Few complete pathways have been established for the biosynthesis of medicinal compounds from plants. Accordingly, many plant-derived therapeutics are isolated directly from medicinal plants or plant cell culture. A lead example is colchicine, a US Food and Drug Administration (FDA)-approved treatment for inflammatory disorders that is sourced from Colchicum and Gloriosa species. Here we use a combination of transcriptomics, metabolic logic and pathway reconstitution to elucidate a near-complete biosynthetic pathway to colchicine without prior knowledge of biosynthetic genes, a sequenced genome or genetic tools in the native host. We uncovered eight genes from Gloriosa superba for the biosynthesis of N-formyldemecolcine, a colchicine precursor that contains the characteristic tropolone ring and pharmacophore of colchicine. Notably, we identified a non-canonical cytochrome P450 that catalyses the remarkable ring expansion reaction that is required to produce the distinct carbon scaffold of colchicine. We further used the newly identified genes to engineer a biosynthetic pathway (comprising 16 enzymes in total) to N-formyldemecolcine in Nicotiana benthamiana starting from the amino acids phenylalanine and tyrosine. This study establishes a metabolic route to tropolone-containing colchicine alkaloids and provides insights into the unique chemistry that plants use to generate complex, bioactive metabolites from simple amino acids.

Medicinal plants are a major source of bioactive small molecules and serve as potent treatments for human disease. A prominent example is the historical usage of Colchicum and Gloriosa plant species, which have been used to treat inflammatory disorders for centuries, and perhaps millennia. The anti-inflammatory action of these plants is due to their production of colchicine, an alkaloid that is approved by the FDA as a pharmaceutical treatment for acute cases of gout and familial Mediterranean fever, and which has also been used to treat pericarditis and Behçet’s disease. The bioactivity of colchicine is thought to result from its interaction with tubulin dimers and inhibition of microtubule growth, although it is not entirely clear how its effect on microtubule dynamics leads to a decrease in inflammation. Notably, together with the taxane and Vinca domains, the colchicine-binding domain on tubulin is a target for a number of small-molecule chemotherapeutics. Colchicine itself is too potent to use in chemotherapy and even for gout treatment, doses of colchicine must be strictly administered to avoid toxicity. However, the anti-mitotic action induced by the interaction of colchicine with tubulin has made colchicine an important research tool; it has been used to identify the number of human chromosomes, study microtubule dynamics and induce polyploidy in plants.

Our understanding of colchicine biosynthesis is based on a multitude of feeding studies that used isotope-labelled substrates in Colchicum plants, as well as on the structural characterization of colchicine-related alkaloids isolated from species of the Colchicaceae family. Collectively, these studies have led to the development of a well-defined biosynthetic hypothesis in which the alkaloid core is built from the amino acids l-phenylalanine and l-tyrosine (Fig. 1; see Supplementary Scheme 1 for full biosynthetic scheme). Phenylalanine and tyrosine are predicted to be processed to 4-hydroxydihydrocinnamaldehyde (4-HDCA) and dopamine, respectively, which are putatively joined through a Pictet–Spengler reaction to form a 1-phenethylisoquinoline scaffold (1). A series of methylations and phenyl ring hydroxylations of this scaffold produces 16 enzymes in total) to N-formyldemecolcine in Nicotiana benthamiana starting from the amino acids phenylalanine and tyrosine. This study establishes a metabolic route to tropolone-containing colchicine alkaloids and provides insights into the unique chemistry that plants use to generate complex, bioactive metabolites from simple amino acids.

Metabolomic and transcriptomic analysis

Biosynthetic pathway genes in plants are often coordinately regulated, a phenomenon that can be leveraged by RNA sequencing (RNA-seq) and de novo transcriptome assembly to compile genes into co-expressed functional units. This approach is most useful when a dedicated pathway gene can be used as a bait to identify candidate genes through co-expression analysis; however, this requires prior knowledge of a gene in the pathway. Furthermore, co-expression patterns may only be apparent after RNA-seq analysis if the spatial, temporal or condition-specific
production of the compound of interest is known and captured in the dataset. Although colchicine alkaloids are found throughout the Colchicum plant, they accumulate to the highest concentrations in the seeds and corms (or in Gloriosa, in the rhizomes). In addition, most of the aforementioned isotope-labelling studies were accomplished by feeding substrates into either the corms or the seed pods of Colchicum plants. Together, these data suggest that the expression of colchicine biosynthetic genes is enriched in these tissues.

At the start of this study, two RNA-seq datasets for colchicine-producing plants (https://medplantrnaseq.org/) were publicly available: one from Colchicum autumnale (autumn crocus) and a second from G. superba (flame lily). We also generated a third RNA-seq dataset that contained multiple replicates of different tissue types (leaf, stem, rhizome and root) from 5-week-old G. superba plants. Notably, accompanying metabolite profiles were also obtained for each tissue to compare alkaloid accumulation with transcript expression profiles. We observed notable differences in colchicine accumulation between rhizome, stem, root and leaf tissues; rhizomes accumulated the highest levels of colchicine and related metabolites (Supplementary Fig. 1).

Discovery of colchicine biosynthetic enzymes

To identify an initial enzyme involved in the colchicine pathway, we first focused on S-adenosylmethionine-dependent methyltransferase enzymes that potentially process the predicted 1-phenethylisoquinoline precursor (1), as labelling studies have indicated that this substrate is initially functionalized with O- and N-methylations. We identified 11 candidate methyltransferase genes in the public G. superba transcriptome that exhibited relatively high expression across any tissue type (Fig. 2). Each methyltransferase was cloned from G. superba CDNA and transiently expressed in N. benthamiana leaves using Agrobacterium-mediated transformation. Crude protein lysates from these leaves were then assayed for in vitro methyltransferase activity with chemically synthesized 1 ([M + H]⁺ = 286,1438) as substrate. Liquid chromatography–mass spectrometry (LC–MS) analysis revealed that one candidate methyltransferase catalysed the S-adenosylmethionine-dependent consumption of 1 and the formation of a methylated compound ([M + H]⁺ = 300,1594 Da) (Extended Data Fig. 1). Further MS/MS analysis led us to assign the structure as the O-methylated product of colchicine (2) in the seeds and corms (or in Gloriosa, in the rhizomes). As such, this O-methyltransferase gene is referred to as GsOMT1.

We noted that the expression of GsOMT1 and GsNMT is highly correlated in the publically available G. superba transcriptome (Pearson’s r > 0.99), which indicates that co-expression analysis could help to prioritize additional candidate enzymes. Of the 11 cloned methyltransferase genes, only 5 (including GsNMT) showed relatively high co-expression (Pearson’s r > 0.5) with GsOMT1 (Fig. 2), suggesting that these methyltransferases may also function in colchicine biosynthesis. Similarly, seven unique cytochrome P450 genes—a class of enzymes implicated in multiple downstream biosynthetic transformations (for
example, hydroxylations, phenol coupling and ring expansion)\(^{31-40}\) – also co-express strongly with \(\text{GsOMT1}\) (Pearson’s \(r > 0.9\) (Fig. 2), and were therefore cloned for transient expression. These candidate methyltransferase and cytochrome P450 genes were each tested through transient co-expression in \(N.\ benthamiana\) with previously identified pathway genes and co-infiltration of 1 as substrate.

Co-expression of one cloned cytochrome P450 (\(\text{GsCYP75A109}\)) in \(N.\ benthamiana\) leaves resulted in partial consumption of 3 and the production of a new compound with a mass that corresponded to a single hydroxylation ([\(M + H\]) = 330.1700 Da) (Extended Data Fig. 3). On the basis of MS/MS analysis, this compound was assigned the structure of 4, in which the new hydroxyl group is at the \(\alpha\) position on ring \(A\).\(^{36-39}\) Relatively low production of 4 led us to examine strategies to improve flux through this system. We identified a putative N-terminal mitochondria-localization signal in the primary amino acid sequence of \(\text{GsNMT}\), and removal of this sequence led to an approximately tenfold increase in the yield of 4 (Extended Data Fig. 3). Therefore, this truncated enzyme (\(\text{NsNMT}\)) was used for all subsequent experiments.

Biosynthesis is next predicted to proceed through the methylation of the \(\alpha\) hydroxyl group on ring \(A\)\(^{36-39}\), and we found that co-expression of another candidate methyltransferase (\(\text{GsOMT2}\)) resulted in consumption of 4. However, the new compound had a mass consistent with the addition of both a methyl and a hydroxyl group ([\(M + H\]) = 360.1805 Da), which MS/MS fragmentation indicated to be on ring \(A\), leading to the assignment of this compound as 6 (Extended Data Fig. 4). One explanation for this result is that \(\text{GsOMT2}\) acts as a \(\alpha\)-O-methyltransferase to produce 5, a confirmed precursor to colchicine\(^{31-35}\), and that \(\text{GsCYP75A109}\) can subsequently install a second hydroxyl group in the other free \(\alpha\) position on ring \(A\) to generate 6. Although we did not observe accumulation of 5, which corresponds to the methylation alone, it is possible that 5 is immediately consumed through the activity of \(\text{GsCYP75A109}\). Alternatively, we cannot exclude the possibility that an endogenous \(N.\ benthamiana\) hydroxylase catalyses this reaction in our heterologous expression system.

Addition of another candidate methyltransferase (\(\text{GsOMT3}\)) to the established pathway resulted in consumption of 6 and formation of a methylated product ([\(M + H\]) = 374.1962 Da) that was confirmed to be autunnaline (7) through LC–MS comparison to a racemic standard\(^{31,39}\) of 7 (Extended Data Fig. 5). This was the first intermediate accessed in this biosynthetic reconstitution that represented an isolated alkaloid from colchicine-producing plants\(^{39}\), and indicated that the pathway that we were assembling in \(N.\ benthamiana\) was probably en route to colchicine. Although our result did not indicate the absolute stereochemistry of 7, (S)-autunnaline has been shown to be the substrate for the subsequent para-para phenol coupling reaction that is catalysed by a cytochrome P450\(^{36,40}\), and we therefore predict that our product is likely to be the (S) enantiomer.

Co-expression of a second cytochrome P450 candidate (\(\text{GsCYP75A110}\)) in this transient expression system led to consumption of 7 and formation of three compound peaks (Extended Data Fig. 6), each with a mass that indicated the loss of two hydrogens ([\(M + H\]) = 372.1805 Da). Correspondingly, further characterization of \(\text{GsCYP75A110}\) expressed alone in both \(N.\ benthamiana\) and yeast demonstrated similar activity. The exact mass of these products matches that of isoaurodbine (8), a confirmed precursor to colchicine\(^{36}\), which is the proposed product from the oxidative para-para phenol coupling of 7 (Extended Data Fig. 6). One of the three compounds produced by \(\text{GsCYP75A110}\) is consumed after the addition of another candidate methyltransferase (\(\text{GsOMT4}\)) to the co-expression system. This activity yielded O-methylandrocymbine (9), ([\(M + H\]) = 386.1962 Da), a known precursor to colchicine\(^{36,37}\), as confirmed by comparison to an authentic standard (Extended Data Fig. 7). Thus, we further tentatively assign the consumed substrate as 8, which matches previously established biosynthetic hypotheses\(^{39,40}\).

**Fig. 3** Combined transcriptomics and metabolomics identify notable co-expression of colchicine biosynthetic genes in \(G.\ superba\).

**Fig. 3** Combined transcriptomics and metabolomics identify notable co-expression of colchicine biosynthetic genes in \(G.\ superba\). a. Tissues from \(G.\ superba\) plants (leaf, stem, rhizome and root) were used to quantify colchicine alkaloid accumulation and to isolate RNA for subsequent RNA-seq analysis. b. Colchicine accumulates in all tissues, but to the highest level in the rhizome, suggesting that the rhizome is the most active site of biosynthesis. The extracted ion abundance of colchicine, m/z 400.1755 ± 20 ppm, is shown. n = 7 independent biological replicates for each tissue type. The centre line indicates the median; box limits indicate upper and lower quartiles; whiskers indicate 1.5× the interquartile range; points outside of the whiskers indicate outliers. c. Hierarchical clustering analysis (distance metric: uncentred Pearson correlation) was performed on contigs with a trimmed mean of \(M\) (TMM)-normalized, counts per million (CPM) expression value greater than 25 (for a total of 11,315 out of 38,466 total contigs compared across 11 tissue samples). This analysis identifies an 89-contig cluster that is populated with a substantial number of colchicine biosynthetic genes, as shown here, which indicates a high level of co-expression.

At this point, we re-assessed the co-expression patterns of the identified biosynthetic genes. Hierarchical clustering of contigs from our in-house-generated \(G.\ superba\) RNA-seq dataset, which had only been filtered by minimum expression, revealed that the seven genes that we demonstrate to process 1 into 9 are present in a single cluster of 89 co-expressed transcripts (Fig. 3 and Supplementary Fig. 2). Notably, this represents only 0.23% of the total contigs in our RNA-seq dataset, suggesting tightly co-regulated expression of colchicine biosynthetic genes.

Expansion of the dienone ring C′ of 9, which ultimately results in the formation of a tropolone ring (ring C), has been suggested to be catalysed by a cytochrome P450 enzyme\(^{32}\). Although the initial cytochrome P450 candidates from the \(G.\ superba\) transcriptomes did not catalyse this reaction, continued analysis of the public \(C.\ autumnale\) dataset (see Supplementary Information for a detailed description of how this cytochrome P450 was identified) revealed a previously overlooked cytochrome P450 (\(\text{CYP71FB1}\)) that shared relatively high co-expression (Pearson’s \(r > 0.8\)) with the other identified biosynthetic genes in our \(G.\ superba\) expression data (Supplementary Figs. 3, 4). Addition of \(\text{CYP71FB1}\) to the transient expression system resulted in consumption of 9 and accumulation of the tropolone-containing compound \(N\)-formyldecolcine (10), ([\(M + H\]) = 400.1755 Da), the proposed product of the ring expansion reaction\(^{12}\), as confirmed by LC–MS comparison to an authentic standard (Extended Data Fig. 8). Additional characterization of \(\text{CYP71FB1}\) expressed individually in...
both *N. benthamiana* and yeast confirmed that this enzyme is responsible for the ring expansion and rearrangement reaction, which involves breaking and reformation of two C–C bonds, a noteworthy cascade reaction promoted by a single oxidative enzyme.

In addition to the authentic standards used to confirm the identities of 7, 9 and 10, we also established that all other proposed intermediates produced heterologously in *N. benthamiana* (2, 3, 4, 6 and 8) co-elute with metabolites found in *G. superba* rhizomes that share identical MS/MS spectra (Extended Data Fig. 9). This suggests that the metabolites produced in our transient expression system represent biologically relevant pathway intermediates. Taken together, the eight enzymes identified in this study from the medicinal plant *G. superba* establish a biosynthetic route from a phenethylisoquinoline substrate (1) to *N*-formyldemecolcine (10) (Fig. 4). Our data corroborate the proposed biosynthetic pathway that had previously been established through many years of rigorous isolate-labelling studies and metabolite isolation from colchicine-producing plants\(^4^6\), and further demonstrate how plants synthesize the tropolone-containing scaffold of colchicine.

**Engineering colchicine alkaloid biosynthesis**

To eliminate the need for a synthetic precursor, we next developed a metabolic engineering strategy to produce 1 in *N. benthamiana*. Production of 1 tentatively requires the Pictet–Spengler condensation of 4-HDCA and dopamine, which are derived from the amino acids phenylalanine and tyrosine, respectively. Labelling studies have indicated that 4-HDCA is produced from phenylalanine through a metabolic route analogous to the biosynthesis of monolignols\(^4^3\),\(^2^3\),\(^2^7\),\(^2^8\) (Supplementary Scheme 3), which are building blocks of the lignin polymers that are ubiquitous to vascular plants\(^4^5\). The only major distinction in 4-HDCA biosynthesis is the reduction of the olefinic (alkenal) double bond. Notably, hierarchical clustering analysis of our *G. superba* transcriptomic data demonstrated co-clustering of many monolignol biosynthetic gene orthologues, as well as a putative alkenal reductase (*GsOMT4*), with the previously characterized methyltransferases and cytochrome P450 genes\(^4^4\) (Fig. 3 and Supplementary Fig. 2). We hypothesized that a co-expressed group of these biosynthetic genes (*GsPAL*, *Gs4CL*, *GsCCR*, *GsAER*, *GsC4H* and *GsDAHP*; comprising module 1) could produce sufficient quantities of 4-HDCA for engineered biosynthesis (Supplementary Scheme 3). Indeed, heterologous co-expression of module 1 in *N. benthamiana* resulted in the production of 4-HDCA, which was otherwise not detected (Extended Data Fig. 10).

For dopamine formation, the incorporation of tyrosine and tyramine into colchicine\(^4^3\),\(^4^6\) suggests the activity of L-tyrosine/L-DOPA decarboxylase (TyDC/DDC) and 3′-hydroxylase enzymes (Supplementary Scheme 1). In support of this, a TyDC/DDC homologue (*GsTyDC/DDC*) was found to be highly co-expressed with the other colchicine biosynthetic genes, as shown in our hierarchical clustering analysis (Fig. 3). To circumvent discovery of a novel 3′-hydroxylase, we used a cytochrome P450 enzyme from *Beta vulgaris* (*BvCYP76ADS*) that hydroxylates tyrosine to form L-DOPA in betalain pigment biosynthesis\(^4^6\). Accordingly, co-expression of *BvCYP76ADS* and *GsTyDC/DDC* (module 2) in *N. benthamiana* led to the production of dopamine (Extended Data Fig. 10).

Although Pictet–Spengler condensation of 4-HDCA and dopamine to produce 1 could occur non-enzymatically\(^4^6\), heterologous co-expression of module 1 and module 2 did not result in accumulation of 1, suggesting that a dedicated ‘Pictet–Spenglerase’ enzyme is required (Extended Data Fig. 10). Given that we could not identify an obvious Pictet–Spenglerase candidate in our dataset, we attempted to use (S)-norcochlorine synthase (NCS), a previously characterized plant Pictet–Spenglerase. NCS condenses 4-hydroxyphenylacetaldehyde (4HPAA) and dopamine in benzylisoquinoline alkaloid biosynthesis\(^4^6\) (Extended Data Fig. 10), and has been shown to condense a wide range

---

**Fig. 4 | Discovery of a pathway for colchicine alkaloid biosynthesis.**

**a**. Transient co-expression of eight identified biosynthetic genes from *G. superba* in *N. benthamiana* enables the step-by-step conversion of a co-infiltrated 1-phenethylisoquinoline substrate (1) into the tropolone-containing alkaloid *N*-formyldemecolcine (10) through the pathway shown here. **b.** Grey boxes to the left of the graphs indicate biosynthetic genes that are included in a co-expression experiment; red box indicates the final acting enzyme in a set of co-expressed genes. For each intermediate, data are mean ± s.d. of the extracted ion abundance (*n* = 6) for the exact ion mass [M + H]\(^+\) (for 10, both [M + H]\(^+\) and [M + Na]\(^+\]) that corresponds to each compound.
of aldehyde substrates with dopamine. Notably, the native NCS substrate (4HPAA) is nearly identical in structure to 4-HDCA, suggesting a potential for NCS to use 4-HDCA. Indeed, heterologous co-expression of an aldehyde substrate (4HPAA) is nearly identical in structure to 4-HDCA, suggesting a potential for NCS to use 4-HDCA. Notably, the native NCS enzyme can catalyze the formation of the (S) enantiomer of 1, these data further indicate that the 1-phenethylisoquinoline intermediates between 1 and 7 that are produced in this system are probably the (S) enantiomers. Notably, the ability of the module-3 enzymes to use these enantiomers aligns with the previous determination of 1-phenethylisoquinoline stereochemistry in colchicine biosynthesis.

We further evaluated whether the pathway exists as a metabolic network, as opposed to a strictly linear route, by performing drop-out experiments in which each individual enzyme from module 3 was removed from the engineered pathway to 10. In these experiments, we primarily observed accumulation of the proposed pathway intermediates illustrated in Fig. 4, and did not detect mass signatures that corresponded to alternative pathway intermediates (Extended Data Fig. 12 and Supplementary Fig. 5). The only exception identified was the order of the initial O- and N-methylations; we found that GSNT activity can precede that of GSOMT1, but that initial activity by GSOMT1 is probably the major route (Extended Data Fig. 12).

In a similar dropout experiment, we determined that nearly all module-1 and module-2 genes contribute significantly to the observed yield of colchicine alkaloids produced heterologously in N. benthamiana (Extended Data Fig. 11). The only exception was Gs4CL, which did not appear to affect product accumulation. Although orthologous N. benthamiana enzymes may contribute to some of the early biosynthetic transformations, inclusion of all described enzymes except Gs4CL (comprising 16 enzymes in total) is necessary for the highest observed biosynthetic capacity (Fig. 5 and Extended Data Fig. 11). Although the module-1 and module-2 genes do not necessarily reflect the native route of colchicine biosynthesis, they do represent a convenient and logical strategy for producing 4-HDCA and dopamine.

**Discussion**

Future investigations are still needed to clarify the final metabolic steps in colchicine biosynthesis. In particular, previous investigations have indicated that N-deformylase, N-demethylase and N-acetyltransferase enzymes are probably required for the conversion of 10 into colchicine. Elucidating these three final biosynthetic transformations, as well as the remaining unidentified native enzymes in early colchicine biosynthesis, will help to finalize our understanding of how this alkaloid is produced in nature.

Through our combined use of metabolomics, transcriptomics and heterologous expression in N. benthamiana, we have been able to rapidly establish a near-complete reconstitution of the complex biosynthetic pathway of colchicine, a plant-derived drug of historical and contemporary importance. By combining biosynthetic genes from the medicinal plant G. superba with enzymes co-opted from other pathways (comprising 16 genes in total), we demonstrate de novo biosynthetic production of N-formyldemecolcine (10) in this commonly used model plant. Ultimately, our results not only provide a metabolic route to the hallmark tropolone scaffold of colchicine, but also highlight a powerful
Methods

Data reporting
No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Chemicals and reagents
Unless stated otherwise, general chemicals and reagents, including several colchicine alkaloids (demecolcine (11), N-deacetylcolchicine (12) and colchicine (13)) were purchased commercially. Authentic standards of (R,S)-autumnale (7) and O-methylandrocymbine (9) were provided from the Kuchan Lab Natural Product Collection (T. Kuchan, Danforth Center).

Plant growth
The _N. benthamiana_ plants were sown in PRO MIX HP Mycorrhizae soil (Premier Tech Horticulture) and maintained on growth shelves with a 16-h light–8-h dark cycle at room temperature. Plants were grown for 4–5 weeks before _Agrobacterium_ infiltration, with periodic watering, as needed. The _G. superba_ seeds were obtained from eBay (seller, therealflorida101). For germination, seeds were soaked in warm water at 30 °C overnight, and then planted in PRO MIX HP Mycorrhizae soil. Plants were grown on a growth shelf with a 16-h light–8-h dark cycle at ambient laboratory temperature with periodic watering, as needed. Plants were harvested 5 weeks after germination to assess the accumulation of putative colchicine biosynthetic intermediates in the leaf, stem, root and rhizome tissues and to extract total RNA from these tissues. Tissues were immediately snap-frozen in liquid nitrogen after removal from the plant and stored at −80 °C for future use.

RNA isolation, library preparation and RNA-seq
Liquid-nitrogen-frozen _G. superba_ tissues were homogenized using a chilled mortar and pestle, and total RNA was extracted using the Spectrum Plant Total RNA Kit (Sigma-Aldrich), according to the manufacturer’s instructions. For cloning of gene candidates, cDNA was prepared from extracted mRNA using the SuperScript III First-Strand Synthesis system (Invitrogen). RNA samples intended for RNA-seq analysis were assessed using an RNA 6000 Nano chip on a 2100 Bioanalyzer (Agilent) to assess RNA quality and quantity. RNA-seq libraries for leaf (2 samples), stem (2 samples), root (2 samples), and rhizome (5 samples) tissues were prepared from high-quality RNA using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) by following the manufacturer’s instructions. The quality and average length of each library (insert size distribution was around 500–1,000 bp) was assessed using a High Sensitivity DNA Chip on a 2100 Bioanalyzer (Agilent). Next-generation sequencing (paired-end, 2 × 300 bp) was performed using single lane on the Illumina MiSeq system at the Stanford Center for Genomics and Personalized Medicine (SCGPM).

RNA-seq data analysis
Transcriptomic data mining and analysis of the Medicinal Plants Consortium datasets. RNA-seq data from _G. superba_ and _C. autumnale_ were downloaded from the Medicinal Plants Consortium database (https://www.medplantrnaseq.org/); the following files were downloaded: Gloriosa_superba.tar.gz, Colchicum_ autumnale.tar.gz). For analysis of the public _G. superba_ data, the previously de novo assembled transcriptome (file: contigs.fa) and associated predicted peptides (file: peptides.fa) were used as provided. The number of fragments per kilobase of contig per million mapped reads (FPKM) for each contig was determined using the number of paired aligned reads provided in the associated ‘readcounts’ folder (four libraries total; two different count files for each library; no description of tissues or conditions). Pfam annotations for each protein-coding contig were provided in the associated ‘annot’ folder. All contigs were annotated with the best BLAST hit (BLASTX) from the _Arabidopsis thaliana_ proteome (The Arabidopsis Information Resource (TAIR), https://www.arabidopsis.org/). Select contigs were further annotated with the best BLASTX hit from the non-redundant Reference Sequence (RefSeq) database available from the National Center for Biotechnology Information (NCBI) website.

The public _C. autumnale_ RNA-seq data contained raw reads from corn (also referred to as a bulb), fruit and leaf tissues (each with single replicates), along with a previously de novo assembled transcriptome (file: contigs.fa) and associated peptides (file: peptides.fa). The provided reads were used to quantify contig abundance for this transcriptome. Read quality for each library was assessed using FastQC (Babraham Bioinformatics, Babraham Institute), after which they were trimmed accordingly using trimmomatic with the following parameters: <ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:HEADCROP:14 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:5 MINLEN:20>. Reads were mapped to the _C. autumnale_ transcriptome using Bowtie2 and quantified with eXpress. The total counts generated by eXpress were normalized using the edgeR trimmed mean of _M_(TMM) method and then transformed into counts per million (CPM) values, which were log twenty-transformed for downstream analyses. Contigs were annotated using Pfam protein family searches and with the best BLAST hit (BLASTX) from the _A. thaliana_ proteome.

Transcriptome assembly and analysis of newly acquired _G. superba_ RNA-seq data. The quality of the raw reads generated from our Illumina MiSeq sequencing of _G. superba_ tissue samples was first assessed with FastQC, then trimmed using trimmomatic with the following parameters: <ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:HEADCROP:20 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:10 MINLEN:50>. The trimmed reads were used to assemble individual transcriptomes de novo using Velvet and Oases with _k_ _mer_ sizes of 101–201 at increments of 10. The clustering tool CD-HIT-EST was used to identify sequences with greater than 99% identity (word size = 11) for each _k_ _mer_-assembly and only the longest representative transcript was kept for further processing and analysis. All of the individual _k_ _mer_ transcriptomes were then combined and clustered further using CD-HIT-EST at a threshold of 99% identity (word size = 11). The combined set of transcripts was further assembled with CAP3 to combine contigs with significant overlaps (minimum 95% identity over at least 100 bp), which resulted in a final working assembly of 38,461 contigs. To provide a unique transcript ID, each contig was annotated with ‘Gsup’ followed by a unique number (Gsup1, Gsup2, and so on). The contigs were then further annotated using sequences against the Pfam database and by BLASTX/BLASTP searches using the _Arabidopsis_ proteome as a reference. In addition, BLASTX and BLASTP searches against the NCBI non-redundant protein database were performed as necessary. The best BLASTP hits from the non-redundant protein database and from the _A. thaliana_ proteome for each _G. superba_ gene that was characterized or used in this manuscript are listed in Supplementary Tables 1 and 2, respectively. The best BLASTP hits from our _G. superba_ transcriptome for each protein characterized in this study are listed in Supplementary Table 3 (these are defined as the closest BLAST hits that do not represent the query gene itself; general, the closest hit with >90% identity is listed). The TargetP1.1 Server was used to predict the presence of transit peptides in the predicted proteins from this transcriptome, as well as from those in the public _Gloriosa_ and _Colchicum_ transcriptomes (Supplementary Table 4).

Trimmed reads for each library were mapped to the assembled transcriptome using Bowtie2 and expression quantification was performed using eXpress. The total counts generated by eXpress were normalized using the TMM method in edgeR, and the resulting values were used to calculate CPM values, which were log twenty-transformed for downstream analyses. It was discovered post hoc that the original transcriptome assembly contained contigs with multiple, distinct coding sequences,
method and plated on LB agar (50 μg/ml kanamycin and 30 μg/ml gentamycin) for selection of positive transformants. After 2 days of growth, positive colonies were identified by colony PCR. Positive transformants were grown in liquid LB medium (50 μg/ml kanamycin and 30 μg/ml gentamycin) at 30 °C on a culture rotary drum for 2 days, after which 25% glycerol stocks were prepared and stored at −80 °C for future use.

For heterologous expression in Saccharomyces cerevisiae (yeast), coding sequences for GsCYP75A110 and GsCYP71F1B1 were PCR amplified from pEAQ-HT constructs with primers that contained overlaps for assembly into the yeast expression plasmid pYeDP60 (Carb² for E. coli selection; ADE2 for yeast selection). Before assembly, pYeDP60 was digested with EcoRI and BamHI restriction enzymes at 37 °C for 2 h, after which the reaction was terminated by heating at 65 °C for 30 min. Agarose-gel-purified PCR amplicons for the genes of interest were inserted into the digested pYeDP60 plasmid using an isothermal DNA assembly reaction, as described above. The assembly mixture was then used to transform NEB10β cells (New England Biolabs), which were plated for selection on LB agar (100 μg/ml carbenicillin). Positive transformants were determined by colony PCR as described above, and grown overnight in liquid LB medium (100 μg/ml carbenicillin) on a culture rotary at 37 °C. Plasmid DNA was isolated as described above and constructs were assessed with Sanger sequencing to confirm the sequence of the inserted gene.

**Transient expression of candidate genes in N. benthamiana**
Agrobacterium strains that contained the gene constructs of interest were streaked individually on LB plates (50 μg/ml kanamycin and 30 μg/ml gentamycin) at 30 °C for 2 days, at which point a lawn of cells had developed. Bacterial colonies were removed from the plate with a 1-ml pipette tip or sterile inoculating loop and resuspended in 1 ml of LB medium. This mixture was centrifuged at 5,000g for 5 min to pellet the cells, after which the supernatant was removed. The cell pellet for each strain was resuspended in Agrobacterium induction buffer (10 mM MES, pH 5.6, 10 mM MgCl₂, 150 μM acetoxyringone) and incubated at room temperature for 1–2 h. For individually tested strains, Agrobacterium suspensions were diluted to an optical density at 600 nm (OD₆₀₀) of 0.6. When strains were used in combination, the Agrobacterium suspensions for each strain were diluted to OD₆₀₀ = 0.3 and were mixed in equal concentration (according to the experiment in question). Prepared Agrobacterium suspensions were infiltrated into the abaxial side of N. benthamiana leaves of 4–5 week-old plants with a needleless 1-ml syringe. Unless stated otherwise, each experimental condition being tested consisted of three replicates, with each replicate corresponding to an independently infiltrated leaf. Each leaf among these three replicates was selected from a different N. benthamiana plant, such that no replicates for one experimental condition were taken from the same plant. In general, infiltrated leaves were harvested 4–5 days after infiltration, snap-frozen in liquid nitrogen and stored at −80 °C for downstream processing.

For substrate co-infiltration studies, 50–250 μM of 1 in water with 0.1–0.5% DMSO (1 was prepared as a 50 mM stock in DMSO) was infiltrated into the abaxial side of previously Agrobacterium-infiltrated leaves with a needleless 1-ml syringe 4 days after infiltration. For co-infiltration of (R:S)-α-tubulin (7) and O-methyllycanthin (9) as substrates, 25 μM of substrate in water with 1% methanol (7 and 9 were prepared as 2.5 mM stock solutions in methanol) was used. Substrate infiltration was no more difficult than Agrobacterium strain infiltration. Leaves were harvested 1 day later, snap-frozen in liquid nitrogen and stored at −80 °C for later processing.

**In vitro assays with N. benthamiana crude leaf extracts**
For in vitro experiments, leaves were infiltrated with Agrobacterium strains that contained the gene construct of interest, as described above. After 4 days, the leaves were excised and snap-frozen in liquid nitrogen. Approximately 1/3 of the infiltrated N. benthamiana leaf was

**Cloning of candidate biosynthetic genes**
Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) or Q5 High-Fidelity DNA polymerase (New England Biolabs) was used for all PCR amplification steps according to the manufacturer’s instructions. Oligonucleotide primers were purchased from Integrated DNA Technologies (IDT). In general, the open-reading frames for candidate genes were cloned from G. superba rhizome cDNA. Native-sequence, full-length clones for GsCYP71F1B1 (Gsp17435-b) and CjNCS (GenBank accession: AB267399.2) were synthesized as gBlocks Gene Fragments (IDT). A clone of the BvCYP76AD5 coding sequence (GenBank accession: KMS92961.1) was provided by A. Lloyd (The University of Texas at Austin). For all cloned genes, overlaps with homology to the appropriate plasmid vector were included on the 5’ and 3’ end of each clone for subsequent Gibson assembly reactions. For PCR-amplified genes, this was accomplished by adding the homology regions to the corresponding oligonucleotide primers used for amplification (see Supplementary Table 5 for a list of all primer sequences used in this study). Synthesized genes were ordered with the corresponding 5’ and 3’ homology regions already included in the sequence to be synthesized. Following PCR amplification of candidate gene sequences, DNA products were analysed on 1% agarose gels, excised and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research).

For use in DNA assembly reactions, the pEAQ-HT plasmid (Kan²) was digested with AgeI and XhoI restriction enzymes (New England Biolabs). PCR amplicons were inserted using an isothermal DNA assembly reaction using NEBuilder HiFi DNA Assembly Mix (New England Biolabs). PCR amplicons were inserted into the digested pYeDP60 plasmid using an isothermal DNA assembly reaction, as described above. The assembly mixture was then used to transform NEB10β cells (New England Biolabs), as per the manufacturer’s instructions. The assembly reaction mixtures were used directly to transform Escherichia coli TOP10 or NEB 5α cells (New England Biolabs), which were plated for selection on LB agar (50 μg/ml kanamycin) at 37 °C. Positive transformants were determined by colony PCR to confirm the presence of the inserted DNA fragment and were then grown overnight in liquid LB medium (50 μg/ml kanamycin) on a culture rotary drum at 37 °C. Plasmid DNA was isolated from E. coli cultures using the QIAprep Spin Miniprep kit (Qiagen) or the ZR Plasmid Miniprep kit (Zymo Research). Isolated plasmids that contained the desired insert were assessed with Sanger DNA sequencing (Elium Biopharm) to confirm the sequence of the clone of interest.

Confirmed pEAQ-HT expression constructs were transformed into Agrobacterium tumefaciens (GV3101:pMP90) using the freeze–thaw method and plated on LB agar (50 μg/ml kanamycin and 30 μg/ml gentamycin) for selection of positive transformants. After 2 days of growth, positive colonies were identified by colony PCR. Positive transformants were grown in liquid LB medium (50 μg/ml kanamycin and 30 μg/ml gentamycin) at 30 °C on a culture rotary drum for 2 days, after which 25% glycerol stocks were prepared and stored at −80 °C for future use.

For heterologous expression in Saccharomyces cerevisiae (yeast), coding sequences for GsCYP75A110 and GsCYP71F1B1 were PCR amplified from pEAQ-HT constructs with primers that contained overlaps for assembly into the yeast expression plasmid pYeDP60 (Carb² for E. coli selection; ADE2 for yeast selection). Before assembly, pYeDP60 was digested with EcoRI and BamHI restriction enzymes at 37 °C for 2 h, after which the reaction was terminated by heating at 65 °C for 30 min. Agarose-gel-purified PCR amplicons for the genes of interest were inserted into the digested pYeDP60 plasmid using an isothermal DNA assembly reaction, as described above. The assembly mixture was then used to transform NEB10β cells (New England Biolabs), which were plated for selection on LB agar (100 μg/ml carbenicillin). Positive transformants were determined by colony PCR as described above, and grown overnight in liquid LB medium (100 μg/ml carbenicillin) on a culture rotary at 37 °C. Plasmid DNA was isolated as described above and constructs were assessed with Sanger sequencing to confirm the sequence of the inserted gene.

**Transient expression of candidate genes in N. benthamiana**
Agrobacterium strains that contained the gene constructs of interest were streaked individually on LB plates (50 μg/ml kanamycin and 30 μg/ml gentamycin) at 30 °C for 2 days, at which point a lawn of cells had developed. Bacterial colonies were removed from the plate with a 1-ml pipette tip or sterile inoculating loop and resuspended in 1 ml of LB medium. This mixture was centrifuged at 5,000g for 5 min to pellet the cells, after which the supernatant was removed. The cell pellet for each strain was resuspended in Agrobacterium induction buffer (10 mM MES, pH 5.6, 10 mM MgCl₂, 150 μM acetoxyringone) and incubated at room temperature for 1–2 h. For individually tested strains, Agrobacterium suspensions were diluted to an optical density at 600 nm (OD₆₀₀) of 0.6. When strains were used in combination, the Agrobacterium suspensions for each strain were diluted to OD₆₀₀ = 0.3 and were mixed in equal concentration (according to the experiment in question). Prepared Agrobacterium suspensions were infiltrated into the abaxial side of N. benthamiana leaves of 4–5 week-old plants with a needleless 1-ml syringe. Unless stated otherwise, each experimental condition being tested consisted of three replicates, with each replicate corresponding to an independently infiltrated leaf. Each leaf among these three replicates was selected from a different N. benthamiana plant, such that no replicates for one experimental condition were taken from the same plant. In general, infiltrated leaves were harvested 4–5 days after infiltration, snap-frozen in liquid nitrogen and stored at −80 °C for downstream processing.

For substrate co-infiltration studies, 50–250 μM of 1 in water with 0.1–0.5% DMSO (1 was prepared as a 50 mM stock in DMSO) was infiltrated into the abaxial side of previously Agrobacterium-infiltrated leaves with a needleless 1-ml syringe 4 days after infiltration. For co-infiltration of (R:S)-α-tubulin (7) and O-methyllycanthin (9) as substrates, 25 μM of substrate in water with 1% methanol (7 and 9 were prepared as 2.5 mM stock solutions in methanol) was used. Substrate infiltration was no more difficult than Agrobacterium strain infiltration. Leaves were harvested 1 day later, snap-frozen in liquid nitrogen and stored at −80 °C for later processing.

**In vitro assays with N. benthamiana crude leaf extracts**
For in vitro experiments, leaves were infiltrated with Agrobacterium strains that contained the gene construct of interest, as described above. After 4 days, the leaves were excised and snap-frozen in liquid nitrogen. Approximately 1/3 of the infiltrated N. benthamiana leaf was
Homogenized with a small scoop of polyvinylpyrrolidone (Sigma) in liquid nitrogen on a ball mill (Retsch MM 400) using 5-mm-diameter stainless steel beads, with shaking at 25 Hz for 2 min. The homogenized leaf was then suspended in 1 ml of Tris buffer (100 mM Tris-HCl, pH 7.4, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride and 10 mM β-mercaptoethanol) and the suspension was incubated for 30 min on ice with periodic, gentle inversion. Extracts were centrifuged at 12,000 g at 4 °C for 10 min to pellet debris, after which the clarified plant extract was removed. A 100-μl volume in vitro assay contained 50 μM of compound I and 1 mM of S-adenosylmethionine, with the remaining volume made up by clarified plant protein extract. Assays were run for 2 h at 30 °C before quenching with one volume of acetonitrile (ACN). Reactions were then flash-frozen in liquid nitrogen and lyophilized overnight. The dried residue was resuspended in 200 μl of ACN with sonication and filtered through 0.45-μm PTFE filters before metabolite analysis.

**Heterologous expression in S. cerevisiae**

For yeast expression of cytochrome P450 genes, we used S. cerevisiae strain WAT11 (ade2), which has a chromosomal copy of the A. thaliana NADPH-cytochrome P450 reductase gene (ATRI)60. This strain as well as the yeast expression plasmid (pYeDP60) were provided by F. Pinot (Institut de Biologie moléculaire des Plantes). The WAT11 strain was routinely cultured with YPAD medium (10 g/l Bacto yeast extract, 20 g/l Bacto peptone, 20 g/l glucose and 80 mg/l adenine hemisulfate; and 18 g/l agar for plates), and unless noted otherwise, yeast cultures were grown at 30 °C, and with shaking at 250 rpm for liquid cultures.

Competent yeast cells were prepared and transformed with expression constructs (empty vector, GscCYP75A110 and GscCYP71FB1) using the Frozen-EZ Yeast Transformation II Kit (Bio-Rad) and were selected by plating on synthetic drop-out medium plates lacking adenine (6.7 g/l yeast nitrogen base without amino acids, 20 g/l glucose, 2 g/l drop-out mix minus adenine, 20 g/l agar). Colony PCR was used to confirm successful insertion of the genes of interest into pYeDP60, positive clones were restreaked on synthetic drop-out medium plates and individual colonies from the restreaked plates were used to inoculate liquid cultures of synthetic drop-out medium. After growth for 2 days, liquid cultures were used to prepare 25% glycerol stocks that were stored at −80 °C.

Growth of yeast cultures and heterologous protein production was performed much as previously described.61 Freshly streaked colonies from each strain of interest were inoculated in 4 ml synthetic drop-out medium and were then grown at 28 °C with 250 rpm shaking. After 2 days of growth, 2 ml of the initial culture was used to inoculate 500 ml of YPGE medium (10 g/l Bacto yeast extract, 10 g/l Bacto peptone, 5 g/l glucose and 3%(v/v) ethanol). Cultures were grown overnight at 28 °C with 250 rpm until reaching a density of 5 × 10^5 cells per ml, as determined by counting with a haemocytometer, at which point a sterile aqueous galactose solution was added to a final concentration of 10% (v/v) to induce heterologous gene expression. Cultures were then grown overnight until a density of around 5 × 10^6 cells per ml was reached, at which point cultures were immediately prepared for microsomal protein isolation.

**Isolation of yeast microsomes**

Preparation of yeast microsomal protein fractions was performed according to an established protocol.61 After reaching the desired cell density following galactose induction, cultures were centrifuged at 5,000 g for 5 min to pellet cells. Cell pellets were then resuspended in 1 ml of TEK buffer (50 mM Tris-HCl, 1 mM EDTA, 100 mM KCl, pH 7.4) per 0.5 g of wet cell pellet mass and incubated at room temperature for 5 min. Cells were then pelleted at 5,000 g for 5 min and resuspended in 5 ml of ice-cold TES B buffer (50 mM Tris-HCl, 1 mM EDTA, 600 mM sorbitol, pH 7.4). All subsequent steps were performed at 4 °C and/or on ice. A volume of glass beads (0.5 mm diameter) roughly equal to that of the cell resuspensions was added to the cells in 500-ml centrifuge tubes, and cells were lysed mechanically by vigorous up-and-down shaking for 10 min (30 s of shaking followed by 30 s on ice). Next, 10 ml of ice-cold TES B buffer was added to the clarified lysate, and the suspension was removed from the glass beads by pipetting and saved. The glass beads and residual lysate were washed an additional two times with 10 ml of TES B buffer, and each of these fractions was combined with the previous 10 ml fraction (for a total of around 30 ml). The combined lysate was then centrifuged at 23,000 g for 10 min at 4 °C to pellet large cell debris. The supernatant was removed from the pellet and the pellet was discarded. The supernatant (around 30 ml) was then diluted twofold with ice-cold TES B buffer (to a final volume of around 60 ml), and microsomes were precipitated by the addition of NaCl to a final volume of 150 mM and polyethylene glycol (PEG)-4000 to a final concentration of 0.1 g/ml. After an incubation of about 1 h on ice with periodic mixing, microsomal protein fractions were collected by centrifugation at 10,000 g for 10 min at 4 °C. The supernatant was discarded, and the microsomal pellet was resuspended in 1 ml of ice-cold TEG storage buffer (50 mM Tris-HCl, 1 mM EDTA, 20% (v/v) glycerol, pH 7.4). Microsomal protein content was quantified using the Bio-Rad Protein Assay Kit, and 60-μl aliquots were snap-frozen in liquid nitrogen and stored at −80 °C for future use.

**In vitro assays with microsomes purified from yeast**

Microsomal enzyme assays were run in potassium phosphate buffer (50 mM potassium phosphate, 100 mM NaCl, pH 7.5) in a volume of 100 μl. Each reaction contained 40 μg microsomal protein, 25 μM substrate (7 or 9) and 1 mM NADPH. Control reactions were performed in which microsomal protein or NADPH were omitted from the reaction. Reactions were incubated at 30 °C for several time points (15, 30 and 60 min), after which the reactions were quenched by the addition of 50 μl ACN with 0.1% formic acid. Quenched reactions were diluted tenfold in 1:1 water:ACN with 0.1% formic acid and filtered through 0.45-μm PTFE filters before LC–MS analysis.

**Metabolite extraction from plant material**

Liquid-nitrogen-frozen tissues (for example, leaves from N. benthamiana transient expression experiments and G. superba samples) were lyophilized to dryness and weighed to calculate the dry mass of each sample. The samples were homogenized on a ball mill (Retsch MM 400) using 5-mm-diameter stainless steel beads, with shaking at 25 Hz for 2 min. Generally, an 80:20 methanol:H₂O solution was used to extract the plant tissue, with 20 μl per mg dry weight used for N. benthamiana leaves and 100–250 μl per mg dry weight used for G. superba samples. The extracted samples were heated at 65 °C for 10 min, centrifuged at 10,000 g for 5 min to pellet plant debris, and the remaining solvent was filtered through 0.45-μm PTFE filters before analysis by high-resolution LC–MS. For detection of aldehydes, samples were derivatized using Girard’s reagent T.64 To accomplish this, 100 μl of filtered plant extract (or standard dissolved in methanol) was mixed with 100 μl of Girard T reagent in methanol (20 mg/ml) and 20 μl acetic acid. The samples were mixed and then incubated at 70 °C for 30 min. Samples were cooled to room temperature before LC–MS analysis.

**LC–MS analysis**

In general, metabolite samples were analysed by reversed-phase liquid chromatography on an Agilent 1260 HPLC, using a 5-μm, 2-mm × 100-mm Gemini NX-C18 column (Phenomenex). Water with 0.1% formic acid (A) and ACN with 0.1% formic acid (B) were used as the mobile phase components at a flow rate of 0.4 ml/min with the following 33-min gradient method: 0–1 min, 3% B; 1–21 min, 3–50% B; 21–22 min, 50–97% B; 22–27 min, 97% B; 27–28 min, 97–3% B; 28–33 min, 3% B. In vitro assays were initially analysed with reversed-phase chromatography as described above with the following 20-min gradient
method: 0–1 min, 3% B; 1–11 min, 3–97% B; 11–15 min, 97% B; 15–16 min, 97–3% B; 16–20 min, 3% B.

Analysis of polar intermediates (for example, tyramine, L-DOPA and dopamine) was performed using hyphrophilic interaction liquid chromatography (HILIC) analysis with either a 1.7-μm, 2.1-mm × 50-mm Acquity UPLC BEH Amide column (Waters) or a 5-μm, 2.1-mm × 100-mm XBridge BEH Amide column (Waters). For HILIC analysis, water with 0.125% formic acid, 10 mM ammonium formate (A) and 95:5 ACN:water with 0.125% formic acid, 10 mM ammonium formate (B) were used as the mobile phase components at a flow rate of 0.6 ml/min with the following 21-min gradient method: 0–2 min, 100% B; 2–12 min, 100–60% B; 12–13 min, 60–100% B; 13–21 min, 100% B. Before HILIC analysis, methanolic extracts (which were prepared as described above) and standards were diluted 1:10 in ACN with 0.1% formic acid to better match the initial mobile phase for adequate chromatography.

For both reversed-phase and HILIC analysis, a coupled Agilent 6520 Accurate-Mass Q-TOF ESI mass spectrometer was used to collect MS data in positive ion mode (parameters: mass range, 100–1700 m/z; drying gas, 300 °C, 11 l/min; nebulizer, 25 psig; capillary, 3.5 kV; fragmentor, 150 V; skimmer, 65 V; octopole RF Vpp, 750 V; 1,000 ms per spectrum). The first 0.5 min of each run was discarded to avoid salt contamination of the MS apparatus. For tandem mass spectrometry (MS/MS) analysis, 5-, 10-, 20- and 40-V collision energies were used to select for mass signatures with a retention time less than 1,800 s and an average peak intensity greater than 5 × 10^4. This filtering enabled the analysis to be focused on a more targeted list of differentially expressed features at specific retention times, along with associated or reported values represent individual.

Metabolomics and MS data analysis

High-resolution mass spectrometry (HRMS) data were analysed using MassHunter Qualitative Analysis software (Agilent) and XCMS-qt (Scripps Center for Metabolomics). For untargeted metabolomics, MassHunter (Agilent) data files were converted to mzXML or mzML format using trapper (Seattle Proteome Center) or MSConvert (ProteoWizard). Grouped mzXML/mzML files were preprocessed and analysed by XCMS, using the following sample R script: library(xcms); xset<-xcmsset(); xset<-group(xset); xset2<-retcor(xset,family="g",plottype="m"); xset2<-group(xset2); xset3<-retcor(xset2,family="s",plottype="m"); xset3<-group(xset3); xset4<-retcor(xset3,family="s",plottype="m"); xset4<-group(xset4,bw=10); xset5<-fillPeaks(xset4); reporttab<-diffreport(xset5, "A", "B", "A vs B", 200).

The output of this analysis contains a list of identified mass signatures at specific retention times, along with associated m/z values, peak intensity fold change, statistical test outputs (P value, two-tailed unequal variance Student’s t-test), retention times and extracted peak intensities. Unless stated otherwise, this list was generally filtered to select for mass signatures with a P value less than 0.1, a value greater than or equal to 0, a fold change greater than 5, a retention time less than 1,800 s and an average peak intensity greater than 5 × 10^4. This filtering enabled the analysis to be focused on a more targeted list of differentially produced mass signatures. The relative ion abundances for a given mass signature reported in this manuscript were generally determined in the MassHunter Qualitative Analysis software by automated integration (Agile method with default settings) of extracted ion chromatograms (EICs) with a 20–50 ppm mass range tolerance.

For all other datasets, statistical analyses of reported data were performed in Microsoft Excel 2016, GraphPad Prism (8.0.2) or JMP Pro (13.0). For each experiment or figure in which statistical tests were performed, the presented or reported values represent individual, distinct samples and not repeated measurements.

Synthesis of 1-phenethylisoquinoline precursor (1)

Synthesis of compound 1 (1-2-(4-hydroxyphenyl)ethyl)-6,7-dihydroxy-1,2,3,4-tetrahydroisoquinoline) was adapted from a previously described protocol. In brief, 151 mg of 4-HDCA (also known as 3-(4-hydroxyphenyl)propanal or 4-hydroxydihydrocinnamaldehyde; Small Molecules) and 144 mg of dopamine (Millipore Sigma) in 4 ml of 1:1 ACN:potassium phosphate buffer (0.1M, pH 6.0) was stirred at 50 °C for 14 h. After this incubation, three volumes of dichloromethane (DCM) were added to the mixture. The mixture was then extracted several times with Milli-Q water. The aqueous fractions were combined and acidified to pH 3.0. The mixture was then purified with a Strata-X-C 100-μm solid-phase extraction column (500 mg/3 ml, Phenomenex). We obtained 6.8 mg of putative 1 and verified the structure by ^1H NMR using a 400-MHz NMR spectrometer (Varian).

1-[2-(4-Hydroxyphenyl)ethyl]-6,7-dihydroxy-1,2,3,4-tetrahydroisoquinoline (1): ^1H NMR (400 MHz, methanol-d$_4$), δ 8.05 (d, J = 8.5 Hz, 2H), 6.70 (d, J = 8.4 Hz, 2H), 6.56 (s, 1H), 6.51 (s, 1H), 3.94 (q, J = 3.8 Hz, 1H), 3.28–3.22 (m, 1H), 3.00–2.93 (m, 1H), 2.83–2.55 (m, 4H), 2.14–1.94 (m, 2H). The NMR spectrum and comparisons to published data are shown in Supplementary Fig. 6.

Isolation and purification of an O-methylandrocymbine (9) standard

C. autumnale bulbs (McClure & Zimmerman and Odyssey Bulbs) were purchased between August and October 2018, at which point flower shoots had begun to emerge. Bulb, shoot and flower tissues were cut into small pieces, snap-frozen in liquid nitrogen and lyophilized to dryness. Dried tissue was frozen with liquid nitrogen and ground to a fine powder using mortar and pestle. In general, purification of putative 9 was based on a previously described purification of this compound. For 10–20 g samples of dried tissue, 500 ml of ethanol was added, and this extraction was stirred using a magnetic stir bar at 200 rpm for 24–48 h. The extract was clarified through filter paper (Fisherbrand Filter Paper, Qualitative PS), and a small aliquot of this was analysed by LC–MS for the presence of colchicine alkaloids, including putative 9 (M + H$^+$ = 386.1962), which were confirmed to be present. After this confirmation, the ethanol extract was dried under vacuum using a rotary evaporator system. The resulting yellow residue was resuspended in 500 ml water with 2% tartaric acid (pH 2). This fraction was partitioned with 50 ml of diethyl ether or tert-butyl methyl ether. The separated organic fraction was washed with 2 × 20 ml water with 2% tartaric acid and 2 × 8 ml deionized water, after which each of these aqueous fractions was combined with the initial 500 ml aqueous fraction. The pH of the aqueous phase was increased to 8 using saturated sodium bicarbonate, and the sample was then partitioned with 3 × 500 ml ethyl acetate. The organic fractions were combined and evaporated to dryness on a rotary evaporator.

The dried residue was resuspended in 10:1 DCM:methanol and filtered through Celite to remove any insoluble particulates. Successive rounds of FLASH chromatography using either 10:1 or 20:1 DCM:methanol with 2% triethylamine as the mobile phase and silica gel as the stationary phase afforded fractions (analysed by LC–MS) that contained most of the desired compound (putative 9), but also a considerable amount of demecoline (II, [M + H$^+$] = 372.1805), which was in much higher abundance, and therefore prevented NMR analysis. Furthermore, the desired product and demecoline (II) also co-eluted under reversed-phase HPLC conditions, thus preventing facile purification. To enable further chromatographic purification, fractions containing the desired product were mixed with acetic formic anhydride, as described below for the synthesis of N-formyldemecoline (10). This resulted in near complete consumption of demecoline (II) and the production of N-formyldemecoline (10), which is easily separable from 9 by reversed-phase HPLC. Essentially no loss of 9 occurred in this process. The desired product was then purified by preparative HPLC using an Agilent 1260 Infinity preparative-scale HPLC system with an Agilent 1100 diode array detector and a Clipere C18 5-μm, 10 mm × 250 mm semi-prep column (Higgins Analytical). Water with 0.1% formic acid (A) and ACN with 0.1% formic acid (B) were used as the mobile phase components at a flow rate of 3 ml/min with the following gradient method: 0–1 min, 20% B; 1–21 min, 20–50% B; 21–22 min, 50–97% B; 22–32 min, 97% B; 32–33 min, 97–20% B; 33–45 min, 20%
Acetylation of DNA sequences for the data shown are in Supplementary Fig. 7.

Synthesis of N-formyldemecolcine (10)

N-formyldemecolcine was synthesized through N-formylation of demecolcine (Millipore Sigma) following an adaptation of a previously established amine formylation protocol66. Formic acid (61 μl, 1.62 mmol) and acetic anhydride (153 μl, 1.62 mmol) were mixed without under inert gas and stirred at 65 °C for 30 min, after which the reaction mix, containing acetic anhydride, was allowed to cool to room temperature. This reagent was then added to demecolcine (3.3 mg, 8.9 μmol) dissolved in 100 μl DCM under inert gas, and the mixture was stirred on ice (0–4 °C) for 1 h. After this incubation, excess methanol was added to quench the reaction, which was then evaporated to dryness using a rotary evaporator. Sufficient reaction was then verified by LC–MS, which revealed that nearly all demecolcine (m/z 372.1805) had been consumed, and that a new peak had appeared with an HRMS m/z value of 400.1755 Da that nearly all demecolcine (m, 4H). The NMR spectrum and comparisons to published data are shown in Supplementary Fig. 8.

Data availability

All reported data in this study are available via database or by request from the corresponding author. Coding DNA sequences for the genes characterized and assessed in this study are provided as FASTA sequences in the Supplementary sequences, and are deposited in the National Center for Biotechnology Information (NCBI) GenBank database with the following accessions: GsOMT1 (MT512039), GsNMT1 (MT512040), GsNMT2 (MT512041), GsCYP75A109 (MT512042), GsOMT2 (MT512043), GsOMT3 (MT512044), GsCYP75A110 (MT512045), GsNMT4 (MT512046), GsCYP77FI1 (MT512047), GsTyDC/DC (MT512048), GsPAI (MT512049), GsGCR (MT512050), GsAER (MT512051), Gs4CL1 (MT512052), GsC4H (MT512053), GsDAH5 (MT512054), A24–CjNCS (MT512055) and A29–CjNCS (MT512056). Raw reads from the RNA-seq profiling analysis of G. superba are deposited in the NCBI Sequence Read Archive (SRA) database under the BioProject accession PRJNA634925. The corresponding Transcriptome Shotgun Assembly (TSA) project has been deposited at DDBJ/EMBL/GenBank under accession GIOZ00000000. The version described in this paper is the first version, GIOZ01000000. Gene constructs described in this manuscript will be made available upon request from the corresponding author. Synthetic substrates and purified compounds will be made available upon request from the corresponding author, as possible. Source data are provided with this paper.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Characterization of GsOMT1. a, LC–MS chromatograms demonstrating activity on substrate 1 by protein lysates from N. benthamiana leaves that transiently express GsOMT1. EICs for 1 (m/z 286.1438; left) and the methylated product (indicated by the asterisk) produced in this experiment (m/z 300.1594; right) are shown. This experiment was performed three times, with similar results each time. b, MS/MS fragmentation spectrum of 1, as well as the generated m/z 300.1594 product (*) at a collision energy of 20 V. Fragmentation of both compounds was performed twice, with similar results observed each time. c, Tabulated list and putative structures for ion fragments from the MS/MS analysis of 1. d, Tabulated list and putative structures for ion fragments from the MS/MS analysis of the m/z 300.1594 product. See Supplementary Information for a detailed analysis of MS/MS results. e, Proposed reaction catalysed by GsOMT1, as supported by MS/MS fragmentation and previously published labelling studies. f, Transient expression of GsOMT1 in N. benthamiana with co-infiltrated substrate 1 results in production of the methylated product 2, as shown by the LC–MS chromatograms. This experiment was performed more than three times with similar results observed each time. g, Untargeted metabolite analysis (XCMS) comparing transient expression of GFP (negative control) to that of GsOMT1 with co-infiltrated substrate 1 (n = 3 independent replicates for each experimental condition). The unique mass signatures (P < 0.1 between samples, as determined by XCMS) are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The mass isotopologues (M₀ and M₁) for the presumed product (m/z 300.1594) are shown in red; the substrate (1, m/z 286.1438) is shown in blue. r.t., retention time.
Extended Data Fig. 2 | Characterization of GsNMT. a, Co-expression of GsOMT1 and GsNMT in N. benthamiana with co-infiltrated 1 leads to consumption of putative 2 (m/z 300.1594) and the production of a new compound that corresponds to a methylation (m/z 314.1751), as shown by the LC–MS chromatograms. Activity of full-length GsNMT was confirmed in three separate experiments. b, MS/MS fragmentation spectrum of the generated m/z 314.1751 product (*) at a collision energy of 20 V. This experiment was performed twice, with similar results observed each time. c, Tabulated list and putative structures for ion fragments from MS/MS analysis of the m/z 314.1751 product. See Supplementary Information for a detailed analysis of MS/MS results. d, Untargeted metabolite analysis (XCMS) comparing the presence and absence of GsNMT in the transient co-expression system (n = 6 independent replicates for each experimental condition). The unique mass signatures (P < 0.1 between samples, as determined by XCMS) are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The mass isotopologues (M_0 and M_1) of the presumed product (m/z 314.1751) are shown in red; the mass isotopologues of the presumed substrate (m/z 300.1594) are shown in blue. e, Proposed reaction catalysed by GsNMT as supported by MS/MS fragmentation and previously published labelling studies.
Extended Data Fig. 3 | Characterization of GsCYP75A109.  

**a**, Addition of GsCYP75A109 to the *N. benthamiana* transient expression system with co-infiltrated 1 leads to the consumption of 3 (m/z 314.1751) and the production of a new compound that corresponds to a hydroxylation (m/z 330.1700), as shown by the LC–MS chromatograms. These results were confirmed in two independent experiments. 

**b**, MS/MS fragmentation spectrum of the generated m/z 330.1700 product (*) at a collision energy of 20 V. Consistent results were obtained in three separate experiments. 

**c**, Tabulated list and putative structures for ion fragments generated from MS/MS analysis of the m/z 330.1700 product. See Supplementary Information for a detailed analysis of MS/MS results. 

**d**, Untargeted metabolite analysis (XCMS) comparing the presence and absence of GsCYP75A109 in the transient co-expression system (n = 6 independent replicates for each experimental condition). The unique mass signatures (P < 0.1 between samples, as determined by XCMS) comparing the presence and absence of GsCYP75A109 in the transient co-expression system are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The presumed product (m/z 330.1700) is shown in red; the mass isotopologues (M₀, M₁) of the presumed substrate (m/z 314.1751) are shown in blue. 

**e**, Proposed reaction catalysed by GsCYP75A109 as supported by MS/MS fragmentation and previously published labelling studies. 

**f**, An N-terminal truncation of a predicted mitochondrial localization signal from GsNMT (yielding GsNMTt) increases the yield of putative 4 (m/z 330.1700) in the transient co-expression system, as shown by the representative LC–MS chromatograms. 

**g**, Quantification of the heterologous production of 3 (m/z 314) or 4 (m/z 330) with the use of GsNMT or GsNMTt in the co-expression system. Filled-in boxes (grey) indicate the presence of a gene in the co-expression experiment; an empty box (white) indicates its absence. For each reaction, data are mean ± s.d. of 3 distinct biological replicates. Statistical comparisons were made using a one-tailed Student’s t-test, with an assumption of unequal variance. n.d., not detected. Direct comparison between the experimental conditions was performed twice with similar results obtained each time. Activity of GsNMTt in pathway engineering was consistent in more than three experiments.
Extended Data Fig. 4 | Characterization of GsOMT2. a, Addition of GsOMT2 into the *N. benthamiana* transient co-expression system with co-infiltrated 1 leads to consumption of putative 4 ([m/z 330.1700]) and the production of a new compound corresponding to both a methylation and a hydroxylation ([m/z 360.1805]), as shown by the LC–MS chromatograms. This activity was confirmed in more than three independent experiments. b, MS/MS fragmentation spectrum of the generated [m/z 360.1805] product (*) at a collision energy of 20 V. MS/MS fragmentation of this peak was performed twice, with similar results each time. c, Tabulated list and putative structures for ion fragments from the MS/MS analysis of the [m/z 360.1805] product. See Supplementary Information for a detailed analysis of MS/MS results. d, Untargeted metabolite analysis (XCMS) comparing the presence and absence of GsOMT2 in the transient co-expression system (*n = 6* independent replicates for each experimental condition). The unique mass signatures (*P < 0.1* between samples, as determined by XCMS) are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The mass isotopologues (*M₀*, *M₁*, and *M₂*) for the presumed product ([m/z 360.1805]) are shown in red; the mass isotopologues (*M₀*, *M₁*) of the presumed substrate ([m/z 330.1700]) are shown in blue. e, Proposed reaction catalysed by GsOMT2, and tentatively GsCYP75A109, as supported by MS/MS fragmentation and previously published labelling studies. Note that compound 5 is not observed in our co-expression system, presumably due to its consumption to 6.
Extended Data Fig. 5 | Characterization of GsOMT3. a, Addition of GsOMT3 into the *N. benthamiana* transient co-expression system with co-infiltrated 1 leads to consumption of 6 (m/z 360.1805) and the production of a new compound corresponding to a methylation (m/z 374.1962), as shown by the LC–MS chromatograms. The new peak was compared to a racemic standard of autumnaline (7) ((R,S)-autumnaline), which supports the identity of this new compound as 7. This experiment was repeated more than three times with similar results observed each time. b, MS/MS fragmentation spectrum of the generated m/z 374.1962 product (*) compared to that of racemic 7, each at a collision energy of 20 V. This experiment was performed three times, with similar results observed each time. c, Tabulated list and putative structures for ion fragments generated from MS/MS analysis of the m/z 374.1962 product. See Supplementary Information for a detailed analysis of MS/MS results. d, Untargeted metabolite analysis (XCMS) comparing the presence and absence of GsOMT3 in the transient co-expression system (n = 6 independent replicates for each experimental condition). The unique mass signatures (P < 0.1 between samples, as determined by XCMS) are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The mass isotopologues (M0, M1) for the product (m/z 374.1962) are shown in red; the mass isotopologues (M0, M1, and M2) of the presumed substrate (m/z 360.1805) are shown in blue. e, Proposed reaction catalysed by GsOMT3, as supported by MS/MS fragmentation, previously published labelling studies and comparison to a 7 standard.
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Characterization of GsCYP75A110.  

a, Addition of GsCYP75A110 into the N. benthamiana transient co-expression system with co-infiltrated 1 leads to consumption of 7 (m/z 374.1962) and the production of three new compounds that each correspond to a loss of two hydrogens (m/z 372.1805), as shown by the LC–MS chromatograms. This experiment was performed more than three times with similar results observed each time.

b, MS/MS fragmentation spectrum of the generated m/z 372.1805 product (*) at a collision energy of 20 V. This spectrum is shown because it represents the only peak consumed in downstream biosynthesis (see Extended Data Fig. 7). MS/MS fragmentation of this peak was performed twice, with similar results observed each time.

c, Tabulated list and putative structures for ion fragments from the MS/MS analysis of the m/z 372.1805 product.

d, Untargeted metabolite analysis (XCMS) comparing the presence and absence of GsCYP75A110 in the transient co-expression system (n = 6 independent replicates for each experimental condition). The unique mass signatures (P < 0.1 between samples, as determined by XCMS) are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The mass signatures for two of the presumed products (m/z 372.1805) are shown in red; the mass signature of the presumed substrate (m/z 374.1962) is shown in blue.

e, Expression of GsCYP75A110 individually with substrate (7) co-infiltration, as shown by the LC–MS chromatograms of the substrate (7, m/z 374.1962) and products (m/z 372.1805). The products produced by pathway reconstitution in N. benthamiana are shown for comparison. This experiment was performed once.

f, In vitro assays using microsomal protein isolated from yeast expressing GsCYP75A110. LC–MS chromatograms of substrate (7) and products (m/z 372.1805) are shown for comparison to the products produced in the N. benthamiana transient expression system. Peak integrations for the substrate (7) are shown in blue text to demonstrate consumption of the substrate in the presence of GsCYP75A110-containing microsomal protein and NADPH. This experiment was performed once.

g, Predicted, alternative phenol coupling isomers may explain the three isomeric peaks detected with m/z 372.1805.

h, Proposed reaction catalysed by GsCYP75A110, as supported by MS/MS fragmentation and previously published labelling studies.
Extended Data Fig. 7 | Characterization of GsOMT4. 

**a**, Addition of GsOMT4 into the *N. benthamiana* transient co-expression system with co-infiltrated 1 leads to consumption of 8 (m/z 372.1805) and the production of a new compound corresponding to a methylation (m/z 386.1962), as shown by the LC–MS chromatograms. Comparison to an O-methylandrocymbine (9) standard purified from *C. autumnale* plants supports the identity of this compound as 9. This result was confirmed in more than three independent experiments. 

**b**, MS/MS fragmentation spectrum of the generated m/z 386.1962 product (*) compared to the purified 9 standard, with both compounds fragmented at a collision energy of 20 V. This was performed twice, with similar results observed each time. 

**c**, Tabulated list and putative structures for ion fragments from the MS/MS analysis of the m/z 386.1962 product. 

**d**, Untargeted metabolite analysis (XCMS) comparing the presence and absence of GsOMT4 in the transient co-expression system (*n* = 6 independent replicates for each experimental condition). The unique mass signatures (*P* < 0.1 between samples, as determined by XCMS) are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The mass signature for the product (9, m/z 386.1962) is shown in red; the mass signature of the presumed substrate (m/z 372.1805) is shown in blue. 

**e**, Proposed reaction catalysed by GsOMT4, as supported by MS/MS fragmentation, previously published labelling studies and comparison to an isolated 9 standard.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Characterization of GsCYP71FB1. a, Addition of GsCYP71FB1 into the N. benthamiana transient co-expression system with co-
infiltrated I leads to consumption of 9 (m/z 386.1962) and the production of a new compound with identified masses of m/z 400.1755 [M + H+] and 422.1574 [M + Na+] as shown by the LC–MS chromatograms. Comparison to an authentic N-formyldemecolcine (10) standard supports formation of this compound. This experiment was performed twice, with similar results observed each time. b, MS/MS fragmentation spectrum of the generated m/z 400.1755 product (*) compared to the 10 standard, with both compounds fragmented at a collision energy of 20 V. This was performed three times, with similar results each time. c, Transient expression of GsCYP71FB1 individually in N. benthamiana with substrate (9) co-infiltration, as shown by the LC–MS chromatograms of the substrate (9, m/z 386.1962) and product (10, m/z 400.1755) in comparison to a 10 standard. This experiment was performed once. d, In vitro assays using microsomal protein isolated from yeast expressing GsCYP71FB1. The LC–MS chromatograms of the substrate (9) and product (10) in comparison to the 10 standard are shown. Peak integrations for the substrate (9) are shown in blue text to demonstrate its consumption in the presence of GsCYP71FB1-containing microsomal protein and NADPH. This experiment was performed once. e, Untargeted metabolite analysis (XCMS) comparing the presence and absence of GsCYP71FB1 in the transient co-expression system (n = 6 independent replicates for each experimental condition). The unique mass signatures (P < 0.1 between samples, as determined by XCMS) are shown in ranked order based on their increasing (top) or decreasing (bottom) fold change in abundance between the two conditions. The mass isotopologues (M₀, M₁) as well as adducts (+Na, +K) for the product (10, m/z 400.1755) are shown in red; the mass isotopologues (M₀, M₁) and adducts (2M+Na) of the substrate (9, m/z 386.1962) are shown in blue. f, Proposed reaction catalysed by GsCYP71FB1, as supported by MS/MS fragmentation, previously published labelling studies and comparison to an authentic 10 standard.
Extended Data Fig. 9 | Comparison of intermediates produced in the N. benthamiana co-expression system to G. superba metabolites. Each biosynthetic product downstream of 1 produced in our co-expression system (black traces) was compared to the equivalent mass ion found in G. superba rhizome extracts (blue traces) or to a verified standard (red traces) by LC–MS analysis. Additionally, MS/MS spectra for co-eluting peaks were compared to demonstrate the chemical similarity between these compounds. Collision energies for all shown MS/MS analyses were 20 V, with the exception of 1, for which fragmentation at 10 V is shown. These LC–MS comparisons were performed once with multiple biological replicates of G. superba metabolite extractions (n = 6 biological replicates from four different tissues: leaf, stem, root and rhizome). The chromatographic traces for G. superba metabolites in this figure are from a representative rhizome extract. Retention time and MS/MS spectra for compounds produced by heterologous expression in N. benthamiana were consistent among individual experiments.
Extended Data Fig. 10 | See next page for caption.
Extended Data Fig. 10 | Engineering early metabolites in colchicine biosynthesis. a, List of module-1 biosynthetic genes and their best BLASTP hit in *A. thaliana*. Note that all genes except for Gs/AER seem to have an orthologue in *Arabidopsis* with >60% identity, suggesting functional equivalence. b, Generalized pathway for the proposed engineered production of 4-HDCA in *N. benthamiana*. c, LC-MS chromatograms demonstrating that co-expression of module 1 in *N. benthamiana* leads to production of 4-HDCA, which was detected as the Girard reagent T derivative (m/z 264.1707). Production of 4-HDCA by module-1 proteins was demonstrated three times with similar results each time. d, LC-MS chromatograms (by HILIC analysis) assessing the production of tyramine (left, m/z 138) and dopamine (right, m/z 154) with individual and co-expression of Gs/TyDC/ DDC and Bv/CYP76ADS (module 2). Next to each set of chromatograms is the corresponding relative quantifications of tyramine (m/z 138), l-DOPA (m/z 198) and dopamine (m/z 154) in each reaction. For each reaction, data are mean ± s.d. of 3 distinct biological replicates. Module 2 activity was confirmed in more than three individual experiments. e, Proposed scheme for the engineered biosynthesis of dopamine. f, Comparison of the native function of Cj/NCS in benzyllisoquinoline alkaloid biosynthesis to the putative reaction required in colchicine alkaloid biosynthesis. g, Co-expression of both module-1 and module-2 genes in *N. benthamiana* leads to concurrent production of the requisite aldehyde (m/z 264.1707, Girard T derivative), as well as dopamine (m/z 154.0863), as shown by the LC-MS chromatograms. Note that dopamine is observed here by C18 chromatography. Co-expression of both modules was performed more than three times, with similar results each time. h, LC-MS chromatograms for the co-expression of module 1 and module 2 with Cj/NCS in comparison to an authentic 1 standard. These experiments were performed more than three times, with similar results observed each time. i, MS/MS fragmentation comparison between the newly identified m/z 286.1438 peak (*) and the 1 standard. Both were analysed with a collision energy of 10 V. This MS/MS comparison was performed twice. j, Comparison of the function of wild-type, full-length Cj/NCS to N-terminal truncations of 24 (Δ24-Cj/NCS) and 29 (Δ29-Cj/NCS) amino acids. Filled-in boxes (grey) indicate the presence of a gene in the co-expression experiment; an empty box (white) indicates its absence. For each reaction, data are mean ± s.d. of 3 biological replicates.
Extended Data Fig. 11 | See next page for caption.
Extended Data Fig. 11 | Metabolic engineering of colchicine alkaloids in Nicotiana benthamiana. a, Biosynthetic schematic of the transient metabolic engineering system in N. benthamiana for the production of 2 and 9. b, LC-MS chromatograms for the co-expression of GsOMT1 with module 1, module 2 and Δ24-CjNCS compared to that of GsOMT1 expressed alone with co-infiltration of 1. The EICs for 1 (blue traces, m/z 286.1438) and the production of 2 (red traces, m/z 300.1594) are shown. c, LC-MS chromatograms demonstrating the production of 9 through co-expression of module 3 (without GsCYP71FBI) with module 1, module 2 and Δ24-CjNCS. This is compared to infiltration of 1 (as substrate) with co-expression of module 3 (without GsCYP71FBI), as well as to a standard of O-methylandrocymbine (9). The EICs specific to the exact mass of 9 (m/z 386.1962) are shown. Engineered production of 2 and 9 was demonstrated three times for each molecule. d, Production of two different colchicine alkaloids (2, m/z 300.1594; 10, m/z 400.1755, 422.1574) through metabolic engineering in N. benthamiana when GsAER is either omitted or included. Filled-in boxes (grey) indicate the presence of a module or gene in the co-expression experiment; an empty box (white) indicates its absence. For each reaction, data are mean ± s.d. of 6 biological replicates for each condition. Statistical significance was assessed using a two-tailed Student’s t-test with an assumption of unequal variance. The production of 2 in this context was assessed once; the production of 10 was performed twice with similar results each time. e, Individual dropout of each module-1 and module-2 gene in the engineered production of 10 (m/z 400.1755, 422.1574). For each reaction, data are the mean ± s.d. for each condition. n = 3 for GFP control; n = 5 for PAL, CCR, AER, C4H, TyDC/DDC and BvCYP76AD5 dropouts; n = 6 for 4CL and DAHPS dropouts and for the no-dropout control. All replicates represent independent biological replicates. Statistical comparisons were made using Dunnett’s test (two-tailed) with comparison to the full pathway control (indicated by arrow). *** = P < 0.001. This experiment was performed twice, with similar results each time.
Extended Data Fig. 12 | See next page for caption.
Extended Data Fig. 12 | Dropout analysis of module-3 biosynthetic genes.

Metabolic engineering of the full pathway to N-formyldemecolcine (10) in N. benthamiana was compared to transient co-expression systems in which individual module-3 enzymes were removed.

a. Accumulation of proposed pathway intermediates in dropout experiments. Grey boxes to the left of the graph indicate biosynthetic genes or modules included in a co-expression experiment; white boxes indicate their absence. For each intermediate, data represent the mean ± s.d. of the extracted ion abundance (n = 3) for the exact ion mass [M + H]+ (for 10, both [M + H]+ and [M + Na]+) at the retention time (r.t.) that corresponds to the compound.

b. Dropout of GsOMT1 from the full engineered pathway leads to accumulation of a new compound with a mass equivalent to 2 (m/z 300.1594), as shown by the LC–MS chromatograms. The newly identified peak is indicated by the arrow.

c. Transient co-expression of GsOMT1 or GsNMTt with module 1, module 2 and Δ24-CjNCS (for production of I). The LC–MS chromatograms for the substrate (I, m/z 286.1438), singly methylated products (2a and 2b, m/z 300.1594) and doubly methylated product (3, m/z 314.1751) are shown.

d. MS/MS fragmentation spectrum of 2a (collision energy of 20 V), as well as a tabulated list and putative structures for the ion fragments.

e. MS/MS fragmentation spectrum of 2b (collision energy of 20 V), as well as a tabulated list and putative structures for the ion fragments.

Note that fragment B (m/z 269) supports the placement of the methyl group on the nitrogen. For reference, compare to fragment B of 2a in d.

f. Comparative consumption of I by GsOMT1 and GsNMTt. Grey boxes indicate the presence of a gene or module in the co-expression experiment; a white box indicates its absence. n = 3 for each reaction; statistical comparisons made using Dunnett’s test with comparison to the module 1/module 2/Δ24-CjNCS control.

g. Proposed scheme for the initial methylations of I. All experiments shown in this figure were performed once.
Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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.

- 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
- 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
- No software was used

Data analysis
- No software was used

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All reported data within this study are available via database or by request. Coding DNA sequences for the genes characterized and assessed in this study are provided as FASTA sequences within the Supplementary Information document, and are deposited in the National Center for Biotechnology Information (NCBI) Genbank database with the following accessions: GsOMT1 (MT512039), GsNMT (MT512040), GsNMTt (MT512041), GsCYP75A109 (MT512042), GsOMT2 (MT512043), GsOMT3 (MT512044), GsCYP75A110 (MT512045), GsOMT4 (MT512046), GsCYP71F81 (MT512047), GsTDC/DDC (MT512048), GsPAL (MT512049), GsCCR (MT512050), GsAER (MT512051), GsC4H (MT512052), GsDAHPS (MT512053), Δ24-CjNCS (MT512054), Δ29-CjNCS (MT512055), Δ24-Δ29-CjNCS (MT512056). Raw reads from RNA-seq profiling of G. superba are deposited in the NCBI Sequence Read Archive (SRA) database under the BioProject accession PRJNA634925. The de novo assembled transcriptome from these reads is deposited in the NCBI Transcriptome Shotgun Assembly (TSA) database with the accession XXXXXXXX000000000000. Gene constructs described in this manuscript will be made available upon request. Synthetic substrates and purified compounds will be made available upon request, as possible. Select data sets that were critical to the conclusions of this study are provided with the online version of this manuscript.
Life sciences study design

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

Sample size
No sample size calculation was performed. Replicates of 3 were chosen for heterologous expression assays, as this is the minimal number required to perform statistical tests for the untargeted detection of differentially produced mass signatures. Three replicates was typically sufficient, as we were usually assessing the presence/absence of a mass signature (representative of a metabolite). Sample size was increased to 4-6 in some cases in which we were interested in determining an effect in the yield of a specific metabolite under certain experimental conditions.

Data exclusions
No data were excluded from the figures/experiments represented in this manuscript in which an average of many measurements is assessed. However, in some figures/experiments, a non-quantitative representative example is shown (e.g. individual chromatograms or mass spectra), as these are intended to only highlight a given result.

Replication
All experiments describing the novel function of a biosynthetic gene were performed at least twice, and in most cases, greater than 3 times, with similar results consistently observed. Experiments verifying the metabolically engineered pathway for colchicine alkaloids has been verified >3 times with several different pathway intermediates. The experiment assessing the effect of truncating the N-terminal sequence from GsNMT was tested twice, and our continued use of this truncated protein replicated similar levels of production that were previously observed. The experiment testing the effect of including or removing GsAER from the metabolic engineering system was performed several times with different pathway intermediates as the measured output. Dropout of module 3 genes was performed once. Expression of GsCYP75A110 and GsCYP71FB1 in yeast, with corresponding microsomal enzyme assays, was performed once for each enzyme.

Randomization
For heterologous expression in Nicotiana benthamiana leaves, 3 replicates (one leaf for each replicate) were typically used for each biosynthetic gene or co-infiltrated set of biosynthetic genes being tested. These three replicates were randomized among different plants such that no plant contained multiple replicates from the same condition being queried. For example - if three different combinations of biosynthetic genes were being tested with three replicates each, three plants would be used, and each of these plants would contain one replicate from each different condition. Due to the difference between top and bottom leaves on these plants (young/small vs. old large), the same randomization would be used to ensure that the three replicates for each condition are spread across the three different "ages" of leaves being utilized. For RNA-sequencing, RNA samples were isolated at the same time using the same kit, the library was amplified with the same number of PCR cycles during library prep, and each library was mixed in equalmolar concentration into a single Illumina MiSeq lane. LC-MS samples were randomized using a number sorter within Microsoft Excel before instrument runs to minimize the effect of instrument drift.

Blinding
The investigators were not blinded during these experiments. Untargeted analysis of mass spectrometry data was performed using XCMS software, which is agnostic to the nature of the samples being analyzed. Other experiments required the insight of the investigators into the enzymesgenes and/or metabolites being analyzed.

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 | Involved in the study |
|-----|-----------------------|
| ☒   | Antibodies            |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology         |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |
| ☒   | Clinical data         |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | ChIP-seq              |
| ☒   | Flow cytometry        |
| ☒   | MRI-based neuroimaging |

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research.

Laboratory animals

The study did not involve laboratory animals.
| Category               | Description                                                                 |
|------------------------|-----------------------------------------------------------------------------|
| Wild animals           | The study did not involve wild animals.                                     |
| Field-collected samples| The study did not involve samples collected from the field.                 |
| Ethics oversight       | No ethical approval or guidance was required. The only living organisms used within our study were generally accessible plants and bacteria. |

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