the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) doi: 10.1111/tpj.16698

The genomic and enzymatic basis for iridoid biosynthesis in cat thyme (*Teucrium marum*)

Samuel J. Smit[1], Sefa Ayten[2], Barbara A. Radzikowska[1,3], John P. Hamilton[4,5], Swen Langer[6], William P. Unsworth[3], Tony R. Larson[6], C. Robin Buell[2,4,5,*] and Benjamin R. Lichman[1,*]

[1]*Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK,*
[2]*Institute of Plant Breeding, Genetics, & Genomics, University of Georgia, Athens, Georgia 30602, USA,*
[3]*Department of Chemistry, University of York, York YO10 5DD, UK,*
[4]*Center for Applied Genetic Technologies, University of Georgia, Athens, Georgia 30602, USA,*
[5]*Department of Crop & Soil Sciences, University of Georgia, Athens, Georgia 30602, USA, and*
[6]*Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, UK*

Received 30 November 2023; revised 30 January 2024; accepted 14 February 2024.
*For correspondence (e-mail robin.buell@uga.edu; benjamin.lichman@york.ac.uk).

SUMMARY

Iridoids are non-canonical monoterpenoids produced by both insects and plants. An example is the cat-attracting and insect-repelling volatile iridoid nepetalactone, produced by *Nepeta* sp. (catmint) and aphids. Recently, both nepetalactone biosynthetic pathways were elucidated, showing a remarkable convergent evolution. The iridoid, dolichodial, produced by *Teucrium marum* (cat thyme) and multiple insect species, has highly similar properties to nepetalactone but its biosynthetic origin remains unknown. We set out to determine the genomic, enzymatic, and evolutionary basis of iridoid biosynthesis in *T. marum*. First, we generated a *de novo* chromosome-scale genome assembly for *T. marum* using Oxford Nanopore Technologies long reads and proximity-by-ligation Hi-C reads. The 610.3 Mb assembly spans 15 pseudomolecules with a 32.9 Mb N50 scaffold size. This enabled identification of iridoid biosynthetic genes, whose roles were verified via activity assays. Phylogenomic analysis revealed that the evolutionary history of *T. marum* iridoid synthase, the iridoid scaffold-forming enzyme, is not orthologous to typical iridoid synthases but is derived from its conserved paralog. We discovered an enzymatic route from nepetalactol to diverse iridoids through the coupled activity of an iridoid oxidase cytochrome P450 and acetyltransferases, via an inferred acylated intermediate. This work provides a genomic resource for specialized metabolite research in mints and demonstration of the role of acetylation in *T. marum* iridoid diversity. This work will enable future biocatalytic or biosynthetic production of potent insect repellents, as well as comparative studies into iridoid biosynthesis in insects.

Keywords: biosynthesis, iridoids, mint family, comparative genomics, enzymology, *Teucrium marum*.

INTRODUCTION

Plants and animals produce and emit a host of volatile compounds for communication and defense (Rosenkranz et al., 2021). In some circumstances, species from both kingdoms can produce the same compounds, having convergently evolved biosynthetic pathways for their formation (Beran et al., 2019). One class of specialized metabolites found in multiple kingdoms are the iridoids, a form of non-canonical cyclic monoterpenoids (Figure 1a), observed in both asterids and insects (El-Naggar & Beal, 1980; Jensen, 1991). In plants, beyond their role as volatiles, iridoids can be found as glycosides and precursors to complex monoterpenoid alkaloids (Figure 1a) (Miettinen et al., 2014; O'Connor & Maresh, 2006).

An example of a multi-kingdom volatile iridoid is nepetalactone (Figure 1a), which is produced by plants including *Nepeta* spp. (catmint), silver vine and valerian (Bol et al., 2017), and by multiple aphid species, where it acts as a sex pheromone (Birkett & Pickett, 2003; Dawson et al., 1987). The specific function of nepetalactone in plants is unresolved but its insect-repellent properties make its likely role in protection against phytophagous insects (Eisner, 1964). Remarkably, cats have evolved to be attracted to and to anoint themselves with nepetalactone due to these insect-repellent properties (Melo et al., 2021; Uenoyama et al., 2021). Recently, the biosynthetic pathways to nepetalactone in both *Nepeta* spp. (Lichman et al., 2019, 2020) and aphids (Köllner et al., 2022; Partridge

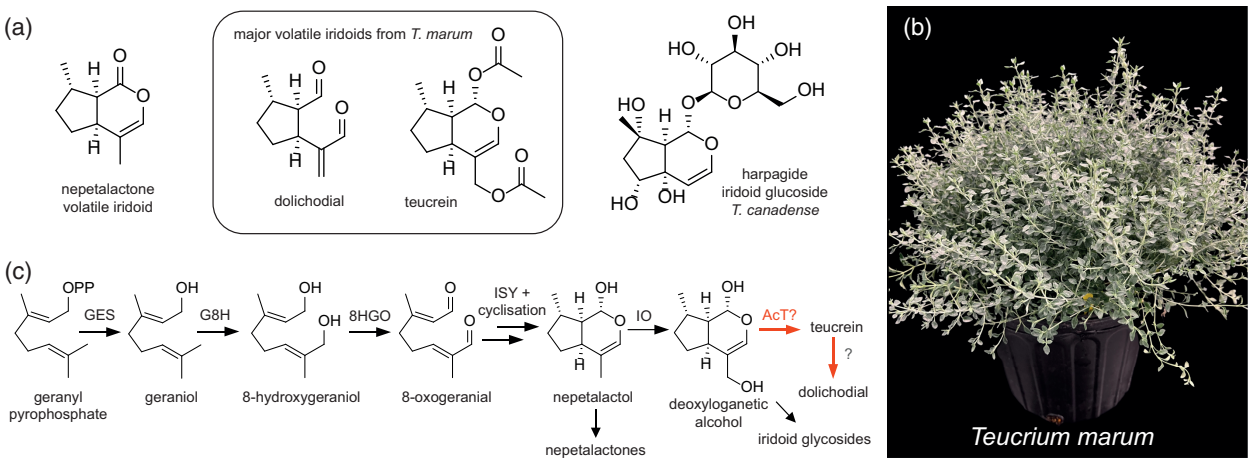

Figure 1. *Teucrium marum* and iridoid metabolites.
(a) Chemical structures of major iridoids from mint family species.
(b) Above ground tissues of *T. marum*.
(c) Iridoid pathway from geranyl pyrophosphate to deoxyloganetic alcohol, characterized across multiple species. GES, geraniol synthase; G8H, geraniol 8-hydroxylase; 8HGO, 8-hydroxygeraniol oxidoreductase; ISY, iridoid synthase; IO, iridoid oxidase; AcT, acetyl transferase. Red arrows represent hypothesized but uncharacterised steps.

et al., 2021) have been elucidated. In a remarkable example of convergent evolution, the chemical steps are identical but the enzymes are of distinct origins.

The compound dolichodial, another volatile iridoid, bears notable similarities to nepetalactone (Figure 1a). It can be found in the plant *Teucrium marum* (Figure 1a; Figure S1), which, like *Nepeta*, is a member of the mint family (Lamiaceae) (Pagnoni et al., 1976). It was first discovered in ants (Cavill & Hinterberger, 1961) and stick insects (Meinwald et al., 1962), and later identified as a component of sex pheromones in aphids (Dewhirst et al., 2008). In plants, conversely dolichodial likely acts as a defensive agent as it repels ants, flies, and cockroaches (Eisner et al., 2000). Furthermore, it has euphoria-inducing effects on cats akin to *Nepeta* spp. (Poli et al., 2007) and hence *T. marum* is known colloquially as 'cat thyme'. These factors provide motivation for further investigation into dolichodial biosynthesis.

The dolichodial biosynthetic pathway is unknown and no biosynthetic genes have been identified from *T. marum* or insect producers. However, dolichodial likely shares the same early biosynthetic steps as other iridoids (Bellesia et al., 1983a, 1983b; Collu et al., 2001; Geu-Flores et al., 2012; Salim et al., 2014; Simkin et al., 2013) (Figure 1c). Iridoids are derived from geraniol which is generated from the monoterpenoid precursor geranyl pyrophosphate (GPP) catalyzed by geraniol synthase (GES). Subsequent oxidations catalyzed by geraniol 8-hydroxylase (G8H) and 8-hydroxygeraniol oxidoreductase (8HGO) cause formation of 8-oxogeranial. This linear dialdehyde is reduced by iridoid synthase (ISY), and subsequent cyclization of the activated intermediate (either with or without enzyme catalysis) results in the formation of nepetalactol, the universal precursor to iridoids. The pathway diverges at this point, with alcohol oxidation of nepetalactol providing nepetalactones (Hallahan et al., 1998; Lichman et al., 2019), whilst hydroxylation of nepetalactol by iridoid oxidase (IO) leads eventually to glycosides and alkaloids (Miettinen et al., 2014; Salim et al., 2014).

The mechanism of dolichodial formation is less obvious as it lacks the typical bicyclic iridoid structure. It has been proposed to be formed via the compound teucrein, also found in *T. marum* (Bellesia et al., 1983a, 1983b; Eisner et al., 2000) (Figure 1a,c). Dolichodial can be formed non-enzymatically from teucrein in mildly alkaline conditions (Bellesia et al., 1983a, 1983b), indicating that teucrein can be a precursor of dolichodial (Eisner et al., 2000). Structurally, teucrein appears to be derived from nepetalactol via a single hydroxylation and double acetylation (Figure 1c). Teucrein, which has no insect-repellent properties, is present at high concentrations in leaf buds yet dolichodial is present at high levels in mature leaves (Eisner et al., 2000).

In Lamiaceae, iridoid glycosides are abundant, including in *Teucreum* spp. (e.g. harpagide in *T. canadense*, Figure 1c), but volatile iridoids have a more irregular distribution, with nepetalactone unique to *Nepeta* spp. and teucrein/dolichodial unique to *T. marum* (Mint Evolutionary Genomics Consortium, 2018). Investigations into the evolution of *Nepeta* nepetalactone biosynthesis revealed that ISY, crucial for iridoid biosynthesis, was lost in the *Nepeta* lineage and then re-emerged from a different clade, in an example of independent, or parallel, evolution (Lichman et al., 2020). The lineage-specific formation of teucrein and dolichodial in *T. marum* may have also been

influenced by gene loss and gain, and other genomic processes.

We set out to discover the biosynthetic pathway to dolichodial in *T. marum* and to investigate its evolutionary origins. We report on the chromosome-scale assembly and annotation of the *T. marum* genome and gain insight into its evolutionary relationship with selected mint species. We use phylogenomics to identify gene candidates for the iridoid biosynthetic pathway and validate these with enzyme assays. We reveal the phylogenetic and genomic origins of the atypical *T. marum* ISY. Finally, we discover a combination of IO and newly reported acetyltransferases that result in the heterologous formation of diverse iridoids from nepetalactol in *Nicotiana benthamiana*. This highlights the inferred contribution of acetylation in *T. marum* iridoid biosynthesis and opens the door to the future heterologous production of the potent insect-repellent dolichodial and investigation into the animal pathway to the same compound.

RESULTS

Genome assembly

To sequence the genome of *T. marum*, we first estimated, by flow cytometry, the haploid genome size of *T. marum* to be 720 Mb. Subsequently, we generated 33.2 Gb of Oxford Nanopore Technology (ONT) long reads (≥10 kb) providing ~46× effective coverage (Table S1). These were assembled using Flye into 12 008 contigs (1.19 Gb total, N50 = 318 kb) (Table S2) (Kolmogorov et al., 2019). After error-correction, the polished assembly contained 7203 contigs (1.13 Gb, N50 = of 345 kb) (Table S2). The larger than anticipated assembly size and high duplicated Benchmarking Universal Single-Copy Orthologs (BUSCO) (55.5%) and heterozygosity revealed in the GenomeScope analysis indicates the presence of haplotigs in the assembly (Figure S2). Thus, the polished assembly was purged of haplotigs (Roach et al., 2018) reducing the contig number to 2499 (609 Mb, N50 = 428 kb) (Table S2).

Hi-C reads were used to scaffold contigs into 15 chromosome-scale pseudomolecules. The final assembly (ger.v1.0) comprised an N50 scaffold size of 32.9 Mb, with a total size of 610 Mb, with 515 Mb contained within the 15 pseudomolecules (Table 1; Figure S3). The quality of the ger.v1.0 assembly was confirmed by alignment of 99.97% of the Illumina WGS reads, of which, 90.6% were properly paired (Table S3). For mRNA-seq data, the overall alignment rates in paired-end mode ranged from 83.5% to 90.6% (Table S4). We also aligned cleaned ONT full-length cDNA reads to the ger.v1.0 assembly, revealing 96.89% of total mapped reads (Table S5). Benchmarking Universal Single-Copy Orthologs (Manni et al., 2021) analysis with 1614 total BUSCO genes revealed 1529 (94.7%) complete orthologs, of which, 1472 (91.2%) single-copy, 57 (3.5%) duplicated, 14 (0.9%) fragmented, and 71 (4.4%) missing orthologs (Table S6). These data demonstrate we generated a chromosome-scale genome assembly of *T. marum*.

Genome annotation

To annotate the genome assembly, a custom repeat library was constructed and used to mask 68% of the assembly as repetitive sequences, which included 31% retrotransposons and 2% DNA transposons. The repeat-masked assembly was annotated using the BRAKER pipeline (v2.1.3) (Hoff et al., 2019) and refinement of predicted gene models using PASA2 (v2.4.1) (Haas et al., 2008; Pham et al., 2020) resulting in a working set of 46 014 genes encoding 84 387 gene models (Table S7), of which, 73 468 gene models from 36 053 genes were identified as high confidence (Table S7). A total of 87% of the high confidence representative gene models were assigned to 15 pseudomolecules, and while the gene density was highly enriched toward the distal region of pseudomolecules ends, repeat density (>1 kb repetitive sequence) was opposite of gene density (Figure 2a). Karyotype analysis of *T. marum* revealed metacentric and submetacentric chromosomes (Castro & Rosselló, 2007), consistent with gene enrichment on only one end of some pseudomolecules. Assessment of the annotation using BUSCO orthologs revealed 92.5% complete in the working set, 91.9% complete in the high confidence set, and 91.6% complete in the high confidence representative set (87.9% complete and single copy, 3.7% duplicated, 2.4% fragmented, 6.0% missing) (Table S6).

We also performed a more comprehensive analysis of repetitive sequences in the *T. marum* genome assembly using the Extensive *de-novo* TE Annotator (EDTA) software

Table 1 Metrics of the final *Teucrium marum* L. genome assembly

Scaffold	
Total assembly size (bp)	610 309 236
Number	1308
N50 size (bp)	32 980 994
L50 count	8
Maximum length (bp)	58 464 141
Minimum length (bp)	1000
Number of pseudomolecules	15
Total pseudomolecule size (bp)	515 839 546
Number of unanchored scaffolds	1293
Total unanchored scaffold size (bp)	94 469 690
Percentage of anchored scaffolds (%)	85
GC content (%)	38.57
Ns (%)	0.17
Contig	
Total assembly size (bp)	609 272 736
Number	3381
N50 size (bp)	378 637
L50 count	393
Maximum length (bp)	7 154 970
Minimum length (bp)	1000

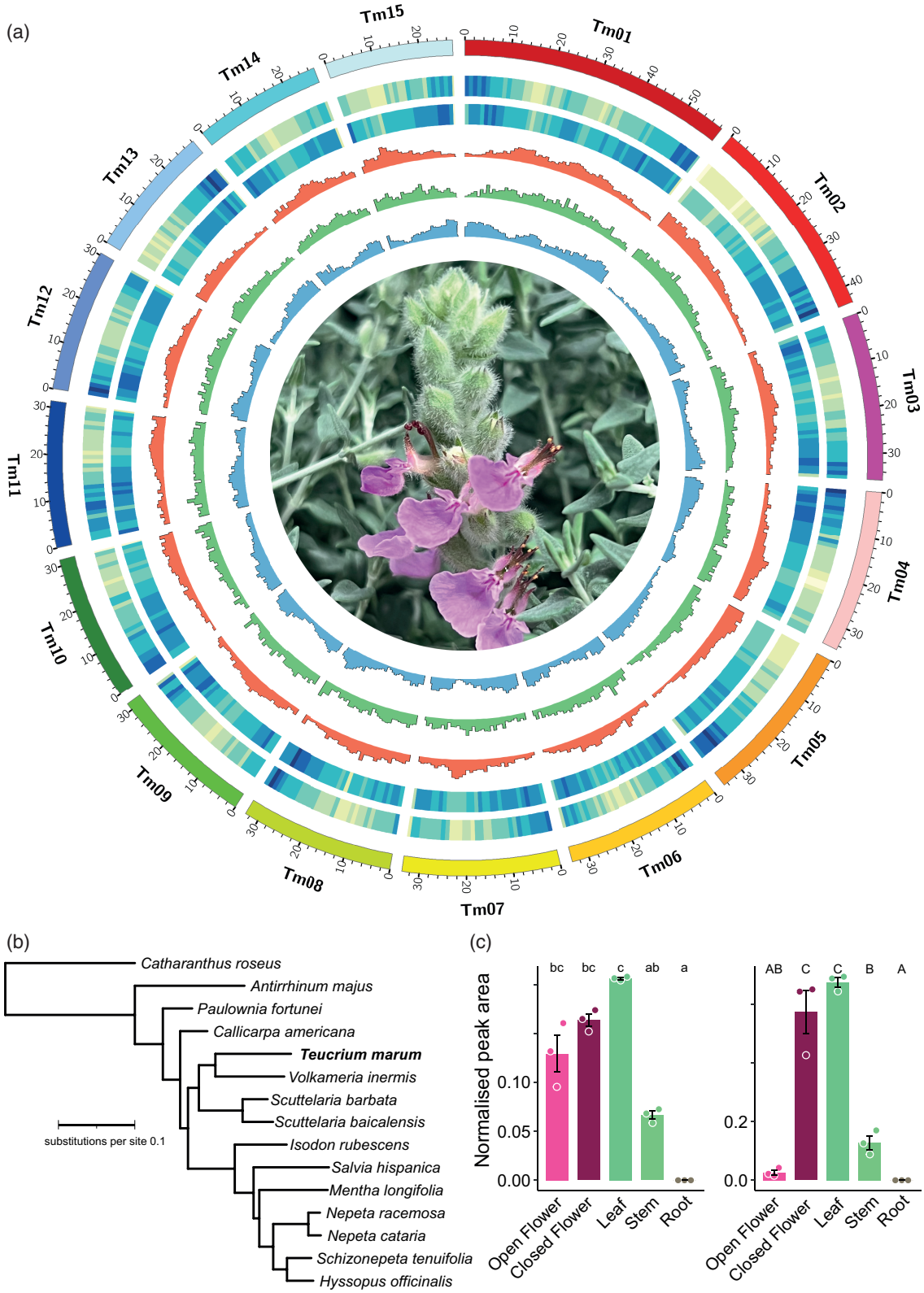

Figure 2. *Teucrium marum* genome sequence and metabolite analysis.
(a) Circos plot of the *Teucrium marum* genome assembly across 15 pseudomolecules in 1 Mb bins. Rings (outermost to innermost) represent 15 pseudomolecules in Mb, heatmap of gene density, total repeat density heatmap, *Gypsy* long terminal repeat retrotransposon (LTR) density, *Copia* LTR density, and *Mariner* DNA transposon density. Exemplar photo of *T. marum* flower is in the center.
(b) Species tree constructed using a supermatrix of single-copy genes derived from genome assemblies and maximum-likelihood tree inference. All branches have 100% support as judged by ultrafast bootstraps and SH-aLRT.
(c) Volatile iridoids (dolichodial and teucrein) concentrations across tissues measured by GC–MS. Each data point is a detected normalized peak area by comparison to a verified standard. Bar heights are means, with standard deviation error bars. Letter grouping is based on Tukey tests at $\alpha = 0.05$.

(Ou et al., 2019). A total of 66.8% of the genome assembly was annotated as a transposable element (Table S8; Figure 2a) with the dominant fraction of transposable elements being long terminal repeats (LTRs), of which, *Gypsy* LTRs occupied 23.2% of the genome assembly (119.9 Mb, 82 230 elements) while *Copia* LTRs occupied 11.0% of the genome assembly (56.8 Mb, 42 417 elements). There were few DNA elements (9.0%) with the exception of *Helitrons* which occupied 16.0% of the genome assembly (82.6 Mb) with 213 391 elements.

Orthogroup and species tree

To determine the species relationships of *T. marum*, we obtained 11 genome assemblies from Lamiaceae alongside two genome assemblies from the wider Lamiales, and the genome assembly of the outgroup *Catharanthus roseus* (Gentianales) (Table S9). Single-copy genes were obtained, their alignments concatenated and a maximum-likelihood tree inferred from the resulting super-matrix (Figure 2b). The resulting species tree had full support on each branch and the relationships matched previous species trees (Mint Evolutionary Genomics Consortium, 2018). *T. marum* was placed in a clade with *Volkameria inermis* (syn *Clerodendrum inerme*), with both genera being part of Ajugoideae (Mint Evolutionary Genomics Consortium, 2018) (Figure 2b).

Metabolite profiling

To identify the type and location of iridoids in *T. marum*, we performed metabolic profiling of open flowers, closed flowers, leaves, stems, and roots. GC–MS was used to explore the levels of the volatiles dolichodial and teucrein (Figure 2c; Figure S4) (Eisner et al., 2000). Dolichodial and teucrein were highest in closed flowers and leaves. Open flowers had high levels of dolichodial but very low quantities of teucrein, suggesting conversion of teucrein to dolichodial during flower development. We did not detect dolichodial and teucrein in roots, consistent with the production of essential oils only in aerial parts of *Teucrium* species (Djabou et al., 2013).

Whilst volatile iridoids like dolichodial and nepetalactone are thought to play a role in repelling insect herbivores, the structurally and biosynthetically related iridoid glucosides damage the gut of herbivores in a manner akin to the mustard oil bomb of glucosinolated (Dobler et al., 2011). Other species in the *Teucrium* genus such as *T. canadense* and *T. chamaedrys* are known to produce iridoid glucosides but do not produce volatile iridoids. No iridoid glucosides have been reported in *T. marum*. To examine whether volatile iridoids have usurped iridoid glucosides in *T. marum*, we investigated the presence of iridoid glucosides by high-resolution LC–MS. We identified putative iridoid glucosides by selecting peaks with high-resolution masses matching entries in an iridoid database, which demonstrated characteristic glucosyl fragmentation in the MS^2 spectrum (Rodríguez-López et al., 2022). We identified 15 compounds matching these criteria, primarily identified in aerial tissues (Table S10, Figure S5). The most abundant iridoid glucoside (by peak area) was annotated as 7-deoxyloganic acid based on mass and fragmentation pattern (Figure S6) (Aničić et al., 2021).

Identification of iridoid synthase

We set out to identify the genes involved in the biosynthesis of the volatile iridoids dolichodial and teucrein. The early iridoid biosynthetic pathway consists of the steps from geranyl pyrophosphate to nepetalactol, the first molecule containing the characteristic bicyclic iridoid scaffold (Figure 1c). Genes in this pathway were first identified in *C. roseus* (Miettinen et al., 2014) but appear to be orthologous across the majority of known iridoid-producing species in the asterids (Alagna et al., 2016; Kang et al., 2021; Lichman et al., 2020; Mint Evolutionary Genomics Consortium, 2018).

The first committed step into the iridoid pathway is catalyzed by iridoid synthase (ISY), the scaffold-forming enzyme, so we elected to first identify the gene encoding this enzyme (Figure 3a) (Geu-Flores et al., 2012). Orthogroup and BLAST analysis revealed two gene candidates closely related to *C. roseus* ISY (Teum.06G020140 and Teum.05G025680). These genes were cloned, and expressed in *E. coli* with a hexahistidine tag, and their protein products were purified through nickel affinity chromatography (Figure S7a). Both enzymes demonstrated *in vitro* NADPH-dependent reduction of 8-oxogeranial to the mixture of products, including *cis-trans*-nepetalactol, characteristic of this enzyme (Figure 3b; Figure S7) (Kries et al., 2017; Lichman et al., 2019). After 30 min of reaction, Teum.06G020140 produced more product and consumed more substrate than Teum.05G025680, though over 3 h no significant difference was observed (Figure S7b).

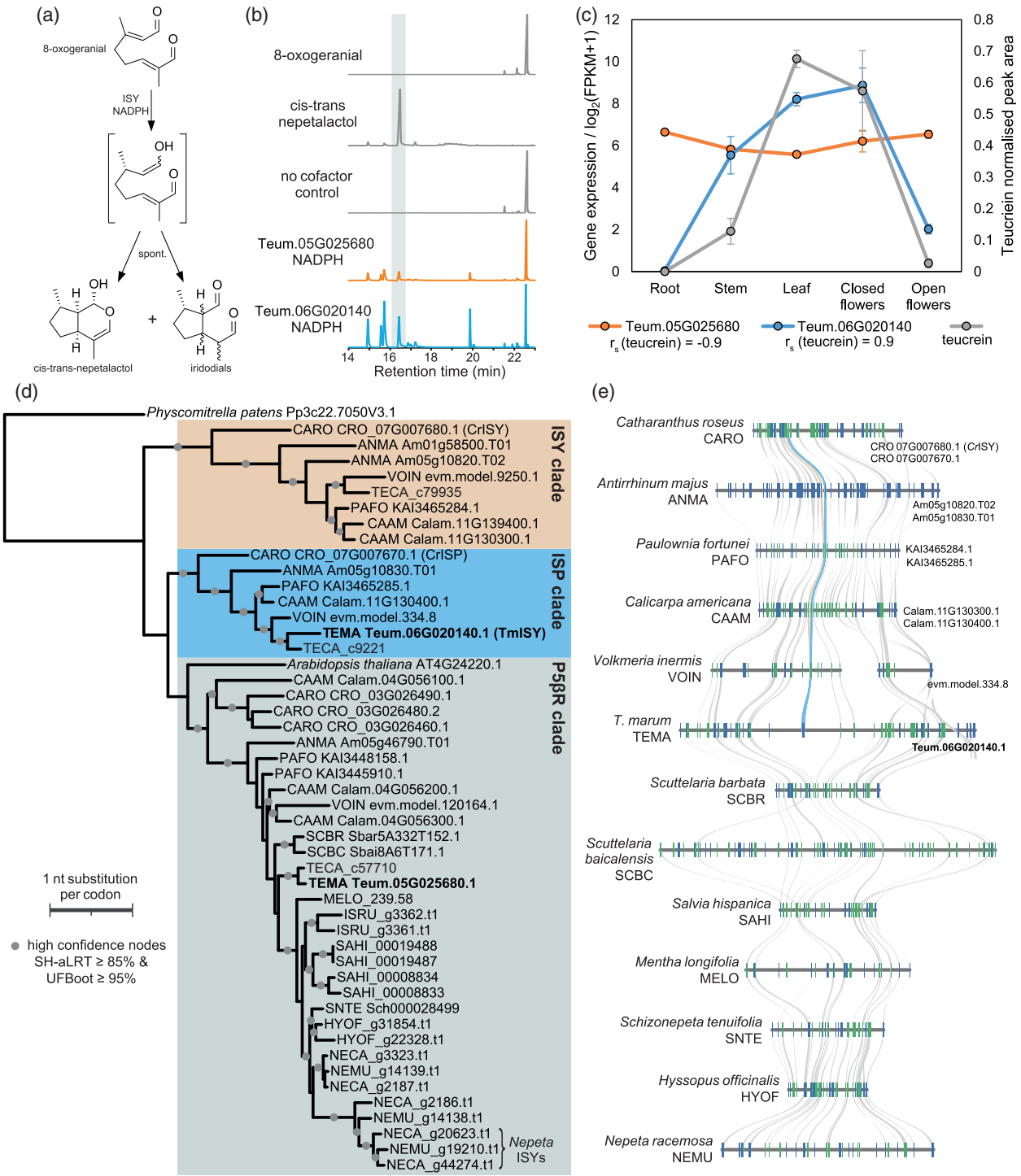

Figure 3. *T. marum* iridoid synthase.
(a) Scheme of *in vitro* ISY reaction.
(b) Conversion (3 h) 8-oxogeranial to products including nepetalactol by ISY candidates.
(c) Average gene expression levels of ISY candidates compared to teucrein accumulation as measured by GC–MS.
(d) Gene tree of PRISE family genes highlighting three major clades: ISY clade (yellow); iridoid synthase paralog (ISP) clade (blue); P5βR clade (gray), related to Arabidopsis VEP1 and containing *Nepeta* ISYs. Abbreviations on Table S9.
(e) Synteny analysis of ISY locus highlighting conservation or loss of tandem ISY and ISP. Blue and yellow lines show ISP and ISY syntenic orthologs respectively.

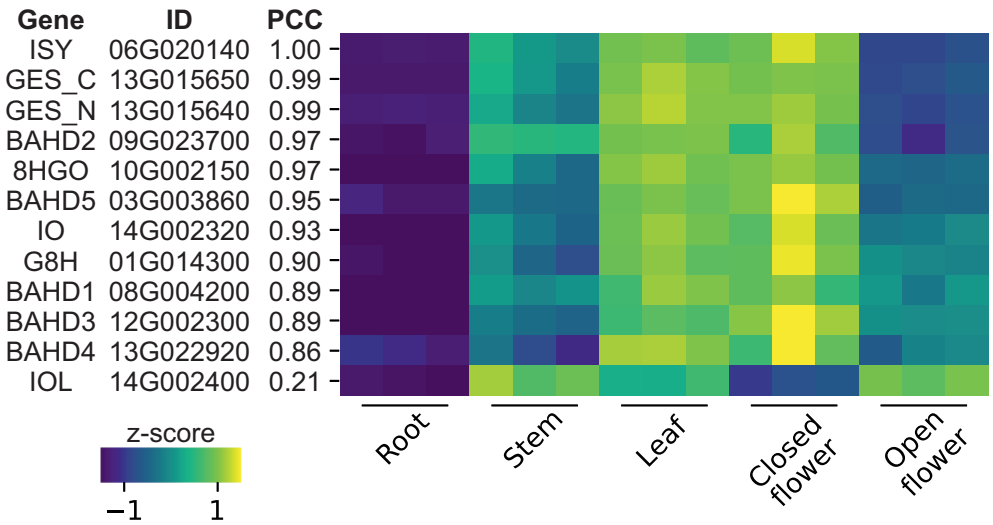

Figure 4. Co-expressed iridoid genes. Tissue gene expression pattern of key iridoid biosynthesis genes and candidates. Genes are ordered according to the Pearson correlation coefficient (PCC) relative to ISY.

Furthermore, the more active homolog, Teum.06G020140, had a pattern of expression matching teucrein accumulation (r_s = 0.9), with expression absent in root, low in stem and open flower, and high in leaf and closed flower (Figure 3c). The other gene, Teum.05G025680, did not show this pattern of expression (r_s = −0.9).

Iridoid synthase is part of the progesterone 5β-reductase/iridoid synthase-like enzymes (PRISE) family of short-chain reductase/dehydrogenases, and the ability to reduce 8-oxogeranial is a common promiscuous activity of PRISE family members beyond verified ISYs (Munkert et al., 2014; Nguyen & O'Connor, 2020). Analysis of orthologous groups of genes from related plants revealed that PRISE genes are grouped into three clades (Figure 3d). Iridoid synthase from *C. roseus* (CRO 07G007680, CrISY) and its tandem paralog (CRO 07G007670, CrISP) form the base for the ISY and ISP (iridoid synthase paralog) clade, respectively. Iridoid producers *Callicarpa americana*, *Antirrhinum majus*, and *Paulownia fortunei* (all Lamiales) similarly have tandemly duplicated orthologs in each of these clades that reside in a colinear region (Figure 3e). *T. marum* and *Volkameria inermis* (syn *Clerodendrum inerme*), however, only have a single gene in this colinear block with phylogenetic analysis revealing it to be a member of the ISP clade. *V. inermis* has a non-syntenic ortholog for ISY outside of the colinear region, whereas *T. marum* lacks an ISY ortholog entirely. Species in *Scutellaria* sp. and Nepetoideae do not make iridoids (except *Nepeta*) and all lack orthologs in the ISY or ISP clade. The third clade in the PRISE tree, comprises P5βR-like genes, except for the independently evolved *Nepeta* spp. ISYs (Lichman et al., 2020).

Analysis of a *T. canadense* transcriptome shows the presence of both ISY and ISP orthologs (c79935 and c9221, respectively) (Figure 3d) (Mint Evolutionary Genomics Consortium, 2018). However, without a genome assembly available it is unknown whether these genes are tandem or dispersed (as in *V. inermis*). Regardless, the presence of both genes in *T. canadense* indicates that the canonical ISY was lost specifically within the *T. marum* lineage, and its function was replaced by ISP (Teum.06G020140), now the *T. marum* iridoid synthase (*Tm*ISY). The role of ISP in typical iridoid-producing plants is unknown but, as shown here, they tend to co-occur as tandem duplicates (Figure 3e).

Therefore, in both the *Nepeta* lineage and the *T. marum* lineage, volatile iridoid biosynthesis emerged independently (Lichman et al., 2020). Furthermore, this was accompanied by the parallel evolution of ISYs, wherein the canonical ISY was lost and its function replaced by a PRISE enzyme from a different clade.

Early iridoid biosynthetic pathway

Using homology and co-expression analysis, we identified candidate genes for other steps in the early iridoid biosynthetic pathway (Figure 4). Cytochrome-P450 genes orthologous to G8H (Teum.01G014300) and IO (Teum.14G002320) were highly expressed and correlated tightly to ISY expression (PCC = 0.90 and 0.93). The gene orthologous to 8HGO was also highly expressed and had a high ISY expression correlation (PCC = 0.97). To verify its function, we cloned the gene, expressed it recombinantly in *E. coli*, and demonstrated the purified enzyme product could oxidize 8-hydroxygeraniol *in vitro* dependent on the NAD^+ cofactor (Figure S8).

Next, we identified a gene candidate for GES, the first step in the iridoid biosynthetic pathway. The genome

annotation incorrectly split the gene model in two (Teum.13G015640 for the N-terminus and Teum.13G015650 for the C-terminus), but despite this, both parts correlated with ISY expression (PCC = 0.99). Based on transcriptome assembly data and manual curation, the full gene model was determined (Table S11). This gene was cloned and the enzyme heterologously expressed in *E. coli* as a GST-tagged N-terminal truncate. The purified enzyme formed geraniol from geranyl pyrophosphate *in vitro*, validating its function as a geraniol synthase (Figure S9).

Iridoid oxidase

In typical iridoid glucoside biosynthesis, IO is hypothesized to catalyze the three-step oxidation from nepetalactol to deoxyloganetic acid via the alcohol and aldehyde intermediates (Miettinen et al., 2014; Salim et al., 2014). It is in this context that we propose that teucrein biosynthesis proceeds via a single oxidation of nepetalactol, a hydroxylation to 7-deoxyloganetic alcohol, without subsequent formation of the aldehyde or acid, catalyzed by an ortholog of IO (Figures 1c and 5a). We identified two genes in *T. marum* that cluster to the IO clade as defined by *C. roseus* IO (Teum.14G002320.1 and Teum.14G002400.1) (Figure S10a). Comparative genomics showed colinear regions harboring IO homologs throughout the Lamiaceae, with at least four copies of IO and its co-linear region present in the family's most recent common ancestor (Figure S10b). Only one *T. marum* IO paralog (Teum.14G002320) had high expression in iridoid-producing tissues and was co-expressed with other iridoid pathway genes (Figure 4; Figure S10), indicating it is the iridoid active enzyme. Its high similarity (97.4% pairwise nucleotide identity) and physical proximity to the other paralog (Teum.14G002400) indicates a local duplication of IO occurred in the *T. marum* lineage (purple clade of Figure S10b).

The IO candidate Teum.14G002320 was selected for recombinant expression and characterization. We initially attempted to use *Saccharomyces cerevisiae* to examine activity but with both *in vivo* and microsomal preparations we were unable to identify reaction products after feeding nepetalactol. The gene was then cloned and transiently expressed in *N. benthamiana* via agrobacteria transformation. Discs were excised from transformed leaves and fed with nepetalactol. In leaf discs transformed with Teum.14G002320, LC–MS analysis showed consumption of the substrate, and formation of a new peak (Figure 5, TmIO trace). This peak had a molecular mass and MS^2 fragmentation matching the glucose ester of the triple oxidation product 7-deoxyloganetic acid (Figure S11). We were unable to observe peaks related to the single oxidation product. It has previously been shown that *N. benthamiana* contains native glycosyltransferases that act on 7-deoxyloganetic acid causing the formation of this off-pathway product (Dudley et al., 2022).

Discovery of iridoid acetyltransferases

The final proposed step of the teucrein biosynthetic pathway is the double acetylation of the intermediate compound 7-deoxyloganetic alcohol (Figure 5a) (Eisner et al., 2000). We used co-expression analysis to identify acyltransferase candidates that correlate with ISY expression (Figure 4a). Five of these (BAHD1-5) were cloned and transiently co-expressed alone and with IO in *N. benthamiana*. We excised discs from transformed leaves and fed these with nepetalactol. In samples containing IO coexpressed with BAHD2, BAHD4, or BAHD5, we saw no notable differences compared to the IO-only sample (Figure 5). However, in samples coexpressing IO with BAHD1 or BAHD3 we were able to observe reduction in the peak area of the 7-deoxyloganetic acid glucose ester (Figure 5b,c), along with the accumulation of multiple peaks with 149.096 *m/z* ($C_{10}H_{13}O^+$) (Figure 5b; Figure S12). In our LC–MS positive mode detection method, iridoids undergo in-source dehydration, with the dominant ion being $[M\text{-}H_2O + H]^+$. Hence, we observe nepetalactol ($C_{10}H_{16}O_2$) as 151.112 *m/z* ($C_{10}H_{15}O^+$) and dolichodial ($C_{10}H_{14}O_2$) as 149.096 *m/z* (Figures S13 and S14). Therefore the peaks with 149.096 *m/z* are diverse oxidized iridoids, structurally related to dolichodial, that undergo in-source dehydration or fragmentation to yield the characteristic ion.

To identify the new 149.096 *m/z* compounds formed in the enzyme assays, we compared them to teucrein and dolichodial standards. We could not detect the formation of teucrein in our enzyme assays. The dolichodial standard appeared as two peaks in LC–MS due to its reaction with methanol during extraction (Figure S14). One peak from the dolichodial standard matched the retention time and MS^2 fragmentation pattern of one of the newly formed peaks (Figure 5b; Figure S15). To verify the identity of dolichodial produced in the enzyme assays, we spiked the standard into the assays and observed the product peak increase (Figure S15). Despite this evidence, we cannot be certain of the assignment due to the high density of isobaric compounds and the lack of a second dolichodial peak in the non-spiked assays.

The 149.096 *m/z* is a characteristic fragment of oxidized iridoids, and peaks containing this mass are likely to be structural isomers of dolichodial (e.g., dolicholactone and teucrimulactone). Specific BAHD acyltransferases plus IO are required for the accumulation of dolichodial isobars: they are very low in no enzyme, IO-only, or BAHD-only controls. Furthermore, the reduction of 7-deoxyloganetic acid glucose ester concentration indicates flux is being diverted away from the IO product by the BAHDs (Figure 5). This suggests that an oxidized and acylated intermediate is present in the pathway, presumably teucrein. However, as teucrein cannot be detected in our *N. benthamiana* enzyme assays we infer it presents as a 'cryptic' intermediate,

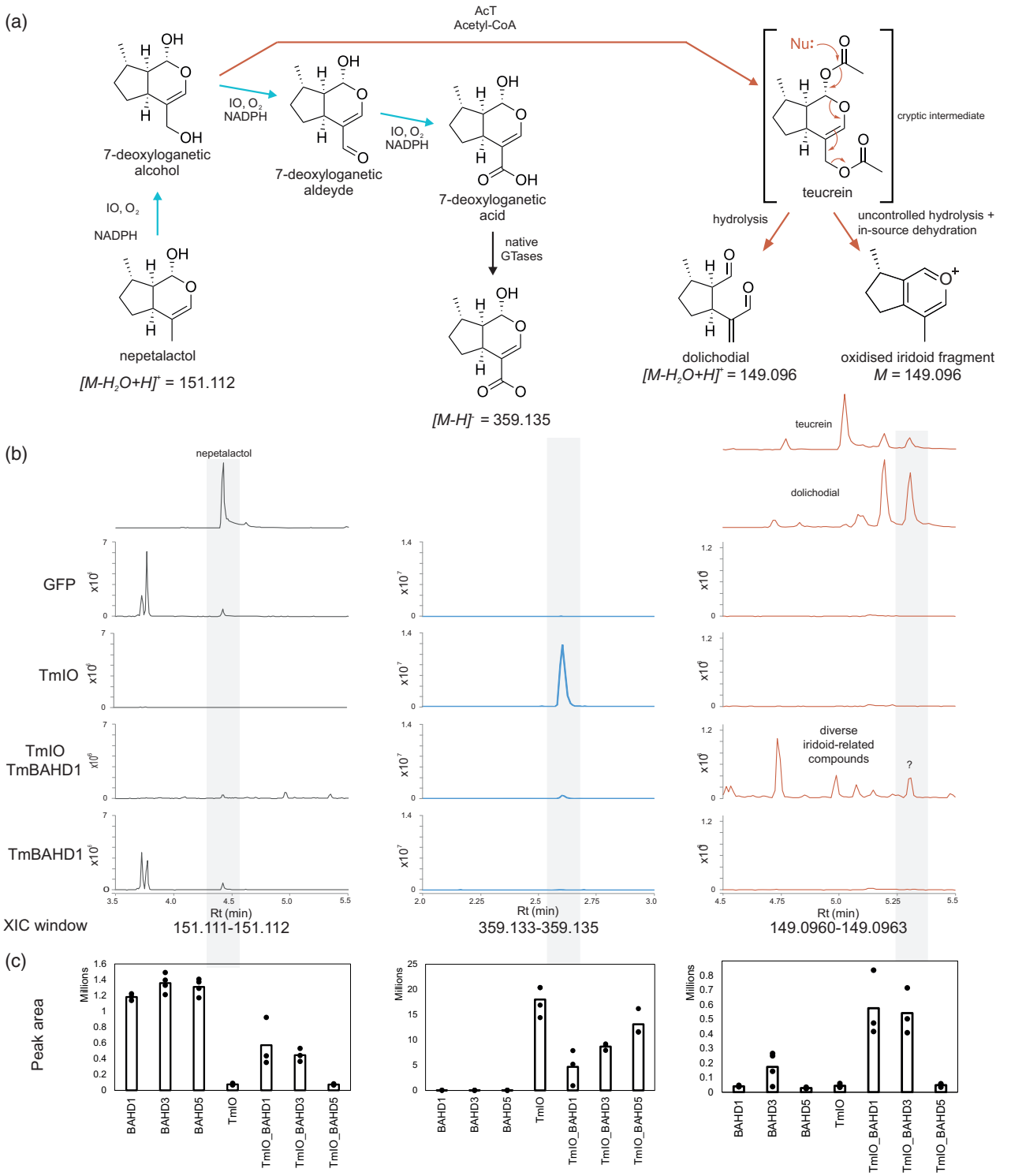

Figure 5. Formation of iridoids from nepetalactol in *N. benthamiana* leaf disc assays.
(a) Reaction scheme. Blue arrows: route to 7-deoxyloganetic Glu-ester via IO and promiscuous glucosyltransferase activity. Red arrows: route to dolichodial and diverse iridoids via coupled IO and acyltransferase activity and inferred teucrein intermediate.
(b) LC–MS analysis of metabolites present in *N. benthamiana* leaf discs. Left column has nepetalactol highlighted, peak confirmed by the verified standard (top). Middle column depicts a 7-deoxyloganetic acid Glu-ester (Figure S11). Right column features dolichodial isomer highlighted, with coeluting peak highlighted (see Figure S15). Nepetalactol and dolichodial standard peak intensities are scaled to enable comparison.
(C) Peak areas of nepetalactol, 7-deoxyloganetic Glu-ester, and iridoid peak coeluting with a dolichodial peak. Dots represent single experiments and bars are the mean of three or four samples (see Figure S12 areas of other iridoid peaks).

The Plant Journal, (2024), doi: 10.1111/tpj.16698

being formed and immediately hydrolyzed into dolichodial and other unknown isomers (Eisner et al., 2000). Promiscuous *N. benthamiana* enzymes may also be contributing to the chemical diversity formed. This data describes acyltransferases contributing to iridoid chemical diversity and their coordinated activity with IO.

DISCUSSION

We set out to investigate how *T. marum* produces iridoids including dolichodial, a potent insect repellant that is also produced *de novo* by insects, and teucrein, its proposed acetylated precursor. We obtained a chromosome-level assembly of *T. marum* that enabled identification of iridoid biosynthetic genes (GES, 8HGO, ISY, and IO), whose roles were verified via activity assays. Phylogenomic analysis revealed the evolutionary history of *Tm*ISY. Finally, we identified acyltransferases capable of working together with IO to produce multiple iridoids, including a peak matching the retention time and mass spectrum of dolichodial. This work describes partial genetic elucidation of the dolichodial biosynthesis pathway, providing future metabolic or biocatalytic access to this potent anti-insectan.

The similarity of *T. marum* (cat thyme) and *Nepeta* sp. (cat mint) was evident at the onset of this work, with both species making volatile iridoids with cat attractant and insect repellent properties (Birkett & Pickett, 2003), and both having acquired this ability relatively recently in their species and genus lineages respectively (Mint Evolutionary Genomics Consortium, 2018). We now reveal that these similarities are also observed in the parallel evolution of ISY, the iridoid scaffold-forming enzyme. In both species, the canonical ISY, orthologous in the majority of plant iridoid producers, is absent, and its function has been replaced by a PRISE enzyme from a different clade (Lichman et al., 2020). In the case of *Nepeta* sp., the new ISY originated from the P5βR clade, whilst in *T. marum* it was from the ISP clade (Figure 3d). Whilst the P5βR clade has no known association with iridoid biosynthesis beyond *Nepeta* sp., the ISP clade appears to be ancestrally related to iridoid biosynthesis.

ISP is conserved in the majority of iridoid-producing plants and typically appears as a tandem paralog to ISY (Figure 3e). In *C. roseus*, it has a similar catalytic efficiency to ISY and is co-expressed with iridoid biosynthetic genes (Li et al., 2023; Munkert et al., 2014). However, silencing ISP did not cause accumulation of putative iridoid precursors as silencing ISY did (Munkert et al., 2014). Whilst ISP appears to be functionally redundant, its conservation may indicate a specific but as yet unknown function. However, in *T. marum* this apparent backup ISY has taken on the major role.

Therefore, it seems that ISYs have had three origins from within the PRISE family: the canonical and presumably ancestral ISY, then two parallel cases of gene loss and acquisition of new function. Curiously, in both *Nepeta* sp. and *T. marum*, the gene evolution has occurred with a shift of end-point metabolite type or function (i.e., formation of volatile iridoids), which may relate to changes to gene regulation during metabolic evolution (Shoji, 2019). Iridoid synthase activity *in vitro* has been demonstrated in PRISE family enzymes across all land plants, including non-iridoid producers (Nguyen & O'Connor, 2020). This may account for multiple instances of functional iridoid synthases arising from this family.

Using the genome sequence, homology, and co-expression, we were able to identify IO and acyltransferases contributing to iridoid biosynthesis in *T. marum*. We examined their function by converting nepetalactol into diverse iridoids in co-infiltrated *N. benthamiana* leaf discs. Whilst we detected a peak with the same retention time and mass spectrum as the dolichodial standard, the complexity of the analytics prevented confident assignment of identity. The identity of the new peaks as iridoids is supported by the presence of a key mass fragment (149.096 *m/z*, Figure 5; Figure S12), MS^2 matches to validated iridoid standards (Figure S15), and concomitant reduction of the iridoid precursor (nepetalactol) and shunt product (7-deoxyloganetic acid glucose ester) (Figure 5).

Whilst the *N. benthamiana* system was instrumental in identifying coupled activity of IO and acyltransferases with nepetalactol, native enzymes appear to be interfering with the heterologous pathway, such as the native glucosyltransferases that modify the IO product (Figure 5) (Dudley et al., 2022; Miettinen et al., 2014). The absence of detectable acylated intermediates indicates native hydrolase activity, and the multiple iridoid peaks may also be indication of some other native enzyme activity. We are working toward increasing the yield and specificity of iridoid accumulation through *de novo* formation of nepetalactol and optimizing infiltration ratios. Modifying the chassis through genome editing might ultimately be required to reduce interfering native enzyme activities (Dudley et al., 2022). Further exploration of alternative expression systems may also be fruitful.

The challenges encountered with the *N. benthamiana* system highlight two aspects of the biosynthetic pathway with unresolved complexity: the oxidation of nepetalactol and the hydrolysis of teucrein (Figure 5a). The pathway requires the single hydroxylation of nepetalactol into deoxyloganetic alcohol. This is catalyzed by IO, but the same enzyme can also perform up to three oxidations on nepetalactol, forming deoxyloganetic acid. This over-oxidation is occurring in *N. benthamiana* assays with IO and nepetalactol in the absence of active acyltransferases (Figure 5b). To form teucrein/dolichodial, acyltransferases must capture the IO product after the first hydroxylation. Therefore, to optimize dolichodial

formation the overoxidation by IO must be minimized. It is possible such a balancing act is occurring in the native plant: we observed 7-deoxyloganic acid, a triple oxidation product, accumulating in the same tissues as teucrein (Figures S5 and S6). Teucrein accumulation correlates to IO expression, and it is therefore possible that IO catalyzes the formation of both single and triple oxidation products, at the branch point between volatile iridoids and iridoid glucosides (Figure 5a).

The control of this branch point could be via a complex mechanism. For example, there may be physical interactions between the BAHDs and IO which ensure single oxidation products are channeled into teucrein biosynthesis. Alternatively, there may be additional enzymes *in planta* that promote the second and third oxidation steps, as known in equivalent three-step oxidation in artemisinin biosynthesis (Paddon et al., 2013; Teoh et al., 2009). A complete understanding of how the oxidation of nepetalactol by IO is catalyzed will enable increased yields of dolichodial and also has implications in iridoid-derived alkaloid pathway metabolic reconstructions where putative helper enzymes have been employed to increase the efficiency of this oxidation step (Brown et al., 2015; Zhang et al., 2022).

The final step of dolichodial biosynthesis is hydrolysis of teucrein (Figures 1c and 5a). In our heterologous reconstructions, we could not detect teucrein, presumably because its hydrolysis was occurring spontaneously, either through chemical processes or promiscuous enzymatic hydrolases. It is also leading to multiple iridoid products. To control teucrein hydrolysis in heterologous systems, we must understand how it is catalyzed in *T. marum*. We found that teucrein to dolichodial conversion was enhanced in open compared to closed flowers (Figure 2c) and previous reports suggest it is increased as leaves mature (Eisner et al., 2000). These observations could aid in the future discovery of a specific hydrolase. The biosynthetic origin of dolichodial in insects is not known but presumably has similarities to insect nepetalactone biosynthesis (Köllner et al., 2022). However, teucrein has never been observed in insects and it is unknown whether they use this as an intermediate to dolichodial.

There are multiple examples of 'cryptic' acetylation steps in plant-specialized metabolite biosynthesis. In nopscapine biosynthesis, an acetyl group acts as a hydroxyl-protecting group prior to deacetylation and rearrangement (Dang et al., 2015). In other examples akin to dolichodial biosynthesis, the addition of an acetyl group converts a hydroxyl into a favorable leaving group. This occurs in the morphine biosynthesis pathway (Lenz & Zenk, 1995) and in Lycopodium alkaloid biosynthesis (Nett et al., 2023).

Iridoids are distributed widely in insects and in the asterid clade of flowering plants, and they have important roles in defense and communication. In this study, we have determined and reconstituted the *T. marum* enzymatic route to iridoids and have analyzed the evolutionary origins of key pathway genes. The unique contribution of acyltransferases to iridoid diversity in *T. marum* highlights how different species have employed different enzyme types and mechanisms to access diverse chemistries from a shared scaffold.

EXPERIMENTAL PROCEDURES

The *T. marum* genome was sequenced and assembled into chromosome-scale pseudomolecules using a combination of ONT, Illumina whole-genome shotgun, and Hi-C sequencing. Transcripts from different tissues were sequenced from cDNA libraries with both Illumina short-reads and ONT. These sequences were used, in combination with other databases and gene prediction algorithms, to obtain high-confidence gene models. Phylogenomic analysis was conducted using orthologous group identification, microsynteny analysis, and maximum likelihood tree inference. Chemical standards were isolated from *T. marum* tissues and validated with NMR spectroscopy. For metabolite profiling, tissues were extracted with hexane or methanol and analyzed by GC–MS or LC–MS respectively. Recombinant enzyme production was performed in *E. coli* or *N. benthamiana*. For *E. coli* expression, enzymes were isolated using His-tag purification, assayed *in vitro*, and the reaction analyzed by GC–MS. For *N. benthamiana* transient expression, plants were infiltrated with agrobacteria containing plasmids encoding genes of interest. Leaf discs of infiltrated plants were isolated, and incubated with substrates and then the metabolites were extracted and analyzed by LC–MS. Detailed methods and materials with associated references can be found in Supplementary Information.

AUTHOR CONTRIBUTIONS

SJS characterized pathway genes and performed metabolite profiling, synteny, and phylogenomic analysis; SA assembled and annotated the genome; BAR and WPU isolated and characterized standards; JPH aided assembly and analysis of genome; SL and TRL performed metabolite profiling and mass spectrometry analysis; CRB and BRL designed research and provided guidance; SJS, SA, CRB and BRL wrote the paper.

ACKNOWLEDGMENTS

We would like to thank Brieanne Vaillancourt, Joshua C. Wood, Mercy Kitavi, and Chenxin Li for their assistance with analyses. We thank Sarah O'Connor at the Max Planck Institute of Chemical Ecology for providing chemical standards, plasmids, and discussions. At the University of York, we acknowledge the Centre for Excellence in Mass Spectrometry for equipment and the Biosciences Technology Facility for metabolomics analysis. This project was undertaken on the Viking Cluster, which is a high-performance computer facility provided by the University of York. We are grateful for computational support from the University of York High-Performance Computing service, Viking, and the Research Computing team.

FUNDING INFORMATION

Funding for this project was provided by funds awarded to C.R.B. from the Georgia Research Alliance, Georgia Seed

Development, The University of Georgia, and Michigan State University. S.A. was supported by a fellowship from the Study Abroad Program of The Republic of Türkiye (Turkey) Ministry of National Education. S.J.S. was supported by the BBSRC (BB/V006452/1) and B.R.L. by UKRI (MR/S01862X/1).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. *Teucrium marum* plant tissues as used for RNA-seq and metabolite profiling.

Figure S2. Heterozygosity level of the *Teucrium marum* genome.

Figure S3. Hi-C contact map of the manually curated chromosome-scale assembly of *T. marum*.

Figure S4. Volatile metabolite analysis of *T. marum*.

Figure S5. Iridoid glucosides putatively identified via LC–MS.

Figure S6. Identification of 7-deoxyloganic acid.

Figure S7. PRISE expression and assay.

Figure S8. HGOA phylogeny and assay.

Figure S9. GES phylogeny and assay.

Figure S10. *T. marum* iridoid oxidase.

Figure S11. Putative identification of 7-deoxyloganetic acid glucose ester.

Figure S12. Acetylation-dependent formation of iridoids.

Figure S13. Mass spectrometry behavior of iridoid standards.

Figure S14. Mass spectrometry behavior of dolichodial standard.

Figure S15. Comparison of putative dolichodial product and standard.

Figure S16. NMR spectrum of teucrein.

Figure S17. NMR spectra of dolichodial.

Figure S18. NMR spectra of dolichodial and anisomorphal.

Figure S19. NOESY spectrum of teucrein.

Figure S20. NOESY spectrum of dolichodial.

Table S1. Genomic DNA libraries used in this study.

Table S2. Assembly contiguity metrics for initial, polished, and purged assembly of *Teucrium marum*.

Table S3. Genomic read (Illumina) alignment metrics to the *Teucrium marum* genome assembly.

Table S4. mRNA-Seq libraries used in this study.

Table S5. ONT cDNA-Seq libraries used in this study.

Table S6. Benchmarking universal single-copy orthologs (BUSCO).

Table S7. Gene annotation metrics.

Table S8. Summary of transposable elements in *Teucrium marum* identified via the EDTA pipeline.

Table S9. Species names and four-letter codes used. References to the papers describing the genomes are provided.

Table S10. Putative iridoid glucosides identified via LC–MS.

Table S11. Primers and expression vectors used in the study.

Dataset S1. FPKM gene expression abundances.

Dataset S2. MCScan synteny analysis.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at: The genome assembly, annotation files and gene expression abundances will be available as supplemental datasets in Figshare (DOI: https://doi.org/10.6084/m9.figshare.25109411). The metabolomics and other mass spectrometry data will be available in MetaboLights (https://www.ebi.ac.uk/metabolights/MTBLS9357). The raw reads for genome and RNAseq will be available in the National Center for Biotechnology Information Sequence Read Archive under BioProject ID PRJNA1076827 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1076827/, SRA SRX23636934-SRX23636940).

DATA AVAILABILITY STATEMENT

Raw read sequences for all generated data in this study are available in the National Center for Biotechnology Information Sequence Read Archive under BioProject ID (PRJNA1076827). The genome assembly, annotation files, and gene expression abundances will be available as supplemental datasets in Figshare (DOI: 10.6084/m9.figshare.25109411). Metabolomics and mass-spectrometry enzyme activity data are available on MetaboLights (Study Identifier: MTBLS9357). Sequences of characterized genes/enzymes are available on GenBank (accessions to be available upon publication): OR570604 (Tm8HGO), OR570605 (TmBAHD1), OR570606 (TmBAHD3), OR570607 (TmGES), OR570608 (TmIO), OR570609 (TmISY) and OR570610 (TmPRISEtrunc).

REFERENCES

Alagna, F., **Geu-Flores, F.**, **Kries, H.**, **Panara, F.**, **Baldoni, L.**, **O'Connor, S.E.** ***et al.*** (2016) Identification and characterization of the iridoid synthase involved in oleuropein biosynthesis in olive (*Olea europaea*) fruits. *The Journal of Biological Chemistry*, **291**, 5542–5554.

Aničić, N., **Gašić, U.**, **Lu, F.**, **Ćirić, A.**, **Ivanov, M.**, **Jevtić, B.** ***et al.*** (2021) Antimicrobial and immunomodulating activities of two endemic nepeta species and their major iridoids isolated from natural sources. *Pharmaceuticals*, **14**, 414. Available from: https://doi.org/10.3390/ph14050414

Bellesia, F., **Pagnoni, U.M.**, **Pinetti, A.** & **Trave, R.** (1983a) Teucrein, a new iridolactol from *Teucrium marum*, and its biosynthetic relationship with dolichodial. *Journal of Chemical Research Synopsis*, **12**, 328–329.

Bellesia, F., **Pagnoni, U.M.**, **Pinetti, A.** & **Trave, R.** (1983b) The biosynthesis of dolichodial in *Teucrium marum*. *Phytochemistry*, **22**, 2197–2201.

Beran, F., **Köllner, T.G.**, **Gershenzon, J.** & **Tholl, D.** (2019) Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. *The New Phytologist*, **223**, 52–67.

Birkett, M.A. & **Pickett, J.A.** (2003) Aphid sex pheromones: from discovery to commercial production. *Phytochemistry*, **62**, 651–656.

Bol, S., **Caspers, J.**, **Buckingham, L.**, **Anderson-Shelton, G.D.**, **Ridgway, C.**, **Buffington, C.A.T.** ***et al.*** (2017) Responsiveness of cats (Felidae) to silver vine (*Actinidia polygama*), Tatarian honeysuckle (*Lonicera tatarica*), valerian (*Valeriana officinalis*) and catnip (*Nepeta cataria*). *BMC Veterinary Research*, **13**, 70.

Brown, S., **Clastre, M.**, **Courdavault, V.** & **O'Connor, S.E.** (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 3205–3210.

Castro, M. & **Rosselló, J.A.** (2007) Karyological observations on plant taxa endemic to the Balearic Islands. *Botanical Journal of the Linnean Society*, **153**, 463–476.

Cavill, G.W.K. & **Hinterberger, H.** (1961) The chemistry of ants. V. Structure and reactions of dolichodial. *Australian Journal of Chemistry*, **14**, 143–149.

Collu, G., **Unver, N.**, **Peltenburg-Looman, A.M.**, **van der Heijden, R.**, **Verpoorte, R.** & **Memelink, J.** (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Letters*, **508**, 215–220.

Dang, T.-T.T., **Chen, X.** & **Facchini, P.J.** (2015) Acetylation serves as a protective group in noscapine biosynthesis in opium poppy. *Nature Chemical Biology*, **11**, 104–106.

Dawson, G.W., **Griffiths, D.C.**, **Janes, N.F.**, **Mudd, A.**, **Pickett, J.A.**, **Wadhams, L.J.** ***et al.*** (1987) Identification of an aphid sex pheromone. *Nature*, **325**, 614–616.

Dewhirst, S.Y., **Birkett, M.A.**, **Fitzgerald, J.D.**, **Stewart-Jones, A.**, **Wadhams, L.J.**, **Woodcock, C.M.** ***et al.*** (2008) Dolichodial: a new aphid sex pheromone component? *Journal of Chemical Ecology*, **34**, 1575–1583.

Djabou, N., **Lorenzi, V.**, **Guinoiseau, E.**, **Andreani, S.**, **Giuliani, M.C.**, **Desjobert, J.M.** ***et al.*** (2013) Phytochemical composition of Corsican *Teucrium* essential oils and antibacterial activity against foodborne or toxi-infectious pathogens. *Food Control*, **30**, 354–363.

Dobler, S., **Petschenka, G.** & **Pankoke, H.** (2011) Coping with toxic plant compounds – the insect's perspective on iridoid glycosides and cardenolides. *Phytochemistry*, **72**, 1593–1604.

Dudley, Q.M., **Jo, S.**, **Guerrero, D.A.S.**, **Chhetry, M.**, **Smedley, M.A.**, **Harwood, W.A.**, **Sherden, N.H.**, **O'Connor, S.E.**, **Caputi, L.** & **Patron, N.J.** (2022) Reconstitution of monoterpene indole alkaloid biosynthesis in genome engineered *Nicotiana benthamiana*. *Communications Biology*, **5**, 949.

Eisner, T. (1964) Catnip: its raison d'Être. *Science*, **146**, 1318–1320.

Eisner, T., **Eisner, M.**, **Aneshansley, D.J.**, **Wu, C.-L.** & **Meinwald, J.** (2000) Chemical defense of the mint plant, *Teucrium marum* (Labiatae). *Chemoecology*, **10**, 211–216.

El-Naggar, L.J. & **Beal, J.L.** (1980) Iridoids. A review. *Journal of Natural Products*, **43**, 649–707.

Geu-Flores, F., **Sherden, N.H.**, **Courdavault, V.**, **Burlat, V.**, **Glenn, W.S.**, **Wu, C.** ***et al.*** (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, **492**, 138–142.

Haas, B.J., **Salzberg, S.L.**, **Zhu, W.**, **Pertea, M.**, **Allen, J.E.**, **Orvis, J.** ***et al.*** (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**, R7.

Hallahan, D.L., **West, J.M.**, **Smiley, D.W.M.** & **Pickett, J.A.** (1998) Nepetalactol oxidoreductase in trichomes of the catmint *Nepeta racemosa*. *Phytochemistry*, **48**, 421–427.

Hoff, K.J., **Lomsadze, A.**, **Borodovsky, M.** & **Stanke, M.** (2019) Whole-genome annotation with BRAKER. In: Kollmar, M. (Ed.) *Gene prediction: methods and protocols*. Springer New York: New York, NY, pp. 65–95.

Jensen, S.R. (1991) Plant iridoids, their biosynthesis and distribution in angiosperms. *Phytochemistry Reviews*, **31**, 33–158.

Kang, M., **Fu, R.**, **Zhang, P.**, **Lou, S.**, **Yang, X.**, **Chen, Y.** ***et al.*** (2021) A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nature Communications*, **12**, 3531.

Köllner, T.G., **David, A.**, **Luck, K.**, **Beran, F.**, **Kunert, G.**, **Zhou, J.-J.** ***et al.*** (2022) Biosynthesis of iridoid sex pheromones in aphids. *Proceedings of the National Academy of Sciences of the United States of America*, **119**, e2211254119.

Kolmogorov, M., **Yuan, J.**, **Lin, Y.** & **Pevzner, P.A.** (2019) Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, **37**, 540–546.

Kries, H., **Kellner, F.**, **Kamileen, M.O.** & **O'Connor, S.E.** (2017) Inverted stereocontrol of iridoid synthase in snapdragon. *The Journal of Biological Chemistry*, **292**, 14659–14667.

Lenz, R. & **Zenk, M.H.** (1995) Acetyl coenzyme A: salutaridinol-7-O-acetyltransferase from *Papaver somniferum* plant cell cultures: the enzyme catalyzing the formation of thebaine in morphine biosynthesis. *The Journal of Biological Chemistry*, **270**, 31091–31096.

Li, C., **Wood, J.C.**, **Vu, A.H.**, **Hamilton, J.P.**, **Rodriguez Lopez, C.E.**, **Payne, R.M.E.** ***et al.*** (2023) Single-cell multi-omics in the medicinal plant *Catharanthus roseus*. *Nature Chemical Biology*, **19**, 1031–1104. Available from: https://doi.org/10.1038/s41589-023-01327-0

Lichman, B.R., **Godden, G.T.**, **Hamilton, J.P.**, **Palmer, L.**, **Kamileen, M.O.**, **Zhao, D.** ***et al.*** (2020) The evolutionary origins of the cat attractant nepetalactone in catnip. *Science Advances*, **6**, eaba0721.

Lichman, B.R., **Kamileen, M.O.**, **Titchiner, G.R.**, **Saalbach, G.**, **Stevenson, C.E.M.**, **Lawson, D.M.** ***et al.*** (2019) Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis. *Nature Chemical Biology*, **15**, 71–79.

Manni, M., **Berkeley, M.R.**, **Seppey, M.**, **Simão, F.A.** & **Zdobnov, E.M.** (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, **38**, 4647–4654.

Meinwald, J., **Chadha, M.S.**, **Hurst, J.J.** & **Eisner, I.** (1962) Defense mechanisms of arthropods – IX anisomorphal, the secretion of a phasmid insect. *Tetrahedron Letters*, **3**, 29–33.

Melo, N., **Capek, M.**, **Arenas, O.M.**, **Afify, A.**, **Yilmaz, A.**, **Potter, C.J.** ***et al.*** (2021) The irritant receptor TRPA1 mediates the mosquito repellent effect of catnip. *Current Biology*, **31**, 1988–1994.e5.

Miettinen, K., **Dong, L.**, **Navrot, N.**, **Schneider, T.**, **Burlat, V.**, **Pollier, J.** ***et al.*** (2014) The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications*, **5**, 3606.

Mint Evolutionary Genomics Consortium. (2018) Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the evolution of chemical diversity in Lamiaceae. *Molecular Plant*, **11**, 1084–1096.

Munkert, J., **Pollier, J.**, **Miettinen, K.**, **van Moerkercke, A.**, **Payne, R.**, **Müller-Uri, F.** ***et al.*** (2014) Iridoid synthase activity is common among the plant progesterone 5beta-reductase family. *Molecular Plant*, **8**, 136–152.

Nett, R.S., **Dho, Y.**, **Tsai, C.**, **Passow, D.**, **Martinez Grundman, J.**, **Low, Y.-Y.** ***et al.*** (2023) Plant carbonic anhydrase-like enzymes in neuroactive alkaloid biosynthesis. *Nature*, **624**, 182–191.

Nguyen, T.-D. & **O'Connor, S.E.** (2020) The progesterone 5β-reductase/iridoid synthase family: a catalytic reservoir for specialized metabolism across land plants. *ACS Chemical Biology*, **15**, 1780–1787.

O'Connor, S.E. & **Maresh, J.J.** (2006) Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Natural Product Reports*, **23**, 532–547.

Ou, S., **Su, W.**, **Liao, Y.**, **Chougule, K.**, **Agda, J.R.A.**, **Hellinga, A.J.** ***et al.*** (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, **20**, 275.

Paddon, C.J., **Westfall, P.J.**, **Pitera, D.J.**, **Benjamin, K.**, **Fisher, K.**, **McPhee, D.** ***et al.*** (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496**, 528–532.

Pagnoni, U.M., **Pinetti, A.**, **Trave, R.** & **Garanti, L.** (1976) Monoterpenes of *Teucrium marum*. *Australian Journal of Chemistry*, **29**, 1375–1381.

Partridge, S.J., **Withall, D.M.**, **Caulfield, J.C.**, **Pickett, J.A.**, **Stockman, R.A.**, **Oldham, N.J.** ***et al.*** (2021) Iridoid sex pheromone biosynthesis in aphids mimics iridoid-producing plants. *Chemistry - A European Journal*, **27**, 7231–7234. Available from: https://doi.org/10.1002/chem.202001356

Pham, G.M., **Hamilton, J.P.**, **Wood, J.C.**, **Burke, J.T.**, **Zhao, H.**, **Vaillancourt, B.** ***et al.*** (2020) Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience*, **9**, giaa100. Available from: https://academic.oup.com/gigascience/article/9/9/giaa100/5910251 [Accessed 27th November 2021].

Poli, F., **Serrilli, A.M.**, **Scartezzini, P.**, **Muzzoli, M.**, **Maxia, A.**, **Ballero, M.** ***et al.*** (2007) Endemic species of sardo-corso-balearic area: molecular composition and biological assay of *Teucrium*. *Natural Product Research*, **21**, 1061–1066.

Roach, M.J., **Schmidt, S.A.** & **Borneman, A.R.** (2018) Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, **19**, 460.

Rodríguez-López, C.E., **Jiang, Y.**, **Kamileen, M.O.**, **Lichman, B.R.**, **Hong, B.**, **Vaillancourt, B.** ***et al.*** (2022) Phylogeny-aware chemoinformatic analysis of chemical diversity in Lamiaceae enables iridoid pathway assembly and discovery of aucubin synthase. *Molecular Biology and Evolution*, **39**, msac057. Available from: https://doi.org/10.1093/molbev/msac057

Rosenkranz, M., **Chen, Y.**, **Zhu, P.** & **Vlot, A.C.** (2021) Volatile terpenes - mediators of plant-to-plant communication. *The Plant Journal*, **108**, 617–631.

Salim, V., **Wiens, B.**, **Masada-Atsumi, S.**, **Yu, F.** & **De Luca, V.** (2014) 7-deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in *Catharanthus roseus* iridoid biosynthesis. *Phytochemistry*, **101**, 23–31.

Shoji, T. (2019) The recruitment model of metabolic evolution: Jasmonate-responsive transcription factors and a conceptual model for the evolution of metabolic pathways. *Frontiers in Plant Science*, **10**, 560.

Simkin, A.J., **Miettinen, K.**, **Claudel, P.**, **Burlat, V.**, **Guirimand, G.**, **Courdavault, V.** ***et al.*** (2013) Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry*, **85**, 36–43.

Teoh, K.H., **Polichuk, D.R.**, **Reed, D.W.** & **Covello, P.S.** (2009) Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany*, **87**, 635–642.

Uenoyama, R., **Miyazaki, T.**, **Hurst, J.L.**, **Beynon, R.J.**, **Adachi, M.**, **Murooka, T.** ***et al.*** (2021) The characteristic response of domestic cats to plant iridoids allows them to gain chemical defense against mosquitoes. *Science Advances*, **7**, eabd9135. Available from: https://doi.org/10.1126/sciadv.abd9135

Zhang, J., **Hansen, L.G.**, **Gudich, O.**, **Viehrig, K.**, **Lassen, L.M.M.**, **Schrübbers, L.** ***et al.*** (2022) A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature*, **609**, 341–347.