Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants

Qidong Jia[a], Guanglin Li[b,c], Tobias G. Köllner[d], Jianyu Fu[b,e], Xinlu Chen[b], Wangdan Xiong[b], Barbara J. Crandall-Stotler[f], John L. Bowman[g], David J. Weston[h], Yong Zhang[i], Li Chen[i], Yinlong Xie[i], Fay-Wei Li[j], Carl J. Rothfels[j], Anders Larsson[k], Sean W. Graham[l], Dennis W. Stevenson[m], Gane Ka-Shu Wong[i,n,o], Jonathan Gershenzon[d], and Feng Chen[a,b,1]

[a]Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN 37996; [b]Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996; [c]College of Life Sciences, Shaanxi Normal University, Xian 710062, China; [d]Department of Biochemistry, Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany; [e]Tea Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310008, China; [f]Department of Plant Biology, Southern Illinois University, Carbondale IL 62901; [g]School of Biological Sciences, Monash University, Melbourne, VIC 3800, Australia; [h]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [i]Beijing Genomics Institute-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China; [j]Department of Integrative Biology, University of California, Berkeley, CA 94720; [k]Systematic Biology, Uppsala University, 752 36 Uppsala, Sweden; [l]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; [m]Plant Genomics Program, New York Botanical Garden, Bronx, NY 10458; [n]Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada; and [o]Department of Medicine, University of Alberta, Edmonton AB T6G 2E1, Canada

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved August 29, 2016 (received for review May 26, 2016)

The vast abundance of terpene natural products in nature is due to enzymes known as terpene synthases (TPSs) that convert acyclic prenyl diphosphate precursors into a multitude of cyclic and acyclic carbon skeletons. Yet the evolution of TPSs is not well understood at higher levels of classification. Microbial TPSs from bacteria and fungi are only distantly related to typical plant TPSs, whereas genes similar to microbial *TPS* genes have been recently identified in the lycophyte *Selaginella moellendorffii*. The goal of this study was to investigate the distribution, evolution, and biochemical functions of microbial terpene synthase-like (*MTPSL*) genes in other plants. By analyzing the transcriptomes of 1,103 plant species ranging from green algae to flowering plants, putative *MTPSL* genes were identified predominantly from nonseed plants, including liverworts, mosses, hornworts, lycophytes, and monilophytes. Directed searching for *MTPSL* genes in the sequenced genomes of a wide range of seed plants confirmed their general absence in this group. Among themselves, MTPSL proteins from nonseed plants form four major groups, with two of these more closely related to bacterial TPSs and the other two to fungal TPSs. Two of the four groups contain a canonical aspartate-rich "DDxxD" motif. The third group has a "DDxxxD" motif, and the fourth group has only the first two "DD" conserved in this motif. Upon heterologous expression, representative members from each of the four groups displayed diverse catalytic functions as monoterpene and sesquiterpene synthases, suggesting these are important for terpene formation in nonseed plants.

terpene synthase | specialized metabolism | nonseed plant | gene evolution

Terpenoids are the largest class of land plant secondary metabolites, but they are not uniformly distributed in the plant kingdom (1). Many seed plants (angiosperms and gymnosperms) produce terpenoids of diverse types in large quantities. However, among nonseed plants, only liverworts are known as copious producers of terpenoids (2). Terpenoids have diverse biological and ecological functions with many serving as chemical defenses against herbivores and pathogens (3, 4). Some have lineage-specific functions, such as the volatile terpenoids in flowers that are involved in attracting pollinators (5). Characterizing the biosynthesis of terpenoids in all plant lineages is therefore an important avenue to understanding their roles in the adaptation of various lineages of terrestrial plants.

Terpene synthases (TPSs) are pivotal enzymes for terpenoid biosynthesis, forming a distinctive superfamily based on both sequence identity and structure classification. However, within this group, typical plant and microbial (bacterial and fungal) TPSs share very low sequence similarity and are therefore only distantly related (6). The typical plant TPSs form subfamilies with individual subfamilies generally associated with specific biochemical functions, such as monoterpene, sesquiterpene, or diterpene biosynthesis (7, 8). Monoterpene synthases and sesquiterpene synthases have been proposed to have evolved independently in gymnosperms and angiosperms from diterpene synthase ancestors (7, 9). Interestingly, the typical plant TPSs in the moss *Physcomitrella patens* (10) and the lycophyte *Selaginella moellendorffii* (11–13), two nonseed plants, were found to be of the diterpene synthase type. Therefore, the molecular basis underlying the biosynthesis of monoterpenes and sesquiterpenes identified in nonseed plants has long been unclear.

Recently, microbial terpene synthase-like (*MTPSL*) genes were identified in *S. moellendorffii* that encode monoterpene and sesquiterpene synthases (13). Unlike typical plant TPSs, which are composed of either three domains (αβγ) or two domains (αβ) (14, 15), MTPSLs contain only an α-domain. Phylogenetic analysis indicated that MTPSLs from *S. moellendorffii* are more closely related to microbial TPSs, in particular fungal TPSs, than to typical plant TPSs (13). So far, MTPSLs have only been identified in *S. moellendorffii* (13), raising intriguing questions

Significance

Terpenoids are ubiquitous products made by land plants with diverse biological functions. Their formation in seed plants is catalyzed by typical plant terpene synthases (TPSs), a well-characterized group of enzymes. In contrast, our knowledge of terpenoid biosynthesis in nonseed plants is very limited. By systematically analyzing the transcriptomes and/or genomes of more than 1000 plant species, we report that microbial terpene synthase-like genes, which are only distantly related to typical plant *TPS* genes, are widely distributed in nonseed plants, but virtually absent in seed plants. The study provides insights into the evolution *of TPS* genes in early land plants and opens the door to investigating the diversity and functions of terpenoids in nonseed plants.

Author contributions: Q.J., G.L., T.G.K., J.G., and F.C. designed research; Q.J., G.L., T.G.K., J.F., X.C., W.X., and F.C. performed research; B.J.C.-S., J.L.B., D.J.W., Y.Z., L.C., Y.X., F.-W.L., C.J.R., A.L., S.W.G., D.W.S., and G.K.-S.W. contributed new reagents/analytic tools; Q.J., G.L., T.G.K., J.G., and F.C. analyzed data; and Q.J., T.G.K., J.G., and F.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences for the biochemically characterized MTPSLs reported in this paper have been deposited in the GenBank database (accession nos. KX230835–KX230843).

[1]To whom correspondence should be addressed. Email: fengc@utk.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1607973113/-/DCSupplemental.

about the origin, evolution, and function of this type of plant terpene synthase genes. The goal of this study was to investigate the distribution of *MTPSL* genes in the green plants, infer their evolution, and determine their biochemical functions.

Results and Discussion

Terpene Synthase Genes of the Microbial Type Are Widespread in the Transcriptomes of Nonseed Land Plants, but Not in Green Algae and Seed Plants. To determine the distribution of *MTPSL* genes in the green plants, the transcriptomes of 1,103 species (*SI Appendix*, Table S1) of green plants (779 species of seed plants, 166 species of nonseed land plants, 47 species of charophytes, and 111 species of chlorophytes) generated from the 1,000 Plant (OneKP) initiative (https://sites.google.com/a/ualberta.ca/onekp/) were searched for microbial type terpene synthase genes using a HMMER method as previously described (13). A total of 712 *MTPSL* genes were identified from the transcriptomes of 146 species. Strikingly, the vast majority of *MTPSL* genes (706 of the 712 *MTPSL* genes or 99.2%) were found in the transcriptomes of nonseed land plants (Fig. 1).

Bryophytes consist of three lineages: hornworts, mosses, and liverworts, which have 7, 41, and 26 species in the OneKP dataset, respectively. The number of hornwort, moss, and liverwort species whose transcriptomes contain *MTPSL* genes was 3, 30, and 24, respectively. Among the 22 lycophyte species, 21 possessed *MTPSL* genes in their transcriptomes. For monilophytes, 65 of the 70 species were found to contain *MTPSL* genes in their transcriptomes. The median number of *MTPSL* genes from the transcriptome of each species for hornworts, mosses, liverworts, lycophytes, and monilophytes was 0, 1, 8, 3, and 4.5, respectively (Fig. 1 and *SI Appendix*, Table S2). Among all species, the monilophyte *Cystopteris utahensis* (a tetraploid) was found to contain the most *MTPSL* genes with 20 members (*SI Appendix*, Table S1).

On the other hand, extremely low occurrences of *MTPSL* genes were found in the transcriptomes of seed plants and charophytes. Among the 779 species of seed plants, only two species, *Phytolacca bogotensis* and *Opuntia* sp., both members of the Caryophyllales, were found to contain *MTPSL* genes in their transcriptomes with one and four members, respectively. Among the 47 species of charophytes, only one species, *Micrasterias fimbriata*, contained a *MTPSL* gene (one member) in its transcriptome. No *MTPSL* genes were found in the transcriptomes of 111 species of chlorophytes.

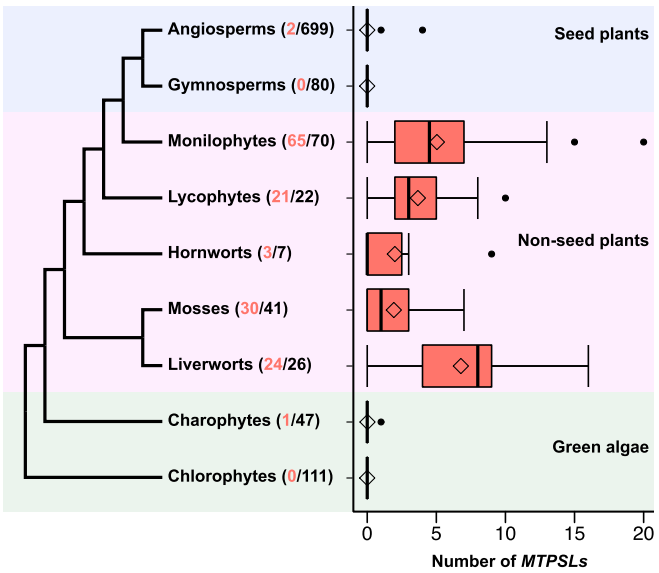

Fig. 1. Distribution of *MTPSL* genes identified from the transcriptomes of 1,103 plant species. The numbers in parentheses represent the number of transcriptomes containing putative *MTPSLs* (in red) and total transcriptomes analyzed in each lineage (in black). The phylogeny of green plants presented was modified from refs. 25 and 26. Each boxplot represents the number of MTPSLs found for individual species in each plant lineage. The solid black lines denote the median number of *MTPSLs* from each species. Whiskers represent 1.5 times the quantile of the data. Points outside of the range of the whiskers are outliers.

The Majority of *MTPSL* Genes Identified in Plant Transcriptomes Belong to Four Groups Clustered with Either Fungal or Bacterial Terpene Synthases. Phylogenetic analysis was performed for the 712 *MTPSL* genes identified from plant transcriptomes together with the 48 known *MTPSL* genes from *S. moellendorffii* (13) and selected terpene synthase genes from bacteria and fungi. The resulting phylogenetic tree indicates that the distribution of *MTPSL* genes in nonseed plants exhibits lineage-specific characteristics and that the majority of them (690 of 712) were clustered into four major groups with either similarity to bacterial TPS (groups I and II) or fungal TPS (groups III and IV) (Fig. 2).

Group I, the second largest *MTPSL* gene group, contains about 86% of *MTPSL* genes (152 of 177) from 23 species of liverworts, 34% of *MTPSL* genes (27 of 79) from 10 species of mosses, and 28% of *MTPSL* genes (23 of 83) from 9 species of lycophytes. Group II was composed of *MTPSL* genes primarily from mosses (about 66% of *MTPSL* genes from 24 species) and hornworts (50% of *MTPSL* genes from all three species in which *MTPSL* genes have been found). There was also one group II *MTPSL* gene found in a liverwort species, *Scapania nemorea*. Members of *MTPSL* genes in this species were also present in groups I and III. Group III, the smallest group, contains 4 *MTPSL* genes from 3 species of hornworts and 14 *MTPSL* genes from 7 species of liverworts. These 18 genes were clustered with the fungal trichodiene synthase (Tri5) genes. Group IV contains almost all *MTPSL* genes found from 65 species of monilophytes (352 of 353) and about 70% of *MTPSL* genes in lycophytes (58 of 83). The known *MTPSL* genes from *S. moellendorffii* were closely related to *MTPSL* genes from the transcriptomes of other lycophyte species.

Twenty-two *MTPSL* genes that lie outside of the four major groups were designated as "unclassified" (*SI Appendix*, Table S3). For example, the *MTPSL* genes found in two seed plant species, *Opuntia* sp. (four members) and *P. bogotensis* (one member), and one green alga, *M. fimbriata* (one member), were included in this list (Fig. 2 and *SI Appendix*, Table S3).

The Majority of *MTPSL* Genes Identified from Plant Transcriptomes Are Genuine Plant Genes. The putative *MTPSL* genes identified in the plant transcriptomes could have one of two possible origins: from plants or from plant-associated microbes. Three lines of evidence were used to judge that the vast majority, if not all, of these genes in the four major *MTPSL* clades are plant genes.

The first line of evidence comes from the analysis of putative *MTPSL* genes from axenic culture. The liverwort *S. nemorea* was selected for this purpose. A total of eight putative *MTPSL* genes were identified in the transcriptome of *S. nemorea* (*SI Appendix*, Table S4), belonging to group I (five genes), group II (one gene), group III (one gene), and unclassified (one gene). An axenic culture of *S. nemorea* was initiated by germinating isolated spores, using sterile culture methods. This culture was therefore free of contamination of endophytic microbes. We extracted genomic DNA from axenically cultured *S. nemorea* and used PCR to amplify DNA fragments for each of the eight putative *MTPSL* genes, and the results were compared with those obtained from the transcriptome analysis. Six of the eight *MTPSL* genes (five from group I and one from group III) were amplified and confirmed by sequencing. However, the amplification of the members from group II and the unclassified group failed. The

PLANT BIOLOGY

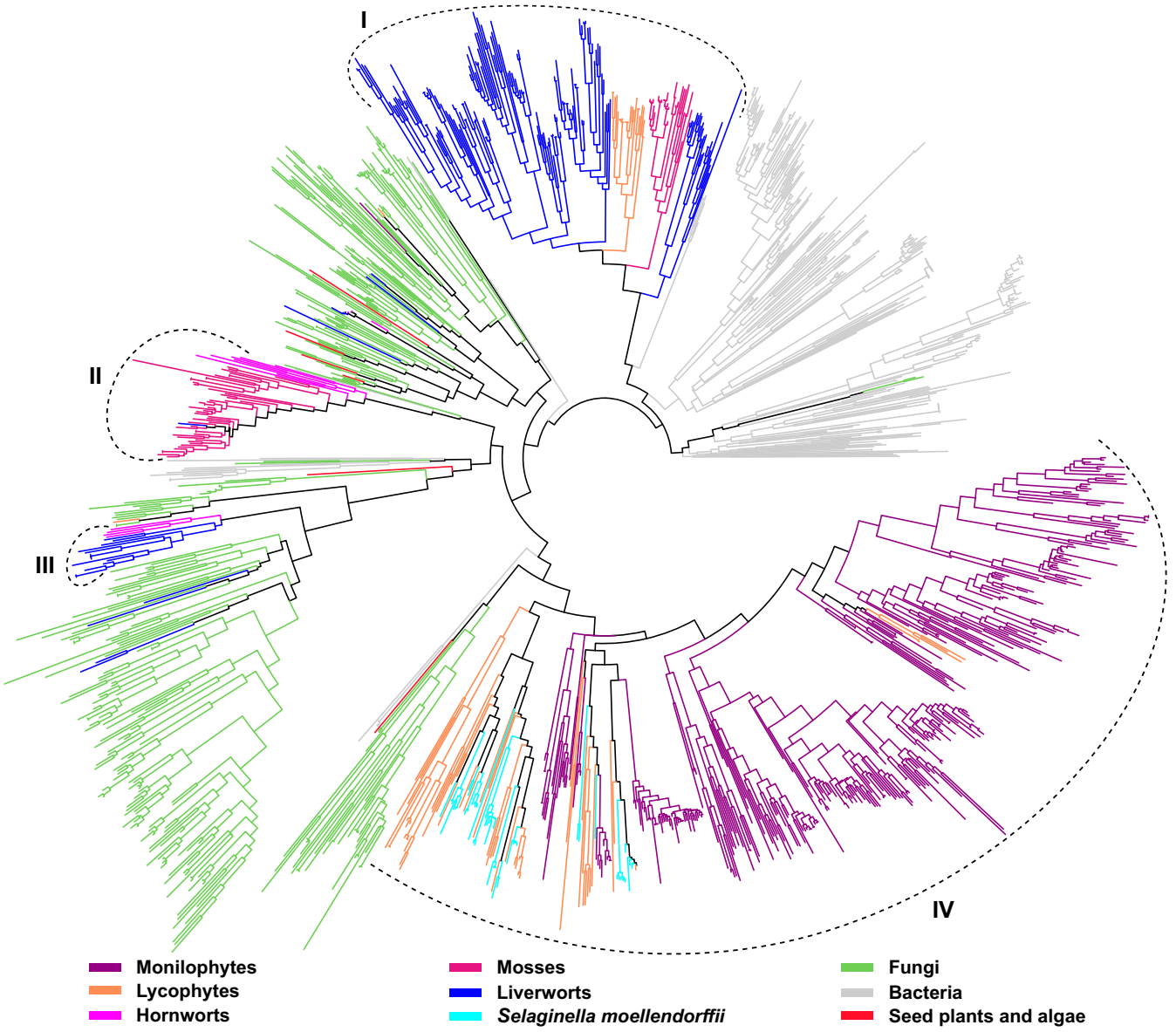

Fig. 2. Phylogeny of MTPSLs identified from OneKP with known MTPSLs from *S. moellendorffii*, bacterial TPSs, and fungal TPSs. Genes are color-coded based on their source. The majority of the newly identified MTPSLs are clustered into four major groups (I–IV).

member from the unclassified group was suggested to be a contaminant from endophytic fungi (*SI Appendix*, Fig. S1). Analysis of the OneKP transcriptomes has suggested that *S. nemorea* was contaminated with an unknown source of plant material (https://pods.iplantcollaborative.org/wiki/display/iptol/Sample+source+and+purity). The putative *MTPSL* from *S. nemorea* in group II is the only *MTPSL* from liverworts to be assigned to this group, suggesting that it was also derived from contamination by other plant material. Overall, this experimental study confirms that the group I and group III MTPSLs from the *S. nemorea* transcriptome are endogenous *S. nemorea* genes.

The second line of evidence for the plant origin of the majority of *MTPSL* genes reported is the identification of their immediate genomic neighbors as bona fide plant genes. We obtained such evidence for group II and III genes of the hornwort *Anthoceros punctatus*, whose genome has been recently sequenced (13). Assembling the raw genome sequence and then searching for *MTPSL* genes using the HMMER search, seven putative full-length *MTPSL* genes were identified; two of the seven genes belong to group II and the remaining five to group III (*SI Appendix*, Table S5). *ApMTPSL1* from group II and *ApMTPSL2* from group III were then selected as representatives for the following study. In the assembled genome of *A. punctatus*, *ApMTPSL1* is the immediate neighbor to a cytochrome P450 gene, for which the most similar gene is from the plant *Oryza brachyantha* (Fig. 3*A*). Similarly, *ApMTPSL2* is the immediate neighbor to a leucine-rich receptor-like kinase gene, for which the most similar gene is from the plant *Theobroma cacao* (Fig. 3*B*). We extracted genomic DNA from *A. punctatus* grown in axenic culture and performed PCR to amplify the coding sequence of *ApMTPSL1* and *-2* and their respective neighbors. The amplified DNA fragment was cloned and fully sequenced confirming that both *ApMTPSL1* (Fig. 3*A*) and *ApMTPSL2* genes (Fig. 3*B*) reside in the *A. punctatus* genome and are neighbors to a plant gene. Similar evidence for a group I gene was obtained from the moss *Sphagnum fallax*. The sequenced *S. fallax* genome contains 21 *MTPSL* genes, all of which belong to group I (*SI Appendix*, Table S6). A representative *MTPSL* gene (*SfMTPSL1*)

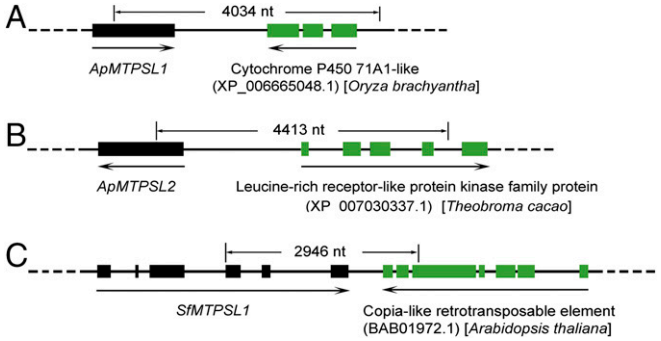

Fig. 3. Validation that representative *MTPSL* genes are of plant origin. Schematic genomic organization of representative *MTPSL* genes with their neighboring genes is shown. The genomic region spanning each *MTPSL* gene, its neighboring gene, and the intergenic region were amplified using PCR from genomic DNA and confirmed by sequencing. (*A*) *ApMTPSL1* with its neighboring gene annotated as a cytochrome P450 71A1-like from the hornwort *A. punctatus*. (*B*) *ApMTPSL2* and its neighboring gene annotated as a leucine-rich receptor-like kinase family protein from the hornwort *A. punctatus*. (*C*) *SfMTPSL1* with its neighboring gene annotated as a Copia-like retrotransposon element from the moss *S. fallax*. In each schematic figure, underneath the neighboring gene, the ID in parentheses represents the accession number of the best hit of this gene by searching the nonredundant database of the National Center for Biotechnology Information. The plant species indicates from which species the best hit was identified.

from *S. fallax* was confirmed to reside in its genome and neighbor with a Copia-like retrotransposable element, to which the most similar sequence is from the plant *Arabidopsis thaliana* (Fig. 3*C*). The genes in group IV form a clade with *SmMTPSL* genes from *S. moellendorffii*, which have already been determined to be plant genes (13). This evidence that representative *MTPSL* genes of groups I–IV originated from plant genomes supports that their apparent orthologs/homologs in each group are also of plant origin.

The third line of evidence for plant origin is that the overall evolutionary relationships of MTPSLs are largely congruent with the relationships of the land plant species from which the *MTPSL* genes are identified. The *MTPSL* genes in each of the four groups are inferred to share a common evolutionary origin (Fig. 2). Within each group, the MTPSLs from the same plant lineage showed higher sequence similarity than with MTPSLs from a different plant lineage. For instance, in group I the MTPSLs from three plant groups, liverworts, mosses, and lycophytes, form three distinct subclades (Fig. 2). The MTPSLs from closely related species are often also most closely related. For instance, the analyzed mosses included three species from the same genus *Sphagnum*: *Sphagnum lescurii*, *Sphagnum palustre*, and *Sphagnum recurvatum*. In group I, the MTPSLs from these three species reside in a clade (Fig. 2). Such fine-scale correlations of MTPSL sequence similarity and plant phylogeny also support a plant origin for most of these genes.

On the other hand, the 22 unclassified *MTPSL* genes (*SI Appendix*, Table S3) have a high probability of being derived from plant-associated microbes because their similarities with microbial TPS genes are extremely high (*SI Appendix*, Table S3), as in the example of the *S. nemorea* gene mentioned above, and so were not considered further in this study. Nevertheless, some of these genes may have been obtained from microbes very recently through horizontal gene transfer (HGT), which will be a subject of future investigation.

***MTPSL* Genes Are Patchily Distributed in Green Plants: Evolutionary Implications.** The confirmation that the vast majority of *MTPSL* genes identified from the OneKP transcriptomes are plant genes indicates that *MTPSL* genes occur widely in nonseed plants. Group I contains *MTPSL* genes from liverworts, mosses, and lycophytes (Fig. 2), which implies the presence of *MTPSL* genes in the common ancestor of land plants. However, our survey found that *MTPSL* genes are generally absent from the transcriptomes of green algae (Fig. 1). To provide further evidence on the presence/absence of *MTPSL* genes in green algae, we conducted a focused search on sequenced genomes for six species of chlorophytes and one species of charophyte (*SI Appendix*, Table S7). No *MTPSL* genes were detected in these sequenced green algae. The absence of *MTPSL* genes in the transcriptomes of a wide range of chlorophytes and the genome of the charophyte, *Klebsormidium flaccidum* (16) suggests that *MTPSL* genes may have their origin in an ancestral land plant rather than an algal ancestor. Broader genome sampling from green algae, especially charophytes, is needed to test this hypothesis. Nonetheless, the evolution of *MTPSL* genes may have been associated with the transition of plants from aquatic to terrestrial habitats. The pioneer land plants faced a harsh environment replete with many new biotic and abiotic stresses. Many products of TPSs are volatile hydrocarbons that may be more useful in a terrestrial environment than in an aquatic one.

MTPSLs from nonseed land plants exhibited different degrees of relatedness to bacterial TPSs and fungal TPSs. Group I is most closely related to bacterial TPSs, whereas groups III and IV are most closely related to fungal TPSs. Group II is most closely related to a number of bacterial TPSs, which, however, reside within a fungal clade (Fig. 2). These patterns suggest a complex evolutionary history of microbial type TPSs. Whereas it is possible that microbial-type *TPS* genes are ancestral in all kingdoms of life and took different evolutionary trajectories, their confinement to bacteria, fungi, and plants implies HGT. Assuming that bacterial and fungal *TPS* genes are ancestral to *MTPSLs*, the distribution pattern of *MTPSLs* can be explained by multiple HGT events from bacteria and fungi. However, it is premature to make strong claims about the donors and recipients of such transfer events because our understanding of phylogenetic relationships of TPSs in bacteria and fungi is still very limited. A better understanding of relationships among *TPS* genes in bacteria and fungi will allow testing of this hypothesis.

The absence of *MTPSL* genes in seed plants is also notable (Fig. 1). To gain additional evidence about the presence/absence of *MTPSL* genes in seed plants, we analyzed the genomic sequences of 48 species of seed plants (*SI Appendix*, Table S7): no *MTPSL* genes were identified in them. As mentioned previously, land plants contain typical plant TPSs, which catalyze similar biochemical reactions for the production of terpenoids as do MTPSLs, but typical plant TPSs are only distantly related to MTPSLs (13). However, in the nonseed plants that have been studied, the typical plant TPSs function as diterpene synthases (7) rather than the full range of monoterpene synthases, sesquiterpene synthases, and diterpene synthases found in seed plants. Several *MTPSL* genes from *S. moellendorffii* have been demonstrated to encode monoterpene synthases and sesquiterpene synthases (13), and we hypothesized that most *MTPSL* genes in nonseed plants function in this way.

Representative *MTPSL* Genes Encode Active TPSs with Diverse Catalytic Activities. The presence of *MTPSL* genes only in nonseed land plants poses an intriguing question about their functions. In seed plants, TPSs are responsible for the production of a diversity of terpenoids important for ecological interactions, especially as defenses against herbivores and pathogens. Some nonseed plants, such as liverworts (17) and mosses (18), also produce a vast diversity of terpenoids. However, little is known about how such terpenoids are synthesized or about their biological functions.

In general, TPSs contain two highly conserved motifs: the DDxxD and NSD/DTE motifs, which are both involved in substrate binding (14, 15, 19). Whereas the NSD/DTE motif is

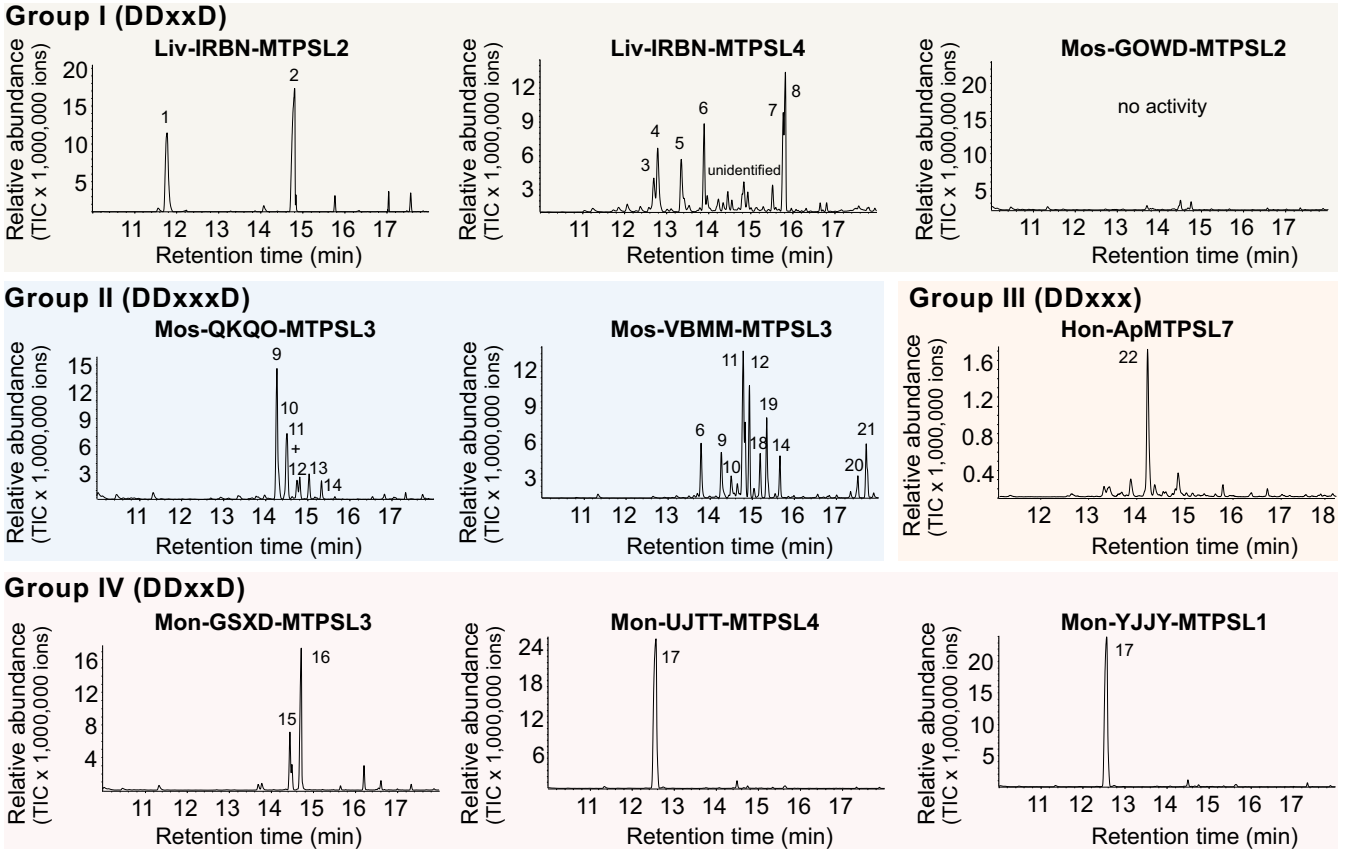

Fig. 4. Representative *MTPSL* genes from four groups (I–IV) encode active TPSs. The aspartate-rich motif associated with each group is indicated. *MTPSL* genes were heterologously expressed in *E. coli* and crude protein extracts were incubated with the potential substrate farnesyl diphosphate [(*E,E*)-FPP]. Enzyme products were collected using solid-phase microextraction and analyzed with gas chromatography/mass spectrometry. 1, bicycloelemene; 2, bicyclogermacrene; 3, α-isocomene; 4, β-elemene*; 5, (*E*)-β-caryophyllene*; 6, (*E*)-β-farnesene*; 7, nerolidol*; 8, dactylol; 9, γ-curcumene; 10, α-zingiberene; 11, β-bisabolene*; 12, β-curcumene; 13, sesquiphellandrene*; 14, (*E*)-α-bisabolene; 15, (*Z,E*)-α-farnesene; 16, (*E,E*)-α-farnesene; 17, protoillud-6-ene*; 18, (*Z*)-γ-bisabolene; 19, (*E*)-γ-bisabolene; 20, β-bisabolol; 21, α-bisabolol; and 22, β-acoradiene. Compounds marked with an asterisk (*) were identified using authentic standards. The origin of each *MTPSL* gene is listed in *SI Appendix*, Table S8. Ap, *A. punctatus*; GOWD, *S. lescurii*; GSXD, *M. eatonii*; Hon, hornwort; IRBN, *S. nemorea*; Liv, liverwort; Mon, monilophyte; Mos, moss; QKQO, *Pseudotaxiphyllum elegans*; UJTT, *Pityrogramma trifoliate*; VBMM, *Anomodon rostratus*; and YJJY, *Woodsia scopulina*.

highly conserved in the MTPSLs, the aspartate-rich DDxxD motif exhibits variations (*SI Appendix*, Fig. S2). Group I and IV proteins contain the canonical DDxxD motif, but group II proteins displayed a conserved DDxxxD motif. In the group III proteins, only the first two aspartates (DD) are conserved.

To gain an initial assessment of the biochemical functions of MTPSLs, a total of nine genes representing the four MTPSL groups (*SI Appendix*, Table S8) were selected for experimental work. Recombinant MTPSLs produced from *Escherichia coli* were tested for TPS activity with geranyl diphosphate [(*E*)-GPP], farnesyl diphosphate [(*E,E*)-FPP], and geranylgeranyl diphosphate [(*E,E,E*)-GGPP], the substrates for monoterpenes, sesquiterpenes, and diterpenes, respectively. With the exception of Mos-GOWD-MTPSL2, all tested enzymes were able to convert (*E,E*)-FPP into individual sesquiterpenes or complex sesquiterpene mixtures (Fig. 4). Whereas most of the MTPSL sesquiterpene products are also known as products from typical plant TPSs, Mon-UJTT-MTPSL4 and Mon-YJJY-MTPSL1, which are closely related to fungal TPS, produced protoillud-6-ene, a sesquiterpene that has only been reported from a fungus so far (20). Mos-GOWD-MTPSL2 showed exclusively monoterpene synthase activity and converted (*E*)-GPP into (*Z*)-β-ocimene and some minor monoterpene products (*SI Appendix*, Fig. S3). In addition to sesquiterpene synthase activity, Liv-IRBN-MTPSL2, Mon-UJTT-MTPSL4, and Hon-Ap-MTPSL7 were able to produce monoterpenes from (*E*)-GPP and Liv-IRBN-MTPSL4 and Mos-VBMM-MTPSL3 were able to convert (*E,E,E*)-GGPP into diterpenes (*SI Appendix*, Fig. S4). Recently, it was shown that certain typical plant TPSs naturally use *cis*-prenyl diphosphates as substrates for terpene production (21). Because most of the representative MTPSLs tested showed sesquiterpene synthase activity using (*E,E*)-FPP (Fig. 4), we performed additional assays to determine whether (*Z,E*)-FPP and (*Z,Z*)-FPP, the *cis*-isomers of FPP, could serve as substrates for these enzymes. Mos-GOWD-MTPSL2 was inactive with either substrate, whereas all of the other eight characterized MTPSLs showed activity with either (*Z,E*)-FPP or (*Z,Z*)-FPP or both (*SI Appendix*, Fig. S5). It is interesting to observe that some enzymes, such as Mos-QKQO-MTPSL3, produced the same products using the *cis*-FPP isomers as with (*E,E*)-FPP, whereas others such as Mon-UJTT-MTPSL4 produced different products (Fig. 4 and *SI Appendix*, Fig. S5). Nevertheless, the fact that the *Myriopteris eatonii* UJTT-MTPSL4 forms a single distinctive product from (*E,E*)-FPP and broad mixtures of over 10 products from each of the *cis*-FPP isomers suggests that (*E,E*)-FPP is the natural substrate. However, more information about the actual occurrence of these assay products in *M. eatonii* is needed before the natural substrates of these enzymes can be determined with certainty.

The substrates actually used by seed plant TPSs depend on their subcellular locations because the various prenyl diphosphate substrates are restricted to different subcellular compartments

(GPP and GGPP to plastids and FPP to the cytosol) (22). Thus, in seed plants, mono- and diterpene synthases are present in plastids and sesquiterpene synthases in the cytosol. However, information on TPS enzyme and substrate localization is not yet available for nonseed plants. To learn more, we used an in silico protein-targeting program (Target P, www.cbs.dtu.dk/services/TargetP) for Mon-UJTT-MTPSL4, an enzyme that produced a single sesquiterpene from (*E*,*E*)-FPP but also formed monoterpenes from (*E*)-GPP. Because the program suggests a cytosolic, nonplastid location for Mon-UJTT-MTPSL4, and FPP is known to be cytosolic in seed plants, this enzyme is likely to act as a sesquiterpene synthase in planta. Under steady-state conditions, the apparent K_M and k_{cat} values of Mon-UJTT-MTPSL4 using (*E*,*E*)-FPP as substrate were determined to be 2.13 ± 0.23 μM and 0.15·s^{-1}, respectively. Such kinetic parameters are very comparable to those of typical plant TPSs (23), suggesting that MTPSL enzymes function in almost the same way as typical plant TPSs.

Based on the in vitro biochemical activities of representative MTPSLs, one could speculate that the MTPSLs have been the primary enzymes to make mono- and sesquiterpenes in early land plants and that the evolution of monoterpene and sesquiterpene synthases among the typical plant TPS family allowed the eventual loss of *MTPSL* genes in seed plants.

Conclusions

In this study, microbial-type terpene synthase genes, once thought to be confined to bacteria and fungi, were systematically mined from large-scale plant transcriptomes. Of 779 seed plant species, only 5 *MTPSL* genes were found in 2 species, whereas 706 *MTPSL* genes were found in 143 nonseed land plant species. So, in addition to the previous report on *S. moellendorffii* (13), *MTPSL* genes are widely distributed in nonseed land plants, but generally absent from seed plants and green algae. Although these genes are also found in fungi and bacteria, their occurrence in plants is in most cases not due to microbial contamination of the plant samples used for sequencing, based on experiments with axenic cultures, phylogenetic analyses, and their embedment in plant genomes with bona fide plant genes as neighbors. *MTPSL* genes form four lineage-specific groups that exhibit diverse structural features, which implies multiple evolutionary origins. Biochemical studies of selected *MTPSL* genes showed that they encode sesquiterpene and monoterpene synthases. However, much more remains to be done to investigate the biological functions of their products and how they have influenced the evolution of the *MTPSL* gene family in nonseed land plants.

Materials and Methods

MTPSL genes were searched against the assembled transcriptomes for 1,103 nonmodel plant species derived from the OneKP (sites.google.com/a/ualberta.ca/onekp/) (24). Fresh materials of three axenically cultured plants, *A. punctatus*, *S. nemorea*, and *S. fallax*, were used for the extraction of genomic DNA that served as template for PCR analysis. For TPS activity assays, crude proteins extracted from *E. coli* expressing individual representative *MTPSL* genes were assayed with individual prenyl diphosphates. The kinetic properties of Mon-UJTT-MTPSL4 were measured with its purified recombinant enzyme following a radiochemical protocol as previously described (23). Details on transcriptome and genome assembly, sequence searches, phylogenetic reconstruction, plant cultures, and biochemical analysis of MTPSLs are provided in *SI Appendix*.

ACKNOWLEDGMENTS. We acknowledge the contributions of Dr. James Leebens-Mack's laboratory at the University of Georgia in obtaining material for RNA sequencing and developing the infrastructure for distribution of the OneKP data. We thank Natascha Rauch (Max Planck Institute for Chemical Ecology) for technical assistance. This project was supported by a University of Tennessee (UT) AgResearch Innovation Grant (to F.C.), the Max Planck Society (to T.G.K. and J.G.), and the Program for New Century Excellent Talents in University (NCET-12-0896 to G.L.). The OneKP initiative led by G.K.-S.W. was funded by the Alberta Ministry of Advanced Education, Alberta Innovates Technology Futures, Innovates Centres of Research Excellence, Musea Ventures, and the China National Genebank. The BGI-Shenzhen group led by Y.Z. was funded by the Shenzhen Supporting Projects Program under Grant CXZZ20140421112021913. *S. fallax* plant and genomic resources were supported by the Laboratory-Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT–Battelle Memorial Institute, for the US Department of Energy under contract DE-AC05-000R22725.

1. Connolly JD, Hill RA (1991) *Dictionary of Terpenoids* (Chapman & Hall, London).
2. Asakawa Y (2004) Chemosystematics of the hepaticae. *Phytochemistry* 65(6):623–669.
3. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. *Nat Chem Biol* 3(7):408–414.
4. Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. *Adv Biochem Eng Biotechnol* 148:63–106.
5. Pichersky E, Gershenzon J (2002) The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr Opin Plant Biol* 5(3):237–243.
6. Cao R, et al. (2010) Diterpene cyclases and the nature of the isoprene fold. *Proteins* 78(11):2417–2432.
7. Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* 66(1):212–229.
8. Bohlmann J, Meyer-Gauen G, Croteau R (1998) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc Natl Acad Sci USA* 95(8):4126–4133.
9. Trapp SC, Croteau RB (2001) Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* 158(2):811–832.
10. Hayashi K, et al. (2006) Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*. *FEBS Lett* 580(26):6175–6181.
11. Mafu S, Hillwig ML, Peters RJ (2011) A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from *Selaginella moellendorffii*. *ChemBioChem* 12(13):1984–1987.
12. Sugai Y, et al. (2011) Enzymatic (13)C labeling and multidimensional NMR analysis of miltiradiene synthesized by bifunctional diterpene cyclase in *Selaginella moellendorffii*. *J Biol Chem* 286(50):42840–42847.
13. Li G, et al. (2012) Nonseed plant Selaginella moellendorffi [corrected] has both seed plant and microbial types of terpene synthases. *Proc Natl Acad Sci USA* 109(36):14711–14715.
14. Starks CM, Back K, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* 277(5333):1815–1820.
15. Köksal M, Jin Y, Coates RM, Croteau R, Christianson DW (2011) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469(7328):116–120.
16. Hori K, et al. (2014) Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nat Commun* 5:3978.
17. Asakawa Y (2008) Liverworts-potential source of medicinal compounds. *Curr Pharm Des* 14(29):3067–3088.
18. Saritas Y, et al. (2001) Volatile constituents in mosses (Musci). *Phytochemistry* 57(3):443–457.
19. Hyatt DC, et al. (2007) Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc Natl Acad Sci USA* 104(13):5360–5365.
20. Hanssen HP, Sprecher E, Abraham WR (1986) 6-Protoilludene, the major molatile metabolite from ceratocystis piceae liquid cultures. *Phytochemistry* 25(8):1979–1980.
21. Matsuba Y, et al. (2013) Evolution of a complex locus for terpene biosynthesis in solanum. *Plant Cell* 25(6):2022–2036.
22. Hemmerlin A, Harwood JL, Bach TJ (2012) A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog Lipid Res* 51(2):95–148.
23. Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. *Plant J* 42(5):757–771.
24. Matasci N, et al. (2014) Data access for the 1,000 Plants (1KP) project. *Gigascience* 3:17.
25. Qiu YL, et al. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci USA* 103(42):15511–15516.
26. Wickett NJ, et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci USA* 111(45):E4859–E4868.

Supplementary Information

Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants

Qidong Jia, Guanglin Li, Tobias G. Köllner, Jianyu Fu, Xinlu Chen, Wangdan Xiong, Barbara Crandall-Stotler, John L. Bowman, David J. Weston, Yong Zhang, Li Chen, Yinlong Xie, Fay-Wei Li, Carl J Rothfels, Anders Larsson, Sean W. Graham, Dennis W. Stevenson, Gane Ka-Shu Wong, Jonathan Gershenzon, and Feng Chen

Materials and Methods

Identification of terpene synthases of the microbial type from transcriptomes and sequenced genomes

The One Thousand Plants Project (OneKP; https://sites.google.com/a/ualberta.ca/onekp/) has sequenced transcriptomes for over 1000 non-model plant species spanning almost all major plant clades (from green algae to flowering plants) (1). All transcriptomes were pre-assembled with the SOAPdenovo-Trans (2) assembler. Transcriptomes of 1103 representative species (Table S1) were analyzed in this study. For all the assembled contigs, the longest regions without stop codons were annotated and translated using the getorf program from the EMBOSS package (3) with a minimum length of 150 amino acids. The resulting peptides were searched against the Pfam-A database locally using HMMER 3.0 hmmsearch (4) with an E-value of $1e^{-5}$. Only sequences with best hits from the following four HMM profiles were considered as putative terpene synthases: Terpene_synth_C (PF03936) and Terpene synthase N-terminal domain (PF01397), TRI5 (PF06330) and SmMTPSLs (a profile created by using 48 microbial type TPSs identified from *S. moellendorffii*). For sequences from the same species that had 100% identity, only the longest one was retained, to reduce redundancy. All the putative TPS sequences were subjected to a BLASTP search against the NCBI's non-redundant database using default parameters. A TPS was annotated as “Microbial TPS-like protein” (MTPSL) if all the top ten best hits were from bacteria and/or fungi or identical/highly similar to SmMTPSLs.

Assembly of hornwort *Anthoceros punctatus* genome and identification of *MTPSL* genes

For *Anthoceros punctatus*, the Illumina paired-end whole genome sequencing data (access number: SRR1278954) (5) were retrieved from NCBI's Sequence Read Archive (SRA) database. The reads were assembled using SPAdes-3.1.1 (6) and the resulting contigs and singletons were further assembled by CAP3 (7). The final CAP3 assembly contains 34448 sequences (16272 contigs and 18176 singletons) a total length of 97Mb, of which 15596 sequences have a minimum length of 500 bp. The N50 contig length based on these 15596 sequences is 12,462 bp. The assemblies were searched for occurrences of terpene synthases using homology-based methods and *ab initio* predictors. A TBLASTN search was performed with an E-value cutoff of $1e^{-30}$ using the 716 *MTPSL* genes identified from OneKP transcriptomes. We also ran SNAP (8) trained for *Arabidopsis thaliana* on the assembly. The resulting protein sequences of predicted genes were subsequently subjected to a HMMER search against four HMM profiles (PF03936, PF01397, PF06330 and SmMTPSLs generated by using 48 microbial type TPSs from *S. moellendorffii*).

Phylogenetic analyses of terpene synthases.

Bacterial and fungal terpene synthases were obtained from Pfam (version 27.0). Considering that certain MTPSLs from plants may contain a transit peptide that is absent in bacterial and fungal TPSs and certain fungal TPSs contain an extended N-terminal domain, to reduce ambiguities in sequence alignment, only the terpene synthase C terminal domains were included. Sequences were aligned using MAFFT (linsi) (9) with 1000 iterations of improvement. ProtTest (10) was used to select the most appropriate protein evolution model for the protein alignment under the Akaike Information Criterion. For the maximum likelihood analyses, we used RAxML (11) with 1000 bootstrap replicates under the best substitution model (LG+G+F) selected by ProtTest.

Plant material, genomic DNA isolation and PCR

Scapania nemorea, Anthoceros punctatus and *Sphagnum fallax* were cultured axenically in Hatcher's medium (12), Knop medium (13) and BCD medium (14), respectively.

Genomic DNA from each species was isolated using the VIOGENE plant genomic DNA isolation kit (Viogene BioTek Corp., Taiwan) and used for PCR with primers listed in Table S9. PCR products were cloned into the pGEM®-T Easy Vector (Promega, USA) and fully sequenced.

Reagents

(*E*)-GPP, (*E,E*)-FPP, (Z,Z)-FPP and (*E,E,E*)-GGPP were purchased from Echelon Biosciences (Salt Lake City). (*Z,E*)-FPP was kindly provided by Nathalie Gatto and Wilhelm Boland from Max Planck Institute for Chemical Ecology, Jena, Germany. [1-^{3}H](*E,E*)-FPP was a product of American Radiolabeled Chemicals (St. Louis).

Terpene synthase enzyme assays and kinetic measurements

Representative members from each of the four groups of *MTPSL* genes were selected for gene synthesis. The synthesized cDNAs were cloned into a protein expression vector pEXP5/CT-TOPO (Thermo Fisher Scientific, USA). Protein expression in *E. coli* and terpene synthase enzyme assays were performed as previously described (15). To determine the kinetic properties of Mon-UJTT-MTPSL4, its cDNA was first amplified via PCR using a pair of primers listed in Table S9. The PCR product was cloned into pET32a, in which the Mon-UJTT-MTPSL4 coding sequence was fused to the his-tag coding sequence at its N-terminal. *E. coli*-expressed recombinant Mon-UJTT-MTPSL4 was purified through the his-tag using the HisTrap HP column (GE Lifesciences, USA). The purified Mon-UJTT-MTPSL4 was used for kinetic measurements using [1-^{3}H](*E,E*)-FPP as substrate via a radiochemical method as previously described (16).

References

1. Matasci N*, et al.* (2014) Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**:17.
2. Xie Y*, et al.* (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **30**(12):1660-1666.
3. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**(6):276-277.
4. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**(Web Server issue):W29-37.
5. Li FW*, et al.* (2014) Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc Natl Acad Sci U S A* **111**(18):6672-6677.
6. Bankevich A*, et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**(5):455-477.
7. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* **9**(9):868-877.
8. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5**:59.
9. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**(14):3059-3066.
10. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**(8):1164-1165.
11. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9):1312-1313.
12. Hatcher RE (1965) Towards the establishment of a pure culture collection of hepaticae. *The Bryologist* **68**(2):227-231.
13. Reski R, Abel WO (1985) Induction of budding on chloronemata and caulonemata of the moss, *Physcomitrella patens*, using isopentenyladenine. *Planta* **165**(3):354-358.
14. Ashton NW, Cove DJ (1977) The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. *Mol Gen Genet* **154**(1):87-95.
15. Li G*, et al.* (2012) Nonseed plant *Selaginella moellendorfii* has both seed plant and microbial types of terpene synthases. *Proc Natl Acad Sci USA* **109**(36):14711-14715.
16. Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. *Plant J* **42**(5):757-771.
17. Hanssen HP, Sprecher E, Abraham WR (1986) 6-Protoilludene, the major molatile metabolite from ceratocystis piceae liquid cultures. *Phytochemistry* **25**(8):1979-1980.

Table S1. List of screened plant species and the number of *MTPSLs* in each species. All these 1103 transcriptomes were from 1KP (https://sites.google.com/a/ualberta.ca/onekp/). MTPSLs are found from 146 species.

Table S1. List of screened transcriptomes and the number of MTPS in each sample. All these 1103 transcriptomes were from 1KP (www.onekp.com). MTPSLs are found from 146 species.

Monilophytes	353
Adiantum aleuticum	5
Adiantum tenerum	8
Anemia tomentosa	12
Angiopteris evecta	5
Argyrochosma nivea	10
Asplenium nidus	6
Asplenium platyneuron	7
Athyrium filix femina	2
Azolla cf. caroliniana	5
Blechnum spicant	15
Botrypus virginianus	6
Cheilanthes arizonica	7
Cibotium glaucum	5
Crepidomanes venosum	0
Cryptogramma acrostichoides	3
Culcita macrocarpa	2
Cyathea spinulosa	1
Cystopteris fragilis	8
Cystopteris protrusa	5
Cystopteris reevesiana	1
Cystopteris utahensis	20
Danaea sp.	7
Davallia fejeensis	8
Dennstaedtia	3
Deparia lobato	8
Didymochlaena truncatula	2
Diplazium wichurae	6
Dipteris conjugata	5
Equisetum diffusum	2
Equisetum hymale	0
Gymnocarpium dryopteris	6
Hemionitis arifolia	3
Homalosorus pycnocarpos	6
Hymenophyllum bivalve	4
Hymenophyllum cupressiforme	1
Leucostegia immersa	8
Lindsaea linearis	7
Lindsaea microphylla	2
Lygodium japonicum	5
Marattia sp.	3
Myriopteris eatonii	11
Nephrolepis exaltata	13
Notholaena montieliae	4
Onoclea sensibilis	1
Ophioglossum vulgatum	7
Osmunda javanica	2
Osmunda regalis	0
Osmunda sp.	4
Osmundastrum cinnamomeum	0
Pilularia globulifera	1
Pityrogramma trifoliata	7
Plagiogyria japonica	2
Pleopeltis polypodioides	9
Polypodium amorphum	9
Polypodium glycyrrhiza	7
Polypodium hesperium	4
Polypodium plectolens	3
Polystichum acrostichoides	5
Psilotum nudum	2
Pteris ensiformis	2
Pteris vittata	10
Sceptridium dissectum	4
Sticherus lobatus	0
Thelypteris acuminata	3
Thyrsopteris elegans	3
Tmesipteris parva	4
Vittaria appalachiana	1
Vittaria lineata	3
Woodsia ilvensis	3
Woodsia scopulina	10
Liverworts	**183**
Barbilophozia barbata	11
Bazzania trilobata	0
Blasia sp.	1
Calypogeia fissa	4
Conocephalum conicum	9
Frullania	8
Lejeuneaceae sp.	14
Lunularia cruciata	9
Marchantia emarginata	4
Marchantia paleacea	8
Marchantia polymorpha	7
Metzgeria crassipilis	1
Monoclea gottschei	0
Odontoschisma prostratum	10
Palavicinia lyellii	4
Pellia cf. Epiphylla	9
Pellia neesiana	16
Plagiochila asplenioides	9
Porella navicularis	8
Porella pinnata	8
Ptilidium pulcherrimum	11
Radula lindenbergia	8
Riccia berychiana	4
Scapania nemorea	8
Schistochila sp.	3
Sphaerocarpos texanus	3
Mosses	**79**
Andreaea rupestris	0
Anomodon attenuatus	1
Anomodon rostratus	4
Atrichum angustatum	1
Aulacomnium heterostichum	1
Bryum argenteum	0
Buxbaumia aphylla	1
Calliergon cordifolium	0
Ceratodon purpureus	0
cf. Physcomicromitrium sp.	0
Climacium dendroides	1
Dicranum scoparium	3
Diphyscium foliosum	3
Encalypta streptocarpa	0
Fontinalis antipyretica	0
Funaria	1
Hedwigia ciliata	0
Hypnum subimponens	5
Leucobryum albidum	2
Leucobryum glaucum	3
Leucodon sciuroides	2
Neckera douglasii	2
Niphotrichum elongatum	1
Orthotrichum lyellii	0
Philonotis fontana	5
Plagiomnium insigne	2
Polytrichum commune	1
Pseudotaxiphyllum elegans	3
Racomitrium varium	4
Rhynchostegium serrulatum	4
Rhytidiadelphus loreus	1
Schwetschkeopsis fabronia	3
Scouleria aquatica	0
Sphagnum lescurii	5
Sphagnum palustre	7
Sphagnum recurvatum	3
Syntrichia princeps	1
Takakia lepidozioides	2
Tetraphis pellucida	0
Thuidium delicatulum	5
Timmia austriaca	2
Hornworts	**14**
Anthoceros formosae	0
Leiosporoceros dussii	0
Megaceros tosanus	0
Megaceros vincentianus	2
Nothoceros aenigmaticus	0
Paraphymatoceros hallii	3
Phaeoceros carolinianus	9

Lycophytes	77
Dendrolycopodium obscurum	2
Diphasiastrum digitatum	0
Huperzia lucidula	1
Huperzia myrisinites	2
Huperzia selago	7
Huperzia squarrosa	2
Isoetes sp.	6
Isoetes sp.	7
Isoetes tegetiformans	5
Lycopodiella apressa	1
Lycopodium annotinum	1
Lycopodium deuterodensum	4
Phylloglossum drummondii	4
Pseudolycopodiella caroliniana	3
Selaginella acanthonota	3
Selaginella apoda	2
Selaginella kraussiana	8
Selaginella lepidophylla	10
Selaginella selaginoides	2
Selaginella stauntoniana	2
Selaginella wallacei	6
Selaginella willdenowii	5
Chlorophytes	**0**
Acrosphonia sp.	0
Ankistrodesmus sp.	0
Aphanochaete repens	0
Asteromonas gracilis	0
Bathycoccus prasinos	0
Blastophysa cf. rhizopus	0
Bolbocoleon piliferum	0
Botryococcus braunii	0
Botryococcus sudeticus	0
Botryococcus terribilis	0
Brachiomonas submarina	0
Bryopsis plumosa	0
Carteria crucifera	0
Carteria obtusa	0
Cephaleuros virescens	0
Chaetopeltis orbicularis	0
Chlamydomonas bilatus	0
Chlamydomonas cribrum	0
Chlamydomonas moewusii	0
Chlamydomonas noctigama	0
Chlamydomonas sp.	0
Chlorella minutissima	0
Chloromonas oogama	0
Chloromonas perforata	0
Chloromonas reticulata	0
Chloromonas rosa	0
Chloromonas subdivisa	0
Chloromonas tughillensi	0
Chlorosarcinopsis halophila	0
Cladophora glomerata	0
coccoid prasinophyte	0
Coccomyxa pringsheimii	0
Codium fragile	0
Cylindrocapsa geminella	0
Cymbomonas sp.	0
Dolichomastix tenuilepi	0
Dunaliella primolecta	0
Dunaliella salina	0
Dunaliella tertiolecta	0
Entocladia endozoica	0
Eremosphaera viridi	0
Eudorina elegans	0
Fritschiella tuberosa	0
Geminella sp.	0
Golenkinia longispicula	0
Gonium pectorale	0
Haematococcus pluvali	0
Haematococcus pluvialis	0
Hafniomonas reticulata	0
Halochlorococcum marinum	0
Helicodictyon planctonicum	0
Heterochlamydomonas inaequalis	0
Hormidiella sp.	0
Ignatius tetrasporus	0
Leptosira obovata	0
Lobochlamys segnis	0
Lobomonas rostrata	0
Mantoniella squamata	0
Micropsora cf. tumidula	0
Microthamnion kuetzigianum	0
Monomastix opisthostigma	0
Nannochloris atomus	0
Neochloris oleoabundans	0
Neochloris sp.	0
Neochlorosarcina sp.	0
Nephroselmis olivace	0
Nephroselmis pyriformis	0
Ochlochaete sp.	0
Oedogonium cardiacu	0
Oedogonium foveolatum	0
Oltmannsiellopsis viridis	0
Oogamochlamys gigantea	0
Pandorina morum	0
Parachlorella kessleri	0
Pediastrum duplex	0
Pedinomonas minor	0
Pedinomonas tuberculata	0
Persursaria percursa	0
Phacotus lenticularis	0
Picocystis salinarum	0
Pirula salina	0
Planophila laetevirens	0
Planophila terrestris	0
Pleurastrum insigne	0
Prasinococcus capsulatus	0
Prasinoderma coloniale	0
Prasiola crispa	0
Prototheca wickerhamii	0
Pseudoscourfieldia marina	0
Pteromonas angulosa	0
Pteromonas sp.	0
Pycnococcus provasolii	0
Pyramimonas parkeae	0
Scenedesmus dimorphus	0
Scherffelia dubia	0
Scourfieldia sp.	0
Spermatozopsis exsultans	0
Spermatozopsis similis	0
Stephanosphaera pluvialis	0
Stichococcus bacillaris	0
Stigeoclonium helveticum	0
Tetraselmis chui	0
Tetraselmis cordiformis	0
Tetraselmis striata	0
Trebouxia arboricola	0
Trentepohlia annulata	0
Unidentified species CCMP 1205	0
Uronema belka	0
Vitreochlamys sp.	0
Volvox aureus	0
Volvox globator	0

Charophytes	1
Bambusina borreri	0
Chaetosphaeridium globosum	0
Chara vulgaris	0
Chlorokybus atmophyticus	0
Closterium lunula	0
Coleochaete irregularis	0
Coleochaete scutata	0
Cosmarium broomei	0
Cosmarium granatum	0
Cosmarium impressulum	0
Cosmarium subtumidum	0
Cosmarium tinctum	0
Cosmocladium cf. constrictum	0
Cylindrocystis brebissonii	0
Cylindrocystis cushleckae	0
Cylindrocystis sp.	0
Desmidium aptogonum	0
Entransia fimbriat	0
Euastrum affine	0
Gonatozygon kinahanii	0
Interfilum paradoxum	0
Klebsormidium subtile	0
Mesostigma viride	0
Mesotaenium braunii	0
Mesotaenium caldariorum	0
Mesotaenium endlicherianum	0
Mesotaenium kramstei	0
Micrasterias fimbriata	1
Mougeotia sp.	0
Netrium digitus	0
Nucleotaenium eifelense	0
Onychonema laeve	0
Penium exiguum	0
Penium margaritaceum	0
Physolobocis nordstedtiana	0
Planotaenium ohtanii	0
Pleurotaenium trabecul	0
Roya obtusa	0
Spirogyra sp.	0
Spirotaenia minuta	0
Spirotaenia sp.	0
Staurastrum sebaldi	0
Staurodesmus convergens	0
Staurodesmus omearii	0
Xanthidium antilopaeum	0
Zygnema sp.	0
Zygnemopsis sp.	0
Gymnosperms	**0**
Abies lasiocarpa	0
Acmopyle pancheri	0
Agathis robusta	0
Amentotaxus argotaenia	0
Araucaria rulei	0
Arucaria sp.	0
Athrotaxis cupressoides	0
Austrocedrus chilensis	0
Austrotaxus spicata	0
Callitris gracilis	0
Callitris macleayana	0
Calocedrus decurrens	0
Cathaya agryrophylla	0
Cedrus libani	0
Cephalotaxus harringtonia	0
Chamaecyparis lawsoniana	0
Cryptomeria japonica	0
Cunninghamia lanceolata	0
Cupressus dupreziana	0
Cycas micholitzii	0
Dacrycarpus compactus	0
Dacrydium balansae	0
Dioon edule	0
Diselma archeri	0
Encephalartos barteri	0
Ephedra sinica	0
Falcatifolium taxoides	0
Fokienia hodginsii	0
Ginkgo biloba	0
Glyptostrobus pensilis	0
Gnetum montanum	0
Halocarpus bidwillii	0
Juniperus scopulorum	0
Keteleeria evelyniana	0
Lagarostrobos franklinii	0
Larix speciosa	0
Manoao colensoi	0
Metasequoia glyptostroboides	0
Microbiota decussata	0
Microcachrys tetragona	0
Microstrobos fitzgeraldii	0
Nageia nagi	0
Neocallitropsis pancheri	0
Nothotsuga longibracteata	0
Papuacedrus papuana	0
Parasitaxus usta	0
Phyllocladus hypophyllus	0
Picea engelmanii	0
Pilgerodendron uviferum	0
Pinus jeffreyi	0
Pinus parviflora	0
Pinus ponderosa	0
Pinus radiata	0
Platycladus orientalis	0
Podocarpus coriaceus	0
Podocarpus rubens	0
Prumnopitys andina	0
Pseudolarix amabilis	0
Pseudotaxus chienii	0
Pseudotsuga menziesii	0
Retrophyllum minus	0
Saxegothaea conspicua	0
Sciadopitys verticillata	0
Sequoia sempervirens	0
Sequoiadendron giganteum	0
Stangeria eriopus	0
Sundacarpus amarus	0
Taiwania cryptomerioides	0
Taxodium distichum	0
Taxus baccata	0
Taxus cuspidata	0
Tetraclinis sp.	0
Thuja plicata	0
Thujopsis dolabrata	0
Torreya nucifera	0
Torreya taxifolia	0
Tsuga heterophylla	0
Welwitschia mirabilis	0
Widdringtonia cedarbergensis	0
Wollemia nobilis	0
Angiosperms	**5**
Acacia argyrophylla	0
Acacia pycnantha	0
Acer negundo	0
Acorus americanus	0
Actinidia chinensis	0
Aerva lanata	0
Aerva persica	0
Aesculus pavia	0
Aextoxicon punctatum	0
Agapanthus africanus	0
Agastache rugosa	0
Agave tequilana	0
Ailanthus altissima	0
Ajuga reptans	0
Akania lucens	0
Akebia trifoliata	0
Alangium chinense	0
Allamanda cathartica	0
Allionia incarnata	0
Allionia spp.	0

Angiosperms (Continued)	
Allium commutatum	0
Allium sativum	0
Alnus serrulata	0
Aloe vera	0
Alternanthera brasileana	0
Alternanthera caracasana	0
Alternanthera sessilis	0
Alternanthera tenella	0
Amaranthus cruentus	0
Amaranthus palmeri	0
Amaranthus retroflexus	0
Amaryllis belladonna	0
Amborella trichopoda	0
Amelanchier canadensis	0
Ancistrocladus tectorius	0
Anemone hupehnensis	0
Anemone pulsatilla	0
Angelica archangelica	0
Aniacanthus quadrifidas	0
Annona muricata	0
Anthemis tinctoria	0
Anthirrhinum majus	0
Anticharis glandulosa	0
Antirrhinum braun	0
Aphanopetalum resinosum	0
Apios americana	0
Apocynum androsaemifolium	0
Arabis alpina	0
Ardisia humilis	0
Ardisia revoluta	0
Argemone mexicana	0
Aristida stricta	0
Aristolochia elegans	0
Aruncus dioicus	0
Ascarina rubricaulis	0
Asclepia curassavica	0
Asclepias syriaca	0
Asparagus densiflorus	0
Aster tataricus	0
Astilbe chinensis	0
Astragalus membranaceus	0
Astragalus propinquus	0
Atriplex hortensis	0
Atriplex prostrata	0
Atriplex rosea	0
Atropa belladonna	0
Aucuba japonica	0
Austrobaileya scandens	0
Avena fatua	0
Azadirachta indica	0
Bacopa caroliniana	0
Balanophora fungosa	0
Basella alba	0
Batis maritima	0
Bauhinia tomentosa	0
Begonia sp.	0
Berberidopsis beckleri	0
Bergenia sp.	0
Beta maritima	0
Betula pendula	0
Bischofia javanica	0
Bituminaria bituminosa	0
Bixa orellana	0
Blutaparon vermiculare	0
Boehmeria nivea	0
Boerhavia cf. spiderwort	0
Boerhavia dominii	0
Borya sphaerocephala	0
Boswellia sacra	0
Bougainvillea spectabilis	0
Boykinia jamesii	0
Brassica nigra	0
Brocchinia reducta	0
Brodiaea sierrae	0
Brugmansia sanguinea	0
Buddleja lindleyana	0
Buddleja sp.	0
Bursera simaruba	0
Buxus sempervirens	0
Byblis gigantea	0
Caiophora chuquitensis	0
Calceolaria pinifolia	0
Calycanthus floridus	0
Camptotheca acuminata	0
Canella winterana	0
Canna sp.	0
Cannabis sativa	0
Capnoides sempervirens	0
Carthamus lanatus	0
Carya glabra	0
Cassytha filiformis	0
Castanea crenata	0
Castanea pumila	0
Casuarina glauca	0
Catharanthus roseus	0
Cavendishia cuatrecasasii	0
Celsia arcturus	0
Celtis occidentalis	0
Centella asiatica	0
Cephalotus follicularis	0
Ceratocapnos vesicaria	0
Ceratophyllum demersum	0
Cercidiphyllum japonicum	0
Cercis canadensis	0
Cercocarpus ledifolius	0
Chamaesyce mesembryanthemum	0
Chelidonium majus	0
Chenopodium quinoa	0
Chionanthus retusus	0
Chlorogalum pomeridianum	0
Chondropetalum tectorum	0
Chrysobalanus icaco	0
Cicerbita plumieri	0
Cimicifuga racemosa	0
Cinnamomum camphora	0
Cissus quadrandularis	0
Cistus inflatus	0
Citrus x paradisi	0
Cladrastis lutea	0
Cleome gynandra	0
Cleome violaceae	0
Cleome viscosa	0
Cocculus laurifolius	0
Cochlearea officinalis	0
Cocos nucifera	0
Codariocalyx motorius	0
Colchicum autumnale	0
Conophalis americana	0
Convolvulus arvensis	0
Conzya canadensis	0
Copaifera officianalis	0
Coriaria nepalensis	0
Cornus floridana	0
Corokia cotoneaster	0
Corydalis linstowiana	0
Cotoneaster transcaucasicus	0
Crassula perforata	0
Crossopetalum rhacoma	0
Croton tiglium	0
Cunonia capensis	0
Curculigo sp.	0
Curcuma olena	0
Curtisia dentata	0
Cuscuta pentagonia	0
Cyanastrum cordifolium	0
Cyanella orchidoformis	0
Cymbopogon nardus	0
Cyperus papyrus	0
Cypselea humifusum	0
Cyrilla racemiflora	0
Daenikera sp.	0

Angiosperms (Continued)	
Daphne geraldii	0
Daphniphyllum macropodum	0
Datura metel	0
Delosperma echinatum	0
Dendropemon caribaeus	0
Deschampsia cespitosa	0
Desmanthus illinoensis	0
Deutzia scabra	0
Dianthus sp.	0
Dichroa febrifuga	0
Digitalis purpurea	0
Dillenia indica	0
Dioscorea villosa	0
Diospyros malabarica	0
Dipsacus sativum	0
Disporopsis pernyi	0
Dombeya burgessiae	0
Draba aizoides	0
Draba hispida	0
Draba magellanica	0
Draba oligosperma	0
Draba ossetica	0
Draba sachalinensis	0
Drakea elastica	0
Drimia altissima	0
Drimys winteri	0
Dryas octopetala	0
Drypetes deplanchei	0
Edgeworthia papyrifera	0
Ehretia acuminata	0
Elaeagnus pungens	0
Elaeocarpus sylvestris	0
Eleusine coracana	0
Epifagus virginiana	0
Epilobium sp.	0
Erigeron speciosus	0
Eriospermum lancifolia	0
Erythroxylum coca	0
Escallonia rubra	0
Escallonia sp. cv. Newport	0
Eschscholzia californica	0
Eucalyptus leucoxylon	0
Eucommia ulmoides	0
Euphorbia pekinensis	0
Eupomatia bennettii	0
Euptelea pleiosperma	0
Exacum affine	0
Exocarpos cupressiformis	0
Fagus sylvatica	0
Ficus religiosa	0
Flaveria angustifolia	0
Flaveria bidentis	0
Flaveria brownii	0
Flaveria cronquestii	0
Flaveria kochiana	0
Flaveria palmeri	0
Flaveria pringlei	0
Flaveria pubescens	0
Flaveria sonorensis	0
Flaveria trinervia	0
Flaveria vaginata	0
Forestiera segregata	0
Fouquieria macdougalii	0
Francoa appendiculata	0
Frankenia laevis	0
Freycinetia multiflora	0
Galax urceolata	0
Galium boreale	0
Galphimia gracilis	0
Garcinia livingstonei	0
Garcinia oblongifolia	0
Gelsemium sempervirens	0
Gentiana acaulis	0
Geranium carolinianum	0
Geranium maculatum	0
Geum quellyon	0
Gleditsia sinensis	0
Gleditsia triacanthos	0
Gloriosa superba	0
Glycine soja	0
Glycyrrhiza glabra	0
Glycyrrhiza lepidota	0
Gomortega keule	0
Gomphrolobium polymorphum	0
Goodyera pubescens	0
Grevillea robusta	0
Greyia sutherlandii	0
Griselinia littoralis	0
Griselinia racemosa	0
Gunnera manicata	0
Gylcine soja	0
Gymnocladus dioicus	0
Gynocarpus americanus	0
Gyrostemon ramulosus	0
Haemaria discolor	0
Hakea drupacear	0
Hakea prostrata	0
Hamamelis virginiana	0
Hedera helix	0
Helenium autumnale	0
Heliconia sp.	0
Heliotropium calcicola	0
Heliotropium convolvulaceum	0
Heliotropium filiforme	0
Heliotropium greggii	0
Heliotropium karwinsky	0
Heliotropium mendocinum	0
Heliotropium racemosum	0
Heliotropium sp.	0
Heliotropium tenellum	0
Heliotropium texanum	0
Helonias bullata	0
Helwingia japonica	0
Hemerocallis sp.	0
Hemerocallis spp.	0
Heracleum lanatum	0
Hesperaloe parviflora	0
Heteropyxis natalensis	0
Heuchera sanguinea	0
Hibbertia grossularifolia	0
Hibiscus cannabinus	0
Hilleria latifolia	0
Hoheria angustifolia	0
Holarrhena pubescens	0
Houttuynia cordata	0
Humulus lupulus	0
Hydrangea quercifolia	0
Hydrastis canadensis	0
Hydrocotyle umbellata	0
Hypecoum procumbens	0
Hypericum perforatum	0
Idiospermum australiense	0
Ilex paraguariensis	0
Ilex sp.	0
Ilex vomitoria	0
Illicium floridanum	0
Illicium parviflorum	0
Impatiens balsamifera	0
Inula helenium	0
Ipomoea coccinea	0
Ipomoea hederacea	0
Ipomoea indica	0
Ipomoea lindheimeri	0
Ipomoea lobata	0
Ipomoea nil	0
Ipomoea pubescens	0
Ipomoea purpurea	0
Ipomoea quamoclit	0
Itea virginica	0
Jacquinia sp.	0
Johnsonia pubescens	0

Table S1 (Continued)

Angiosperms (Continued)	
Joinvillea ascendens	0
Juglans nigra	0
Juncus inflexus	0
Kadsura heteroclite	0
Kalanchoe crenato diagremontiana	0
Kalighora madagascariensis	0
Kerria japonica	0
Kigelia africana	0
Kirkia wilmsii	0
Kochia scoparia	0
Koeberlina spinosa	0
Krameria lanceolata	0
Lactuca graminifolia	0
Lagerstroemia indica	0
Lantana camara	0
Larrea tridentata	0
Lathyrus sativus	0
Laurelia sempervirens	0
Lavandula angustifolia	0
Ledum palustre	0
Lennoa madreporoides	0
Leontopodium alpinum	0
Leonurus japonicus	0
Lepidosperma gibsonii	0
Licania michauxii	0
Ligustrum sinense	0
Lilium sargentiae	0
Limnanthes douglassii	0
Limonium spectabile	0
Lindenbergia philippensis	0
Lindera benzoin	0
Linum bienne	0
Linum flavum	0
Linum grandiflorum	0
Linum hirsutum	0
Linum leonii	0
Linum lewisii	0
Linum macraei	0
Linum perenne	0
Linum strictum	0
Linum tenuifolium	0
Linum usitatissimum	0
Liquidambar styraciflua	0
Litchi chinensis	0
Lobelia siphilitica	0
Lomandra longifolia	0
Lonicera japonica	0
Lophophora williamsii	0
Loropetalum chinense	0
Ludovia sp.	0
Lupinus angustifolius	0
Lupinus polyphyllus	0
Lycium barbarum	0
Lycium sp.	0
Lycopersicon cheesmanii	0
Maesa lanceolata	0
Magnolia grandiflora	0
Maianthemum canadense	0
Maianthemum sp.	0
Malesherbia fasciculata	0
Malus baccata	0
Manihot grahamii	0
Manilkara zapota	0
Mansoa alliacea	0
Mapania palustris	0
Maranta leuconeura	0
Marrubium vulgare	0
Masdevallia yuangensis	0
Matricaria matricariodes	0
Medinilla magnifica	0
Melaleuca quinquenervia	0
Melia azedarach	0
Meliosma cuneifolia	0
Melissa officinalis	0
Menyanthes trifoliata	0
Mertensia paniculata	0
Michelia maudiae	0
Micromeria fruticosa	0
Microstegium vimineum	0
Microtea debilis	0
Mirabilis jalapa	0
Mitella pentandra	0
Mollugo cerviana	0
Mollugo nudicaulis	0
Mollugo pentaphylla	0
Mollugo verticillata	0
Monotropa uniflora	0
Morinda citrifolia	0
Moringa oleifera	0
Morus nigra	0
Mumea americana	0
Muntingia calabura	0
Mydocarpus sp.	0
Myrica cerifera	0
Myriophyllum aquaticum	0
Myristica fragrans	0
Nandina domestica	0
Narcissus viridiflorus	0
Nelumbo nucifera	0
Nelumbo sp.	0
Nepenthes alata	0
Nepeta cataria	0
Neurachne alopecuroidea	0
Neurachne annularis	0
Neurachne lanigera	0
Neurachne minor	0
Neurachne munroi	0
Neurachne tenuifolia	0
Nicotiana sylvestris	0
Nolina atopocarpa	0
Nolina bigelorii	0
Nothofagus obliqua	0
Nuphar advena	0
Nypa fruticans	0
Nyssa ogeche	0
Ochna mossambicensis	0
Ochna serrulata	0
Oenothera affinis	0
Oenothera berlandieri	0
Oenothera biennis	0
Oenothera clelandii	0
Oenothera elata	0
Oenothera elata hookeri	0
Oenothera filiformis	0
Oenothera gaura	0
Oenothera grandiflora	0
Oenothera grandis	0
Oenothera laciniata	0
Oenothera longituba	0
Oenothera nana	0
Oenothera picensis	0
Oenothera rhombipetala	0
Oenothera rosea	0
Oenothera serrulata	0
Oenothera speciosa	0
Oenothera suffulta suffulta	0
Oenothera villaricae	0
Olea europaea	0
Oncidium sphacelatum	0
Oncotheca balansae	0
Opuntia sp.	4
Orchidantha maxillaroides	0
Oresitrophe rupifraga	0
Orobanche fasciculata	0
Oxalis sp.	0
Oxera neriifolia	0
Oxera pulchella	0
Paeonia lactiflora	0
Panicum miliaceum A	0
Papaver bracteatum	0
Papaver rhoeas	0
Papaver setigerum	0
Papaver somniferum	0
Paraneurachne muelleri	0
Passiflora caerulea	0
Passiflora edulis	0
Paulownia fargesii	0
Peganum harmala	0
Peliosanthese minor	0
Peltoboykinia watanabei	0
Pennantia corymbosa	0
Peperomia fraseri	0
Pereskia aculeata	0
Persea borbonia	0
Petiveria alliacea	0
Peumus boldus	0

Angiosperms (Continued)	
Phacelia campanularia	0
Phelline lucida	0
Phellodendron amurense	0
Philadelphus inodorus	0
Phlox drummondii	0
Phlox sp.	0
Pholisma arenarium	0
Phoradendron serotinum	0
Phormium tenax	0
Physcella aff. cyrtanthoides	0
Phyla dulcis	0
Phyllanthus sp.	0
Physena madagascariensis	0
Physocarpus opulifolius	0
Phytolacca americana	0
Phytolacca bogotensis	1
Pilostyles thunbergii	0
Pinguicula agnata	0
Pinguicula caudata	0
Piper auritum	0
Pistia stratioides	0
Pittosporum resiniferum	0
Pittosporum sahnianum	0
Plantago coronopis	0
Plantago maritima	0
Plantago virginica	0
Platanthera clavellata	0
Platanus occidentalis	0
Platycodon grandiflorus	0
Platyspermation crassifolium	0
Plumbago auriculata	0
Podophyllum peltatum	0
Pogostemon sp.	0
Polansia trachysperma	0
Poliomintha bustamanta	0
Polycarpaea repens	0
Polygala lutea	0
Polygonum convolvulus	0
Polypremum procumbens	0
Polyscias fruticosa	0
Portulaca amilis	0
Portulaca cryptopetala	0
Portulaca grandiflora	0
Portulaca mauii	0
Portulaca molokaiensis	0
Portulaca oleracea	0
Portulaca pilosa	0
Portulaca suffruticosa	0
Portulaca umbraticola	0
Posidonia australis	0
Prunella vulgaris	0
Prunus prostrata	0
Psychotria douarrei	0
Psychotria ipecacuanha	0
Psychotria marginata	0
Punica granatum	0
Pycnanthemum tenuifolium	0
Pyrenacantha malvifolia	0
Quassia amara	0
Quercus shumardii	0
Quillaja saponaria	0
Rauvolfia tetraphyla	0
Rehmannia glutinosa	0
Reseda odorata	0
Rhamnus caroliniana	0
Rhamnus japonica	0
Rhizophora mangle	0
Rhodiola rosea	0
Rhododendron scopulorum	0
Rhodophiala pratensis	0
Rhus radicans	0
Ribes aff. giraldii	0
Ricinus communis	0
Roridula gorgonias	0
Rosa palustris	0
Rosmarinus officinalis	0
Ruellia brittoniana	0
Ruscus sp.	0
Sabal bermudana	0
Sagittaria latifolia	0
Saintpaulia ionantha	0
Salix acutifolia	0
Salix dasyclados	0
Salix eriocephala	0
Salix fargesii	0
Salix purpurea	0
Salix sachalinensis	0
Salix viminalis	0
Salvadora sp.	0
Salvia spp.	0
Sambucus canadensis	0
Sanchezia sp.	0
Sanguinaria canadensis	0
Sanguisorba minor	0
Sansevieria trifasciata	0
Santalum acuminatum	0
Saponaria officianalis	0
Sarcandra glabra	0
Sarcobatus vermiculatus	0
Sarcodes sanguinea	0
Saruma henryi	0
Sassafras albidum	0
Saururus cernuus	0
Saxifraga stolonifera	0
Scaevola sp.	0
Schiedea membranacea	0
Schizolaena sp.	0
Schlegelia parasitica	0
Schlegelia parasitica B	0
Schlegelia violacea	0
Scutellaria montana	0
Senecio rowleyanus	0
Senna hebecarpa	0
Serenoa repens	0
Sessuvium portulacastrum	0
Sessuvium ventricosum	0
Sideroxylon reclinatum	0
Silene latifolia	0
Silybum marianum	0
Simmondsia chinensis	0
Sinapis alba	0
Sinningia tuberosa	0
Sinojackia xylocarpa	0
Sisyrinchium angustifolium	0
Smilax bona nox	0
Solanum dulcamara	0
Solanum lasiophyllum	0
Solanum ptychanthum	0
Solanum sisymbriifolium	0
Solanum xanthocarpum	0
Solenostemon scutellaroides	0
Solidago canadensis	0
Sorbus koehneana	0
Souroubea exauriculata	0
Spergularia media	0
Stachyurus praecox	0
Stackhousia spathulata	0
Staphylea trifolia	0
Stemona tuberosa	0
Strelitzia reginae	0
Strobilanthes dyerianus	0
Strychnos spinosa	0
Stylidium adnatum	0
Symphoricarpos sp.	0
Symplocus sp.	0
Synsepalum dulcificum	0
Syzygium macranthum	0
Syzygium paniculatum	0
Tabebuia umbellate	0
Talbotia elegans	0
Talinum sp.	0
Tamarix chinensis	0
Tanacetum parthenium	0
Tapiscia sinensis	0
Tellima breviflora	0
Terminalia neotaliala	0
Ternstroemia gymnanthera	0
Tetrastigma obtectum	0
Tetrastigma voinierianum	0
Tetrazygia bicolor	0
Teucrium chamaedrys	0
Thalictrum thalictroides	0
Thladiantha villosula	0
Thymus vulgaris	0
Thyridolepis mitchelliana	0
Thyridolepis multiculmis	0

Angiosperms (Continued)	
Tiarella polyphylla	0
Tragopogon castellanus	0
Tragopogon dubius	0
Tragopogon porrifolius	0
Tragopogon pratensis	0
Traubia modesta	0
Trianthemum portulacastrum	0
Triglochin maritimum	0
Triodia aff. bynoei	0
Trochodendron araliodes	0
Tropaeolum peregrinum	0
Trubulus eichlerianus	0
Typha angustifolia	0
Typha latifolia	0
Typhonium blumei	0
Ulmus alata	0
Uncaria grandidieri	0
Uniola paniculata	0
Urginea maritima	0
Urtica dioica	0
Utricularia sp.	0
Uvaria microcarpa	0
Valeriana officianalis	0
Vanilla planifolia	0
Verbascum sp.	0
Verbena hastata	0
Viburnum odoratissimum	0
Viola canadensis	0
Viola tricolor	0
Vitex agnus castus	0
Wikstroemia indica	0
Wisteria floribunda	0
Wrightia natalensis	0
Xanthicercis zambesiaca	0
Xanthuium strumarium	0
Xeronema callistemon	0
Xerophyllum asphodeloides	0
Xerophyta villosa	0
Ximenia americana	0
Yucca brevifolia	0
Yucca filamentosa	0
Zaleya pentandra	0
Zephyranthes treatiae	0
Zingiber officinale	0
Ziziphus jujuba	0

Table S2. Summary statistics of *MTPSLs* in 9 plant lineages.

Lineage	Species Count	MTPSL Count	Mean	Median	St. Dev.	Min	Max
Angiosperms	699	5	0.01	0	0.16	0	4
Gymnosperms	80	0	0	0	0	0	0
Monilophytes	70	353	5.04	4.50	3.81	0	20
Lycophytes	22	83	3.77	3	2.65	0	10
Hornworts	7	14	2	0	3.32	0	9
Mosses	41	79	1.93	1	1.82	0	7
Liverworts	26	177	6.81	8	4.15	0	16
Charophytes	47	1	0.02	0	0.15	0	1
Chlorophytes	111	0	0	0	0	0	0

Table S3. 22 *MTPSL* genes designated as Unclassified and their top hits from Non redundant (NR) database of NCBI

MTPSL_id	Lineages	NR_top_hit	evalue	bit_score	Scientific_name	Kingdom
(RXRQ_WCZB)_MTPSL7	Hornworts	gi|751680917|gb|KIM31075.1|	0	1019	Serendipita vermifera MAFF 305830	Fungi
(RXRQ_WCZB)_MTPSL8	Hornworts	gi|353240956|emb|CCA72799.1|	0	981	Piriformospora indica DSM 11827	Fungi
(RXRQ_WCZB)_MTPSL9	Hornworts	gi|353240956|emb|CCA72799.1|	7E-130	391	Piriformospora indica DSM 11827	Fungi
MCHJ_MTPSL1	Charophytes	gi|913451420|ref|WP_050430829.1|	0.001	50.8	Chondromyces crocatus	Bacteria
MRKX_MTPSL1	Angiosperms	gi|588255974|ref|XP_006957163.1|	0	558	Wallemia mellicola CBS 633.66	Fungi
QAIR_MTPSL1	Angiosperms	gi|751178290|gb|KIL64254.1|	5E-177	508	Amanita muscaria Koide BX008	Fungi
QAIR_MTPSL2	Angiosperms	gi|751175026|gb|KIL61028.1|	0	620	Amanita muscaria Koide BX008	Fungi
QAIR_MTPSL3	Angiosperms	gi|751174784|gb|KIL60790.1|	7E-143	427	Amanita muscaria Koide BX008	Fungi
QAIR_MTPSL4	Angiosperms	gi|751181225|gb|KIL67171.1|	0	691	Amanita muscaria Koide BX008	Fungi
JKAA_MTPSL1	Lycophytes	gi|927407765|ref|XP_013949969.1|	0	541	Trichoderma virens Gv29-8	Fungi
ZYCD_MTPSL1	Lycophytes	gi|238496645|ref|XP_002379558.1|	1E-130	390	Aspergillus flavus NRRL3357	Fungi
AEXY_MTPSL1	Liverworts	gi|389636521|ref|XP_003715910.1|	6E-178	513	Magnaporthe oryzae 70-15	Fungi
IRBN_MTPSL6	Liverworts	gi|751680917|gb|KIM31075.1|	0	934	Serendipita vermifera MAFF 305830	Fungi
JHFI_MTPSL16	Liverworts	gi|751680917|gb|KIM31075.1|	6E-101	313	Serendipita vermifera MAFF 305830	Fungi
LGOW_MTPSL3	Liverworts	gi|629725325|ref|XP_007822988.1|	7E-56	196	Metarhizium robertsii	Fungi
NWQC_MTPSL8	Liverworts	gi|629725325|ref|XP_007822988.1|	6E-36	144	Metarhizium robertsii	Fungi
OFTV_MTPSL7	Liverworts	gi|751680917|gb|KIM31075.1|	0	931	Serendipita vermifera MAFF 305830	Fungi
RTMU_MTPSL4	Liverworts	gi|667838359|ref|XP_007783348.1|	3E-128	386	Coniosporium apollinis CBS 100218	Fungi
WJLO_MTPSL3	Liverworts	gi|549052256|emb|CCX30236.1|	4E-68	231	Pyronema omphalodes CBS 100304	Fungi
WJLO_MTPSL4	Liverworts	gi|549052256|emb|CCX30236.1|	1E-71	240	Pyronema omphalodes CBS 100304	Fungi
YBQN_MTPSL8	Liverworts	gi|648165817|gb|KDR79494.1|	0	596	Galerina marginata CBS 339.88	Fungi
QIAD_MTPSL2	Monilophytes	gi|629662947|ref|XP_007805277.1|	2E-77	251	Endocarpon pusillum Z07020	Fungi

Table S4. Eight MTPSLs identified from the transcriptome of the liverwort *Scapania nemorea*

Gene	Group
Liv-IRBN-MTPSL1	I
Liv-IRBN-MTPSL2	I
Liv-IRBN-MTPSL3	I
Liv-IRBN-MTPSL4	I
Liv-IRBN-MTPSL5	I
Liv-IRBN-MTPSL6	**II**
Liv-IRBN-MTPSL7	III
Liv-IRBN-MTPSL8	**U**[a]

[a]unclassified

Table S5. *MTPSL* genes from the genome of the hornwort *Anthoceros punctatus*

Gene	Protein	Group
ApMTPSL1	408	II
ApMTPSL2	430	III
ApMTPSL3	436	II
ApMTPSL4	401	III
ApMTPSL5	413	III
ApMTPSL6	421	III
ApMTPSL7	427	III

Table S6. *MTPSL* genes from the genome of the moss *Sphagnum fallax*

Gene	Protein size	Group
SfMTPSL1	472	I
SfMTPSL2	341	I
SfMTPSL3	328	I
SfMTPSL4	482	I
SfMTPSL5	489	I
SfMTPSL6	477	I
SfMTPSL7	481	I
SfMTPSL8	484	I
SfMTPSL9	377	I
SfMTPSL10	481	I
SfMTPSL11	440	I
SfMTPSL12	487	I
SfMTPSL13	455	I
SfMTPSL14	456	I
SfMTPSL15	472	I
SfMTPSL16	472	I
SfMTPSL17	341	I
SfMTPSL18	457	I
SfMTPSL19	453	I
SfMTPSL20	455	I
SfMTPSL21	455	I

Table S7. A list of sequenced genomes searched for *MTPSL genes*

Species	Data version
Amaranthus hypochondriacus	v1.0
Amborella trichopoda	v1.0
Ananas comosus	v3
Aquilegia coerulea	v1.1
Aquilegia coerulea	v3.1
Arabidopsis halleri	v1.1
Arabidopsis lyrata	v1.0
Arabidopsis thaliana	TAIR10
Boechera stricta	v1.2
Brachypodium distachyon	v3.1
Brachypodium stacei	v1.1
Brassica rapa	FPsc v1.3
Capsella grandiflora	v1.1
Capsella rubella	v1.0
Carica papaya	ASGPBv0.4
Chlamydomonas reinhardtii	v5.5
Citrus clementina	v1.0
Citrus sinensis	v1.1
Coccomyxa subellipsoidea C-169	v2.0
Cucumis sativus	v1.0
Eucalyptus grandis	v2.0
Eutrema salsugineum	v1.0
Fragaria vesca	v1.1
Glycine max	Wm82.a2.v1
Gossypium raimondii	v2.1
Kalanchoe marnieriana	v1.1
Klebsormidium flaccidum	v1.0
Linum usitatissimum	v1.0
Malus domestica	v1.0
Manihot esculenta	v6.1
Medicago truncatula	Mt4.0v1
Micromonas pusilla CCMP1545	v3.0
Micromonas sp. RCC299	v3.0
Mimulus guttatus	v2.0
Musa acuminata	v1
Oryza sativa	v7_JGI
Ostreococcus lucimarinus	v2.0
Panicum hallii	v2.0
Panicum virgatum	v1.1
Phaseolus vulgaris	v1.0
Physcomitrella patens	v3.3
Populus trichocarpa	v3.0
Prunus persica	v2.1
Ricinus communis	v0.1
Salix purpurea	v1.0
Selaginella moellendorffii	v1.0
Setaria italica	v2.2
Setaria viridis	v1.1
Solanum lycopersicum	iTAG2.3
Solanum tuberosum	v3.4
Sorghum bicolor	v3.1
Sphagnum fallax	v0.5
Spirodela polyrhiza	v2
Theobroma cacao	v1.1
Triticum aestivum	v2.2
Vitis vinifera	Genoscope.12X
Volvox carteri	v2.1
Zea mays	6a
Zostera marina	v2.2

Table S8. A list of representative *MTPSL* genes experimentally studied.

Sequence ID	Lineage	Species	Group
Liv-IRBN-MTPSL2	Liverworts	*Scapania nemorea*	I
Liv-IRBN-MTPSL4	Liverworts	*Scapania nemorea*	I
Mos-GOWD-MTPSL2	Mosses	*Sphagnum lescurii*	I
Mos-QKQO-MTPSL3	Mosses	*Pseudotaxiphyllum elegans*	II
Mos-VBMM-MTPSL3	Mosses	*Anomodon rostratus*	II
Hon-ApMTPSL7	Hornworts	*Anthoceros punctatus*	III
Mon-GSXD-MTPSL3	Monilophytes	*Myriopteris eatonii*	IV
Mon-UJTT-MTPSL4	Monilophytes	*Pityrogramma trifoliata*	IV
Mon-YJJY-MTPSL1	Monilophytes	*Woodsia scopulina*	IV

Table S9. Primers used in this study

Region amplified	Primer	DNA sequences
ApMTPSL1 and its	Forward	5'-CACTACTGCGTCGGCTTCATG-3'
neighoring gene	Reverse	5'-CGCACAGCATTCACAATTTCACTT-3'
ApMTPSL2 and its	Forward	5'-CAGGTAGGAGCCCGCGATTT-3'
neighoring gene	Reverse	5'-AGGGAAAAGGAGGGTGGTG-3'
SfMTPSL1 and its	Forward	5'-CAGAAGCAAAGTATCGGTCTCTTAC -3'
neighoring gene	Reverse	5'-CACTGTTAGCAGGGTATGGTGAAC-3'
Liv-IRBN-MTPSL1	Forward	5'-TTCTGAGGACGAGCGTATTCTTC-3'
	Reverse	5'-GCAAAACGTCAACTAAACGAGAAG-3'
Liv-IRBN-MTPSL2	Forward	5'-TCATACTCGCCTCCATATCCTGTG-3'
	Reverse	5'-GATTTGAAATGTCAGTCATGTGTGC-3'
Liv-IRBN-MTPSL3	Forward	5'-GATGCCAACGCAGCCATACAGAC-3'
	Reverse	5'-GCCTGATACCCAGTTTCTGACGG-3'
Liv-IRBN-MTPSL4	Forward	5'-CAGTATGTGTGAACTCCTCTTGGGTC-3'
	Reverse	5'-GCACTCCTTTTCTGTACCGACTGG-3'
Liv-IRBN-MTPSL5	Forward	5'-TCAAAGGCATCACCTGAAGTCTG-3'
	Reverse	5'-ATATTATCGGTGTTCCAATCCTCC-3'
Liv-IRBN-MTPSL6	Forward	5'-AATGCTTGGTGTGTGTTCGTCTC-3'
	Reverse	5'-CCTCCATGTGATTTCGCAAAGTAG-3'
Liv-IRBN-MTPSL7	Forward	5'-CTGGCAGATGATTTAGATGAGATAGC-3'
	Reverse	5'-CAGAAACAACCCGCAAACCATTC-3'
Liv-IRBN-MTPSL8	Forward	5'-TCTCCCTGTTGCCACTGCTTTCC-3'
	Reverse	5'-GTTGGTCCTGGTACGGCGACTGA-3'
Mon-UJTT-MTPSL4-pET32a	Forward	5'-CATGCCATGGCATCCATTATATTAGGAAGCTC-3'
	Reverse	5'-CCCAAGCTTAGTTAAAGGCCATCATGACAC-3'

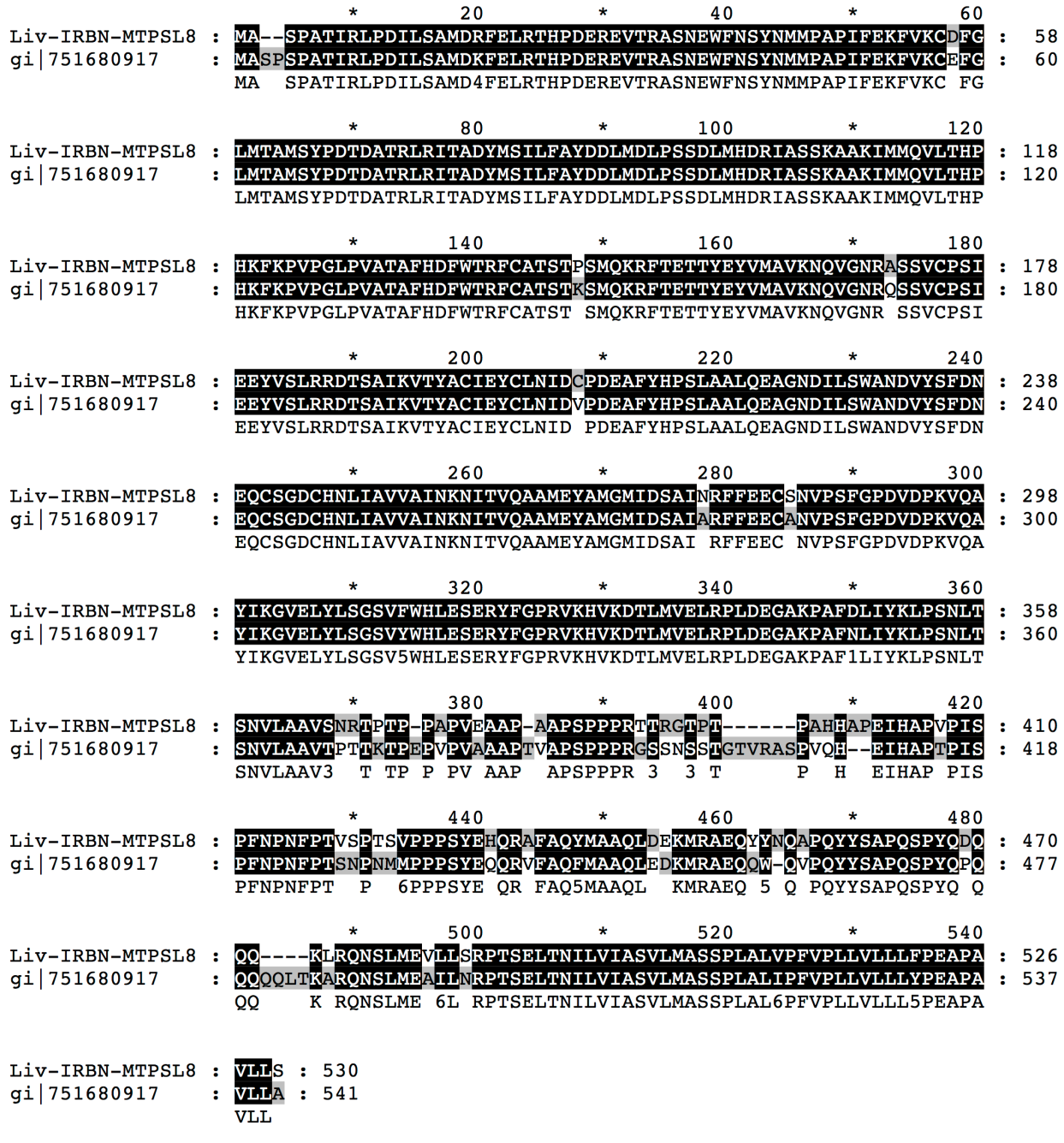

Fig. S1. Sequence alignment of IRBN_MTPSL6 identified from the transcriptome of the liverwort *Scapania nemorea* with its top hit in nr database at NCBI. "gi:751680917" is putative terpene synthase gene identified in the fungus *Serendipita vermifera*. The two sequences are 91% identical.

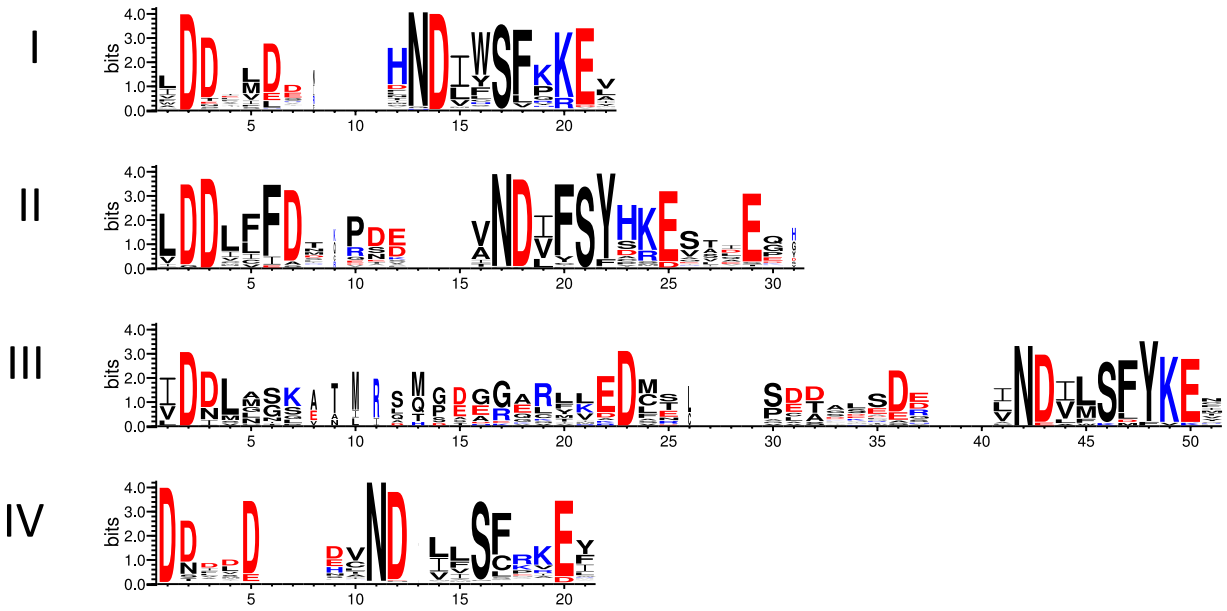

Fig. S2. Identification of conserved motifs among the four groups (I, II, III and IV) of microbial type (MTPSL) terpene synthases. The 'NDxxSxxxD/E'motif is highly conserved among all MTPSLs. The canonical 'DDxxD' motif was present in group I and IV proteins, but group II proteins displayed a conserved 'DDxxxD' motif, whereas in the group III enzymes only the first two aspartates ('DD') are conserved. Sequence motif logos made using weblogo 3.0, showing the conserved motifs found in each group of terpene synthase genes of microbial type.

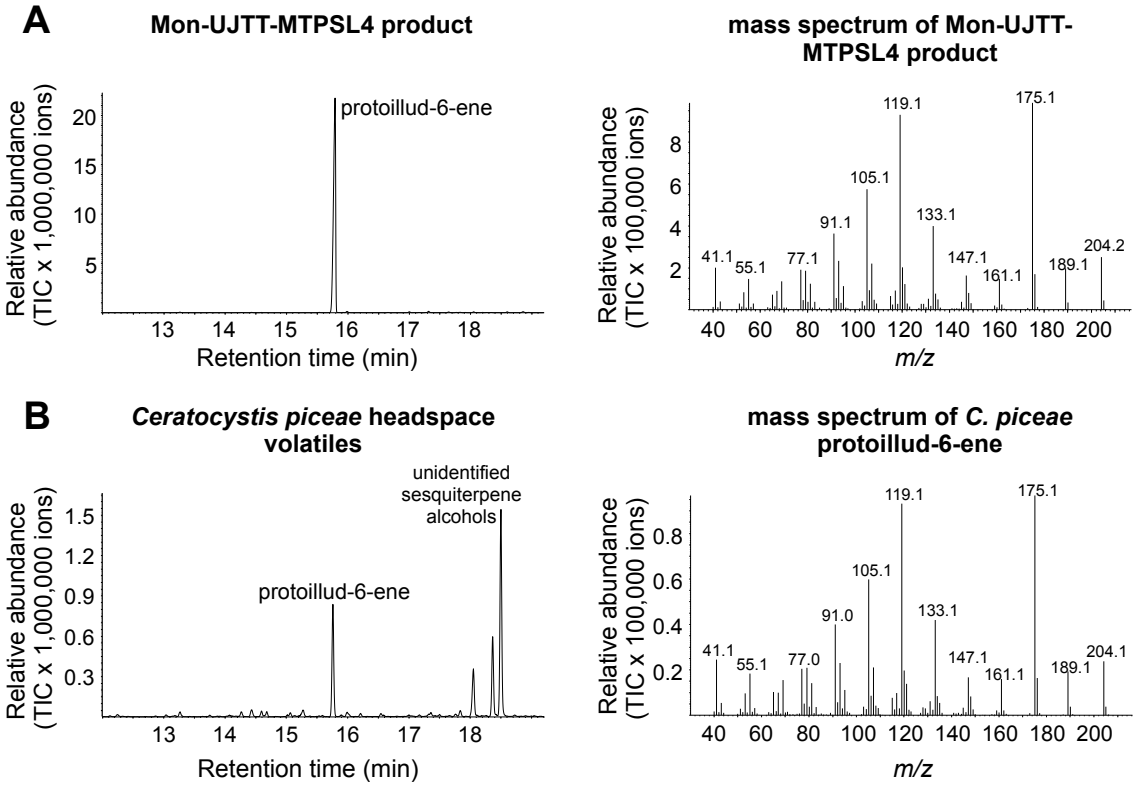

Fig. S3. Product identification of Mon-UJTT-MTPSL4. (*A*) The gene was heterologously expressed in *Escherichia coli* and the crude bacterial protein extract was incubated with FPP. The enzyme product was collected using solid phase micro extraction (SPME) and analyzed with gas chromatography/mass spectrometry (GC-MS). (*B*) Volatiles from the headspace of a liquid culture of *Ceratocystis piceae* were collected using SPME and analyzed with GC-MS. *C. piceae* has been reported to produce protoillud-6-ene as main sesquiterpene hydrocarbon (17).

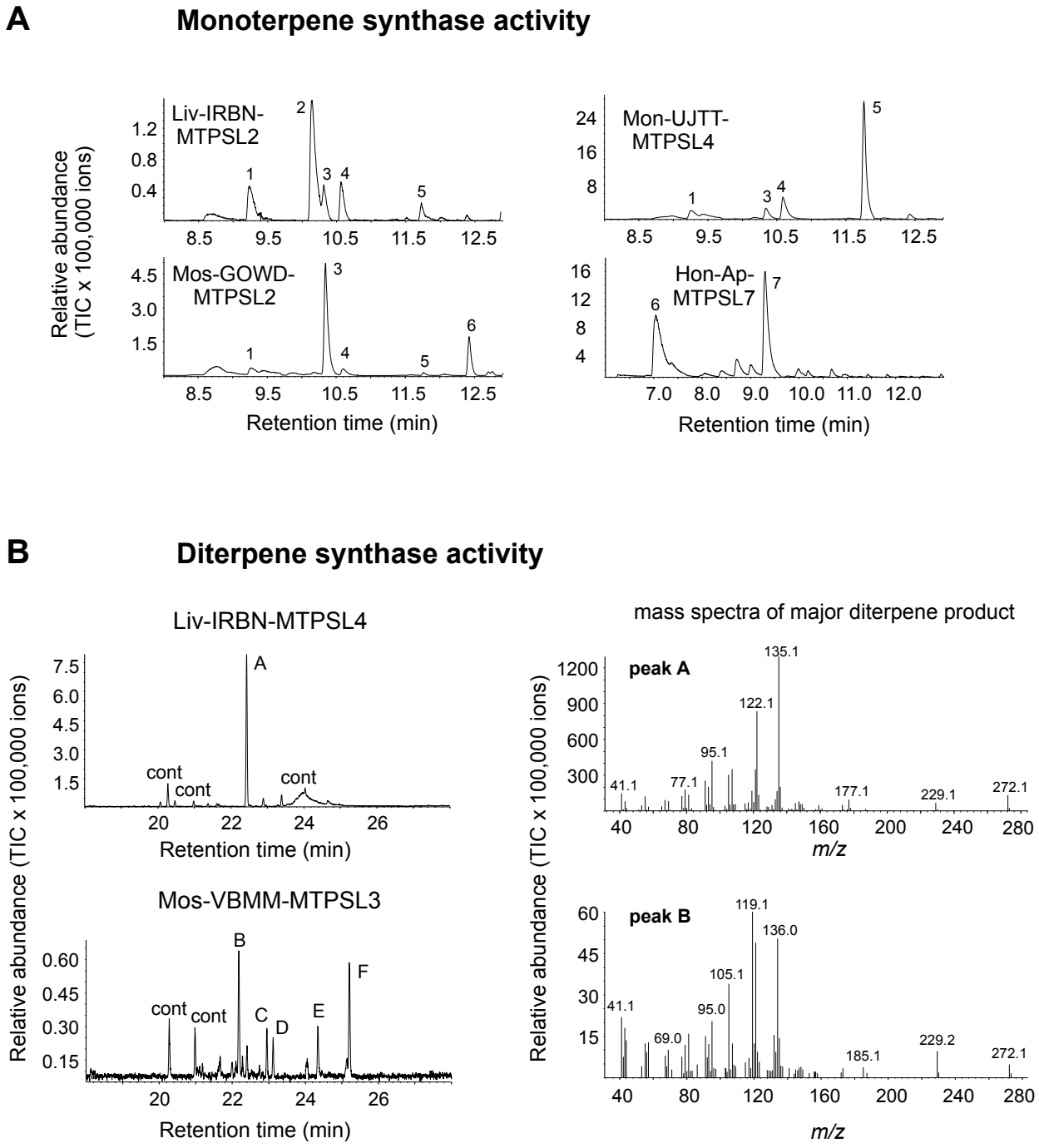

Fig. S4. Monoterpene synthase and diterpene synthase activities of MTPSLs. *MTPSL* genes were heterologously expressed in *Eschericha coli* and and crude protein extracts were incubated with the potential substrates GPP (*A*) and GGPP (*B*), respectively. Monoterpene products were collected using solid-phase micro-extraction and diterpene products were extracted with hexane. Products were analyzed using gas chromatography/mass spectrometry. 1, myrcene*; 2, limonene*; 3, (*Z*)-β-ocimene; 4, (*E*)-β-ocimene*; 5, linalool*; 6, allo-ocimene; 6, α-pinene; 7, β-phellandrene. A-F, unidentified diterpenes. Compounds marked with an asterisk (*) were identified using authentic standards.

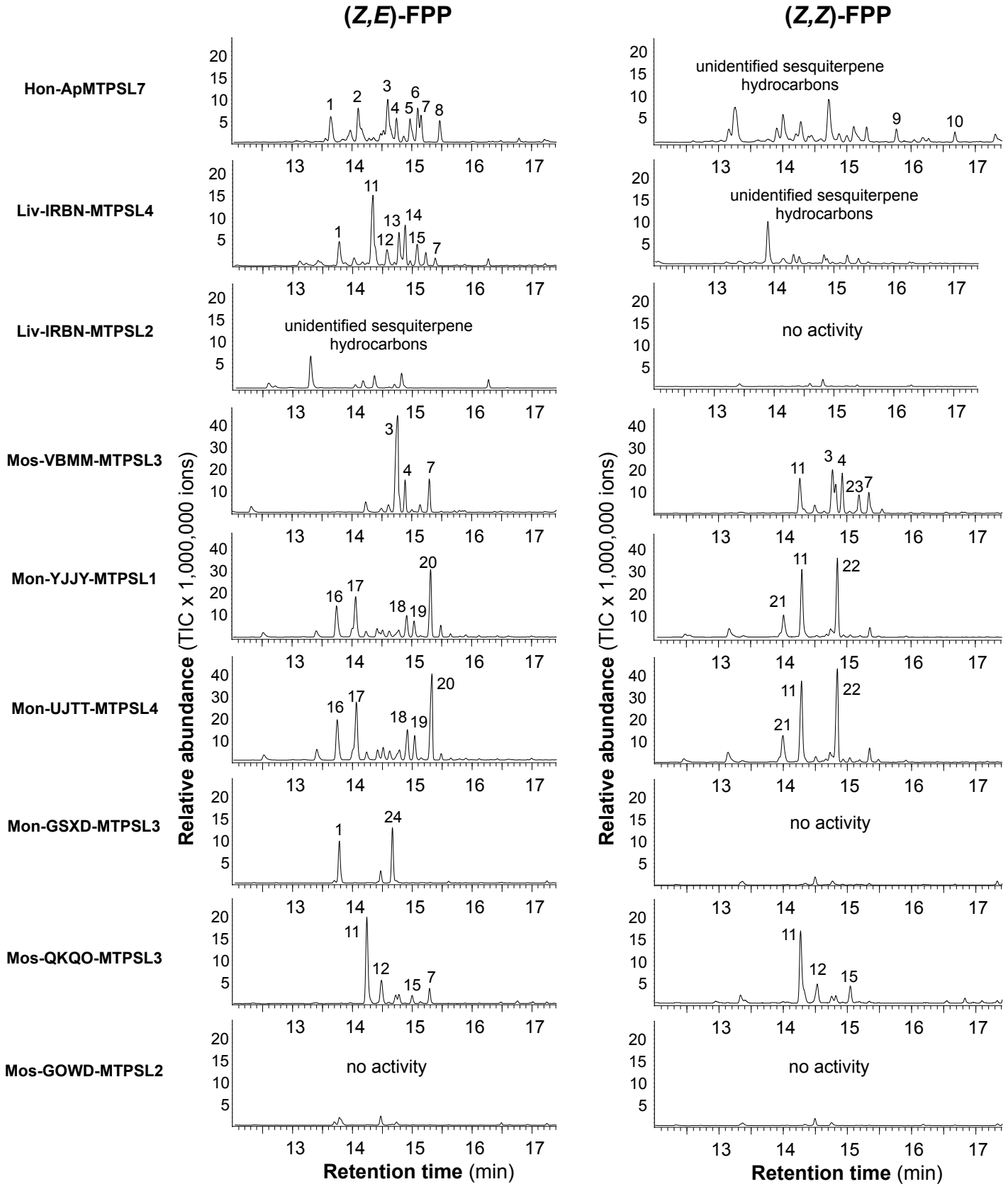

Fig. S5. Sesquiterpene synthase activity of representative MTPSLs with (*Z,E*)-FPP and (*Z,Z*)-FPP. *MTPSL* genes were heterologously expressed in *Eschericha coli* and crude protein extracts were incubated with the potential substrates (*Z,E*)-FPP and (*Z,Z*)-FPP. Enzyme products were collected using solid-phase micro-extraction and analyzed with gas chromatography/ mass spectrometry. 1, (*E*)-β-farnesene; 2, β-acoradiene; 3, β-bisabolene; 4, (*Z*)-γ-bisabolene; 5, unidentified ST; 6, unidentified ST; 7, (*Z*)-α-bisabolene; 8, nerolidol; 9, unidentified oxygenated ST; 10, unidentified oxygenated ST; 11, γ-curcumene; 12, zingiberene; 13, unidentified ST; 14, unidentified ST; 15, β-sesquiphellandrene; 16, unidentified ST; 17, *epi*-bicyclosesquiphellandrene; 18, γ-cadinene; 19, δ-cadinene; 20, α-cadinene; 21, unidentified ST; 22, unidentified ST; 23, (*E*)-γ-bisabolene; 24, (*E,E*)-α-farnesene. ST, sesquiterpene.