# The biosynthesis of the anti-microbial diterpenoid leubethanol in *Leucophyllum frutescens* proceeds via an all-*cis* prenyl intermediate

Garret P. Miller (iD), Wajid Waheed Bhat (iD), Emily R. Lanier (iD), Sean R. Johnson (iD), Davis T. Mathieu and Björn Hamberger[*] (iD)

*Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, USA*

## SUMMARY

Serrulatane diterpenoids are natural products found in plants from a subset of genera within the figwort family (Scrophulariaceae). Many of these compounds have been characterized as having anti-microbial properties and share a common diterpene backbone. One example, leubethanol from Texas sage (*Leucophyllum frutescens*) has demonstrated activity against multi-drug-resistant tuberculosis. Leubethanol is the only serrulatane diterpenoid identified from this genus; however, a range of such compounds have been found throughout the closely related *Eremophila* genus. Despite their potential therapeutic relevance, the biosynthesis of serrulatane diterpenoids has not been previously reported. Here we leverage the simple product profile and high accumulation of leubethanol in the roots of *L. frutescens* and compare tissue-specific transcriptomes with existing data from *Eremophila serrulata* to decipher the biosynthesis of leubethanol. A short-chain *cis*-prenyl transferase (*Lf*CPT1) first produces the rare diterpene precursor nerylneryl diphosphate, which is cyclized by an unusual plastidial terpene synthase (*Lf*TPS1) into the characteristic serrulatane diterpene backbone. Final conversion to leubethanol is catalyzed by a cytochrome P450 (CYP71D616) of the CYP71 clan. This pathway documents the presence of a short-chain *cis*-prenyl diphosphate synthase, previously only found in Solanaceae, which is likely involved in the biosynthesis of other known diterpene backbones in *Eremophila*. *Lf*TPS1 represents neofunctionalization of a compartment-switching terpene synthase accepting a novel substrate in the plastid. Biosynthetic access to leubethanol will enable pathway discovery to more complex serrulatane diterpenoids which share this common starting structure and provide a platform for the production and diversification of this class of promising anti-microbial therapeutics in heterologous systems.

Keywords: terpene biosynthesis, all-*cis* prenyl diphosphate, cytochrome P450, leubethanol, anti-microbial.

## INTRODUCTION

Terpenoids are a major class of specialized metabolites in plants, with applications ranging from fragrances and cosmetics to pesticides and pharmaceuticals. This wide variety of uses can be attributed to the incredible structural diversity of terpenoid compounds, resulting from sequential and combinatorial modifications of common starting molecules. Typically beginning with three common C10, C15, and C20 *trans*-prenyl diphosphate substrates, hundreds of mono-, sesqui-, and diterpene backbones are cyclized by terpene synthases (TPSs) (Degenhardt *et al.*, 2009; Durairaj *et al.*, 2019; Johnson *et al.*, 2019a). These backbones are further diversified to thousands of terpenoids (Zeng *et al.*, 2019) through successive modification by enzymes such as cytochrome P450 mono-oxygenases, aldehyde dehydrogenases, and acetyl transferases (Luo *et al.*, 2016; Pateraki *et al.*, 2017; Johnson *et al.*, 2019b; Karunanithi and Zerbe, 2019). Diterpenoids (C20) make up more than 13 000 known plant terpenoids (Johnson *et al.*, 2019a), and the vast majority with known biosynthetic pathways are derived from all-*trans* (*E,E,E*)-geranylgeranyl diphosphate (GGPP). A given diterpene backbone can be the source of anywhere from one to hundreds of diterpenoids following downstream modification (Johnson *et al.*, 2019a), and the biosynthetic routes for the majority of these backbones remain unknown (Johnson *et al.*, 2019b).

The serrulatane diterpenoids are one such example of a range of compounds derived from a single diterpene backbone (Figure 1), with more than 30 identified within the Scrophulariaceae family (order Lamiales) (DNP v28.2).
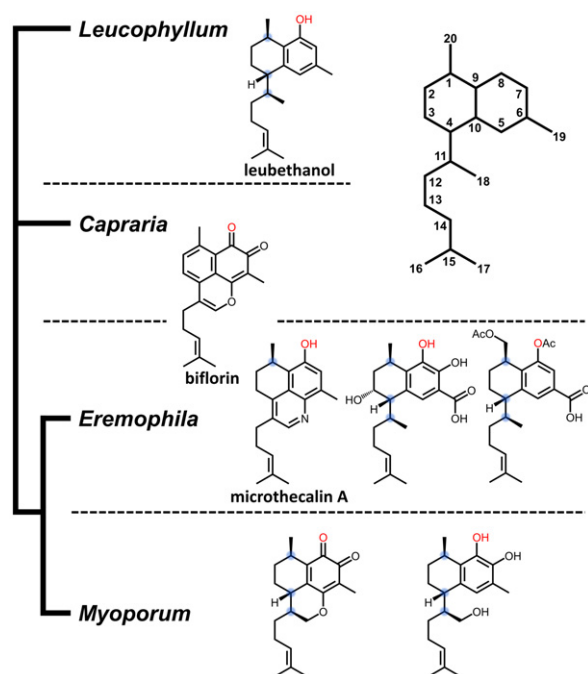
**Figure 1.** Distribution of serrulatane diterpenoids in members of the Scrophulariaceae family. Shared backbone structure in top right. Leubethanol contains common stereochemistry highlighted in blue, and common oxygenation highlighted in red. Only one serrulatane diterpenoid each has been identified in *Leucophyllum* (leubethanol) and *Capraria* (biflorin), while only a few representatives are shown for the *Eremophila* and *Myoporum* genera; *Eremophila* alone harbors more than 30.

Many of these compounds have been shown to be bioactive. Leubethanol from *Leucophyllum frutescens* is active against multi-drug-resistant tuberculosis (Molina-Salinas et al., 2011), biflorin from *Capraria biflora* and *Eremophila neglecta* has both anti-tumor (Carvalho et al., 2013) and anti-microbial (Ndi et al., 2007a) properties, and microthecaline A from *Eremophila microtheca* is active against malaria (Kumar et al., 2018). While relatively few have been identified in other genera, the *Eremophila* genus is especially rich in these anti-microbial compounds (Ndi et al., 2007b; Anakok et al., 2012; Barnes et al., 2013; Mon et al., 2015; Algreiby et al., 2018; Hossain et al., 2019) and at least three are found in the *Myoporum* genus (Aminimoghadamfarouj and Nematollahi, 2017). Given their promise in therapeutic applications, there has been a substantial effort to devise total chemical syntheses (Best and Wege, 1986; Lu et al., 2013; Yu et al., 2016; R. Kumar et al., 2017; Tenneti et al., 2018; Penjarla et al., 2019) as an alternative to extraction and purification from natural sources (Figure 2).

Despite the efforts invested into natural product discovery, anti-microbial screens, and total chemical syntheses, the biosynthetic pathway to these serrulatane diterpenoids has remained elusive. Identifying the enzymes responsible would pave the way for production of these serrulatane

diterpenoids in heterologous systems, offering an appealing alternative to formal chemical synthesis. To address this, we sought to elucidate the biosynthetic pathway to the serrulatane diterpenoid leubethanol in *L. frutescens*. Three properties of this species made it an ideal target for studying this pathway. First, leubethanol is the only serrulatane diterpenoid known to be produced by this plant and accumulates in high quantities in root tissue (Molina-Salinas et al., 2011). Second, RNA-seq data are publicly available for the closely related species *Eremophila serrulata* (Kracht et al., 2017), allowing for comparative transcriptomics between genera. Third, leubethanol shares a common hydroxylation with the majority of other known serrulatane diterpenoids (Figure 1) and is the likely intermediate in their biosynthesis. Beyond these advantages, leubethanol is itself an appealing target, with activity against multi-drug-resistant tuberculosis (minimum inhibitory concentration 22 μM) comparable to those of isoniazid (23 μM) and ethambutol (39 μM) (Molina-Salinas et al., 2011), two drugs commonly used in combination therapy.

Through comparing transcriptomes between tissue types and genera, we have identified three enzymes which constitute the full biosynthetic pathway to leubethanol in *L. frutescens*. While the vast majority of known diterpenoids originate from GGPP, this pathway involves a short-chain *cis*-prenyl transferase (*cis*-PT) which produces the uncommon diterpene precursor (*Z,Z,Z*)-nerylneryl diphosphate (NNPP – the all-*cis* stereoisomer of GGPP). This is then cyclized to the shared serrulatane backbone by a TPS which exclusively converts NNPP and is a member of the primarily sesquiterpene (C15) synthase TPS-a subfamily. Finally, this backbone is converted to leubethanol by a cytochrome P450 of the CYP71 clan, which harbors many recently discovered P450s involved in terpene specialized metabolism (Hamberger and Bak, 2013; Bathe and Tissier, 2019). The identification of a short-chain *cis*-PT which makes NNPP clarifies the likely origin of other diterpenes identified in the *Eremophila* genus based on the presence of *cis* double bonds in these backbones. Reconstruction of the full pathway in *Nicotiana benthamiana* allowed for production of leubethanol in a heterologous system and provides access to a plausible key intermediate in the biosynthesis of other serrulatane diterpenoids.

## RESULTS

### Accumulation of leubethanol-guided tissue-specific RNA sequencing

To begin our search for the biosynthetic pathway to leubethanol, we took advantage of its tissue-specific accumulation in *L. frutescens*. Previous work on the medicinal properties of this species has shown that root extracts were most potent against multi-drug-resistant tuberculosis, while leaves showed some activity and flowers showed

**Figure 2.** Maximum likelihood phylogenetic tree of TPS candidates. Candidate terpene synthases from *L. frutescens* and *E. serrulata* are shown in purple and yellow, respectively, with reference TPSs in black. Putative transit peptides, predicted by N-terminal extensions, are denoted within the TPS-a subfamily.
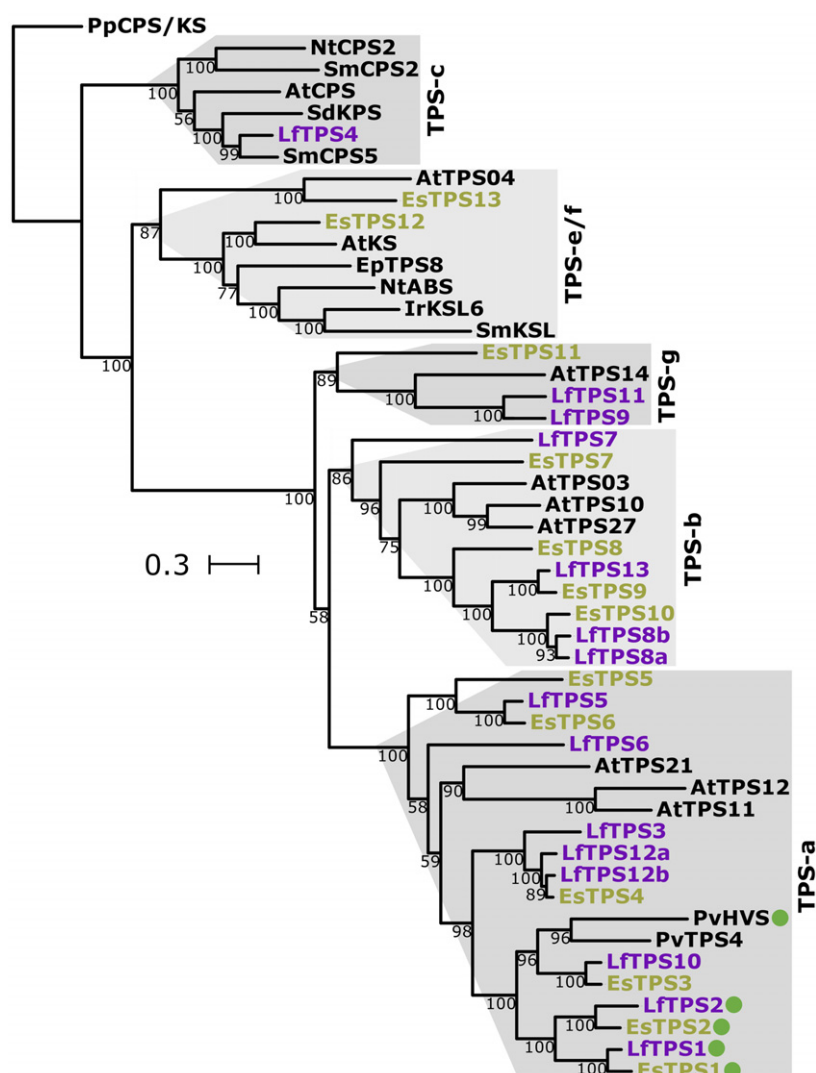
none (Molina-Salinas *et al.*, 2007). To confirm the tissue-specific accumulation of leubethanol, extracts of the leaves, roots, and flowers were analyzed by GC-MS. Leubethanol was found to accumulate in both root and leaf tissue, while none was detected in flower tissue (Figure S1). Consequently, we isolated and sequenced RNA from both roots and flowers to allow for comparative transcriptomics between tissue types. Serrulatane diterpenoids are also found in the closely related *Eremophila* genus (Ghisalberti, 1993). RNA-seq data are publicly available from the leaves of *E. serrulata* (Sequence Read Archive [SRA]: ERX1321488; (Kracht *et al.*, 2017)), and serrulatanes are known to accumulate in this tissue (Ndi, 2007b). These data were also included to allow for comparison between genera.

**Identification of TPS candidates from *L. frutescens***

We began our search by identifying TPS candidates from *L. frutescens* through a homology-based search of our

transcriptomic data against a reference set of TPSs (Dataset S1). Fifteen candidates were identified, and a phylogenetic tree was constructed to group each candidate by TPS subfamily (Figure 1). One candidate (*Lf*TPS13) was not expressed in root tissue and was eliminated from further consideration.

While containing a bicyclic decalin core, the structure of leubethanol is inconsistent with the labdane group of plant diterpenoids, the most common type of backbone which results from cyclization by pairs of class II and class I diTPS (Peters, 2010). In contrast, the cyclization pattern of leubethanol indicates activity of a class I enzyme, which catalyzes cyclization via removal of the diphosphate moiety. Out of the 14 root-expressed candidates, only one was predicted to be a class II TPS (*Lf*TPS4; TPS-c subfamily), and therefore 13 possibilities remained.

A number of non-labdane diterpenes have been shown previously to be made by TPS-a enzymes which are localized to the plastid (Mau *et al.*, 1994; Ennajdaoui *et al.*, 2010;

Kirby *et al.*, 2010; Vaughan *et al.*, 2013; Zerbe *et al.*, 2013; Wang *et al.*, 2016; Luo *et al.*, 2016; Johnson *et al.*, 2019b). The majority of TPS-a enzymes are sesquiterpene synthases localized to the cytosol (Chen *et al.*, 2011), and the presence of an N-terminal plastidial transit peptide in the primary amino acid sequence can therefore aid in prediction of diterpene synthase activity in this subfamily (Johnson *et al.*, 2019b). Two *L. frutescens* candidates (*Lf*TPS1 and *Lf*TPS2) in the TPS-a subfamily were found to carry N-terminal extensions. Additionally, both have an ortholog in *E. serrulata* with nearly identical sequence length and homology through these N-terminal extensions (Figure S2). Of these two candidates, only *Lf*TPS1 is exclusively expressed in root tissue and was therefore considered the more likely candidate; however, both were tested.

Full-length genes for both *Lf*TPS1 and *Lf*TPS2 were cloned from root cDNA for transient expression in an *N. benthamiana* system engineered for increased levels of the presumed substrate GGPP (Andersen-Ranberg *et al.*, 2016). N-terminal truncated constructs, removing the putative transit peptides, were cloned into pET-28b(+) for expression of pseudomature variants in *Escherichia coli*. Assays were extracted with hexane and analyzed by GC-MS.

To account for uncertainty of the predicted plastidial targeting signals, transient expression assays in *N. benthamiana* were carried out separately with co-expression of either plastidial or cytosolic GGPP terpene precursor pathway enzymes. Co-expression of both candidates with either cytosolic or plastidial precursor enzymes did not yield detectable products (Figure S3). To independently verify activity, each enzyme was expressed in *E. coli* with a C-terminal histidine tag and purified through Ni-affinity chromatography. Consistent with the results of the transient *N. benthamiana* assays, incubation of both *Lf*TPS1

and *Lf*TPS2 with GGPP in *in vitro* assays yielded no measurable activity. Additionally, no activity was seen when incubated with farnesyl diphosphate (FPP, precursor for sesquiterpenes) or geranyl diphosphate (GPP, precursor for monoterpenes) (Figure S3).

## *Lf*TPS1 exclusively cyclizes nerylneryl diphosphate into the serrulatane backbone

Following these results, we considered two routes forward: first, to expand testing to each other class I TPS candidate, and second, to test *Lf*TPS1 and *Lf*TPS2 against uncommon terpene precursors. The former route was considered because even very closely related TPSs can have activities which differ substantially (Durairaj *et al.*, 2019), and there are many examples of TPSs which have different functions than would be predicted by their subfamily (Johnson *et al.*, 2019b). The latter route was considered because of the absence of activity against each common substrate. GPP, FPP, and GGPP contain exclusively *trans* double bonds. All-*cis* stereoisomers of each have been reported in members of the nightshade (Solanaceae) family (Akhtar *et al.*, 2013), together with TPSs which can convert these to terpene products (Sallaud *et al.*, 2009; Schilmiller and Schauvinhold, 2009; Zi *et al.*, 2014; Matsuba *et al.*, 2015).

The serrulatane backbone is ambiguous with respect to the original stereochemistry of its precursor; however, closer inspection of diterpenoids from the *Eremophila* genus shows that acyclic, bisabolane, and cembrane type diterpenoids (Figure 3(a)) in various *Eremophila* species contain internal *cis* double bonds (Ghisalberti, 1993). This prompted us to test NNPP (the all-*cis* stereoisomer of GGPP) as the precursor for the serrulatane backbone in *L. frutescens*. Since NNPP is not commercially available, truncated constructs of *Lf*TPS1 and *Lf*TPS2 in pET-28b(+) were used for co-expression with *Sl*CPT2, the plastidial
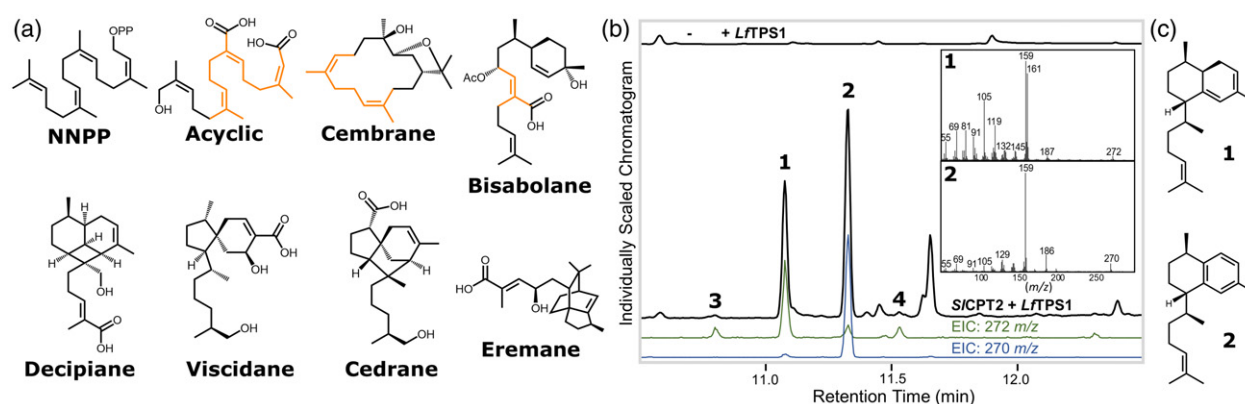


**Figure 3.** Dihydroserrulatene production by *Lf*TPS1. (a) Structure of NNPP and representative non-serrulatane diterpenoids from *Eremophila* with labeled backbone structure types. Isoprenyl subunits found in acyclic, cembrane, and bisabolane type diterpenoids, which can be used to infer the stereochemistry of their prenyl diphosphate precursor, are highlighted in orange. (b) Transient expression of *Lf*TPS1 in *N. benthamiana* with *Sl*CPT2 (a plastidial NNPP synthase), and mass spectra of major products. Each assay has *Cf*DXS co-expressed in addition to those listed. (c) Structures of dihydroserrulatene (**1**) and serrulatane (**2**).

*Solanum lycopersicum cis*-PT (Akhtar *et al.*, 2013), in an *E. coli* system engineered to increase terpene precursor availability (Cyr *et al.*, 2007). Following hexane extraction and analysis by GC-MS, *Lf*TPS1 was found to convert NNPP (Figure S4). This activity was independently confirmed in *N. benthamiana* (Figures 3(b) and S4).

Four diterpene products were observed, with only one major product (**1**) in the *E. coli* system, and a relative amount of **2** exceeding **1** in *N. benthamiana*. Diterpene olefins typically have a molecular ion of 272 *m/z*; however, **2** has a molecular ion of 270 *m/z*. The fragmentation pattern for **2** is consistent with an aromatic product, and is similar to that of leubethanol (286 *m/z*) with major peaks shifted by 16, consistent with a difference of one hydroxylation (Figure S1). Given that TPSs are not known to catalyze redox reactions, **2** is likely derived from spontaneous aromatization of the major product **1**, a phenomenon seen previously in diterpene biosynthesis (Zi *et al.*, 2013). To confirm the structure of **1**, production in the *E. coli* system was scaled up for NMR analysis (Table S1 and Figure S5), revealing that *Lf*TPS1 makes dihydroserrulatene (peak **1**), and supporting the identity of peak **2** as serrulatane (Figure 3(c)). A proposed mechanism for *Lf*TPS1 conversion of NNPP to dihydroserrulatene is given in Scheme S1.

In parallel to the testing against NNPP, we began working towards testing the remaining class I candidate TPSs. While we cloned each of these candidates out of *L. frutescens* cDNA, we received the positive results for *Lf*CPT1 conversion of NNPP to dihydroserrulatene before we characterized these other candidates. These were, however, cloned and sequence-verified, and are given here with GenBank accession numbers for reference.

### *Lf*CPT1, a short-chain *cis*-prenyl transferase, supplies NNPP in serrulatane biosynthesis

We next sought out the source of NNPP in *L. frutescens* by searching for a *cis*-prenyl transferase. *Cis*-PTs are ubiquitous throughout plants and are typically involved in the synthesis of long-chain polyisoprenoids (Akhtar *et al.*, 2013), although very few which make short-chain products (fewer than 35 carbons) have been identified. Three short-chain *cis*-PTs which yield NPP (neryl diphosphate; 10 carbons), (*Z-Z*)-FPP (*Z-Z*-farnesyl diphosphate; 15 carbons), and NNPP (20 carbons) have been identified from *S. lycopersicum* through functional characterization of the entire family of *cis*-PTs from this species (Akhtar *et al.*, 2013). We identified candidate *cis*-PTs from both the *L. frutescens* and *E. serrulata* transcriptomes through a homology-based search against the entire family of *cis*-PTs from *S. lycopersicum*. Ten candidate *cis*-PTs were identified from *L. frutescens*, and phylogenetic analysis revealed that six are closely related to the short-chain *cis*-PTs from *S. lycopersicum* (Figure 4). *Lf*TPS1 has a predicted plastidial transit peptide, and successfully converts NNPP in *N.*

*benthamiana* assays when co-expressed with *Sl*CPT2, which is known to be targeted to the plastid (Akhtar *et al.*, 2013). Therefore, we looked for a *cis*-PT candidate that is likely targeted to the plastid. Three of these candidates were found to carry predicted plastidial transit peptides and are expressed in root tissue (*Lf*CPT1-3). *Lf*CPT1 was considered to be the most likely candidate as it is the only of these three to have a direct ortholog in our *E. serrulata* transcriptome assembly (*Es*CPT1); however, all three were tested.

*Lf*CPT1-3 were cloned from *L. frutescens* root cDNA. Each candidate *cis*-PT was co-expressed in *N. benthamiana* with *Lf*TPS1, and products were analyzed by GC-MS following hexane extraction. Co-expression with *Lf*CPT1 yielded the same diterpene product profile as with the NNPP synthase from *S. lycopersicum* (*Sl*CPT2) (Figures 4 and S6). In addition, direct comparison of *Lf*CPT1 with *Sl*CPT2 without co-expression of a TPS showed the same peak and mass spectrum for dephosphorylated NNPP (Figure 4).

### A cytochrome P450 converts the serrulatane backbone to leubethanol

Leubethanol is oxidized twice relative to dihydroserrulatene, presumably through hydroxylation by a cytochrome P450 and aromatization. Given the propensity for dihydroserrulatene to spontaneously aromatize to serrulatane, we set out to identify P450 candidates for the required oxidation at C8. A homology-based search of both the *L. frutescens* and *E. serrulata* transcriptomes was carried out against a reference set of plant P450s (Dataset S1). A total of 165 candidates were identified from *L. frutescens*. We first narrowed our search by focusing on those within the CYP71 clan. While P450s in other clans have been identified in diterpenoid specialized metabolism, we began our search here based on the CYP71 clan containing the majority of previously characterized examples (Hamberger and Bak, 2013; Bathe and Tissier, 2019). Clustering each P450 candidate by family and eliminating those outside of the CYP71 clan reduced the list of candidates to 59 (Figure S7). Considering only those that were expressed in root tissue but not flower tissue and those that had an ortholog in our *E. serrulata* transcriptome assembly, only five candidates remained. One additional candidate (CYP71D615), which did not have a direct ortholog in *E. serrulata*, was included based on its root-exclusive expression and location among a cluster of other *L. frutescens* and *E. serrulata* candidates in the phylogenetic tree (Figure S7).

These six P450 candidates were cloned from *L. frutescens* root cDNA. Co-expression with *Lf*CPT1 and *Lf*TPS1 in *N. benthamiana* revealed that CYP71D616 facilitates the conversion of dihydroserrulatene to leubethanol (Figures 5 and S8). A relative decrease of dihydroserrulatene over
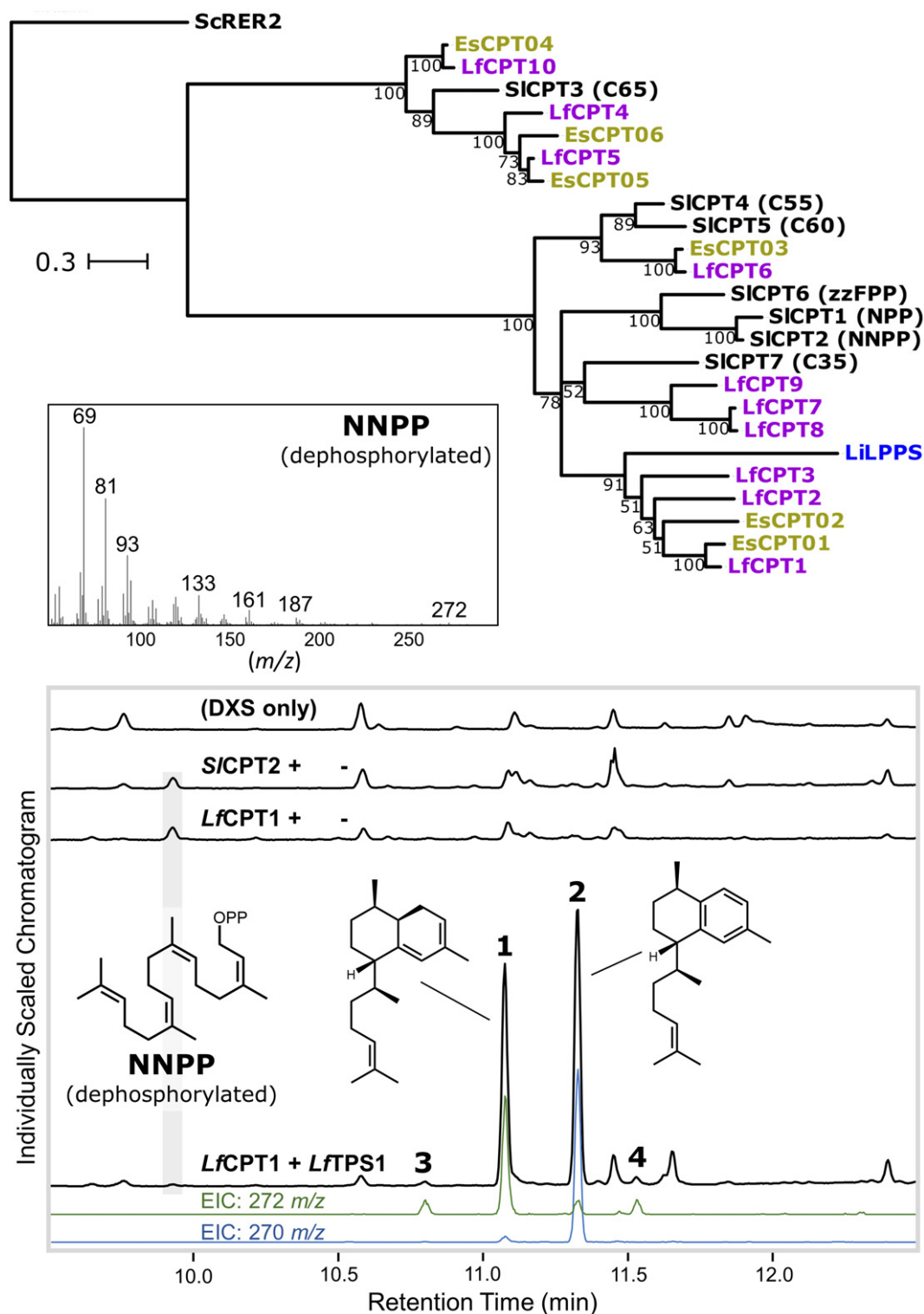
**Figure 4.** NNPP production by *Lf*CPT1. (Top) Maximum likelihood phylogenetic tree of candidate *cis*-prenyl transferases from *L. frutescens* (purple) and *E. serrulata* (yellow) transcriptome assemblies, and *S. lycopersicum* (black) with products given in parentheses. An unusual head-to-middle condensation *cis*-PT from *Lavandula* x *intermedia* (*Li*LPPS; blue) is included. The scale bar represents substitutions per site, and branch numbers represent percent support from 1000 bootstrap replicates; branches with less than 50% support have been collapsed. The *cis*-PT *Sc*RER2 from *Saccharomyces cerevisiae* is used as an outgroup. Reference names, accessions, and amino acid sequences are available in Dataset S1. (Bottom) GC-MS chromatograms for *N. benthamiana* assay of *Lf*CPT1 function, both alone and in combination with *Lf*TPS1, and mass spectrum of dephosphorylated NNPP found in the highlighted region (gray) of each sample except DXS control. Each assay has *Cf*DXS co-expressed in addition to those listed.

**Figure 5.** Leubethanol production by CYP71D616. GC-MS chromatograms of CYP71D616 assay in *N. benthamiana* and *L. frutescens* root extract, with mass spectrum of leubethanol (**5**) from heterologous expression of all three enzymes in the pathway. Total ion chromatograms are shown in black, and each assay has *Cf*DXS co-expressed in addition to those listed.
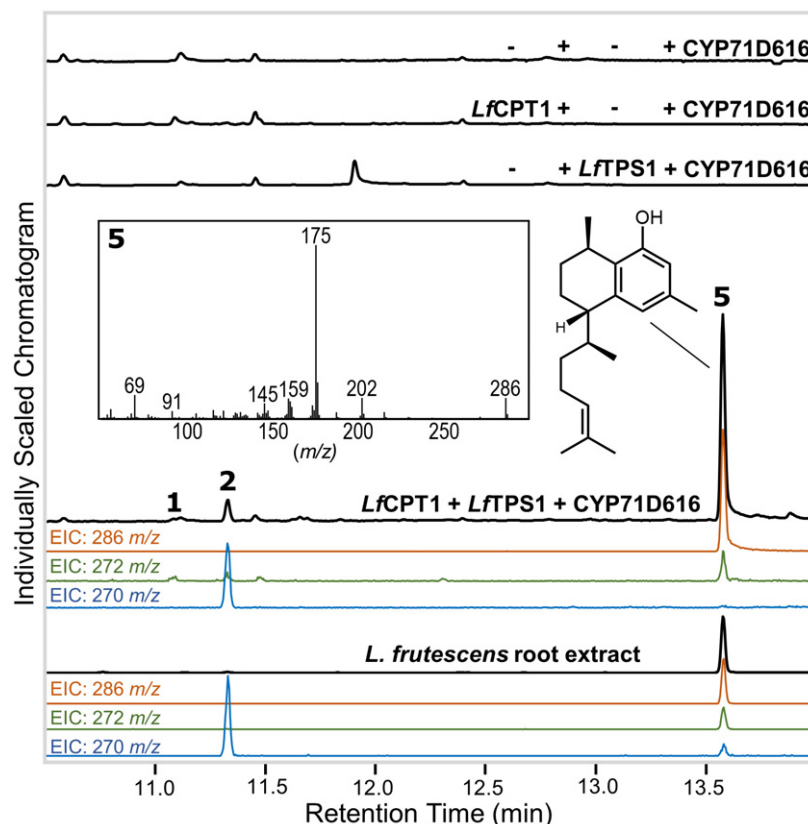


serrulatane indicates that the preferred substrate for CYP71D616 is dihydroserrulatene. The observed minor reduction in serrulatane is plausibly due to P450-mediated turnover of dihydroserrulatene preceding spontaneous aromatization. This is supported by the metabolomic data from root tissue extracts, which show an accumulation of serrulatane but no detectable quantities of dihydroserrulatene (Figures 5 and S1).

The interdependence of each enzyme in the pathway is demonstrated (Figures 4 and 5), showing that all three are necessary for leubethanol production when expressed in *N. benthamiana*. To determine whether the TPS activity is conserved in the *Eremophila* genus, we tested a synthetic homolog of *Lf*TPS1 (*Es*TPS1; 85% amino acid identity) from the *E. serrulata* transcriptome assembly. Replacing *Lf*TPS1 with *Es*TPS1 yields the same products in each combination (Figure S9), demonstrating orthology between the enzymes and conservation of this pathway in the serrulatane-rich *Eremophila* genus.

## DISCUSSION

Through comparative transcriptomics between tissue types and genera, we have identified three enzymes responsible for the biosynthesis of the serrulatane diterpenoid leubethanol in *L. frutescens*. The stereochemistry at all three chiral centers in dihydroserrulatene matches that of every serrulatane diterpenoid identified from the Scrophulariaceae family wherever the stereocenter is retained in the final diterpenoid product (see examples in Figure 1). This, together with the conserved function between *Lf*TPS1 and *Es*TPS1, suggests that dihydroserrulatene is in fact the common precursor to all serrulatanes. During the preparation of this manuscript, Gericke *et al.* (2020) reported a similar pathway to dihydroserrulatene involving a *cis*-PT and plastidial TPS-a in *Eremophila drummondii and Eremophila denticulata*, further supporting the conservation of this pathway. Nearly all of the serrulatane diterpenoids in Scrophulariaceae share a common hydroxylation (or derivative thereof) with leubethanol, suggesting that leubethanol itself is a common precursor. Given this commonality, the CYP71D616-catalyzed hydroxylation is likely the entry step between the diterpene backbone and diversification toward other anti-microbial serrulatane diterpenoids from other genera such as biflorin and microthecaline A.

This pathway is unusual in that it involves the all-*cis* prenyl diphosphate precursor NNPP rather than the common diterpene precursor GGPP. Prenyl diphosphate substrates are synthesized by members of either the *trans*- or *cis*-prenyl transferase families, typically in a head-to-tail condensation of the five-carbon molecules isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP)

(Zhou and Pichersky, 2020). These two enzyme families are distinct with no sequence (Shimizu *et al.*, 1998) or structural (Fujihashi *et al.*, 2001) homology. The evolution of members of the *cis*-PT family to make uncommon terpene precursors has been found in two other cases, with the series of NPP (*Sl*CPT1), (*Z,Z*)-FPP (*Sl*CPT6), and NNPP (*Sl*CPT2) in *S. lycopersicum* (Solanaceae) (Akhtar *et al.*, 2013), and lavandulyl diphosphate (head-to-middle condensation catalyzed by *Li*LPPS) in *Lavandula* x *intermedia* (Lamiaceae) (Demissie *et al.*, 2013). *Lf*CPT1, *Li*LPPS, and the *S. lycopersicum* short-chain *cis*-PT are phylogenetically closely related when compared to the overall characterized *cis*-PT family in *S. lycopersicum* (Figure 4). This may indicate a shared common ancestry of the short-chain *cis*-PTs in Solanaceae, Lamiaceae, and Scrophulariaceae. Scrophulariaceae diverged from Solanaceae between 75 to 88 million years ago (MYA) and from Lamiaceae between 44 and 67 MYA, based on molecular time estimates (timetree.org (Kumar *et al.*, 2017) and 32 references therein, accessed February 12, 2020), which is consistent with the divergence pattern of the short-chain *cis*-PTs (Figure 4): *Lf*CPT1 appears to be more closely related to *Li*LPPS (Lamiaceae) than any of the *Solanum cis*-PTs, despite being closer to the *Solanum* enzymes in product profile. Additionally, it has been suggested that the shorter product length of the *S. lycopersicum cis*-PTs may be due in part to a shortened alpha helix not present in the long-chain *cis*-PTs from this species (Akhtar *et al.*, 2013). This is not present in either *Li*LPPS or *Lf*CPT1 based on homology modeling (Figure S10) and a sequence alignment (Figure S11), suggesting that the evolution towards smaller precursors is independent and follows different trajectories from an ancestral sequence.

In addition to finding a similar pathway to dihydroserrulatene, Gericke *et al.* (2020) identified TPSs which make the cembrane and viscidane backbones in *Eremophila lucida*, and showed that these exclusively use NNPP over GGPP as well. To identify where the TPSs and *cis*-PTs from these three other *Eremophila* species (*E. denticulata, E. drummondii,* and *E. lucida*) lie relative to our candidates, we generated phylogenetic trees including each candidate identified from these species and our sequences (TPS: Figure S12; *cis*-PT: Figure S13). Each other *Eremophila* NNPP synthase is a direct ortholog of *Lf*CPT1, while *Lf*CPT2 and *Lf*CPT3 have no orthologs in any of these *Eremophila* species (Figure S13). Interestingly, a (*Z,Z*)-FPP synthase (*Ed*CPT2) was found; however, a TPS in *Eremophila* which converts (*Z,Z*)-FPP has yet to be identified (Gericke *et al.*, 2020). The cembratrienol synthase (*El*TPS31) is a member of the TPS-b subfamily, commonly involved in monoterpene synthesis, and *L. frutescens* does not have an ortholog. The hydroxyviscidane synthase *El*TPS3 lines up closely with *Lf*TPS2 and another enzyme from *E. denticulata* (*Edt*TPS5); however, neither of these candidates were found to have this same function. Interestingly, more TPS-a candidates which are putatively targeted to the plastid, but do not convert GGPP or NNPP, are present in these three *Eremophila* species. The function of *Lf*TPS2 and these other plastidial TPS-a enzymes remains to be seen, and may suggest that other precursors that were not taken into account in either study may be present in the plastids of these plants.

The identification of a short-chain *cis*-PT in Scrophulariaceae clarifies the likely origin of other diterpene backbones present in the *Eremophila* genus. Acyclic and bisabolane type diterpenoids identified in this genus contain internal alkenes in *cis* configuration (Figure 4). As serrulatanes and viscidanes have now both been shown to be derived from NNPP, it is likely that the decipiane, cycloserrulatane, and cedrane backbones (Ghisalberti, 1993) are derived from NNPP as well. The backbones for decipianes and cycloserrulatanes resemble a tricyclic serrulatane backbone, and the cedrane backbone resembles a tricyclic viscidane backbone. Beyond Scrophulariaceae, there are hundreds of other diterpene backbones with unknown biosynthetic routes. In Lamiaceae alone there are at least 200 (Johnson *et al.*, 2019a), and in *Salvia sclarea* (Lamiaceae), two previously reported diterpenoids salviatriene A and B (Laville *et al.*, 2012) resemble a cycloserrulatane and tricyclic viscidane, respectively. Given the independent emergence of *cis*-PTs which yield NNPP in different plant families, it may be that some of these unknown diterpenoid pathways involve NNPP as well.

Numerous diterpene backbones that differ from the more common labdane structure have been shown to be formed by enzymes in the TPS-a subfamily, which is mostly comprised of cytosolic sesquiterpene synthases. *Lf*TPS1 provides another example of a compartment and substrate-switching TPS from this subfamily, but differs from these previous examples in that it does not convert GGPP. In contrast to earlier work in *P. vulgaris* (Lamiaceae), where the enzyme *Pv*HVS showed acceptance of both GGPP and the presumed non-native NNPP (Johnson *et al.*, 2019b), *Lf*TPS1 showed a high specificity towards NNPP. *Pv*TPS5 and *Pv*TPS2 (both TPS-a) could also convert NNPP to a diterpene product in addition to their native functions as sesquiterpene and diterpene synthases, respectively (Johnson *et al.*, 2019b). This could plausibly arise from negative selection against GGPP, as both substrates are available in *L. frutescens* and presumably only GGPP is available in *P. vulgaris*. The presence of competing substrates in *L. frutescens* may introduce a strong selective pressure for specificity (Tawfik, 2014), while the absence of NNPP in *P. vulgaris* means that no such selective pressure exists. Such specificity can also be seen in *Solanum* where these all-*cis* substrates are present, where PHS1 (Schilmiller *et al.*, 2009), SBS (Sallaud *et al.*, 2009), and *Sl*TPS21 (Matsuba *et al.*, 2015) all showed high

specificity towards NPP, (*Z,Z*)-FPP, and NNPP, respectively, compared to their all-*trans* counterparts.

Even some class II diTPSs (TPS-c) have been shown to have promiscuous activities in converting NNPP into irregular labdane structures (Jia *et al.*, 2017). The substrate promiscuity of these TPSs suggests that the evolution of a prenyl transferase to afford an unusual terpene precursor may not require the co-evolution of a TPS, as the ability to convert a novel substrate may already be present in lineages where promiscuity was never selected against. Additionally, the occurrence of TPSs which natively convert *cis*-prenyl substrates is widespread throughout different TPS subfamilies. Examples have now been seen in the TPSe/f (*Solanum* species), TPS-b (*E. lucida*), and TPS-a (*L. frutescens* and three *Eremophila* species) subfamilies (Figure S12), showing that evolution towards specificity for these substrates has happened independently in vastly different lineages of TPSs. Taken together, the presence of uncommon substrates may be more widespread than generally assumed, and the search for biosynthetic routes to new terpene backbones should involve a consideration of other possible precursors beyond the all-*trans* substrates which are typical.

## MATERIALS AND METHODS

### Plant material, RNA isolation and cDNA synthesis, and metabolite analysis

*L. frutescens* plants were obtained from Stokes Tropicals (Homestead, FL, USA) and grown in a greenhouse under ambient photoperiod and 24°C day/17°C night temperatures. Total RNA from flower, leaf, and root tissues was extracted following methods described in Hamberger *et al.*, 2011 using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA). RNA extraction was followed by DNase I digestion using DNA-free™ DNA Removal Kit (Thermo Fisher Scientific). Total RNA was assessed for quantity and integrity by Qubit™ (Thermo Fisher Scientific) and RNA-nano assays (Agilent Bioanalyzer 2100), prior to whole transcriptome sequencing (Novogene, Sacramento, CA, USA). First-strand cDNA was synthesized from 2 µg of root total RNA using SuperScript III (Invitrogen). For GC-MS-based metabolomics, approximately 1 g of root, leaf, or flower tissue was extracted in 1 ml MTBE for 3 h and analyzed by GC-MS with the same method described below for analysis of enzyme assays.

### *L. frutescens* and *E. serrulata de novo* transcriptome assembly and analysis

RNA-seq data were obtained through tissue-specific RNA sequencing on an Illumina HiSeq 4000 for *L. frutescens* and the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra (ERX1321488)) for *E. serrulata* (Kracht *et al.*, 2017). Quality of sequencing data was checked with

FastQC (v0.11.4), and adapters were trimmed with Trimmomatic (v0.39; Bolger *et al.*, 2014). A transcriptome was assembled with Trinity (v2.8.4; Grabherr *et al.*, 2011), expression levels calculated with Salmon (v.0.11.2; Patro *et al.*, 2017), and open reading frames picked out with TransDecoder (v5.5.0; Haas *et al.*, 2013). A BLAST (v2.7.1+) search against reference databases of respective enzyme families (Dataset S1) was done to pick out candidates. Phylogenetic trees were made with Clustal Omega (v1.2.4; Sievers *et al.*, 2011) and RAxML (v8.0.0; Stamatakis, 2014) and visualized with Interactive Tree of Live (Letunic and Bork, 2019). Plastidial transit peptides were predicted between TargetP (v 1.1; Emanuelsson *et al.*, 2000) and sequence alignments with Clustal Omega (v1.2.4; Sievers *et al.*, 2011).

### Cloning and sources of genes used

Synthetic oligonucleotides, GenBank accession numbers, and sequences of each enzyme characterized in this study are listed in Dataset S1. Candidate enzymes were PCR-amplified from root cDNA, and coding sequences were cloned through In-Fusion cloning into the plant expression vector pEAQ-HT (Sainsbury *et al.*, 2009) for transient expression assays in *N. benthamiana*, or into pET-28b(+) for expression in *E. coli*. *Lf*TPS1 and *Lf*TPS2 were cloned into pET-28b(+) as N-terminal truncated constructs omitting the first 23 amino acid residues, removing their putative transit peptides. For *in vitro* assays, constructs for *Pv*TPS4, *Pv*TPS5, and *Pv*HVS(Δ43) in pET-28b(+) made in Johnson *et al.*, 2019b were used as positive controls. For *in vivo* *E. coli* assays, the same truncated *Lf*TPS constructs described above were used. TPS constructs were co-transformed with pIRS (Morrone *et al.*, 2010) and pNN (Johnson *et al.*, 2019b).

For all assays in *N. benthamiana*, full-length candidates were cloned into pEAQ-HT. For cytosolic tests, TPS candidates were co-expressed with *Euphorbia lathyris* HMGR and *Methanothermobacter thermautotrophicus* GGPPS (Sadre *et al.*, 2019) in the pEarlygate vector (Early *et al.*, 2006). As a positive control for cytosolic tests, an N-terminal truncated construct of *Pv*HVS (*Pv*HVS(Δ43)) was cloned into pEAQ-HT in this study. For plastidial tests, each candidate was co-expressed with *Coleus forskohlii* DXS (Andersen-Ranberg *et al.*, 2016) in pEarlygate. TPS candidate tests involved either co-expression of *C. forskohlii* GGPPS (Andersen-Ranberg *et al.*, 2016) in pEarlygate or *S. lycopersicum* CPT2 in pEAQ-HT (Johnson *et al.*, 2018b), with a full-length construct of *Pv*HVS in pEAQ-HT as a positive control (Johnson *et al.*, 2018b).

### *In vitro* assays

TPS expression and purification were carried out as described in Johnson *et al.*, 2019b. *Lf*TPS1 and *Lf*TPS2 constructs in pET-28b(+) were transformed into the *E. coli* C41

OverExpress strain. Primary cultures (5 ml LB plus 50 μg/ml kanamycin) were grown overnight 37°C, and 1 ml was used to inoculate a bulk culture (100 ml TB plus 50 μg/ml kanamycin). This culture was grown to an $OD_{600}$ of 0.6 at 37°C, and expression was induced with 0.2 mM IPTG. Expression was carried out overnight at 17°C, and cells were collected by centrifugation and resuspended in Buffer A (20 mM HEPES, pH 7.2, 25 mM imidazole, 500 mM NaCl, 5% [v/v] glycerol) plus 10 μl/ml protease inhibitor cocktail (Sigma) and 0.1 mg/ml lysozyme (VWR). Cells were lysed by sonication and centrifuged at 11,000 ×**g** for 30 min. Supernatants were loaded onto Ni-NTA columns (His Spin-Trap; GE Healthcare) pre-equilibrated with Buffer A, washed with two column volumes of Buffer A, and protein was eluted with Buffer B (Buffer A with 350 mM imidazole). Samples were de-salted with a PD MidiTrap G-25 column (GE Healthcare) pre-equilibrated with Buffer C (20 mM HEPES, pH 7.2, 1 mM $MgCl_2$, 350 mM NaCl, and 5% [v/v] glycerol). Purified enzymes were frozen in liquid nitrogen and stored at −80°C prior to in vitro assays.

*In vitro* assays were carried out with 1 μM enzyme and 30 μM substrate (GPP, FPP, or GGPP; Cayman Chemical) in 750 μl Buffer D (50 mM HEPES, pH 7.2, 7.5 mx $MgCl_2$, and 5% [v/v] glycerol), with 500 μl hexane overlay. Reactions were carried out for 16 h at 30°C, vortexed to extract products, and centrifuged to re-separate the aqueous and organic layers. The organic layer was directly removed for GC-MS analysis.

### Transient expression in *N. benthamiana*

Transient expression assays in *N. benthamiana* were carried out as described earlier (Johnson *et al.*, 2019a). *N. benthamiana* plants were grown for 5 weeks in a controlled growth room under 16 h light (24°C) and 8 h dark (17°C) cycle before infiltration. Constructs of candidates in pEAQ and others used for co-expression were separately transformed into *Agrobacterium tumefaciens* strain LBA4404. Cultures were grown overnight at 30°C in 10 ml LB plus 50 μg/ml kanamycin and 50 μg/ml rifampicin, collected by centrifugation, and washed with 10 ml water twice. Cells were resuspended and diluted to an $OD_{600}$ of 1.0 in water plus 200 μM acetosyringone and incubated at 30°C for 2–3 h. Separate cultures were mixed in a 1:1 ratio for each combination of enzyme tested (e.g., for leubethanol production, equal volumes of cultures were mixed harboring *Cf*DXS, *Lf*CPT1, *Lf*TPS1, and CYP71D616). Mixed cultures were infiltrated with a syringe into the abaxial side of *N. benthamiana* leaves, and plants were returned to the controlled growth room for 5 days. Approximately 200 mg fresh weight from infiltrated leaves was extracted with 1 ml hexane overnight at 18°C, plant material was collected by centrifugation, and the organic phase was removed for GC-MS analysis.

### *E. coli in vivo* assays

For *in vivo E. coli* assays, an engineered *E. coli* system (Cyr *et al.*, 2007) was used. *Lf*TPS1(Δ23) and *Lf*TPS2(Δ23) were co-transformed with pIRS and pNN and grown overnight at 37°C in 5 ml LB plus 25 μg/ml kanamycin, 17 μg/ml chloramphenicol, and 25 μg/ml streptomycin. A culture of 10 ml TB including the same antibiotics (same concentrations) was inoculated with 100 μl of the overnight culture and grown to an $OD_{600}$ of 0.6 at 37°C. The incubation temperature was lowered to 16°C for 1 h, expression was induced with 0.5 mM IPTG, and cultures were supplemented with 1 mM $MgCl_2$ and 40 mM pyruvate. Cultures were incubated at 16°C for an additional 60 h before extraction with an equal volume of hexane and 2% (v/v) EtOH. The organic phase was separated by centrifugation and analyzed by GC-MS.

### Dihydroserrulatene production scale-up and NMR

To generate enough of the major *Lf*TPS1 product (dihydroserrulatene) for NMR analysis, production in the *E. coli* system was carried out as detailed above, scaled up to 1 L. Following extraction, the organic layer was separated by centrifugation, concentrated under $N_2$ gas, and analyzed by GC-MS to confirm the presence of the *Lf*TPS1 product. This product was purified by silica gel flash column chromatography with a mobile phase of 10% ethyl acetate in hexane. NMR spectra were measured on an Agilent DirectDrive2 500 MHz spectrometer using $CDCl_3$ as the solvent. $CDCl_3$ peaks were referenced to 7.26 and 77.00 ppm for $^1H$ and $^{13}C$ spectra, respectively.

### GC-MS

All GC-MS analyses were performed on an Agilent 7890A GC with an Agilent VF-5 ms column (30 m × 250 μm × 0.25 μm, with 10 m EZ-Guard) and an Agilent 5975C detector. The inlet was set to 250°C splitless injection of 1 μl and He carrier gas (1 ml/min), and the detector was activated following a 3 min solvent delay. All assays and tissue analysis, with the exception of *in vitro* assays against GPP, used the following method: temperature ramp start 40°C, hold 1 min, 40°C/min to 200°C, hold 4.5 min, 20°C/min to 240°C, 10°C/min to 280°C, 40°C/min to 320°C, and hold 5 min (3 min hold for *in vitro* assays). For *in vitro* assays against GPP, the following method was used: temperature ramp start 40°C, 10°C/min to 180°C, 40°C/min to 320°C, hold 3 min.

### Homology modeling

Homology models for *Lf*CPT1 (Figure S10) were generated using I-TASSER (v. 5.1; Yang *et al.*, 2015) with either *Solanum habrochaites* (Z-Z)-FPPS (PDB ID: 5HXN; Chan *et al.*, 2017) or *Li*LPPS (PDB ID: 5HC6; Liu *et al.*, 2016) as the template structure. Figures were generated in PyMOL (v2.3).

## AUTHOR CONTRIBUTIONS

GM and BH: conceptualization; SJ: resources; GM, DM, and SJ: data curation; GM, EL, and WWB: formal analysis and validation; GM: methodology; GM: writing original draft; GM, EL, and BH: writing, revision, and editing; BH: supervision; BH: funding acquisition.

## CONFLICT OF INTEREST

BH, SJ, WWB, GM, EL, and DM have filed a provisional patent application (US Patent Application Serial No.: 62/986,286) describing the method to produce leubethanol and related diterpenes.

## DATA AVAILABILITY STATEMENT

RNA-seq data for *L. frutescens* have been submitted to the NCBI SRA under the accession numbers SRX8371655 (root) and SRX8371656 (flower). GenBank accession numbers for nucleotide sequences of all enzymes tested in this study are as follows: LfTPS1: MT136608; LfTPS2: MT136609; LfCPT1: MT136610; LfCPT2: MT136611; LfCPT3: MT136612; CYP706G22: MT136613; CYP76A112: MT136614; CYP736A294: MT136615; CYP736A295: MT136616; CYP71D615: MT136617; CYP71D616: MT136618 EsTPS1: MT136619. The following additional *L. frutescens* class I TPS candidates were cloned but not characterized: LfTPS3: MT521506; LfTPS5: MT521507; LfTPS6: MT521505; LfTPS7: MT521508; LfTPS8a: MT521515; LfTPS8b: MT521516; LfTPS9: MT521509; LfTPS10: MT521511; LfTPS11: MT521510; LfTPS12a: MT521512; LfTPS12b: MT521513; LfTPS13: MT521514.

## SUPPORTING INFORMATION

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

**Figure S1.** GC-MS analysis of *L. frutescens* flower, leaf, and root tissue extracts.
**Figure S2.** Sequence alignment of TPS-a enzymes shown in Figure 2 in the main text.
**Figure S3.** Initial screening of *Lf*TPS1 and *Lf*TPS2 against all-*trans* prenyl substrates.
**Figure S4.** Initial screening of *L. frutescens* TPS candidates against NNPP.
**Figure S5.** ¹H, ¹³C, HSQC, H2BC, HMBC, COSY, and NOESY NMR spectra for dihydroserrulatene.
**Figure S6.** Initial screening of *L. frutescens cis*-PT candidates.
**Figure S7.** Phylogenetic tree of candidate cytochrome P450s in the CYP71 clan.
**Figure S8.** Initial screening of *L. frutescens* cytochrome P450 candidates.
**Figure S9.** Replacement of *Lf*TPS1 with *Es*TPS1 in the biosynthetic pathway for leubethanol.
**Figure S10.** Homology models of *Lf*CPT1.
**Figure S11.** Sequence alignment of candidate and reference *cis*-PTs.
**Figure S12.** Phylogenetic tree of TPSs with sequences from three other *Eremophila* species.
**Figure S13.** Phylogenetic tree of *cis*-PTs with sequences from three other *Eremophila* species.
**Scheme S1.** Proposed mechanism for *Lf*TPS1 conversion of NNPP to dihydroserrulatene.
**Table S1.** ¹³C and ¹H chemical shifts for NMR spectra of dihydroserrulatene.
**Dataset S1.** Sequences for primers, reference enzymes, and candidate enzymes.

## OPEN RESEARCH BADGE

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://doi.org/10.17605/OSF.IO/58HJT.

## REFERENCES

**Akhtar, T.A., Matsuba, Y., Schauvinhold, I., Yu, G., Lees, H.A., Klein, S.E. and Pichersky, E.** (2013) The tomato cis-prenyltransferase gene family. *Plant J.* **73**, 640–652.
**Algreiby, A.A., Hammer, K.A., Durmic, Z., Vercoe, P. and Flematti, G.R.** (2018) Antibacterial compounds from the Australian native plant *Eremophila glabra*. *Fitoterapia*, **126**, 45–52.
**Aminimoghadamfarouj, N. and Nematollahi, A.** (2017) Structure elucidation and botanical characterization of diterpenes from a specific type of bee glue. *Molecules*, **22**.
**Anakok, O.F., Ndi, C.P., Barton, M.D., Griesser, H.J. and Semple, S.J.** (2012) Antibacterial spectrum and cytotoxic activities of serrulatane compounds from the Australian medicinal plant *Eremophila neglecta*. *J. Appl. Microbiol.* **112**, 197–204.
**Andersen-Ranberg, J., Kongstad, K.T., Nielsen, M.T. et al.** (2016) Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angew. Chem. Int. Ed.* **55**, 2142–2146.
**Barnes, E.C., Kavanagh, A.M., Ramu, S., Blaskovich, M.A., Cooper, M.A. and Davis, R.A.** (2013) Antibacterial serrulatane diterpenes from the Australian native plant *Eremophila microtheca*. *Phytochemistry*, **93**, 162–169.
**Bathe, U. and Tissier, A.** (2019) Cytochrome P450 enzymes: a driving force of plant diterpene diversity. *Phytochemistry*, **161**, 149–162.
**Best, W.M. and Wege, D.** (1986) Intramolecular Diels-Alder Additions of Benzynes to Furans. Application to the Total Synthesis of Biflorin, and the Mansonone-E, I and F. *Aust. J. Chem.* **39**, 647–666.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Carvalho, A.A., da Costa, P.M., Da Silva Souza, L.G. *et al.* (2013) Inhibition of metastatic potential of B16–F10 melanoma cell line in vivo and in vitro by biflorin. *Life Sci.* **93**, 201–207.

Chan, Y., Haas, B.J., Yassour, M. *et al.* (2017) Crystal Structure and Potential Head-to-Middle Condensation Function of a Z. Z-Farnesyl Diphosphate Synthase. *ACS Omega*, **2**, 930–936.

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.

Cyr, A., Wilderman, P.R., Determan, M. and Peters, R.J. (2007) A Modular approach for facile biosynthesis of labdane-related diterpenes. *J. Am. Chem. Soc.* **129**, 6684–6685.

Degenhardt, J., Köllner, T.G. and Gershenzon, J. (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, **70**, 1621–1637.

Demissie, Z.A., Erland, L.A.E., Rheault, M.R. and Mahmoud, S.S. (2013) The biosynthetic origin of irregular monoterpenes in *Lavandula*: isolation and biochemical characterization of a novel cis-prenyl diphosphate synthase gene, lavandulyl diphosphate synthase. *J. Biol. Chem.* **288**, 6333–6341.

Durairaj, J., Di Girolamo, A., Bouwmeester, H.J., de Ridder, D., Beekwilder, J. and van Dijk, A.D.J. (2019) An analysis of characterized plant sesquiterpene synthases. *Phytochemistry*, **158**, 157–165.

Earley, K., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**(4), 616–629.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016.

Ennajdaoui, H., Vachon, G., Giacalone, C., Besse, I., Sallaud, C., Herzog, M. and Tissier, A. (2010) Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing cis-regions. *Plant Mol. Biol.* **73**, 673–685.

Fujihashi, M., Zhang, Y.-W., Higuchi, Y., Li, X.-Y., Koyama, T. and Miki, K. (2001) Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. *PNAS*, **98**, 4337–4342.

Gericke, O., Hansen, N.L., Pedersen, G.B. *et al.* (2020) Nerylneryl diphosphate is the precursor of serrulatane, viscidane and cembrane-type diterpenoids in *Eremophila* species. *BMC Plant Biol.* **20**, 91.

Ghisalberti, E.L. (1993) The phytochemistry of the myoporaceae. *Phytochemistry*, **35**, 7–33.

Grabherr, M.G., Haas, B.J., Yassour, M. *et al.* (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.*, **29**, 644–652.

Haas, B.J., Papanicolaou, A., Yassour, 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.

Hamberger, B. and Bak, S. (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**, 20120426.

Hamberger, B., Ohnishi, T., Hamberger, B., Séguin, A. and Bohlmann, J. (2011) Evolution of Diterpene Metabolism: Sitka Spruce CYP720B4 Catalyzes Multiple Oxidations in Resin Acid Biosynthesis of Conifer Defense against Insects. *Plant Physiol.* **157**, 1677–1695.

Hossain, M.A., Biva, I.J., Kidd, S.E. *et al.* (2019) Antifungal Activity in Compounds from the Australian Desert Plant *Eremophila alternifolia* with Potency Against Cryptococcus spp. *Antibiotics (Basel)*, **8**, 34.

Jia, M. and Peters, R.J. (2017) Cis or Trans with class II diterpene cyclases. *Org. Biomol. Chem.* **15**, 3158–3160.

Johnson, S.R. *et al* .A database-driven approach identifies additional diterpene synthase activities in the mint family (Lamiaceae). J. Biol. Chem., jbc.RA118.006025 (2019a).

Johnson, S.R., Bhat, W.W., Sadre, R., Miller, G.P., Garcia, A.S. and Hamberger, B. (2019b) Promiscuous terpene synthases from *Prunella vulgaris* highlight the importance of substrate and compartment switching in terpene synthase evolution. *New Phytol.* **223**, 323–335.

Karunanithi, P.S. and Zerbe, P. (2019) Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Front. Plant Sci.* **10**.

Kirby, J. Nishimoto, M., Park, J.G. *et al.* (2010) Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevisiae*. *Phytochemistry*, **71**, 1466–1473.

Kracht, O.N., Ammann, A.-C., Stockmann, J. *et al.* (2017) Transcriptome profiling of the Australian arid-land plant *Eremophila serrulata* (A.DC.) Druce (Scrophulariaceae) for the identification of monoterpene synthases. *Phytochemistry*, **136**, 15–22.

Kumar, R., Duffy, S., Avery, V.M., Carroll, A.R. and Davis, R.A. (2018) Microthecaline a, a quinoline serrulatane alkaloid from the roots of the Australian desert plant *Eremophila microtheca*. *J. Nat. Prod.* **81**, 1079–1083.

Kumar, R., Duffy, S., Avery, V.M. and Davis, R.A. (2017) Synthesis of antimalarial amide analogues based on the plant serrulatane diterpenoid 3,7,8-trihydroxyserrulat-14-en-19-oic acid. *Bioorg. Med. Chem. Lett.* **27**, 4091–4095.

Kumar, S., Stecher, G., Suleski, M. and Hedges, S.B. (2017) TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819.

Laville, R., Callari, R., Hamberger, B. *et al.* (2012) Amphilectane Diterpenes from *Salvia sclarea*: Biosynthetic Considerations. *J. Nat. Prod.*, **75**, 121–126.

Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.*, **47**, W256–W259.

Lu, J.M.H., Perkins, M.V. and Griesser, H.J. (2013) Total synthesis and structural confirmation of the antibacterial diterpene leubethanol. *Tetrahedron*, **69**, 6468–6473.

Liu, M., Chen, C.-C., Chen, L. *et al.* (2016) Structure and Function of a "Head-to-Middle" Prenyltransferase: Lavandulyl Diphosphate Synthase. *Angew. Chem. Int. Ed.*, **55**, 4721–4724.

Luo, D., Callari, R., Hamberger, B. *et al.* (2016) Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from mature seeds of *Euphorbia lathyris* L. *PNAS*, **113**, E5082–E5089.

Matsuba, Y., Zi, J., Jones, A.D., Peters, R.J. and Pichersky, E. (2015) Biosynthesis of the Diterpenoid Lycosantalonol via Nerylneryl Diphosphate in *Solanum lycopersicum*. *PLoS One*, **10**, e0119302

Mau, C.J. and West, C.A. (1994) Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. *PNAS*, **91**, 8497–8501.

Molina-Salinas, G.M., Pérez-López, A., Becerril-Montes, P., Salazar-Aranda, R., Said-Fernández, S. and Torres, N.W.d.. (2007) Evaluation of the flora of Northern Mexico for in vitro antimicrobial and antituberculosis activity. *J. Ethnopharmacol.* **109**, 435–441.

Molina-Salinas, G.M., Rivas-Galindo, V.M., Said-Fernández, S. *et al.* (2011) Stereochemical Analysis of Leubethanol, an Anti-TB-Active Serrulatane, from *Leucophyllum frutescens*. *J. Nat. Prod.* **74**, 1842–1850.

Mon, H.H., Christo, S.N., Ndi, C.P. *et al.* (2015) Serrulatane diterpenoid from *Eremophila neglecta* exhibits bacterial biofilm dispersion and inhibits release of pro-inflammatory cytokines from activated macrophages. *J. Nat. Prod.* **78**, 3031–3040.

Morrone, D., Lowry, L., Determan, M.K., Hershey, D.M., Xu, M. and Peters, R.J. (2010) Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl. Microbiol. Biotechnol.* **85**(6), 1893–1906.

Ndi, C.P., Semple, S.J., Griesser, H.J., Pyke, S.M. and Barton, M.D. (2007a) Antimicrobial compounds from the Australian desert plant *Eremophila neglecta*. *J. Nat. Prod.*, **70**, 1439–1443.

Ndi, C.P., Semple, S.J., Griesser, H.J., Pyke, S.M. and Barton, M.D. (2007b) Antimicrobial compounds from *Eremophila serrulata*. *Phytochemistry*, **68**, 2684–2690.

Pateraki, I. Andersen-Ranberg, J., Jensen, N.B. *et al.* (2017) Total biosynthesis of the cyclic AMP booster forskolin from Coleus forskohlii. *eLife*, **6**, e23001.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**, 417–419.

Penjarla, T.R., Kundarapu, M., Baquer, S.M. and Bhattacharya, A. (2019) Total synthesis of the plant alkaloid racemic microthecaline A: first example of a natural product bearing a tricyclic quinoline-serrulatane scaffold. *RSC Adv.* **9**, 23289–23294.

Peters, R.J. (2010) Two rings in them all: the labdane-related diterpenoids. *Nat. Prod. Rep.* **27**, 1521–1530.

**Sadre, R., Kuo, P., Chen, J., Yang, Y., Banerjee, A., Benning, C. and Hamberger, B.** (2019) Cytosolic lipid droplets as engineered organelles for production and accumulation of terpenoid biomaterials in leaves. *Nat. Commun.* **10**, 1–12.

**Sainsbury, F., Thuenemann, E.C. and Lomonossoff, G.P.** (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693.

**Sallaud, C., Rontein, D., Onillon, S. et al.** (2009) A Novel Pathway for Sesquiterpene Biosynthesis from Z, Z-Farnesyl Pyrophosphate in the Wild Tomato *Solanum habrochaite*s. *Plant Cell*, **21**, 301–317.

**Schilmiller, A.L., Schauvinhold, I., Larson, M. et al.** (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *PNAS*, **106**, 10865–10870.

**Sievers, F., Wilm, A., Dineen, D. et al.** (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.

**Shimizu, N., Koyama, T. and Ogura, K.** (1998) Molecular Cloning, Expression, and Purification of Undecaprenyl Diphosphate Synthase: NO SEQUENCE SIMILARITY BETWEEN *E* - AND *Z* -PRENYL DIPHOSPHATE SYNTHASES. *J. Biol. Chem.* **273**, 19476–19481.

**Stamatakis, A.** (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

**Tawfik, D.S.** (2014) Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* **21**, 73–80.

**Tenneti, S., Biswas, S., Cox, G.A., Mans, D.J., Lim, H.J., & RajanBabu, T.V.** (2018) Broadly applicable stereoselective syntheses of serrulatane, amphilectane diterpenes, and their diastereoisomeric congeners using asymmetric hydrovinylation for absolute stereochemical control. *J. Am. Chem. Soc.* **140**, 9868–9881.

**Vaughan, M.M., Wang, Q., Webster, F.X. et al.** (2013) Formation of the Unusual Semivolatile Diterpene Rhizathalene by the Arabidopsis Class I Terpene Synthase TPS08 in the Root Stele Is Involved in Defense against Belowground Herbivory. *Plant Cell*, **25**, 1108–1125.

**Wang, Q., Jia, M., Huh, J.-H., Muchlinski, A., Peters, R.J. and Tholl, D.** (2016) Identification of a dolabellane type diterpene synthase and other root-expressed diterpene synthases in *Arabidopsis*. *Front. Plant Sci.* **7**.

**Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y.** (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, **12**, 7–8.

**Yu, X., Su, F., Liu, C. et al.** (2016) Enantioselective total syntheses of various amphilectane and serrulatane diterpenoids via cope rearrangements. *J. Am. Chem. Soc.* **138**, 6261–6270.

**Zeng, T., Liu, Z., Liu, H., He, W., Tang, X., Xie, L. and Wu, R.** (2019) Exploring chemical and biological space of terpenoids. *J. Chem. Inf. Model.* **59**, 3667–3678.

**Zerbe, P., Hamberger, B., Yuen, M.M.S. et al.** (2013) Gene discovery of modular diterpene metabolism in nonmodel systems. *Plant Physiol.* **162**, 1073–1091.

**Zhou, F. and Pichersky, E.** (2020) More is better: the diversity of terpene metabolism in plants. *Curr. Opin. Plant Biol.*, **55**, 1–10.

**Zi, J., Matsuba, Y., Hong, Y.J., Jackson, A.J., Tantillo, D.J., Pichersky, E. and Peters, R.J.** (2014) Biosynthesis of Lycosantalonol, a *cis*-Prenyl Derived Diterpenoid. *J. Am. Chem. Soc.* **136**, 16951–16953.

**Zi, J. and Peters, R.J.** (2013) Characterization of CYP76AH4 clarifies phenolic diterpenoid biosynthesis in the Lamiaceae. *Org. Biomol. Chem.* **11**, 7650–7652.