
Article:
Boachon, Benoît, Buell, C. Robin, Crisovan, Emily et al. (17 more authors) (2018) Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. Molecular plant. pp. 1084-1096. ISSN 1674-2052

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

CellPress
PARTNER JOURNAL

Molecular Plant
Research Article

Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae

Mint Evolutionary Genomics Consortium*

*Correspondence: buell@msu.edu

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

ABSTRACT

The evolution of chemical complexity has been a major driver of plant diversification, with novel compounds serving as key innovations. The species-rich mint family (Lamiaceae) produces an enormous variety of compounds that act as attractants and defense molecules in nature and are used widely by humans as flavor additives, fragrances, and anti-herbivory agents. To elucidate the mechanisms by which such diversity evolved, we combined leaf transcriptome data from 48 Lamiaceae species and four outgroups with a robust phylogeny and chemical analyses of three terpenoid classes (monoterpenes, sesquiterpenes, and iridoids) that share and compete for precursors. Our integrated chemical–genomic–phylogenetic approach revealed that: (1) gene family expansion rather than increased enzyme promiscuity of terpene synthases is correlated with mono- and sesquiterpene diversity; (2) differential expression of core genes within the iridoid biosynthetic pathway is associated with iridoid presence/absence; (3) generally, production of iridoids and canonical monoterpenes appears to be inversely correlated; and (4) iridoid biosynthesis is significantly associated with expression of geraniol synthase, which diverts metabolic flux away from canonical monoterpenes, suggesting that competition for common precursors can be a central control point in specialized metabolism. These results suggest that multiple mechanisms contributed to the evolution of chemodiversity in this economically important family.

Key words: chemodiversity, evolution, terpene, iridoid

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.

INTRODUCTION

Chemical diversity, a consequence of specialized metabolic pathways that branch from primary metabolism, is a key driver in angiosperm diversification, as exemplified by the evolution of glucosinolates (mustard oils) as anti-herbivory compounds in Brassicales (Ettlinger and Kjaer, 1968; Edger et al., 2015), betalains as colorful pigments in Caryophyllales (Brockington et al., 2011), and sesquiterpene compounds that function in biotic stress in Asteraceae (Heywood et al., 1977; Chadwick et al., 2013). Specialized metabolites typically originate from a small group of precursors, which are subjected to multiple sequential modifications to yield an array of distinct end-products. Terpenes are an outstanding example of this process: consider, for example, monoterpenes, sesquiterpenes, and iridoids (non-canonical monoterpenes). All are derived from the same precursors, but through the formation of an initial divergent scaffold and subsequent modifications, a wide variety of chemical structures and properties are produced (Figure 1A–1D).

Lamiaceae (mint family) is the sixth largest family of angiosperms and includes many culturally and economically important species used as sources of flavor additives (e.g., basil [*Ocimum* L.], rosemary [*Rosmarinus* L.], peppermint [*Mentha* × *piperita* L.]), fragrance (e.g., English lavender [*Lavandula angustifolia* Mill.]), cosmetics (e.g., mint [*Mentha* L.]), ornamentals (e.g., coleus [*Plectranthus* L'Hér.]), and anti-herbivory compounds (e.g., sage [*Salvia officinalis* L.]). This family exhibits an exceptionally high degree of secondary metabolite diversity, providing antimicrobial, anti-herbivory, and pollinator attraction functions in nature, and resources for human health, food, and agricultural purposes. Monoterpenes, iridoids, and sesquiterpenes are responsible for many of these functions.

Monoterpenes and iridoids are both derived from geranyl pyrophosphate (GPP), produced by the methylerythritol phosphate

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

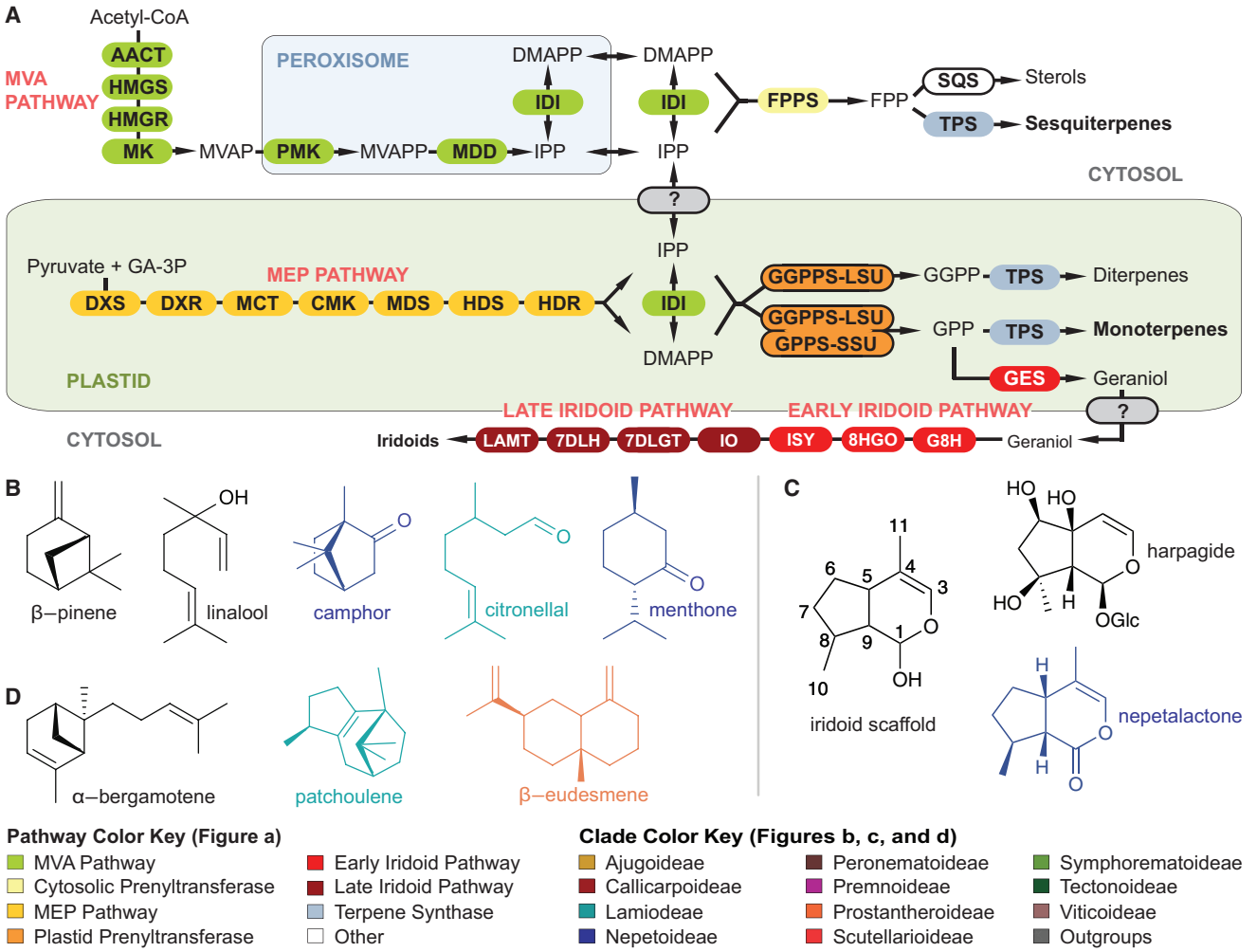

Figure 1. Biosynthetic Pathways Leading to Production of Monoterpenes, Sesquiterpenes, Iridoids, and Key Specialized Metabolites of Interest within Lamiaceae.

(A) Schematic of mevalonate (MVA), 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP), iridoid, and terpene metabolic pathways including their subcellular localization. MVA pathway leading to sterols and sesquiterpenes: acetyl-CoA acetyltransferase activity (AACT); 3-hydroxy-3-methylglutaryl coenzyme A synthase (HMGS); hydroxymethylglutaryl-CoA reductase (HMGR); ATP:mevalonate phosphotransferase (MK); mevalonate 5-phosphate (MVAP); phosphomevalonate kinase (PMK); mevalonate 5-diphosphate (MVAPP); mevalonate diphosphate decarboxylase (MDD); isopentenyl diphosphate (IPP); isopentenyl diphosphate delta-isomerase (IDI); dimethylallyl diphosphate (DMAPP); geranyltranstransferase (FPPS); farnesyl diphosphate (FPP); squalene synthase (SQS); TPS (terpene synthase). MEP pathway leading to geraniol, monoterpenes, and diterpenes: 1-deoxy-D-xylulose-5-phosphate synthase (DXS); 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR); 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT); 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase (CMK); 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS); 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS); 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (HDR); geranyl pyrophosphate synthase small subunit (GPPS-SSU); geranylgeranyl pyrophosphate synthase large subunit (GGPPS-LSU); geranyl diphosphate (GPP); geranylgeranyl diphosphate (GGPP); geraniol synthase (GES). Iridoid pathway: geraniol 8-hydroxylase (G8H); 8-hydroxygeraniol oxidoreductase (8HGO); iridoid synthase (ISY); iridoid oxidase (IO); 7-deoxyloganetic acid glucosyltransferase (7DLGT); 7-deoxyloganic acid hydroxylase (7DLH); loganic acid methyltransferase (LAMT). Terpene synthases: TPS a, b, c, g, e/f.

(B–D) Representative metabolites assayed in this study: **(B)** monoterpenes, **(C)** iridoids, and **(D)** sesquiterpenes.

Widely distributed terpenoids and iridoids are drawn in black; other colors represent clade-specific terpenes and iridoids and match the color coding shown in Figure 2.

(MEP) pathway (Figure 1A). In the plastid, monoterpene synthases transform GPP to many different monoterpene scaffolds in a single enzymatic reaction (Figure 1B). In iridoid biosynthesis, the dedicated monoterpene synthase, geraniol synthase (GES), converts GPP to geraniol, which is then exported to the cytosol where it is hydroxylated, oxidized, and cyclized, forming the iridoid scaffold, which can undergo a wide array of derivatizations (Figure 1C). Sesquiterpenes are synthesized in the cytosol mainly via the mevalonate (MVA) pathway. The sesquiterpene pathway competes for isopentenyl diphosphate (IPP) derived from the MEP pathway, which is transported from the plastid to the cytosol and converted to farnesyl diphosphate (FPP), which is then utilized by sesquiterpene synthases to produce a diverse array of sesquiterpenes (Figure 1D).

Lamiaceae is rich in terpenoids (Lange et al., 2000), making this family an excellent target for an evolutionarily based study of specialized metabolism. The evolution of plant metabolite diversity can be achieved via gene or genome duplication followed by sub- or neo-functionalization, an increase in enzyme promiscuity, and/or utilization of scaffold-decorating enzymes (Kliebenstein et al., 2001; O'Maille et al., 2008; Matsuno et al., 2009; Ono et al., 2010; Schilmiller et al., 2015; Moghe et al., 2017; Xu et al., 2017; for review see Weng et al., 2012). However, the selective pressures responsible for the generation of extensive chemical diversity within a set of related taxa, and the biochemical, molecular, and genetic mechanisms by which such diversity evolves, remain largely unknown. We performed a transcriptomic and metabolomic survey of the leaf transcriptome in 48 phylogenetically diverse species of Lamiaceae and constructed a molecular phylogeny that was coupled with transcript sequences, expression abundances, gene orthology, and terpenoid metabolite profiles. Our analysis suggests that multiple evolutionary mechanisms collectively led to the extensive chemodiversity in Lamiaceae.

RESULTS

Phylogeny of Lamiaceae as Revealed through Transcribed Sequences

To capture and adequately reconstruct chemical evolution across the mint family, we selected 48 species from 11 of 12 recognized subclades representing the phylogenetic diversity of Lamiaceae; four species from Lamiales families were selected as outgroups (Supplemental Table 1). As the leaf is the primary site of photosynthesis and carbon fixation, as well as terpenoid production (Turner and Croteau, 2004), we generated transcriptomes for all 52 species from young leaves, resulting in an average of 64 381 transcripts representing 51.3 Mb per species (Supplemental Table 1). The number of transcripts and total base pairs per assembled transcriptome were highly variable across the 52 species (Supplemental Table 1), suggestive of polyploidy and/or extensive heterozygosity, both of which have been reported in the family (Morton, 1973; Xu et al., 2016; Vining et al., 2017) and would result in assembly of distinct transcripts for alleles, paralogs, homologs, and/or homeologs.

Previous plastid-based molecular studies of Lamiaceae identified major subclades (recognized as subfamilies), although interrelationships among these subclades remain unclear (Bendiksby et al., 2011; Li et al., 2012, 2016; Chen et al., 2014) (Supplemental Figure 1). We mined sequences representing nuclear genes that are generally single-copy across all angiosperms (De Smet et al., 2013) from our filtered transcriptomes and used them to reconstruct species relationships. We identified 520 putatively single-copy nuclear genes for Lamiaceae. Phylogenetic analyses based on these genes (>652 kb; Supplemental Tables 2 and 3) supported the monophyly of Lamiaceae and confirmed 10 of 12 previously identified subclades (Li et al., 2016); Premnoideae were not recovered, and Cymarioideae were not sampled. Relationships were well resolved at nearly all nodes in both the maximum-likelihood (ML) and species trees (Figure 2 and Supplemental Figure 2). A clade of *Callicarpa* (Callicarpoideae) + Prostantheroideae is sister to the rest of Lamiaceae, which form two clades: Nepetoideae (with approximately half of all Lamiaceae species) and a clade of *Tectona* (Tectonoideae), Symphorematoideae, Scutellarioideae, Ajugoideae, Lamioideae, *Petraeovitex* Oliv. (Peronematoideae), and a grade of species previously classified in Viticoideae sensu Harley et al. (2004) that is not monophyletic in our analysis. The monophyly of Premnoideae and its phylogenetic position were not supported, and the species tree indicated gene tree discordances for relationships among *Petraeovitex*, Lamioideae, and Ajugoideae (Figure 2; Supplemental Figures 1 and 2). Significantly, both topologies differed considerably from those of Li et al. (2016), with the exception of strong support for Viticoideae + Symphorematoideae and for Prostantheroideae + Callicarpoideae as sister to all remaining Lamiaceae.

Chemical Diversity of Lamiaceae

Since Lamiaceae leaves produce essential oils rich in terpenoids (as reviewed in Lange, 2015), we performed metabolic profiling of young leaf tissue to capture terpenoid diversity across the Lamiaceae phylogeny. Using gas chromatography–mass spectrometry (GC–MS), we identified a total of 44 monoterpenes and 39 sesquiterpenes (Supplemental Tables 4 and 5; Figure 3A and 3B) and obtained metabolic profiles that were generally in good agreement with literature reports (Gershenzon et al., 2000; Bozin et al., 2006; Crocoll et al., 2010). Of the 83 volatile compounds detected by GC–MS, 14 were unique to a single species. We reconstructed the presence/absence of these metabolites across the ML phylogeny; select ancestral state reconstructions are presented in Supplemental Figure 3. Evolutionary patterns of the remaining monoterpenes and sesquiterpenes are complex, with parallel gains and losses of individual compounds. For example, the distribution of linalool across Lamiaceae shows little phylogenetic signal, with frequent gains and losses throughout the family (Supplemental Figure 3D). In contrast, many other compounds appear to have originated or been lost just a few times in parallel, often in distantly related species (e.g., across the deep major split in the family between Nepetoideae and its sister clade; Supplemental Figure 3). Examples of this latter pattern include the loss of the sesquiterpene β-caryophyllene in parallel in *Homskioldia*, *Clerodendrum*, *Ajuga*, *Lycopus* + *Prunella*, and *Nepeta* (Supplemental Figure 3F) and the parallel gains of caryophyllene oxide, which is found in *Tectona*, *Petraeovitex*, *Lamium*, *Lavandula*, *Plectranthus*, *Thymus*, and *Origanum*, as well as guaiol, which is produced by *Cornutia* and *Agastache*.

We also quantified the phylogenetic signal in binary (presence/absence) and continuous metabolite data (Supplemental Table 6) and tested whether the traits of species were structured as the result of shared ancestry. Sesquiterpenes were distributed evenly throughout the family, with 46 of 48 Lamiaceae species producing these compounds (Figure 3B). However, sesquiterpenes were below detection level in leaves of the four outgroup species; consequently, we detected a very strong phylogenetic signal in binary trait data for these metabolites ($D = -1.419$, $P < 0.001$; Supplemental Table 6). Although sesquiterpenes have been reported in several species of Lamiales outside Lamiaceae (Sena Filho et al., 2010; Hudaib et al., 2013), the evolution of the sesquiterpenes in Lamiaceae remains unresolved with this dataset. Binary monoterpene

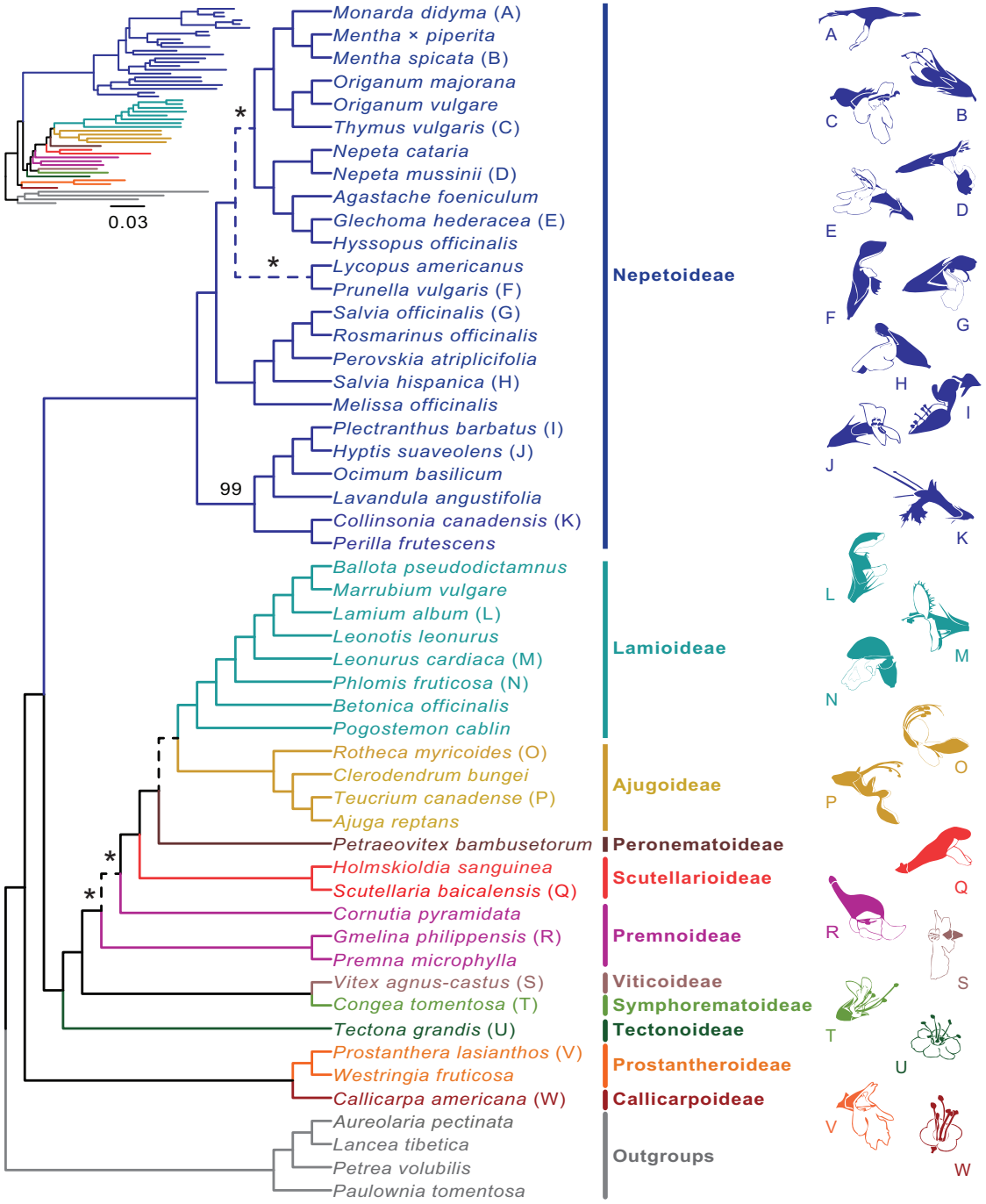

Figure 2. Phylogenetic Relationships among Sampled Species.

Maximum-likelihood (ML) tree inferred from 520 single-copy nuclear genes derived from leaf transcriptomes. The topology is shown as a cladogram, with a phylogram illustrating the tree shape shown as an inset. Subfamilial classification for all species is color coded according to the key provided (Li et al., 2016; Li and Olmstead, 2017). All nodes in the tree are fully supported by the ML bootstrap results (BS = 100%), except where indicated by a number or an asterisk (BS < 50%). Dashed lines show topological incongruences with ASTRAL-II results (Supplemental Figure 2 and Figure 3; see Methods). Representative species denoted in the tree (A through W) are shown as floral silhouettes.

traits exhibited a random distribution across the phylogeny (Supplemental Table 6), but the level of diversity of these metabolites differed between the largest mint clades. The greatest levels of monoterpenes were found in Nepetoideae, with species-specific chemical diversity ranging from 1 to 17 compounds, and the highest diversity was observed in *Perovskia atriplicifolia* Benth. (Russian sage) and *L. angustifolia*. Linalool and α- and β-pinene were frequently detected (up to 27 of 48 species) across Lamiaceae, while some monoterpenes were species specific among our samples (e.g., menthol in peppermint, thymol

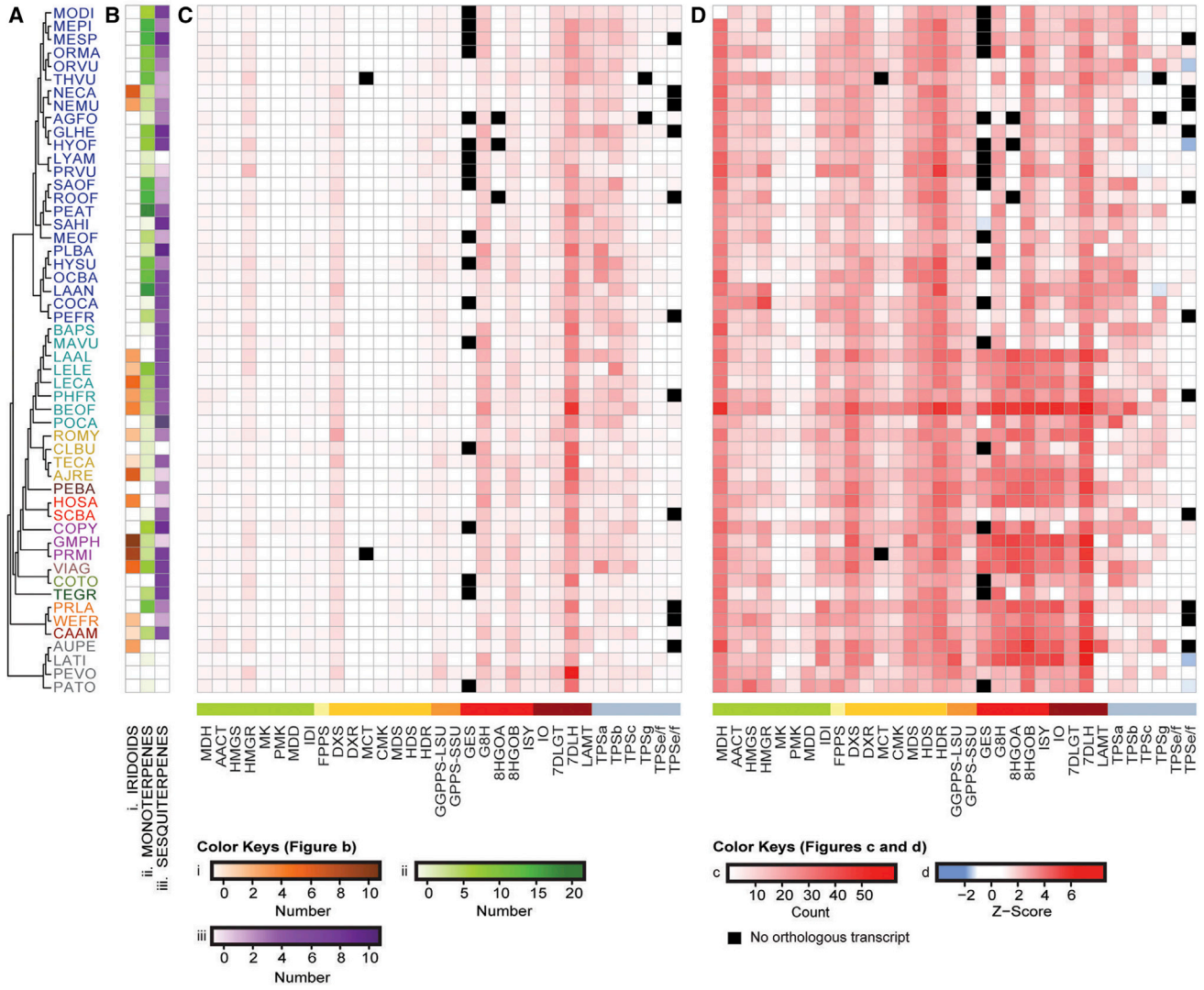

Figure 3. Relationship between Chemical Diversity, Gene Duplication, and Expression Abundances across Lamiaceae.

(A) Phylogeny of the 48 surveyed Lamiaceae and four outgroup species colored by clade as shown in Figure 2.

(B) Diversity of iridoid (i, orange), monoterpene (ii, green), and sesquiterpene (iii, purple) compounds in the surveyed species.

(C) Orthogroup occupancy of genes involved in iridoid and terpene biosynthesis.

(D) Gene expression levels shown as *Z*-scores of genes within orthologous/paralogous groups that encode enzymes within the iridoid and terpene biosynthetic pathways as identified through OrthoFinder (Emms and Kelly, 2015). Species abbreviations are shown in Supplemental Table 1 and gene family name abbreviations are from Figure 1; orthogroup identifiers are from Supplemental Table 8.

in *Thymus vulgaris* L. [thyme]). In all, the presence/absence of 12 metabolites exhibited a moderate to very strong phylogenetic signal (Supplemental Table 6).

Our liquid chromatography (LC)–MS analysis tentatively identified 64 iridoid glycosides, distributed throughout Lamiaceae, and we detected phylogenetic structure in binary data for these metabolites (Figure 3B; Supplemental Tables 6 and 7; D = 0.160, P = 0.005). These metabolic profiles were largely consistent with previous literature reports of mint-derived iridoids (Supplemental Table 7), although we did not observe all iridoids that had been previously observed, likely due to differences in developmental stage or cultivar (Supplemental Table 7). GC–MS analysis also identified two iridoid aglycones that were volatile: nepetalactone, in both species of *Nepeta* (*Nepeta cataria* L. [catnip] and *Nepeta mussinii* Spreng. ex Henckel [dwarf catmint]), and iridodiol, in *Callicarpa americana* L. (American beautyberry) (see Supplemental Information). As noted in previous chemosystematic studies of Lamiaceae (El-Gazzar and Watson, 1970; Kooiman, 1972; Hegnauer, 1989; Taskova et al., 1997; Wink, 2003), *Nepeta* is remarkable as an iridoid producer within the large, non-iridoid-producing clade, Nepetoideae. Thus, metabolomics profiling has consistently shown that iridoids appear to have been lost in Nepetoideae, but then re-emerged in *Nepeta* (see below and Supplemental Figure 3). Furthermore, most iridoids identified in this study were the 8*R*-type glycosides typical of Lamiaceae, while *Nepeta* iridoids were of the *S*-type as previously reported for *Nepeta* (Jensen, 1991). In addition, we found lineage-specific losses of iridoids in Lamiaceae outside of Nepetoideae

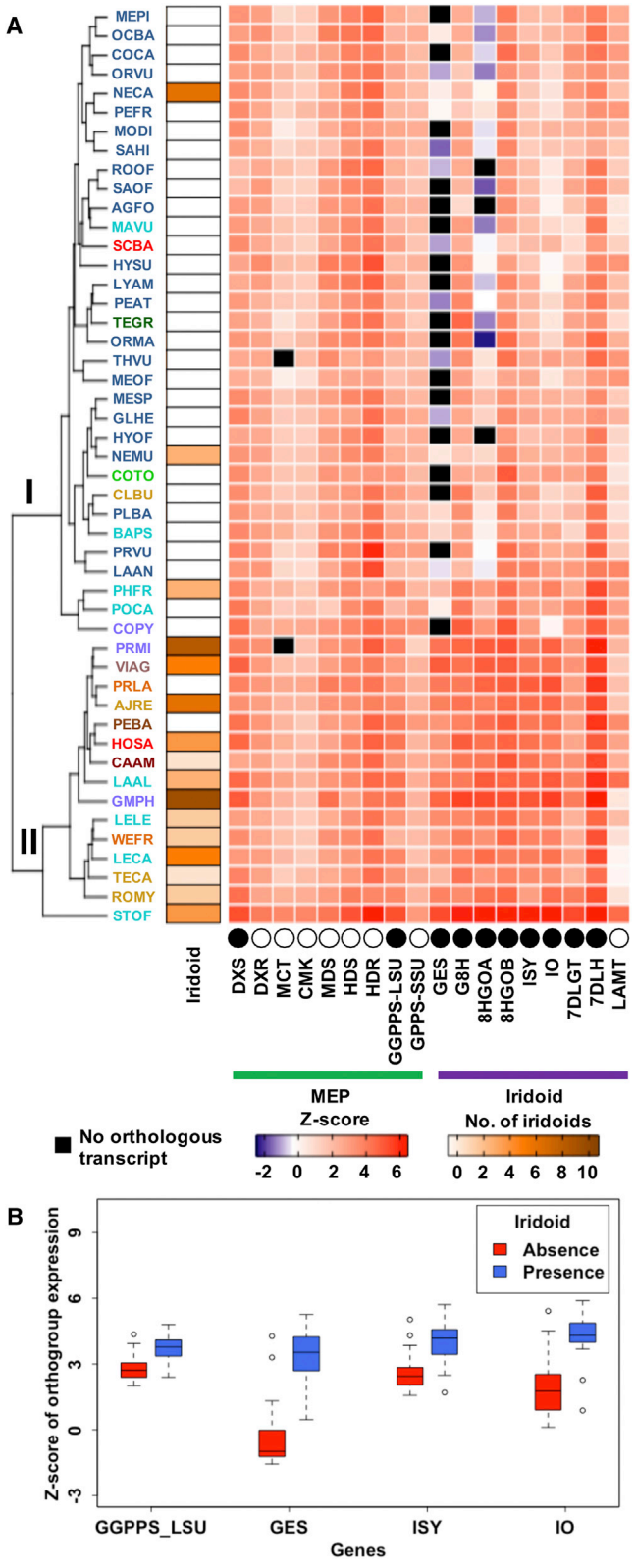

Figure 4. Iridoid Biosynthesis Is Associated with Gene Expression of Key Pathway Genes.

(A) Clustering of 48 surveyed Lamiaceae species based on expression abundance of transcripts within orthogroups of genes encoding 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP) and iridoid biosynthetic enzymes. Species abbreviations are listed in Supplemental Table 1 and are color coded based on subfamily as shown in Figure 2. Closed circles at the bottom of the heatmap represent statistically significant differences in expression levels of the defined orthologous group between species that produce iridoids and species that do not produce iridoids; open circles indicate that expression levels were not significantly different.

(B) Box plots of expression abundances of transcripts within orthologous groups for genes within the MEP and iridoid biosynthetic pathway—GPPS-LSU, GES, ISY, and IO—for iridoid producers and non-iridoid producers; all differences in expression abundance between iridoid producers and non-producers were significant. Abbreviations for genes are listed in Figure 1; orthogroup identifiers are from Supplemental Table 8.

(Supplemental Figure 3). Species within Nepetoideae exhibit the greatest monoterpene diversity (Figure 3B; El-Gazzar and Watson, 1970; Kooiman, 1972) and lack iridoid glucosides, which suggests that canonical monoterpene diversity in Nepetoideae may arise in part through more available carbon flux from the common GPP precursor.

Biochemical Pathway Diversity in Lamiaceae

To understand the molecular basis for the loss and gain of terpenoids and their diversity across Lamiaceae, we examined gene family expansion, contraction, and expression across metabolic pathways. We generated putative orthologous and paralogous groups using the predicted proteomes of all 52 species studied here and the predicted proteomes of five additional land plants: *Amborella trichopoda* Baill., *Arabidopsis thaliana* (L.) Heynh., *Catharanthus roseus* (L.) G. Don, *Physcomitrella patens* (Hedw.) Bruch and Schimp., and *Selaginella moellendorffii* Hieron. In total, 1 838 621 proteins clustered into 86 548 orthologous/paralogous groups. To validate the representation of genes within metabolic pathways, we assessed the canonical genes involved in the tricarboxylic acid (TCA) cycle in the transcriptomes of all 52 species. For all analyzed species, a similar number of genes with limited variation in copy number was identified within 10 orthogroups encoding enzymes of the TCA cycle (Supplemental Tables 8 and 9). Furthermore, expression of genes within the TCA orthogroups did not significantly vary across all species, as shown by the consistent *Z*-scores (Supplemental Tables 8 and 9). This expected expression pattern for a primary metabolic pathway indicates that our assemblies and expression abundances provide a robust and accurate assessment of leaf metabolism.

Iridoid Biosynthetic Pathway

Clustering of species based on the gene expression levels of orthogroups encoding iridoid biosynthetic pathway enzymes (Supplemental Table 8) resulted in two distinct expression clusters: primarily non-iridoid-producing species (Figure 4A: Group I) and primarily iridoid-producing species (Figure 4A: Group II) with the presence/absence of iridoids significantly associated with expression abundances of genes previously known to be involved in iridoid biosynthesis. Group II contains species (e.g., *Petraeovitex bambusetorum* [PEBA]) that do not produce iridoids, and this trait may have been lost relatively recently as these plants still express genes of the iridoid biosynthetic pathway; alternatively, they may contain low levels of iridoids undetectable by our methods. *Phlomis fruticosa* (PHFR) and *Nepeta* species (NECA, NEMU)

produce iridoids, yet cluster with primarily non-iridoid-producing species (Figure 4A: Group I), highlighting the unique nature of the iridoid pathway in these genera. The clustering of these three iridoid-producing species with non-iridoid producers may be due to differences in gene expression patterns, the use of genes from different orthogroups to produce iridoids, and/or independent evolution of the iridoid pathway.

Orthogroups encoding enzymes involved in iridoid biosynthesis were analyzed individually to test for associations in both orthogroup occupancy (using phylogenetic Poisson regression) and expression (using phylogenetic logistic regression) among iridoid-producing and non-iridoid-producing Lamiaceae species (Supplemental Table 9; Figure 3C and 3D). Ten orthogroups showed significant associations ($P < 0.05$) in orthogroup expression (Figure 4A). We observed the most significant associations ($P \leq 0.001$) in expression of orthogroups that encode GES, iridoid synthase (ISY), 8-hydroxygeraniol oxidoreductase (8HGOA), and iridoid oxidase (IO; Figure 4B–4D and Supplemental Table 9). The 7DLH and loganic acid methyltransferase (LAMT) orthogroups, which encode cytochrome P450 and methyltransferase enzymes, are very large, suggesting that these groups likely contain many genes that encode alternative functions in addition to iridoid biosynthesis. In addition to gene expression differences, we identified one positive relationship between iridoid diversity and occupancy in the orthogroup encoding 8-hydroxygeraniol oxidoreductase (8HGOB; $P \leq 0.01$; Figure 3C and Supplemental Table 9).

Many species analyzed in this study that did not produce iridoids had no detectable GES expression, suggesting that GES is a key gatekeeping step in iridoid biosynthesis. Despite ISY being responsible for the iridoid scaffold formation and catalyzing the first committed step in iridoid biosynthesis, GES is responsible for diverting metabolic flux away from canonical monoterpenes by converting GPP to geraniol. The association of increased expression of genes encoding the upstream enzymes 1-deoxyxylulose 5-phosphate synthase (DXS; $P = 0.0260$) and geranylgeranyl pyrophosphate synthase large subunit (GGPPS-LSU; $P = 0.0011$) (Figure 4A and Supplemental Table 9) with iridoid biosynthesis is also notable. DXS is a rate-limiting enzyme in the MEP pathway, which provides precursors for iridoid formation (Estevez et al., 2001). GGPPS-LSU can form GPP when acting with geranyl pyrophosphate synthase small subunit (GPPS-SSU; Rai et al., 2013), although it is surprising that no correlation with GPPS-SSU was observed as this protein causes the enzyme complex to produce GPP as its major product. In *Catharanthus roseus*, two different enzymes have been proposed to act as 8-hydroxygeraniol oxidoreductase, 8-hydroxygeraniol oxidoreductase A (8HGOA; KF302069; Miettinen et al., 2014) and 8-hydroxygeraniol oxidoreductase B (8HGOB; AY352047; Krithika et al., 2015). In this study, both enzyme expression patterns correlate significantly with the presence of iridoids, although 8HGOA shows a more significant association ($P = 0.0001$ versus $P = 0.0042$) and only 8HGOB shows a significant association with orthogroup occupancy ($P = 0.0083$); the precise physiological and biochemical roles of 8HGOA and 8HGOB remain to be determined experimentally.

Mono- and Sesquiterpene Biosynthetic Pathways

Based on phylogenetic Poisson regression analyses of orthogroup occupancy and mono- and sesquiterpene diversity (total number of compounds in each species) for 32 orthogroups involved in mono- and sesquiterpene biosynthesis, a positive relationship between occupancy of the terpene synthase b (TPSb) orthogroup and monoterpene diversity ($P = 0.0000498$) (Figure 5A and Supplemental Table 9) was detected. As most angiosperm monoterpene synthases belong to the TPSb subfamily (Chen et al., 2011), this association suggests that monoterpene diversity is the result of action of different monoterpene synthases rather than product promiscuity of a limited number of TPS enzymes. Occupancy of orthogroups encoding GPPS-SSU ($P = 0.01152$), LAMT ($P = 0.02226$), and TPSc ($P = 0.036571$) was also positively associated with monoterpene diversity, albeit with larger P values than observed with TPSb occupancy (Figure 5C and Supplemental Table 9). While increased GPPS-SSU can increase flux to monoterpenes, the contribution of LAMT and TPSc to monoterpene diversity may be attributable to the large size of these orthologous groups (424 and 328, respectively) and a higher degree of diversity and associated enzymatic function.

While we observed a positive relationship between iridoid production and GES expression (Figure 4C), no significant association was found between monoterpenes and GES orthogroup expression (Supplemental Table 9). The absence of such an association is not surprising, since monoterpenes are widely spread across non-iridoid and iridoid-producing species. In contrast, phylogenetic logistic regression analyses revealed no significant relationships ($P > 0.05$; Figure 5B and Supplemental Table 9) between monoterpene presence/absence and expression of any orthogroup encoding upstream pathway enzymes or typical monoterpene synthases, suggesting that the primary driver of monoterpene diversity is gene family expansion, not altered expression. Lineage-specific expansion of terpene synthases occurred in *Eucalyptus grandis* (Myburg et al., 2014), a major source of eucalyptus oil composed primarily of the monoterpene 1,8-cineole, resulting in the largest number of terpene synthases (113) reported in any plant genome.

We also found a positive relationship between sesquiterpene diversity and orthogroup occupancy for TPSa and TPSe/f, the subfamilies that include sesquiterpene synthases (Figure 5D and Supplemental Table 9; $P = 0.08328$ and $P = 0.0517$, respectively) (Chen et al., 2011). No other terpene synthase orthogroup were significantly associated with either gene expression or gene family size as measured through orthogroup expression and occupancy, respectively (Supplemental Table 9). However, there was a significant negative relationship between orthogroup occupancy and sesquiterpene diversity for the orthogroup encoding GGPPS-LSU ($P = 0.0151$) along with the MEP biosynthetic pathway genes (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [MDS], 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [HDS], $P = 0.0007$), whereas for mevalonate diphosphate decarboxylase (MDD; $P = 0.0329$) there was a positive relationship between orthogroup occupancy and sesquiterpene diversity (Supplemental Table 9). The negative correlation of sesquiterpene biosynthesis with the MEP pathway orthogroups, along with the positive correlation with occupancy of MDD, an enzyme of the MVA pathway, suggests that the

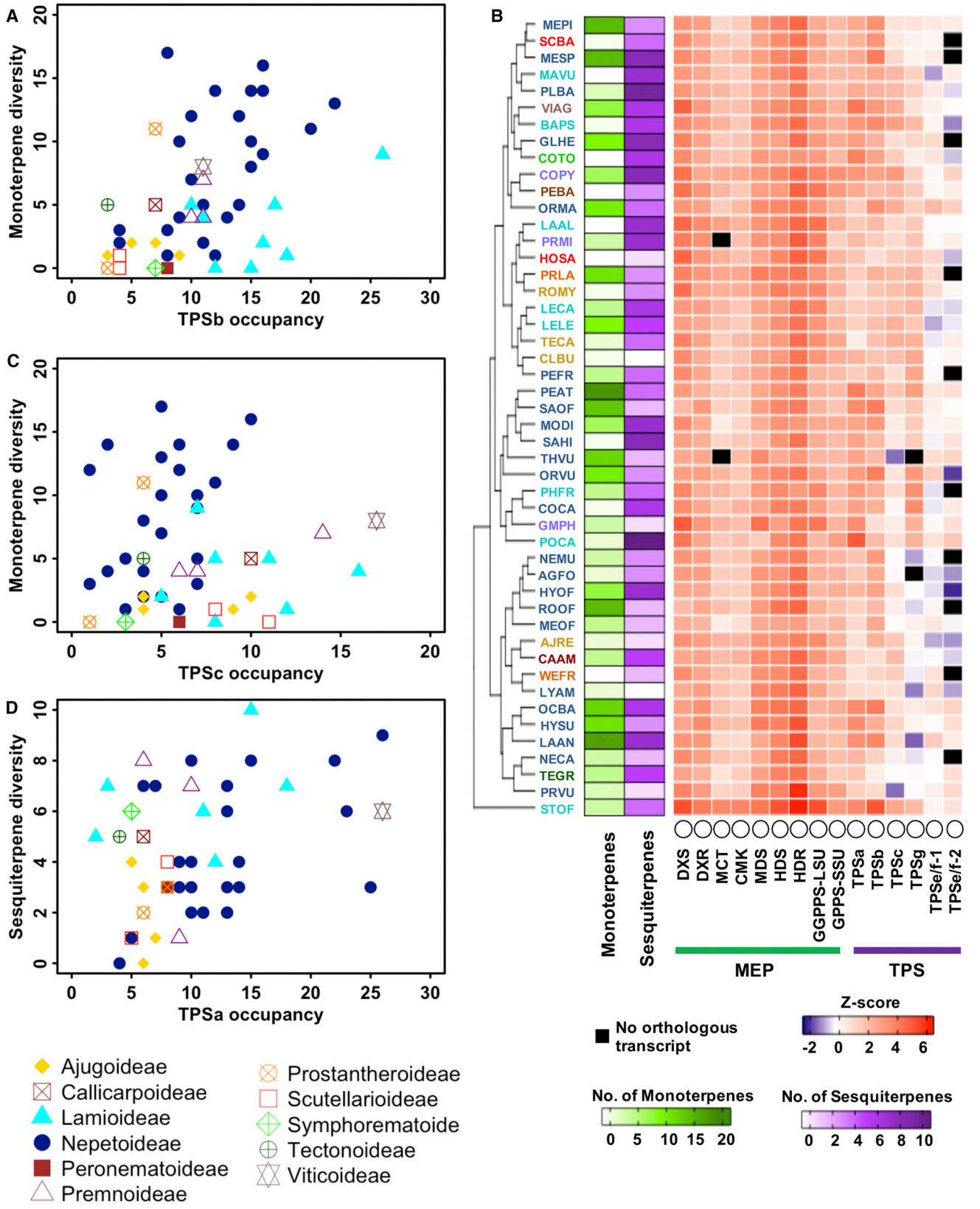

Figure 5. Monoterpene Diversity Is Associated with Gene Family Expansion.

(A) Orthogroup occupancy within TPSb orthologous groups of Lamiaceae species as reflected by a plot of gene number versus diversity of monoterpenes produced. Species are color coded based on the subfamily classification from Figure 2.

(legend continued on next page)

precursors for sesquiterpene formation are derived mainly from the MVA pathway. Sesquiterpene production also appeared to be positively correlated with the early iridoid biosynthetic gene geraniol 8-hydroxylase (G8H; $P = 0.0078$), suggesting that genes within the G8H orthogroup may also contain enzymes that derivatize sesquiterpenes.

In our search for additional genes associated with modifications of terpenoids, we used a set of cloned Lamiaceae cytochrome P450s known to be involved in terpene modifications to identify relevant P450 orthologous groups (Supplemental Table 8). Regression analyses with the presence/absence of iridoids, monoterpenes, and sesquiterpenes only revealed a significant correlation between iridoid presence/absence and orthologous group OG0000115, which encodes CYP76-related genes (Supplemental Table 9). This result most likely reflects the large size of the orthologous group (n = 606) and the presence of G8H within this group. In addition, sesquiterpene diversity, but not iridoid or monoterpene diversity, was significantly associated with orthogroup occupancy of OG0000047, which encodes CYP71A-related genes, and orthogroup occupancy of OG0000115. While these data suggest a role for a subset of these genes in terpenoid metabolism, more detailed analyses exploring the substrate specificity of these orthologs is required.

DISCUSSION

Results from this study, combined with previously reported metabolite profiles of mints, have led to new insights into how the diversity of terpenes correlates with phylogeny, and how the combined chemical, genomic, and phylogenetic approach employed here can elucidate the mechanisms controlling this chemical diversity in leaf tissue. It was previously proposed that the absence of iridoids in the majority of the Nepetoideae species is due to inactivation of the biosynthetic genes involved (Wink, 2003). However, we found that most orthogroups containing iridoid biosynthetic genes were expressed in all Nepetoideae species. The stereochemistry of iridoids in *Nepeta* compared with those from other Lamiaceae (8*S* versus 8*R*) also highlights the unique nature of iridoid cyclization in *Nepeta*. Furthermore, iridoids in *Nepeta* have likely been repurposed for different roles than the iridoid glucosides typical of Lamiaceae. The nepetalactones of *Nepeta* are volatile and resemble sex pheromones of certain insects, and therefore seem to have a role in plant–insect communication (Birkett et al., 2011). Iridoid glycosides are known to serve as defensive compounds (Konno et al., 1999), although ecological roles for iridoid glycosides in Lamiaceae species have not been extensively explored. However, it is likely that the biological function and mode of action of non-volatile iridoids are different from those of the volatile nepetalactones. Regardless, in Nepetoideae, mono- and sesquiterpene volatiles appear to have usurped non-volatile iridoid glycosides as key defense compounds.

In contrast, monoterpene diversity was associated with gene expansion in the TPSb orthogroup, suggesting that the diversity of monoterpenes may be caused by the presence of additional monoterpene synthases that generate a range of monoterpene structures instead of increasing product promiscuity of a limited number of TPS enzymes. Neither orthogroup expression levels nor numbers of terpene synthases are associated with sesquiterpene presence or diversity. However, the positive correlation of sesquiterpene diversity with orthogroup occupancy of MDD, which is involved in the MVA pathway, as well as its negative relationships with orthogroups associated with the parallel MEP pathway, suggests that sesquiterpenes are predominantly derived from the MVA pathway and that limited crosstalk between the plastidic (MEP) and cytosolic (MVA) pools of terpene precursors (Figure 1) occurs in Lamiaceae.

Expression levels of the GES orthogroup, which is located at the first branch point of iridoid synthesis, where the central GPP precursor is diverted to geraniol, appear to play a key role in controlling iridoid production. The lack of detection of volatile geraniol in iridoid-producing plants suggests that geranyl-8-hydroxylase efficiently directs geraniol toward iridoid formation without its release. Although GES is present in a few species that, based on our analyses, do not produce iridoids, the loss of GES or lack of its expression is, in general, associated with loss of iridoid production. We hypothesize that GES is a major point of control for GPP flux from canonical monoterpenes to iridoids.

Our study shows the power of conducting detailed chemical analyses in a phylogenetic context with robust sequence and expression datasets. This combined chemical–genomic–phylogenetic approach has provided novel insights into chemical and genome evolution in the mint family (Figure 6). In leaf tissue, production of iridoids and canonical monoterpenes appeared to be, in general, inversely correlated, with the production of iridoids being strongly associated with enhanced gene expression while, in contrast, expansion of a terpene synthase family appeared to account for increased monoterpene diversity. Overall, these results suggest that multiple mechanisms contribute to the chemical diversity within the Lamiaceae. Chemical diversity was also significantly associated with GES, a key step controlling iridoid biosynthesis that diverts metabolic flux away from canonical monoterpenes, suggesting that competition for common precursors can be a central control point in specialized metabolism. In summary, this multifaceted approach highlights the multiple evolutionary mechanisms involved in generating biochemical diversity and the evolutionary dynamics in a chemically rich clade of plants.

(B) Clustering of 48 surveyed Lamiaceae species based on expression abundance of transcripts within orthologous groups encoding 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP) and terpene biosynthesis. Species abbreviations are listed in Supplemental Table 1 and are color coded based on subfamily as shown in Figure 2. Heat maps to the right of the four-letter species codes reflect the number of monoterpenes (green) and sesquiterpenes (purple) detected from GC–MS. Abbreviations for genes are listed in Figure 1; orthogroup identifiers are from Supplemental Table 8.

(C and D) Orthogroup occupancy within TPSc relative to monoterpene diversity **(C)** and orthogroup occupancy of TPSa relative to sesquiterpene diversity **(D)**. Species are color coded based on the subfamily classification from Figure 2.

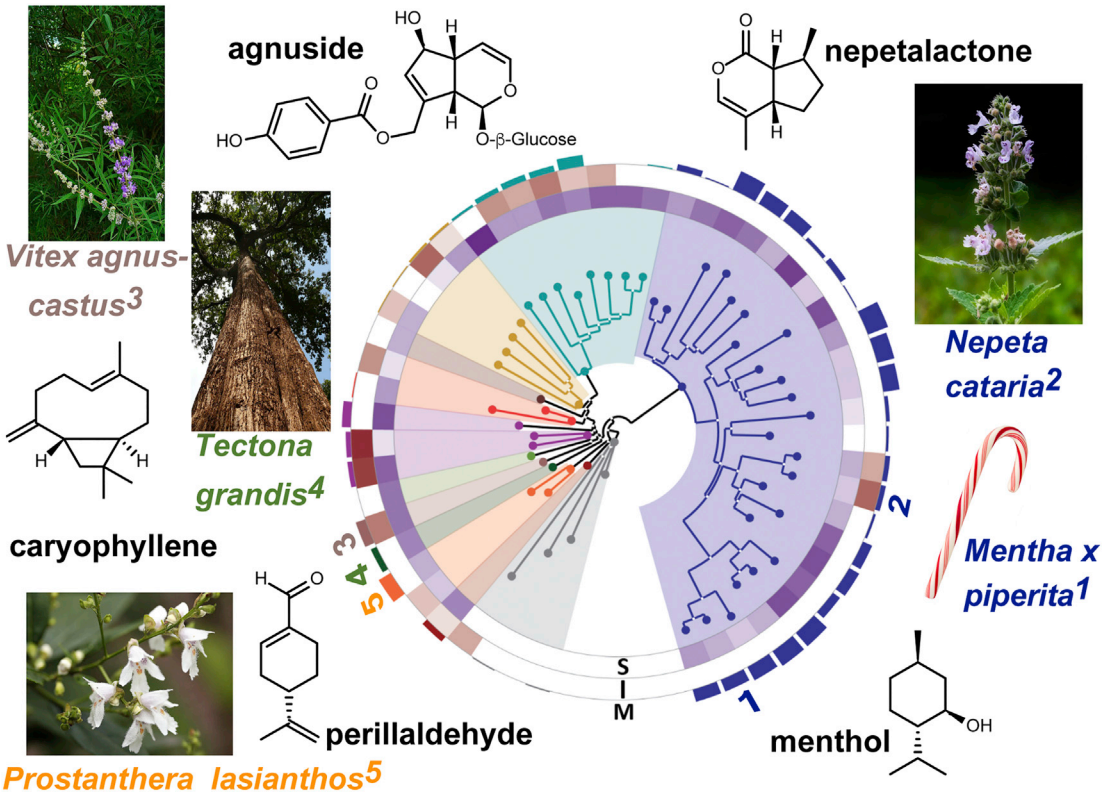

Figure 6. Chemical Diversity across Lamiaceae.
Phylogeny of Lamiaceae showing the diversity of sesquiterpenes and iridoids (heatmaps) and monoterpenes (bar graph) prepared with GraPhlAn (Asnicar et al., 2015). Species are color coded as shown in Figure 2 and include *Mentha* × *piperita*, known for menthol production, which is associated with the flavor of peppermint, *Nepeta cataria*, which produces the volatile iridoid nepetalactone, the active compound in catnip, *Vitex agnus-castus* (chaste tree), which produces the iridoid glycoside agnuside, *Tectona grandis* (teak), which produces caryophyllene, and *Prostanthera lasianthos* (Victorian Christmas bush), which produces perrillaldehyde. Photos courtesy of H. Zell (*V. agnus-castus*), L. Allison (*P. lasianthos*), P. Jeganathan (*T. grandis*), and Evan-Amos (peppermint) via the Creative Commons Attribution Share Alike license.

METHODS

Selection of the Species

Phylogenetic relationships within Lamiaceae have proven difficult to disentangle (Bendiksby et al., 2011; Li et al., 2012, 2016; Chen et al., 2014) (Supplemental Figure 1). The most recent multi-gene analyses provide support for the presence of five major subclades (e.g., Ajugoideae [~760 species]), Lamioideae [~1200 species], Prostantheroideae [~320 species], Scutellarioideae [~380 species], and Nepetoideae [~3500 species]) and a phylogenetic framework for the family (Li et al., 2016). A summary tree that reflected our best understanding of the Lamiaceae phylogeny guided an evolutionarily based sampling scheme to analyze chemical diversity (Supplemental Table 1). Species from 11 of 12 recognized subclades (48 total species) represent the phylogenetic diversity of Lamiaceae; Cymarioideae were recently described and are not represented in our sampling (Li et al., 2016).

Germplasm and Tissues Used in This Study

Accessions for nearly all 52 species of Lamiales investigated were obtained from commercial nurseries or botanical gardens (Supplemental Table 1) and grown in a greenhouse prior to verification of species identity and sampling. A few samples, such as the tree *Tectona grandis* L. f. (teak), were obtained from botanical gardens and sampled *in situ*. Fully expanded young leaves were harvested from mature, vegetative plants of each species; however, for a few species, the plants were flowering at the time of leaf tissue collection (Supplemental Table 1). To obtain sufficient tissue for both chemical profiling and transcriptome analysis, it was occasionally necessary to include older leaves as well as the targeted young leaves. Tissue was sampled from a single individual (see photographs on the Dryad Digital Repository), except for four species with extremely small leaves for which it was necessary to combine leaves from multiple individuals (Supplemental Table 1). Tissue was ground into a fine powder, and divided into aliquots to permit sampling of the same tissue pool for metabolite and transcriptome analyses. Genome size (Supplemental Table 1) was estimated from leaf tissues at the Flow Cytometry Facility at the Benaroya Research Institute at Virginia Mason.

RNA Isolation, Transcriptome Sequencing, Assembly, and Annotation

RNA isolation, construction of RNA-sequencing libraries, sequencing, assembly, annotation, and expression abundance estimations are described in the Supplemental Information. Expression abundances of orthogroups in the MEP, iridoid, and terpene synthase pathways were converted to *Z*-scores ([$\log_2$(orthogroup FPKM) − (species mean FPKM)]/[SD of species FPKMs]; FPKM = fragments per kilobase per million mapped reads), which were used to generate an expression heatmap using the ComplexHeatmap package in RStudio (v1.0.136).

Orthologous/Paralogous Groups, Single-Copy Nuclear Gene Discovery, and Phylogeny Reconstruction

Orthogroups were identified with OrthoFinder (v 0.7.1; Emms and Kelly, 2015) using Transdecoder-predicted peptide sequences (Haas et al.,

2013; see Supplemental Information for details). Coding sequences (CDS) predicted by Transdecoder from representative transcripts for 48 Lamiaceae and four outgroup taxa were processed with MarkerMiner (v1.2; Chamala et al., 2015) using default settings to identify single-copy nuclear (SCN) genes for phylogenetic analysis. Phylogenetic relationships were reconstructed from the 520 aligned loci (Supplemental Tables 2 and 3) using (i) supermatrix and (ii) coalescent-based species tree methods (see Supplemental Information for details).

Metabolite Analyses

For analysis of monoterpenes, sesquiterpenes, and volatile iridoids, metabolites were extracted from ground plant tissue as described in Supplemental Information and injected on the GC–MS instrument as previously described (Dudareva et al., 2005). Monoterpenes were identified using *m/z* values characteristic of isoprenes (*m/z* 69), monoterpenes (*m/z* 137), and monoterpene alcohols (*m/z* 155) and confirmed by Agilent GC/MSD ChemStation. Sesquiterpenes were identified using *m/z* values for isoprenes and sesquiterpenes (*m/z* 204). Iridoid glycosides were extracted as described in Supplemental Information, and initial analysis of all 52 leaf tissue samples was performed using an ultra-performance liquid chromatography (UPLC) Shimadzu IT-ToF/MS (ion-trap time-of-flight–MS; Shimadzu, Kyoto, Japan), equipped with an electrospray ionization source, detailed methods for which are described in Supplemental Information. An additional, second analysis (30 species: all outgroups, all non-Nepetoideae species, and both *Nepeta* species) was performed using a high-definition UPLC Acquity Synapt G2Si Q-ToF (quadrupole time-of-flight) tandem MS instrument (see Supplemental Information).

Ancestral State Reconstructions and Phylogenetic Signal

Chemical diversity data (3 chemical classes and 83 compounds; Supplemental Tables 4, 5, and 7) for all species were scored as binary (presence/absence) traits, and the evolutionary history of each compound or compound class was reconstructed across the ML topology using ML methods as implemented in Mesquite 3.11 (http://mesquiteproject.org; Supplemental Information). Phylogenetic signal was measured for 92 selected traits (Supplemental Table 6) and used to test whether species' traits are randomly distributed across the phylogeny (Figure 2) or clustered as the result of shared ancestry. All analyses were conducted using an ultrametric version of our ML topology (Figure 2) that was generated using Sanderson (2002) penalized likelihood with cross-validation to estimate ë ("chronopl" function in the APE package; https://cran.r-project.org/web/packages/ape/index.html; Supplemental Information).

Statistical Analyses of Chemical and Genomic Traits

Species are part of a hierarchically structured phylogeny, and, for statistical purposes, any trait data observed across species should be considered non-independent (Felsenstein, 1985). When available, statistical methods that accounted for phylogenetic relationships among species were used to test hypotheses and to model relationships using discrete and continuous chemical (e.g., presence/absence, diversity, and abundance) and genomic (e.g., orthogroup occupancy and expression) traits. All analyses were conducted using an ultrametric version of the ML tree with outgroups removed (see aforementioned Methods). Phylogenetic regression models were built for chemical traits (presence/absence and diversity) against the two genomic variables (orthogroup occupancy and expression). The analyses were implemented with the "phyloglm" function of the "phylolm" R package (https://cran.r-project.org/web/packages/phylolm/phylolm.pdf). To investigate the influence of gene family size on chemical diversity traits, phylogenetic Poisson regression (Paradis and Claude, 2002) models were built using species count data representing the number of iridoid, monoterpene, and sesquiterpene compounds (response variables) and species orthogroup occupancies (predictor variables). The influence of orthogroup expression on binary chemical traits was also investigated. For these analyses, phylogenetic logistic regression (Ives and Garland, 2010; Ho and Ane, 2014) models were built using iridoid, monoterpene, and sesquiterpene presence/absence data (binary response variables) and orthogroup expression *Z*-scores (continuous predictor variables). The "logistic_MPLE" option was used to maximize the penalized likelihood in all logistic regressions.

ACCESSION NUMBERS

Raw sequence reads have been deposited in the National Center for Biotechnology Information under BioProject ID PRJNA359989. Due to the large files of a number of datasets, we have deposited the following files on the Dryad Digital Repository (https://doi.org/10.5061/dryad.tj1p3).

(i) Ancestral state reconstructions; (ii) Expression abundances and functional annotation of all 52 Lamiales species; (iii) GC–MS chromatograms of 52 Lamiales species; (iv) LC–MS data of 52 Lamiales species; (v) OrthoFinder-derived orthologous and paralogous groups; (vi) Predicted CDS of all 52 Lamiales species; (vii) Predicted peptides of all 52 Lamiales species; (viii) Transcriptomes of all 52 Lamiales species; and (ix) Photographs of plants used for this study.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

Funds for this study were provided by a grant to C.R.B., N.D., S.E.O., D.S., and P.S. from the National Science Foundation Plant Genome Research Program (IOS-1444499) and from Hatch Funds (MICL02431) to C.R.B.

CONSORTIA

The members of the Mint Evolutionary Genomics Consortium for this project are, Benoît Boachon, C. Robin Buell, Emily Crisovan, Natalia Dudareva, Nicolas Garcia, Grant Godden, Laura Henry, Mohamed O. Kamileen, Heather Rose Kates, Matthew B. Kilgore, Benjamin R. Lichman, Evgeny V. Mavrodiev, Linsey Newton, Carlos Rodriguez-Lopez, Sarah E. O'Connor, Douglas Soltis, Pamela Soltis, Brieanne Vaillancourt, Krystle Wiegert-Rininger, Dongyan Zhao.

AUTHOR CONTRIBUTIONS

C.R.B., N.D., S.E.O., D.S., and P.S. conceived this study. E.C., E.V.M., G.G., N.G., L.H., L.N., H.R.K., M.B.K., M.O.K., B.R.L., K.W.-R., and D.Z. generated data. B.B., C.R.B., N.D., G.G., N.G., L.H., H.R.K., M.O.K., B.R.L., E.V.M., S.E.O., C.R.-L., D.S., P.S., K.W.-R., B.V., and D.Z. analyzed data. All authors wrote portions of the manuscript.

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: March 24, 2018
Revised: June 5, 2018
Accepted: June 10, 2018
Published: June 16, 2018

REFERENCES

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ **3**:e1029.

Bendiksby, M., Thorbek, L., Scheen, A.-C., Lindqvist, C., and Ryding, O. (2011). An updated phylogeny and classification of Lamiaceae subfamily Lamioideae. Taxon **60**:471–484.

Birkett, M.A., Hassanali, A., Hoglund, S., Pettersson, J., and Pickett, J.A. (2011). Repellent activity of catmint, *Nepeta cataria*, and iridoid nepetalactone isomers against Afro-tropical mosquitoes, ixodid ticks and red poultry mites. Phytochemistry **72**:109–114.

Bozin, B., Mimica-Dukic, N., Simin, N., and Anackov, G. (2006). Characterization of the volatile composition of essential oils of some Lamiaceae spices and the antimicrobial and antioxidant activities of the entire oils. J. Agric. Food Chem. **54**:1822–1828.

Brockington, S.F., Walker, R.H., Glover, B.J., Soltis, P.S., and Soltis, D.E. (2011). Complex pigment evolution in the Caryophyllales. New Phytol. **190**:854–864.

Chadwick, M., Trewin, H., Gawthrop, F., and Wagstaff, C. (2013). Sesquiterpenoids lactones: benefits to plants and people. Int. J. Mol. Sci. **14**:12780–12805.

Chamala, S., Garcia, 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**. https://doi.org/10.3732/apps.1400115.

Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. **66**:212–229.

Chen, Y.-P., Li, B., Olmstead, R.G., Cantino, P.D., Liu, E.-D., and Xiang, C.-L. (2014). Phylogenetic placement of the enigmatic genus *Holocheila* (Lamiaceae) inferred from plastid DNA sequences. Taxon **63**:355–366.

Crocoll, C., Asbach, J., Novak, J., Gershenzon, J., and Degenhardt, J. (2010). Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. Plant Mol. Biol. **73**:587–603.

De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc. Natl. Acad. Sci. USA **110**:2898–2903.

Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., Boland, W., and Gershenzon, J. (2005). The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proc. Natl. Acad. Sci. USA **102**:933–938.

Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glockner, G., Platts, A.E., Heckel, D.G., Der, J.P., Wafula, E.K., Tang, M., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. Proc. Natl. Acad. Sci. USA **112**:8362–8366.

El-Gazzar, A., and Watson, L. (1970). Some economic implications of the taxonomy of Labiatae. New Phytol. **69**:487–492.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. **16**:157.

Estevez, J.M., Cantero, A., Reindl, A., Reichler, S., and Leon, P. (2001). 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. J. Biol. Chem. **276**:22901–22909.

Ettlinger, M., and Kjaer, A. (1968). Sulfur compounds in plants. Recent Adv. Phytochem. **1**:59–144.

Felsenstein, J. (1985). Phylogenies and the comparative method. Am. Nat. **125**:1–15.

Gershenzon, J., McConkey, M.E., and Croteau, R.B. (2000). Regulation of monoterpene accumulation in leaves of peppermint. Plant Physiol. **122**:205–214.

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

Harley, R., Atkins, S., Budantsev, A.L., Cantino, P.D., Conn, B.J., Grayer, R., Harley, M.M., de Kok, R., Krestovskaja, T., Morales, R., et al. (2004). Labiatae. In The Families and Genera of Vascular Plants, K. Kubitzki, ed. (Berlin: Springer Verlag), pp. 167–275.

Hegnauer, R. (1989). Labiatae. In Chemotaxonomie der Pflanzen. Lehrbücher und Monographien aus dem Gebiete der Exakten Wissenschaften, R. Hegnauer, ed. (Basel: Birkhäuser), pp. 579–631.

Heywood, V.H., Harborne, J.B., and Turner, B.L. (1977). The Biology and Chemistry of the Compositae (London: New York Academic Press).

Ho, L., and Ane, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst. Biol. **63**:397–408.

Hudaib, M., Tawaha, K., and Bustanji, Y. (2013). Chemical profile of the volatile oil of lemon verbena (*Aloysia citriodora* Paláu) growing in Jordan. J. Essent. Oil Bear. Pl. **16**:568–574.

Ives, A.R., and Garland, T. (2010). Phylogenetic logistic regression for binary dependent variables. Syst. Biol. **59**:9–26.

Jensen, S.R. (1991). Plant iridoids, their biosynthesis and distribution in angiosperms. Proc. Phytochem. Soc. Eur. **31**:33–158.

Kliebenstein, D.J., Lambrix, V.M., Reichelt, M., Gershenzon, J., and Mitchell-Olds, T. (2001). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell **13**:681–693.

Konno, K., Hirayama, C., Yasui, H., and Nakamura, M. (1999). Enzymatic activation of oleuropein: a protein crosslinker used as a chemical defense in the privet tree. Proc. Natl. Acad. Sci. USA **96**:9159–9164.

Kooiman, P. (1972). The occurrence of iridoid glycosides in the Labiatae. Acta Bot. Neerland. **21**:417–427.

Krithika, R., Srivastava, P.L., Rani, B., Kolet, S.P., Chopade, M., Soniya, M., and Thulasiram, H.V. (2015). Characterization of 10-hydroxygeraniol dehydrogenase from *Catharanthus roseus* reveals cascaded enzymatic activity in iridoid biosynthesis. Sci. Rep. **5**:8258.

Lange, B.M. (2015). The evolution of plant secretory structures and emergence of terpenoid chemical diversity. Ann. Rev. Plant Biol. **66**:139–159.

Lange, B.M., Wildung, M.R., Stauber, E.J., Sanchez, C., Pouchnik, D., and Croteau, R. (2000). Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. Proc. Natl. Acad. Sci. USA **97**:2934–2939.

Li, B., Cantino, P.D., Olmstead, R.G., Bramley, G.L., Xiang, C.L., Ma, Z.H., Tan, Y.H., and Zhang, D.X. (2016). A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. Sci. Rep. **6**:34343.

Li, B., and Olmstead, R.G. (2017). Two new subfamilies in Lamiaceae. Phytotaxa **313**:222–226.

Li, B., Xu, W., Tu, T., Wang, Z., Olmstead, R.G., Peng, H., Francisco-Ortega, J., Cantino, P.G., and Zhang, D. (2012). Phylogenetic position of *Wenchengia* (Lamiaceae): a taxonomically enigmatic and critically endangered genus. Taxon **61**:392–401.

Matsuno, M., Compagnon, V., Schoch, G.A., Schmitt, M., Debayle, D., Bassard, J.E., Pollet, B., Hehn, A., Heintz, D., Ullmann, P., et al. (2009). Evolution of a novel phenolic pathway for pollen development. Science **325**:1688–1692.

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., et al. (2014). The seco-iridoid pathway from *Catharanthus roseus*. Nat. Commun. **5**:3606.

Moghe, G.D., Leong, B.J., Hurney, S.M., Daniel Jones, A., and Last, R.L. (2017). Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. Elife **6**. https://doi.org/10.7554/eLife.28468.

Morton, J.K. (1973). A cytological study of the British Labiatae (excluding *Mentha*). Watsonia **9**:239–246.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., et al. (2014). The genome of *Eucalyptus grandis*. Nature **510**:356–362.

O'Maille, P.E., Malone, A., Dellas, N., Andes Hess, B., Jr., Smentek, L., Sheehan, I., Greenhagen, B.T., Chappell, J., Manning, G., and Noel, J.P. (2008). Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. Nat. Chem. Biol. **4**:617–623.

Ono, E., Homma, Y., Horikawa, M., Kunikane-Doi, S., Imai, H., Takahashi, S., Kawai, Y., Ishiguro, M., Fukui, Y., and Nakayama, T. (2010). Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). Plant Cell **22**:2856–2871.

Paradis, E., and Claude, J. (2002). Analysis of comparative data using generalized estimating equations. J. Theor. Biol. **218**:175–185.

Rai, A., Smita, S.S., Singh, A.K., Shanker, K., and Nagegowda, D.A. (2013). Heteromeric and homomeric geranyl diphosphate synthases from *Catharanthus roseus* and their role in monoterpene indole alkaloid biosynthesis. Mol. Plant **6**:1531–1549.

Sanderson, M.J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. **19**:101–109.

Schilmiller, A.L., Moghe, G.D., Fan, P., Ghosh, B., Ning, J., Jones, A.D., and Last, R.L. (2015). Functionally divergent alleles and duplicated Loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. Plant Cell **27**:1002–1017.

Sena Filho, J.G., Xavier, H.S., Barbosa Filho, J.M., and Duringer, J.M. (2010). A chemical marker proposal for the *Lantana* genus: composition of the essential oils from the leaves of *Lantana radula* and *L. canescens*. Nat. Prod. Commun. **5**:635–640.

Taskova, R., Mitova, M., Evstatieva, L., Ancev, M., Peev, D., Handjieva, N., Bankova, V., Popov, S., et al. (1997). Iridoids, flavonoids and terpenoids as taxonomic markers in Lamiaceae, Scrophulariaceae, and Rubiaceae. Bocconea **5**:631–636.

Turner, G.W., and Croteau, R. (2004). Organization of monoterpene biosynthesis in *Mentha*. Immunocytochemical localizations of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase. Plant Physiol. **136**:4215–4227.

Vining, K.J., Johnson, S.R., Ahkami, A., Lange, I., Parrish, A.N., Trapp, S.C., Croteau, R.B., Straub, S.C., Pandelova, I., and Lange, B.M. (2017). Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. Mol. Plant **10**:323–339.

Weng, J.K., Philippe, R.N., and Noel, J.P. (2012). The rise of chemodiversity in plants. Science **336**:1667–1670.

Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry **64**:3–19.

Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., Xu, J., Li, Y., Song, C., Wang, B., et al. (2016). Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. Mol. Plant **9**:949–952.

Xu, S., Brockmoller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., Zhou, W., Kreitzer, C., Stanke, M., Tang, H., et al. (2017). Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc. Natl. Acad. Sci. USA **114**:6133–6138.