



Origin and early evolution of the plant terpene synthase family

Qidong Jia^{a,1,2}, Reid Brown^{b,1}, Tobias G. Köllner^c, Jianyu Fu^{d,e}, Xinlu Chen^f, Gane Ka-Shu Wong^{g,h,i}, Jonathan Gershenzon^c, Reuben J. Peters^{b,2}, and Feng Chen^{a,f,2}

Edited by Dean DellaPenna, Michigan State University, East Lansing, MI; received July 29, 2021; accepted February 24, 2022

As a mid-sized gene family conserved more by lineage than function, the typical plant terpene synthases (*TPSs*) could be a valuable tool to examine plant evolution. *TPSs* are pivotal in biosynthesis of gibberellins and related phytohormones as well as in formation of the extensive arsenal of specialized plant metabolites mediating ecological interactions whose production is often lineage specific. Yet the origin and early evolution of the *TPS* family is not well understood. Systematic analysis of an array of transcriptomes and sequenced genomes indicated that the *TPS* family originated after the divergence of land plants from charophytic algae. Phylogenetic and biochemical analyses support the hypothesis that the ancestral *TPS* gene encoded a bifunctional class I and II diterpene synthase producing the *ent*-kaurene required for phytohormone production in all extant lineages of land plants. Moreover, the ancestral *TPS* gene likely underwent duplication at least twice early in land plant evolution. Together these two gave rise to three *TPS* lineages leading to the extant *TPS-c*, *TPS-e/f*, and the remaining *TPS (h/d/a/b/g)* subfamilies, with the latter dedicated to secondary rather than primary metabolism while the former two contain those genes involved in *ent*-kaurene production. Nevertheless, parallel evolution from the *ent*-kaurene-producing class I and class II diterpene synthases has led to roles for *TPS-e/f* and *-c* subfamily members in secondary metabolism as well. These results clarify *TPS* evolutionary history and provide context for the role of these genes in producing the vast diversity of terpenoid natural products observed today in various land plant lineages.

nonseed plants | terpenoids | secondary metabolites | diterpene synthases

The conquest of land by pioneering land plants and their subsequent diversification forever changed the terrestrial ecosystem on Earth. The evolution of land plants relied on continual genetic innovations, among which is the elaboration of a varied arsenal of natural products (1). Terpenoids constitute the largest class of natural products made by land plants. While a few terpenoids such as gibberellins are phytohormones with essential roles in regulating growth and development (2), the vast majority are specialized or secondary metabolites involved in diverse plant interactions with the environment (3). Thus, continued innovation of terpenoid biosynthesis has played an important role in the adaptations of land plants during diversification, that is, into the extant bryophytes (hornworts, liverworts, and mosses), lycophytes, ferns, gymnosperms, and angiosperms (4).

One important avenue to understanding the evolution of terpenoid biosynthesis in land plants is the investigation of terpene synthases, which are pivotal enzymes that initiate terpenoid biosynthesis. These generally catalyze lysis of the allylic diphosphate ester bond in acyclic isoprenyl diphosphate precursors, leading to cyclization that generates enormous structural diversity (5). Such enzymes have been termed class I terpene synthases (Fig. 1) and catalyze their reactions in the cavity of an alpha-helical bundle protein fold utilizing a pair of motifs, DDxxD and NSE/DTE (5) that coordinate the requisite trio of divalent magnesium (Mg^{2+}) cofactors (6). Plants contain two families of genes encoding class I terpene synthases: typical plant terpene synthase (*TPS*) genes and microbial-terpene synthase-like (*MTPSL*) genes. *MTPSL* genes resemble terpene synthase genes from bacteria and fungi (7) and occur only in nonseed plants (8). In contrast, *TPS* genes, the focus of this study, occur widely in land plants, where they form a mid-sized gene family with striking evolutionary plasticity (9). Intriguingly, *TPSs* invariably contain a substantial additional domain (9). This N-terminal addition has been termed the β domain to distinguish it from the helical bundle class I fold α domain (10) that is also present in *MTPSLs* (11). Notably, this β domain appears to be relictual, a remnant from an ancestral gene fusion event, in particular of a class I diterpene synthase with a class II diterpene cyclase that gave rise to a bifunctional diterpene cyclase (12) (Fig. 1). While the β domain is not involved in class I catalysis, it has been shown that the interface between the α and β domains is required for folding and stability of the α domain in the *TPS* family (13).

Significance

Land plants produce numerous terpenoids that regulate development and mediate environmental interactions. Thus, how typical plant terpene synthase (*TPS*) genes originated and evolved to create terpenoid diversity is of fundamental interest. By investigating *TPSs* from the genomes and transcriptomes of diverse taxa of green plants, it was demonstrated here that the ancestral *TPS* gene originated in land plants after divergence from green algae and encoded a bifunctional *ent*-kaurene synthase for phytohormone biosynthesis. This ancestral *TPS* then underwent gene duplication at least twice early in land plant evolution, leading to three ancient *TPS* lineages reflecting sub-functionalization of class I and II activities for phytohormone biosynthesis and neo-functionalization from primary to secondary metabolism, followed in each case by dynamic functional divergence.

Author contributions: Q.J., T.G.K., J.G., R.J.P., and F.C. designed research; Q.J., R.B., T.G.K., J.F., and X.C. performed research; Q.J., R.B., T.G.K., J.F., X.C., G.K.-S.W., J.G., R.J.P., and F.C. analyzed data; and Q.J., R.B., J.G., R.J.P., and F.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹Q.J. and R.B. contributed equally to this work.

²To whom correspondence may be addressed. Email: qjia2@vols.utk.edu, rjpeters@iastate.edu, or fengc@utk.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2100361119/-DCSupplemental>.

Published April 8, 2022.

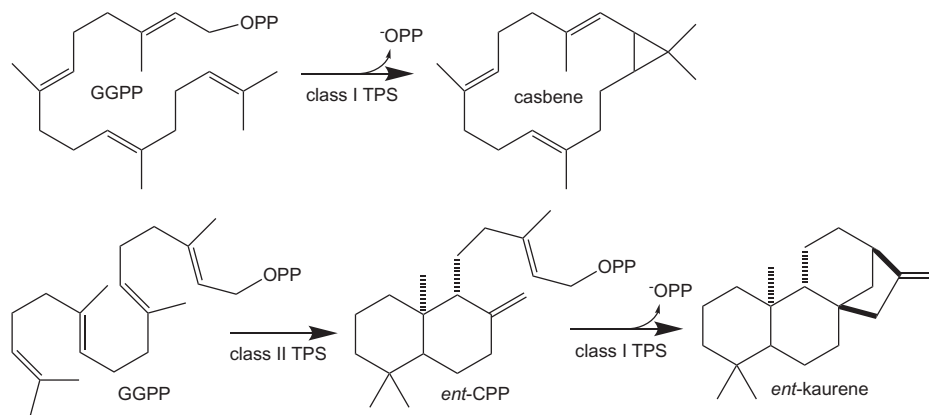


Fig. 1. Biochemical reactions catalyzed by representative class I terpene synthase (class I TPS) and class II terpene synthase (class II TPS) enzymes. Some terpene synthases are bifunctional, having both class I and class II activities. GGPP: geranylgeranyl diphosphate; OPP, diphosphate group; ent-CPP: ent-copalyl diphosphate.

Class II diterpene cyclases catalyze protonation-initiated bicyclization of the general diterpenoid precursor (*E,E,E*)-geranylgeranyl diphosphate (GGPP). In plants, all known class II terpene synthases fall into the TPS and not the MTPSL family. The class II active site resides between a pair of alpha-helical double-barrel domains, the β domain and an additionally N-terminal γ domain (14). The site utilizes a DxDD motif, wherein the middle aspartate acts as the catalytic acid (15). Class II activity is present in all land plants since production of *ent*-copalyl diphosphate (*ent*-CPP) is a necessary step in the formation of *ent*-kaurene (Fig. 1). This tetracyclic diterpene is a precursor to the gibberellin phytohormones required for normal growth and development in all vascular plants and other phytohormones as well that are found in nonvascular land plants (bryophytes) (2). Perhaps not surprisingly then, it has been hypothesized that the ancestral TPS was a bifunctional diterpene synthase/cyclase (class I and II activity) that produced *ent*-kaurene (9, 16). The class II activity would then serve as an *ent*-CPP synthase (CPS), while the class I activity serves an *ent*-kaurene synthase (KS). However, while examples of such fused CPSKSs are present in certain extant species [e.g., in the moss *Physcomitrella patens* (17)], seed plants contain separate CPS and KS activities, which clearly arose from gene duplication and subfunctionalization but still contain the ancestral $\gamma\beta\alpha$ tri-domain architecture (2). While the CPSs phylogenetically group with the extant CPSKSs in what has been termed the TPS-c subfamily, the KSs have given rise to the TPS-*e/f* subfamily, which no longer exhibits class II diterpene cyclase activity (9). In both cases, CPSs and KSs have repeatedly undergone gene duplication and neofunctionalization in various plant lineages, giving rise to enzymes involved in secondary metabolism (2).

Besides the TPS-c and TPS-*e/f* subfamilies, other terpene synthase subfamilies have been described (18). Three are confined to the angiosperms (TPS-a, -b, and -g), while gymnosperm TPSs are all grouped together in one subfamily (TPS-d), which has since been broken up into three distinct subgroups (9). The lycophyte TPSs not involved in phytohormone biosynthesis were assigned to the TPS-h subfamily (9). Within this scheme, the evolutionary relationships of seed plant TPSs are relatively clear, with early loss of the γ domain from a class I $\gamma\beta\alpha$ tridomain progenitor, leading to the TPS-a/b/g subfamilies and TPS-d1 group. In contrast, the origin and early evolution of the TPS family is still opaque. This is mainly due to the lack of knowledge about TPSs from nonseed plants. Here, large-scale mining of TPS genes from green plants of wide taxonomic

scope, particularly nonseed plants, was undertaken to infer the origin of the TPS family. The genes assembled were then subjected to phylogenetic analysis and large-scale biochemical characterization to develop a model for the early evolution of the TPS family in the context of land plant diversification.

Results and Discussion

Ubiquitous Occurrence of TPS Genes in Land Plants and Their Absence in Green Algae Imply the Origination of TPS Genes in Ancestral Land Plants. To determine the distribution of TPS genes in green plants, genome sequences of 31 species of green algae (*SI Appendix, Table S1*), 14 species of land plants (*SI Appendix, Table S2*), and transcriptomes of 1,178 species of green plants that include 1,109 species of the OneKP dataset (8) and 69 species of ferns (19) (*SI Appendix, SI Material and Methods*) were mined for TPS genes using a hidden Markov model (HMM) search. TPS genes were detected in the transcriptomes and genomes of all major lineages of land plants (Fig. 2): angiosperms, gymnosperms, ferns, lycophytes, and all three lineages of bryophytes (hornworts, liverworts, and mosses). They were identified from every genome investigated, but not in each transcriptome. TPS genes were detected in 85 to 100% of the transcriptomes from angiosperms, gymnosperms, lycophytes, and liverworts, but the detection rate in the transcriptomes of other plant lineages was only 63% for ferns, 54% for mosses, and 50% for hornworts. In contrast to land plants, no TPS genes were detected in green algae after searching 158 transcriptomes (47 charophytes and 111 chlorophytes) and 31 genomes. While several hits with significant e-values were identified in the genome of the charophyte *Klebsormidium flaccidum*, none had the conserved features of TPS proteins when the sequences were examined in detail. The absence of TPS genes in red algae has been previously noted (20). The lack of TPS genes in a wide range of green algae, particularly in charophytes, presumed to be the closest relatives of land plants, contrasts with their ubiquitous occurrence in land plants. Accordingly, it can be hypothesized that the TPS family originated in ancestral land plants after their divergence from green algae.

Phylogenetic Analysis Infers the Existence of Three Lineages of Tridomain TPS Genes in Land Plants. To understand the early evolution of TPSs in land plants, we created a dataset of TPS sequences enriched in those from nonseed plants. This included all TPSs identified from the transcriptomes of nonseed

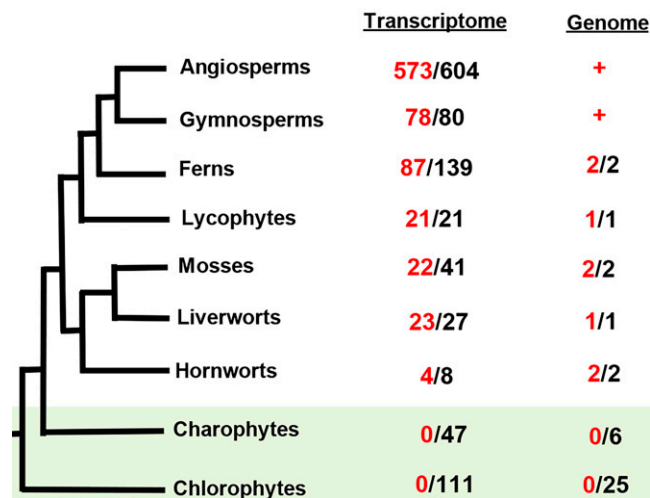


Fig. 2. Occurrence of *TPS* genes among green plants. The numbers represent the number of genomes or transcriptomes from which *TPS* genes were identified (in red) and the total number of genomes or transcriptomes that were analyzed in each class (in black). The phylogeny of the major lineages of land plants was drawn according to a recent large-scale phylogenomic analysis (4). "+" indicates the known ubiquitous occurrence of *TPS* genes in angiosperm and gymnosperm genomes.

plants in the OneKP database and the additional 69 fern transcriptomes. In addition, *TPS*s from eight nonseed plants with sequenced genomes (*SI Appendix, Table S3*) were included. Since the evolutionary relationship of the *TPS*s from seed plants is relatively well understood, only *TPS*s from three angiosperms (*Arabidopsis thaliana*, *Oryza sativa*, and *Amborella trichopoda*) and four gymnosperms (*Ginkgo biloba*, *Pseudotsuga menziesii*, *Picea abies*, and *Pinus lambertiana*) were included, plus several other sequences to anchor the various *TPS* subfamilies (*SI Appendix, Table S2*). To ensure the quality of phylogenetic reconstructions, our dataset was confined to sequences with a length longer than 350 amino acids.

The unrooted *TPS* phylogenetic tree made from our *TPS* dataset (Fig. 3 and *SI Appendix, Fig. S1*) has a number of deep divisions that could be separated in various ways, assuming no early gene loss. However, given the distribution of the *TPS* subfamilies in the major lineages of land plants combined with their biochemical and physiological function, three divisions were defined that correspond to the *TPS-c* subfamily, the *TPS-e/f* subfamily, and the rest of the *TPS* subfamilies (*h/d/a/b/g*). The separation of the rest of the subfamilies from the *TPS-c* and *-e/f* subfamilies has strong bootstrap support (96%), while the separation of *TPS-c* and *-e/f* is based on their previous definitions (i.e., groups of monofunctional class I *TPS*s are placed into *TPS-e/f*, while those with class II activity fall within *TPS-c*). To gain additional evidence for the observed phylogeny, two additional phylogenetic trees were reconstructed with longer *TPS* sequences. As shown in *SI Appendix, Fig. S2* (made with *TPS*s longer than 475 amino acids) and *SI Appendix, Fig. S3* (made with *TPS*s longer than 500 amino acids), the topologies of the phylogenetic trees are largely consistent. As such, our interpretation and discussion about *TPS* evolution in this work is mainly based on the phylogenetic tree presented in Fig. 3.

The *TPS-c* subfamily is uniquely present in all groups of land plants (Fig. 3). By definition, the *TPS-c* subfamily contains the extant examples of bifunctional CPSKs, which are functionally analogous to the ancestral *TPS* required for biosynthesis of gibberellins and related phytohormones (9, 16). In

addition to the bifunctional CPSKs, the *TPS-c* subfamily also includes large numbers of *TPS*s from nonseed plants, which must have radiated out from the ancestral CPSKs in this lineage. The majority of the newly identified *TPS*s in this study fall within this group. While the presence of a fern *TPS* (MON_CQPW_PTSP4) in a branch otherwise composed of all hornwort *TPS*s (Fig. 3) is somewhat surprising, the significance of the placement of this isolated sequence remains to be determined.

The *TPS-e/f* subfamily is anchored by the KSs from phytohormone biosynthesis. While these *TPS*s almost all still retain the ancestral $\gamma\beta\alpha$ tridomain architecture, they no longer exhibit class II diterpene cyclase activity and the associated DxDD motif. There is a deep division in the *TPS-e/f* subfamily between two distinct branches, which leaves the origins of the subfamily uncertain. The two branches are largely divided between seed and nonseed plants, indicative of parallel subfunctionalization within each lineage. However, inclusion of one *TPS* from hornworts in the branch otherwise containing only those from seed plants argues against this interpretation (Fig. 3). Indeed, the *TPS-e/f* subfamily in the phylogenetic tree made with *TPS*s longer than 500 amino acids (*SI Appendix, Fig. S3*) no longer exhibits such a deep division. The observed separation may stem from relaxed selection in the $\gamma\beta$ didomains

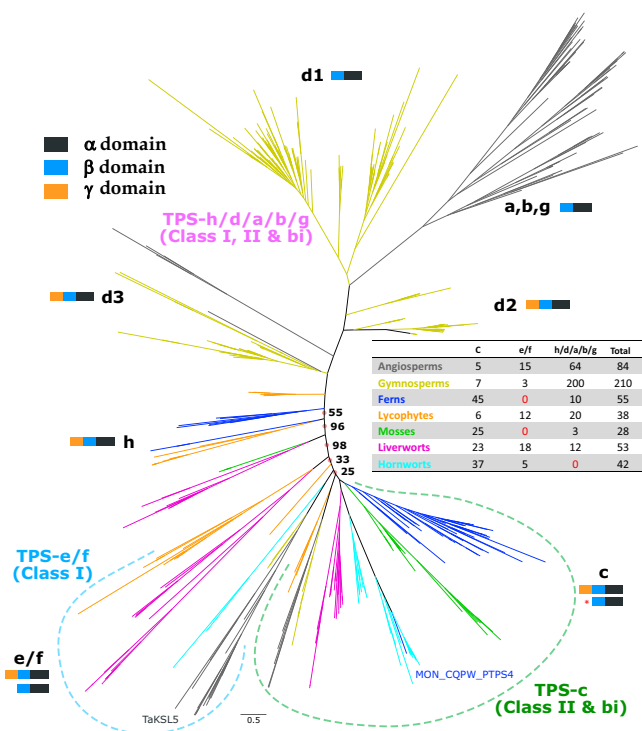


Fig. 3. Unrooted phylogenetic tree of typical plant *TPS* genes. The phylogeny is based especially on the nonseed plant sequences characterized here. The tree was constructed from *TPS*s identified from nonseed plants with transcriptomes, *TPS*s from nonseed plants with sequenced genomes and selected seed plants with sequenced genomes (*SI Appendix, Table S2*), and selected known *TPS*s (*SI Appendix, Table S4*), as indicated in the inset table. Genes are color-coded based on their source. The seven previously defined *TPS* subfamilies (a, b, c, d, e/f, g, and h) are indicated. These *TPS*s are further clustered into three lineages leading to the extant *TPS-c* subfamily, *TPS-e/f* subfamily, and *TPS-h/d/a/b/g* subfamilies. Each lineage contains *TPS* genes from bryophytes, lycophytes and/or ferns, gymnosperms, and angiosperms: The composition of domains (α , β , and γ) is noted for each subfamily. The presence of $\beta\alpha$ didomain *TPS*s in the *TPS-e/f* and *TPS-c* subfamilies is specified, with the red asterisk indicating the putative nature of these *TPS*s in the *TPS-c* subfamily. Bootstrap values are indicated for a few key branch points (all bootstrap values >50% are presented in *SI Appendix, Fig. S1*).

given the loss of the associated class II (CPS) activity in this subfamily. Regardless, TPSs from this subfamily do not seem to be present in either the mosses or ferns. This absence is correlated with the presence of bifunctional CPSKSs in these plant lineages [e.g., the moss PpCPSKS (17) and fern LjCPSKS (21)], which would preclude the need for a separate (monofunctional) KS. Yet in some plant lineages, monofunctional KSs of the TPS-e/f subfamily and bifunctional CPSKSs are both present, implying metabolic redundancy. A separate KS might indeed have occurred alongside a bifunctional CPSKS for an extended time prior to subfunctionalization of the CPSKS to a monofunctional CPS due to the ability of KS to react with *ent*-CPP released by the CPSKS from its class II diterpene cyclase active site. Analogous reasoning has been used to support the existence of monofunctional class I diterpene synthases alongside bifunctional class I and II enzymes in gymnosperm resin acid biosynthesis (22), where the release of the CPP intermediate has been shown (23).

The remaining TPS subfamilies form a distinct group termed here TPS-h/d/a/b/g. As none of these genes are known to be involved in phytohormone biosynthesis, they are all apparently dedicated to secondary metabolism. The TPS-h/d/a/b/g subfamilies have members in all plant lineages except hornworts, suggesting that the common ancestor of this TPS lineage arose from early gene duplication and neo-functionalization. This group also contains bifunctional diterpene synthases, but these are not involved in phytohormone biosynthesis. Nevertheless, since the bifunctional enzymes of the TPS-h/d/a/b/g subfamilies are present near the root of this group, it is assumed that they are derived from gene duplication of the ancestral bifunctional CPSKS with neo-functionalization to more specialized metabolism.

The TPS phylogenetic tree presented here clarifies the origin of the $\beta\alpha$ didomain architecture dominant in angiosperm TPSs, which results from the loss of the γ domain from the ancestral $\gamma\beta\alpha$ TPS. While it had been suggested that $\beta\alpha$ TPSs (which all have only class I activity) arose from the TPS-e/f subfamily, which also has only class I activity (2), analysis of the domain composition of the TPSs in the phylogenetic tree clearly indicates that this loss of the γ domain occurred in the lineage leading to the TPS-h/d/a/b/g subfamilies. In particular, this loss seems to have occurred early in the seed plant lineage, giving rise to the TPS-d1 group in gymnosperms and the TPS-a/b/g subfamilies in angiosperms. The loss of class II diterpene cyclase activity seems to have preceded γ domain loss since there is no class II activity in the TPS-a, b, g, d1, and d2 subfamilies (9). A branch of *Amborella* TPSs, which was previously postulated as the tentative TPS-x subfamily and placed between the TPS-h and TPS-d subfamilies (24), is here placed between the TPS-d3 and TPS-d2 groups (Fig. 3). This indicates that additional TPS sequences, especially those from basal flowering plants, are needed to resolve the early evolution of TPSs in seed plants.

Loss of the γ domain has independently occurred at least twice in the TPS-e/f subfamily (25) and also in the TPS-c subfamily. Though this loss frequently follows the loss of class II TPS activity, a parallel loss of the α domain following the loss of class I activity has not been observed, presumably due to the previously reported mutual structural interdependence of the α and β domains as well as their associated class I and class II activities (13).

Characterization of Representative TPSs from Nonseed Plants Suggests Dynamic Functional Evolution. Building on the hypothesis that all extant plant TPS genes were derived from the three ancestral lineages in the common ancestor of

land plants, the functional evolution of TPSs within and among the three groups was then investigated, with a total of 31 genes from nonseed plants selected for biochemical characterization (*SI Appendix, Table S5*). Given the expansion observed in the TPS-c subfamily, the majority of those selected here for characterization were from this subfamily, along with substantial numbers from the putative bifunctional TPSs in the TPS-h subfamily, enabling analysis of whether loss of class I or class II activity in these subfamilies is correlated with the loss of the relevant DDxxD and NSE/DTE or DxDD motifs.

Beyond the motifs required for essential TPS catalytic activity, additional motifs hypothesized to be specifically conserved in the CPSs and KSs involved in the biosynthesis of gibberellins and related *ent*-kaurene-derived phytohormones were also monitored. In the case of the CPSs, it has been shown that the catalytic base terminating bicyclization is a water molecule ligated in part by a histidine-asparagine dyad, conserved as LHS and PNV motifs (26, 27). In addition, it has been suggested that these motifs are susceptible to synergistic inhibition by GGPP and Mg^{2+} (28), which also serves as a cofactor for CPSs (29), as mediated by the presence of histidine at a particular position (30). In the case of KSs, it has been shown that formation of *ent*-kaurene is dependent on a particular isoleucine that is conserved in such TPSs (31, 32). Moreover, the KSs specifically involved in phytohormone biosynthesis are marked by a pair of threonines just upstream of the first of the two Mg^{2+} -binding motifs, which is then TTxxDDxxD, although the functional importance of these residues is unclear (33). Here, the correlation of these additional motifs with CPS and KS activity was also examined (*SI Appendix, Table S5*). For this purpose, the histidine hypothesized to impose synergistic GGPP/ Mg^{2+} inhibition on CPS was defined as falling within a FEHxW motif, while the key isoleucine for *ent*-kaurene production by KS was defined as falling within a PIx motif.

The TPSs chosen for study were functionally characterized by use of a previously reported modular metabolic engineering system (34) that enables facile coexpression with a GGPP synthase in *Escherichia coli* to provide the necessary substrate. For TPSs that exhibit only class I activity, the modular system can supply each of the three known stereoisomers of CPP (35). In total, 20 of the 31 TPS genes investigated here were demonstrated to encode functional diterpene synthases. The results of these studies are summarized in Table 1, with the underlying data reported as Supporting Information (*SI Appendix, Figs. S4–S20*). Note that class II diterpene cyclase products are detected as their corresponding dephosphorylated derivatives resulting from the activity of endogenous (*E. coli*) phosphatases. When the TPSs investigated here are indicated to produce the primary alcohol derivative of the phosphorylated intermediate, the dephosphorylation itself is not directly demonstrated and is only based on the presence of the Mg^{2+} -binding motifs required for class I activity (36) (Table 1, observed activity indicated with a number sign). Nevertheless, these TPSs are still considered to be bifunctional here. The 20 newly characterized diterpene synthases in this study, together with 13 previously characterized diterpene synthases from nonseed plants (*SI Appendix, Table S6*), are plotted to the TPS phylogenetic tree with their catalytic activity, domain architecture, and catalytic motifs (Fig. 4).

The members of the TPS-c subfamily investigated here were found to include bifunctional CPSKSs (Fig. 4), whose stereospecific production of *ent*-kaurene was verified by inactivation of their class I (KS) activity (via alanine substitution of the first two aspartates of the DDxxD motif) and coupling of the

Table 1. TPSs from nonseed plants characterized in this study

Gene name	Lineage*	Subfamily	Motifs [†]							Activity class [‡]	Products	
			LHS	PNV	FEHxW	DxDD	(TTxx)	DDxxD	PIx			NSE/DTE
PcCPS	Hornwort	c	✓	✓	✓	✓					II	<i>ent</i> -CPP
PcSS	Hornwort	c				✓			✓	✓	bi	sandaracopimaradiene
LcCPS	Liverwort	c	✓	✓	✓	✓					II	<i>ent</i> -CPP
OpKOS	Liverwort	c	LYS	✓	✓	✓	✓	DDxxE	✓	✓	bi [#]	kolavenol
PICPSKS	Liverwort	c	✓	✓	✓	✓	✓	✓	✓	✓	bi	<i>ent</i> -kaurene
MpCPS	Liverwort	c	✓	✓	✓	✓	✓				II	<i>ent</i> -CPP
OIIAS	Moss	c	✓	PSV	✓	✓		✓		✓	bi	(<i>ent</i> ?)isoabienol
LjuCPSKS	Moss	c	✓	✓	✓	✓	✓	✓	✓	✓	bi	<i>ent</i> -kaurene
SwCPS	Lycophyte	c	✓	✓	✓	✓					II	<i>ent</i> -CPP
AfKS	Fern	c	✓	✓		✓	✓	✓	✓	✓	bi	<i>ent</i> -kaurene
VaCPS	Fern	c	✓	✓	✓	✓					II	<i>ent</i> -CPP
VaPS	Fern	c	✓	PSV		✓		✓		✓	bi	<i>ent</i> -isopimara 8,15 diene
MpDTPS6	Liverwort	c				✓		✓		✓	bi [#]	<i>cis</i> -kolavenol
MpDTPS2	Liverwort	c				✓		✓		✓	bi [#]	kolavenol
MpDTPS7	Liverwort	c	LYS			✓		✓		✓	bi [#]	terpentedienol
MpDTPS5	Liverwort	h	✓			✓					II	terpentedienyl-PP
HsLS	Lycophyte	h	LYS	PCH		✓		✓		✓	bi	levopimaradiene
PdLS	Lycophyte	h	LYS	PCH		✓		✓		✓	bi	levopimaradiene
OscPS	Fern	h				✓					II	8-endo-CPP
OjMS	Fern	h						✓		✓	I	<i>syn</i> -manool

*Refer to *SI Appendix, Table S5* for the species from which the gene was obtained.

[†]For all motifs, "✓" indicates conservation, while red text indicates important changes discussed in text.

[‡]"I", "II", and "bi" depict class I, class II, and class I/class II bifunctional TPS, respectively, with green text indicating expected production of *ent*-CPP and/or *ent*-kaurene (CPS and/or KS activity, respectively) for phytohormone biosynthesis. "bi[#]" indicates observed activity for which dephosphorylation was not directly demonstrated.

remaining CPS activity to a stereospecific KS (37). Other TPSs from this subfamily were found to have subfunctionalized their ancestral CPSKS activities (Fig. 4). For example, subfunctionalization via loss of class I (KS) activity to yield a monofunctional CPS occurred early in the evolution of vascular plants (as evidenced here by the lycophyte TPS SwCPS), and this appears to have occurred independently in the plant lineages, leading to the extant liverworts (MpCPS, LcTPS and MpDTPS3) and hornworts (PcCPS) as well as ferns (VaCPS), indicating parallel evolution within this subfamily. In each case, these TPSs were found to stereoselectively produce *ent*-CPP, shown as previously described (38). Each of these CPS activities also retained the specific catalytic base dyad motif of the CPSKS ancestor and so are hypothesized to have kept the ancestral CPS activity. The histidine and associated motif for synergistic GGPP/Mg²⁺ substrate inhibition is also conserved in these newly identified nonseed CPSs, but this histidine residue is not present in the liverwort CPSKS identified here (PICPSKS), consistent with two other previously characterized CPS(KS)s as well, the gymnosperm PgCPS and moss PpCPSKS (17, 39). Thus, the wider import of this histidine remains uncertain, and the implications of its retention or loss for evolutionary derivation remains unclear. Regardless, all the CPSKSs identified here contain the ancestral class II catalytic base dyad motifs (i.e., LHS and PNV) identified in all previously characterized CPS(KS)s to date (26).

The subfunctionalized CPSs in ferns apparently arose independently of those found in lycophytes, gymnosperms, and angiosperms (Fig. 3), which is incongruent with the accepted evolutionary history of these plant taxa (Fig. 2). This may reflect the retention of a bifunctional CPSKS in ferns, as noted above. The subfunctionalization of CPS requires a KS activity, which could be provided by the bifunctional CPSKS. Alternatively, this might be provided by an independently arising KS.

For example, Asu_TPS1 is a KS whose sequence indicates that it falls within the TPS-c subfamily instead of the TPS-e/f subfamily, where all other plant KSs are found. Nevertheless, consistent with a retained function in phytohormone biosynthesis, this KS, along with all the bifunctional CPSKSs identified here, contains the KS specific motifs (i.e., PIx and the TTxx extension of DDxxD).

Some of the bifunctional TPSs of the TPS-c subfamily from liverworts and hornworts characterized here do not act as CPSKSs, suggesting that neofunctionalization occurred in this subfamily following the evolutionary event that founded the lineage leading to the TPS-h/d/a/b/g subfamilies. In most of these cases, the class II active site catalyzes rearrangement of the initially formed bicycle to yield kolavenyl diphosphate (KPP) of various stereochemical configurations. There is also an example of a monofunctional KPP synthase (OpKOS) that no longer contains the class I Mg²⁺-binding motifs. It can be hypothesized that the monofunctional KPP synthase (OpKOS) forms *ent*-KPP, as this contains a tyrosine in place of the histidine of the ancestral CPS catalytic base dyad, and such substitution has been previously shown to be sufficient to alter product outcome from *ent*-CPP to *ent*-KPP (40). Indeed, the presence of an aromatic residue at this position has been found to be predictive, as demonstrated by reverse engineering of both known *ent*-KPP synthases to produce *ent*-CPP instead (41, 42).

Another neofunctionalized TPS-c family member from the mosses (OIIAS) retains bifunctionality and produces isoabienol (Fig. 4). This requires production of a hydroxylated derivative of CPP by its class II active site, which retains the histidine but contains a serine in place of the asparagine from the ancestral CPS catalytic base dyad. Such substitution has been previously shown to be sufficient to alter product outcome from *ent*-CPP to 8β-hydroxy-*ent*-CPP (26), suggesting that this enzyme may produce *ent*-isoabienol.

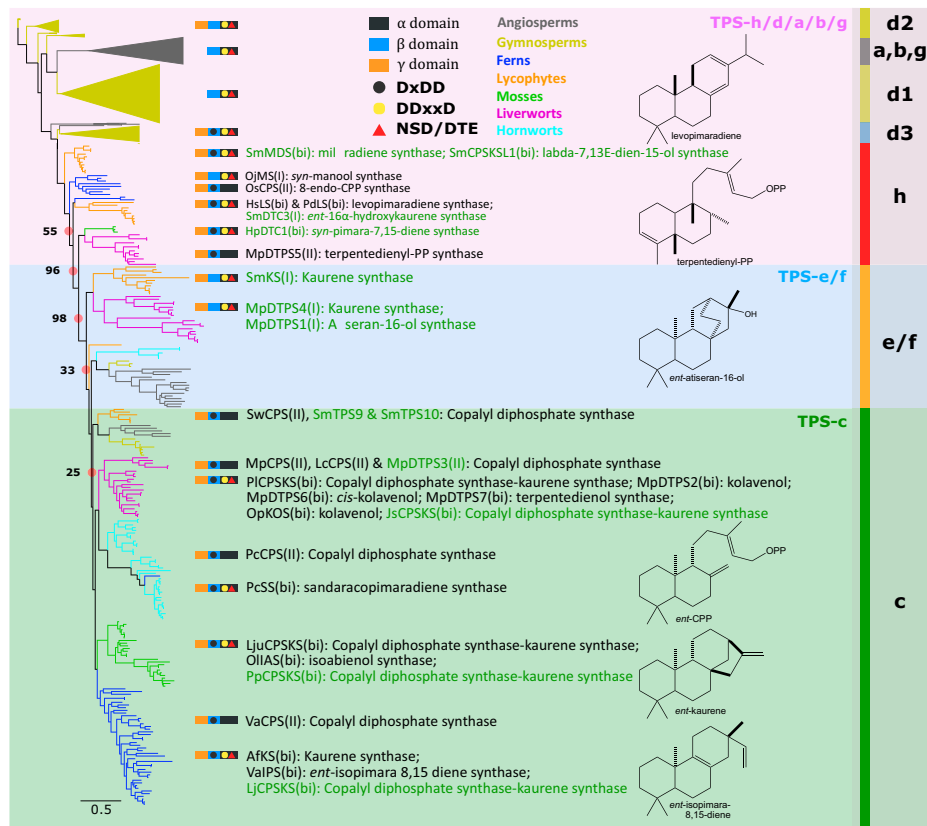


Fig. 4. Chief characteristics of newly described TPSs: phylogenetic placement, major products, land plant lineage, domain architecture, and motif composition. All gene names followed by functions in black are diterpene synthases functionally characterized in this study (Table 1 and *SI Appendix, Table S5*). Some previously characterized representatives of the TPS-c, -e/f, and -h subfamilies are also listed (in green) (*SI Appendix, Table S6*). Also presented are structures of selected diterpenes that are products of newly characterized TPSs. PP, diphosphate.

TPSs of the TPS-h/d/a/b/g subfamilies investigated here were all found to yield products other than the *ent*-CPP or *ent*-kaurene required for phytohormone biosynthesis, consistent with dedication of the TPS-h/d/a/b/g subfamilies to secondary metabolism. For example, two bifunctional levopimaradiene synthases were found in lycophytes, HsLS and PdLS, that exhibit intriguing parallels to the functionally analogous TPS-d3 group members from conifers involved in resin acid biosynthesis containing similar features. The catalytic base in the class II diterpene cyclase active site of gymnosperm resin acid TPSs is formed by direct hydrogen-bonding between the side chains of a tyrosine and histidine, where the tyrosine is from a LYS motif that replaces the ancestral (CPS) LHS, while the histidine is from a PCH motif that similarly replaces the ancestral (CPS) PNV (26, 43, 44). Both motifs are also present in these lycophyte levopimaradiene synthases. Moreover, a key alanine in the class I active site of the gymnosperm TPS-d3 subfamily members that plays an equivalent role to the key isoleucine in the ancestral KS but is found four residues upstream (45), is also conserved in these lycophyte levopimaradiene synthases, along with surrounding residues that form a VSIAL motif. Thus, it is tempting to hypothesize that levopimaradiene synthase activity evolved early in the vascular plant lineage. Alternatively, the same arrangement of key residues and surrounding motifs independently evolved in both the class II and class I active sites in TPSs from the lycophyte and gymnosperm lineages respectively.

Also notable among the TPS-h/d/a/b/g subfamilies members characterized here are two TPSs from a fern and a liverwort TPS, both of which have lost class I activity, consistent with their loss of the associated Mg^{2+} -binding motifs. Both then act as

monofunctional class II diterpene cyclases, with the fern enzyme (OsCPS) producing 8-*endo*-CPP, while the liverwort enzyme (MpDTPS5) produces rearranged KPP. These represent monofunctional class II diterpene cyclases outside the TPS-c subfamily, demonstrating that parallel evolutionary loss of class I activity has occurred.

Other TPS-h/d/a/b/g subfamilies members have lost class II activity and now function as monofunctional class I diterpene synthases (46). For example, a fern *syn*-manool synthase was characterized from the TPS-h subfamily that reacts with *syn*-CPP to produce *syn*-manool. Interestingly, this monofunctional class I TPS is not closely related to the previously described lycophyte 16 α -hydroxykaurene synthase, also a class I TPS from the TPS-h subfamily, and may reflect independent loss of class II activity. Such an evolutionary event in the lineage leading to the TPS-h/d/a/b/g subfamilies parallels the loss of class II activity from CPSKS that initiated the lineage leading to the TPS-e/f subfamily, as well as the equivalent losses of class II activity that occurred in the transition from TPS-d3 to TPS-d2 and within the TPS-d3 groups (22). Thus, the loss of class II activity is a repeated theme in TPS evolution, which then enables loss of the γ domain involved in such activity, as has now been observed in all three main groups of TPSs.

Conclusions

To understand the enormous skeletal diversity of terpenoid natural products present in plants, more knowledge of the evolution of the plant TPS family is required. The molecular,

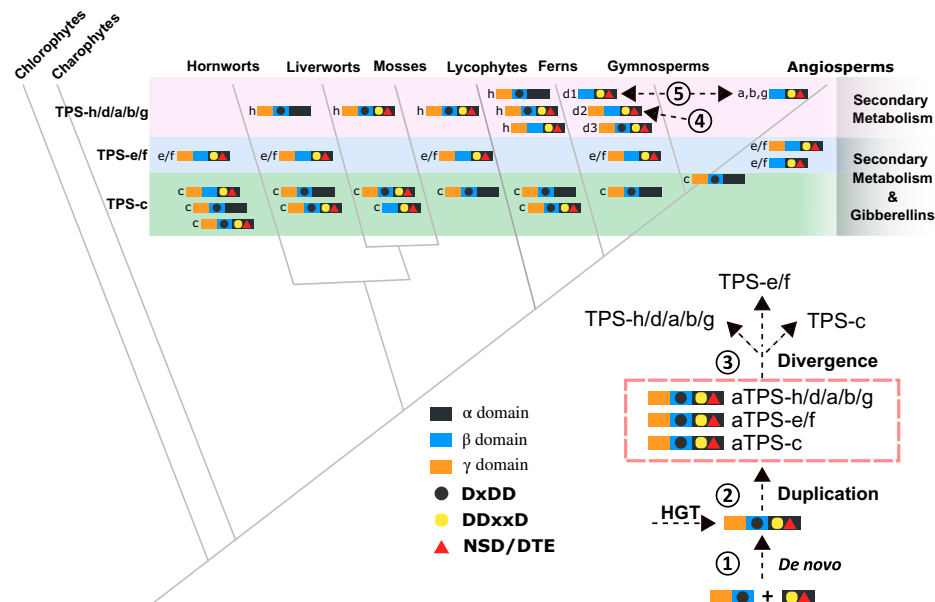


Fig. 5. A model for the origin and evolution of *TPS* genes in land plants. The phylogeny of green plants was drawn according to a recent large-scale phylogenomic analysis (4). The circled numbers 1 to 5 indicate the major events hypothesized for the evolution of plant *TPS* family, as described in *Conclusions*. The ancestral *TPS* gene may have originated through fusion of α and γ domain proteins indicated as “*De novo*.” Another possible mechanism is horizontal gene transfer, indicated as “HGT” from a bacterial host. “aTPS-c,” “aTPS-e/f,” and “aTPS-h/d/a/b/g” refer to the ancestor of the extant TPS-c, TPS-e/f, and TPS-h/d/a/b/g subfamilies, respectively.

phylogenetic, and biochemical analyses reported here allow insight into the major events in the origin of the *TPS* family (Fig. 5).

- 1) Our results are consistent with the hypothesis that the ancestral *TPS* encoded a bifunctional CPSKS producing the *ent*-kaurene precursor for the biosynthesis of gibberellins and related phytohormones. In particular, this is supported by the widespread retention of CPS- and KS-specific motifs throughout all land plants. The ancestral CPSKS originated in land plants from the probable fusion of a γ β didomain CPS and an α domain KS that occurred either in the common ancestor of land plants or in bacteria. In the latter case, this CPSKS would have been acquired by ancestral land plants through horizontal gene transfer.
- 2) The ancestral γ β α tridomain CPSKS appears to have undergone at least two gene duplication events early in land plant evolution.
- 3) How the three ancestral *TPS* lineages evolved from these two gene duplication events is unclear, as each could have been derived from either event. Nevertheless, together these gave rise to three ancient *TPS* lineages leading to the extant TPS-c, TPS-e/f, and TPS-h/d/a/b/g subfamilies. Although the TPS-c and TPS-e/f subfamilies contain the CPS(KS) and KSs, respectively, required for phytohormone biosynthesis, they also have given rise to numerous proteins involved in secondary metabolism. On the other hand, all members of the TPS-h/d/a/b/g subfamilies appear to be dedicated to secondary metabolism and represent by far the largest radiation in this area. Given the presence of bifunctional diterpene synthases in the TPS-h/d/a/b/g subfamilies from the earlier diverging plant lineages, this group almost certainly arose from duplication of the ancestral CPSKS. However, the retention of CPSKS even in the presence of the ancestral KS of the TPS-e/f subfamily leaves uncertain the relative timing for establishment of the TPS-e/f subfamily versus the TPS-h/d/a/b/g or TPS-c subfamilies. In part, the ancestral CPSKS retained bifunctionality at least until divergence of

the fern lineage but kept only CPS (class II) and lost KS (class I) activity in other major plant lineages (except the mosses), forming the TPS-c subfamily.

- 4) The ancestral bifunctional *TPS* in the TPS-h/d/a/b/g subfamilies underwent subfunctionalization early in vascular plant evolution (within the TPS-d subfamily), with loss of the class II activity giving rise to the TPS-d2 group.
- 5) This was followed by loss of the γ domain involved in class II activity, which gave rise to the TPS-d1 group as well as the angiosperm-specific TPS-a/b/g subfamilies.

Thus, it is now apparent that divergence of the ancestral *CPSKS* gene, both sub- and neo-functionalization, has independently occurred in various plant lineages. This has given rise to various examples of parallel evolution in separate *TPS* subfamilies resulting in functionally equivalent activity in unrelated lineages. A notable example is the occurrence of class I diterpene synthases that act on (normal) CPP to produce the cyclohexa-1,4-diene abietane miltiradiene, not only in the expected TPS-e/f subfamily but also in the TPS-a subfamily, which falls within the TPS-h/d/a/b/g lineage (47–49). Thus, throughout its evolution, the *TPS* family has been shaped by selective pressure for both conserved production of *ent*-kaurene for phytohormone biosynthesis as well as diversification of terpenoid natural product skeletons in lineage-specific fashion, resulting in the complex patterns of function observed today in land plants.

Materials and Methods

Sequence Retrieval and Identification of Terpene Synthase Genes. The OneKP transcriptome dataset was described previously (8). The additional transcriptomes of 69 fern species also have been previously reported (19). The sources of genomes analyzed in this study are listed in *SI Appendix, Table S1* (green algae) and *SI Appendix, Table S2* (land plants). A detailed description of various databases is provided in *SI Appendix, SI Materials and Methods*. The proteomes for all the datasets were searched against the Pfam-A database locally using HMMER 3.0 hmmsrch (50) with an E value of $1e-5$. Only sequences with best hits from the following two HMM profiles were considered as putative TPSs:

Terpene_synth_C (PF03936) and Terpene synthase N-terminal domain (PF01397). For sequences from the same species that have 100% identity, only the longest one was kept as the representative sequence to reduce redundancy. All the putative TPS sequences were subjected to BLASTP (51) search against the National Center for Biotechnology Information's nonredundant database using default parameters.

Phylogenetic Analyses. Sequences were aligned using MAFFT (einsi) with 1,000 iterations of improvement. ProtTest (52) was used to select the most appropriate protein evolution model for alignment under the Akaike information criterion. For the maximum likelihood analyses, RAxML (53) was used with 1,000 bootstrap replicates under the best substitution model (JTT+G+F).

Biochemical Characterization. For each TPS gene selected for biochemical characterization, the encoded protein was first analyzed using TargetP (<https://services.healthtech.dtu.dk/service.php?TargetP-2.0>) to predict the putative transit peptide. Then, with the sequence for putative transit peptide removed and a codon for Met added to the beginning of the coding sequence, a complementary DNA for each pseudomature TPS was synthesized and cloned into pEXP-5-CT/TOPO. The resulting expression construct was incorporated into a previously described modular metabolic engineering system (39) for expression of recombinant TPSs in *E. coli* and for assays of class I, class II, or class I/class II bifunctional diterpene synthase activities. Terpene products were identified using gas chromatography-mass spectrometry. The detailed procedure for TPS enzyme assays is provided in *SI Appendix, SI Materials and Methods*.

1. G. D. Moghe, B. J. Leong, S. M. Hurney, A. Daniel Jones, R. L. Last, Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife* **6**, e28468 (2017).
2. J. Zi, S. Mafu, R. J. Peters, To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. *Annu. Rev. Plant Biol.* **65**, 259–286 (2014).
3. J. Gershenzon, N. Dudareva, The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414 (2007).
4. J. H. Leebens-Mack *et al.*, One Thousand Plant Transcriptomes Initiative, One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
5. D. W. Christianson, Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442 (2006).
6. J. A. Aaron, D. W. Christianson, Trinuclear metal clusters in catalysis by terpenoid synthases. *Pure Appl. Chem.* **82**, 1585–1597 (2010).
7. G. Li *et al.*, Nonseed plant *Selaginella moellendorffii* [corrected] has both seed plant and microbial types of terpene synthases. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14711–14715 (2012).
8. Q. Jia *et al.*, Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12328–12333 (2016).
9. F. Chen, D. Tholl, J. Bohlmann, E. Pichersky, 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 (2011).
10. R. Cao *et al.*, Diterpene cyclases and the nature of the isoprene fold. *Proteins* **78**, 2417–2432 (2010).
11. Q. Jia, T. G. Köllner, J. Gershenzon, F. Chen, MTPSs: New terpene synthases in nonseed plants. *Trends Plant Sci.* **23**, 121–128 (2018).
12. Y. Gao, R. B. Honzato, R. J. Peters, Terpenoid synthase structures: A so far incomplete view of complex catalysis. *Nat. Prod. Rep.* **29**, 1153–1175 (2012).
13. R. J. Peters, O. A. Carter, Y. Zhang, B. W. Matthews, R. B. Croteau, Bifunctional abietadiene synthase: Mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. *Biochemistry* **42**, 2700–2707 (2003).
14. M. Köksal, H. Hu, R. M. Coates, R. J. Peters, D. W. Christianson, Structure and mechanism of the diterpene cyclase *ent*-copalyl diphosphate synthase. *Nat. Chem. Biol.* **7**, 431–433 (2011).
15. S. Prisic, J. Xu, R. M. Coates, R. J. Peters, Probing the role of the DXDD motif in class II diterpene cyclases. *ChemBioChem* **8**, 869–874 (2007).
16. D. Morrone *et al.*, Gibberellin biosynthesis in bacteria: Separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. *FEBS Lett.* **583**, 475–480 (2009).
17. K. Hayashi *et al.*, Identification and functional analysis of bifunctional *ent*-kaurene synthase from the moss *Physcomitrella patens*. *FEBS Lett.* **580**, 6175–6181 (2006).
18. J. Bohlmann, G. Meyer-Gauen, R. Croteau, Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4126–4133 (1998).
19. H. Shen *et al.*, Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *Gigascience* **7**, 1–11 (2018).
20. G. Wei *et al.*, Terpene biosynthesis in red algae is catalyzed by microbial type but not typical plant terpene synthases. *Plant Physiol.* **179**, 382–390 (2019).
21. J. Tanaka *et al.*, Antheridiogen determines sex in ferns via a spatiotemporally split gibberellin synthesis pathway. *Science* **346**, 469–473 (2014).
22. D. E. Hall *et al.*, Evolution of conifer diterpene synthases: Diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. *Plant Physiol.* **161**, 600–616 (2013).
23. R. J. Peters, M. M. Ravn, R. M. Coates, R. B. Croteau, Bifunctional abietadiene synthase: Free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.* **123**, 8974–8978 (2001).
24. A. G. Project, Amborella Genome Project, The Amborella genome and the evolution of flowering plants. *Science* **342**, 1241089 (2013).

Data Availability. DNA sequence data have been deposited in GenBank (accession Nos. [OL989431](https://doi.org/10.1093/ol989431)–[OL989450](https://doi.org/10.1093/ol989450)). All other study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Dr. Fay-Wei Li for making the genome and transcriptome assemblies of two fern species *Salvinia cucullata* and *Azolla filiculoides* available to us prior to their publication. We thank Dr. Chi Zhang for his assistance with compiling the genome sequences of green algae and transcriptome sequences of ferns. We thank Dr. Wangdan Xiong for her effort in the initial diterpene synthase activity screening for a subset of the selected TPSs. We also thank Dr. Yin-Long Qiu for his helpful discussion about phylogenetic tree interpretation. This work was supported by a grant from the NIH (Grant No. GM131885 to R.J.P.), an innovation grant from the Institute of Agriculture, The University of Tennessee (to F.C.), and the Max Planck Society (funds to T.G.K. and J.G.).

Author affiliations: ^aGraduate School of Genome Science and Technology, The University of Tennessee, Knoxville, TN 37996; ^bRoy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011; ^cDepartment of Biochemistry, Max Planck Institute of Chemical Ecology, 07745 Jena, Germany; ^dTea Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310008, China; ^eKey Laboratory of Tea Quality and Safety Control, Ministry of Agriculture and Rural Affairs, Hangzhou 310008, China; ^fDepartment of Plant Sciences, The University of Tennessee, Knoxville, TN 37996; ^gBGI-Shenzhen, Shenzhen 518083, China; ^hDepartment of Biological Sciences, University of Alberta, Edmonton, AB, Canada T6G 2E9; and ⁱDepartment of Medicine, University of Alberta, Edmonton, AB, Canada T6G 2E1

25. M. L. Hillwig *et al.*, Domain loss has independently occurred multiple times in plant terpene synthase evolution. *Plant J.* **68**, 1051–1060 (2011).
26. C. Lemke, K. C. Potter, S. Schulte, R. J. Peters, Conserved bases for the initial cyclase in gibberellin biosynthesis: From bacteria to plants. *Biochem. J.* **476**, 2607–2621 (2019).
27. K. Potter, J. Criswell, J. Zi, A. Stubbs, R. J. Peters, Novel product chemistry from mechanistic analysis of *ent*-copalyl diphosphate synthases from plant hormone biosynthesis. *Angew. Chem. Int. Ed. Engl.* **53**, 7198–7202 (2014).
28. S. Prisic, R. J. Peters, Synergistic substrate inhibition of *ent*-copalyl diphosphate synthase: A potential feed-forward inhibition mechanism limiting gibberellin metabolism. *Plant Physiol.* **144**, 445–454 (2007).
29. R. G. Frost, C. A. West, Properties of kaurene synthetase from *Marah macrocarpus*. *Plant Physiol.* **59**, 22–29 (1977).
30. F. M. Mann *et al.*, A single residue switch for Mg(2+)-dependent inhibition characterizes plant class II diterpene cyclases from primary and secondary metabolism. *J. Biol. Chem.* **285**, 20558–20563 (2010).
31. M. Xu, P. R. Wilderman, R. J. Peters, Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7397–7401 (2007).
32. M. Jia, R. J. Peters, Extending a single residue switch for abbreviating catalysis in plant *ent*-kaurene synthases. *Front. Plant Sci.* **7**, 1765 (2016).
33. R. Brown, M. Jia, R. J. Peters, A pair of threonines mark *ent*-kaurene synthases for phytohormone biosynthesis. *Phytochemistry* **184**, 112672 (2021).
34. A. Cyr, P. R. Wilderman, M. Determan, R. J. Peters, A modular approach for facile biosynthesis of labdane-related diterpenes. *J. Am. Chem. Soc.* **129**, 6684–6685 (2007).
35. R. J. Peters, Two rings in them all: The labdane-related diterpenoids. *Nat. Prod. Rep.* **27**, 1521–1530 (2010).
36. S. Mafu, M. L. Hillwig, R. J. Peters, A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from *Selaginella moellendorffii*. *ChemBioChem* **12**, 1984–1987 (2011).
37. M. Xu, M. L. Hillwig, M. S. Tiernan, R. J. Peters, Probing labdane-related diterpenoid biosynthesis in the fungal genus *Aspergillus*. *J. Nat. Prod.* **80**, 328–333 (2017).
38. Y. Wu *et al.*, Functional characterization of wheat copalyl diphosphate synthases sheds light on the early evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry* **84**, 40–46 (2012).
39. C. I. Keeling *et al.*, Identification and functional characterization of monofunctional *ent*-copalyl diphosphate and *ent*-kaurene synthases in white spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. *Plant Physiol.* **152**, 1197–1208 (2010).
40. K. C. Potter *et al.*, Blocking deprotonation with retention of aromaticity in a plant *ent*-copalyl diphosphate synthase leads to product rearrangement. *Angew. Chem. Int. Ed. Engl.* **55**, 634–638 (2016).
41. N. L. Hansen, J. N. Nissen, B. Hamberger, Two residues determine the product profile of the class II diterpene synthases TPS14 and TPS21 of *Tripterygium wilfordii*. *Phytochemistry* **138**, 52–56 (2017).
42. K. A. Pelot *et al.*, Biosynthesis of the psychotropic plant diterpene salvinorin A: Discovery and characterization of the *Salvia divinorum* clerodienyl diphosphate synthase. *Plant J.* **89**, 885–897 (2017).
43. S. Mafu *et al.*, Efficient heterocyclisation by (di)terpene synthases. *Chem. Commun. (Camb.)* **51**, 13485–13487 (2015).
44. K. Zhou *et al.*, Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from *Abies grandis*. *J. Biol. Chem.* **287**, 6840–6850 (2012).
45. P. R. Wilderman, R. J. Peters, A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. *J. Am. Chem. Soc.* **129**, 15736–15737 (2007).
46. M. Shimane *et al.*, Molecular evolution of the substrate specificity of *ent*-kaurene synthases to adapt to gibberellin biosynthesis in land plants. *Biochem. J.* **462**, 539–546 (2014).

47. N. L. Hansen *et al.*, The terpene synthase gene family in *Tripterygium wilfordii* harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. *Plant J.* **89**, 429–441 (2017).
48. F. S. Inabuy *et al.*, Biosynthesis of diterpenoids in *Tripterygium* adventitious root cultures. *Plant Physiol.* **175**, 92–103 (2017).
49. P. Su *et al.*, Identification and functional characterization of diterpene synthases for triptolide biosynthesis from *Tripterygium wilfordii*. *Plant J.* **93**, 50–65 (2018).
50. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–37 (2011).
51. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
52. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
53. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).