RESEARCH ARTICLE

American Society of Plant Biologists | SEB SOCIETY FOR EXPERIMENTAL BIOLOGY | WILEY

Transcriptome-based identification and functional characterization of iridoid synthase involved in monotropein biosynthesis in blueberry

Lovely Mae F. Lawas[1] | Mohamed O. Kamileen[2] | C. Robin Buell[3,4] | Sarah E. O'Connor[2] | Courtney P. Leisner[1,5]

[1]Department of Biological Sciences, Auburn University, Auburn, Alabama, USA

[2]Department of Natural Product Biosynthesis, Max Planck Institute for Chemical Ecology, Jena, Germany

[3]Department of Plant Biology, Michigan State University, East Lansing, Michigan, USA

[4]Department of Crop and Soil Sciences, Institute of Plant Breeding, Genetics, & Genomics, University of Georgia, Athens, Georgia, USA

[5]School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, Virginia, USA

Correspondence
Courtney P. Leisner, School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA 24061, USA.
Email: cleisner@vt.edu

Abstract

Blueberries (*Vaccinium* spp.) are well known for their nutritional quality, and recent work has shown that *Vaccinium* spp. also produce iridoids, which are specialized metabolites with potent health-promoting benefits. The iridoid glycoside monotropein, which has anti-inflammatory and antinociceptive activities, has been detected in several wild blueberry species but in only a few cultivated highbush blueberry cultivars. How monotropein is produced in blueberry and the genes involved in its biosynthesis remain to be elucidated. Using a monotropein-positive (M+) and monotropein-negative (M−) cultivar of blueberry, we employed transcriptomics and comparative genomics to identify candidate genes in the blueberry iridoid biosynthetic pathway. Orthology analysis was completed using de novo transcript assemblies for both the M+ and M− blueberry cultivars along with the known iridoid-producing plant species *Catharanthus roseus* to identify putative genes involved in key steps in the early iridoid biosynthetic pathway. From the identified orthologs, we functionally characterized iridoid synthase (ISY), a key enzyme involved in formation of the iridoid scaffold, from both the M+ and M− cultivars. Detection of nepetalactol suggests that *ISY* from both the M+ and M− cultivars produce functional enzymes that catalyze the formation of iridoids. Transcript accumulation of the putative *ISY* gene did not correlate with monotropein production, suggesting other genes in the monotropein biosynthetic pathway may be more directly responsible for differential accumulation of the metabolite in blueberry. Mutual rank analysis revealed that *ISY* is co-expressed with UDP-glucuronosyltransferase, which encodes an enzyme downstream of the ISY step. Results from this study contribute new knowledge in our understanding of iridoid biosynthesis in blueberry and could lead to development of new cultivars with increased human health benefits.

KEYWORDS
(cultivated) blueberry, iridoid biosynthesis, iridoid synthase, monotropein, transcriptomics

1 | INTRODUCTION

Blueberry, a species in the genus *Vaccinium* which is native to North America, is grown globally and has seen an increase in both harvested area and yield over the last decades (Food and Agriculture Organization [FAO], 2022). This can be attributed to its economic importance and well-known nutritional quality. Blueberry fruits contain health-promoting compounds such as anthocyanins and flavonoids (Moyer et al., 2002; Rossi et al., 2022; Skrovankova et al., 2015), which have antioxidant properties that help prevent chronic diseases such as cancer, diabetes, cardiovascular, and neurological disorders (Grace et al., 2009; Seeram et al., 2006; Zafra-Stone et al., 2007). Additionally, blueberries and other *Vaccinium* species contain iridoid compounds (Heffels et al., 2017; Jensen et al., 2002; Leisner et al., 2017; Ma et al., 2013; Swiatek & Komorowski, 1972), a class of plant natural products with known human health benefits (Dinda et al., 2011; Tundis et al., 2008; Viljoen et al., 2012; Wang et al., 2020). Recent work has detected the iridoid glycoside monotropein in blueberry (Leisner et al., 2017). In a survey of 13 wild and 71 cultivated blueberries, it was shown that all wild *Vaccinium* species contained monotropein, yet this compound was only present in select cultivated blueberries (Leisner et al., 2017). The highest abundance of monotropein was found in floral buds and stems but was also present in all tissues examined in cultivars that were monotropein-positive (M+). On the other hand, monotropein-negative (M−) cultivars did not produce the compound in any tissues surveyed (Leisner et al., 2017) with a cutoff for quantitation (10 ng/mg) for classifying cultivars as M+ or M−. Most of the M+ cultivars were of the southern highbush ecotype, while the majority of M− cultivars surveyed were northern highbush. Monotropein is also found in other plant species, including Indian mulberry (*Morinda officinalis*), where it has been shown to have antinociceptive, anti-inflammatory, anti-osteoporosis, and wound healing properties (Chen et al., 2020; Choi et al., 2005; Karna et al., 2019; Shen et al., 2018; Wang et al., 2018; Zhang et al., 2016). Due to its potential positive human health benefits and differential production across blueberry cultivars, it is of interest to understand more about the biosynthesis of this metabolite at the molecular level.

To identify key genes in iridoid biosynthesis in blueberry, transcriptomics, comparative genomics, and biochemistry were employed. The use of multiple "omic" tools have been used in elucidation of the monoterpene indole alkaloid (MIA) pathway in the model medicinal plant *Catharanthus roseus* (Kellner et al., 2015; Miettinen et al., 2014; van Moerkercke et al., 2013), as well as identification of a key

cyclization enzyme (iridoid synthase, ISY) in iridoid biosynthesis in olive and *Nepeta* species (Alagna et al., 2016; Sherden et al., 2018). Outside of the iridoid biosynthetic pathway, comparative genomics has been critical in identification of genes in the terpene biosynthetic pathway in tomato (Matsuba et al., 2013) and the glucosinolate pathway in *Arabidopsis* (Bekaert et al., 2012). Furthermore, it has been observed that specialized metabolite biosynthetic pathway genes have the proclivity to be physically clustered in plant genomes (DellaPena and O'Connor, 2012). Taken together, the use of comparative genomics integrated with transcriptomics and metabolite profile analysis is a powerful tool to identify genes involved in the biosynthesis of specialized metabolites.

Among plants, the most complete elucidation of the iridoid biosynthetic pathway is in *C. roseus* (Collu et al., 2001; Geu-Flores et al., 2012; Irmler et al., 2000; Miettinen et al., 2014; Murata et al., 2008; Simkin et al., 2013). The iridoid pathway stems from the methylerythritol phosphate (MEP) pathway through the conversion of geranyl pyrophosphate (GPP) to geraniol in the early iridoid pathway. The early portion of the iridoid pathway is then linked to the late iridoid pathway by the activity of ISY (Mint Evolutionary Genomics Consortium, 2018). Beyond this first committed step in the formation of the iridoid skeleton by ISY, tailoring enzymes account for the biosynthesis of specific iridoid compounds.

ISY is part of the larger progesterone 5β-reductase (P5βR)/iridoid synthase (PRISE) family (Nguyen & O'Connor, 2020; Petersen et al., 2016) and performs a key transition step by conducting reductive cyclization that generates the iridoid ring scaffold, converting 8-oxogeranial to nepetalactol (Geu-Flores et al., 2012). Due to its role in this key step in iridoid biosynthesis, we hypothesize that either the presence or absence of the *ISY* gene or its gene expression in blueberry may account for differential monotropein production among cultivars. Validation of the putative *ISY* gene identified from targeted orthology analysis was carried out using heterologous expression in *Escherichia coli* and in vitro end-point enzyme assays, as this has proven effective to validate ISY in other plant species (Alagna et al., 2016; Geu-Flores et al., 2012; Lichman et al., 2020), as well as monoterpene synthase in *Camelina sativa* (Borghi & Xie, 2018). We also propose here a possible pathway for monotropein biosynthesis (Figure 1). Overall, elucidation of the iridoid biosynthetic pathway in blueberry would increase our basic knowledge of natural product biosynthesis in an economically important fruit crop, shed light on the mechanisms leading to iridoid production in specific germplasm, and could aid in developing cultivars with increased iridoid/monotropein content for human health.
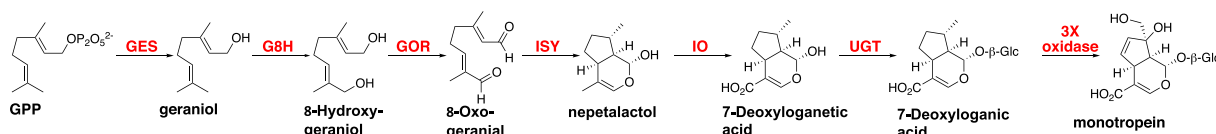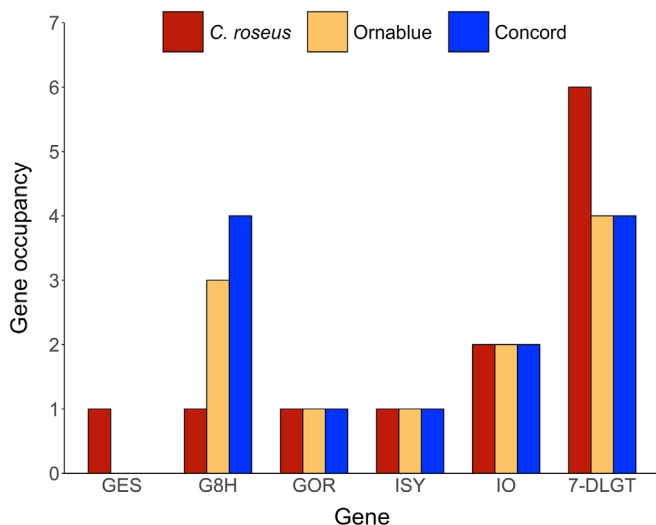

FIGURE 1 Proposed monotropein biosynthetic pathway in *Vaccinium corymbosum*. GPP—geranyl diphosphate; GES—geraniol synthase; G8H—geraniol 8-hydroxylase; GOR—8-hydroxygeraniol oxidoreductase; ISY—iridoid synthase; IO—iridoid oxidase; UGT—UDP-glucuronosyltransferase.

2 | RESULTS AND DISCUSSION

2.1 | De novo transcriptome assembly of differential monotropein cultivars and transcriptome-enabled gene discovery

A diverse set of tissues from two highbush blueberry cultivars with contrasting monotropein content (Table S1) were subjected to RNA-Sequencing (RNA-Seq) and when aligned to the *V. corymbosum* cv. Draper reference genome (Colle et al., 2019), had an overall alignment rate of 89.1%–92.3% (Table S2). As the reference genome (cv. Draper) is M−, independent transcriptome assemblies were completed for both Ornablue (M+) and Concord (M−) (Table S3) using young leaf, mature leaf, floral bud, open flower (Ornablue only), stem, unripe fruit, and ripe fruit tissues (Table S1) to ensure the difference in monotropein biosynthesis was not due to a presence/absence variant. Benchmarking Universal Single-Copy Orthologue (BUSCO) analysis revealed 98.8% (420/425) and 98.4% (418/425) of the Viridiplantae BUSCO genes for Ornablue (M+) and Concord (M−), respectively (Table S3), suggesting of a highly representative transcriptome for gene discovery.

Translated peptide sequences generated from representative transcripts for both Ornablue (M+) and Concord (M−) were used to identify orthologs of *C. roseus* iridoid biosynthetic genes using OrthoFinder (Figure 2). OrthoFinder assigned 158,620 genes (50.5% of total) to 56,571 orthogroups; 50% of all genes were in orthogroups with two or more genes (G50 = 2). There were 6,432 orthogroups with all species present and 2,158 of these consisted

entirely of single-copy genes. Gene occupancy analysis revealed orthologs in both Ornablue (M+) and Concord (M−) for all genes involved in the early portion of the iridoid biosynthetic pathway, except for geraniol synthase (GES) (Table S4). An additional BLAST analysis using *C. roseus* GES as a query against custom blueberry peptide databases resulted in the identification of four potential orthologs of GES in blueberry, all of which were present in Ornablue (M+) (Table S5). Alignment of these potential orthologs to the *V. corymbosum* genome (cv. Draper) identified six possible candidate GES genes (Table S5). The orthogroup containing the *C. roseus ISY* gene contained only a single ortholog of *ISY* for both Ornablue (M+) and Concord (M−), respectively (Figure 2). Orthologs of the late iridoid pathway genes including 7-deoxyloganetic acid glucosyltransferase (7-DLGT), the UDP-glucuronosyltransferase (UGT) in *C. roseus*, were also identified (Figure 2), although the specific UGT for iridoid/monotropein production in blueberry is yet to be determined (Figure 1).

2.2 | Multiple sequence alignment and analysis

To characterize the candidate *ISY* gene in blueberry (*VcISY*), multiple sequence alignment of the amino acid sequences of the Ornablue (M+) and Concord (M−) VcISY was completed and revealed 98% identity to each other and 63% identity to the *C. roseus* ISY (CrISY) (Figure S1). Further comparison of other ISY amino acid sequences demonstrated that VcISY contained the amino acid residues previously identified to be critical for ISY function (Table 1) (Kries et al., 2016, 2017). The VcISY sequence was also compared with other known ISY sequences that are structurally different from CrISY (Table 1). Analysis of critical residues revealed that except for the amino acid at residue 346, VcISY had the same amino acid residue with CrISY and *Olea europaea* ISY (OeISY) compared with the ISY residues for *Antirrhinum majus* (AmISY), which catalyzes formation of an alternative iridoid stereoisomer. This also suggests the specificity of VcISY to 8-oxogeranial as substrate since the Ala343-Ser354 loop has been shown to determine substrate specificity of CrISY, particularly in substrate recognition (Kries et al., 2016).

Phylogenetic analysis of the Ornablue *VcISY* sequence using the maximum likelihood method revealed clustering with functionally validated *ISY* genes from *C. roseus* (Madagascar periwinkle), *O. europaea*

FIGURE 2 Number of genes from *C. roseus* and blueberry (Ornablue [M+] and Concord [M−]) in each orthogroup identified using OrthoFinder. Additional possible orthologs identified by BLAST are presented in Table S5. GES—geraniol synthase; G8H—geraniol 8-hydroxylase; GOR—8-hydroxygeraniol oxidoreductase; ISY—iridoid synthase; IO—iridoid oxidase; 7-DLGT—7-deoxyloganetic acid glucosyltransferase (the likely UDP-glucuronosyltransferase involved in monotropein biosynthesis).

TABLE 1 Comparison of amino acid residues at the binding pocket of characterized iridoid synthase (ISY) in different species. Column headers indicate amino acid position. Cr—*C. roseus*; Oe—*Olea europaea*; Vc—*Vaccinium corymbosum*; Am—*Antirrhinum majus*.

|	149	246	342	345	346	349	352
CrISY	F	A	F	I	A	S	L
OeISY	I	V	F	I	A	L	V
VcISY	F	A	F	I	V	S	L
AmISY	W	W	L	V	V	N	T

(olive), *Digitalis purpurea* (foxglove), *A. majus* (snapdragon), and *Sesamum indicum* (sesame) and not with *ISY* genes associated with the VEP1 clade, other progesterone-5-β-reductase genes (including the CrP5βR6 clade), or Nepetoideae-specific VEP1 clade (Figure 3). Although P5βRs have intrinsic iridoid synthase activities (Munkert et al., 2015), *VcISY* did not cluster with *ISY* genes associated with the VEP1 clade, other P5βR genes, or Nepetoideae-specific VEP1 clade, as these ISYs have unique iridoid synthase structures compared with the majority of characterized iridoid synthases (Lichman et al., 2020). This shows that *VcISY* has a higher sequence similarity with the other functionally validated *ISYs*.

2.3 | Biochemical assay of blueberry ISY

The *VcISY* gene from both Ornablue (M+) and Concord (M−) were expressed in *E. coli* and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (Figure S2) showed that the purified protein corresponds to the expected size of 42 kDa and that the His-tagged protein was successfully isolated from other expressed proteins. The purified protein was used for functional enzyme analysis. End-point enzyme assays and analysis of the end product using gas chromatography–mass spectrometry (GC-MS) revealed that both the Ornablue (M+) and Concord (M−) VcISY catalyzed the reaction converting 8-oxogeranial (8-OG) to *cis,trans*-nepetalactol (Figure 4), with the substrate fully consumed in both reactions. Analysis of the enzyme reaction products also showed the presence of other peaks in the GC-MS chromatograms (Figures 4 and S3), which correspond to the partially cyclized iridodials that are typically found in the ISY reaction (Nguyen & O'Connor, 2020; Sherden et al., 2018).

2.4 | Expression analysis of iridoid biosynthetic genes

Gene expression analysis of orthologs identified using OrthoFinder and BLAST was performed to determine if transcript abundance mirrored metabolite levels across tissues and cultivars (Figure 5; Table S6). To minimize differences in transcript assembly between the two de novo transcriptomes, we aligned our Ornablue (M+) and Concord (M−) transcriptomes to the Draper reference genome (Colle et al., 2019), and annotated genes located in the aligned regions of the Draper genome were then analyzed for their gene expression using the RNA-Seq data. Some orthologs mapped to more than one annotated gene in the Draper genome, while others mapped to regions without any annotated genes (Table S4). The highest transcript expression, represented as fragments per kilobase of exon per million mapped fragments (FPKM), across all samples and genes was in Ornablue floral buds for the majority of the orthologs in the iridoid biosynthetic pathway (Figure 5; Table S6). Transcript expression of *ISY* in Ornablue (M+) was highest in unripe fruits followed by floral buds (Figure 5), which demonstrate a different trend in comparison with metabolite production across tissues analyzed (Figure S4) (Spearman Rank coefficient $R = 0.45$, $p = 0.14$) (Leisner et al., 2017). Transcript levels of *ISY* in Concord (M−) had a different pattern compared with the metabolite data. Notably, the relative expression levels in mature leaves of Concord are comparable with Ornablue floral buds, but metabolite levels are lower in Concord (Figures 5 and S4). Since both Ornablue (M+) and Concord (M−) show expression of the *ISY* ortholog, but Concord does not produce monotropein, it is possible that the monoterpene cyclase in blueberry is structurally different than ISY in *C. roseus*. There is a precedent for this hypothesis, as the
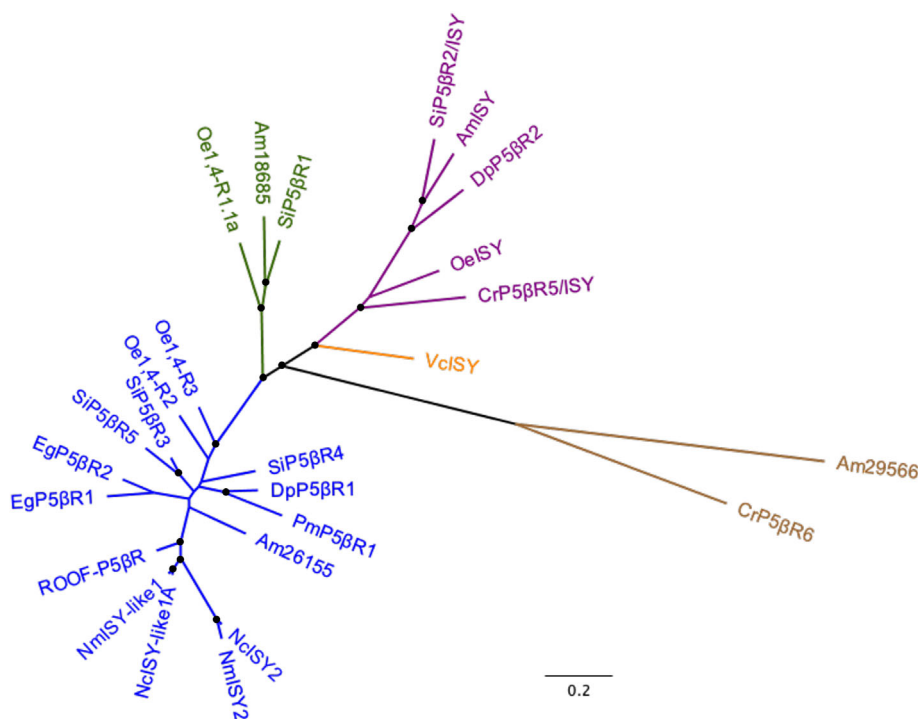
FIGURE 3 Phylogenetic tree of ISY and related P5βR-like enzymes. Relationships are based on coding sequences of characterized and putative ISY and P5βR/ISY-like enzymes (PRISE) analyzed by maximum likelihood method. Clades with bootstrap values ≥90 are labeled with black circles. The VEP1 clade is shown in blue; *Cr*P5βR4 clade in green; ISY clade in purple including blueberry ISY (*VcISY*) in orange; and *Cr*P5βR6 clade in brown. Species acronyms: *Antirrhinum majus* (Am), *Catharanthus roseus* (Cr), *Digitalis purpurea* (Dp), *Erythranthe guttatus* (Eg), *Nepeta cataria* (Nc), *Nepeta mussinii* (Nm), *Olea europaea* (Oe), *Plantago major* (Pm), *Sesamum indicum* (Si), *Vaccinium corymbosum* (Vc). *VcISY* sequence is based on the monotropein-producing cultivar Ornablue.
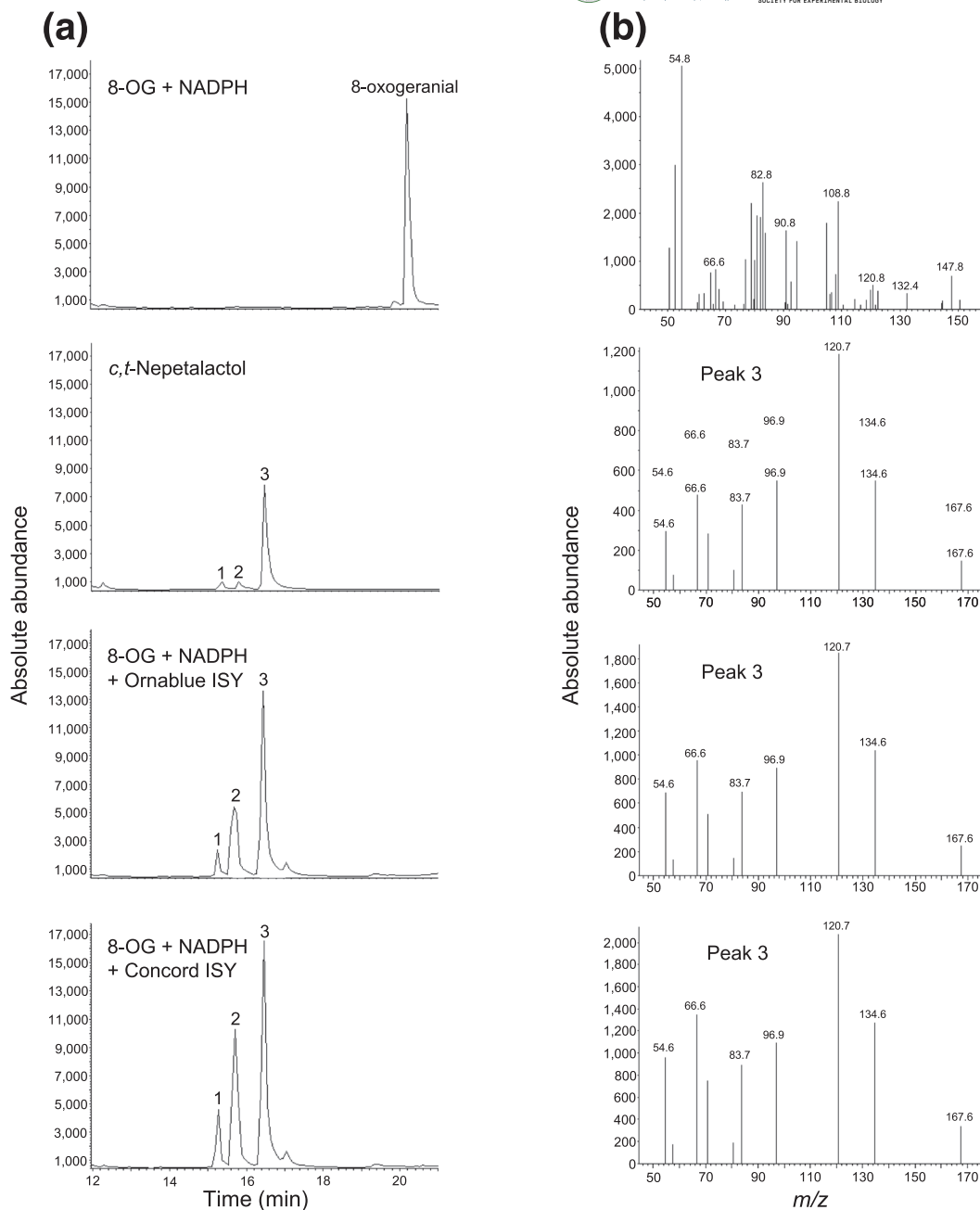
FIGURE 4 Functional characterization of blueberry ISY by gas chromatography–mass spectrometry analysis of in vitro enzyme assay reaction products. (a) Total ion chromatograms of negative control reaction (substrate [8-oxo-geranial, 8-OG] and NADPH only; no enzyme), c,t-nepetalactol (reference standard), and reaction products with ISY from Ornablue (M+) and Concord (M−) in the presence of NADPH are compared. ISY converted 8-OG to c,t-nepetalactol (peak 3) and iridodials (peaks 1 and 2). (b) Mass spectra of c,t-nepetalactol (peak 3) labeled in panel a. Mass spectra of peaks 1 and 2 are shown in Figure S3. Chromatograms and mass spectra were generated using the Agilent Enhanced ChemStation software.

monoterpene cyclase in *C. roseus* uses an alternative biochemical mechanism to create cyclic terpenes (Geu-Flores et al., 2012).

Due to the discordant metabolite and transcript expression levels, we hypothesize that *VcISY* is expressed in both cultivars, yet the expression of genes encoding enzymes catalyzing the other steps of the biosynthetic pathway may affect monotropein levels. For example, in Lamiaceae, GES and ISY were significantly associated with gene expression and iridoid production across multiple Lamiaceae species

(Mint Evolutionary Genomics Consortium, 2018). Low or no detectable expression of *GES* was observed in non-iridoid producing species, while expression of *GES* was highly correlated with iridoid production. Regardless of iridoid production, *ISY* had some expression in all species, although iridoid-producing species had higher gene expression levels compared with non-producing species (Mint Evolutionary Genomics Consortium, 2018). In blueberry, we observed a similar pattern with iridoid (monotropein) production and *ISY*, where it is
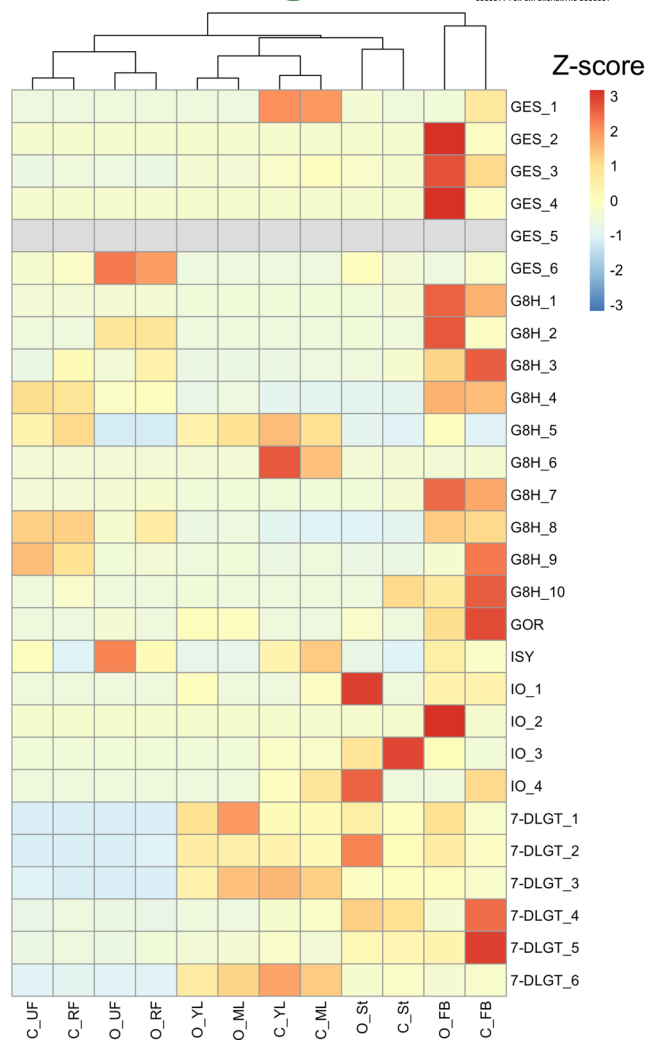
FIGURE 5 Transcript abundance of iridoid biosynthetic genes in Ornablue (M+) and Concord (M−). Transcript data is shown as Z-scores of FPKM values of transcripts across tissue samples in Ornablue (M+) and Concord (M−) with hierarchical clustering across samples. FPKM values were based on alignment to the *V. corymbosum* cv. Draper genome (Colle et al., 2019). Alignment of orthologs to the *V. corymbosum* genome was done using GMAP (v.20130509), and transcripts located in the aligned regions were then analyzed for their gene expression (FPKM). Expression values of orthologs are indicated in Table S6. Cells shown in gray depict identified transcripts that are not expressed (FPKM = 0) in the analyzed samples. Transcript abbreviations: GES—geraniol synthase; G8H—geraniol 8-hydroxylase; GOR—8-hydroxygeraniol oxidoreductase; ISY—iridoid synthase; IO—iridoid oxidase; 7-DLGT—7-deoxyloganetic acid glucosyltransferase. Sample acronyms: O—Ornablue; C—Concord; FB—floral buds; St—stems; UF—unripe fruits; RF—ripe fruits; YL—young leaves; ML—mature leaves.

TABLE 2 Top 10 co-expressed genes with iridoid synthase in blueberry. Co-expressed genes were identified by mutual rank analysis. Functional annotation was based on Wisecaver and Edger (2019). The top 200 co-expressed genes are listed in Table S8.

Gene ID	Annotation
maker-VaccDscaff17-snap-gene-35.46	Adenylate kinase/UMP-CMP kinase
snap_masked-VaccDscaff48-processed-gene-44.18	UDP-glycosyltransferase family
maker-VaccDscaff2-snap-gene-409.60	UDP-glucuronosyl/UDP-glucosyltransferase
maker-VaccDscaff31-snap-gene-15.28	Armadillo-type fold
maker-VaccDscaff34-augustus-gene-236.23	3,4-dihydroxy-2-butanone 4-phosphate synthase, RibB
maker-VaccDscaff28-augustus-gene-338.25	ABC transporter type 1, transmembrane domain
maker-VaccDscaff2-augustus-gene-25.22	LOG family (possible lysine decarboxylase)
augustus_masked-VaccDscaff34-processed-gene-61.1	Stress up-regulated Nod 19
maker-VaccDscaff24-augustus-gene-156.30	Vacuolar protein sorting-associated protein 13, N-terminal domain
maker-VaccDscaff20-snap-gene-300.25	Protein kinase domain

expression in leaves from Concord (Figure 5), although expression levels are generally low (FPKM <5) (Table S6). This follows a similar pattern as that found in Lamiaceae species (Mint Evolutionary Genomics Consortium, 2018), but further analysis and experimentation is needed to determine functionality of these orthologs.

Additionally, post-transcriptional modifications and/or post-translational modifications could lead to the absence of monotropein. This discordant pattern between metabolite and transcript expression levels has been observed in other plant species, including metabolites in the central metabolic pathways in *Brassica napus* (Schwender et al., 2014). It is also possible that formation of other iridoid compounds that have the same stereochemistry as monotropein, such as gardenoside which has been previously identified in neotropical blueberry (Ma et al., 2013), is favored in M− cultivars such as Concord as opposed to production of monotropein. Additionally, low expression of iridoid biosynthetic genes could lead to low levels of monotropein that is below the limit of quantitation (10 ng/mg monotropein), which was the basis of classifying cultivars as M+ or M− (Leisner et al., 2017).

Other steps in the iridoid biosynthetic pathway may also limit monotropein production as genes downstream of *ISY* are required for the complete biosynthesis of monotropein (Figure 1). To identify additional candidate genes in blueberry, mutual rank-based co-expression analysis using expression data across the transcriptome (Table S7) was also performed (Table S8). The 10 best co-expressed transcripts with *VcISY* from the mutual rank analysis are shown in Table 2, which

expressed in both M+ and M− cultivars (Figure 5). In our current dataset, we observed that both Ornablue (M+) and Concord (M−) have detectable levels of potential *GES* orthologs, with high expression (FPKM >50) of two *GES* orthologs (GES_2 and GES_4) and moderate expression of one *GES* ortholog (GES_3) in Ornablue floral buds (Figure 5; Table S6). Another *GES* ortholog (GES_1) had the highest

includes UDP-glucuronosyltransferase (*UGT*). This gene has higher expression in Ornablue (M+) tissues than in Concord (M−), except for leaves (Figure S5). In the proposed monotropein biosynthetic pathway (Figure 1) UGT is downstream of ISY, following enzyme action of iridoid oxidase (IO). Based on analysis in *C. roseus*, the likely UGT is 7-DLGT, which we identified potential candidate genes of from our orthology analysis (Figure 2; Tables S4 and S6). The *UGT* identified from mutual rank analysis, however, is not one of the *7-DLGT* candidate genes identified from orthology analysis (Tables 2 and S6). Analysis of the *7-DLGT* candidate genes demonstrated these orthologs have the lowest expression in fruits across all tissues in both Ornablue (M+) and Concord (M−) (Figure 5; Table S6). Altogether, we have identified candidates for three biosynthetic genes (*ISY*, *GES*, and *UGT/7-DLGT*) in the blueberry monotropein biosynthetic pathway using transcriptomics, orthology, and co-expression analysis that will allow us to investigate the mechanism behind differential monotropein production between M+ and M− blueberry cultivars.

3 | CONCLUSION

Overall, we have shown that multiple "omic" (transcriptomics, co-expression analysis, and comparative genomics) approaches identified candidate genes for the iridoid biosynthetic pathway in blueberry. Among the candidate genes identified, we functionally characterized ISY using in vitro enzyme assays and showed that the putative *VcISY*s from both Ornablue (M+) and Concord (M−) produce functional enzymes that catalyze the formation of the iridoid *c,t*-nepetalactol. With the success of the methodology pipeline, other genes in the pathway will also be validated using the same approach. Among the suite of genes involved in iridoid/monotropein production, *GES* in the early iridoid pathway as well as *UGT* and oxidases in the tailoring steps leading to monotropein synthesis are of special interest for the next analysis based on orthology and co-expression analyses. Findings from this study expand our knowledge on iridoid production in blueberry that could be used in developing new cultivars with increased nutritional quality.

4 | MATERIALS AND METHODS

4.1 | Sample collection

Sampling was performed on blueberry plants grown by the Michigan Blueberry Growers Association in Grand Junction, MI, USA (42°24′09.4″N, 86°04′20.9″W) in 2016. Tissue samples were collected from Ornablue, a monotropein-positive cultivar, and Concord, a monotropein-negative cultivar (Leisner et al., 2017). Unripe and ripe fruits, young and mature leaves (at fruit ripening), floral buds, and stems were harvested from both Ornablue and Concord, with additional open flowers from Ornablue. Tissues were sampled from one plant per cultivar. Samples were placed on ice and transported to East Lansing, MI, where they were subsequently flash frozen with liquid nitrogen and stored at −80°C until use. Monotropein content of the same set of samples (except open flowers) was previously quantified by liquid chromatography-mass spectrometry (Leisner et al., 2017).

4.2 | RNA-sequencing (RNA-Seq) and de novo transcriptome assembly

Total RNA was extracted from each of the tissue samples for both Ornablue and Concord. Frozen tissue was homogenized into a fine powder using a mortar and pestle. RNA from young leaves, mature leaves, floral buds, open flowers (Ornablue only), and stems was extracted using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) according to manufacturer's protocol. RNA from unripe and ripe fruits was extracted using a modified protocol from Vashisth et al. (2011). Genomic DNA contamination was removed from RNA samples using Turbo DNase treatment (Applied Biosystems by Life Technologies, Austin, TX, USA) according to the manufacturer's protocol. RNA quantity was determined with a spectrophotometer (NanoDrop 1000, Thermo Fisher Scientific, Waltham, MA, USA), while quality was assessed by running samples on a 1.5% nondenaturing agarose gel and on a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). All cDNA libraries were constructed by Novogene (Novogene Corporation, Chula Vista, CA, USA). RNA-Seq was also completed by Novogene generating 150-nt paired-end reads for all tissues of both cultivars using an Illumina HiSeq 4000 (Illumina Inc, San Diego, CA, USA). FASTQ files from all sequencing runs have been deposited in the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gove/sra) under BioProject PRJNA477303.

Paired-end reads were assessed for quality using FASTQC (v.0.11.5) and cleaned using Cutadapt (v.2.10; Martin, 2011) with a minimum quality score of 20 (-q 20), minimum length of 36 (-m 36) and custom adapter sequences. Libraries were checked for contamination by comparison to the NCBI non-redundant nucleotide database using BLAST (v.2.9.0+; Camacho et al., 2009). Separate transcriptome assemblies were made for Ornablue and Concord utilizing all tissue libraries for each cultivar using Trinity (v. 2.10.0; Haas et al., 2013) with a minimum contig length of 200 bp (--min_contig_length 200) and --full_cleanup mode with remaining default parameters. Completeness of the assembled transcriptomes based on expected gene content was assessed by BUSCO (v.5.3.2; Manni et al., 2021) using the dataset viridiplantae_odb10 (2022-05-19).

4.3 | Gene expression and co-expression analysis

Gene expression analysis was performed by aligning each RNA-Seq library to the *V. corymbosum* genome (cv. Draper; Colle et al., 2019) using HISAT2 (v.2.2.1; Kim et al., 2019) with default parameters, followed by determination of FPKM values using StringTie (v.2.2.0; Pertea et al., 2016) (Table S7). Transcript abundances of potential orthologs of iridoid biosynthetic genes were converted to Z-scores,

which were plotted on an expression heatmap with hierarchical clustering across samples using the pheatmap package (v.1.0.12; Kolde, 2019) in RStudio (v.4.1.1; RStudio Team, 2021). Spearman correlation test between monotropein levels and transcript expression (FPKM) of *ISY* in all tissues was performed using the cor.test function in RStudio (v.4.1.1; RStudio Team, 2021).

Co-expression analysis by mutual rank was also performed. First, the Pearson correlation coefficient was calculated using the filtered FPKM matrix (FPKM >1). A mutual rank list was then generated following Geu-Flores et al. (2012). The top 200 genes co-expressed with *ISY* were picked (Table S8). Annotation of the top 10 co-expressed genes was determined based on Wisecaver and Edger (2019).

4.4 | Orthology analysis

Orthology analysis using OrthoFinder (v.0.7.1; Emms & Kelly, 2015) was performed with representative peptide sequences from *C. roseus* and Transdecoder-derived peptides (Haas et al., 2013) from the Ornablue and Concord de novo transcriptomes. Potential orthologs were analyzed for location in the genome by aligning transcript sequences to the *V. corymbosum* reference genome (cv. Draper; Colle et al., 2019) using GMAP (v.20130509; Wu & Watanabe, 2005). Annotated genes located in the aligned regions of the Draper genome were then analyzed for their gene expression as stated above. In the case of absence of an ortholog identified using Orthofinder (e.g., GES), BLAST (v.2.2.31+; Camacho et al., 2009) was also used to identify potential orthologs. BLAST was performed using a custom database built using either the Ornablue or Concord peptide sequences derived from Transdecoder. The *C. roseus* peptide sequence for GES was used as the query to search the custom databases. Functional annotation of the top BLAST hits was determined by BLAST analysis using The Arabidopsis Information Resource (TAIR) database (TAIR10; https://www.arabidopsis.org/Blast/). Transcript abundance of the GES orthologs determined by BLAST was also determined by aligning transcript sequences to the *V. corymbosum* reference genome (cv. Draper; Colle et al., 2019) using GMAP (v.20130509) and then annotated genes located in the aligned regions of the Draper genome were analyzed for their gene expression as stated above.

4.5 | Analysis and sequence alignment of ISY

To determine sequence similarity, the putative blueberry ISY (VcISY) was compared with *C. roseus* ISY (CrISY) at the amino acid level using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/). Additionally, the amino acids at the enzyme binding pocket site of VcISY and CrISY were compared based on the structure of CrISY (Kries et al., 2016).

The coding sequences of *VcISY* (Ornablue [M+], DN135366_c0_g1_i1) and *Cr*ISY/P5βR5 (JX974564) from the ISY orthogroup were used for alignment using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). A phylogenetic tree of known ISYs and P5ßR enzymes was constructed using additional sequences from GenBank from *C. roseus*: CrP5βR6 (KJ873887.1); *A. majus* (*Am*): *Am*ISY (MF281392.1), *Am*18685 (MF281393.1), *Am*26155 (MF281394.1), Am29566 (MF281395.1); *O. europaea* (*Oe*): *Oe*ISY (KT954038.1), *Oe*1,4-R1.1a (KT954039.1), *Oe*1,4-R2 (KT954043.1), *Oe*1,4-R3 (KT954044.1); *D. purpurea* (*Dp*): *Dp*P5βR1 (AJ310673.1), *Dp*P5βR2 (GU062787.1); *Rosmarinus officinalis*: ROOF-P5βR (JQ894868.1); *S. indicum* (*Si*): *Si*P5βR1 (XM 011071474.2), *Si*P5βR2/ISY (XM 011071475.2), *Si*P5βR3 (XM 011089869.2), *Si*P5βR4 (XM 011094221.2), *Si*P5βR5 (XM 020695852.1); *Erythranthe guttatus* (*Eg*): *Eg*P5βR1 (XM 012981417.1), *Eg*P5βR2 (XM 012981418.1); *Plantago major* (*Pm*): *Pm*P5βR1 (GU354233.1); *Nepeta cataria* (*Nc*): *Nc*ISYlike1A (KY882233.1), *Nc*ISY2 (KY882234.1); and *Nepeta mussinii* (*Nm*): *Nm*ISYlike1 (KY882235.1), *Nm*ISY2 (38) (KY882236.1). Using the inferred guide tree, sequences were realigned using PRANK (v.170427; Löytynoja & Goldman, 2008). Maximum likelihood (ML) tree inference was performed using IQ-TREE (v.1.6.12; Nguyen et al., 2015) with the best-fit substitution model (TIM3 + F + I + G4 based on Bayesian Information Criterion) determined by ModelFinder (Kalyaanamoorthy et al., 2017). The ML tree was generated with 200 iterations and ultrafast bootstrap approximation was used to obtain branch supports (1000 replicates; Hoang et al., 2018). The tree was visualized using FigTree (v1.4.4; http://tree.bio.ed.ac.uk/software/figtree/).

4.6 | Iridoid synthase protein expression and purification

Both the full-length Ornablue and Concord *VcISY* could not be amplified; hence, genes were synthesized (Twist Bioscience, San Francisco, CA, USA) and cloned in the pET-28a(+) expression vector with N-terminal 6× His-tag and codon-optimized for *E. coli* (Table S9). Enzyme expression and purification was done following Lichman et al. (2020) with modifications. The plasmids were subcloned into *E. coli* (BL21 DE3) which were grown overnight in LB medium with kanamycin at 37°C with shaking at 200 rpm. The overnight starter culture was used to inoculate 1 L of LB medium with kanamycin and grown at 37°C at 200 rpm until $OD_{600} = 0.7$. The cultures were then cooled to 18°C with shaking before adding 0.25 mM IPTG to induce protein expression. The bacterial cultures were incubated at 18°C for 17 h. Cells were harvested by centrifugation (4000*g*, 15 min, 4°C), resuspended in lysis buffer, and incubated on ice for 15 min. Cells were then lysed using a sonicator (Qsonica, Newtown, CT, USA; amplitude, 50% pulses) for two 5-min cycles of 20 s on, 40 s off for a total of 10 min. The lysate was centrifuged (35,000*g*, 30 min, 4°C) and the resulting supernatant was added with 500 μL Ni-nitrilotriacetic acid (Ni-NTA) agarose (QIAGEN, Hilden, Germany) washed in binding buffer. The mixture was incubated in a rocking incubator at 4°C for 1 h (30 rpm) and then centrifuged (2000*g*, 3 min, 4°C) to collect the Ni-NTA pellet. The pellet was washed three times with 10 mL binding buffer, then eluted with 600 μL elution buffer. Following

centrifugation (2000g, 5 min, 4°C), the supernatant was collected and filtered using a centrifugal filter (0.1 μm, Ultrafree®-MC, Merck Millipore, Tullagreen Carrigtwohill Co. Cork, Ireland). The elution buffer was exchanged with sample buffer using a Ultracel-10 regenerated cellulose membrane (Amicon Ultra-0.5, Merck Millipore, Tullagreen Carrigtwohill Co. Cork, Ireland). Protein concentration was measured using NanoDrop. The purified proteins were divided into 20 μL aliquots and flash-frozen in liquid nitrogen before storage at −80°C. The purity of the samples was checked by running samples (≤5 μg) on an SDS-PAGE gel and stained with Coomassie blue.

4.7 | End-point enzyme assay and GC-MS

Purified proteins were used in an end-point enzyme assay with 8-oxogeranial (8-OG) as substrate. The reactions were based on Lichman et al. (2020). Briefly, buffer (HEPES [50 mM, pH 8.0]), NADPH (1 mM), ISY (0.5 mM), and 8-OG (0.5 mM) were mixed and incubated at 30°C with shaking at 200 rpm for 3 h. The reactions were quenched with 100 μL ethyl acetate mixed with D-camphor (1 mM, 10 μL) as internal standard, and the organic phase was analyzed by GC-MS. An Agilent 6890N GC and 5975 MS with a Restek Rxi-5Sil MS column with Integra-Guard column (15 m, 0.25 mm ID, 0.25 μm df) was used to analyze the samples. The temperature of the injection port was 220°C and 1 μL of sample was injected in split mode with a ratio of 5:1. The pressure was constant at 2.2 psi and flow was set to 1.2 mL/min with Helium gas. The GC oven temperature was programmed with an initial temperature of 40°C for 3 min, ramp to 110°C at 5°C/min, and held for 2 min and then ramp to 280°C at 120°C/min with a 4 min hold, thus requiring a total run time of 24.4 min. The MS transfer line was set at 250°C. The MS source was set to 230°C and the EI was set to 70 eV emission with a scan range of 45 to 300 mz. Selected ion monitoring was used to monitor D-camphor at 80.8 and 108.8 mz and c,t-nepetalactol as reference standard at 54.6 and 120.7 mz beginning at 14 min. The raw data were processed and analyzed using Agilent Enhanced ChemStation software.

ORCID
Lovely Mae F. Lawas https://orcid.org/0000-0001-9141-1304
C. Robin Buell https://orcid.org/0000-0002-6727-4677
Courtney P. Leisner https://orcid.org/0000-0002-8285-686X

REFERENCES
Alagna, F., Geu-Flores, F., Kries, H., Panara, F., Baldoni, L., O'Connor, S. E., & Osbourn, A. (2016). Identification and characterization of the iridoid synthase involved in oleuropein biosynthesis in olive (Olea europaea) fruits. Journal of Biological Chemistry, 291, 5542–5554. https://doi.org/10.1074/jbc.M115.701276

Bekaert, M., Edger, P. P., Hudson, C. M., Pires, J. C., & Conant, G. C. (2012). Metabolic and evolutionary costs of herbivory defense: Systems biology of glucosinolate synthesis. New Phytologist, 196, 596–605. https://doi.org/10.1111/j.1469-8137.2012.04302.x

Borghi, M., & Xie, D. Y. (2018). Cloning and characterization of a monoterpene synthase gene from flowers of Camelina sativa. Planta, 247, 443–457. https://doi.org/10.1007/s00425-017-2801-x

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10, 421. https://doi.org/10.1186/1471-2105-10-421

Chen, Y., Lu, Y., Pei, C., Liang, J., Ding, P., Chen, S., & Hou, S. Z. (2020). Monotropein alleviates secondary liver injury in chronic colitis by regulating TLR4/NF-κB signaling and NLRP3 inflammasome. European Journal of Pharmacology, 883, 173358. https://doi.org/10.1016/j.ejphar.2020.173358

Choi, J., Lee, K. T., Choi, M. Y., Nam, J. H., Jung, H. J., Park, S. K., & Park, H. J. (2005). Antinociceptive anti-inflammatory effect of monotropein isolated from the root of Morinda officinalis. Biological & Pharmaceutical Bulletin, 28, 1915–1918. https://doi.org/10.1248/bpb.28.1915

Colle, M., Leisner, C. P., Wai, C. M., Ou, S., Bird, K. A., Wang, J., Wisecaver, J. H., Yocca, A. E., Alger, E. I., Tang, H., Xiong, Z., Callow, P., Ben-Zvi, G., Brodt, A., Baruch, K., Swale, T., Shiue, L., Song, G. Q., Childs, K. L., ... Edger, P. P. (2019). Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. Gigascience, 8, 1–15. https://doi.org/10.1093/gigascience/giz012

Collu, G., Peltenburg-Looman, A. M. G., 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. https://doi.org/10.1016/S0014-5793(01)03045-9

DellaPena, D., & O'Connor, S. E. (2012). Plant gene clusters and opiates. Science, 336, 1648–1649. https://doi.org/10.1126/science.1225473

Dinda, B., Debnath, S., & Banik, R. (2011). Naturally occuring iridoids and secoiridoids. An updated review, part 4. Chemical & Pharmaceutical Bulletin, 59, 803–833. https://doi.org/10.1248/cpb.59.803

Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*, 157. https://doi.org/10.1186/s13059-015-0721-2

FAO. (2022) FAOSTAT, Rome.

Geu-Flores, F., Sherden, N. H., Glenn, W. S., O'connor, S. E., Courdavault, V., Burlat, V., Nims, E., Wu, C., & Cui, Y. (2012). An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, *492*, 138–142. https://doi.org/10.1038/nature11692

Grace, M. H., Ribnicky, D. M., Kuhn, P., Poulev, A., Logendra, S., Yousef, G. G., Raskin, I., & Lila, M. A. (2009). Hypoglycemic activity of a novel anthocyanin-rich formulation from lowbush blueberry, *Vaccinium angustifolium* Aiton. *Phytomedicine*, *16*, 406–415. https://doi.org/10.1016/j.phymed.2009.02.018

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Heffels, P., Müller, L., Schieber, A., & Weber, F. (2017). Profiling of iridoid glycosides in *Vaccinium* species by UHPLC-MS. *Food Research International*, *100*, 462–468. https://doi.org/10.1016/j.foodres.2016.11.018

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*, 518–522. https://doi.org/10.1093/molbev/msx281

Irmler, S., Schröder, G., St-Pierre, B., Crouch, N. P., Hotze, M., Schmidt, J., Strack, D., Matern, U., & Schröder, J. (2000). Indole alkaloid biosynthesis in *Catharanthus roseus*: New enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *The Plant Journal*, *24*, 797–804. https://doi.org/10.1046/j.1365-313x.2000.00922.x

Jensen, H. D., Krogfelt, K. A., Cornett, C., Hansen, S. H., & Christensen, S. B. (2002). Hydrophilic carboxylic acids and iridoid glycosides in the juice of American and European cranberries (*Vaccinium macrocarpon* and *V. oxycoccos*), lingonberries (*V. vitis-idaea*), and blueberries (*V. myrtillus*). *Journal of Agricultural and Food Chemistry*, *50*, 6871–6874. https://doi.org/10.1021/jf0205110

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589. https://doi.org/10.1038/nmeth.4285

Karna, K. K., Choi, B. R., You, J. H., Shin, Y. S., Cui, W. S., Lee, S. W., Kim, J. H., Kim, C. Y., Kim, H. K., & Park, J. K. (2019). The ameliorative effect of monotropein, astragalin, and spiraeoside on oxidative stress, endoplasmic reticulum stress, and mitochondrial signaling pathway in varicocelized rats. *BMC Complementary and Alternative Medicine*, *19*, 333. https://doi.org/10.1186/s12906-019-2736-9

Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., McLay, K., Buell, C. R., & O'Connor, S. E. (2015). Genome-guided investigation of plant natural product biosynthesis. *The Plant Journal*, *82*, 680–692. https://doi.org/10.1111/tpj.12827

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kolde, R. (2019) pheatmap: Pretty heatmaps. Available at: https://CRAN.R-project.org/package=pheatmap

Kries, H., Caputi, L., Stevenson, C. E. M., Kamileen, M. O., Sherden, N. H., Geu-Flores, F., Lawson, D. M., & O'Connor, S. E. (2016). Structural determinants of reductive terpene cyclization in iridoid biosynthesis. *Nature Chemical Biology*, *12*, 6–8. https://doi.org/10.1038/nchembio.1955

Kries, H., Kellner, F., Kamileen, M. O., & O'Connor, S. E. (2017). Inverted stereocontrol of iridoid synthase in snapdragon. *Journal of Biological Chemistry*, *292*, 14659–14667. https://doi.org/10.1074/jbc.M117.800979

Leisner, C. P., Kamileen, M. O., Conway, M. E., O'Connor, S. E., & Buell, C. R. (2017). Differential iridoid production as revealed by a diversity panel of 84 cultivated and wild blueberry species. *PLoS ONE*, *12*, e0179417. https://doi.org/10.1371/journal.pone.0179417

Lichman, B. R., Godden, G. T., Hamilton, J. P., Palmer, L., Kamileen, M. O., Zhao, D., Vaillancourt, B., Wood, J. C., Sun, M., Kinser, T. J., Henry, L. K., Rodriguez-Lopez, C., Dudareva, N., Soltis, D. E., Soltis, P. S., Buell, C. R., & O'Connor, S. E. (2020). The evolutionary origins of the cat attractant nepetalactone in catnip. *Science Advances*, *6*, eaba0721. https://doi.org/10.1126/sciadv.aba0721

Löytynoja, A., & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, *320*, 1632–1635. https://doi.org/10.1126/science.1158395

Ma, C., Dastmalchi, K., Flores, G., Wu, S. B., Pedraza-Peñalosa, P., Long, C., & Kennelly, E. J. (2013). Antioxidant and metabolite profiling of north american and neotropical blueberries using LC-TOF-MS and multivariate analyses. *Journal of Agricultural and Food Chemistry*, *61*, 3548–3559. https://doi.org/10.1021/jf400515g

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Current Protocols*, *1*, e323. https://doi.org/10.1002/cpz1.323

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Jounal*, *17*, 10–12. Available at: http://www-huber.embl.de/users/an-. https://doi.org/10.14806/ej.17.1.200

Matsuba, Y., Nguyen, T. T. H., Wiegert, K., Falara, V., Gonzales-Vigil, E., Leong, B., Schäfer, P., Kudrna, D., Wing, R. A., Bolger, A. M., Usadel, B., Tissier, A., Fernie, A. R., Barry, C. S., & Pichersky, E. (2013). Evolution of a complex locus for terpene biosynthesis in *Solanum*. *Plant Cell*, *25*, 2022–2036. https://doi.org/10.1105/tpc.113.111013

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., Verpoorte, R., Oksman-Caldentey, K. M., Martinoia, E., Bouwmeester, H., Goossens, A., Memelink, J., & Werck-Reichhart, D. (2014). The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications*, *5*, 3606. https://doi.org/10.1038/ncomms4606

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. https://doi.org/10.1016/j.molp.2018.06.002

Moyer, R. A., Hummer, K. E., Finn, C. E., Frei, B., & Wrolstad, R. E. (2002). Anthocyanins, phenolics, and antioxidant capacity in diverse small fruits: *Vaccinium*, *Rubus*, and *Ribes*. *Journal of Agricultural and Food Chemistry*, *50*, 519–525. https://doi.org/10.1021/jf011062r

Munkert, J., Pollier, J., Miettinen, K., van Moerkercke, A., Payne, R., Müller-Uri, F., Burlat, V., O'Connor, S. E., Memelink, J., Kreis, W., & Goossens, A. (2015). Iridoid synthase activity is common among the plant progesterone 5β-reductase family. *Molecular Plant*, *8*, 136–152. https://doi.org/10.1016/j.molp.2014.11.005

Murata, J., Roepke, J., Gordon, H., & de Luca, V. (2008). The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell*, *20*, 524–542. https://doi.org/10.1105/tpc.107.056630

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. https://doi.org/10.1021/acschembio.0c00220

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274. https://doi.org/10.1093/molbev/msu300

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*, 1650–1667. https://doi.org/10.1038/nprot.2016.095

Petersen, J., Lanig, H., Munkert, J., Bauer, P., Müller-Uri, F., & Kreis, W. (2016). Progesterone 5β-reductases/iridoid synthases (PRISE): Gatekeeper role of highly conserved phenylalanines in substrate preference and trapping is supported by molecular dynamics simulations. *Journal of Biomolecular Structure & Dynamics*, *34*, 1667–1680. https://doi.org/10.1080/07391102.2015.1088797

Rossi, G., Woods, F. M., & Leisner, C. P. (2022). Quantification of total phenolic, anthocyanin, and flavonoid content in a diverse panel of blueberry cultivars and ecotypes. *HortScience*, *57*, 901–909. https://doi.org/10.21273/HORTSCI16647-22

RStudio Team. (2021) RStudio: Integrated development environment for R.

Schwender, J., König, C., Klapperstück, M., Heinzel, N., Munz, E., Hebbelmann, I., Hay, J. O., Denolf, P., De Bodt, S., Redestig, H., & Caestecker, E., (2014). Transcript abundance on its own cannot be used to infer fluxes in central metabolism. *Frontiers in Plant Science*, *5*, 668.

Seeram, N. P., Adams, L. S., Zhang, Y., Lee, R., Sand, D., Scheuller, H. S., & Heber, D. (2006). Blackberry, black raspberry, blueberry, cranberry, red raspberry, and strawberry extracts inhibit growth and stimulate apoptosis of human cancer cells in vitro. *Journal of Agricultural and Food Chemistry*, *54*, 9329–9339. https://doi.org/10.1021/jf061750g

Shen, Y., Zhang, Q., Wu, Y., He, Y.-q., Han, T., Zhang, J.-h., Zhao, L., Hsu, H.-y., Song, H.-t., Lin, B., Xin, H.-l., Qi, Y.-p., & Zhang, Q.-y. (2018). Pharmacokinetics and tissue distribution of monotropein and deacetyl asperulosidic acid after oral administration of extracts from *Morinda officinalis* root in rats. *BMC Complementary and Alternative Medicine*, *18*, 288. https://doi.org/10.1186/s12906-018-2351-1

Sherden, N. H., Lichman, B., Caputi, L., Zhao, D., Kamileen, M. O., Buell, C. R., & O'Connor, S. E. (2018). Identification of iridoid synthases from Nepeta species: Iridoid cyclization does not determine nepetalactone stereochemistry. *Phytochemistry*, *145*, 48–56. https://doi.org/10.1016/j.phytochem.2017.10.004

Simkin, A. J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., St-Pierre, B., Giglioli-Guivarc'h, N., Fischer, M. J. C., Memelink, J., & Clastre, M. (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. https://doi.org/10.1016/j.phytochem.2012.09.014

Skrovankova, S., Sumczynski, D., Mlcek, J., Jurikova, T., & Sochor, J. (2015). Bioactive compounds and antioxidant activity in different types of berries. *International Journal of Molecular Sciences*, *16*, 24673–24706. https://doi.org/10.3390/ijms161024673

Swiatek, L., & Komorowski, T. (1972). The occurance of monotropein and of asperulosdie in some species of the families: Ericaceae, Empetraceae and Rubiaceae. *Herba Polonica*, *2*, 168–173.

Tundis, R., Loizzo, M., Menichini, F., Statti, G., & Menichini, F. (2008). Biological and pharmacological activities of iridoids: Recent developments. *Mini-Reviews in Medicinal Chemistry*, *8*, 399–420. https://doi.org/10.2174/138955708783955926

van Moerkercke, A., Fabris, M., Pollier, J., Baart, G. J. E., Rombauts, S., Hasnain, G., Rischer, H., Memelink, J., Oksman-Caldentey, K.-M., & Goossens, A. (2013). CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-seq data. *Plant & Cell Physiology*, *54*, 673–685. https://doi.org/10.1093/pcp/pct039

Vashisth, T., Johnson, L. K., & Malladi, A. (2011). An efficient RNA isolation procedure and identification of reference genes for normalization of gene expression in blueberry. *Plant Cell Reports*, *30*, 2167–2176. https://doi.org/10.1007/s00299-011-1121-z

Viljoen, A., Mncwangi, N., & Vermaak, I. (2012). Anti-inflammatory iridoids of botanical origin. *Current Medicinal Chemistry*, *19*, 2104–2127. https://doi.org/10.2174/092986712800229005

Wang, C., Gong, X., Bo, A., Zhang, L., Zhang, M., Zang, E., Zhang, C., & Li, M. (2020). Iridoids: Research advances in their phytochemistry, biological activities, and pharmacokinetics. *Molecules*, *25*, 287. https://doi.org/10.3390/molecules25020287

Wang, C., Mao, C., Lou, Y., Xu, J., Wang, Q., Zhang, Z., Tang, Q., Zhang, X., Xu, H., & Feng, Y. (2018). Monotropein promotes angiogenesis and inhibits oxidative stress-induced autophagy in endothelial progenitor cells to accelerate wound healing. *Journal of Cellular and Molecular Medicine*, *22*, 1583–1600. https://doi.org/10.1111/jcmm.13434

Wisecaver, J.H. and Edger, P. (2019). Global coexpression analysis of genes in haplotype-phased genome of tetraploid blueberry. Purdue University Research Repository. https://doi.org/10.4231/0FAA-GK65

Wu, T. D., & Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*, 1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Zafra-Stone, S., Yasmin, T., Bagchi, M., Chatterjee, A., Vinson, J. A., & Bagchi, D. (2007). Berry anthocyanins as novel antioxidants in human health and disease prevention. *Molecular Nutrition & Food Research*, *51*, 675–683. https://doi.org/10.1002/mnfr.200700002

Zhang, Z., Zhang, Q., Yang, H., Liu, W., Zhang, N., Qin, L., & Xin, H. (2016). Monotropein isolated from the roots of *Morinda officinalis* increases osteoblastic bone formation and prevents bone loss in ovariectomized mice. *Fitoterapia*, *110*, 166–172. https://doi.org/10.1016/j.fitote.2016.03.013

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.