Wang, Z. Towards a better understanding of medicinal uses of *Carthamus tinctorius* L. in traditional Chinese medicine: a phytochemical and pharmacological review. *J. Ethnopharmacol.* **151**, 27–43 (2014).
2. Bai, X. et al. Therapeutic potential of hydroxysafflor yellow A on cardio-cerebrovascular diseases. *Front. Pharmacol.* **11**, 01265 (2020).
3. Wang, W. et al. Hydroxysafflor yellow A ameliorates alcohol-induced liver injury through PI3K/Akt and STAT3/NF-κB signaling pathways. *Phytomedicine* **132**, 155814 (2024).
4. Zhang, X. et al. Pharmacological actions, molecular mechanisms, pharmacokinetic progressions, and clinical applications of hydroxysafflor yellow A in antidiabetic research. *J. Immunol. Res.* **2021**, 560012 (2021).
5. Feng, Z.-M. et al. NMR solution structure study of the representative component hydroxysafflor yellow A and other quinochalcone *C*-glycosides from *Carthamus tinctorius*. *J. Nat. Prod.* **76**, 270–274 (2013).
6. Wu, Z. H. et al. Current advances of *Carthamus tinctorius* L.: a review of its application and molecular regulation of flavonoid biosynthesis. *Med. Plant Biol.* **3**, e004 (2024).
7. Li, F., He, Z. & Ye, Y. Isocartormin, a novel quinochalcone *C*-glycoside from *Carthamus tinctorius*. *Acta Pharm. Sin. B* **7**, 527–531 (2017).
8. Hayashi, T., Ohmori, K. & Suzuki, K. Synthetic study on carthamin. 2. stereoselective approach to *C*-glycosyl quinochalcone via desymmetrization. *Org. Lett.* **19**, 866–869 (2017).
9. Suzuki, T., Ishida, M., Kumazawa, T. & Sato, S. Oxidation of 3,5-di-*C*-(per-*O*-acetylglucopyranosyl)phloroacetophenone in the synthesis of hydroxysafflor yellow A. *Carbohyd. Res.* **448**, 52–56 (2017).
10. Zhang, Y. Q., Zhang, M., Wang, Z. L., Qiao, X. & Ye, M. Advances in plant-derived *C*-glycosides: phytochemistry, bioactivities, and biotechnological production. *Biotechnol. Adv.* **60**, 18 (2022).
11. Guo, D. et al. Overexpression of CtCHS1 increases accumulation of quinochalcone in safflower. *Front. Plant Sci.* **8**, 1409 (2017).
12. Wu, Z. et al. The chromosome-scale reference genome of safflower (*Carthamus tinctorius*) provides insights into linoleic acid and flavonoid biosynthesis. *Plant Biotechnol. J.* **19**, 1725–1742 (2021).
13. Chen, J. et al. Whole-genome and genome-wide association studies improve key agricultural traits of safflower for industrial and medicinal use. *Hortic. Res.* **10**, 12 (2023).
14. Shinozaki, J. et al. Cloning and functional analysis of three chalcone synthases from the flowers of safflowers *Carthamus tinctorius*. *Nat. Prod. Commun.* **11**, 787–790 (2016).
15. Chen, C. J. et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).
16. Hansen, C. C., Nelson, D. R., Moller, B. L. & Werck-Reichhart, D. Plant cytochrome P450 plasticity and evolution. *Mol. Plant* **14**, 1244–1265 (2021).
17. Wang, Z. L. et al. Dissection of the general two-step di-*C*-glycosylation pathway for the biosynthesis of (iso)schaftosides in higher plants. *Proc. Natl Acad. Sci. USA* **117**, 30816–30823 (2020).
18. 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).
19. Huang, J. Q. et al. Aromatization of natural products by a specialized detoxification enzyme. *Nat. Chem. Biol.* **16**, 250–256 (2020).
20. Hong, B. K. et al. Biosynthesis of strychnine. *Nature* **607**, 617–622 (2022).
21. Reed, J. et al. Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. *Science* **379**, 1252–1264 (2023).
22. Nett, R. S., Lau, W. & Sattely, E. S. Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* **584**, 148–153 (2020).
23. Wang, H. T. et al. Insights into the missing apiosylation step in flavonoid apiosides biosynthesis of Leguminosae plants. *Nat. Commun.* **14**, 6658 (2023).
24. Lech, K., Nawala, J. & Popiel, S. Mass spectrometry for investigation of natural dyes in historical textiles: unveiling the mystery behind safflower-dyed fibers. *J. Am. Soc. Mass Spectrom.* **32**, 2552–2566 (2021).
25. Sui, S., Xie, K., Guo, R., Dai, J. & Yang, L. Molecular characterization of a stereoselective and promiscuous flavanone 3-hydroxylase from *Carthamus tinctorius* L. *J. Agric. Food Chem.* **71**, 1679–1689 (2023).
26. He, Q. et al. *Arabidopsis* TIE1 and TIE2 transcriptional repressors dampen cytokinin response during root development. *Sci. Adv.* **8**, 18 (2022).
27. Liu, C. et al. The protein phosphatase PC1 dephosphorylates and deactivates CatC to negatively regulate H_2O_2 homeostasis and salt tolerance in rice. *Plant Cell* **35**, 3604–3625 (2023).
28. Zhao, Q. et al. A specialized flavone biosynthetic pathway has evolved in the medicinal plant, *Scutellaria baicalensis*. *Sci. Adv.* **2**, e1501780 (2016).
29. Leebens-Mack, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
30. Wang, X., Zhang, Z. & Wu, S. C. Health benefits of *Silybum marianum*: phytochemistry, pharmacology, and applications. *J Agric. Food Chem.* **68**, 11644–11664 (2020).
31. Pena, R. D. et al. Complex scaffold remodeling in plant triterpene biosynthesis. *Science* **379**, 361–368 (2023).
32. Nett, R. S. et al. Plant carbonic anhydrase-like enzymes in neuroactive alkaloid biosynthesis. *Nature* **624**, 182–191 (2023).

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grants No. 82273806 to M.Y., 81725023 to M.Y., 82274039 to J.C., 82304326 to Z.-L.W., and 823B2093 to H.-T.W.), the National Key Research and Development Program of China (No. 2023YFA0914100 to M.Y.), Yunnan Provincial Science and Technology Project at Southwest United Graduate School (No. 202302AP370006 to M.Y.), and China National Postdoctoral Program for Innovation Talents (Grant No. BX20220022 to Z.-L.W.). The authors thank Dr. Jing Wang at State Key Laboratory of Natural and Biomimetic Drugs of Peking University for technical help in the MST experiments. The authors wish to thank Dr. David Nelson (The University of Tennessee Health Science Center) of the P450 nomenclature committee for the naming of CtF6H. The authors thank Prof. Xiao-Ya Chen and Prof. Jin-Quan Huang at CAS Center for Excellence in Molecular Plant Sciences, Prof. Zhen-Hua Liu at Shanghai Jiao Tong University, Prof. Zhi-Hua Liao and Prof. Fang-Yuan Zhang at Southwest University for providing VIGS system and modified pEAQ-HT vector.

Author contributions

M.Y. and Z.-L.W. designed the research. M.Y., Z.-L.W., H.-T.W., and J.C. acquired funding. Z.-L.W. and H.-T.W. performed experiments and analyzed the data. G.C., G.Y., and M.Z. assisted with experiments. J.C. provided the seeds of safflower. Z.-L.W., H.-T.W., and M.Y. wrote the manuscript. All authors have approved the final version of the manuscript.

Competing interests

M.Y., Z.-L.W., and H.-T.W. are inventors listed on a patent application related to this work (CN 202510459063.1) that was submitted by Peking University. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59774-3.

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

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.