White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/195102/

Version: Published Version

Article:
Liu, Chanchan, Smit, Samuel, Dang, Jingjie et al. (2023) A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in Japanese catnip. Molecular plant. ISSN: 1674-2052

https://doi.org/10.1016/j.molp.2023.01.004

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

Molecular Plant
Research Article

CellPress
Partner Journal

A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*

Chanchan Liu[1,2,6], Samuel J. Smit[3,6], Jingjie Dang[1,2,6], Peina Zhou[1,2,6], Grant T. Godden[4], Zheng Jiang[1,2], Wukun Liu[1], Licheng Liu[1,2], Wei Lin[2,5], Jinao Duan[1,2], Qinan Wu[1,2,*] and Benjamin R. Lichman[3,*]

[1]School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, China

[2]Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Nanjing University of Chinese Medicine, Nanjing, China

[3]Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK

[4]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

[5]Department of Pathogen Biology, School of Medicine and Holistic Integrative Medicine, Nanjing University of Chinese Medicine, Nanjing, China

[6]These authors contributed equally to this article.

***Correspondence: Qinan Wu (wuqn@njucm.edu.cn), Benjamin R. Lichman (benjamin.lichman@york.ac.uk)**

https://doi.org/10.1016/j.molp.2023.01.004

ABSTRACT

Biosynthetic gene clusters (BGCs) are regions of a genome where genes involved in a biosynthetic pathway are in proximity. The origin and evolution of plant BGCs as well as their role in specialized metabolism remain largely unclear. In this study, we have assembled a chromosome-scale genome of Japanese catnip (*Schizonepeta tenuifolia*) and discovered a BGC that contains multiple copies of genes involved in four adjacent steps in the biosynthesis of *p*-menthane monoterpenoids. This BGC has an unprecedented bipartite structure, with mirrored biosynthetic regions separated by 260 kilobases. This bipartite BGC includes identical copies of a gene encoding an old yellow enzyme, a type of flavin-dependent reductase. *In vitro* assays and virus-induced gene silencing revealed that this gene encodes the missing isopiperitenone reductase. This enzyme evolved from a completely different enzyme family to isopiperitenone reductase from closely related *Mentha* spp., indicating convergent evolution of this pathway step. Phylogenomic analysis revealed that this bipartite BGC has emerged uniquely in the *S. tenuifolia* lineage and through insertion of pathway genes into a region rich in monoterpene synthases. The cluster gained its bipartite structure via an inverted duplication. The discovered bipartite BGC for *p*-menthane biosynthesis in *S. tenuifolia* has similarities to the recently described duplicated *p*-menthane biosynthesis gene pairs in the *Mentha longifolia* genome, providing an example of the convergent evolution of gene order. This work expands our understanding of plant BGCs with respect to both form and evolution, and highlights the power of BGCs for gene discovery in plant biosynthetic pathways.

Key words: specialized metabolism, monoterpenoids, biosynthetic gene cluster, Lamiaceae, biosynthesis, chromosomal inversion

Liu C., Smit S.J., Dang J., Zhou P., Godden G.T., Jiang Z., Liu W., Liu L., Lin W., Duan J., Wu Q., and Lichman B.R. (2023). A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*. Mol. Plant. **16**, 1–16.

INTRODUCTION

Plants interact with their environment through the production, sequestration, and emission of organic molecules. Specialized metabolites are taxon restricted and can vary both between and within species, with potential roles in ecological adaptation (Moore et al., 2014). For a subset of metabolite pathways, genes encoding biosynthetic enzymes can be found in

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

biosynthetic gene clusters (BGCs), genomic regions where multiple non-homologous pathway genes are adjacent (Nützmann et al., 2016; Smit and Lichman, 2022). BGCs may persist due to enhanced co-regulation of clustered pathways, or due to the suppression of recombination that could unlink a polygenic trait, or some combination of the two (Smit and Lichman, 2022). The explanation as to why certain pathways are in BGCs, whereas some are not, remains unclear, as does their specific contribution to the remarkable variation of plant chemistry. Practically, BGCs can aid in the discovery of biosynthetic enzymes that can have use for synthetic biology and biocatalysis (Lichman et al., 2020a; Polturak et al., 2022).

Highly aromatic plants gain their specific odor through the accumulation and emission of volatile terpenoids, often 10-carbon monoterpenoids. These compounds have ecological roles in repelling herbivores and attracting pollinators, and are major components of commercially available fragrances and flavors (Pichersky and Gershenzon, 2002; Unsicker et al., 2009). Many monoterpenoids are bioactive, which is reflected in the traditional medicinal use of aromatic plants and in the use of purified monoterpenes in pharmaceutical formulations (Zielińska-Błajet and Feder-Kubis, 2020). Elucidation of monoterpenoid pathways and identification of their constituent enzymes can provide access to valuable compounds and feedstocks via plant-based (Vining et al., 2020), biocatalytic (Leferink et al., 2019), and synthetic biology approaches (Zebec et al., 2016; Gao et al., 2020).

Members of the mint family (Lamiaceae), especially those from the large subclade (same to traditional subfamily) Nepetoideae, are rich in monoterpenes, and include many culinary and medicinal herbs, and plants harvested for their essential oils (Liu et al., 2018; Mint Evolutionary Genomics Consortium, 2018; Lichman et al., 2020b; Dang et al., 2022). These include the medicinal herb *Schizonepeta tenuifolia* (Benth.) Briq. (Japanese catnip), which is native to China, Korea, and Japan, and is used in traditional medicine for its proposed immunomodulatory, anti-inflammatory, and anti-viral properties (Shan et al., 2016, 2021). The major component in its essential oil is pulegone, a monoterpene with a *p*-menthane structure (Figure 1A).

Plants from the closely related genus *Mentha* also produce *p*-menthane compounds, though in the opposite enantiomeric series (Figure 1A). This stereochemical divergence is caused by the selectivity of the initial limonene synthase (LS) catalyzed step: LSs from *Mentha* spp. produce (−)-limonene, whereas LS from *S. tenuifolia* produces (+)-limonene (Alonso et al., 1992; Rajaonarivony et al., 1992; Maruyama et al., 2001; Srividya et al., 2020). Recently, a transcriptome for *S. tenuifolia* has enabled identification and characterization of downstream enzymes active in *p*-menthane biosynthesis through comparison with well characterized *Mentha* enzymes (Liu et al., 2021a, 2021b; Srividya et al., 2022). For known steps in the *Mentha* pathway up to menthone, a closely related homolog was identified in *S. tenuifolia* and shown to be active. The exception to this is isopulegone isomerase, which is unknown in both pathways (Croteau et al., 2005; Currin et al., 2018), and isopiperitenone reductase (IPR), which is described in *Mentha* but no active homolog could be identified in *S. tenuifolia* transcriptomes (Srividya et al., 2022).

In this work, we set out to elucidate the genetic, genomic, and enzymatic basis for *p*-menthane biosynthesis in *S. tenuifolia*. We assembled a chromosome-scale genome that revealed a monoterpenoid BGC with a unique bipartite structure. In addition to known pathway genes, this BGC contains genes encoding the missing IPR. This old yellow enzyme (OYE) evolved convergently to the short chain reductase that catalyzes the equivalent step in *Mentha* spp. Comparison with related species' genomes showed that the BGC was assembled uniquely in the *S. tenuifolia* lineage through recruitment of genes into a region enriched in monoterpene synthases, and later gained its unique bipartite structure via a duplication-inversion event. This work provides an example of gene discovery through BGCs and highlights how phylogenomic analysis can reveal the origins of BGCs in dynamic genomic neighborhoods.

RESULTS

Schizonepeta tenuifolia genome, phylogenetic placement, and ancient whole-genome duplication

To obtain an *S. tenuifolia* genome assembly, we combined PacBio (80.68 Gb), DNBSEQ (67.57 Gb), and Hi-C sequencing reads (102.96 Gb) into a 762 Mb (contig N50 length = 1.37 Mb, scaffold N50 length = 101 Mb) assembly, of which 696 Mb were mapped onto six pseudomolecules (Supplemental Figure 1 and Supplemental Tables 1, 2, and 3). A total of 30 283 protein-coding genes supported by the Illumina RNA-seq reads was annotated, and 91% of these were functionally annotated (Supplemental Table 4). The BUSCO score, which assesses genome completeness, was 94.5% (Supplemental Table 5).

S. tenuifolia is considered synonymous to *Nepeta tenuifolia* Benth. by some sources (Chang et al., 2014). As an understanding of the phylogenetic placement of species is important for effective interpretation of comparative genomic analysis, we set out to confirm the phylogenetic placement of *S. tenuifolia* within Lamiaceae, especially its relationship to species of *Nepeta* L. To do this, we inferred a molecular phylogeny using a set of species that had high-quality, well-annotated genomes (Dong et al., 2018; Lichman et al., 2020a; Bornowski et al., 2020; Hamilton et al., 2020; Vining et al., 2022). We only included species from Nepetoideae, except *Callicarpa americana* L., an early diverging lineage within Lamiaceae, which acted as an outgroup (Mint Evolutionary Genomics Consortium, 2018). *Schizonepeta tenuifolia* was recovered as a sister to *Hyssopus officinalis* L., which together were sister to two *Nepeta* spp. (Figure 1B). This topology supports previous morphology- and DNA-based phylogenetic results that placed *Schizonepeta* within Nepetinae but not within *Nepeta* (Moon et al., 2010; Serpooshan et al., 2018). Thus, until contrary phylogenetic evidence based on broader sampling within *Nepetinae* becomes available, we recognize *S. tenuifolia* as a distinct lineage and, for convenience, use the name proposed by Briquet (1896) to distinguish this taxon from other species recognized within *Nepeta* (Engler and Prantl, 1896).

Given the widespread but asymmetrical levels of gene duplication and ancient whole-genome duplication (WGD) (polyploidy) reported in Lamiaceae (Godden et al., 2019) and their

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

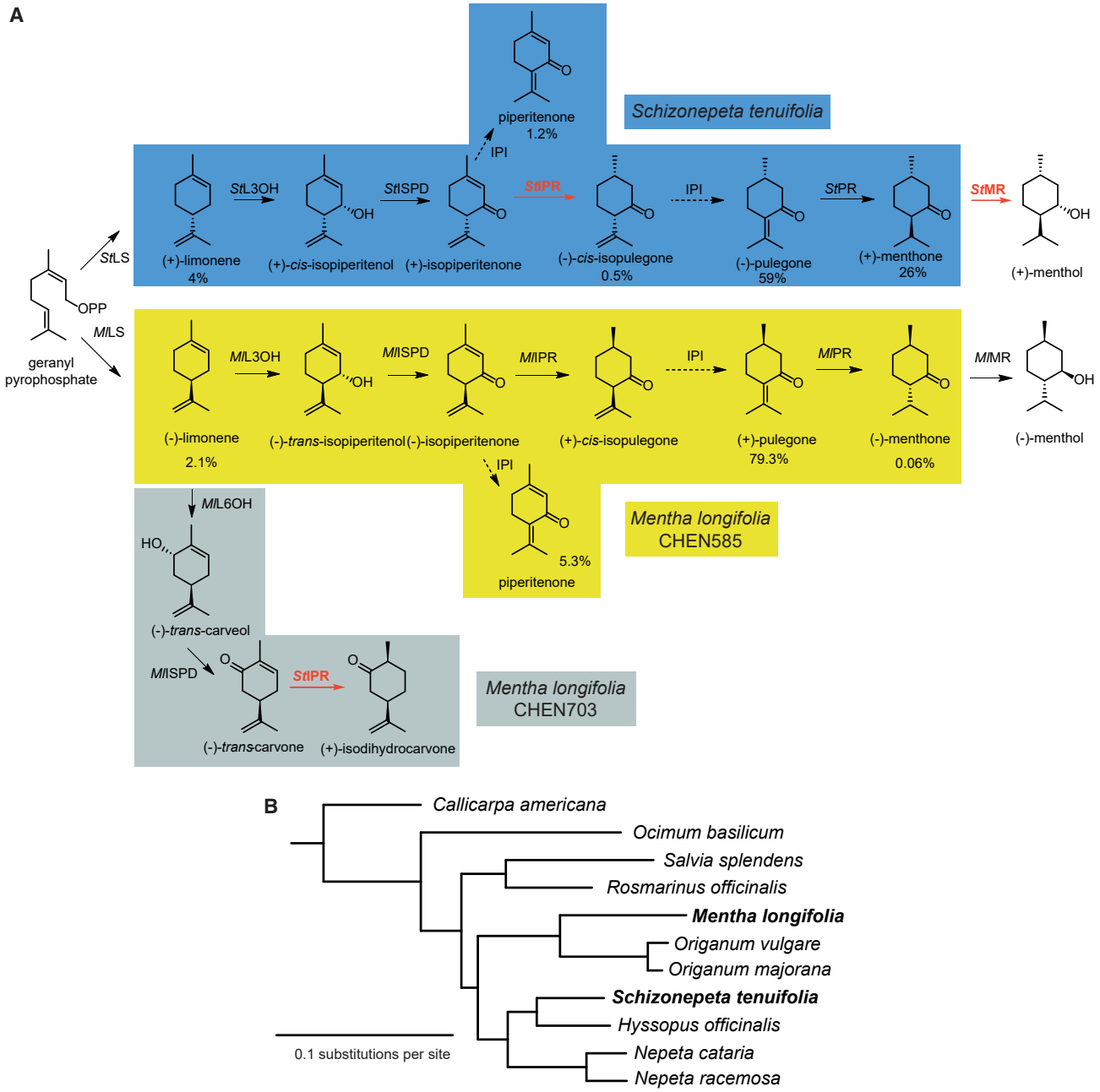

Figure 1. Monoterpene biosynthesis in *Schizonepeta tenuifolia*.

(A) Overview of *p*-menthane monoterpene structures and their biosynthesis in *S. tenuifolia* and *Mentha longifolia*. The boxes indicate *p*-menthane biosynthesis in *S. tenuifolia* and *M. longifolia* CHEN585 and CHEN703 varietals. Percentages under *S. tenuifolia* compounds represent relative abundance in essential oil (Liu et al., 2018), those under *M. longifolia* compounds represent relative abundance of monoterpenoid content (Vining et al., 2022). Dotted arrow indicates the reaction is unknown. Red text/arrows are reactions described for the first time in this manuscript. LS, limonene synthase; L3OH, limonene 3-hydroxylase; L6OH, limonene-6-hydroxylase; ISPD, isopiperitenone dehydrogenase; IPR, isopiperitenone reductase; PR, pulegone reductase; MR, menthone reductase.

(B) Maximum likelihood species tree of selected mint family species in the Nepetoideae, with *Callicarpa americana* used as a non-Nepetoideae outgroup. Tree inference was based on a supermatrix alignment of single-copy genes and was inferred using iQ-Tree. All branches show 100% support as judged by 1000X SH-aLRT and UltraFast Boostrapping replicates.

potential impacts on chemical evolution, we investigated ancient WGDs in Nepetinae using the newly available genomes for *S. tenuifolia* and three additional species (i.e., *H. officinalis, Nepeta cataria* L., *N. racemosa* Lam. subsp. *racemosa*. [syn. *N. mussinii* Spreng.]) (Lichman et al., 2020a). Distributions of estimates of synonymous substitutions per synonymous site (K_S) among pairs of paralogous sequences present in *S. tenuifolia* revealed no unique events. The four genomes showed similar patterns, each with two punctuated episodes of large-scale gene duplication

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

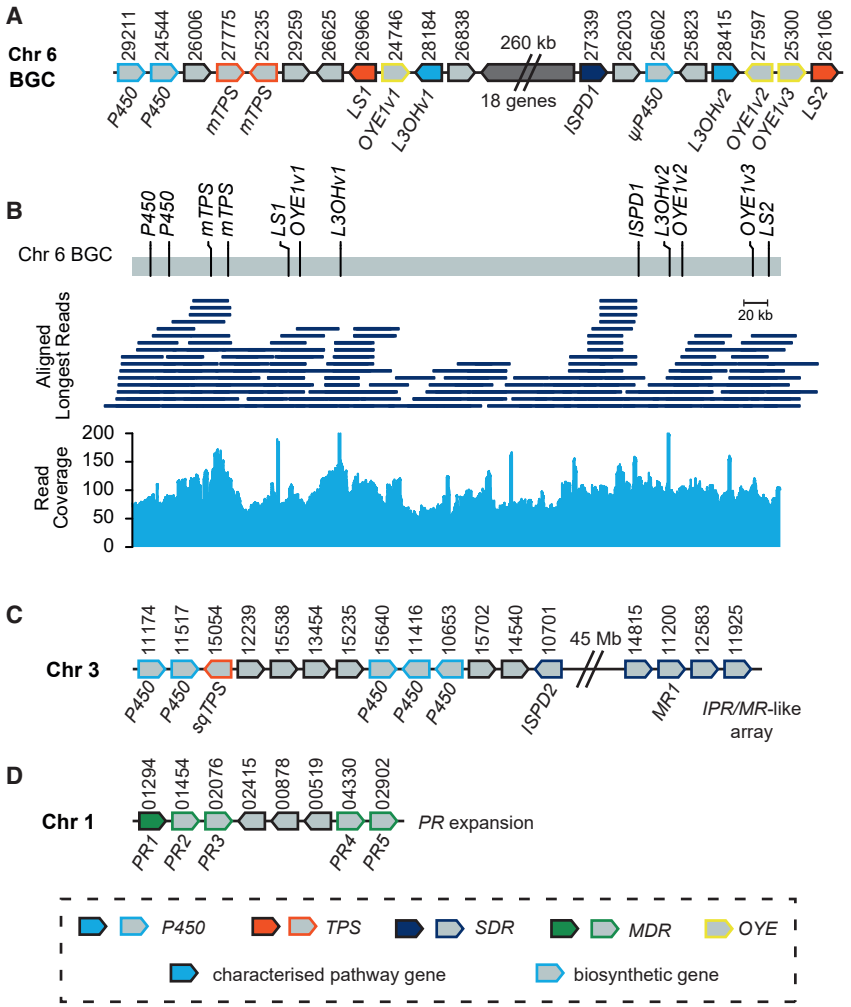

Figure 2. Genomic basis for *p*-methane biosynthesis in *S. tenuifolia*.

(A) Chromosome 6 bipartite biosynthetic gene cluster (BGC). Functional annotations or assigned names noted below gene symbols; gene index number noted above (with Sch0000 removed).

(B) Re-mapped PacBio reads onto the chromosome 6 BGC to assess assembly quality.

(C) Chromosome 3 containing biosynthetic gene-rich region including *ISPD* and *L3OH* homologs. An *IPR/MR*-like array is also present.

(D) Chromosome 1 *PR* array. Genes colored by protein family. Color outline/gray interior represents uncharacterized biosynthetic genes and color fill refers to characterized pathway genes. *ISPD*, isopiperitone dehydrogenase; *L3OH*, limonene 3-hydroxylase; *LS*, limonene synthase; *PR*, pulgeone reductase; *IPR*, isopiperitenone reductase; *MR*, menthone reductase; *P450*, Cytochrome P450; *TPS*, terpene synthase (sq, sesqui-; m, mono-); *SDR*, short-chain dehydrogenase/reductase; *MDR*, medium chain reductase; *OYE*, old yellow enzyme.

characterizing putative WGDs at $K_S \simeq 0.1$–0.2 and 1 (Supplemental Figure 2 and Supplemental Table 6), respectively, corroborating previously reported estimates for Nepetinae based on transcriptome data (Godden et al., 2019).

Bipartite BGC for *p*-menthane biosynthesis

Examination of the genome sequence revealed that certain genes encoding enzymes involved in the biosynthesis of *p*-menthane monoterpenoids were present in a bipartite BGC on chromosome 6 (Figure 2A). The cluster contains copies of genes encoding the previously characterized enzymes LS (Maruyama et al., 2001) and limonene 3-hydroxylase (L3OH) (Liu et al., 2021a), as well a gene with 92% nucleotide identity (nt id) to the previously described isopiperitenol dehydrogenase (*ISPD*) (Srividya et al., 2022) (Supplemental Figures 3–5). The BGC features two regions separated by 260 kb. The 5′ region features *LS1* (Sch000026966) and *L3OHv1* (Sch000028184) separated by a gene annotated as 12-oxophytodienoate reductase, a member of the OYE family (Sch000024746, *OYE1v1*) (Supplemental Figure 6) (Schaller and Weiler, 1997; Toogood et al., 2010). Just upstream of *LS1* are two uncharacterized monoterpene synthases and two cytochrome P450s. The 3′ BGC contains *ISPD1* (Sch00002 7339), three intervening genes and then *L3OHv2* (Sch000028415) and *LS2* (Sch000026106), separated by two copies of *OYE* (Sch000027597, *OYE1v2*; Sch000025300, *OYE1v3*).

There is a degree of symmetry across the unusual bipartite cluster, with the *LS1-OYE1v1-L3OHv1* sequence in the 5′ region reflected in the *L3OHv2-OYE1v2-OYE1v3-LS2* sequence in the 3′ region. Alongside the reversal of the *LS* and *L3OH* gene order, the gene orientations are also mirrored. Furthermore, the paralogs are also remarkably similar: the *L3OH* nucleotide coding sequences are identical, as are the three copies of *OYE1*. The two *LS* paralogs have 98% nt id, with all changes occurring at the N terminus, within the first 9% of the sequence's length (Supplemental Figure 3).

To assess the possibility of an assembly error causing the unusual symmetric bipartite BGC, we re-mapped PacBio reads onto all assembled pseudomolecules, and then extracted the top 10% of the longest reads mapping to the BGC region. The mapping shows good coverage over the whole region (>50×) (Figure 2B). Remarkably, the whole region is spanned by just two long reads, and the boundary between these reads has good coverage. This indicates that the BGC is not an assembly error but a real feature of the genome sequence.

The genes contained in functional BGCs typically demonstrate coordinated expression across different tissues and treatments (Nützmann and Osbourn, 2014; Smit and Lichman, 2022). Using RNA-seq, we examined gene expression across different tissues (seedlings, leaves, or roots), ages (10, 20, or 35 days), and treatments (0, 100, or 300 μM methyl jasmonate). Compared with genes in the wider chromosomal region, the genes in the bipartite cluster co-express (Figure 3). This indicates that they are under similar regulatory control and are functionally related, supporting the identification of this region as a functional *p*-menthane BGC.

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

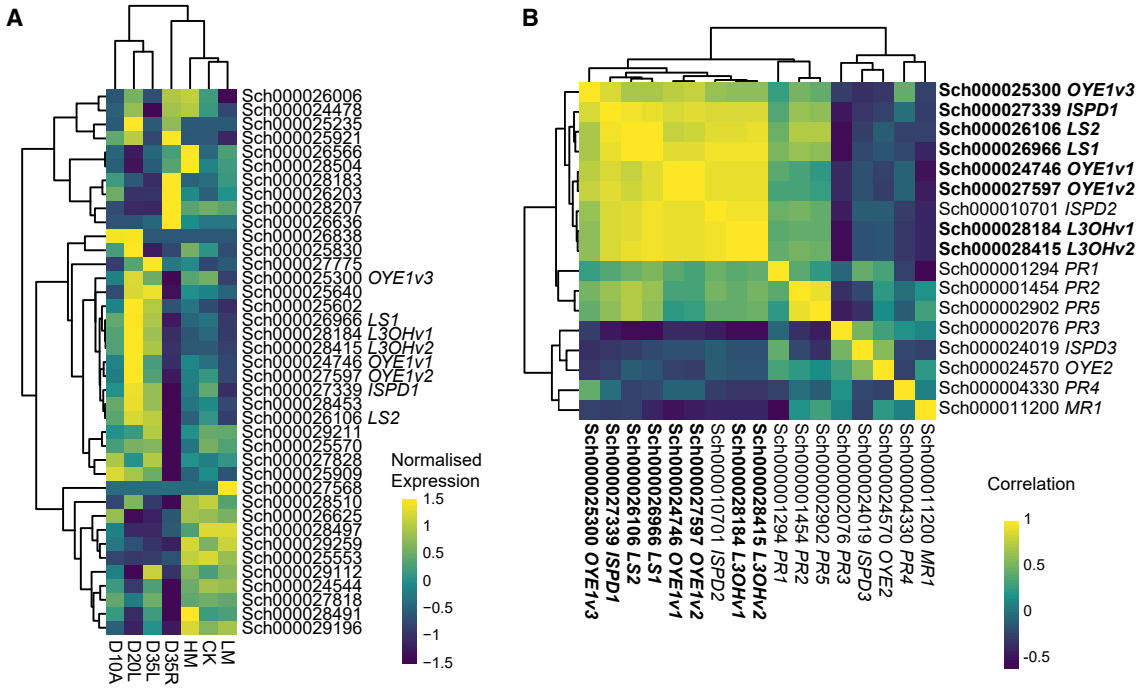

Figure 3. Coordinated expression of *p*-menthane biosynthesis genes in BGC.

(A) Expression levels of genes in the chromosome 6 synteny block encapsulating the BGC region. Heatmap clustered by hierarchical clustering (relationships shown with dendrogram). Conditions: D10A, 10-day-old aerial tissues; D20L, 20-day-old leaves; D35L, 35-day-old leaves; D35R, 35-day-old roots; HM, high methyl jasmonate-treated leaves; LM, low methyl jasmonate-treated leaves; CK, untreated leaves. RNA-seq experiments conducted as biological triplicate and values averaged and normalized.

(B) Expression correlation of *p*-menthane pathway genes and homologs. Heatmap of pairwise Pearson correlation coefficients calculated from normalized gene expression from RNA-seq experiments, clustered by hierarchical clustering (relationships shown with dendrogram). BGC genes shown in bold.

Biosynthetic regions outside the BGC

Genes with connection to the *p*-menthane pathway were also found in other genomic locations. Most notably, a region on chromosome 3 highly enriched in biosynthetic genes including P450s and an *ISPD* paralog (*ISPD2*, Sch000010701) (Figure 2C). Curiously, the previously characterized *StISPD* (Srividya et al., 2022) is more closely related to this chromosome 3 *ISPD2* (98% nt id) than it is to the chromosome 6 *ISPD1* (92% nt id) (Supplemental Figure 5). Another *ISPD* paralog (*ISPD3*, Sch000024019) is present on unplaced scaffold 59 and has 83% nt id with the previous reported *StISPD* (Srividya et al., 2022).

The genome also features two arrays of reductases potentially connected to *p*-menthane biosynthesis. Chromosome 3 features a tandem array of four short-chain dehydrogenase/reductases (*SDRs*), with homology to *Mentha IPRs* and menthone/isomenthone reductases (*MR*) (Figure 2C and Supplemental Figure 7) (Ringer et al., 2003; Srividya et al., 2022). On chromosome 1 is an array of five closely related (80% nt id) pulegone reductase (*PR*) homologs, with one of these (Sch000001294, *PR1*) being identical to the previously characterized *StPR* gene (Figure 2D and Supplemental Figure 8) (Liu et al., 2021b).

Experimental validation of gene activity and function

To validate the function of the genes associated with *p*-menthane biosynthesis, we characterized them through recombinant protein assays and virus-induced gene silencing (VIGS) (Supplemental Tables 7 and 8). The two full-length LSs found encoded in the BGC were expressed in *Escherichia coli* and demonstrated limonene formation from GPP (Figure 4A). The previously described *St*LS has 99% and 97% amino acid identity with LS1 and LS2, respectively, so this result was expected (Maruyama et al., 2001). Suppressing the expression of *LS* genes using VIGS led to a significant reduction of limonene and pulegone, validating its role in the pathway (Figure 4B and Supplemental Figure 9A).

The L3OHs encoded by the BGC genes *L3OHv1* and *L3OHv2* are identical to each other and to previously characterized *St*L3OH (CYP71D544) (Liu et al., 2021a). We confirmed their activity, hydroxylation of limonene to *cis*-isopiperitenol, using yeast feeding assays (Figure 4C). VIGS targeting both *L3OH*s confirmed their role in the pathway as reduction in gene expression led to accumulation of limonene and reduction of pulegone compared with the empty vector control (Figure 4D and Supplemental Figure 9B).

The enzyme encoded by *ISPD1* (Sch000027339), was characterized using recombinant expression in *E. coli*. This enzyme was active on isopiperitenol, forming isopiperitenone in the presence of NAD^+ (Figure 4E). We targeted the gene expression using VIGS but were only able to knock down gene expression by around 50% (Supplemental Figure 9C). Consequently, we saw no

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

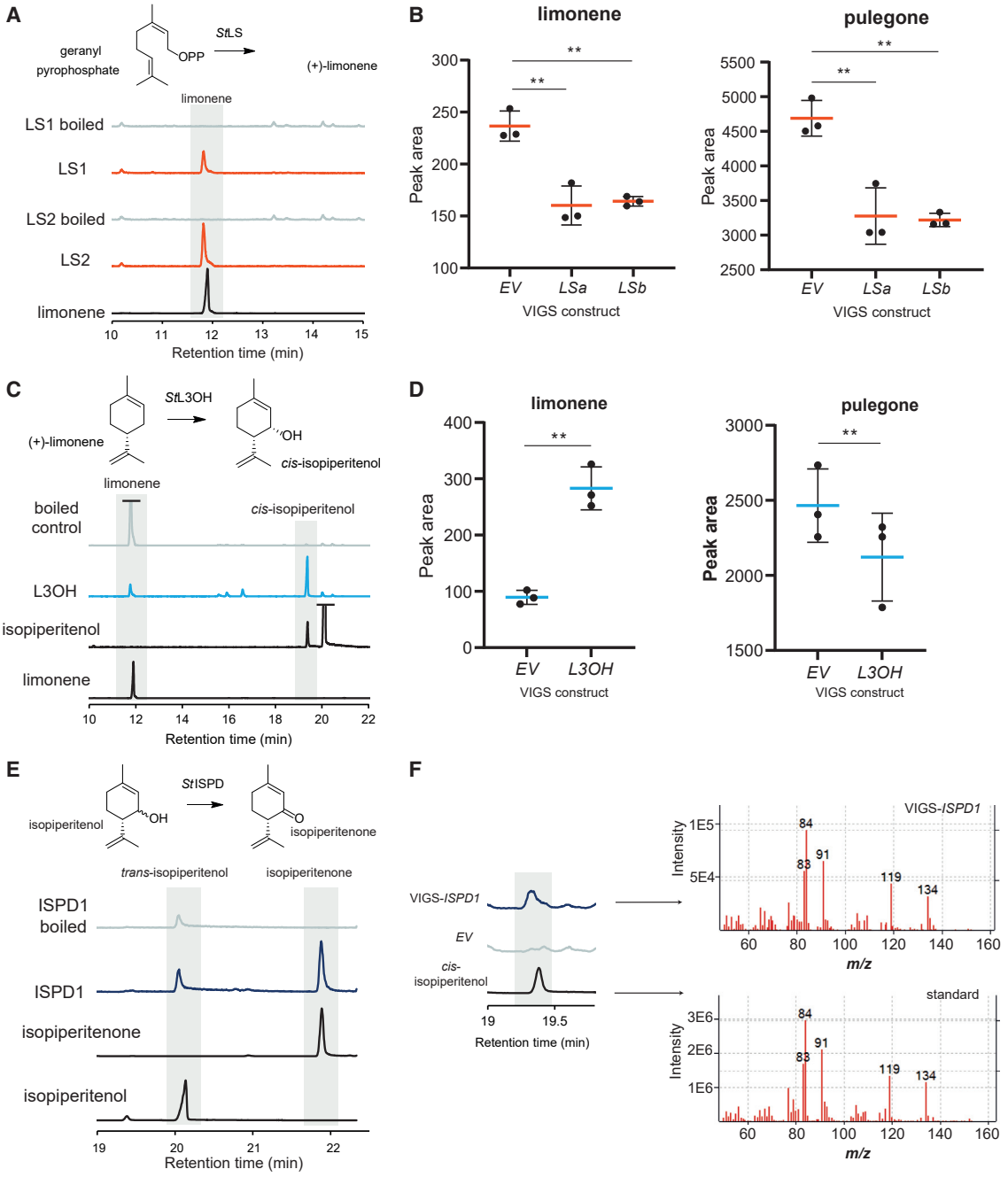

Figure 4. Validation of biosynthetic pathway steps.

(A) Limonene synthase (LS) enzyme activities, converting geranyl pyrophosphate to limonene.

(B) Metabolite content of VIGS silenced leaves measured by GC–MS. Two different regions for *LS* silencing were used (LSa and LSb).

(C) Limonene 3-hydroxylase (L3OH) enzyme activity, converting (+)-limonene to *cis*-isopiperitenol, with additional NADPH. The isopiperitenol standard consists of two isomers, the *trans* major component, and *cis* minor component.

(D) Metabolite content of VIGS silenced leaves with *L3OH* silencing construct.

(E) Isopiperitenol dehydrogenase (ISPD) enzyme activity. The isopiperitenol substrate consists of a *trans* major isomer and *cis* minor isomer. NAD^+ was added to the reaction.

(F) Metabolite content of leaves silenced with *ISPD* targeting construct. A peak corresponding to *cis*-isopiperitenol emerges in silenced plants. Electron impact (EI) spectrum of isopiperitenol standard (inset, bottom) matches the new peak in the VIGS-ISPD silenced plants (inset, top). Significance tests are one-tailed *t*-tests (*$p < 0.05$, **$p < 0.01$), EV, empty vector. qRT–PCRs to verify silencing are found in Supplemental Figure 9, and details of regions targeted in Supplemental Figures 3–5. Chromatograms of standards have been scaled for ease of comparison. All chromatogram traces are total ion chromatograms (TICs). Peak identities were verified using EI spectral matches to standards and NIST library.

notable changes to the concentration of the majority of compounds in the *p*-methane pathway. However, we did observe a new peak emerge in the chromatogram, corresponding to isopiperitenol, the substrate of ISPD (Figure 4F). With less ISPD present, the rate of isopiperitenol consumption appears to have decreased and it has accumulated. The combination of activity, expression, and VIGS data support that, in the specific plant investigated here, the clustered *ISPD1* is the active gene in the *p*-menthane pathway.

The PR from *S. tenuifolia*, catalyzing reduction of (−)-pulegone to (+)-menthone and (−)-isomenthone, is a member of the medium chain dehydrogenase family and has been previously characterized (Liu et al., 2021b). On the genome, there is a tandem array of *PR* paralogs (Figure 2D and Supplemental Figure 8). We investigated their activity using recombinant expression in *E. coli*. PR1 (Sch000001294) is identical to the previously described enzyme, and was able to reduce pulegone to menthone (Supplemental Figure 10A). Enzymes encoded by the paralogous genes *PR2* (Sch000001454) and *PR5* (Sch000002902) were also examined and made the same products.

In *Mentha*, the reduction of isopiperitenone to pulegone is catalyzed by IPR, a NADPH-dependent member of the SDR family (Ringer et al., 2003). The IPRs are closely related to (−)-menthone:(+)-neomenthol dehydrogenase enzymes (MR) (Lygidakis et al., 2016). Unlike other pathway steps, no clear ortholog to *Mentha IPR* could be found in the *S. tenuifolia* transcriptome (Srividya et al., 2022); searching the genome also failed to yield close orthologs. Genes homologous to *Mentha IPR/MR* were found in an array of four genes on chromosome 6, which shared 67%–88% nt id (Supplemental Figure 7). The enzyme encoded by the most highly expressed gene (*MR1*, Sch000011200) was produced recombinantly in *E. coli* and tested with NADPH and various substrates. No products could be detected with isopiperitenone or pulegone (Supplemental Figure 10B and 10C). However, the enzyme was able to reduce menthone to produce menthol, despite menthol not being a known naturally occurring compound in *S. tenuifolia* (Supplemental Figure 10D) (Liu et al., 2018).

Discovery of an old yellow enzyme as an isopiperitenone reductase

With the key candidate enzymes for IPR failing to show expected activity, we searched for other enzymes capable of catalyzing the reaction. We had noted the presence of the three identical copies of *OYE1* in the BGC. OYEs are flavin-dependent reductases known for their ability to reduce the double bond of an α,β-unsaturated carbonyl (1,4- or Michael reduction) (Williams and Bruce, 2002). The IPR reaction is of the same reaction class. The BGC localization and enzyme class led us to suspect that the *OYE* genes encoded the missing IPR.

We cloned *OYE1* and tested it using recombinant expression in *E. coli*. The purified enzyme was capable of reducing isopiperitenone to isopulegone in the presence of NADPH (Figure 5A). We also examined a closely related paralog, *OYE2* (Sch000024570), also on chromosome 6 but not in the cluster. The *OYE2* enzyme product has 69% amino acid identity to OYE1 and is also able to act as an IPR (Figure 5A). We also tested OYEs for activity with piperitenone, but failed to see activity (Supplemental Figure 10E), and with carvone, which was successfully turned over by OYE1 (Supplemental Figure 10F).

The function of *OYE1* paralogs in the *p*-menthane biosynthetic pathway is supported by their co-expression with other pathway genes (Figure 3). In contrast, *OYE2* does not share a similar expression pattern (Figure 3B). To further validate the *in planta* role of *OYE*s, we targeted them with VIGS silencing. We examined three VIGS constructs, one targeting *OYE1*, another targeting *OYE2*, and a hybrid construct combining both genes (Supplemental Figure 6 and Supplemental Table 8). Due to the high identity of the genes, off-target silencing was observed and the specific contributions of *OYE1* and *OYE2* could not be resolved with this method (Supplemental Figure 9D), although co-expression supports *OYE1*s as the active pathway genes (Figure 3). Nevertheless, with all constructs we observed an increase in isopiperitenone and a reduction in isopulegone, the substrate and the product, respectively. We also saw a decrease in pulegone, a downstream product, and an increase in piperitonone, a shunt product (Figure 5B and 5C). The combination of *in vitro*, expression, and VIGS data confirm that the OYEs act as IPRs in *S. tenuifolia*, and these were renamed accordingly.

Therefore, unlike the other *p*-menthane pathway genes, the *StIPR* is not homologous to *Mentha* genes. Instead, *Mentha* and *S. tenuifolia* have evolved an IPR enzyme convergently from different enzyme families, SDR and OYE, respectively. While convergence in plant biosynthetic pathways is not rare, it is more typically found as parallel evolution, where the enzymes originate from the same families (Pichersky and Lewinsohn, 2011). The result highlights the power of BGC analysis for gene discovery.

Phylogenomic analysis of the BGC and related regions

To examine the evolutionary origin of the *S. tenuifolia* BGC and *p*-menthane pathway, we compared the *S. tenuifolia* genome with other Lamiaceae genomes (Supplemental Tables 9 and 10). Synteny analysis revealed that, in other species, the region syntenic to the *S. tenuifolia* bipartite BGC contains monoterpene synthase genes (*mTPS*), including putative *LS*s, but lacks any evidence of *L3OH*, *ISPD*, and *OYE* (Figure 6A and 6B). This indicates that the *p*-menthane BGC is unique to the *S. tenuifolia* lineage, emerging in a locus already rich in monoterpene synthases. The BGC emerged in the *S. tenuifolia* lineage through the recruitment of pathway genes.

In *S. tenuifolia*, the two *LS* paralogs flank a region of approximately 300 kb, with the two parts of the BGC at either end. Compared with other genomes, this whole region is inverted in *S. tenuifolia*, with the boundaries of the inversion approximately marked by *LS1* and *LS2*. Despite some gene gain/loss within each lineage, the inversion is clearly marked by reversal of gene order and orientation in *S. tenuifolia* compared with other genomes (Figure 6A and Supplemental Figure 11). The monoterpene synthase content shows variation between species, with *H. officinalis* only having a single *mTPS* copy in the region, *N. cataria* and *Mentha longifolia* (L.) Huds. having two copies flanking the region, and *N. racemosa* containing three *mTPS*s but on one side of the syntenic region.

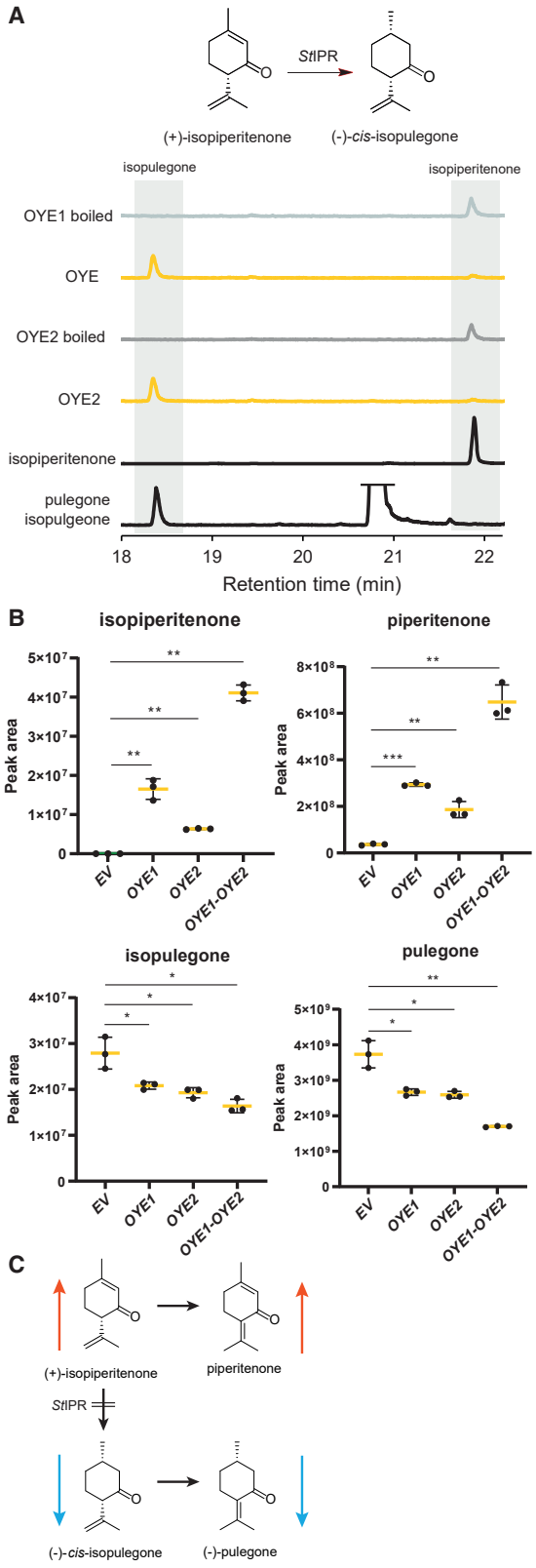

Figure 5. Activity of the isopiperitenone reductase from *S. tenuifolia*.

(A) IPR enzyme activity catalyzed OYEs, with addition of NADPH. Chromatograms of standards have been scaled for ease of comparison. All peak identities were verified using EI spectral matches to standards and NIST library.

(B) Metabolite content of VIGS silenced leaves, with IPR silencing constructs. Significance tests are one-tailed *t*-tests (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). EV, empty vector. qRT–PCR to verify silencing is found in Supplemental Figure 9, and details of regions targeted on Supplemental Figure 6.

(C) Schematic of effect of VIGS silencing of IPR. When the IPR reaction is blocked the substrate and its shunt product piperitenone increases, and the product isopulegone and its downstream product pulegone decreases.

Monoterpenoid biosynthesis in *Schizonepeta tenuifolia*

Phylogenomic analysis of the *mTPS*s in this region shows syntenic *mTPS* orthologs are present in the early diverging Lamiaceae species, *C. americana* (Figure 6B and Supplemental Figure 11). The locus therefore appears to be a region associated with monoterpene diversity across the Lamiaceae, with variation caused by sequence changes in *mTPS*s alongside local duplications.

The *S. tenuifolia* chromosome 3 region described above (Figure 2C), which contains paralogs of *L3OH* and *ISPD*, lies on the edge of a larger region containing many *P450* and *TPS* genes (Supplemental Figure 12). This region corresponds to a diterpenoid-related BGC described in multiple Lamiaceae species (Zhao et al., 2019; Ma et al., 2021; Li et al., 2022). Combined phylogenetic and synteny analysis indicates that the *L3OH* paralog (CYP71) duplications in this region predate the *S. tenuifolia* lineage (Supplemental Figures 12 and 13A). For example, closely related genes are found in syntenic locations in *H. officinalis* and *N. cataria*. In contrast, no other genomes contain syntenic *ISPD* homologs, indicating that the *ISPD2* gene was introduced into this region uniquely in the *S. tenuifolia* lineage (Supplemental Figures 12 and 13B).

There is no syntenic region to the chr 3 diterpenoid BGC or *L3OH*/*ISPD* region in the *M. longifolia* genome. Furthermore, the *mTPS*s in the region syntenic to the *S. tenuifolia* chr 6 BGC are uncharacterized *LS*-like paralogs with 77% nucleotide identity compared with the bone fide *M. longifolia LS* (Figure 6A and 6B). Instead, the active *L3OH* and *LS* in *M. longifolia* can be found as duplicated gene pairs on chr 5 (Figure 6C), in a region that has no syntenic relationship to *S. tenuifolia* chr 6 BGC or chr 3 *L3OH*/*ISPD*s. The *M. longifolia LS*s therefore have originated from a dispersed duplication, which moved them from the *mTPS*/BGC syntenic region to chr 5 where they formed gene pairs (Figure 6B and 6C). Based on this it appears that the genomic association of *L3OH* and *LS* occurred independently in *S. tenuifolia* and *M. longifolia* (Figure 7). This is an unusual example of the convergent evolution of genome structures, essentially independent evolution of collinearity.

DISCUSSION

This work demonstrates how genomics can enhance our understanding of plant specialized metabolism, aiding biosynthetic gene discovery and revealing the evolutionary origins of lineage-specific pathways. While the majority of the genes in the *S. tenuifolia* pulegone pathway had previously been described based on homology to *Mentha* spp. enzymes (Maruyama et al., 2002; Liu et al., 2021a; Srividya et al., 2022), the *IPR* was missing. The chromosome-level *S. tenuifolia* genome assembly revealed the presence of a monoterpenoid-related BGC.

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

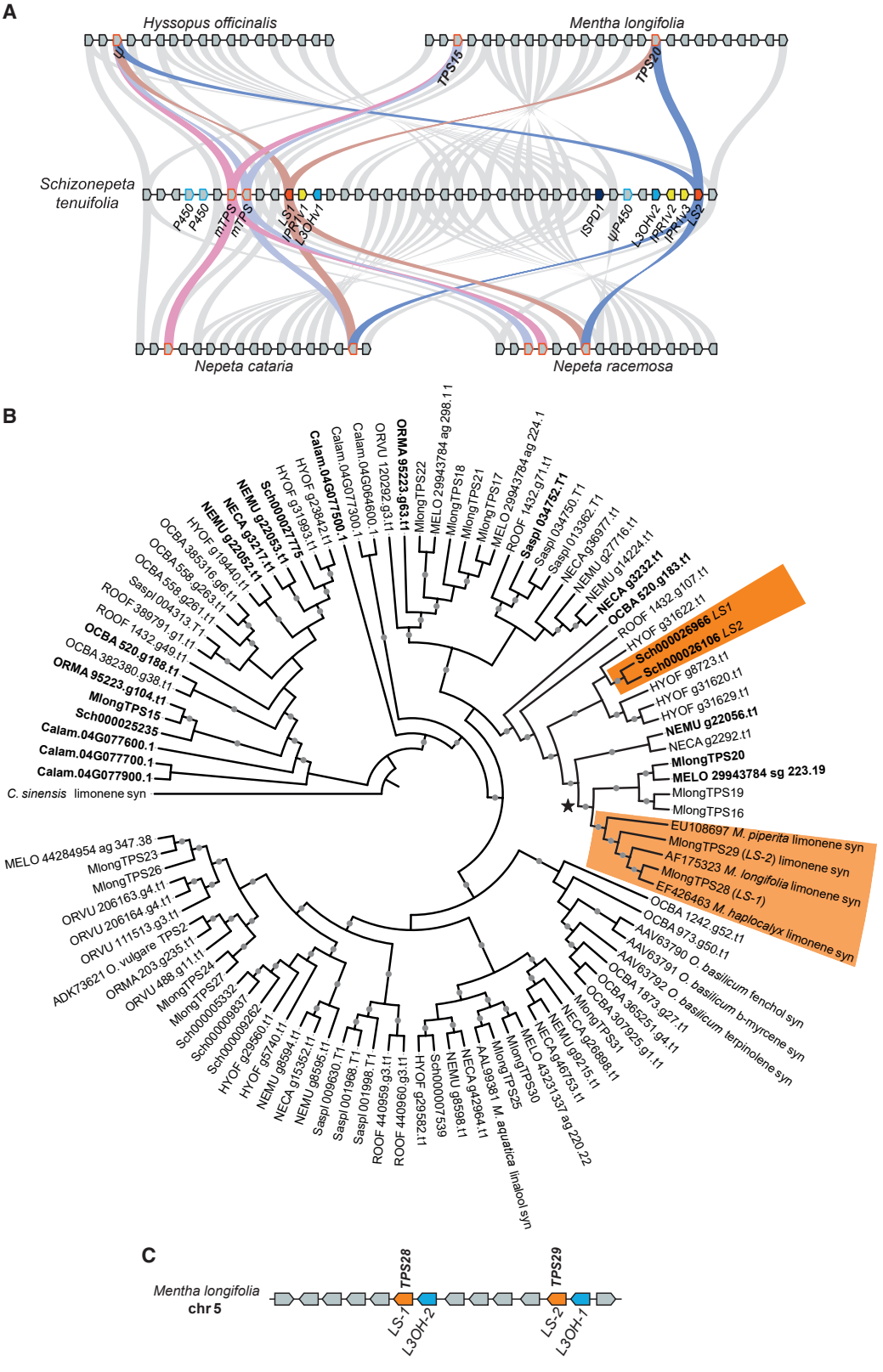

Figure 6. Phylogenomic analysis of BGC in *S. tenuifolia*.

(A) Syntenic analysis of BGC. Collinearity (synteny) analysis of bipartite BGC compared with other high-quality Lamiaceae genomes. Lines indicate homologous genes. Colored lines are monoterpene synthases. Gene size and space has been normalized. Note the lack of *L3OH*, *IPR*, and *ISPD* homologs in other genomes indicating unique assembly of the BGC in *S. tenuifolia*. Also, note the relative inversion of the region between the two BGC regions.

(legend continued on next page)

Alongside the known pathway genes *LS*, *L3OH*, and *ISPD*, the cluster contained copies of a gene encoding an *OYE*. This gene, which has no homology to the *Mentha IPR* (Supplemental Figure 14), was shown to encode the missing IPR through *in vitro* (enzyme assays) and *in vivo* (VIGS) methods (Figure 5).

The presence of non-homologous IPRs in *M. longifolia* and *S. tenuifolia* is an example of convergent evolution, with the two lineages evolving enzymes from different superfamilies to catalyze the same reaction. This stands in contrast to the preceding pathway enzymes, LS, L3OH, and ISPD, which are very closely related in the two lineages (Figure 6B and Supplemental Figure 13). These enzyme pairs would be considered true orthologs if the common ancestor of *M. longifolia* and *S. tenuifolia* had a similar monoterpenoid pathway. Indeed, closely related Lamiaceae species are also rich in monoterpenoids derived from limonene, including *Agastache rugosa* (Fisch. & C.A. Mey.) Kuntze, which also produces pulegone (Mint Evolutionary Genomics Consortium, 2018; Dang et al., 2022). However, the monoterpenoid pathways in *Mentha* spp. and *S. tenuifolia* are of the opposite enantiomeric series set by LS (Figure 1A) (Srividya et al., 2020, 2022). Kinetic analysis of *p*-menthane enzymes from *Mentha x piperita* L. and *S. tenuifolia* indicate that they (with the exception of L3OH) are more active (i.e., higher k_{cat}/K_m) with the substrate enantiomers derived from their pathway (Srividya et al., 2022). However, whether these homologs are derived from convergent, parallel, or divergent evolution has not yet been determined.

The BGC containing the first four steps of pulegone biosynthesis has an unusual bipartite structure, with the core cluster (*LS*, *L3OH*, and *IPR1*) duplicated and separated by over 260 kb. *ISPD* appears only once, indicating it is a late acquisition to the BGC, possibly occurring after the duplication. While this bipartite BGC structure has not be described previously, similar BGC structures are known. For example, split BGCs, spread over multiple loci, have been described, such as momilactone biosynthesis in rice (Miyamoto et al., 2016). Large, relatively diffuse BGCs are also known, such as the 260 kb taxol-associated locus in *Taxus* genomes, sitting in a wider 72 Mb biosynthetically rich region (Xiong et al., 2021). Duplicated BGCs have also been described in polyploids, such as the nepetalactone BGC in *N. cataria* (Lichman et al., 2020a), although these would occur on separate chromosomes.

The duplication of the *S. tenuifolia* monoterpenoid BGC appears to be relatively recent, judging by the near-identical nature of genes in the two regions (Supplemental Figures 3, 4, and 6). A BGC duplication may initially be advantageous through a dosage effect, although eventually one BGC copy may decay or diversify in a manner akin to gene duplication (Lichman et al., 2020b). While it appears that genes in both regions of the BGC co-express and contribute to *p*-menthane biosynthesis, some paralogs (i.e., *L3OH*s, *IPR1*s) are so similar it is challenging to determine the exact contribution of each specific gene using the methods employed here (VIGS, qRT–PCR, and RNA-seq). Ultimately, genome editing may be required to examine each gene specifically.

Comparisons with related genomes indicate the BGC structure arose through an inverted duplication (Figures 6A and 7). Inversions as a mechanism of BGC growth has recently been described in *Arabidopsis* (Liu et al., 2020). Intriguingly, chromosomal inversions are also a feature of supergenes, loci that are similar to BGCs in that they contain multiple genes contributing to a specific adaptive trait (Thompson and Jiggins, 2014). Inversions in supergenes are proposed to be a mechanism for suppressing recombination and fixing genes in adaptive proximity (Kim et al., 2022).

The mechanism of the inverted duplication is unknown. Inversions can occur via ectopic recombination of homologous sequences at different loci, inverting the region between the breakpoints. However, this would not account for the duplication. Transposable elements (TEs) may be involved, either serving as homologous sequences for recombination breakpoints, or perhaps as active partners in the rearrangement (Huang and Rieseberg, 2020). There are indeed plenty of both DNA and RNA TE signatures in the BGC, although it is not known if they are enriched or active in this region (Supplemental Figure 15).

The phylogenomic comparison highlights the importance of particular biosynthetic genomic "hotspot" regions across a multiple lineages and pathways. The *S. tenuifolia* BGC formed around *LS* which was present in a region enriched in *mTPS*. This region is found across the Lamiaceae and originates from a *mTPS* tandem expansion, as seen in *C. americana* (Supplemental Figures 11 and Figure 6B). Furthermore, *L3OH* and *ISPD* paralogs are found in a chr 3 region which is highly enriched in other *TPS* and *P450* genes, and where a diterpene BGC resides in closely related species (Supplemental Figure 12) (Zhao et al., 2019; Ma et al., 2021). The link between the chr 5 BGC and chr 3 biosynthetic region points toward genomes containing multiple biosynthetic hotspots that exchange genetic material (Liu et al., 2019; Smit and Lichman, 2022).

In the *M. longifolia* lineage, rather than receiving genes into the *LS* locus to build a BGC, instead a *LS* has undergone a dispersed duplication into chr 5, alongside *L3OH* (Figure 6B, 6C, and 7) (Vining et al., 2022). This *LS-L3OH* gene pair has then undergone a duplication. However, no other pathway genes are present in this locus so it cannot be strictly termed a BGC (Nützmann et al., 2016; Smit and Lichman, 2022). The genomic association of *LS* and *L3OH* in *M. longifolia* and *S. tenuifolia* are independent events. This is an example of the convergent, or

(B) Limonene synthase maximum likelihood phylogenetic tree, depicted as a cladogram without branch lengths. All genes in chr 6 BGC syntenic region are in bold **(A)** and Supplemental Figure 11. The *S. tenuifolia LS* paralogs are in a colored box. The characterized *Mentha LS* genes found as gene pairs **(C)** are in a colored box. The star shows a dispersed duplication in *M. longifolia* leading to non-syntenic *LS*s **(C)**. Circles show branches with >85% and >95% support as judged by 1000X SH-aLRT and UltraFast Boostrapping replicates, respectively. See Supplemental Tables 9 and 10 for sources of genes and acronyms.

(C) Duplicated *p*-menthane biosynthesis gene pairs in *M. longifolia* genome. These appear unique to *Mentha* spp. and therefore independently assembled compared with the *S. tenuifolia* BGC.

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

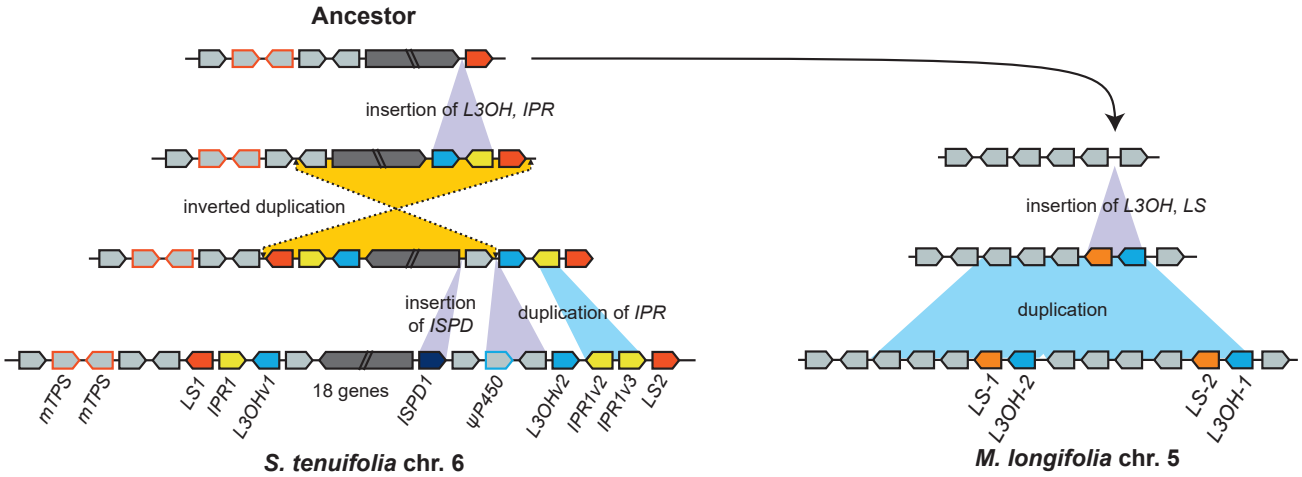

Figure 7. Proposed evolutionary origin of *S. tenuifolia* bipartite BGC and *M. longifolia* duplicated gene pairs.
Based on the syntenic and phylogenetic analysis, we propose the origins of *p*-menthane-related genome structures in the *S. tenuifolia* and *M. longifolia* lineages. In the *S. tenuifolia* lineage, genes encoding *L3OH* (blue) and *IPR* (from old yellow enzyme family, yellow) were inserted into a region containing monoterpene synthases (red) to form a *p*-menthane BGC. An inverted duplication of the region led to its development into a bipartite BGC. Subsequently *ISPD* moved into the cluster and *IPR* underwent tandem duplication. In contrast, in the *Mentha longifolia* lineage, *LS* and *L3OH* inserted into a non-syntenic region to form a gene pair. A five-gene region containing the gene pair underwent segmental duplication.

parallel, evolution of gene order. The convergent evolution of a BGC has previously been described, with a bryophyte and rice independently evolving a momilactone BGC (Mao et al., 2020).

This work highlights the role of BGCs in plant specialized metabolism. We identified a BGC for *p*-menthane biosynthesis in *S. tenuifolia*, enabled by a high-quality genome assembly. BGC genes' involvement in pulegone biosynthesis was verified by co-expression and VIGS. Enzyme activities were verified by recombinant expression and enzyme assays. We discovered a missing *p*-menthane biosynthesis pathway gene in the BGC: an *IPR* from the OYE family. The BGC has an unprecedented "bipartite" structure, which, through comparative phylogenomics, appeared to be formed through an inverted duplication. We identified two examples of convergent evolution between the *M. longifolia* and *S. tenuifolia* lineages, at the enzymatic (independent origin of IPRs) and genomic level (independent *LS* and *L3OH* associations in non-syntenic regions). The Lamiaceae (mint family), has large chemical diversity; species with cultural, medicinal and economic importance; and relatively manageable genome sizes and cultivation (Mint Evolutionary Genomics Consortium, 2018). This makes it an excellent model for examining metabolic evolution in plants (Lichman et al., 2020a, 2020b; Ma et al., 2021). Such studies reveal how genomes, enzymes, and metabolism have co-evolved to provide the rich chemistry of the natural world.

METHODS

Genome assembly

The *S. tenuifolia* genome was assembled using reads from PacBio, DNBseq, and Hi-C. Details are available as supplemental methods. The BGC was identified by first finding the genes corresponding to previously described *S. tenuifolia* pathway enzymes using sequence similarity (Maruyama et al., 2001; Liu et al., 2021a; Srividya et al., 2022). Then, using genome position annotations (as described in the genome GFF file), the genomic position of the pathway genes, both in the context of a chromosome and relative to each other, were determined.

Chemicals

Commercially available monoterpene standards, (−)-limonene (CAS: 5989-54-8, Sigma-Aldrich), (+)-limonene (CAS: 5989-27-5, Sigma-Aldrich), (−)-pulegone (CAS: 3391-90-0, Sigma-Aldrich), (+)-pulegone (CAS: 89-82-7, Sigma-Aldrich), (−)-menthone (CAS: 14073-97-3, Sigma-Aldrich), (+)-menthone (CAS: 3391-87-5, Sigma-Aldrich), (−)-menthol (CAS: 2216-51-5, shyuanye), (+)-menthol (CAS: 25356-60-2, shyuanye), (−)-isomenthone (CAS: 18309-28-9, Toronto Research Chemicals), (+)-isomenthone (CAS: 1196-31-2, ZZSRM), and camphor (CAS: 76-22-2, Macklin) were used for annotation peak or normalization. Isopiperitenol and isopiperitenone was synthesized as described previously (Dethe et al., 2018). D-Glucose 6-phosphate disodium salt hydrate (CAS: 3671-99-6, Solarbio), G-6PDH (CAS: 9001-40-5, shyuanye), NADPH tetrasodium salt hydrate (CAS: 2646-71-1, shyuanye), NAD^+ (CAS: 53-84-9, shyuanye), methanol (CAS: 67-56-1, shhuishi), n-hexane (CAS: 110-54-3, yonghuachem), D-Sorbitol (CAS: 50-70-4, shyuanye), NaCl (CAS: 50-70-4, qschem), glycerol (CAS: 56-82-5, shhuishi), and KH_2PO_4 (CAS: 7778-77-0, nanjingchem) were used for enzyme expression and enzyme assays.

Plant growth conditions

S. tenuifolia seeds were obtained from Hebei province, China. They were grown in a growth incubator under 10 000 l× intensity and 50% humidity with 16/8 h light/dark photoperiod at 25°C (Liu et al., 2021a). Plant tissues were carefully removed, immediately snap-frozen in liquid nitrogen, and stored at −80°C for RNA extraction.

Gene cloning and enzyme expression

Total RNA was extracted using a FastPure Plant Total RNA Isolation Kit (Polysaccharides/Polyphenolics-rich) (Vazyme Biotech, Nanjing, China). cDNA libraries were formed from young leaf RNA using a HiScript III 1st Strand cDNA Synthesis Kit (+gDNA wiper) according to kit instructions. Single-strand cDNA was used as the template for PCR amplification of the target cDNA with 2× PrimeSTAR Max DNA Polymerase (Takara, Japan) and gene-specific primers designed based on annotated results from the transcriptome database. The PCR products were separated

using the GeneJET Gel Extraction Kit (Thermo Scientific), and the purified DNA fragment was subcloned into the pCE2TA/Blunt-Zero vector (Vazyme Biotech) for sequencing. The primers and vectors used for cloning and expressing are listed in Supplemental Table 7. All cloned and subcloned genes were sequenced by Sanger Sequencing (Sangon Biotech, Shanghai, China). The gene model sequence of Sch000024570 was different to the cloned *OYE2* as it had an extended N terminus and extra inserted exon.

Single colonies were used to inoculate LB medium (5 ml) containing 100 ng/l kanamycin. The culture (1 ml) was transferred into 100 ml of the same medium and continued to grow at 37°C to reach an OD_{600} of approximately 1.0. Protein expression was induced by adding isopropyl β-thiogalactopyranoside to a final concentration of 1 mM. After 24 h incubation at 16°C, the cells were harvested by centrifugation (15 min at 4°C and 5000 rpm). The cells were resuspended in 5 ml lysis buffer containing 10 mM Tris–HCl (pH 8.0), 200 mM NaCl, and 5% (v/v) glycerine. The suspension was transferred to an ultrasonic cell disruptor system and ultrasonically disrupted for 25 s each with 35 s on ice between disruptions. The mixture was centrifuged (30 min, 4°C, 12 000 rpm) and the supernatant was carefully collected as crude protein.

Then, 2 ml of Ni-NTA Sefinose (TM) Resin 6FF (C600033, BBI) was taken into a centrifuge tube, centrifuged (4°C, 3000 rpm, 2 min) and the supernatant was carefully discarded. Four milliliters of Binding/Wash Buffer (10 mM imidazole, NaCl, $NaH_2PO_4.2H_2O$ [pH 7.9–8.1]) (C600303-0500, BBI) was added and shaken at 4°C for 30 min to make Binding/Wash Buffer and resin completely homogenous, and the supernatant was discarded by centrifugation (4°C, 3000 rpm, 2 min). The His-tagged crude protein (Sch000025300 and Sch000024570) (2 ml) was mixed with Binding/Wash Buffer (2 ml). The mixture was added to the resin and shaken at 4°C for 30 min. The resin was collected by centrifugation (4°C, 3000 rpm, 2 min) and supernatant removed. The pellet was washed by repeated addition of 4 ml of Binding/Wash Buffer followed by centrifugation (4°C, 3000 rpm, 2 min) until the supernatant elution had an absorbance at 280 nm close to the baseline. The resin was eluted with 2 ml of Elution Buffer (250 mM imidazole, NaCl, $NaH_2PO_4.2H_2O$ [pH 7.9–8.1]) (C600304-0500, BBI), and the supernatant was collected by centrifugation (4°C, 3000 rpm, 2 min). This step was repeated twice, and the supernatants were the purified enzyme protein. SDS–polyacrylamide gel electrophoresis and spectrophotometric analysis (absorbance at 280 nm) were used to check purity and approximate quantity of protein. The crude and purified enzyme protein were stored at −20°C until required for further enzyme activity testing.

For eukaryotic expression of *NtL3OH*, the plasmid pYeDP60-L3OH was introduced into *Saccharomyces cerevisiae* WAT11 using a transformation kit (Frozen-EZ Yeast Transformation II Kit, Zymo Research). The empty vector was used as control. The transformation mixture was plated on SGI medium containing 20 g/l glucose, 6.7 g/l yeast nitrogen base without amino acids, 1 g/l bactocasamino acids (Difco), 15 g/l agar, and 40 mg/l DL-tryptophan. A single colony from SGI medium was grown in 50 ml of SGI liquid medium at 30°C for 12 h. The cells were centrifuged at 1630 *g* for 5 min and resuspended in SGI medium containing 20 g/l galactose. The suspension was diluted to an OD_{600} of 0.4 and induced at 16°C for 12 h.

Enzyme assays

For NADPH-dependent reduction reactions, the reaction (0.4 ml) consisted of buffer B (50 mM KH_2PO_4, 10% sorbitol, 1 mM DTT [pH 7.5]), containing 20 μM substrate, 10 mM NADPH tetrasodium salt hydrate, 6 mM glucose-6-phosphate, 20 U glucose-6-phosphate dehydrogenase (Solarbio), and 50 μl protein extract. For NAD-dependent reduction reactions, 10 mM NAD was used instead of NADPH, and no glucose-6-phosphate or glucose-6-phosphate dehydrogenase was added. For Sch000025300 and Sch000024570, purified enzyme was used in place of protein extract, whereas all other reductase/dehydrogenase reactions were conducted with protein lysate. n-Hexane (0.2 ml) was added on the top of the reaction solution. Reaction was carried out at 31°C for 16 h with slow stirring. The reaction was terminated by placing the reaction vial at −80°C for 2 h. The upper organic phase was transferred into a new 2 ml glass vial containing a conical glass insert, which was then analyzed by GC–MS (Dang et al., 2022).

Activities of L3OH enzymes were determined using yeast feeding assays. Limonene solution was added to the culture to a final concentration of 0.2 mM. The reaction was stopped by sonication for 15 min after 12 h incubation. The products from each reaction were extracted with 200 μl n-hexane. After the enzyme reactions, the glass vial was placed in a freezer (−80°C) for 2 h and the upper organic phase was transformed into a new 2 ml glass vial containing a conical glass insert (while the still frozen aqueous phase of the original glass vial was discarded). Negative controls were generated by heating the reconstitution for 5 min at 100°C (Dang et al., 2022).

GC–MS and chiral-GC method

Gas chromatography–mass spectrometry (GC–MS) was performed using a 6890N GC interfaced with 5973 inert MS instrument (Agilent Technologies, Santa Clara, CA) and chromatographic column (Agilent, 19091S-433-HP-5ms, 30 m × 250 μm × 0.25 μm). Helium was used as the carrier gas at a linear velocity of 1.2 ml/min. The injector temperature was kept at 220°C. Mass spectra were recorded in electron impact ionization mode at 70 eV. The quadrupole mass detector, ion source, and transfer line temperatures were set, respectively, at 150°C, 230°C, and 280°C. The MS was selected ion monitoring mode was used for the identification and quantification of analytes (Liu et al., 2021a). The GC–MS procedure for enzyme assays was maintained at 50°C for 3 min, elevated to 90 °C at 3°C/min, and then raised to 150°C at 5°C/min. The injection volume was 1.0 μl and splitless. The GC–MS procedure for VIGS was maintained at 50°C for 3 min, elevated to 90°C at 3°C/min, then raised to 150°C at 5°C/min, and raised to 220°C at 10°C/min, and maintained at 220°C for 5 min. The injection volume was 1.0 μl and splitless.

VIGS assay

Fragments of open reading frames encoding *S. tenuifolia* phytoene desaturase (460 bp) were amplified with PrimeSTAR Max DNA Polymerase. The target genes were similarly amplified from *S. tenuifolia* cDNA and inserted in pTRV2 vector at EcoRI and SacI sites with a ClonExpress II One Step Cloning Kit (https://www.vazyme.com/product/81.html). The plasmid was transformed into *E. coli* DH5a (Vazyme Biotech). The recombinant plasmid was purified using a GeneJET Plasmid Miniprep Kit and verified by sequencing, using pTRV2 vector-specific primers (Supplemental Table 8). The vectors pTRV1 and pTRV2 target genes were introduced into the Agrobacterium strain *GV3101*, and resuspended in infiltration medium (10 mM $MgCl_2$, 10 mM MES, 200 μM acetosyringone) and adjusted to an OD_{600} of 1.0. Strains harboring pTRV2 constructs were then mixed in a 1:1 ratio with strains harboring pTRV1 with dark activation at room temperature for 3 h and then injected into the cotyledons of 30 10-day-old *S. tenuifolia* seedlings. The plants were placed in the dark for 24 h and then incubated at 25°C with a 16 h/8 h light/dark cycle. As experimental controls, the empty vectors pTRV1 and pTRV2 were injected. The *phytoene desaturase* gene was used as a marker of the silencing effect. At about 3 weeks post-injection, the leaves were collected for analysis. Transformants were screened by PCR with the pTRV2 vector-specific primers to ensure that the cultures contained the expected construct. Then, the metabolites were extracted with 2 ml n-hexane including camphor as an internal standard (final concentration at 30 ng/μl) from 0.2 g fresh leaves, after leaf samples were frozen in liquid nitrogen and pulverized with a tissue grinder (Yamamoto et al., 2021). VIGS experiments were conducted with three biological replicates for initial screening experiments.

Gene expression and essential oil measurement

RNA samples were extracted using a FastPure Plant Total RNA Isolation Kit (Polysaccharides/Polyphenolics-rich) (Vazyme Biotech), and aliquots of 0.5 μg of total RNA were used in reverse transcriptase reactions with the HiScript III 1st Stand cDNA Synthesis Kit (+gDNA wiper) (Vazyme Biotech). The qRT–PCR reactions were performed in triplicate on a QuantStudio 3 Real-Time PCR Systems (Thermo Fisher Scientific, Waltham, MA) using ChamQ Universal SYBR qPCR (Vazyme Biotech). *St*Actin was the reference gene, and the relative expression levels are $2^{-\Delta Ct}$ values of target compared with actin, and then normalized to the average value of empty vector negative control. All the primers used in the qRT–PCR are listed in Supplemental Table 8. Metabolite identification was achieved by comparison with the chromatographic retention and electron impact mass spectrum characteristics of authentic standards. The peak areas of the compounds were corrected by internal standard and then the relative content of the substance can be calculated. Three biological repeats were measured for each sample and significant differences were determined by one-tailed *t*-tests.

RNA extraction, sequencing, and expression analyses

Arial tissue of 10-day-old seedlings, leaves of 20-day-old plants, leaves and roots of 35-day-old plants, and leaves treated with methyl jasmonate (0, 100, and 300 μM) were collected and frozen in liquid nitrogen. Total RNA was extracted from different tissues using TRIzol reagent (Sigma). Sequencing libraries were generated using a NEBNext Ultra RNA Library Prep Kit for Illumina (NEB) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample, in which the mRNA was purified from total RNA using poly-T oligo-attached magnetic beads, and PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina). After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform and paired-end reads were generated.

Adapters and low-quality reads were removed from the raw reads followed by quality assessment using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned using STAR v2.7.10a with the number of reads counted using the "quantMode" option (Dobin et al., 2013). MultiQC was used to assess the quality of the STAR alignments (Ewels et al., 2016). Reads per gene count (counts for unstranded RNA-seq) were concatenated into a single matrix and genes with no expression removed. The DESeq2 (release 3.15) R package was used to normalize the counts data based on library size followed by variance stabilization (Love et al., 2014). Co-expression analyses were performed using the CoExpNetViz application (Tzfadia et al., 2016) within Cytoscape v3.9.1 (Shannon et al., 2003) and GraphBio (Zhao and Wang, 2022).

Syntenic relationships

Interspecies syntenic blocks were identified using the mcscan pipeline of the JCVI suite (Molinari et al., 2008; Wang et al., 2012). Pairwise orthologs to *S. tenuifolia* were identified in full mode using default settings. Macrosyntenic blocks were identified using the default settings. Microsyntenic regions were identified using default settings with a maximum of six iterations to ensure that possible genome duplications were identified.

Orthogroups, single-copy gene discovery, and phylogenetics

Orthogroups were identified with OrthoFinder (Emms and Kelly, 2019) using the protein sequences accompanied with each of the genomes. Where protein sequences were not provided, the CDS regions were extracted from the genome sequences using available GFF annotations. Exonic regions were concatenated and the longest ORF translated using Geneious Prime. Single-copy genes were identified with MarkerMiner v1.2 (Chamala et al., 2015) using default settings and the *A. thaliana* reference. Codon alignments of GOIs and nucleotide alignments of single-copy genes were performed using MAFFT L-INS-i. Alignments were iteratively curated and initial trees inferred using FastTree in Geneious Prime. Gaps were removed for single-copy alignments using trimAl (Capella-Gutiérrez et al., 2009) and individual alignments concatenated into a supermatrix. Maximum likelihood trees were inferred using IQ-Tree 2 (Minh et al., 2020) with ModelFinder (Kalyaanamoorthy et al., 2017), ultrafast bootstraps (UFBoot2, X1000) (Hoang et al., 2017), and SH-aLRT supports (X1000) (Guindon et al., 2010). Partitioning was performed for the single-copy supermatrix (Chernomor et al., 2016). Gene trees were visualized using iTOL 6.3.2 (Letunic and Bork, 2021).

Ancient WGSs

Ancient WGD events were inferred from estimates of divergence at synonymous sites (K_S) among paralogous gene pairs present in the genomes of *S. tenuifolia, H. officinalis* (Lichman et al., 2020a), *N. cataria* (Lichman et al., 2020a), and *N. racemosa* (Lichman et al., 2020a). The DupPipe pipeline (Barker et al., 2010) was used with its default settings to analyze coding sequences representing the longest isoforms of genes filtered from high-confidence (HC) gene sets. For *S. tenuifolia*, the HC gene set included coding sequences representing the longest open reading frames predicted by TransDecoder that had detectable expression and an InterProScan annotation (Haas et al., 2013). For the remaining genomes, coding sequences representing the longest isoform of each gene were filtered from existing HC gene sets downloaded from Dryad (https://doi.org/10.5061/dryad.88tj450). The analyses followed a workflow used previously with Lamiaceae (Godden et al., 2019; Zhao et al., 2019; Hamilton et al., 2020). Peaks in the observed K_S distribution for each species were identified with Gaussian mixture models, as implemented with the mixtools R package (Benaglia et al., 2009), and corroborated with results from a SiZer analysis (Chaudhuri and Marron, 1999).

TEs

TEs were predicted using EDTA (Ou et al., 2019) with the –anno and –evaluate options selected. The *S. tenuifolia* CDS were specified and gene regions were excluded from the prediction using the BED annotation for the CDS regions. Predicted TEs were extracted using the *getfasta* tool from the BEDTools suite (Quinlan and Hall, 2010). The default parameters of the search workflow of MMseqs2 were used to align scaffold_6 extracted TEs to the *S. tenuifolia* TE library. Intergenic coordinates were used to subset TEs for genomic regions of interest. The percentage identity, e value, and bit scores were used as edge weights to construct homology networks in Cytoscape v3.7.2 (Shannon et al., 2003). The kpPlotDensity function of karyoploteR (Gel and Serra, 2017) was used to visualize TE density and genomic positions for the Scaffold 6 BGC. TE densities were calculated using a sliding window of 5000 bp.

Long-read coverage on the BGC

Minimap2 (Li, 2018) was used to re-map PacBio reads to assembled scaffolds. BEDTools (Quinlan and Hall, 2010) was used to extract the coordinates of the top 10% longest reads mapping to the BGC region and visualized using the kpPlotRegions function of karyoploteR. Read coverage of reads longer than 4000 bp was visualized with kpPlotCoverage function karyoploteR (Gel and Serra, 2017).

DATA AVAILABILITY

The cloned genes described in the manuscript have been deposited into NCBI GenBank (OP022342-OP022350). The genome sequence and raw sequencing reads are available at NCBI (BioProject PRJNA859855) and CNGB (BioProject PRJCA014205). The assembled genome is available in the whole-genome sequence database (JANLIY000000000) and the raw reads are available in the Sequence Read Archive (PacBio: SRR20609656; Hi-C: SRR20649371; DNBseq: SRR20648152-SRR20648159; RNA-seq: SRR21857433-44, SRR21822890-8, SRR20649371,

SRR20648152-9, SRR20609656). The genome annotation is provided as supplemental data with this manuscript. All other data are available from the authors upon reasonable request.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This research was supported by the National Natural Science Foundation of China (grant nos. 81973435 and 81473313), the National Natural Science Foundation for Young Scientists of China (grant no. 81903756), the Open Project of the Natural Science Foundation of Nanjing University of Chinese Medicine (no. NZY81903756), research on ecological planting and quality assurance of Jiangsu Dao-di herbs (2021), and a Jiangsu Government Scholarship for Overseas Studies (JS-2020-044). We also acknowledge support from the BBSRC (BB/V006452/1) and UKRI (MR/S01862X/1).

AUTHOR CONTRIBUTIONS

C.L. contributed to genome sequencing, experimental design, and writing. S.J.S. contributed to phylogenetic and genome bioinformatics analyses, and writing. J. Dang contributed to gene cloning, protein overexpression, and enzyme assays. P.Z. contributed to VIGS development and experiments of VIGS. G.T.G. conducted whole-genome duplication analysis. Z.J. contributed to GC–MS operation and data analysis. W. Liu synthesized compounds. L.L. aided with the VIGS experiment. W. Lin aided with protein purification. J. Duan and Q.W. contributed to experimental design and funding acquisition. B.R.L. contributed to experimental design, data analysis, writing, and funding acquisition.

ACKNOWLEDGMENTS

Aspects of this project was undertaken on the Viking Cluster, which is a high-performance compute 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. We thank Achang Wang from Nanjing University of Chinese Medicine for chemical analysis. We also thank the Academician Xiaoya Chen and all the staff at the Shanghai Institute of Plant Physiology and Ecology (Chinese Academy of Sciences) for providing plasmids (pTRV1, pTRV2, pYeDP60), yeast cells (WAT11), *Agrobacterium* strain *GV3101*, and experimental guidance. We are grateful for Professor Sihai Yang and Long Wang from NanJing University for genome sequence analysis. We thank David Nelson from the University of Tennessee for naming P450s. No conflict of interest declared.

Received: July 26, 2022
Revised: December 6, 2022
Accepted: January 4, 2023
Published: January 5, 2023

REFERENCES

Alonso, W.R., Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992). Purification of 4*S*-limonene synthase, a monoterpene cyclase from the glandular trichomes of peppermint (*Mentha x piperita*) and spearmint (*Mentha spicata*). J. Biol. Chem. **267**:7582–7587.

Barker, M.S., Dlugosch, K.M., Dinh, L., Sashikiran Challa, R., Kane, N.C., King, M.G., and Rieseberg, L.H. (2010). Evopipes.net: bioinformatic tools for ecological and evolutionary genomics. Evol. Bioinf. Online **2010**:143–149.

Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D. (2009). Mixtools: an R package for analyzing finite mixture models. J. Stat. Softw. **32**:1–29.

Bornowski, N., Hamilton, J.P., Liao, P., Wood, J.C., Dudareva, N., and Buell, C.R. (2020). Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. DNA Res. **27**:dsaa016.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972–1973.

Chamala, S., García, N., Godden, G.T., Krishnakumar, V., Jordon-Thaden, I.E., De Smet, R., Barbazuk, W.B., Soltis, D.E., and Soltis, P.S. (2015). MarkerMiner 1.0: a new application for phylogenetic marker development using angiosperm transcriptomes. Appl. Plant Sci. **3**:1400115.

Chang, C.-S., Kim, H., and Chang, K.S. (2014). Provisional Checklist of Vascular Plants for the Korea Peninsula Flora (KPF) (Pajo: Designpost).

Chaudhuri, P., and Marron, J.S. (1999). SiZer for exploration of structures in curves. J. Am. Stat. Assoc. **94**:807–823.

Chernomor, O., Von Haeseler, A., and Minh, B.Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. **65**:997–1008.

Croteau, R.B., Davis, E.M., Ringer, K.L., and Wildung, M.R. (2005). (−)-Menthol biosynthesis and molecular genetics. Naturwissenschaften **92**:562–577.

Currin, A., Dunstan, M.S., Johannissen, L.O., Hollywood, K.A., Vinaixa, M., Jervis, A.J., Swainston, N., Rattray, N.J.W., Gardiner, J.M., Kell, D.B., et al. (2018). Engineering the "missing link" in biosynthetic (-)-Menthol production: bacterial isopulegone isomerase. ACS Catal. **8**:2012–2020.

Dang, J., Lin, G., Liu, L., Zhou, P., Shao, Y., Dai, S., Sang, M., Jiang, Z., Liu, C., and Wu, Q. (2022). Comparison of pulegone and estragole chemotypes provides new insight into volatile oil biosynthesis of *Agastache rugosa*. Front. Plant Sci. **13**:850130.

Dethe, D.H., Das, S., Kumar, V.B., and Mir, N.A. (2018). Enantiospecific total syntheses of (+)-Hapalindole H and (−)-12-epi-Hapalindole U. Chemistry **24**:8980–8984.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**:15–21.

Dong, A.X., Xin, H.B., Li, Z.J., Liu, H., Sun, Y.Q., Nie, S., Zhao, Z.N., Cui, R.F., Zhang, R.G., Yun, Q.Z., et al. (2018). High-quality assembly of the reference genome for scarlet sage, *Salvia splendens*, an economically important ornamental plant. GigaScience **7**:giy068.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238.

Engler, H.G.A., and Prantl, K.A.E. (1896). *Schizonepeta* (Benth.) Briq. In In Naturlichen Pflanzenfamilien, p. 235.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics **32**:3047–3048.

Gao, Q., Wang, L., Zhang, M., Wei, Y., and Lin, W. (2020). Recent advances on feasible strategies for monoterpenoid production in *Saccharomyces cerevisiae*. Front. Bioeng. Biotechnol. **8**:609800.

Gel, B., and Serra, E. (2017). KaryoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics **33**:3088–3090.

Godden, G.T., Kinser, T.J., Soltis, P.S., and Soltis, D.E. (2019). Phylotranscriptomic analyses reveal asymmetrical gene duplication dynamics and signatures of ancient polyploidy in mints. Genome Biol. Evol. **11**:3393–3408.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the

Please cite this article in press as: Liu et al., A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*, Molecular Plant (2023), https://doi.org/10.1016/j.molp.2023.01.004

Trinity platform for reference generation and analysis. Nat. Protoc. **8**:1494–1512.

Hamilton, J.P., Godden, G.T., Lanier, E., Bhat, W.W., Kinser, T.J., Vaillancourt, B., Wang, H., Wood, J.C., Jiang, J., Soltis, P.S., et al. (2020). Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing Lamiaceae species, *Callicarpa americana*. GigaScience **9**:giaa093.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2017). UFBoot2 : improving the ultrafast bootstrap approximation. Mol. Biol. Evol. **35**:518–522.

Huang, K., and Rieseberg, L.H. (2020). Frequency, origins, and evolutionary role of chromosomal inversions in plants. Front. Plant Sci. **11**:296.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods **14**:587–589.

Kim, K.-W., De-Kayne, R., Gordon, I.J., Omufwoko, K.S., Martins, D.J., Ffrench-Constant, R., and Martin, S.H. (2022). Stepwise evolution of a butterfly supergene via duplication and inversion. Philos. Trans. R. Soc. Lond. B Biol. Sci. **377**:20210207.

Leferink, N.G.H., Dunstan, M.S., Hollywood, K.A., Swainston, N., Currin, A., Jervis, A.J., Takano, E., and Scrutton, N.S. (2019). An automated pipeline for the screening of diverse monoterpene synthase libraries. Sci. Rep. **9**:11936.

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. **49**:W293–W296.

Li, C.Y., Yang, L., Liu, Y., Xu, Z.G., Gao, J., Huang, Y.B., Xu, J.J., Fan, H., Kong, Y., Wei, Y.K., et al. (2022). The sage genome provides insight into the evolutionary dynamics of diterpene biosynthesis gene cluster in plants. Cell Rep. **40**:111236.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094–3100.

Lichman, B.R., Godden, G.T., and Buell, C.R. (2020b). Gene and genome duplications in the evolution of chemodiversity: perspectives from studies of Lamiaceae. Curr. Opin. Plant Biol. **55**:74–83.

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., et al. (2020a). The evolutionary origins of the cat attractant nepetalactone in catnip. Sci. Adv. **6**:eaba0721.

Liu, C., Gao, Q., Shang, Z., Liu, J., Zhou, S., Dang, J., Liu, L., Lange, I., Srividya, N., Lange, B.M., et al. (2021b). Functional characterization and structural insights into stereoselectivity of pulegone reductase in menthol biosynthesis. Front. Plant Sci. **12**:780970.

Liu, C., Srividya, N., Parrish, A.N., Yue, W., Shan, M., Wu, Q., and Lange, B.M. (2018). Morphology of glandular trichomes of Japanese catnip (*Schizonepeta tenuifolia* Briquet) and developmental dynamics of their secretory activity. Phytochemistry **150**:23–30.

Liu, L., Yin, M., Lin, G., Wang, Q., Zhou, P., Dai, S., Sang, M., Liu, C., Wu, Q., and Wu, Q. (2021a). Integrating RNA-seq with functional expression to analyze the regulation and characterization of genes involved in monoterpenoid biosynthesis in *Nepeta tenuifolia* Briq. Plant Physiol. Biochem. **167**:31–41.

Liu, Z., Cheema, J., Vigouroux, M., Hill, L., Reed, J., Paajanen, P., Yant, L., and Osbourn, A. (2020). Formation and diversification of a paradigm biosynthetic gene cluster in plants. Nat. Commun. **11**:5354.

Liu, Z., Suarez Duran, H.G., Harnvanichvech, Y., Stephenson, M.J., Schranz, M.E., Nelson, D., Medema, M.H., and Osbourn, A. (2019). Drivers of metabolic diversification: how dynamic genomic neighbourhoods generate new biosynthetic pathways in the Brassicaceae. New Phytol. **227**:1109–1123.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**:550.

Lygidakis, A., Karuppiah, V., Hoeven, R., Ní Cheallaigh, A., Leys, D., Gardiner, J.M., Toogood, H.S., and Scrutton, N.S. (2016). Pinpointing a mechanistic switch between ketoreduction and "Ene" reduction in short-chain dehydrogenases/reductases. Angew. Chem. Int. Ed. Engl. **55**:9596–9600.

Ma, Y., Cui, G., Chen, T., Ma, X., Wang, R., Jin, B., Yang, J., Kang, L., Tang, J., Lai, C., et al. (2021). Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*. Nat. Commun. **12**:685.

Mao, L., Kawaide, H., Higuchi, T., Chen, M., Miyamoto, K., Hirata, Y., Kimura, H., Miyazaki, S., Teruya, M., Fujiwara, K., et al. (2020). Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. Proc. Natl. Acad. Sci. USA. **117**:12472–12480.

Maruyama, T., Ito, M., Kiuchi, F., and Honda, G. (2001). Molecular cloning, functional expression and characterization of *d*-limonene synthase from *Schizonepeta tenuifolia*. Biol. Pharm. Bull. **24**:373–377.

Maruyama, T., Saeki, D., Ito, M., and Honda, G. (2002). Molecular cloning, functional expression and characterization of *d*-Limonene synthase from *Agastache rugosa*. Biol. Pharm. Bull. **25**:661–665.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., and Teeling, E. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**:1530–1534.

Mint Evolutionary Genomics Consortium Electronic address buell@msuedu; Mint Evolutionary Genomics Consortium. (2018). Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. Mol. Plant **11**:1084–1096.

Miyamoto, K., Fujita, M., Shenton, M.R., Akashi, S., Sugawara, C., Sakai, A., Horie, K., Hasegawa, M., Kawaide, H., Mitsuhashi, W., et al. (2016). Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. Plant J. **87**:293–304.

Molinari, N.A., Petrov, D.A., Price, H.J., Smith, J.D., Gold, J.R., Vassiliadis, C., Dudley, J.W., Biradar, D.P., Devos, K.M., Bennetzen, J.L., et al. (2008). Synteny and collinearity in plant genomes. Science **320**:486–488.

Moon, H.K., Smets, E., and Huysmans, S. (2010). Phylogeny of tribe Mentheae (Lamiaceae): the story of molecules and micromorphological characters. Taxon **59**:1065–1076.

Moore, B.D., Andrew, R.L., Külheim, C., and Foley, W.J. (2014). Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytol. **201**:733–750.

Nützmann, H.W., and Osbourn, A. (2014). Gene clustering in plant specialized metabolism. Curr. Opin. Biotechnol. **26**:91–99.

Nützmann, H.W., Huang, A., and Osbourn, A. (2016). Plant metabolic clusters – from genetics to genomics. New Phytol. **211**:771–789.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. **20**:275.

Pichersky, E., and Gershenzon, J. (2002). The formation and function of plant volatiles: perfumes for pollinator attraction and defense. Curr. Opin. Plant Biol. **5**:237–243.

Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. Annu. Rev. Plant Biol. **62**:549–566.

Polturak, G., Dippe, M., Stephenson, M.J., Chandra Misra, R., Owen, C., Ramirez-Gonzalez, R.H., Haidoulis, J.F., Schoonbeek, H.J., Chartrain, L., Borrill, P., et al. (2022). Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. Proc. Natl. Acad. Sci. USA. **119**. e2123299119.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**:841–842.

Rajaonarivony, J.I., Gershenzon, J., and Croteau, R. (1992). Characterization and mechanism of (4*S*)-Limonene synthase, A monoterpene cyclase from the glandular trichomes of peppermint (*Mentha x piperita*). Arch. Biochem. Biophys. **296**:49–57.

Ringer, K.L., McConkey, M.E., Davis, E.M., Rushing, G.W., and Croteau, R. (2003). Monoterpene double-bond reductases of the (-)-menthol biosynthetic pathway: Isolation and characterization of cDNAs encoding (-)-isopiperitenone reductase and (+)-pulegone reductase of peppermint. Arch. Biochem. Biophys. **418**:80–92.

Schaller, F., and Weiler, E.W. (1997). Molecular cloning and characterization of 12-oxophytodienoate reductase, an enzyme of the octadecanoid signaling pathway from *Arabidopsis thaliana*. Structural and functional relationship to yeast old yellow enzyme. J. Biol. Chem. **272**:28066–28072.

Serpooshan, F., Jamzad, Z., Nejadsattari, T., and Mehregan, I. (2018). Molecular phylogenetics of Hymenocrater and allies (Lamiaceae): new insights from nrITS, plastid trnL intron and trnL-F intergenic spacer DNA sequences. Nord. J. Bot. **36**:e01600.

Shan, M.Q., Qian, Y., Yu, S., Guo, S.C., Zhang, L., Ding, A.W., and Wu, Q.N. (2016). Anti-inflammatory effect of volatile oil from *Schizonepeta tenuifolia* on carrageenin-induced pleurisy in rats and its application to study of appropriate harvesting time coupled with multi-attribute comprehensive index method. J. Ethnopharmacol. **194**:580–586.

Shan, M., Jiang, Y., Fu, Y., Zhou, Y., Lu, Z., Yu, S., Yan, H., Liu, C., Chen, P., Bao, B., et al. (2021). A review of the botany, traditional uses, phytochemistry and pharmacology of *Nepeta tenuifolia* Briq. Phytochem. Rev. **20**:991–1012.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models. Genome Res. **13**:2498–2504.

Smit, S.J., and Lichman, B.R. (2022). Plant biosynthetic gene clusters in the context of metabolic evolution. Nat. Prod. Rep. **39**:1465–1482.

Srividya, N., Lange, I., and Lange, B.M. (2020). Determinants of enantiospecificity in limonene synthases. Biochemistry **59**:1661–1664.

Srividya, N., Lange, I., Richter, J.K., Wüst, M., and Lange, B.M. (2022). Selectivity of enzymes involved in the formation of opposite enantiomeric series of *p*-menthane monoterpenoids in peppermint and Japanese catnip. Plant Sci. **314**:111119.

Thompson, M.J., and Jiggins, C.D. (2014). Supergenes and their role in evolution. Heredity **113**:1–8.

Toogood, H., Gardiner, J., and Scrutton, N. (2010). Biocatalytic reductions and chemical versatility of the old yellow enzyme family of flavoprotein oxidoreductases. ChemCatChem **2**:892–914.

Tzfadia, O., Diels, T., De Meyer, S., Vandepoele, K., Aharoni, A., and Van De Peer, Y. (2016). CoExpNetViz: comparative co-expression networks construction and visualization tool. Front. Plant Sci. **6**:1194.

Unsicker, S.B., Kunert, G., and Gershenzon, J. (2009). Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. Curr. Opin. Plant Biol. **12**:479–485.

Vining, K.J., Hummer, K.E., Bassil, N.V., Lange, B.M., Khoury, C.K., and Carver, D. (2020). Crop wild relatives as germplasm resource for cultivar improvement in mint (Mentha L.). Front. Plant Sci. **11**:1217.

Vining, K.J., Pandelova, I., Lange, I., Parrish, A.N., Lefors, A., Kronmiller, B., Liachko, I., Kronenberg, Z., Srividya, N., Lange, B.M., et al. (2022). Chromosome-level genome assembly of *Mentha longifolia* L. Reveals gene organization underlying disease resistance and essential oil traits. G3 **12**:jkac112.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40**:e49.

Williams, R.E., and Bruce, N.C. (2002). "New uses for an old enzyme" - the Old Yellow Enzyme family of flavoenzymes. Microbiology **148**:1607–1614.

Xiong, X., Gou, J., Liao, Q., Li, Y., Zhou, Q., Bi, G., Li, C., Du, R., Wang, X., Sun, T., et al. (2021). The *Taxus* genome provides insights into paclitaxel biosynthesis. Nat. Plants **7**:1026–1036.

Yamamoto, K., Grzech, D., Koudounas, K., Stander, E.A., Caputi, L., Mimura, T., Courdavault, V., and O'Connor, S.E. (2021). Improved virus-induced gene silencing allows discovery of a serpentine synthase gene in *Catharanthus roseus*. Plant Physiol. **187**:846–857.

Zebec, Z., Wilkes, J., Jervis, A.J., Scrutton, N.S., Takano, E., and Breitling, R. (2016). Towards synthesis of monoterpenes and derivatives using synthetic biology. Curr. Opin. Chem. Biol. **34**:37–43.

Zhao, D., Hamilton, J.P., Bhat, W.W., Johnson, S.R., Godden, G.T., Kinser, T.J., Boachon, B., Dudareva, N., Soltis, D.E., Soltis, P.S., et al. (2019). A chromosomal-scale genome assembly of *Tectona grandis* reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways. GigaScience **8**:giz005.

Zhao, T., and Wang, Z. (2022). GraphBio: a shiny web app to easily perform popular visualization analysis for omics data. Front. Genet. **13**:957317.

Zielińska-Błajet, M., and Feder-Kubis, J. (2020). Monoterpenes and their derivatives—recent development in biological and medical applications. Int. J. Mol. Sci. **21**:7078.