



Promiscuous CYP87A enzyme activity initiates cardenolide biosynthesis in plants

Received: 10 January 2023

Accepted: 16 August 2023

Published online: 18 September 2023

 Check for updates

Maritta Kunert, Chloe Langley, Rosalind Lucier , Kerstin Ploss, Carlos E. Rodríguez López , Delia A. Serna Guerrero , Eva Rothe, Sarah E. O'Connor   & Prashant D. Sonawane  

Cardenolides are specialized, steroidal metabolites produced in a wide array of plant families^{1,2}. Cardenolides play protective roles in plants, but these molecules, including digoxin from foxglove (*Digitalis* spp.), are better known for treatment of congenital heart failure, atrial arrhythmia, various cancers and other chronic diseases^{3–9}. However, it is still unknown how plants synthesize ‘high-value’, complex cardenolide structures from, presumably, a sterol precursor. Here we identify two cytochrome P450, family 87, subfamily A (CYP87A) enzymes that act on both cholesterol and phytosterols (campesterol and β -sitosterol) to form pregnenolone, the first committed step in cardenolide biosynthesis in the two phylogenetically distant plants *Digitalis purpurea* and *Calotropis procera*. *Arabidopsis* plants overexpressing these CYP87A enzymes ectopically accumulated pregnenolone, whereas silencing of *CYP87A* in *D. purpurea* leaves by RNA interference resulted in substantial reduction of pregnenolone and cardenolides. Our work uncovers the key entry point to the cardenolide pathway, and expands the toolbox for sustainable production of high-value plant steroids via synthetic biology.

Members of the Apocynaceae (for example, *Calotropis procera*, *Nerium oleander*, *Strophanthus gratus* and *Strophanthus kombe*), Plantaginaceae (for example, *Digitalis lanata* and *Digitalis purpurea*) and Brassicaceae (for example, *Erysimum cheiranthoides*) families produce substantial levels of cardenolides, also known as cardiac glycosides^{1,10–13}. The renowned cardenolide example is digoxin, a cardiotonic drug recommended as an essential medicine by the World Health Organization. The use of cardenolide-containing *Digitalis* (foxglove) extracts for the treatment of congestive heart diseases was first reported in 1785¹⁴. However, only a partial biosynthetic pathway of cardenolides has been proposed, which was based on feeding studies of radioisotope-labelled precursors in *D. lanata* and *D. purpurea* performed in the 1960s^{15–18}. These findings demonstrated that cardenolide biosynthesis proceeds through pregnane derivatives (pregnenolone or progesterone) that are probably derived from either cholesterol or phytosterols (for example, β -sitosterol). In contrast, a

preferred route for cardenolide biosynthesis involving phytosterols (24-alkyl sterols), not cholesterol, was suggested because of feeding studies in *D. lanata* using inhibitors of phytosterol biosynthesis¹⁹. Therefore, the specific sterol precursor(s) involvement in cardenolide biosynthesis remains unconfirmed. Despite decades of extensive research on biosynthesis of sterol-derived cardenolides in *Digitalis* spp.^{20,21}, only two biosynthetic enzymes, 3 β -hydroxysteroid dehydrogenase (3 β HSD) and progesterone 5 β -reductase (P5 β R), have been identified and partially characterized from *Digitalis*^{22,23} and *Erysimum*^{24,25} species. The biosynthetic steps starting from the initial sterol precursor to pregnenolone, and those steps downstream of pregnenolone towards cardiac aglycones (for example, digitoxigenin, calotropagenin) and their corresponding glycosides (for example, digitoxin), remain unidentified in any cardenolide-producing plant species (Fig. 1a; detailed pathway shown for *D. purpurea* in Supplementary Fig. 1).

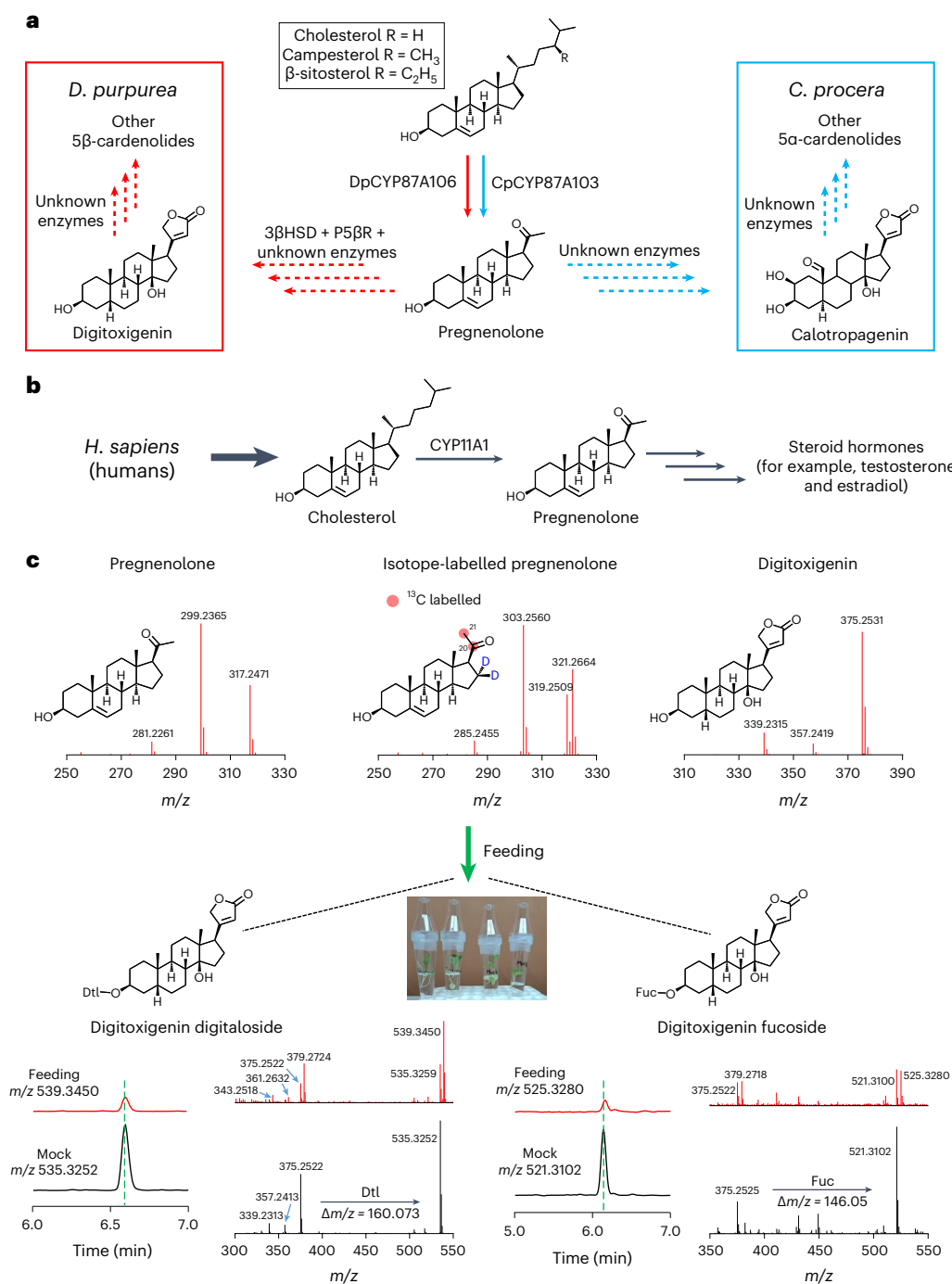


Fig. 1 | Summary of cardenolide biosynthesis and transformation of pregnenolone to cardenolides by stable isotope-labelled feeding studies.

a, The proposed biosynthetic pathway for cardenolides, starting from an unconfirmed sterol precursor in *D. purpurea* (red) and *C. procera* (blue). A simplified pathway scheme is presented. Dashed arrows indicate uncharacterized steps in the pathway. DpCYP87A106 and CpCYP87A103 enzymatic steps shown in this study are represented by solid arrows. See Supplementary Fig. 1 for more details of the proposed biosynthetic pathway. **b**, CYP11A1 catalyses the conversion of cholesterol to pregnenolone, a precursor of all other steroid hormones in humans. **c**, LC-MS analysis of cardenolides after feeding of isotope-labelled pregnenolone to 4-week-old *D. purpurea* grown

in vitro. Mass fragmentation spectra $[M + H]^+$ for authentic pregnenolone, isotope-labelled pregnenolone and digitoxigenin standards are shown at the top. Comparison of extracted ion chromatograms and corresponding tandem mass spectrometry fragmentation spectra of two cardenolides, digitoxigenin digitaloside (left) and digitoxigenin fucoside (right) in feeding (treated with isotope-labelled pregnenolone; red) and mock (treated with de-ionized water; black) samples, respectively. Exogenously fed isotope-labelled pregnenolone was successfully incorporated into cardenolides in *D. purpurea*. m/z ions are indicated for each cardenolide under feeding and mock treatments. Dtl, digitalose; Fuc, fucose.

Formation of pregnenolone from the hypothesized sterol precursor is the first proposed step in the cardenolide biosynthetic pathway in plants (Fig. 1a). In mammals, pregnenolone is formed from cholesterol through the action of cytochrome P450 (CYP), family 11,

subfamily A, member 1 (CYP11A1) (Fig. 1b), also known as cholesterol side-chain cleavage enzyme, which catalyses a three-step reaction during steroid hormone biosynthesis (Supplementary Fig. 2). CYP11A1 is a multifunctional mitochondrial enzyme that catalyses

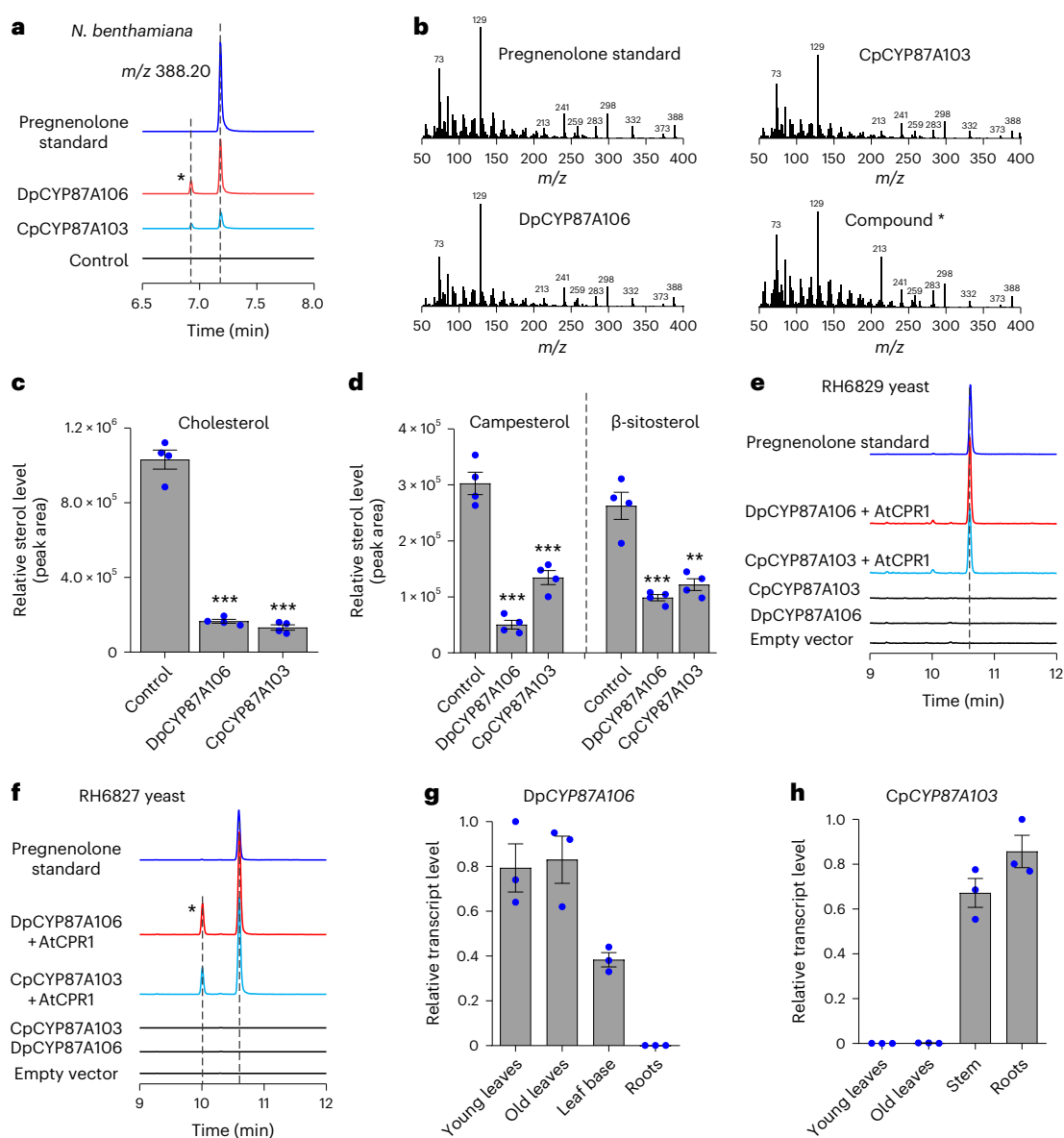


Fig. 2 | Discovery of DpCYP87A106 and CpCYP87A103 enzymes catalysing pregnenolone biosynthesis in cardenolide-producing plants. a, Aligned extracted ion chromatograms presenting the accumulation of pregnenolone (m/z 388.20, trimethylsilylated) following transient expression of DpCYP87A106 (red) or CpCYP87A103 (turquoise blue) in *N. benthamiana* compared with an authentic standard (dark blue). The asterisk shows the presence of a new compound at a different retention time, with the same mass spectrum as that of pregnenolone (see **b**). **b,** GC–MS spectra of pregnenolone and the new compound (indicated by the asterisk in **a**) produced in *N. benthamiana* transient experiments compared with pregnenolone standard. **c,d,** Levels of cholesterol (**c**) and phytosterols (campesterol and β -sitosterol) (**d**) in *N. benthamiana* leaves that transiently express DpCYP87A106 or CpCYP87A103 compared with control. Leaves infiltrated with empty vector were used as control (in **a**, **c** and **d**). The values in **c** and **d** indicate the mean \pm s.e.m. of four biological replicates ($n = 4$) obtained from four independently infiltrated plants. The asterisks indicate significant changes compared with control samples, as calculated by

two-tailed Student's *t*-test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **e,f,** Expression of DpCYP87A106 and CpCYP87A103 in the yeast strains RH6829 (**e**) and RH6827 (**f**), which were engineered to accumulate cholesterol and campesterol, respectively³⁵. Co-expression of DpCYP87A106 (red) or CpCYP87A103 (turquoise blue) together with AtCPR1 resulted in pregnenolone formation in RH6829 (**e**) and RH6827 (**f**) yeast strains, respectively. Aligned extracted ion chromatograms for pregnenolone (m/z 388.20, trimethylsilylated) are presented. Black represents yeast controls (empty vector (pESC-HIS), DpCYP87A106 alone and CpCYP87A103 alone); blue represents pregnenolone authentic standard. As seen in *N. benthamiana* transient assays (in **a** and **b**), we also detected the compound (shown as asterisk) with the same mass spectrum as that of pregnenolone, but with a different retention time, in the RH6827 yeast experiments. GC–MS was used for sterol profiling. **g,h,** DpCYP87A106 (**g**) and CpCYP87A103 (**h**) gene expression in selected tissues of *D. purpurea* (**g**) and *C. procera* (**h**), as determined by qRT-PCR. The values in **g** and **h** indicate the mean \pm s.e.m. of biological replicates obtained from three independent plants ($n = 3$).

two hydroxylation (C22 and C20) reactions followed by a C20–C22 bond cleavage²⁶ (Supplementary Fig. 2). In plants, sterol side-chain cleavage enzyme activity on different sterol substrates was evidenced decades ago^{27,28}. Although no such plant enzyme has been identified and characterized unambiguously, sterol C22 hydroxylation and other C–C bond cleavage activities are known to be catalysed by diverse CYP

classes in the plant kingdom^{29–32}. Moreover, previous work showed that nearly all enzymes involved in cholesterol and phytosterol biosynthesis in plants are localized to the endoplasmic reticulum (ER) membrane³³. Therefore, we speculated that the sterol pool used for cardenolide biosynthesis would be available in the ER membrane, and thus pregnenolone biosynthesis would probably be carried out by a

single ER-localized CYP enzyme (as in mammals) or by a set of two or three CYP enzymes.

Before searching for CYP genes, we first wanted to validate that pregnenolone is in fact the precursor for cardenolide biosynthesis in *D. purpurea* because previous feeding studies of isotope-labelled pregnenolone and derivatives showing incorporation into cardenolides were performed in *D. lanata*^{15,17,34}, a related *Digitalis* species. We fed stable isotope-labelled pregnenolone ([20,21-¹³C₂][16,16-D₂] pregnenolone) to the cut stems of 4-week-old *D. purpurea*. After 2–3 weeks, cardenolides were extracted from collected leaves and analysed using liquid chromatography–mass spectrometry (LC–MS). Labelled digitoxigenin glycosides (digitoxigenin digitaloside and digitoxigenin fucoside) and gitaloxigenin glycoside (verodoxin) were detected by monitoring the [M + H]⁺ ions with mass to charge ratio (*m/z*) values of 539.3450, 525.3280 and 582.3279. For plants treated only with water (mock), the corresponding non-labelled cardenolides were detected by monitoring the [M + H]⁺ ions with *m/z* values of 535.3252, 521.3102 and 579.3154 (Fig. 1c and Supplementary Fig. 3a). These results suggest that both carbons (C20 and C21) and hydrogens (16 α and 16 β) of pregnenolone were retained during biosynthesis of digitoxigenin-derived cardenolides (that is, digitoxigenin digitaloside and fucoside). In verodoxin, a derivative of gitaloxigenin, it appeared that one hydrogen (16 β) from pregnenolone was eliminated during the biosynthesis, which is in agreement with the gitaloxigenin structure (16 β -hydrogen modified to 16R; R = OCHO; Supplementary Fig. 3b). As previously observed in *D. lanata*^{15,17,34}, our findings confirm that pregnenolone is the central, committed precursor in cardenolide biosynthesis in *D. purpurea*, regardless of the unconfirmed starting sterol substrate.

To identify candidate CYP gene(s) involved in pregnenolone biosynthesis, we selected two cardenolide producers belonging to two distantly related plant families, *D. purpurea* (family Plantaginaceae) and *C. procera* (family Apocynaceae). We reasoned that comparative profiling of cardenolides in different tissues of *D. purpurea* and *C. procera*, coupled with transcriptome datasets from the same tissues, would facilitate gene discovery in these plants. Using ultra-high-performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry, we examined the cardenolide content in multiple tissue types of *D. purpurea* (young leaves, old leaves, leaf base (winged petiole) and roots) and *C. procera* (young leaves, old leaves, stem and roots). We observed increased accumulation of digitoxin and digitoxigenin fucoside in young and old leaves compared with leaf base and roots of *D. purpurea* (Supplementary Fig. 4a). Although cardenolides were found in all analysed tissue types in *C. procera*, they primarily accumulated in the young leaves (Supplementary Fig. 4b). Finally, we generated corresponding tissue-specific

RNA sequencing (RNA-seq) datasets from *D. purpurea* and *C. procera* plants (5–7 weeks old).

We initially selected 13 candidate CYP genes from the *D. purpurea* transcriptome that were highly expressed in young and old leaves compared with roots (Supplementary Fig. 5, left). Each CYP candidate was cloned and expressed in *Nicotiana benthamiana* leaves using an *Agrobacterium tumefaciens*-mediated transient expression system. *N. benthamiana* produces a diverse array of sterols (cholesterol, β -sitosterol, stigmasterol, campesterol and isofucosterol) in substantial amounts³³; therefore, this host does not require administration of exogenous sterol substrate, making it a highly efficient platform for rapid screening of candidate enzymes in pregnenolone biosynthesis. Metabolic profiling of the leaf extracts by gas chromatography–MS (GC–MS) showed that transient expression of one of these candidates (*D. purpurea* CYP family 87, subfamily A, member 106; DpCYP87A106) led to the production of pregnenolone (Fig. 2a,b), along with substantially decreased levels of three sterols, namely, cholesterol, campesterol and β -sitosterol (Fig. 2c,d). Another compound with the same mass spectrum as that of pregnenolone, but with a different retention time (shown as * in Fig. 2a,b) was also observed but in quantities too low for characterization.

Next we identified five candidate CYP genes in the *C. procera* transcriptome based on their increased expression in young leaves (Supplementary Fig. 5, right), the tissue with the highest accumulation of cardenolide (Supplementary Fig. 4b). However, none of these candidates produced pregnenolone when expressed transiently in *N. benthamiana* and, notably, this candidate list did not contain any member of the CYP87 family. As a next step, we searched for homologues of the *D. purpurea* gene by performing a BLAST search using DpCYP87A106 as the query against the *C. procera* transcriptome. A single hit (*C. procera* CYP family 87, subfamily A, member 103; CpCYP87A103) with 54% amino acid sequence identity to DpCYP87A106 was identified. The expression pattern of this gene was poorly correlated with cardenolide presence as it was highly expressed in the stem and roots but had negligible expression in young and old leaves (Supplementary Fig. 5, right). However, when CpCYP87A103 was transiently expressed in *N. benthamiana* leaves, GC–MS analysis of the leaf extract clearly showed the formation of pregnenolone, along with concomitant reduction in cholesterol, campesterol and β -sitosterol (Fig. 2a–d). Therefore, both DpCYP87A106 and CpCYP87A103 enzymes are capable of catalysing pregnenolone formation from an array of sterols. Moreover, cholesterol, campesterol and β -sitosterol appear to be capable of serving as sterol precursors for pregnenolone biosynthesis in *N. benthamiana*. To further support the pregnenolone-forming activity of DpCYP87A106 and CpCYP87A103 enzymes from multiple sterol

Fig. 3 | Silencing of DpCYP87A106 by RNAi in *D. purpurea* confirms its role in cardenolide biosynthesis. a,b, De novo production of pregnenolone in transgenic *Arabidopsis* leaves overexpressing DpCYP87A106 (red) (a) or CpCYP87A103 (blue) (b) compared with wild-type leaves (black). Lines 2 and 3 are independent DpCYP87A106-Ox transgenic *Arabidopsis* lines, whereas lines 1 and 3 are independent CpCYP87A103-Ox transgenic *Arabidopsis* lines. Extracted ion chromatograms are shown. GC–MS was used for sterol profiling. c, Levels of pregnenolone in leaves of DpCYP87A106-RNAi transgenic *D. purpurea* lines compared with wild type. Pregnenolone was not detected in the DpCYP87A106-RNAi lines. d, Knockdown of DpCYP87A106 resulted in almost complete loss of cardenolides in transgenic *D. purpurea* leaves. The relative cardenolide levels shown are the sum of the peak areas obtained from six cardenolides (verodoxin, digitoxigenin fucoside, digitoxin, purpurea glycoside A, purpurea glycoside B and glucogitaloxin) that typically accumulate in *D. purpurea*¹⁰. Four independent DpCYP87A106-RNAi transgenic *D. purpurea* lines (lines 1, 2, 3 and 4) were generated using a stable transformation approach and analysed in the T₀ generation. Values in c and d indicate means \pm s.e.m. of biological replicates ($n = 4$ for wild type and $n \geq 2$ for individual transgenic lines). Biological replicates (for example, for line 1, $n = 2$) refer to the individual plants that emerged from the same callus explant. Asterisks indicate significant changes compared with

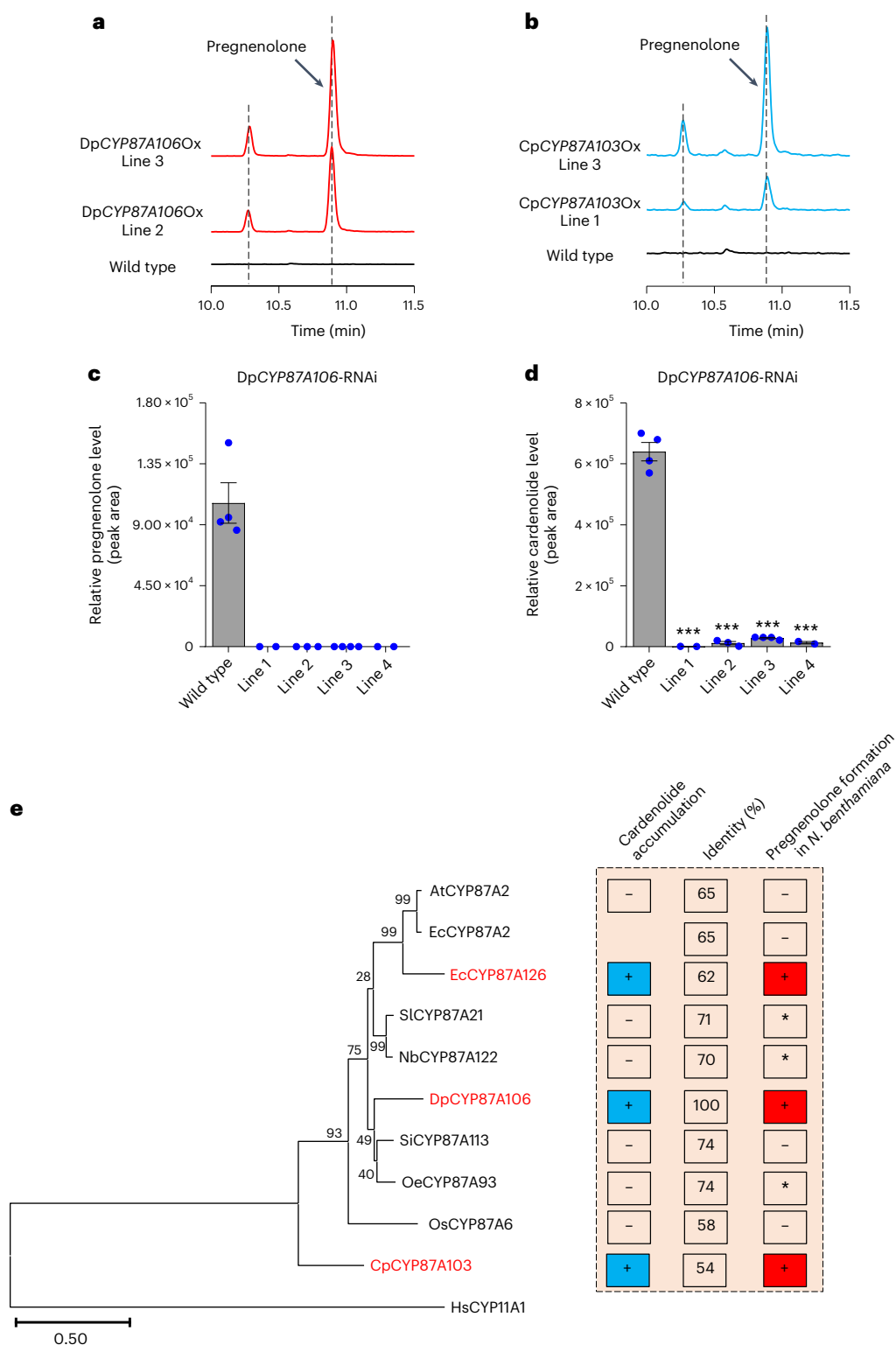
wild-type samples, as calculated by two-tailed Student's *t*-test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. e, Phylogenetic context of CYP87A proteins obtained from cardenolide-producing and cardenolide-free plants. The percentage identity at the amino acid level for each CYP87A protein compared with DpCYP87A106 is presented. The blue squares (with the + sign) indicate presence of chemotype (that is, cardenolide) in a particular plant species. In planta pregnenolone-producing activity of CYP87A clade proteins seen in *N. benthamiana* transient expression experiments is shown by red squares (with the + sign). Increased pregnenolone accumulation by three CYP87A proteins (marked in red), EcCYP87A126, DpCYP87A106 and CpCYP87A103, correlated well with the presence of cardenolides in these corresponding plant species. The asterisk represents negligible pregnenolone-formation activity in transient experiments by CYP87A enzymes (for example, tomato) from non-cardenolide-producing plants. CYP87A sequences from the following species were used in phylogenetic analysis: *S. lycopersicum* (Sl; tomato), *N. benthamiana* (Nb), *O. sativa* (Os; rice), *A. thaliana*, *S. indicum* (Si; sesame), *O. europaea* (Oe; olive), *E. cheiranthoides*, *D. purpurea*, *C. procera* and *H. sapiens* (humans). The amino acid sequences used in the phylogenetic analysis are provided in Supplementary Data 1. Percentage bootstrap values are shown at the nodes of each branch. Scale bar represents branch lengths measured in the number of amino acid substitutions per site.

wild-type samples, as calculated by two-tailed Student's *t*-test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. e, Phylogenetic context of CYP87A proteins obtained from cardenolide-producing and cardenolide-free plants. The percentage identity at the amino acid level for each CYP87A protein compared with DpCYP87A106 is presented. The blue squares (with the + sign) indicate presence of chemotype (that is, cardenolide) in a particular plant species. In planta pregnenolone-producing activity of CYP87A clade proteins seen in *N. benthamiana* transient expression experiments is shown by red squares (with the + sign). Increased pregnenolone accumulation by three CYP87A proteins (marked in red), EcCYP87A126, DpCYP87A106 and CpCYP87A103, correlated well with the presence of cardenolides in these corresponding plant species. The asterisk represents negligible pregnenolone-formation activity in transient experiments by CYP87A enzymes (for example, tomato) from non-cardenolide-producing plants. CYP87A sequences from the following species were used in phylogenetic analysis: *S. lycopersicum* (Sl; tomato), *N. benthamiana* (Nb), *O. sativa* (Os; rice), *A. thaliana*, *S. indicum* (Si; sesame), *O. europaea* (Oe; olive), *E. cheiranthoides*, *D. purpurea*, *C. procera* and *H. sapiens* (humans). The amino acid sequences used in the phylogenetic analysis are provided in Supplementary Data 1. Percentage bootstrap values are shown at the nodes of each branch. Scale bar represents branch lengths measured in the number of amino acid substitutions per site.

substrates, we also expressed these enzymes in *Saccharomyces cerevisiae*. As yeast do not produce cholesterol or phytosterols (for example, campesterol), we used previously reported stable yeast strains, RH6829 and RH6827, which were engineered to produce cholesterol and campesterol, respectively³⁵. Expression of either DpCYP87A106 or CpCYP87A103 together with the *Arabidopsis* CYP reductase 1 (AtCPR1) in cholesterol-producing RH6829 yeast resulted in formation of pregnenolone (Fig. 2e). Similarly, RH6827 yeast (campesterol generating) cells expressing individual CYP87A enzymes together with

AtCPR1 were also able to produce pregnenolone (Fig. 2f). Therefore, both DpCYP87A106 and CpCYP87A103 enzymes produce pregnenolone from either cholesterol or campesterol and, importantly, are sufficient for pregnenolone biosynthesis from sterol precursors in heterologous hosts.

Although cardenolides are found in the leaves of both *D. purpurea* (Supplementary Fig. 4a) and *C. procera* (Supplementary Fig. 4b), the corresponding transcriptome datasets (one replicate per tissue) indicate that DpCYP87A106 is predominantly expressed in young and old



leaves, whereas CpCYP87A103 is mainly expressed in stem and roots (Supplementary Fig. 5). This expression pattern was further validated by quantitative real-time PCR (qRT-PCR) measurements performed in an independent experiment (three biological replicates per tissue, $n = 3$; Fig. 2g,h). Therefore, stem and root tissues appear to be the primary site of pregnenolone biosynthesis in the early cardenolide pathway in *C. procera*, and either pregnenolone or downstream intermediates are probably transported to young and old leaves for further biosynthetic steps. In contrast, in *D. purpurea*, the pregnenolone biosynthetic enzyme and the cardenolides are both located in the leaf, suggesting that intermediates are not transported for cardenolide biosynthesis in this plant. Moreover, the accumulation of cholesterol and phytosterols (for example, campesterol and β -sitosterol) in these tissues of *C. procera* and *D. purpurea* indicate that the starting sterol precursors required for pregnenolone biosynthesis are available in these plants (Supplementary Fig. 6).

We next assessed whether DpCYP87A106 and CpCYP87A103 could produce pregnenolone in a non-cardenolide-producing model plant such as *Arabidopsis thaliana*. We generated stable transgenic *A. thaliana* lines overexpressing either DpCYP87A106 (named DpCYP87A106-Ox) or CpCYP87A103 (named CpCYP87A103-Ox). We observed de novo production of pregnenolone in the leaves of DpCYP87A106-Ox and CpCYP87A103-Ox plants (Fig. 3a,b), confirming that both CYP87A subfamily enzymes catalyse formation of pregnenolone in a plant unrelated to *N. benthamiana*.

To definitively establish whether these enzymes play a direct role in cardenolide biosynthesis, we used RNA-mediated interference (RNAi) to silence DpCYP87A106 (DpCYP87A106-RNAi) in *D. purpurea*. DpCYP87A106 transcript levels were significantly reduced in the leaves of DpCYP87A106-RNAi plants (Supplementary Fig. 7a). Furthermore, we could not detect pregnenolone in the leaves (6–8 weeks old) of DpCYP87A106-RNAi lines (Fig. 3c). DpCYP87A106-RNAi leaves also showed a substantial decrease in cardenolide levels compared with wild type *D. purpurea* leaves (Fig. 3d). Analysis of sterols in DpCYP87A106-RNAi leaves showed a significant increase in campesterol levels, with no major change in cholesterol (Supplementary Fig. 7b), suggesting that campesterol could be the preferred sterol precursor for pregnenolone formation in *D. purpurea*. To examine whether the low cardenolide chemotype in DpCYP87A106-silenced lines could be rescued, we fed the pregnenolone precursor to cut leaf discs of DpCYP87A106-RNAi lines. After 4–6 days, leaf discs were extracted and analysed for cardenolides using LC–MS. The cardenolide content in pregnenolone-treated DpCYP87A106-RNAi lines remain unaffected compared with that of control DpCYP87A106-RNAi lines (water treated). However, pregnenolone-treated DpCYP87A106-RNAi plants (lines 1 and 2) showed accumulation of 5 β -pregnane-3,20-dione and 5 β -pregnane-3 β -ol-20-one, which are known downstream intermediates of the cardenolide pathway (Supplementary Fig. 8). We did not detect these pathway intermediates in the control DpCYP87A106-RNAi (water treated) plants. As the cut leaf discs are only metabolically active for less than a week, we hypothesize that this was not enough time for all the downstream enzymatic steps required for complete rescue of cardenolide biosynthesis. Unfortunately, these feeding studies were not technically possible with the mature transformed *Digitalis* plants (9–10 months old).

The cholesterol side-chain cleavage enzyme in humans, *Homo sapiens* CYP11A1 (HsCYP11A1), catalyses a well-studied, three-step reaction sequence in which two stereoselective successive hydroxylations of cholesterol (at C20 and C22), followed by C20–C22 cleavage of 20,22-dihydroxycholesterol, results in the formation of pregnenolone³⁶. To compare the potential mechanism of the plant enzyme with the human one, we generated structural models of DpCYP87A106 and CpCYP87A103 (ref. 37). The catalytic pockets of DpCYP87A106 and CpCYP87A103 (Supplementary Fig. 9, shown in blue and green, respectively) were highly conserved despite the

overall low homology of the two proteins (54% amino acid identity). Next we performed docking of cholesterol, 22-hydroxycholesterol and 20,22-dihydroxycholesterol substrates in the active site of DpCYP87A106 (Supplementary Fig. 10a,c,e) and CpCYP87A103 (Supplementary Fig. 10b,d,f). The large number of non-polar residues within the catalytic pocket suggests steric hindrance is important to correctly orientate the substrate, as reported for HsCYP11A1 (ref. 36). For each docked sterol, we could achieve a substrate orientated with the hydroxylation (for cholesterol or 22-hydroxycholesterol) or C–C cleavage position (for 20,22-dihydroxycholesterol) between 2.7 Å and 6.1 Å of the ferric haem, suggesting that a three-step reaction sequence, as proposed for HsCYP11A1, is plausible³⁶. Similar results were achieved when DpCYP87A106 and CpCYP87A103 homology models were docked with campesterol, 22-hydroxycampesterol and 20,22-dihydroxycampesterol (Supplementary Fig. 11a–f). From the substrate-docking studies, we identified Ile210 in DpCYP87A106 (Ile209 in CpCYP87A103) as a residue that might control substrate orientation by steric hindrance (Supplementary Fig. 12a,b). Site-directed mutagenesis of DpCYP87A106 Ile210 and CpCYP87A103 Ile209 to either phenylalanine or tryptophan led to enzyme mutants that produced negligible or no pregnenolone when expressed transiently in *N. benthamiana* (Supplementary Fig. 12c,e). Both point mutations resulted in restoration of the levels of cholesterol and the phytosterols campesterol and β -sitosterol to those observed in control plants (Supplementary Fig. 12d,f), suggesting that Ile209 or Ile210 is important for sterol substrate binding in pregnenolone formation.

To explore the evolutionary aspect of the CYP87A family enzymes involved in cardenolide biosynthesis, we conducted a BLAST search of DpCYP87A106 against National Center for Biotechnology Information (NCBI) and public databases and extracted the putative orthologous coding sequences from an additional cardenolide-producing plant (*E. cheiranthoides*), and several cardenolide-free plants (*A. thaliana*, *Oryza sativa* (rice), *Solanum lycopersicum* (tomato), *Sesamum indicum* (sesame), *Olea europaea* (olive) and *N. benthamiana*). These identified CYP87A proteins share 54–74% homology with DpCYP87A106, and their existence in both cardenolide-free and cardenolide-producing species suggests that CYP87A is widely conserved in plants (Fig. 3e). However, so far no CYP87A subfamily member had been functionally characterized. We identified two CYP87A proteins (EcCYP87A2 and EcCYP87A126) sharing 75% amino acid identity in *E. cheiranthoides*, a substantial cardenolide producer from Brassicaceae family. We assayed each of these CYP87A proteins (10 in total including DpCYP87A106 and CpCYP87A103; Fig. 3e and Supplementary Fig. 13) for pregnenolone-forming activity via transient expression in *N. benthamiana*. In addition to DpCYP87A106 and CpCYP87A103, transient expression of EcCYP87A126 led to increased accumulation of pregnenolone (Supplementary Fig. 13), suggesting its involvement in *Erysimum* cardenolide biosynthesis. In contrast, we detected only negligible levels of pregnenolone with CYP87A members from tomato, olive and *N. benthamiana* (Supplementary Fig. 13). Phylogenetic analysis shows a separate small clade containing *A. thaliana* CYP87A2 (AtCYP87A2) and EcCYP87A2 enzymes, both of which lack pregnenolone formation activity (Fig. 3e). Clear separation of EcCYP87A126 from EcCYP87A2 and the *Arabidopsis* enzyme (AtCYP87A2) suggest that EcCYP87A126 may have arisen via gene duplication from an EcCYP87A2-like ancestor before undergoing neo-functionalization to acquire pregnenolone-formation activity (Fig. 3e). However, we observed no such clades for *Digitalis* and *Calotropis* enzymes. Thus, the evolutionary mechanism by which DpCYP87A106 and CpCYP87A103 acquired the catalytic activity required for pregnenolone biosynthesis remains undetermined.

CYP87D subfamily members, such as CYP87D16 (*Maesa lanceolata*), CYP87D20 (*Cucumis sativus*) and CYP87D18 (*Siraitia grosvenorii*), catalyse single or multiple hydroxylation and/or oxidation reactions in triterpene biosynthesis³⁸. Our study shows that the CYP87A subfamily

members DpCYP87A106 from *D. purpurea* and CpCYP87A103 from *C. procera* act on sterol molecules in cardenolide biosynthesis. Characterization of CYP87A activity in pregnenolone formation is a significant step towards resolving one of the most complex and lengthy biosynthetic pathways of sterol-derived cardenolides in distantly related plant families (for example, Apocynaceae (*Calotropis*, *Nerium*, *Asclepias*), Plantaginaceae (*D. lanata*), Asparagaceae (*Convallaria majalis*) and Moraceae (*Antiaris toxicaria*)). Furthermore, pregnenolone serves as a precursor for many high-value steroidal molecules, including hormones (for example, progesterone, testosterone, androsterone, estradiol and many more) and drugs (for example, corticosteroids and anti-inflammatory and anti-allergic medications). Discovery of the CYP87A enzymes provides a crucial tool for building a platform for engineering the sustainable production of high-value steroids using state-of-the-art synthetic biology applications.

Methods

Plant material

D. purpurea (foxglove) and *C. procera* plants were grown in a climate-controlled greenhouse at 24 °C during the day and 18 °C during the night, with natural light. *N. benthamiana* plants were grown in a growth room maintained at 23 ± 2 °C with a 16 h day and 8 h night regime.

Analytical standards

Digitoxigenin and digitoxin standards (Sigma-Aldrich) were dissolved in methanol at a concentration of 1 mg ml⁻¹. Sterol standards, cholesterol, campesterol, β -sitosterol, pregnenolone and [20,21-¹³C₂] [16,16-D₂]pregnenolone were purchased from Sigma-Aldrich and, unless stated otherwise, were dissolved in ethanol to a concentration of 1 mg ml⁻¹. 5 β -Pregnane-3,20-dione and 5 β -pregnane-3 β -ol-20-one (pregnanolone) standards were purchased from Steraloids and dissolved in ethanol to a concentration of 1 mg ml⁻¹ for further use.

Plant extract preparation and LC–MS-based targeted profiling of cardenolides

Preparation of extracts of *D. purpurea* (young leaves, old leaves, leaf base (winged petiole) and roots) and *C. procera* (young leaves, old leaves, stem and roots) tissues ($n = 3$) was performed as previously described^{10,11}. Briefly, selected plant tissues were frozen in liquid nitrogen and ground to a fine powder using a TissueLyser II (Qiagen) homogenizer or mortar and pestle. Frozen tissue (100 mg) was extracted with 80% methanol, briefly vortexed for 2 min and then incubated at 65 °C for 12 min. Finally, the extracts were centrifuged for 15 min at 20,000g and filtered through 0.22 μ m filters. Samples were analysed using a Thermo Scientific UltiMate 3000 ultra-high-performance liquid chromatography system coupled to an Impact II UHR-Q-ToF (ultra-high-resolution quadrupole time of flight) mass spectrometer (Bruker Daltonics) with the standard (22 min, positive mode) run conditions as follows: 5% B for 1 min, 5% B to 95% B in 12 min, 95% B for 2 min, changing to 100% B within 0.2 min and continuing at 100% B for 2.8 min, and finally returning to the initial conditions (5% phase B) within 0.5 min. The column was equilibrated with 5% B for another 3.8 min before next injection. The mobile phase consisted of 0.1% formic acid in water (phase A) and acetonitrile (phase B). Separation of metabolites was performed on a 50 × 2.1 mm, 1.7 μ m Acquity UPLC C18 column (Waters). The flow rate was 0.3 ml min⁻¹, and the column temperature was kept at 35 °C. Mass spectrometry was performed in positive electrospray ionization mode (capillary voltage 4,000 V; end plate offset 500 V; nebulizer pressure 2.2 bar; drying gas, nitrogen at 250 °C and 10 l min⁻¹). Mass spectrometry data were recorded at 12 Hz, ranging from 100 m/z to 1,200 m/z in auto tandem mass spectrometry mode with an active exclusion window of 0.2 min. Fragmentation was triggered on an absolute threshold of 400 counts and restricted to a total cycle time of 0.5 s, with dynamic collision energy (20–50 eV). To calibrate spectrum recording for mass spectrometry, each run was

initiated with the direct source infusion of a sodium formate–isopropanol calibration solution (using external syringe pump at 0.18 ml h⁻¹). The initial 1 min of the chromatographic gradient was directed towards the waste. Cardenolides were identified by comparing the retention time and mass spectra of authentic standards analysed on the same instrument (see ‘Analytical standards’ above). When the corresponding standards were not available, metabolites were putatively identified by comparing their retention times, elemental composition and mass fragmentation pattern with those described in the literature^{10,12,13}. Relative quantification of the cardenolides was performed using Bruker Compass Data Analysis (v.5.3) software.

Transcriptome analysis

Total RNA from 5–7-week-old *D. purpurea* (young leaves, old leaves, leaf base (winged petiole) and roots) and *C. procera* (young leaves, old leaves, stem and roots) was extracted using the RNeasy Mini Kit (Qiagen), according to the manufacturer’s instructions. High-quality RNA (260 nm:280 nm ratio of 2.0–2.2; 260 nm:230 nm ratio of 2.1–2.4) samples were submitted to Novogene (<https://en.novogene.com/>) for preparing mRNA libraries and further RNA-seq (paired-end 2 × 150, -40M reads per sample) using the company’s standard protocols. The raw sequencing data were assembled using an in-house pipeline developed for transcriptome analysis. In brief, de novo transcriptome assemblies were generated for *D. purpurea* and *C. procera* from cleaned, trimmed reads using Trinity³⁹. The transcriptome assemblies were refined using the CD-HIT Suite to group transcripts with greater than 90% identity, and only the longest transcript was retained in each case. Transdecoder (<https://github.com/TransDecoder/TransDecoder>) was used to identify candidate-coding regions within transcript sequences. Functional annotation was then performed by running BLAST against the UniProt and Pfam database. Finally, gene expression, measured in transcripts per million, was calculated using Salmon⁴⁰.

Feeding experiment

Isotope-labelled pregnenolone (2 mg) was dissolved in 100 μ l ethanol. *D. purpurea* plants grown in vitro were used in feeding experiments. Briefly, 4-week-old plants were trimmed by cutting off the base stem, placed into plastic tubes with de-ionized water (~2 ml) and then fed with stable isotope-labelled pregnenolone (final concentration, 0.23 mM). Plants treated with de-ionized water were used as control (mock). Treated and mock plants were further maintained in a growth room at 25 °C, with a 16:8 hour photoperiod under LED light. After 2–3 weeks, leaves were collected and analysed by LC–MS for the accumulated cardenolides, as described above.

Generation of DpCYP87A106 transgenic *D. purpurea* plants

The DpCYP87A106-silencing construct (DpCYP87A106-RNAi) was generated by introducing a 321 bp DpCYP87A106 fragment (forward primer, GCGGCCGCGAGCTGGTATCCAAGCACCTTC; reverse primer, GCGGCCGCGCTGATCAAACCATCTATGAACGCC) to pENTR/D-TOPO (Invitrogen; by NotI and AscI) and then cloning to the pK7GWIWG2 (II) binary vector using the Gateway LR Clonase II enzyme mix (Invitrogen). The vector was stably transformed to *D. purpurea* as follows: seeds of *D. purpurea* were surface sterilized by dipping in 70% ethanol for 30 s, followed by 15 min in 1% sodium hypochlorite and then rinsing three to five times in sterile distilled water. The sterile seeds were germinated in half-strength MS221 (Murashige and Skoog) medium, supplemented with 1% sucrose and 0.8% plant agar (Duchefa) adjusted to pH 5.8. Plants were maintained in a growth room at 25 °C, with a 16:8 hour photoperiod under LED light. Overnight-grown *A. tumefaciens* (strain GV3101) bacteria containing the binary vector pK7GWIWG2 (II)–DpCYP87A106 were centrifuged and the pellet was resuspended (OD₆₀₀ = 0.4) in CT medium containing 3% glucose and 0.2 g l⁻¹ KH₂PO₄, used for liquid co-cultivation. Well-developed first and second pairs of leaves from 5-week-old *D. purpurea* plants germinated

in vitro were used as explants. The distal edges of leaf explants were cut out and briefly dipped in the bacterial suspension (supplemented with 100 μM acetosyringone), incubated in the dark for 20–30 min, placed on CT solid media, already supplemented with 100 μM acetosyringone, 75 $\mu\text{g ml}^{-1}$ DTT, 2 $\mu\text{g ml}^{-1}$ zeatin and 0.2 $\mu\text{g ml}^{-1}$ indole-3-acetic acid (IAA). After 48 h of co-cultivation in dark conditions, explants were transferred to MSI (shoot induction) medium containing zeatin (2 $\mu\text{g ml}^{-1}$), IAA (0.2 $\mu\text{g ml}^{-1}$), kanamycin (50 $\mu\text{g ml}^{-1}$) and ticarcillin (250 $\mu\text{g ml}^{-1}$) for 3 weeks, and then transferred onto MSII (shoot elongation) medium containing zeatin (1 $\mu\text{g ml}^{-1}$), zeatin riboside (1 $\mu\text{g ml}^{-1}$), IAA (0.2 $\mu\text{g ml}^{-1}$), kanamycin (50 $\mu\text{g ml}^{-1}$) and ticarcillin (100 $\mu\text{g ml}^{-1}$). Well-developed shoots (after 3–4 weeks) were excised and transferred to MSIII (rooting medium) containing indole-3-butyric acid (2 $\mu\text{g ml}^{-1}$), kanamycin (50 $\mu\text{g ml}^{-1}$) and ticarcillin (100 $\mu\text{g ml}^{-1}$). Finally, plantlets with roots were transferred to soil for hardening in greenhouse and maintained further. Wild-type (non-transformed) *D. purpurea* plants germinated in vitro were used as a control in transgenic plant analysis. Positive transgenic lines were selected by qRT-PCR and then used for LC-MS and GC-MS-based metabolite analysis. Four independent transgenic lines (lines 1, 2, 3 and 4) were successfully generated after transformation and analysed in the T_0 generation. Each transgenic plant line was generated from callus (produced from leaf explants) during the plant transformation process. In the case of *D. purpurea*, multiple shoots emerged from a single callus explant in tissue culture, and these shoots became multiple plantlets after passing through the shoot elongation, rooting, acclimatization and hardening steps of plant transformation. Therefore, biological replicates here (for example, for line 1, $n = 2$ biological replicates) refer to the individual plants that each emerged from the same callus explant.

qRT-PCR

Total RNA was isolated from different tissues of *D. purpurea* and *C. procera* plants and from transgenic *D. purpurea* (leaves) using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Unless stated otherwise, at least three biological replicates from each genotype were used for gene expression analysis ($n \geq 3$) by qRT-PCR. For transgenic *D. purpurea* lines, biological replicates ($n \geq 2$) were used. DNase I-treated (Sigma-Aldrich) RNA was reverse transcribed using a high-capacity cDNA reverse transcription kit (Applied Biosystems). Gene-specific oligonucleotides were designed with Primer BLAST software (NCBI). *UBIQUITIN10* (ref. 41) and *ACTIN12* genes were used as the reference genes for *Digitalis* and *Calotropis*, respectively, in expression analysis.

Transient expression in *N. benthamiana*

For transient overexpression, all candidate genes, including *DpCYP87A106* and *CpCYP87A103* (list provided in Supplementary Fig. 5), were cloned into destination vector 3 Ω 1 using GoldenBraid cloning⁴² and transformed into *A. tumefaciens* (GV3101) by electroporation. Single clones with each target construct were inoculated into 10 ml of LB medium supplemented with antibiotics (250 $\mu\text{g ml}^{-1}$ spectinomycin and 50 $\mu\text{g ml}^{-1}$ gentamicin) and cultures were grown overnight at 28 °C with shaking (200 rpm). Cells were centrifuged at 2,000g for 15 min and the pellet was resuspended in 5 ml of infiltration buffer (50 mM MES buffer (pH 5.6), 10 mM MgCl₂, 150 μM acetosyringone). After another round of centrifugation, the pellet was resuspended again in 10 ml of infiltration buffer and incubated at room temperature for 2 h. *Agrobacterium* suspensions ($\text{OD}_{600} = 0.4$ for each strain) were infiltrated into 4–6-week-old *N. benthamiana* leaves. After 5 days, leaves were collected for further LC-MS-based cardenolide analysis and GC-MS-based sterol analysis. Biological replicates consisted of several leaves collected from different infiltrated plants. Leaves infiltrated with empty vector were used as controls. In other experiments, first site-directed mutants were generated for *DpCYP87A106* (I210F and I210W) and *CpCYP87A103* (I209F and I209W) and subsequently cloned into the 3 Ω 1 vector. The

resulting constructs were used for transient expression in *N. benthamiana* as described above.

GC-MS analysis of sterols

Powdered, frozen tissues (100 mg) were saponified at 70 °C for 2 h in 0.6 ml of 20% KOH (w/v) in 50% ethanol. Samples were mixed every 20 min during this procedure. Upon cooling to room temperature, samples were extracted three times with 0.5 ml hexane. The combined hexane phases were evaporated to dryness using a gentle stream of nitrogen, resuspended in 60 μl of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA), and incubated for 10 min at room temperature and then for 20 min at 65 °C. Each sample was transferred to the glass insert, and 1 μl was injected onto GC-MS with or without dilution, depending on the run conditions and analyte concentrations. The GC-MS system comprised a GC PAL autosampler (CTC Analytics), a trace 1310 GC ultra-gas chromatograph equipped with a split-splitless injector and ISQ quadrupole mass spectrometer (Thermo Scientific). GC was performed on two different columns: column 1, 30 m \times 0.25 mm \times 0.1 μm Zebron ZB-5HT with 5 m guard mass spectrometry column (Phenomenex; for example, data generated using this column was shown in Fig. 2a); column 2, 30 m \times 0.25 mm \times 0.25 μm Zebron ZB-5 with 10 m guard mass spectrometry column (Phenomenex; generated data shown in Fig. 3a,b). Samples were analysed in the split mode and the inlet temperature was set at 250 °C. Analytes were separated using the following chromatographic conditions: helium was used as carrier gas at a flow rate of 1.1 ml min⁻¹. The thermal gradient used for column 1 started at 160 °C (hold 1.5 min), ramped up to 270 °C at 30 °C min⁻¹ and then ramped up to 290 °C at 1.5 °C min⁻¹ (hold 15 min); the thermal gradient used for column 2 started at 160 °C (hold 1.5 min), ramped up to 270 °C at 30 °C min⁻¹, ramped up to 290 °C at 1 °C min⁻¹ (hold 10 min) and then finally up to 300 °C at 30 °C min⁻¹ (hold 2.5 min). Eluents were fragmented in the electron impact mode with an ionization voltage of 70 eV, and the mass spectrometry transfer line temperature was set at 290 °C and the ion source at 250 °C. The reconstructed ion chromatograms and mass spectra were evaluated using Xcalibur software (v.4.2.47; Thermo Scientific). Sterol compounds were identified by comparing their retention index and mass spectrum with those generated for authentic standards (trimethylsilylated) analysed on the same instrument (see 'Analytical standards' above) and those reported in the literature^{43–45}.

Expression of *DpCYP87A106* and *CpCYP87A103* genes in yeast and metabolite extraction

Both *DpCYP87A106* and *CpCYP87A103* genes were cloned separately into pESC-HIS plasmid using BamHI and XhoI restriction enzymes. The resulting pESC-HIS plasmid harbouring either *DpCYP87A106* or *CpCYP87A103* was used as a template further to clone *AtCPRI* (AT4G24520) using NotI and SacI restriction enzymes. Final pESC-HIS-*DpCYP87A106* and pESC-HIS-*CpCYP87A103* constructs with or without *AtCPRI* were transformed into RH6827 and RH6829 yeast strains using the Yeastmaker yeast transformation system (Clontech). Both RH6827 and RH6829 yeast strains were engineered to accumulate campesterol and cholesterol, respectively³⁵. Transformed yeast was grown on synthetic drop-out medium supplemented with appropriate amino acids (without histidine) and 2% glucose. Colony PCR was used to confirm the presence of the transgene in transformed yeast strains. Positive yeast clones were grown for 36 h at 30 °C in 2 ml synthetic drop-out liquid medium supplemented with appropriate amino acids (without histidine) and 2% glucose. Then the yeast cell cultures were centrifuged at 700g for 10 min, and pellets were resuspended in 10 ml H₂O and centrifuged again. Finally, washed yeast cells were transferred to synthetic drop-out medium supplemented with appropriate amino acids (without histidine) and 2% galactose, and grown further for 24 h at 30 °C. After induction of gene expression, yeast cells were centrifuged at 700g for 10 min, the pellet was resuspended

in 0.6 ml of saponification solution (20% KOH (w/v) in 50% ethanol) and transferred to a 2 ml Eppendorf tube. A volume of glass beads (0.5 mm diameter) roughly half of the cell resuspension was added to the Eppendorf tube, and this was vortexed 10 times for 1 min. Between each vortexing cycle, cells were kept on ice for 1 min. Lysed cells were saponified at 70 °C for 2 h, extracted three times with hexane, dried and resuspended in 50 µl of MSTFA. Derivatized samples were incubated for 20 min at 65 °C and analysed on GC–MS for sterol analysis, as described above with one modification in the GC column. For the yeast experiment, GC was performed on the following newly installed column: 30 m × 0.25 mm × 0.25 µm Zebtron ZB-5 with 10 m guard mass spectrometry column (Phenomenex; generated data shown in Fig. 2e,f).

Generation of *Arabidopsis* transgenic lines

A. tumefaciens GV3101 strains harbouring the 3Ω1–DpCYP87A106 and 3Ω1–CpCYP87A103 constructs were transformed separately into *Arabidopsis* (Col-0) plants using the floral dip method as previously described⁴⁶. Leaves from positive transgenic lines were analysed further for pregnenolone formation using GC–MS as described above.

Phylogenetic analysis

CYP87A clade sequences from various plants were obtained from NCBI and public databases using the BlastP program (using DpCYP87A106 as the query). Sequence alignments were performed using ClustalO-mega⁴⁷. The maximum-likelihood tree was inferred in MEGAX⁴⁸ with the following parameters: 1,000 bootstrap replications, Poisson model, discrete gamma distribution (five categories) and partial deletion. Evolutionary distances are in units of number of amino acid substitutions per site. The amino acid sequences used in the phylogenetic analysis are provided in Supplementary Data 1.

Synthetic gene sequences

De-novo-synthesized CYP87A gene sequences were obtained from Twist Bioscience for the following plant species: *O. sativa* (rice), *S. indicum* (sesame), *O. europaea* (olive) and *E. cheiranthoides*. As we could not amplify CYP87A from *S. lycopersicum* (tomato) and *N. benthamiana*, synthetic versions of these genes were also acquired from Twist Bioscience. All the synthetic CYP87A genes were cloned into 3Ω1 and used for transient expression in *N. benthamiana* as described above. The primers used for cloning are provided in Supplementary Table 1.

DpCYP87A106 and CpCYP87A103 modelling and docking studies

Homology models of DpCYP87A106 and CpCYP87A103 were generated using ColabFold, accessed via Google Colaboratory, and run using default parameters³⁷. The highest-scoring model of each protein, generated based on the predicted local distance difference test, was used for further study. Docking was performed using AutoDock Vina on the Webina webserver with default parameters (exhaustiveness, 8)⁴⁹. Substrates and receptors were prepared for docking using PDBQTConvert. Models with ligand orientations in which the hydroxylation or C–C cleavage site was in proximity to the ferric heme were selected for further study; this orientation was not always the lowest possible energy solution. Results for modelling and docking studies were visualized using PyMol (v2.0; PyMOL Molecular Graphics System; Schrödinger).

Pregnenolone feeding to DpCYP87A106-RNAi *D. purpurea* lines

Pregnenolone (20 mg) was dissolved in 1 ml ethanol. Leaf discs from DpCYP87A106-RNAi *D. purpurea* plants (for example, lines 1 and 2) were used in feeding experiments. Briefly, 10 mm leaf discs (5–7 discs for each genotype) were cut using a borer, placed into 5 ml Eppendorf tubes with de-ionized water (~2 ml) and then fed with pregnenolone (final concentration, 0.32 mM). Leaf discs treated with de-ionized water were used as control. Eppendorf tubes with treated and control

leaf discs were further incubated in a growth chamber at 25 °C, with 16:8 hour photoperiod under LED light. After 4–6 days, leaf discs were collected and analysed by LC–MS for the accumulated cardenolides or pathway intermediates as described above.

Statistical analysis

Microsoft Excel 2016 and GraphPad Prism 8 software were used for regular statistical analysis. A two-tailed Student's *t*-test was used to calculate significant differences among samples or genotypes. Details of biological or technical replicates and statistical parameters used in various experiments are provided in Methods and the captions of the figures and supplementary figures, wherever necessary.

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 work are available within the article and its Supplementary Information. *D. purpurea* and *C. procera* RNA-seq data associated with this article have been deposited into the NCBI Sequence Read Archive with BioProject identifiers PRJNA929980 and PRJNA1003839, respectively. The databases used in this study are UniProt Swiss-Prot (release 2022_02) and the Pfam database (Pfam-A.hmm release 34.0). Correspondence and requests for materials should be addressed to P.D.S. or S.E.O.C. Source data are provided with this paper.

References

- Wen, S. et al. Cardenolides from the Apocynaceae family and their anticancer activity. *Fitoterapia* **112**, 74–84 (2016).
- Agarwal, A., Petschenka, G., Bingham, R., Weber, M. & Rasmann, S. Toxic cardenolides: chemical ecology and coevolution of specialized plant-herbivore interactions. *New Phytol.* **194**, 28–45 (2012).
- Prassas, I. & Diamandis, E. P. Novel therapeutic applications of cardiac glycosides. *Nat. Rev. Drug Discov.* **7**, 926–935 (2008).
- Prassas, I., Paliouras, M., Datti, A. & Diamandis, E. P. High-throughput screening identifies cardiac glycosides as potent inhibitors of human tissue kallikrein expression: implications for cancer therapies. *Clin. Cancer Res.* **14**, 5778–5784 (2008).
- Juncker, T., Schumacher, M., Dicato, M. & Diederich, M. UNBS1450 from *Calotropis procera* as a regulator of signaling pathways involved in proliferation and cell death. *Biochem. Pharmacol.* **78**, 1–10 (2009).
- Newman, R. A., Yang, P., Pawlus, A. D. & Block, K. I. Cardiac glycosides as novel cancer therapeutic agents. *Mol. Interv.* **8**, 36–49 (2008).
- Mekhail, T. et al. Phase 1 trial of Anvirzel in patients with refractory solid tumors. *Invest. New Drugs* **24**, 423–427 (2006).
- Wong, R., Lingwood, C., Ostrowski, M., Cabral, T. & Cochrane, A. Cardiac glycoside/aglycones inhibit HIV-1 gene expression by a mechanism requiring MEK1/2-ERK1/2 signaling. *Sci. Rep.* **8**, 850 (2018).
- Van Quaquebeke, E. et al. Identification of a novel cardenolide (2'-oxovoruscharin) from *Calotropis procera* and the hemisynthesis of novel derivatives displaying potent in vitro antitumor activities and high in vivo tolerance: structure–activity relationship analyses. *J. Med. Chem.* **48**, 849–856 (2005).
- Gopal, B. R., Guardian, M., Dickman, R. & Wang, Z. Profiling and structural analysis of cardenolides in two species of *Digitalis* using liquid chromatography coupled with high-resolution mass spectrometry. *J. Chromatogr. A* **1618**, 460903 (2020).
- Züst, T. et al. Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae). *eLife* **9**, e51712 (2020).

12. Pandey, A. et al. Transcriptome and metabolite analysis reveal candidate genes of the cardiac glycoside biosynthetic pathway from *Calotropis procera*. *Sci. Rep.* **6**, 34464 (2016).
13. Kanojiya, S. & Madhusudanan, K. Rapid identification of calotropagenin glycosides using high-performance liquid chromatography electrospray ionisation tandem mass spectrometry. *Phytochem. Anal.* **23**, 117–125 (2012).
14. Withering, W. *An Account of Foxglove, and Some of Its Medicinal Uses* (M. Swinney, 1785).
15. Tschesche, R. Biosynthesis of cardenolides, bufadienolides and steroid sapogenins. *Proc. R. Soc. Lond. B* **180**, 187–202 (1972).
16. Caspi, E., Lewis, D. O., Piatak, D. M., Thimann, K. V. & Winter, A. Biosynthesis of plant sterols. Conversion of cholesterol to pregnenolone in *Digitalis purpurea*. *Experientia* **22**, 506–507 (1966).
17. Caspi, E. & Lewis, D. O. Progesterone: its possible role in biosynthesis of cardenolides in *Digitalis lanata*. *Science* **156**, 519–520 (1967).
18. Bennett, R. D., Heftmann, E. & Winter, B. J. A function of sitosterol. *Phytochemistry* **8**, 2325–2328 (1969).
19. Milek, F., Reinhard, E. & Kreis, W. Influence of precursors and inhibitors of the sterol pathway on sterol and cardenolide metabolism in *Digitalis lanata* EHRH. *Plant Physiol. Biochem.* **35**, 111–121 (1997).
20. Kreis, W. The foxgloves (*Digitalis*) revisited. *Planta Med.* **83**, 962–976 (2017).
21. Kreis, W., Hensel, A. & Stuhlemmer, U. Cardenolide biosynthesis in foxglove. *Planta Med.* **64**, 491–499 (1998).
22. Herl, V., Frankenstein, J., Meitingner, N., Müller-Uri, F. & Kreis, W. $\Delta 5$ - 3β -Hydroxysteroid dehydrogenase (3β HSD) from *Digitalis lanata*. Heterologous expression and characterisation of the recombinant enzyme. *Planta Med.* **73**, 704–710 (2007).
23. Herl, V. et al. Using progesterone 5β -reductase, a gene encoding a key enzyme in the cardenolide biosynthesis, to infer the phylogeny of the genus *Digitalis*. *Plant Syst. Evol.* **271**, 65–78 (2008).
24. Munkert, J., Ernst, M., Müller-Uri, F. & Kreis, W. Identification and stress-induced expression of three 3β -hydroxysteroid dehydrogenases from *Erysimum crepidifolium* Rchb. and their putative role in cardenolide biosynthesis. *Phytochemistry* **100**, 26–33 (2014).
25. Munkert, J. et al. Progesterone 5β -reductase genes of the Brassicaceae family as function-associated molecular markers. *Plant Biol.* **17**, 1113–1122 (2015).
26. Simpson, E. R. & Boyd, G. S. The cholesterol side chain-cleavage system of bovine adrenal cortex. *Eur. J. Biochem.* **2**, 275–285 (1967).
27. Pilgrim, H. 'Cholesterol side-chain cleaving enzyme' aktivität in keimlingen und *in vitro* kultivierten gewebe von *Digitalis purpurea*. *Phytochemistry* **27**, 1725–1728 (1972).
28. Lindemann, P. & Luckner, M. Biosynthesis of pregnane derivatives in somatic embryos of *Digitalis lanata*. *Phytochemistry* **46**, 507–513 (1997).
29. Ohnishi, T., Yokota, T. & Mizutani, M. Insights into the function and evolution of P450s in plant steroid metabolism. *Phytochemistry* **70**, 1918–1929 (2009).
30. Umamoto, N. et al. Two cytochrome P450 monooxygenases catalyze early hydroxylation steps in the potato steroid glycoalkaloid biosynthetic pathway. *Plant Physiol.* **171**, 2458–2467 (2016).
31. Nett, R. S., Lau, W. & Sattely, E. S. Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* **584**, 148–153 (2020).
32. Guengerich, F. P. & Yoshimoto, F. K. Formation and cleavage of C–C bonds by enzymatic oxidation–reduction reactions. *Chem. Rev.* **118**, 6573–6655 (2018).
33. Sonawane, P. D. et al. Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants* **3**, 16205 (2017).
34. Tschesche, R. & Kleff, U. Beiträge zur biochemischen 14β -hydroxylierung von C21-steroiden zu cardenoliden. *Phytochemistry* **12**, 2375–2380 (1973).
35. Souza, C. M. et al. A stable yeast strain efficiently producing cholesterol instead of ergosterol is functional for tryptophan uptake, but not weak organic acid resistance. *Metab. Eng.* **13**, 555–569 (2011).
36. Strushkevich, N. et al. Structural basis for pregnenolone biosynthesis by the mitochondrial monooxygenase system. *Proc. Natl Acad. Sci. USA* **108**, 10149–10143 (2011).
37. Mirdita, M. et al. CoLabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
38. Ghosh, S. Triterpene structural diversification by plant cytochrome P450 enzymes. *Front. Plant Sci.* **8**, 1886 (2017).
39. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
40. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
41. Raghavan, I., Gopal, B. R., Carroll, E. & Wang, Z. Q. Cardenolide increase in foxglove after 2,1,3-benzothiadiazole treatment reveals a potential link between cardenolide and phytosterol biosynthesis. *Plant Cell Physiol.* **64**, 107–116 (2022).
42. Sarrion-Perdigones, A. et al. GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* **162**, 1618–1631 (2013).
43. Wretensjö, I. & Karlberg, B. Characterization of sterols in borage oil by GC–MS. *J. Am. Chem. Oil Soc.* **79**, 1069–1074 (2004).
44. Yang, B., Karlsson, R. M., Oksman, P. H. & Kallio, H. P. Phytosterols in sea buckthorn (*Hippophae rhamnoides* L.) berries: identification and effects of different origins and harvesting times. *J. Agric. Food Chem.* **49**, 5620–5629 (2001).
45. Kamal-Eldin, A., Appelqvist, L. A., Yousif, G. & Iskander, G. M. Seed lipids of *Sesamum indicum* and related wild species in Sudan. The sterols. *J. Sci. Food Agric.* **59**, 327–334 (1992).
46. Clough, S. & Bent, A. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
47. Sievers, F. et al. Fast scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
48. Kumar, S., Stecher, G., Li, M., Nuyt, K. & Tamura, K. MEGA: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
49. Kochnev, Y., Hellemann, E., Cassidy, K. C. & Durrant, J. D. Webina: an open-source library and web app that runs AutoDock Vina entirely in the web browser. *Bioinformatics* **36**, 4513–4515 (2020).

Acknowledgements

We thank A. Aharoni and B. Negin for providing *C. procera* seeds, and D. Nelson for CYP nomenclature. We thank H. Reizman for providing RH6827 and RH6829 yeast strains for analysis. We acknowledge the Max Planck Society and the European Research Council (788301) for funding.

Author contributions

M.K. performed the research and wrote the paper. C.L. performed structural modelling and docking studies. K.P. assisted with the generation of transgenic plants. C.E.R.L. helped with bioinformatics analysis. D.A.S.G. assisted with metabolomics data analysis and

operated the LC–MS. E.R. assisted in growing and maintaining various plants. R.L. assisted with transient expression experiments and performed the yeast experiments and analysis. S.E.O.C. and P.D.S. designed the research and wrote the paper.

Funding

Open access funding provided by Max Planck Society

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/s41477-023-01515-9>.

Correspondence and requests for materials should be addressed to Sarah E. O'Connor or Prashant D. Sonawane.

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

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

All presented data have been acquired using existing and routinely used softwares. These were mentioned in the respective parts of manuscript. LC-MS data presented in this study was collected and analyzed using Thermo Scientific UltiMate 3000 ultra-high performance liquid chromatography (UHPLC) system coupled to an Impact II UHR-Q-ToF (Ultra-High Resolution Quadrupole-Time-of-Flight) mass spectrometer (Bruker Daltonics). GC-MS data was collected and analyzed using the GC-MS system comprised a GC PAL autosampler (CTC Analytics), a trace 1310 GC ultra-gas chromatograph equipped with a split-splitless injector and ISQ quadrupole MS (Thermo Scientific).

Data analysis

Bruker Compass Data Analysis (Version 5.3) and Xcalibur (Thermo Scientific version 4.2.47) softwares were used for LC-MS and GC-MS data analysis respectively. Microsoft Excel 2016 and GraphPad Prism 8 softwares were used for statistical analysis. Chemical structures were generated in ChemDraw Professional 17.1. Phylogenetic tree was constructed using MEGAX (version 10.2.2). Primer Blast (NCBI) was used for designing qRT-PCR primers. AlphaFold2 and Webina were used for protein modeling and docking studies.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data supporting the findings of this work are available within the paper and its Supplementary Information files. Source data are provided with this paper. D. purpurea and C. procera RNA-seq data associated with this manuscript have been deposited into the NCBI Sequence Read Archive with BioProject IDs PRJNA929980 and PRJNA1003839, respectively. The databases used in this study are Uniprot Swiss-Prot (release 2022_02) and the Pfam database (Pfam-A.hmm release 34.0). Correspondence and requests for materials should be addressed to P.D.S. or S.E.O'C.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="Not applicable to our study"/>
Population characteristics	<input type="text" value="Not applicable to our study"/>
Recruitment	<input type="text" value="Not applicable to our study"/>
Ethics oversight	<input type="text" value="Not applicable to our study"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="Statistical based sample size determination was not applicable to our study. Details of biological/technical replicates related to various experiments are provided in the paper, wherever necessary. Minimum of three biological replicates (except in some cases of transgenic analysis performed in T0 generation) were used while performing the experiments. Sample size information for each experiment is present in the paper."/>
Data exclusions	<input type="text" value="No data was excluded in our analysis"/>
Replication	<input type="text" value="Details of biological/technical replicates used in various experiments are provided in methods section as well as in Main Figures and Supplementary Figures legends, wherever necessary."/>
Randomization	<input type="text" value="Minimum three independently grown N. benthamiana, D. purpurea and C. procera plants were chosen randomly for experiments and analysis. For yeast expression studies, positive transformants were selected based on colony PCR experiments for each characterized gene."/>
Blinding	<input type="text" value="Not applicable. Same treatment/parameter was used while comparing control/wild type versus treated/transgenic samples in experiments"/>

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | n/a | Included in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | n/a | Included in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |