UC Davis

UC Davis Previously Published Works

Title

Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution

Permalink

https://escholarship.org/uc/item/6sc0f87f

Journal

Proceedings of the National Academy of Sciences of the United States of America, 114(29)

ISSN

0027-8424

Authors

Huang, Ancheng C

Kautsar, Satria A

Hong, Young J

et al.

Publication Date

2017-07-18

DOI

10.1073/pnas.1705567114

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution

Ancheng C. Huang[a], Satria A. Kautsar[b], Young J. Hong[c], Marnix H. Medema[b], Andrew D. Bond[d], Dean J. Tantillo[c], and Anne Osbourn[a,1]

[a]Department of Metabolic Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom; [b]Bioinformatics Group, Wageningen University, 6708 PB Wageningen, The Netherlands; [c]Department of Chemistry, University of California, Davis, CA 95616; and [d]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

Edited by Rodney B. Croteau, Washington State University, Pullman, WA, and approved June 7, 2017 (received for review April 4, 2017)

Sesterterpenoids are a rare terpene class harboring untapped chemodiversity and bioactivities. Their structural diversity originates primarily from the scaffold-generating sesterterpene synthases (STSs). In fungi, all six known STSs are bifunctional, containing C-terminal *trans*-prenyltransferase (PT) and N-terminal terpene synthase (TPS) domains. In plants, two colocalized PT and TPS gene pairs from *Arabidopsis thaliana* were recently reported to synthesize sesterterpenes. However, the landscape of PT and TPS genes in plant genomes is unclear. Here, using a customized algorithm for systematically searching plant genomes, we reveal a suite of physically colocalized pairs of PT and TPS genes for the biosynthesis of a large sesterterpene repertoire in the wider Brassicaceae. Transient expression of seven TPSs from *A. thaliana*, *Capsella rubella*, and *Brassica oleracea* in *Nicotiana benthamiana* yielded fungal-type sesterterpenes with tri-, tetra-, and pentacyclic scaffolds, and notably (−)-*ent*-quiannulatene, an enantiomer of the fungal metabolite (+)-quiannulatene. Protein and structural modeling analysis identified an amino acid site implicated in structural diversification. Mutation of this site in one STS (*AtTPS19*) resulted in premature termination of carbocation intermediates and accumulation of bi-, tri-, and tetracyclic sesterterpenes, revealing the cyclization path for the pentacyclic sesterterpene (−)-retigeranin B. These structural and mechanistic insights, together with phylogenetic analysis, suggest convergent evolution of plant and fungal STSs, and also indicate that the colocalized PT–TPS gene pairs in the Brassicaceae may have originated from a common ancestral gene pair present before speciation. Our findings further provide opportunities for rapid discovery and production of sesterterpenes through metabolic and protein engineering.

sesterterpene biosynthesis | plant natural products | cyclization mechanism | convergent evolution | Brassicaceae

Sesterterpenoids are a largely unexplored class of terpenes with only around 1,000 members isolated from nature so far, representing a mere <2% of the reported terpene family members (>70,000). These compounds are structurally diverse and have a wide spectrum of biological activities, ranging from antiinflammatory, anticancer, cytotoxic, and antimicrobial bioactivities to phytotoxicity and plant defense (1). Recent work on sesterterpenoids has suggested that this terpene class could be an important new source of anticancer drugs (2, 3). The majority of sesterterpenoids characterized to date are from marine sponges and terrestrial fungi, with only ∼60–70 being of plant origin. Many of these plant sesterterpenes were isolated from the mint family (Lamiaceae) and have been implicated in plant defense (4–7). Although large transcriptome datasets for the mint family have been generated (8), very limited genome sequences for the Lamiaceae are currently available (9, 10).

Like other classes of terpenes, the structural diversity of sesterterpenes largely originates from the very first scaffold-generating step, which in this case is catalyzed by sesterterpene synthases (STSs), a class of terpene synthase (TPS). In fungi, six STSs have recently been identified and shown to synthesize ophiobolin F (in *Aspergillus clavatus*) (11), sesterfisherol (in *Neosartorya fischeri*) (12), and stellata-2,6,19-triene, (2*E*)-α-cericerene, quiannulatene, and astellifadiene (in *Emericella variecolor*) (13–16) (Fig. 1*A*). These fungal enzymes are all bifunctional, containing a C-terminal *trans*-prenyltransferase (PT) and an N-terminal TPSs domain. The PT domain synthesizes the universal C25 sesterterpene precursor geranylfarnesyl diphosphate (GFPP, **1**) from dimethyl allyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP), which is then cyclized by the TPS domain to form diverse scaffolds. Some fungal diterpene synthases (DTSs), such as *Phomopsis amygdali* fusicoccadiene synthase (17, 18) and phomopsene synthase (19), are also bifunctional. Fusion of PT and TPS activities has been proposed to contribute to more efficient production of bioactive terpenes from precursor supply (18, 20). In most cases, however, the condensation and cyclization steps for diterpene biosynthesis in fungi are carried out by individual enzymes encoded by genes that are sometimes colocalized within fungal biosynthetic gene clusters (21).

In plants, two TPSs from *Arabidopsis thaliana* were recently reported to synthesize the sesterterpenes arathanatriene (**2**) and retigeranin B (**3**) (22). Intriguingly, these two TPSs are colocalized with PTs that synthesize the precursor substrate GFPP. To our knowledge, bifunctional TPSs that contain both PT and TPS domains have not been reported from any plant species.

Significance

Sesterterpenoids are a rare terpene class with largely unexplored chemical diversity and bioactivities, representing a potential new drug source. Sesterterpene synthases catalyze the first committed biosynthesis step and shape sesterterpene structural diversity. Here, using a customized algorithm for systematically surveying plant genomes, we unearth a sesterterpene synthase gene repertoire in the Brassicaceae that synthesizes fungal-type sesterterpenes with diverse scaffolds, thus fueling the drug-discovery pipeline. Mechanistic studies shed light on the origin of structural diversification and offer insights for expanding nature's diversity by protein engineering. Our work also supports the concept of convergent evolution in natural product biosynthesis.

Author contributions: A.C.H., S.A.K., M.H.M., D.J.T., and A.O. designed research; A.C.H., S.A.K., Y.J.H., and A.D.B. performed research; A.C.H. and S.A.K. contributed new reagents/analytic tools; A.C.H., S.A.K., Y.J.H., M.H.M., A.D.B., D.J.T., and A.O. analyzed data; and A.C.H., S.A.K., Y.J.H., M.H.M., A.D.B., D.J.T., and A.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The crystallographic data have been deposited at the Cambridge Crystallographic Data Centre (Accession nos. CCDC 1539191–1539193 for compounds **7–9** and CCDC 1535306–1535309 for compounds **10–13**).

[1]To whom correspondence should be addressed. Email: anne.osbourn@jic.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1705567114/-/DCSupplemental.

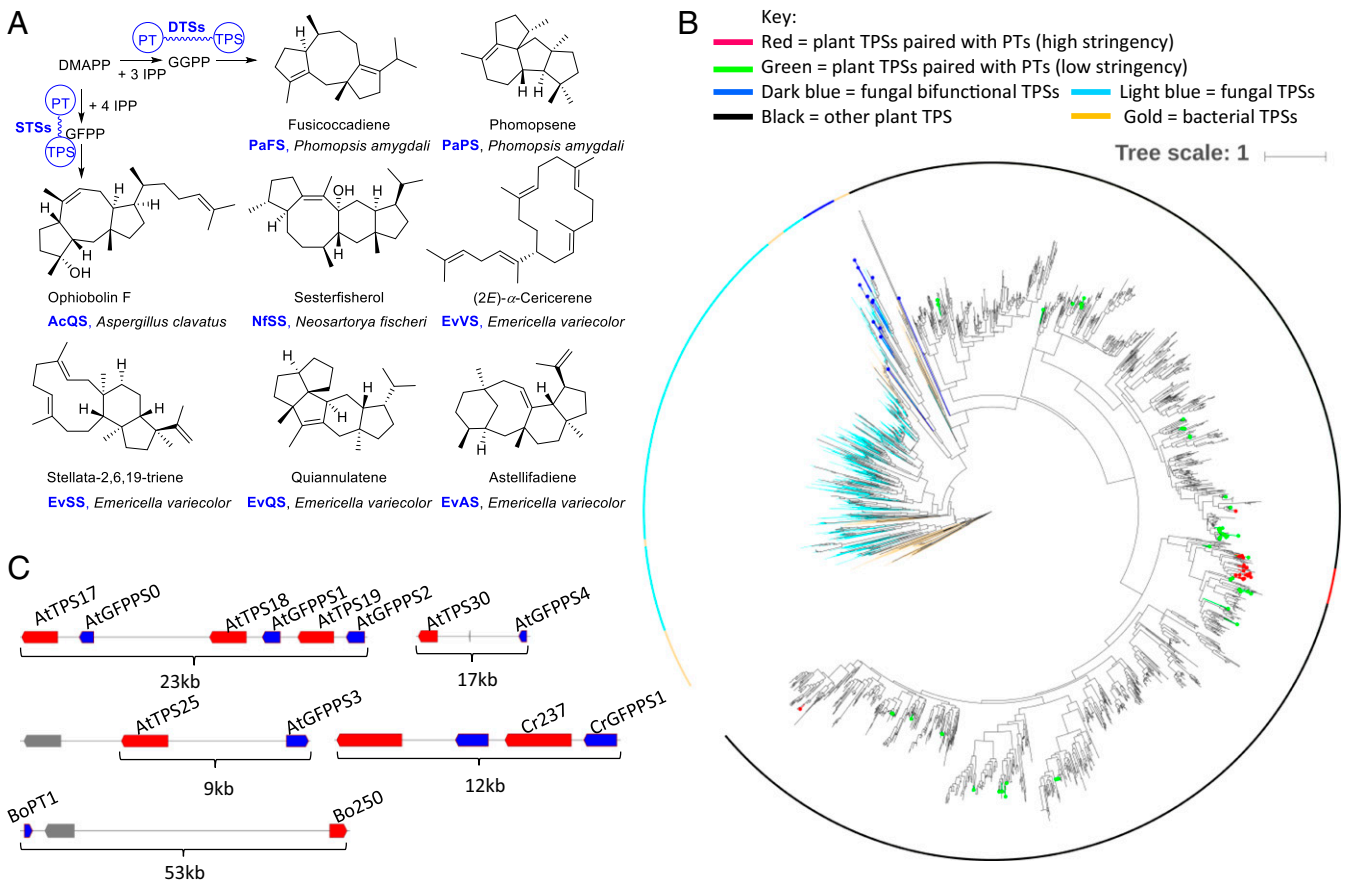


Fig. 1. Identification of candidate STS genes from plant genomes. (*A*) Chemical structures of diterpenes and sesterterpenes synthesized by fungal bifunctional DTSs and STSs. (*B*) Phylogenetic tree constructed using TPSs from 55 plant genomes together with fungal and bacterial TPSs. The TPSs for which the genes were found to be colocalized with PT genes with high stringency are shown in red, and those with low stringency in green. Bifunctional fungal TPSs are shown in dark blue. (*C*) A graphic representation of PT–TPS gene pairs from *A. thaliana*, *C. rubella*, and *B. oleracea* selected for functional validation in this work.

This raises questions regarding the landscape of PT and TPS genes and their related functions in plant genomes. We have previously developed algorithms for mining plant genomes for genes encoding natural product biosynthetic pathways (23, 24). Here, we carried out a systematic search of 55 sequenced plant genomes using a customized version of the plantiSMASH genome mining algorithm (24) to identify genes predicted to encode PT or TPS domains. Although we did not find evidence for bifunctional genes encoding both PT and TPS domains, we identified a pool of colocalized pairs of PT and TPS genes using this algorithm. Phylogenetic analysis of all TPSs from the 55 plant genomes that were investigated identified a single clade containing those TPS genes that were physically clustered with PT genes, greatly expanding on previous observations (25). Transient expression of selected TPS genes from this group from *A. thaliana* (thalecress), *Capsella rubella* (pink shepherd's purse), and *Brassica oleracea* in heterologous host *Nicotiana benthamiana* revealed five additional TPSs that make a variety of previously unknown fungal-type sesterterpenes with diverse scaffolds, including 11/6/5 tricyclic (−)-caprutriene (**7**), 5/12/5 tricyclic (−)-variculatriene A (**5**), 5/8/6/5 tetracyclic (−)-aleurodiscalene A (**9**), 5/5/5/6/5 pentacyclic (−)-*ent*-quiannulatene (**4**) and (+)-boleracene (**8**), and 5/4/7/6/5 pentacyclic (+)-astellatene (**6**) (Fig. 2*B*, *SI Appendix*, Tables S1–S8, and Dataset S1). Among these scaffolds, (−)-*ent*-quiannulatene (**4**) is an enantiomer of the fungal metabolite (+)-quiannulatene. Homology modeling and sequence comparisons implicated a key amino acid as being likely to be important for scaffold diversification. Site-directed mutagenesis of this residue further expanded scaffold diversity and, coupled with quantum chemical calculations, shed light on the cyclization mechanisms leading to the formation of the pentacyclic sesterterpenes (−)-retigeranin B (**3**) and (+)-astellatene (**6**). The carbocation cascade sequences following protonation of the GFPP substrate in the active site of plant STSs likely mirror those of the distantly related fungal bifunctional STSs. However, phylogenetic analysis suggests the independent evolution of fungal and plant STSs, indicating that the plant and fungal enzymes have arisen by convergent evolution.

Although mining plant genomes and transcriptomes for individual PTs and TPSs has proven to be effective for uncovering terpene diversity (26–28), to our knowledge systematic genome mining for colocalized PT and TPS genes has not previously been attempted. Collectively, our work sheds light on the genomic organization of plant PT and STS genes and demonstrates an effective pipeline for plant sesterterpene discovery, from finding genes to production by engineering in the heterologous host *N. benthamiana* and further manipulation by protein engineering.

Results

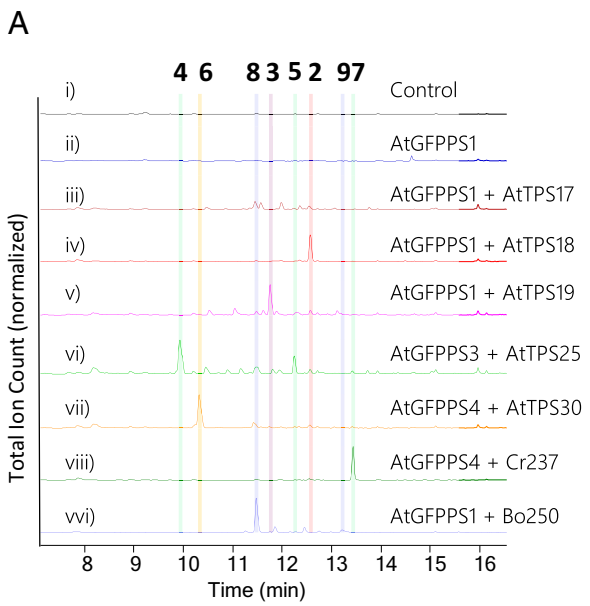
Landscape of PT and TPS Genes in Plant Genomes and Identification of a Phylogenetic Clade of Candidate Plant STSs Using a Customized Genome Mining Algorithm. To search for genes with predicted PT and TPS domains across 55 sequenced high-quality plant genomes (Dataset S2), we used the Polyprenyl_synt (PT) and Terpene_synth C (TPS) models from the Pfam database (29) in a version of plantiSMASH (24) with gene cluster detection logic customized for this specific purpose. The Terpene_synth_C pHMM (pfam.xfam.org/family/Terpene_synth_C#tabview=tab7) was seeded from plant, fungal, and bacterial TPSs and so is applicable across plants

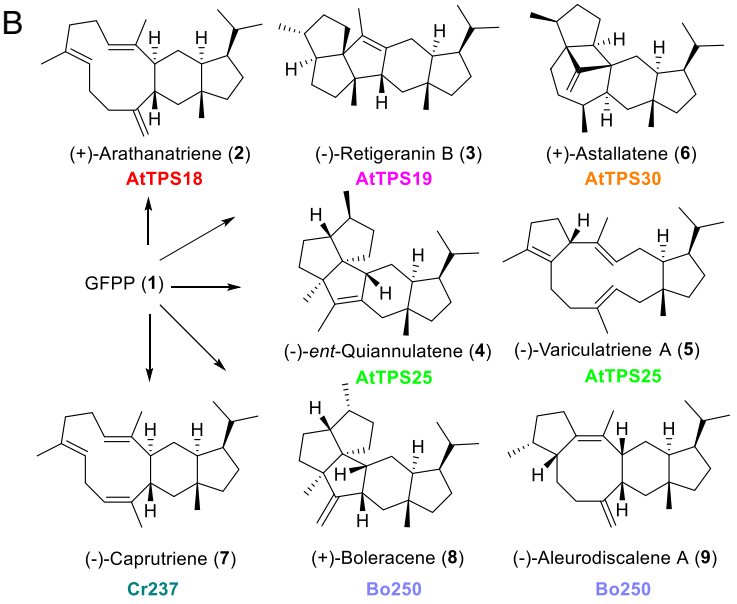
and microbes. Although we could not identify any candidate bifunctional TPS genes, we were able to identify 21 colocalized PT–TPS gene pairs using our default (high-stringency) parameters [cut-off 25 (±5) kb; deviation calculated using gene density/average intergenic distance for the neighboring 10 genes as a multiplier] (Dataset S2). An additional 75 candidate PT–TPS gene pairs were identified at lower stringency (allowing for distances between the PT and TPS domains that were up to five times larger) (Dataset S2). We next constructed a phylogenetic tree (Fig. 1*B*) using 2,846 TPS sequences extracted from the 55 plant genomes (Dataset S2), together with 724 fungal TPSs (including the known bifunctional TPSs). A total of 278 bacterial TPSs was included as an outgroup. This phylogenetic analysis enabled us to identify a single TPS clade in which the TPS genes that were colocalized with PT genes as identified using the high-stringency cut-off were heavily represented (Fig. 1*B*). Nine of the PTs in these PT–TPS gene pairs have indeed been previously shown to be functional plant GFPPSs (25, 30), and two of the TPSs (*AtTPS18* and *AtTPS19* from *A. thaliana*) were shown to biosynthesize the sesterterpenes (+)-arathanatriene and (−)-retigeranin B (22). However, two other TPSs from these pairs [*AtTPS25* (*At3g29410*) and *AtTPS30* (*At3g32030*) from *A. thaliana*] have recently been reported to have diterpene or sesquiterpene activities when expressed in *Escherichia coli* (31). A total of 18 TPS genes, including *AtTPS25* and *AtTPS30*, that are colocalized with PT genes were grouped together in this clade (marked in red, Fig. 1*B* and *SI Appendix*, Fig. S1*A*), greatly outnumbering previous observations (25). Two outlier TPS genes (one from *Vigna angularis* and one from *Brassica napus*, marked as red dots in Fig. 1*B*) from the 21 PT–TPS gene pairs identified using high-stringency cut-off were located in other clades (Fig. 1*B* and *SI Appendix*, Fig. S1*B*). One TPS gene from these 21 PT–TPS gene pairs was filtered out during tree construction. The characterized fungal bifunctional TPSs (including STSs and DTSs) grouped together in two closely related clades that were clearly distinct from this plant TPS clade, suggesting that the plant TPS (red in Fig. 1*B*) and the fungal bifunctional STS clades (blue in Fig. 1*B*) have evolved independently. Other TPS genes that colocalized with PT genes over larger distances (low stringency) were spread across different clades (green dots in Fig. 1*B*). To understand the functions of the TPSs that emerged from our search and investigate the significance of this phylogenetic grouping, we chose to functionally test seven TPSs (five from *A. thaliana*, one from *C. rubella*, and one from *B. oleracea*) for which the genes are clustered with PT genes from the main clade (Fig. 1*C* and *SI Appendix*, Fig. S1*A*), along with two other *A. thaliana* TPSs [*AtTPS22* (*At1g33750*) and *AtTPS29* (*At1g31950*)] (*SI Appendix*, Fig. S1*A*) for which the genes are not clustered with a PT gene yet are part of the same monophyletic branch within the TPS phylogeny.

***AtTPS17*, *-18*, *-19*, *-25*, and *-30* from *A. thaliana*, *Cr237* from *C. rubella*, and *Bo250* from *B. oleracea* Are all Functional STSs.** The selected TPSs, *AtTPS17* (*At3g14490*), *AtTPS18* (*At3g14520*), *AtTPS19* (*At3g14520*), *AtTPS25* (*At3g29410*), and *AtTPS30* (*At3g32030*) from *A. thaliana*, *Cr237* (*CARUB_v10016237mg*) from *C. rubella*, and *Bo250* (*LOC106343250*) from *B. oleracea* all have corresponding clustered PT genes, namely *AtGFPPS0* (*At3g14510*), *AtGFPPS1* (*At3g14530*), *AtGFPPS2* (*At3g14550*), *AtGFPPS3* (*At3g29430*), *AtGFPPS4* (*At3g32040*), *CrGFPPS1* (*CARUB_v10014047mg*), and *BoPT1* (*LOC106293690*), respectively (Fig. 1*C*). In contrast, *AtTPS22* and *AtTPS29* are not paired with any PT genes in the genome. *AtGFPPS0* is a pseudo gene encoding a nonfunctional protein with a frame shift, whereas *AtGFPPS1–4* and *CrGFPPS1* are characterized GFPPSs (25, 30). *BoPT1* has not been previously identified or characterized. N-terminal transit peptide (N-tp)–truncated recombinant *AtTPS18* and *AtTPS19* proteins were recently coexpressed with N-tp *AtGFPPS1* and *AtGFPPS2* in *E. coli* and reported to synthesize (+)-arathanatriene and (−)-retigeranin B, respectively (22). Recombinant proteins encoded by N-tp–truncated *AtTPS22* and *AtTPS25* were also recently expressed in *E. coli* but reported to have sesquiterpene synthase activity, making sesquiterpene mixtures when incubated with *E,E*-farnesyl diphosphate (*E,E*-FPP) as a substrate in vitro (31). Recombinant proteins encoded by N-tp–truncated *AtTPS30* were found to have DTS activities when assayed with GGPP or *ent*-copalyl diphosphate (ent-CPP) as substrates in vitro, whereas *AtTPS29* was not active in these assays (31). *AtTPS17*, *Cr237*, and *Bo250* have not been investigated previously. We generated expression constructs for each of seven TPSs from the PT–TPS gene pairs (*A. thaliana AtTPS17*, *-18*, *-19*, *-25*, and *30*; *C. rubella Cr237*; *B. oleraceae Bo250*), three AtGFPPSs (*A. thaliana AtGFPPS1*, *-3*, *-4*), and the two nonclustered *A. thaliana* TPS *AtTPS22* and *AtTPS29* genes using a vector designed for transient expression in *N. benthamiana* (pEAQ-HT) (32). These constructs, each harboring individual genes, were transformed into *Agrobacterium tumefaciens*. *A. tumefaciens* strains containing the different expression constructs were then used for transient expression in *N. benthamiana* leaves. This tobacco transient-expression platform allows high-level expression of genes of interest, especially those of plant origin, thus enabling us to rapidly test the biochemical functions of individual genes (33). Furthermore, different combinations of genes can be coexpressed simply by mixing *A. tumefaciens* strains containing the different expression constructs and coinfiltrating these into the leaves (33).

When we expressed all nine TPSs (*AtTPS17*, *-18*, *-19*, *-25*, *-30*, *-22*, *-29*, *Cr237*, and *Bo250*) individually in *N. benthamiana*, trace compounds **2**, **3**, **4**, **6**, **7**, and **8** were detected from extracts of leaves expressing *AtTPS18*, *-19*, *-25*, *-30*, *Cr237*, and *Bo250* (*SI Appendix*, Fig. S2*B*). However, we did not observe the formation of sesquiterpenes or diterpenes, even for *AtTPS22*, *-25*, *-29*, and *-30*, which had previously been reported to synthesize these types of terpenes in vitro (31). This outcome may be because of the low activity of the encoding enzymes, inadequate precursor supply, or the loss of the volatile/semivolatile products in *N. benthamiana*, especially if these products were produced in very low abundance. Expression of *AtGFPPS1*, *-3*, and *-4* alone in *N. benthamiana* resulted in the accumulation of two minor new peaks (**1c** and **1d** in *SI Appendix*, Fig. S2*B*), which were tentatively identified as β-geranylfarnesene and geranylfarnesol, based on their mass spectra (*SI Appendix*, Fig. S2*C*). These are likely to be hydrolyzed products of GFPP generated by endogenous *N. benthamiana* enzymes. However, when *AtTPS17*, *-18*, *-19*, *-25*, *-30*, *Cr237*, or *Bo250* were coexpressed with a GFPPS (regardless of whether it was the corresponding paired one) in *N. benthamiana*, new peaks with a characteristic sesterterpene mass fragment at m/z = 340 (calculated for $C_{25}H_{40}$ = 340) were detected (Fig. 2*A* and *SI Appendix*, Figs. S2*E* and S3 *A–H*), indicating production of sesterterpenes by these TPSs. The electron-ionization (EI)-MS spectra of compounds **2** and **3** agreed with those of (+)-arathanatriene and (−)-retigeranin B characterized recently (22). We did not detect any sesterterpenes when *AtTPS22* or *AtTPS29* were coexpressed with a GFPPS. Apart from *AtTPS17*, which produced multiple products in very low yields, *AtTPS18*, *-19*, *-25*, *-30*, *Cr237*, and *Bo250* all produced dominant peaks corresponding to compounds **2–8** (Fig. 2 *A* and *C*), suggesting that GFPP was limiting in *N. benthamiana* leaves in the absence of a coexpressed GFPPS. Compounds **2–8** were then isolated for structural elucidation.

***AtTPS18*, *-19*, *-25*, *-30*, *Cr237*, and *Bo250* Synthesize the Fungal-Type Sesterterpenes (+)-Arathanatriene (2), (−)-Retigeranin B (3), (−)-*Ent*-Quiannulatene (4), (−)-Variculartriene A (5), (+)-Astellatene (6), (−)-Caprutriene (7), (+)-Boleracene (8), and (−)-Aleurodiscalene A (9).** To isolate compounds **2–9** for structural elucidation, we scaled up the transient expression of these different GFPPS–TPS pairs in *N. benthamiana* (34). Following extraction, repeated chromatography, and recrystallization, we obtained quantities of pure compounds **2** (102.0 mg), **3** (13.0 mg), **4** (3.7 mg), **5** (3.5 mg), **6** (4.2 mg), **7** (21.3 mg), **8** (17.3 mg), and **9** (2.1 mg) for collecting full 1D and 2D NMR datasets (Dataset S1). In addition to NMR data, we also used X-ray diffraction analysis to solve these complex structures unambiguously. We synthesized the epoxides **10**, **12**, and **13** of compounds **2**, **4**, and **6** and the enone **11** of compound **3**





C

Species	TPS	Compounds											
		2	3	4	5	6	7	8	9	14	15	16	Sum %
A. thaliana	AtTPS17	-	-	-	-	-	-	-	-	-	-	-	-
A. thaliana	AtTPS18	89	-	-	-	-	-	-	-	-	-	-	89
A. thaliana	AtTPS19	5.5	42	-	-	-	-	-	-	5.4	4.8	0	57.7
A. thaliana	AtTPS25	3.4	-	44	16	-	-	-	-	-	-	-	63.4
A. thaliana	AtTPS30	5.3	-	-	-	73	-	-	-	-	-	-	78.3
C. rubella	Cr237	-	-	-	-	-	91	-	-	-	-	-	91
B. oleracea	Bo250	-	-	-	-	-	-	81	4.5	-	-	-	85.5
A. thaliana	AtTPS19^{428Asp}	12	30	-	-	-	-	-	-	14	9.7	28	93.7
A. thaliana	AtTPS18^{431Tyr}	86	-	-	-	-	-	-	-	-	-	-	86

Fig. 2. Functional analysis of candidate STSs using transient tobacco expression. (*A*) Comparative GC-MS total ion chromatograms (TICs) of extracts from *N. benthamiana* leaves transiently coexpressing different combinations of GFPPS–TPS gene pairs. *A. tumefaciens* LBA4404 carrying the green florescent protein gene in pEAQ-HT was used as a control. (*B*) Chemical structures of compounds **2**–**9** synthesized by the selected TPSs (*AtTPS17*, *-18*, *-19*, *-25*, *-30*, *-Cr237*, and *Bo250*). (*C*) Extracted ion chromatogram (*m/z* = 340) area percentages of the identified compounds in the total sesterterpene mixtures generated by the different TPSs. Sum area percentages represent the percentages of all characterized compounds in the total sesterterpenes made by each TPS. Note that the sum does not amount to 100% because some uncharacterized and very minor sesterterpenes are not included.

(Fig. 3) to obtain single crystals for X-ray diffraction analysis. We also succeeded in obtaining single crystals of **7**, **8**, and **9** for crystallography. The crystal structures of **7**–**13** enabled us to establish the relative structures of (+)-arathanatriene (**2**), (−)-retigeranin B (**3**), (−)-*ent*-quiannulatene (**4**), and (+)-astellatene (**6**), (−)-caprutriene (**7**), (+)-boleracene (**8**), and (−)-aleurodiscalene A (**9**) unambiguously, as depicted in Fig. 2*B*. The spectroscopic data of compounds **2** and **3** agreed with those reported in the literature (22). The most likely absolute structures of the compounds (**2**–**4** and **6**–**9**) were established using Bayesian statistics on the Bijovet differences (35) of crystal structures **7**–**13**. (−)-*ent*-Quiannulatene (**4**) is an enantiomer of the fungal sesterterpene (+)-quiannulatene, sharing an identical relative structure, and its absolute structure was also confirmed by comparing its specific optical rotation ($[\alpha]_D^{20}$ = −43.9 in benzene) with that of quiannulatene ($[\alpha]_D^{28}$ = +41.7 in benzene-d_6) reported in the literature (15). The absolute structure of compound **5**, which is a minor product of *AtTPS25*, was elucidated by 1D and 2D NMR analysis (*SI Appendix*, Table S4), as well as the fact that it is a product derived from one carbocation in the cyclization pathway toward (−)-*ent*-quiannulatene (**4**) (Fig. 4). This compound is named (−)-variculatriene A based on the 5/12/5 scaffold of variculanol, which was isolated from *E. variecolor* in 1991 (36). (−)-Retigeranin B (**3**) and (+)-astellatene (**6**) are direct precursors of fungal sesterterpenoids retigeranic acid B [isolated from lichens in 1972 (37–39)] and astellatol [isolated from *E. variecolor* in 1989 (40)], respectively (*SI Appendix*, Fig. S4). (−)-Aleurodiscalene A (**9**) shares the same scaffold with the antimicrobial sesterterpene glycoside aleurodiscal isolated from *Aleurodiscus mirabilis* in 1989 (41) (*SI Appendix*, Fig. S4).

With these pure compounds in hand, we also detected compounds **2**, **3**, **4**, and **6**, (Fig. 5), which are products of *A. thaliana* TPSs in the plant of origin. We found that these four sesterterpenes are present across different *A. thaliana* tissues at various levels. (+)-Arathanatriene (**2**) and (−)-retigeranin B (**3**) were more abundant in the siliques and seeds, whereas (−)-*ent*-quiannulatene (**4**) and (+)-astellatene (**6**) were detected only in roots. The metabolite profiles are in good agreement with the expression profiles of the corresponding TPSs, suggesting that *AtTPS18*, *-19*, *-25*, and *-30* are active and could function as STSs *in planta* (Fig. 5 and *SI Appendix*, Figs. S5 and S6).

Divergence of Carbocation Cyclization Pathways Gives Rise to the Structural Diversity of Sesterterpenes 2–9. The diverse fungal-type sesterterpene products of the plant STSs identified in this work arise

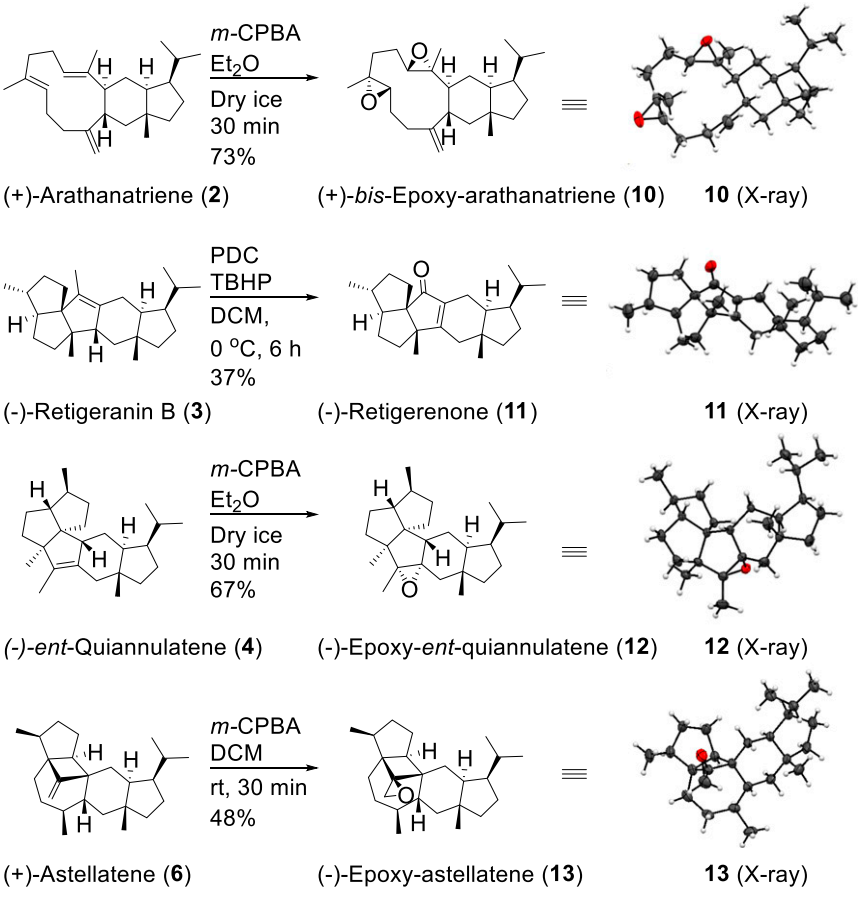


Fig. 3. Synthesis of sesterterpene derivatives (**10**–**13**) for X-ray diffraction analysis. Epoxides **10**, **12**, and **13** were synthesized by epoxidation with *meta*-chloroperoxybenzoic acid (*m*-CPBA). Enone **11** was synthesized by allylic oxidation with pyridinium dichromate (PDC) and *tert*-butyl hydroperoxide (TBHP) in dichloromethane (DCM). Crystal structures (**10**–**13**) are presented with displacement ellipsoids shown at 50% probability.

from the divergence of carbocation cyclization pathways following the formation of a common ancestral bicyclic C12 carbocation **1b** (Fig. 4). Feeding experiments with ^{13}C-labeled precursors in fungi have suggested the formation of a quiannulatene scaffold via a 5/12/5 tricyclic carbocation and an intriguing cyclobutane intermediate (15, 42). With the isolation of tricyclic (5/12/5) and tetracyclic (5/8/6/5) carbocation-derived (−)-variculatriene A (**5**) and (−)-aleurodiscalene A (**9**) as minor products of *AtTPS25* and *Bo250*, which produce mainly the quiannulatene scaffold sesterterpenes **4** and **8**, respectively (Fig. 2*C*), it is very likely that cyclization of GFPP toward (−)-*ent*-quiannulatene (**4**) and (+)-boleracene (**8**) in plant STSs adopts the same path as (+)-quiannulatene in fungal bifunctional STS (EvQS), as shown in Fig. 4. Although the astellatene and retigeranin B scaffolds have recently been proposed to also be derived from 5/12/5 tricyclic carbocations (12), previous feeding experiments with ^{13}C-labeled precursor (43) and our finding that the 11/6/5 scaffold (**2**) is the major product of AtTPS18 and is present as a very minor product of *AtTPS19*, *-25*, and *-30* (Fig. 2*C*) suggests that the bicyclic C12 carbocation may be diverged into tricyclic 5/12/5 and 11/6/5 cations en route to the formation of pentacyclic sesterterpenes, such as (−)-retigeranin B (**3**) and (+)-astellatene (**6**). Such divergence is driven by inherent energetics of the carbocation reactions involved, the conformations allowed in STS active sites and interactions of carbocations with key amino acid residues in the active sites of STSs.

Mutation of Tyrosine428 to Aspartic Acid in AtTPS19 Reveals Intermediates in the Cyclization Pathway Toward the Formation of (−)-Retigeranin B (3). Phylogenetic analysis of the seven STSs identified in this work showed that *AtTPS18* is a relatively recently evolved homolog of *AtTPS19* (*SI Appendix*, Fig. S1*A*). By comparing the protein sequences of these STSs (Fig. 6*A* and *SI Appendix*, Fig. S7), together with homology modeling of the encoding enzymes (Fig. 6*B*), we identified a conserved amino acid site present as an aromatic tyrosine (Tyr; Y) in the active sites of *AtTPS17*, *-19*, *-25*, *-30*, and *Bo250*. However, in *AtTPS18*, the corresponding residue is a negatively charged aspartic acid (Asp; D) in close proximity (approximately 2.75 Å) to the C12 of the GFPP substrate (Fig. 6*B*), suggesting that this site might be important for driving the formation of the 11/6/5 tricyclic scaffold of (+)-arathanatriene (**2**). We then carried out site-directed mutagenesis of the STSs from *A. thaliana*. We mutagenized the Tyr at the corresponding amino acid sites of *AtTPS17*, *-19*, *-25*, and *-30* to Asp. This approach resulted in inactivation of *AtTPS17*, *-25*, and *-30*, as well as significant functional changes in *AtTPS19*, with (−)-retigeranin B dropping to 30% from 42% in wild-type *AtTPS19* and the appearance of peak **16** and the increase of peaks **14** and **15** (Figs. 2*C* and 6*C*), whereas altering Asp to Tyr in *AtTPS18* did not lead to metabolite profile change (Figs. 2*C* and 6*C*). The Tyr residue in *AtTPS17*, *-25*, and *-30* therefore appears to be important for enzyme function and changing it to Asp could have possibly caused incorrect folding. Although *AtTPS19* is a closer homolog to *AtTPS18*, sharing 87% protein sequence identity (Fig. 5*A* and *SI Appendix*, Fig. S8), the mutant form of *AtTPS19*, *AtTPS19*^{428Asp}, was still active after changing Tyr428 to Asp428. We then scaled-up coinfiltration of *A. tumefaciens* strains carrying pEAQ-HT/*AtTPS19*^{428Asp} and pEAQ-HT/*AtGFPPS1* in *N. benthamiana*, isolated the compounds corresponding to peaks **14**, **15**, and **16**, and established their structures as (−)-fusaproliferene (**14**), (−)-variculatriene (**15**), and (+)-aleurodiscalene (**16**), as depicted in Fig. 4 by NMR (*SI Appendix*, Tables S13–S15). Intriguingly, changing Tyr428 to Asp428 in *AtTPS19* resulted in premature termination of the intermediate carbocations at different stages of the cyclization pathway toward (−)-retigeranin B (**3**) and led to the accumulation of the corresponding bi-, tri-, and tetracyclic products (−)-fusaproliferene (**14**), (−)-variculatriene B (**15**), (+)-arathanatriene (**2**), and (+)-aleurodiscalene B (**16**). We hypothesize that Tyr428 of *AtTPS19* might play a role in stabilizing the carbocation intermediates in

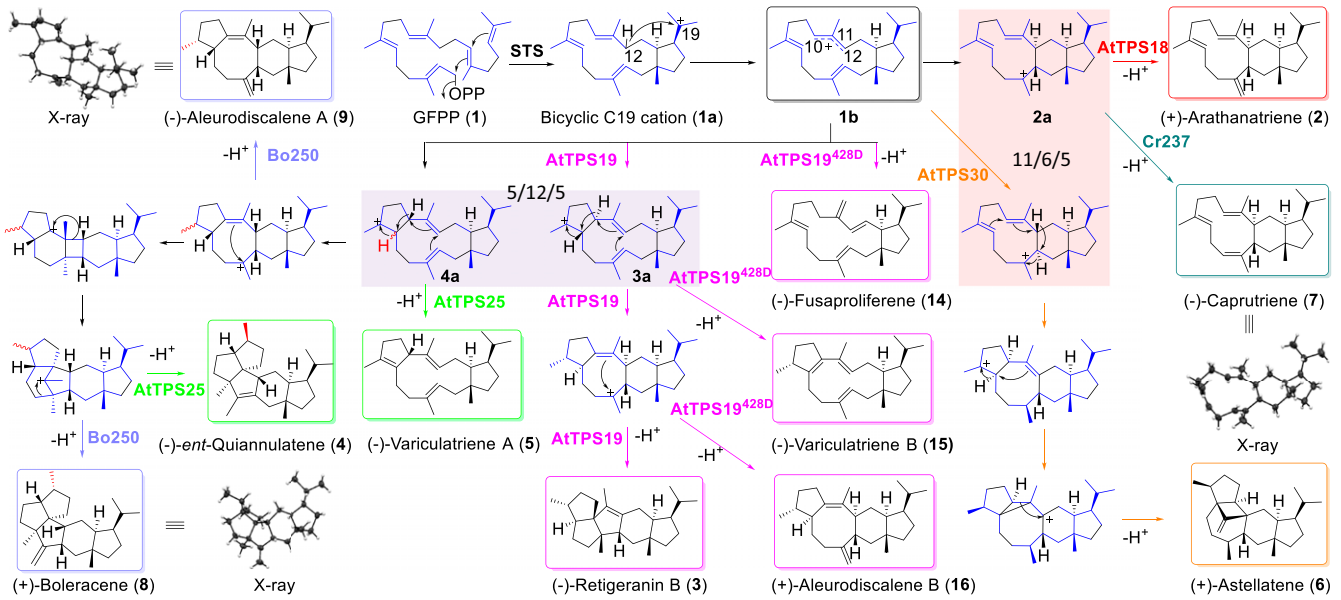


Fig. 4. Proposed cyclization paths toward the formation of fungal-type sesterterpenes **2–9** and **14–16** by plant STSs. The universal sesterterpene precursor GFPP is cyclized to form the unified bicyclic C12 cation **1b** (black box) following protonation in the active sites of plant STSs and mutated *AtTPS19* (*AtTPS19*^{428D}). Cation **1b** diverges to 5/12/5 and 11/6/5 tricyclic carbocations en route to the formation of (+)-arathanatriene (**2**), (−)-retigeranin B (**3**), (−)-*ent*-quiannulatene (**4**), (−)-variculatriene A (**5**), (+)-astellatene (**6**), (−)-caprutriene (**7**), (+)-boleracene (**8**), (−)-aleurodiscalene A (**9**), (−)-fusaproliferene (**14**), (−)-variculatriene B (**15**), and (+)-aleurodiscalene B (**16**). Compounds isolated and characterized are highlighted in colored boxes. Different colors indicate different cyclization paths. Crystal structures **7–9** are presented with displacement ellipsoids shown at 50% probability.

transition states via cation–π interactions (44), such that cyclization toward pentacyclic (−)-retigeranin B (**3**) can be completed. However, it is also possible that the enzyme activity of the *AtTPS19*^{428Asp} mutant is compromised such that not all of the intermediates could be cyclized to the final pentacyclic product. The fact that the composition of (+)-arathanatriene (**2**) increased from 5.5% of wild-type *AtTPS19* products to 12% for the mutagenized *AtTPS19*^{428Asp} suggests that a common ancestral bicyclic C12 cation may have been diverted in the enzyme active sites into 5/12/5 and 11/6/5 tricyclic carbocations en route to the formation of pentacyclic sesterterpenes, such as (−)-retigeranin B (**3**), and that such divergence could occur within a single enzyme. Although changing Asp431 to Tyr431 in *AtTPS18* did not alter its function and we have not yet identified the amino acid sites that drive the exclusive formation of the 11/6/5 tricyclic scaffold of **2**, it is reasonable to hypothesize that *AtTPS18* has evolved to be more specific and to have a more compact active site that would impose steric force on the bicyclic C12 cation (**1b**) to allow for the tight formation of the C2–C12 bond and subsequently the 11/6/5 tricyclic scaffold (**2** and **7**).

Discussion

The characterized bifunctional STSs from fungi and the plant STSs identified in this work are very distantly related (around 10% amino acid sequence identity) (*SI Appendix*, Fig. S8). However, both types of STSs are able to cyclize the universal precursor GFPP into similar or in some cases even identical scaffolds [e.g., (+)-quiannulatene by EvQS and (−)-*ent*-quiannlatene by *AtTPS25*] via virtually equivalent cyclization paths. This striking example of convergent evolution of terpenoids in plants and fungi has parallels with the well-known example of the diterpenoid gibberellin phytohormones (45). By delving into the carbocation cyclization mechanisms of the plant STSs that we have identified, we have demonstrated that cyclization of the acyclic GFPP precursor to pentacyclic sesterterpenes [e.g., (−)-retigeranin B and (−)-*ent*-quiannulatene] can diverge into 5/12/5 and 11/6/5 tricyclic scaffolds in one single enzyme after the formation of a common ancestral bicyclic C12 cation, supported experimentally by the isolation and characterization of intermediate cation derived compounds **2**, **5**, **9**, and **14–16**.

Quantum chemical calculations [mPW1PW91/6–31+G(d,p)//B3LYP/6–31+G(d,p)] (see *Materials and Methods* and Dataset S3 for details) were also performed on two reasonable cyclization pathways for the formation of (+)-astellatene (**6**) (Fig. 7). In path 1, a 11/6/5 ring system is formed first (carbocation **D**), whereas path 2 involves initial formation of a 5/12/5 ring system (carbocation **K**). On the basis of the computed energetics for these two rearrangements (Fig. 8), both pathways are energetically viable, but path 1 is inherently preferred, given that the highest energy transition-state structure along this pathway is predicted to be 6-kcal/mol lower than that along path 2. In both pathways, instances of the merging of chemical events (e.g., hydride shifts, carbocation-alkene cyclizations) into concerted processes are observed (e.g., **E** → **G**, **G** → **I**, **L** → **N**, **N** → **I**) (46). The results of these calculations indicate that several putative intermediates

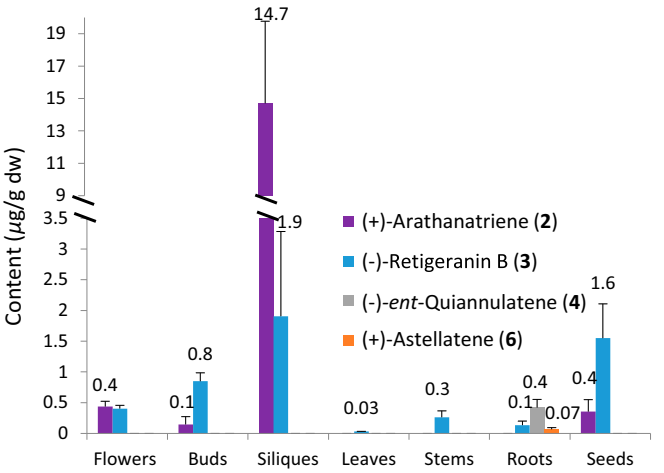


Fig. 5. Detection of compounds **2**, **3**, **4**, and **6** from different tissues of *A. thaliana* accession Col-0. SD bars for three biological replicates are shown.

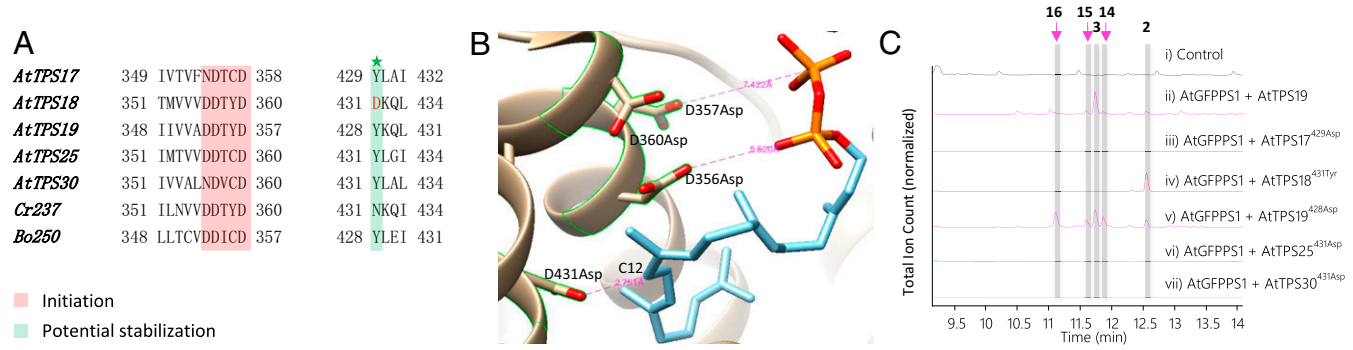


Fig. 6. Protein sequence analysis, homology modeling and site-directed mutagenesis of *A. thaliana* STSs. (*A*) Sequence analysis of *AtTPS17*, *-18*, *-19*, *-25*, *-30*, *Cr237*, and *Bo250* identified a conservation site implicated in structural diversification (protein sequences were aligned using MUSCLE). (*B*) Homology modeling of *AtTPS18* with substrate GFPP (carbon backbone in cyan and phosphate in orange and red) docked in the active site showing the motif DDXXD and 431Asp suggested that 431Asp might be involved in making the 11/6/5 tricyclic scaffold of arathanatriene. The homology model of *AtTPS18* was generated based on the crystal structure of 5-epi-aristolochene synthase on the Phyre[2] server (www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). (*C*) Comparative GC-MS total ion chromatograms of extracts of *N. benthamiana* leaves transiently coexpressing AtGFPPS1 with mutagenized $AtTPS17^{429Asp}$, $AtTPS18^{431Tyr}$, $AtTPS19^{428Asp}$, $AtTPS25^{431Asp}$, and $AtTPS30^{431Asp}$.

are unlikely to have lifetimes long enough for deprotonation to occur, allowing us to predict which carbocations are most likely to lead directly to sesterterpenes.

By identifying a key amino acid residue Tyr428 in *AtTPS19* important for the complete the cyclization toward (−)-retigeranin B (**3**), we have unveiled the bi-/tri-/tetracyclic structures **14**, **15**, **2**, and **16** that are derived from likely carbocations in the cyclization path toward (−)-retigeranin B (**3**), demonstrating the capacity of plant STSs to generate these diverse scaffolds and showcasing the promising potential for directing the functional plasticity of STSs for creating chemical diversity via rational protein engineering (47). Solving the crystal structures of these enzymes will be key in guiding such approaches in the future. The fact that (−)-fusaproliferene (**14**) and (+)-aleurodiscalene B (**16**) are direct precursors of the known bioactive sesterterpenoids fusaproliferin (toxic) and aleurodiscal (antimicrobial) (*SI Appendix*, Fig. S4) illustrates the importance of harnessing the capacity to generate scaffolds of this type.

We present in this work the convergent evolution of STSs in fungi and plants as inferred from both metabolic and genomic data. Both fungi and plants appear to have independently evolved similar strategies for biosynthesizing structurally related sesterterpenes. Although, unlike the situation in fungi, plant STSs and GFPPSs are encoded by separate genes rather than bifunctional ones, it is intriguing that the genes encoding the STSs that we have characterized are colocalized in plant genomes with PT genes. These gene pairs are not obviously coexpressed (*SI Appendix*, Fig. S5) and the GFPPSs are interchangeable (Fig. 2*A*), suggesting that these gene pairs have not undergone intimate coevolution. The genomic colocalization of PT and TPS genes appears to be particularly prevalent in the Brassicaceae based on our analysis of currently available plant genomes. From the topology of the paired TPSs within the phylogenetic tree (Fig. 1*B*) and our functional analysis of selected TPSs from *A. thaliana*, *C. rubella*, and *B. oleracea*, it seems that the functions of closely related TPSs from these three species are quite conserved and that close homologs make sesterterpene analogs (Fig. 9). It is likely that these PT and TPS gene pairs may have originated from a common ancestral PT–TPS gene pair by duplication and functional divergence (23, 48–50). Although the biological roles of these sesterterpenes in plants are as yet unknown and inevitably many downstream products in the biosynthetic pathways remain to be elucidated, we expect that the biosynthetic and mechanistic insights provided by this work will facilitate the understanding and discovery of the biology and chemistry of this important class of compounds in plants.

Materials and Methods

Bioinformatics Algorithms and Construction of Phylogenetic Trees. A collection of 55 high-quality plant genomes (Dataset S2) was used as input for analysis with a customized version of plantiSMASH (plantismash.secondarymetabolites.org) (24). Terpene_synth_C (PF03936) and Polyprenyl_synt (PF00348) profile hidden Markov models (HMMs; pfam.xfam.org, database v30.0) were fitted into a custom gene cluster detection rule (Dataset S2).

Subsequently, the PF03936 profile HMM was used to collect all TPS-encoding genes from the 55 plant genomes, which resulted in the identification of 2,846 genes. A total of 724 fungal genes, including 17 bifunctional TPS genes (15), were also collected with the same method but queried against the reference proteome database with taxonomic filtering set to "Fungi." Finally, 278 bacterial genes (same method, with filtering set to "Bacteria") were included to act as an outgroup.

All sequences were aligned using the hmmalign program from the HMMer suite (51) with default parameters. Predicted repetitive subsequences were discarded. The remaining sequences were manually trimmed for tree reconstruction, which was performed using FastTree (v2.1) (52). The phylogenetic tree was visualized using iTOL software (53). Reliability tests were done using the FastTree-provided Shimodaira–Hasegawa test with 1,000× resamples without branch optimization.

Cloning and Transient Expression. *AtTPS19*, *AtTPS25*, *AtTPS30*, *AtGFPPS1*, and *AtGFPPS3* were amplified from a root cDNA library of *A. thaliana* accession Col-0. Coding sequences for *AtTPS17*, *AtTPS18*, *AtTPS22*, *AtTPS29*, and *AtGFPPS4* were retrieved from The Arabidopsis Information Resource and those of *Cr237* (*CARUB_v10016237mg*) and *Bo250* (*LOC106343250*) from the National Center for Biotechnology Information and synthesized by Integrated DNA Technologies to include the 5′ and 3′ attB sites for Gateway cloning. These sequences were cloned into the pEAQ-HT expression vector for transient expression in *N. benthamiana*, as detailed in *SI Appendix*.

Metabolite Extraction and GC-MS Analysis. *N. benthamiana* leaves expressing genes of interest were harvested 5 d postinfiltration, lyophilized, extracted with EtOAc, and analyzed directly by GC-MS, as described in *SI Appendix*. Qualitative and quantitative metabolite analysis were performed on an Agilent GC (7890B)-MSD (5977A) equipped with a robotic multipurpose autosampler (MPS) and a Zebra-5HT INFERO capillary column (35-m × 0.25-mm × 0.1-μm film thickness; phenomenex).

Isolation of Sesterterpenes Following Large-Scale Vacuum-Infiltration of *N. benthamiana* Leaves. *A. tumefaciens* LBA4404 harboring expression constructs for the relevant GFPPSs and STSs were infiltrated into *N. benthamiana* leaves as described previously (34), with slight modifications using a custom-built vacuum infiltration system (*SI Appendix, Materials and Methods*). Leaves of *N. benthamiana* expressing desired constructs were harvested 5 d post-infiltration, lyophilized, and extracted with *n*-hexane. Pure compounds **2**–**9** and **14**–**16** were isolated from the extract as described in *SI Appendix*.

Chemical Synthesis of Sesterterpene Derivatives (10–13) for X-Ray Diffraction Analysis. Compounds **10**, **12**, and **13** were synthesized by epoxidation of sesterterpenes **2**, **4**, and **5** with *meta*-chloroperoxybenzoic acid. Enone **11** was synthesized by allylic oxidation of compound **3** with pyridinium dichromate and *tert*-butyl hydroperoxide in dichloromethane (*SI Appendix, Materials and Methods*).

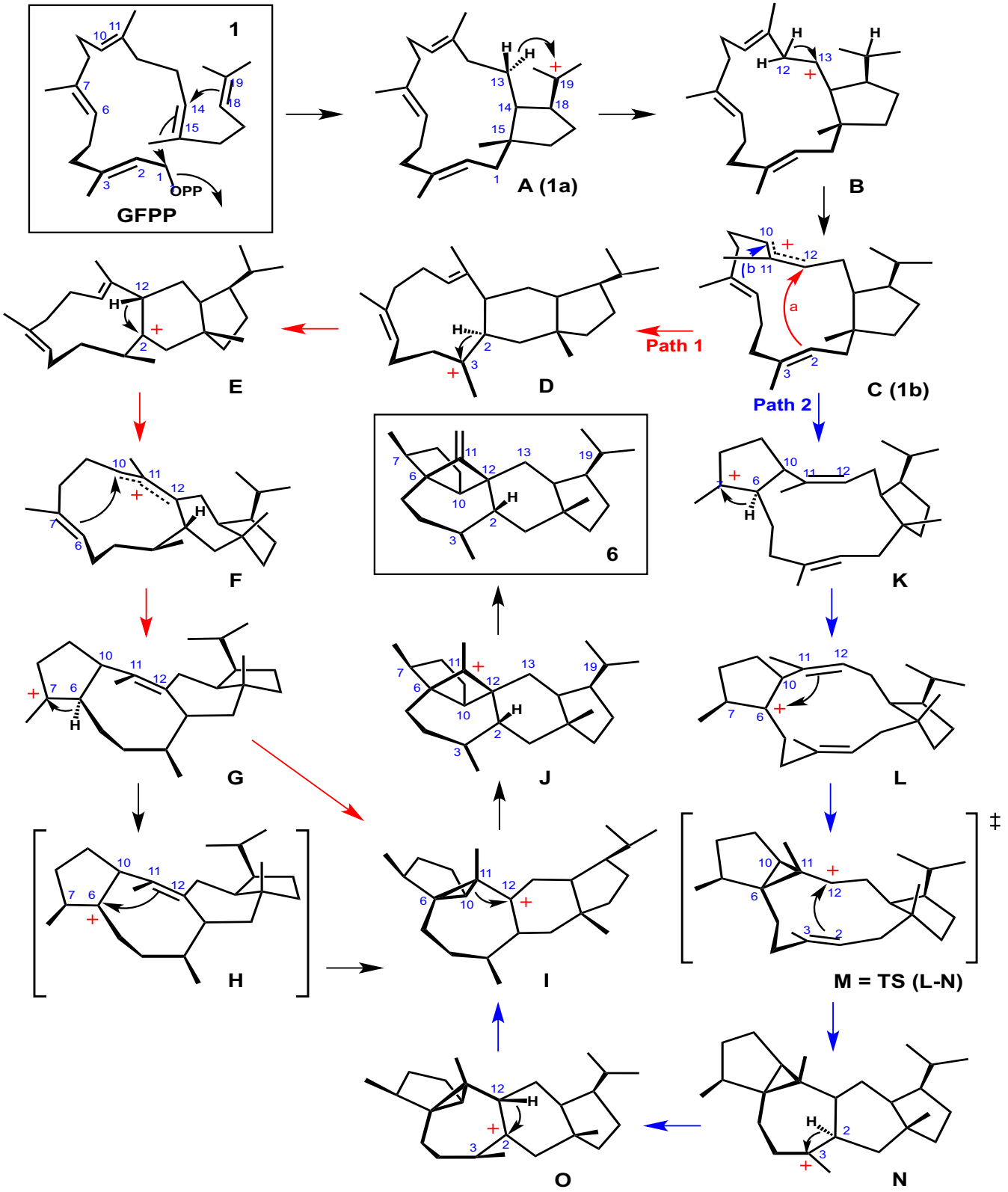


Fig. 7. (*A*–*O*) Calculated structures for two possible cyclization paths en route to (+)-astellatene (**6**).

Structural Characterization of Sesterterpenes by NMR and Other Spectroscopic Techniques. Standard 1D and 2D NMR spectra, including ^{1}H, ^{13}C, DEPT135, COSY, HSQC, HMBC, and NOESY were acquired on a Bruker 400-MHz Topspin NMR spectrometer. All signals were acquired at 298 K. Samples were dissolved in $CDCl_3$ or benzene-d_6 for data acquisition and calibrated by referencing to either residual solvent ^{1}H and ^{13}C signal or TMS. Detailed structural assignments and tabulated NMR data for compounds **2**–**16** are presented in *SI Appendix*, Tables S1–S15. Optical rotations of compounds **2**–**16** were measured in benzene using a PerkinElmer Polarimeter (Model 341) with a 100-mm path cell (1 mL) at 20 °C and converted to

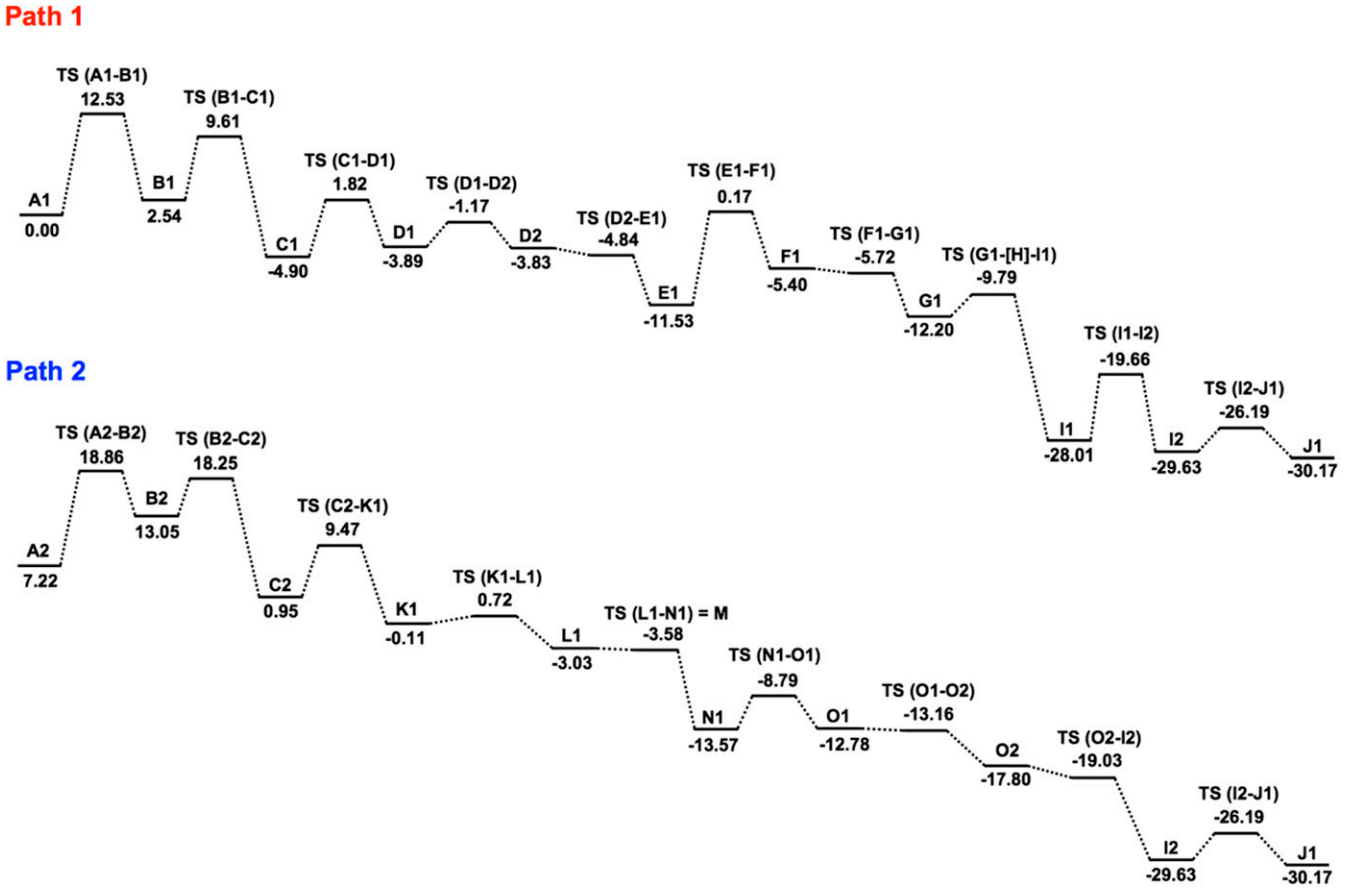


Fig. 8. Computed energetics for two possible paths to (+)-astellatene (**6**). Energies computed with mPW1PW91/6–31+G(d,p)//B3LYP/6–31+G(d,p) are shown for the two pathways depicted in Fig. 7. Path 1 is energetically preferred. Numbers following carbocation letters indicate different conformations. See Dataset S3 for computed geometries.

specific optical rotation using equation $[\alpha]_D^{20}$ = 100 × (measured optical rotation)/concentration (g/100 mL).

Single Crystal X-Ray Diffraction Analysis. Single-crystal X-ray analysis was carried out for **7**, **8**, **9**, **10**, **11**, **12**, and **13** on a Bruker D8-QUEST instrument equipped with a PHOTON-100 area detector and Incoatec IμS Cu microsource (wavelength = 1.5418 Å, beam diameter at the crystal approximately 100 μm). Crystals were mounted on an X-ray transparent loop (Mitegen) using an inert oil and cooled to 180(2) K using an open-flow N_2 cryostat. Data were collected and processed using the *APEX3* software package (Bruker). Structures were solved and refined using *SHELXT* and *SHELXL* (Bruker). In the absence of any significant anomalous scatterers, reliable indications of the absolute structure could not be obtained by conventional refinement of the Flack parameter. However, Bayesian statistical methods (35) indicate the following probabilities that the absolute structure is correct, under the assumption that the crystal is enantiopure [*p2*(true)]: **7** 0.886, **8** 0.980, **9** 0.682, **10** 1.000, **11** 1.000, **12** 1.000, and **13** 0.745. See Datasets S4 and S5 for crystal structures and crystallographic raw data of compounds **7**–**13**.

Sequence Comparisons and Homology Modeling. Amino acid sequences of *AtTPS17*, *-18*, *-19*, *-25*, *-30*, *Cr237*, and *Bo250* were aligned using Uniprot

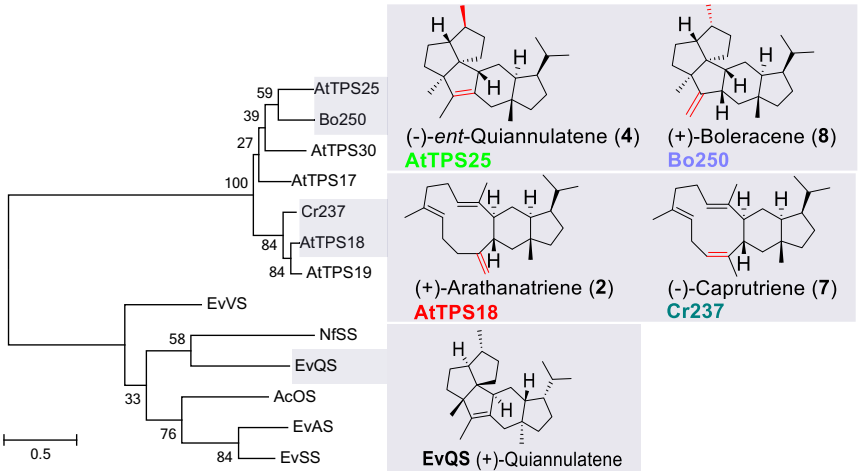


Fig. 9. Phylogenetic tree of the characterized plant and fungal STSs showing the phylogeny of these distantly related proteins. The chemical structures of analogs synthesized by plant STS homologs and the distantly related fungal STS EvQS are shown to aid in evolutionary interpretation. The tree was constructed using MEGA5.0 with 1,000 bootstrap resampling.

(www.uniprot.org/). Homology models of *AtTPS18* were generated on modeling server Phyre[2] (www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) using the crystal structure of 5-epi-aristolochene synthase as template. Docking of STSs substrate GFPP was performed using Autodock4.0 and AutoDockTool developed by The Scripps Research Institute (54).

Mutagenesis of *AtTPSs17*, *-18*, *-19*, *-25*, and *-30*. Site-directed mutagenesis was performed by PCR amplification using the entry vector pDONR207 harboring the wild-type genes as templates and the mutated complementary sequences as primers, as listed in the *SI Appendix*, Table S17. Mutagenized genes were Gateway-cloned into the pEAQ-HT expression vector, transformed into LBA4404 and coexpressed with LBA4404/GFPPS1 in our transient tobacco expression system, as described in *Cloning and Transient Expression*, above.

Quantum Chemical Calculations. All quantum calculations were performed with the *GAUSSIAN09* software suite (55). Geometries were optimized using the B3LYP density functional theory method and the 6–31+G(d,p) basis set (56). All stationary points were characterized as minima or transition-state structures using frequency calculations. All reported energies include zero-point energy corrections (unscaled) from these frequency calculations. mPW1PW91 single-point energies were calculated for all structures (57). These methods are well established for examining carbocation rearrangement reactions (46). Structures of intermediate carbocations, transition state structures, and corresponding coordinates are presented in Dataset S3.

ACKNOWLEDGMENTS. We thank A. Leveau, H. Nutzmann, R. Thimmappa, and other members of the A.O. laboratory for helpful discussion and suggestions on experiments. Research in D.J.T.'s laboratory is supported by US National Science Foundation CHE-1565933 and computing support from CHE030089 (XSEDE). A.O.'s laboratory is supported by the United Kingdom Biotechnological and Biological Sciences Research Council Institute Strategic Programme Grant "Understanding and Exploiting Plant and Microbial Metabolism" (BB/J004561/1) and the John Innes Foundation. This work is supported by the NIH Genome to Natural Products Network Award U101GM110699 (to A.O. and A.C.H.). A.C.H. is supported by European Commission Marie Skłodowska-Curie Individual Fellowship H2020-MSCA-IF-EF-ST-702478-TRIGEM. M.H.M. is supported by VENI Grant 863.15.002 from The Netherlands Organization for Scientific Research. S.A.K. is supported by a grant from the Graduate School for Experimental Plant Sciences (to M.H.M.).

1. Wang L, Yang B, Lin X-P, Zhou X-F, Liu Y (2013) Sesterterpenoids. *Nat Prod Rep* 30:455–473.
2. Evidente A, et al. (2015) Sesterterpenoids with anticancer activity. *Curr Med Chem* 22: 3502–3522.
3. Zhang C, Liu Y (2015) Targeting cancer with sesterterpenoids: The new potential antitumor drugs. *J Nat Med* 69:255–266.
4. Luo S-H, et al. (2010) Glandular trichomes of *Leucosceptrum canum* harbor defensive sesterterpenoids. *Angew Chem Int Ed Engl* 49:4471–4475.
5. Li C-H, et al. (2013) Peltate glandular trichomes of *Colquhounia coccinea* var. mollis harbor a new class of defensive sesterterpenoids. *Org Lett* 15:1694–1697.
6. Luo S-H, et al. (2011) Defensive sesterterpenoids with unusual antipodal cyclopentenones from the leaves of *Leucosceptrum canum*. *Org Lett* 13:1864–1867.
7. Luo S-H, et al. (2013) Defense sesterterpenoid lactones from *Leucosceptrum canum*. *Phytochemistry* 86:29–35.
8. Ahkami A, Johnson SR, Srividya N, Lange BM (2015) Multiple levels of regulation determine monoterpenoid essential oil compositional variation in the mint family. *Mol Plant* 8:188–191.
9. Upadhyay AK, et al. (2015) Genome sequencing of herb Tulsi (*Ocimum tenuiflorum*) unravels key genes behind its strong medicinal properties. *BMC Plant Biol* 15:212.
10. Vining KJ, et al. (2017) Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. *Mol Plant* 10:323–339.
11. Chiba R, Minami A, Gomi K, Oikawa H (2013) Identification of ophiobolin F synthase by a genome mining approach: A sesterterpene synthase from *Aspergillus clavatus*. *Org Lett* 15:594–597.
12. Ye Y, et al. (2015) Genome mining for sesterterpenes using bifunctional terpene synthases reveals a unified intermediate of di/sesterterpenes. *J Am Chem Soc* 137:11846–11853.
13. Matsuda Y, et al. (2016) Astellifadiene: Structure determination by NMR spectroscopy and crystalline sponge method, and elucidation of its biosynthesis. *Angew Chem Int Ed Engl* 55:5785–5788.
14. Qin B, et al. (2016) An unusual chimeric diterpene synthase from *Emericella variecolor* and its functional conversion into a sesterterpene synthase by domain swapping. *Angew Chem Int Ed Engl* 55:1658–1661.
15. Okada M, et al. (2016) Genome-based discovery of an unprecedented cyclization mode in fungal sesterterpenoid biosynthesis. *J Am Chem Soc* 138:10011–10018.
16. Matsuda Y, Mitsuhashi T, Quan Z, Abe I (2015) Molecular basis for stellatic acid biosynthesis: A genome mining approach for discovery of sesterterpene synthases. *Org Lett* 17:4644–4647.
17. Chen M, Chou WKW, Toyomasu T, Cane DE, Christianson DW (2016) Structure and function of fusicoccadiene synthase, a hexameric bifunctional diterpene synthase. *ACS Chem Biol* 11:889–899.
18. Toyomasu T, et al. (2007) Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi. *Proc Natl Acad Sci USA* 104:3084–3088.
19. Toyomasu T, et al. (2009) Biosynthetic gene-based secondary metabolite screening: A new diterpene, methyl phomopsenonate, from the fungus *Phomopsis amygdali*. *J Org Chem* 74:1541–1548.
20. Brodelius M, Lundgren A, Mercke P, Brodelius PE (2002) Fusion of farnesyldiphosphate synthase and epi-aristolochene synthase, a sesquiterpene cyclase involved in capsidiol biosynthesis in *Nicotiana tabacum*. *Eur J Biochem* 269:3570–3577.
21. Fischer MJ, Rustenhloz C, Leh-Louis V, Perrière G (2015) Molecular and functional evolution of the fungal diterpene synthase genes. *BMC Microbiol* 15:221.
22. Shao J, et al. (2017) (+)-Thalianatriene and (–)-retigeranin B catalyzed by sesterterpene synthases from *Arabidopsis thaliana*. *Org Lett* 19:1816–1819.
23. Boutanaev AM, et al. (2015) Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci USA* 112:E81–E88.
24. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH (2017) plantiSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*, 10.1093/nar/gkx305.
25. Wang C, et al. (2016) Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in Brassicaceae plants. *Mol Plant* 9:195–204.
26. Keeling CI, et al. (2011) Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). *BMC Plant Biol* 11:43.
27. Schilmiller AL, et al. (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc Natl Acad Sci USA* 106:10865–10870.
28. Falara V, et al. (2011) The tomato terpene synthase gene family. *Plant Physiol* 157:770–789.
29. Finn RD, et al. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285.
30. Nagel R, et al. (2015) *Arabidopsis thaliana* isoprenyl diphosphate synthases produce the C25 intermediate geranylfarnesyl diphosphate. *Plant J* 84:847–859.
31. Wang Q, et al. (2016) Identification of a dolabellane type diterpene synthase and other root-expressed diterpene synthases in *Arabidopsis*. *Front Plant Sci* 7:1761.
32. Sainsbury F, Thuenemann EC, Lomonossoff GP (2009) pEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J* 7:682–693.
33. Sainsbury F, Lomonossoff GP (2014) Transient expressions of synthetic biology in plants. *Curr Opin Plant Biol* 19:1–7.
34. Geisler K, et al. (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc Natl Acad Sci USA* 110:E3360–E3367.
35. Hooft RWW, Straver LH, Spek AL (2008) Determination of absolute structure using Bayesian statistics on Bijvoet differences. *J Appl Crystallogr* 41:96–103.
36. Singh SB, et al. (1991) Variculanol: Structure and absolute stereochemistry of a novel 5/12/5 tricyclic sesterterpenoid from *Aspergillus variecolor*. *J Org Chem* 56:5618–5622.
37. Kaneda M, Takahashi R, Iitaka Y, Shibata S (1972) Retigeranic acid, a novel sesterterpene isolated from the lichens of *Lobaria retigera* group. *Tetrahedron Lett* 13:4609–4611.
38. Corey EJ, Desai MC, Engler TA (1985) Total synthesis of (+–)-retigeranic acid. *J Am Chem Soc* 107:4339–4341.
39. Kaneda M, Iitaka Y, Shibata S (1974) X-ray studies of C25 terpenoids. IV. The crystal structure of retigeranic acid p-bromoanilide. *Acta Crystallogr B* 30:358–364.
40. Sadler IH, Simpson TJ (1989) The determination by n.m.r. methods of the structure and stereochemistry of astellatol, a new and unusual sesterterpene. *J Chem Soc Chem Commun* 21:1602–1604.
41. Lauer U, Anke T, Sheldrick WS, Scherer A, Steglich W (1989) Antibiotics from basidiomycetes. XXXI. Aleurodiscal: An antifungal sesterterpenoid from *Aleurodiscus mirabilis* (Berk. & Curt.) Höhn. *J Antibiot (Tokyo)* 42:875–882.
42. Hong YJ, Tantillo DJ (2014) How cyclobutanes are assembled in nature—Insights from quantum chemistry. *Chem Soc Rev* 43:5042–5050.
43. Simpson TJ (1994) Biosynthesis of astellatol, a novel rearranged sesterterpenoid metabolite of *Aspergillus variecolor*. *J Chem Soc Perkin Trans 1* (21):3055–3056.
44. Hong YJ, Tantillo DJ (2013) C–H–pi interactions as modulators of carbocation structure—Implications for terpene biosynthesis. *Chem Sci* 4:2512–2518.
45. Hedden P, Sponsel V (2015) A century of gibberellin research. *J Plant Growth Regul* 34:740–760.
46. Tantillo DJ (2011) Biosynthesis via carbocations: Theoretical studies on terpene formation. *Nat Prod Rep* 28:1035–1053.
47. Keeling CI, Weisshaar S, Lin RPC, Bohlmann J (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc Natl Acad Sci USA* 105:1085–1090.
48. Field B, et al. (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci USA* 108:16116–16121.
49. Matsuba Y, et al. (2013) Evolution of a complex locus for terpene biosynthesis in *Solanum*. *Plant Cell* 25:2022–2036.
50. Trapp SC, Croteau RB (2001) Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* 158:811–832.
51. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195.
52. Price MN, Dehal PS, Arkin AP (2010) FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490.
53. Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245.
54. Morris GM, et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791.
55. Frisch MJ, et al. (2013) *Gaussian 09 Revision D.01* (Gaussian Inc., Wallingford, CT).
56. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652.
57. Matsuda SPT, Wilson WK, Xiong Q (2006) Mechanistic insights into triterpene synthesis from quantum mechanical calculations. Detection of systematic errors in B3LYP cyclization energies. *Org Biomol Chem* 4:530–543.