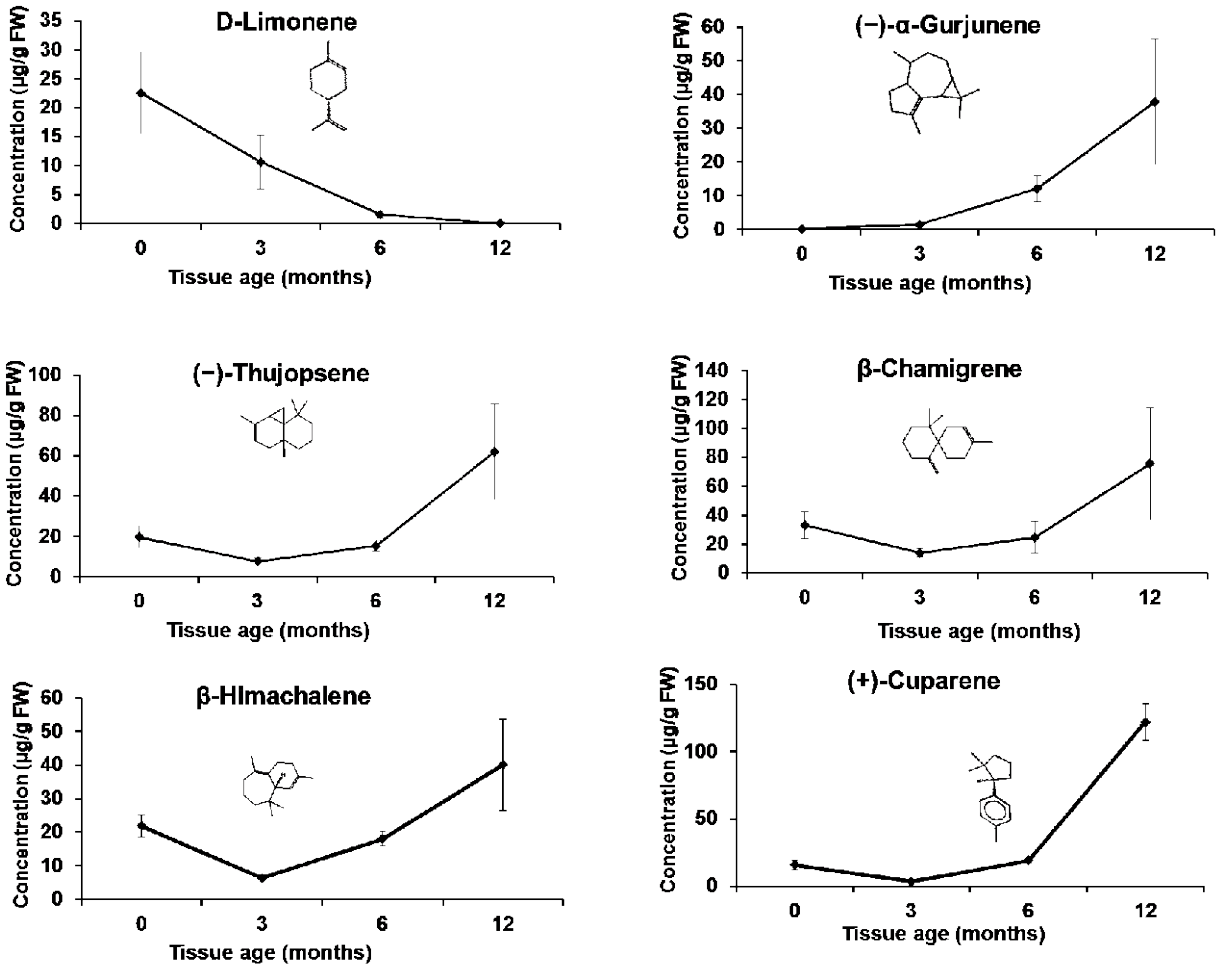
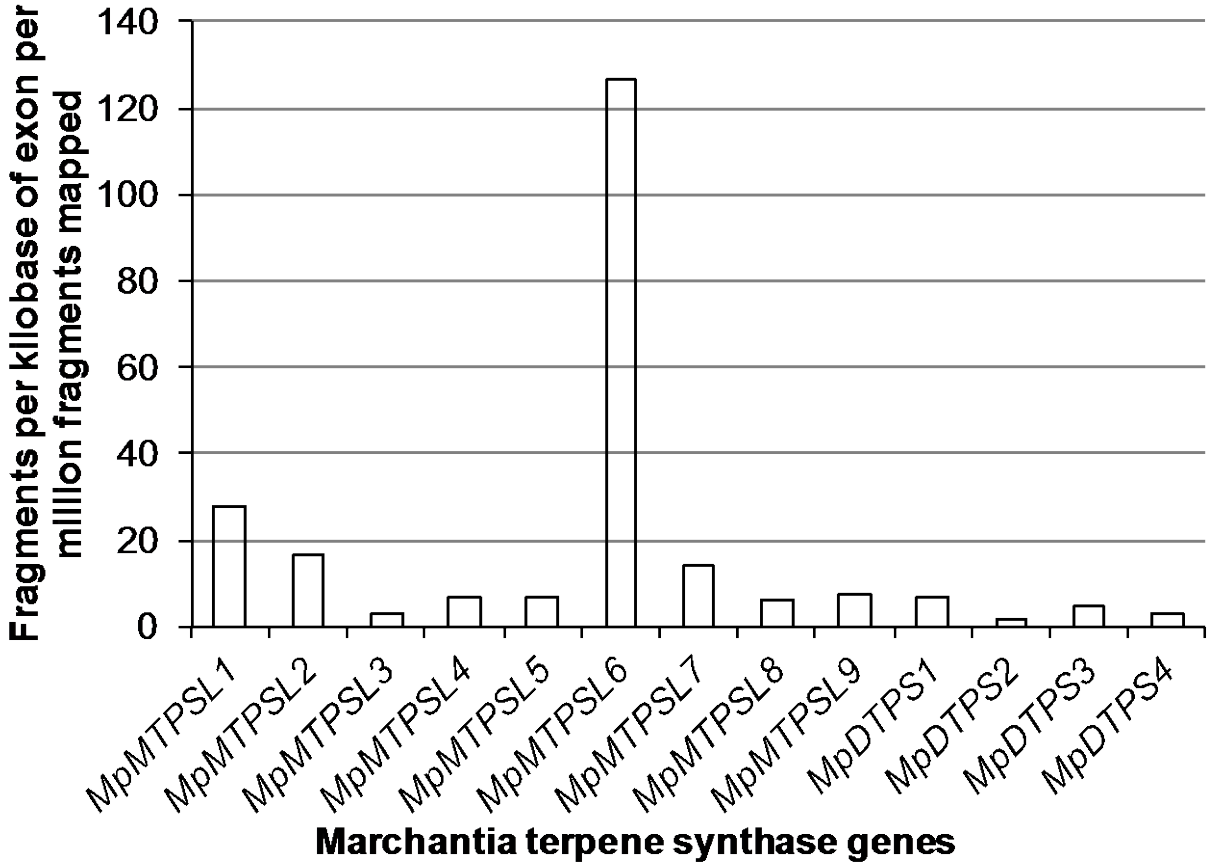
SUPPLEMENTAL DATA

SUPPLEMENTAL FIGURES



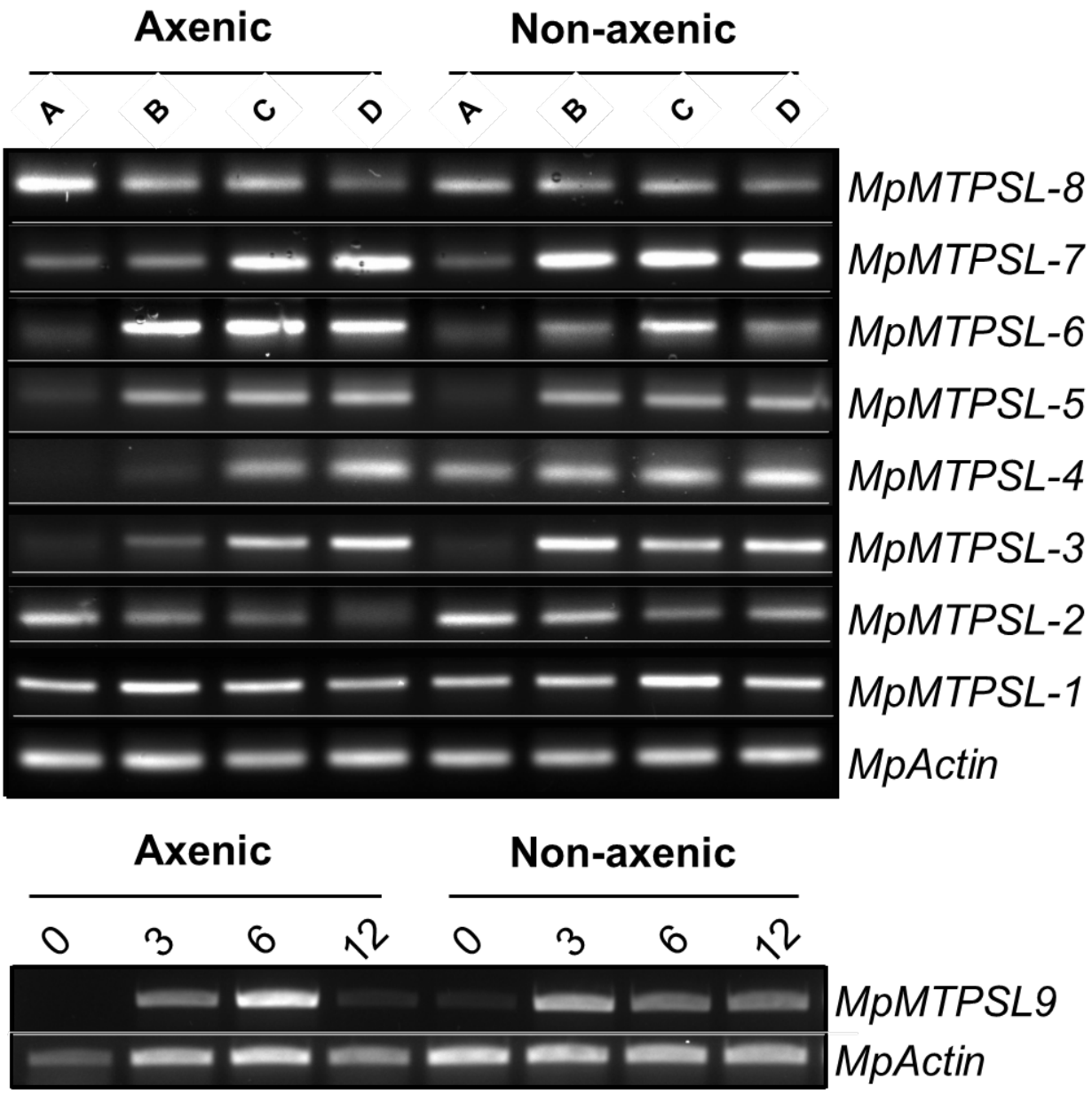
Supplemental Figure 1. Major terpenes present in axenic (sterile) cultures of *Marchantia polymorpha* harvested after 0, 3, 6, 12 months of growth.



Supplemental Figure 2. Measurement of terpene synthase-like gene expression based on the Fragments per kilobase of exon per million fragments mapped (FPKM).

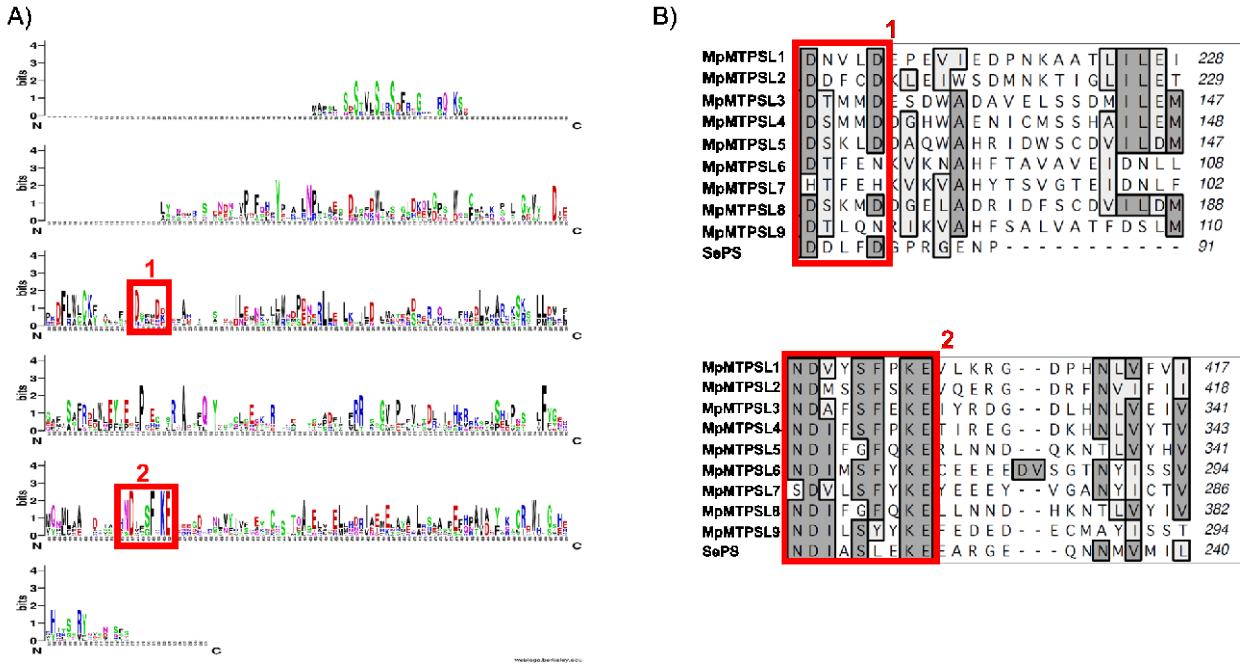
We relied on the methods of Yeo et al. (2013) to validate the use of map reads to gain a relative appreciation for transcripts levels in *M. polymorpha*, which were corroborated by RT-PCR measurements (see Supplemental Figure 5).

Yeo, Y.S. et al. (2013) Functional identification of valerena-1,10-diene synthase, a terpene synthase catalyzing a unique chemical cascade in the biosynthesis of biologically active sesquiterpenes in *Valeriana officinalis*. J. Biol. Chem. 288: 3163-73.



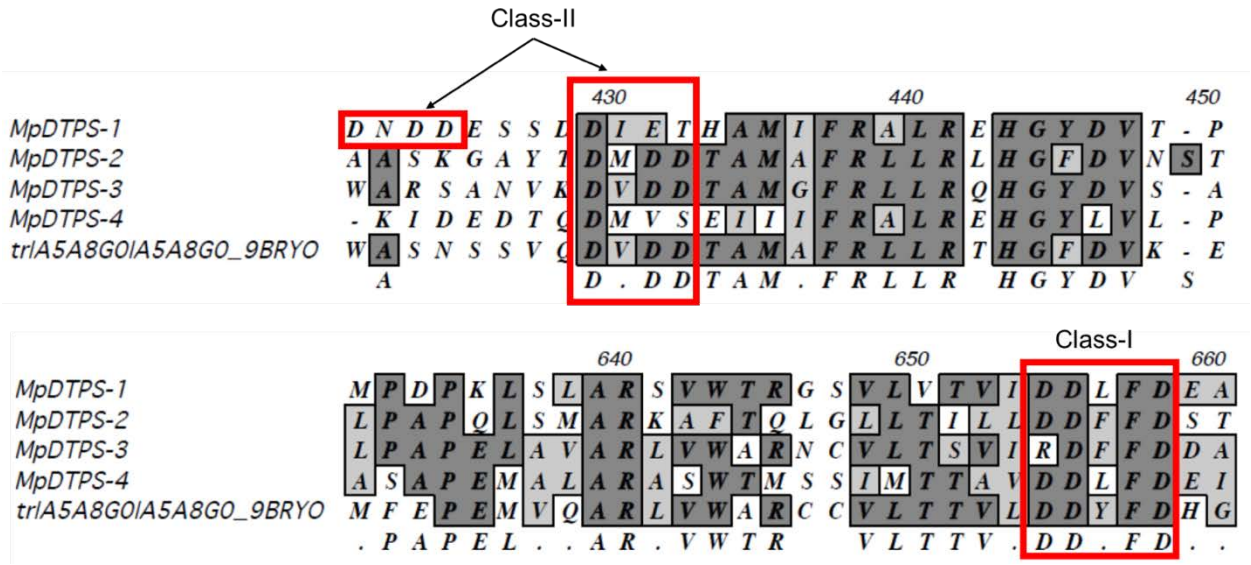
Supplemental Figure 3. Relative expression of the terpene synthase-like genes (Mp*MTPSL*) from axenic as well as non-axenic *M. polymorpha* tissue grown for 0 (A), 3 (B), 6 (C) and 12 (D) months.

Qualitative PCR was performed for terpene synthase like genes using 250 ng of cDNA templates from axenic as well as non-axenic cultures of *M. polymorpha* as described in detail in the Materials and Methods.



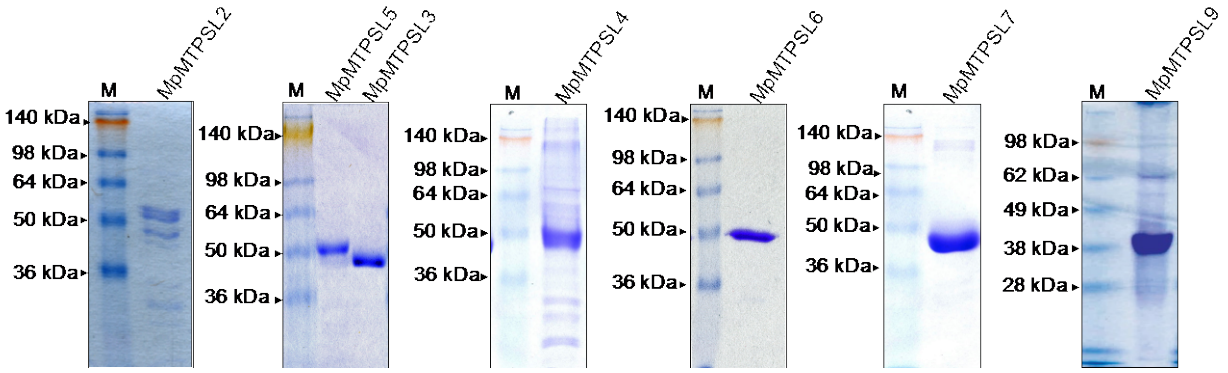
Supplemental Figure 4. Web logo based consensus sequence pattern for *Marchantia* terpene synthase like genes and Aspartate-rich substrate binding motif description using clustalW alignment.

Aspartate-rich substrate binding motif was calculated using WebLogo based on an alignment of the 9 *Marchantia* terpene synthase like genes (MpMTPSL1-9) (A). The bigger the letter, the more conserved the amino acid site. The alignment was produced from commercially available Mac-vector plugins for Clustal W using MpMTPSL1-9 along with the bacterial pentalene synthase gene (GenBank Accession AAA19131) (B). The conserved metal binding motifs, the DDXXD motif (1) and the NDXXSXXXE motif (2) are highlighted in red.



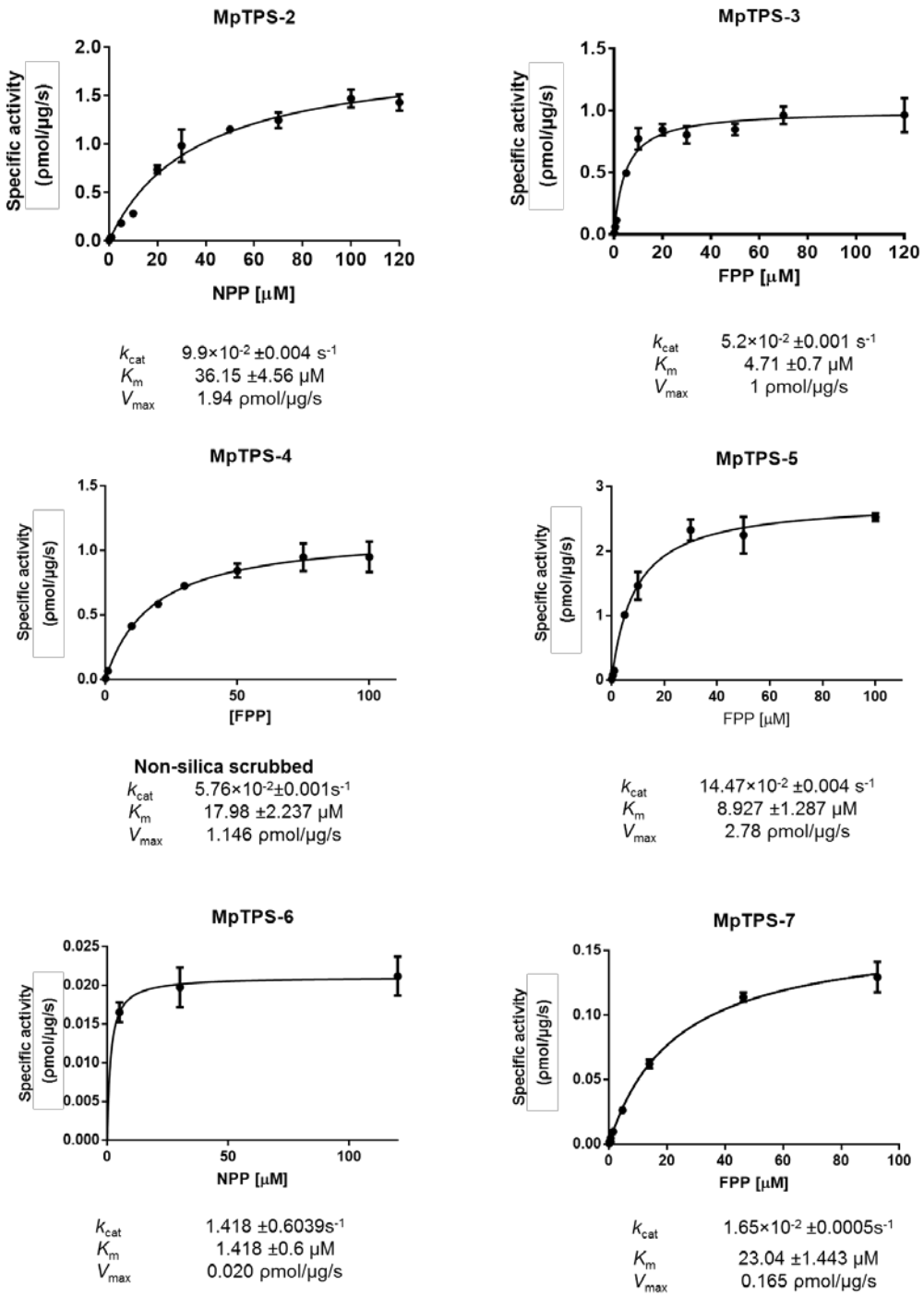
Supplemental Figure 5. Clustal W alignment of the diterpene synthase-like genes present in *M. polymorpha* (MpDTPS 1-4) centered on Class I (DXDD) and Class II (DDXXD) divalent metal binding motifs.

Alignment was produced from commercially available Clustal W plugins for MacVector using MpDTPS 1-4 along with *Physcomitrella patens* ent-kaurene synthase gene (PpCPS/KS) (GenBank Accession AB302933.1). The conserved aspartate-rich motifs for Class-I and Class-II are highlight by red blocks.



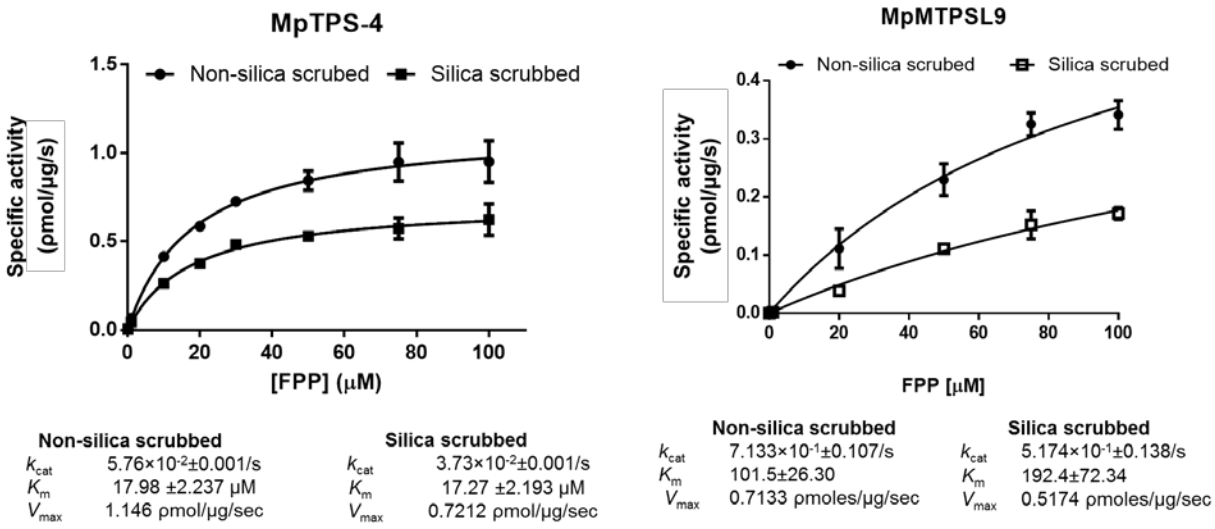
Supplemental Figure 6. Purification of recombinant MpMTPSL terpene synthases.

Coomassie Blue-stained SDS-PAGE gel, showing recombinant MpMTPSL-*2, 3, 4*, 5, 6 and 7 expressed in *E. coli* BL21-DE3 and after Co^{2+}-affinity purification. Separation of 2.5 µg of each protein was performed on 10% discontinuous SDS-polyacrylamide gels.



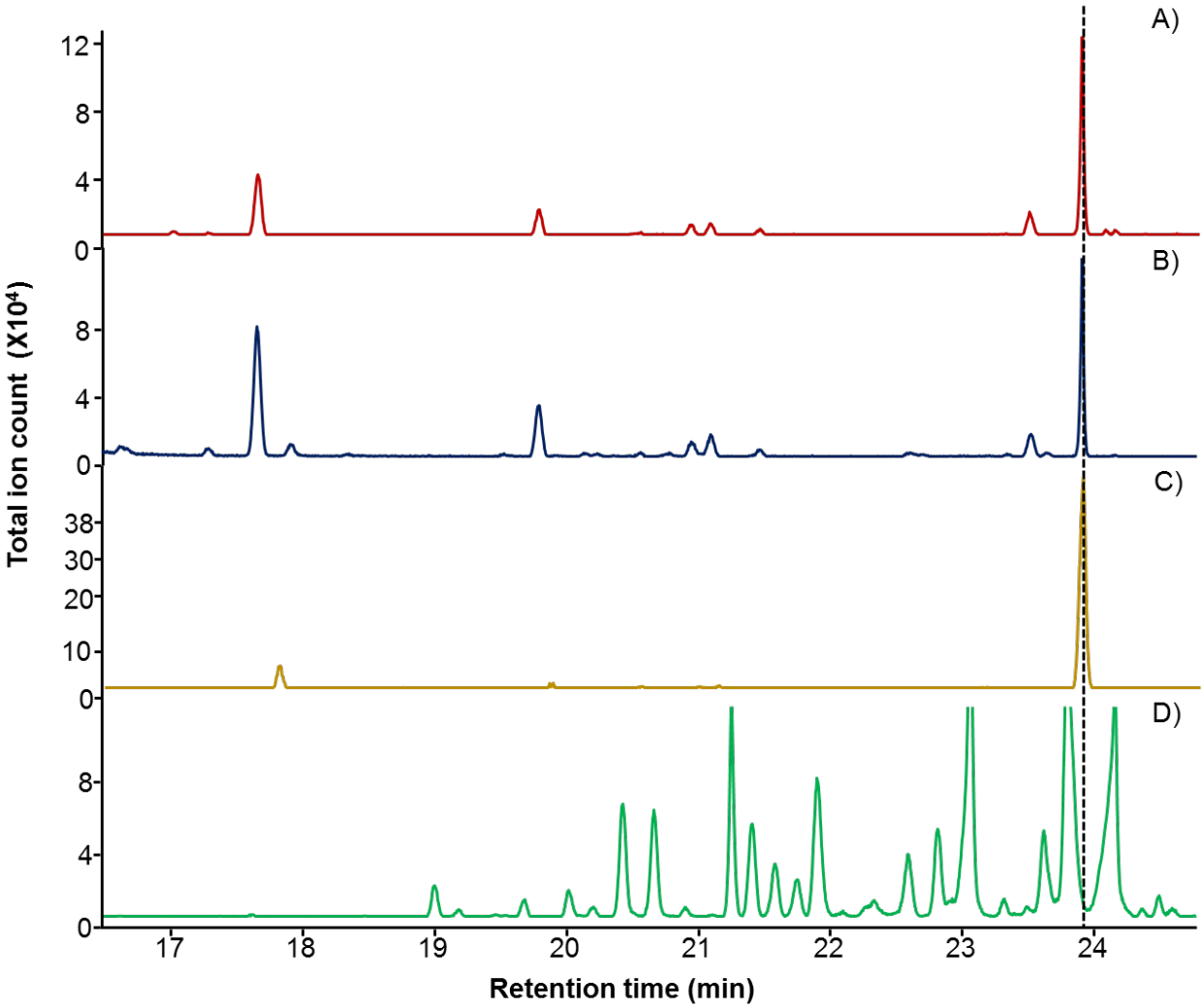
Supplemental Figure 7A. Enzyme kinetic determinations for MpMTPSL2, 3, 4, 5, 6, 7.

Enzyme assays (50 µl) were set up with purified MpMTPSL1-7 at 100 nM and the indicated concentration of ^{3}H-NPP, ^{3}H-GPP, and ^{3}H-FPP. Assays were incubated for 5 min at 37°C and stopped by addition of 50 µl stop buffer. The reactions were then extracted with 200 µl of hexane and radioactivity determined in aliquots by scintillation spectrometry. The data was analyzed using the Prism Graphpad 6.0. Data represents mean of triplicate assays.



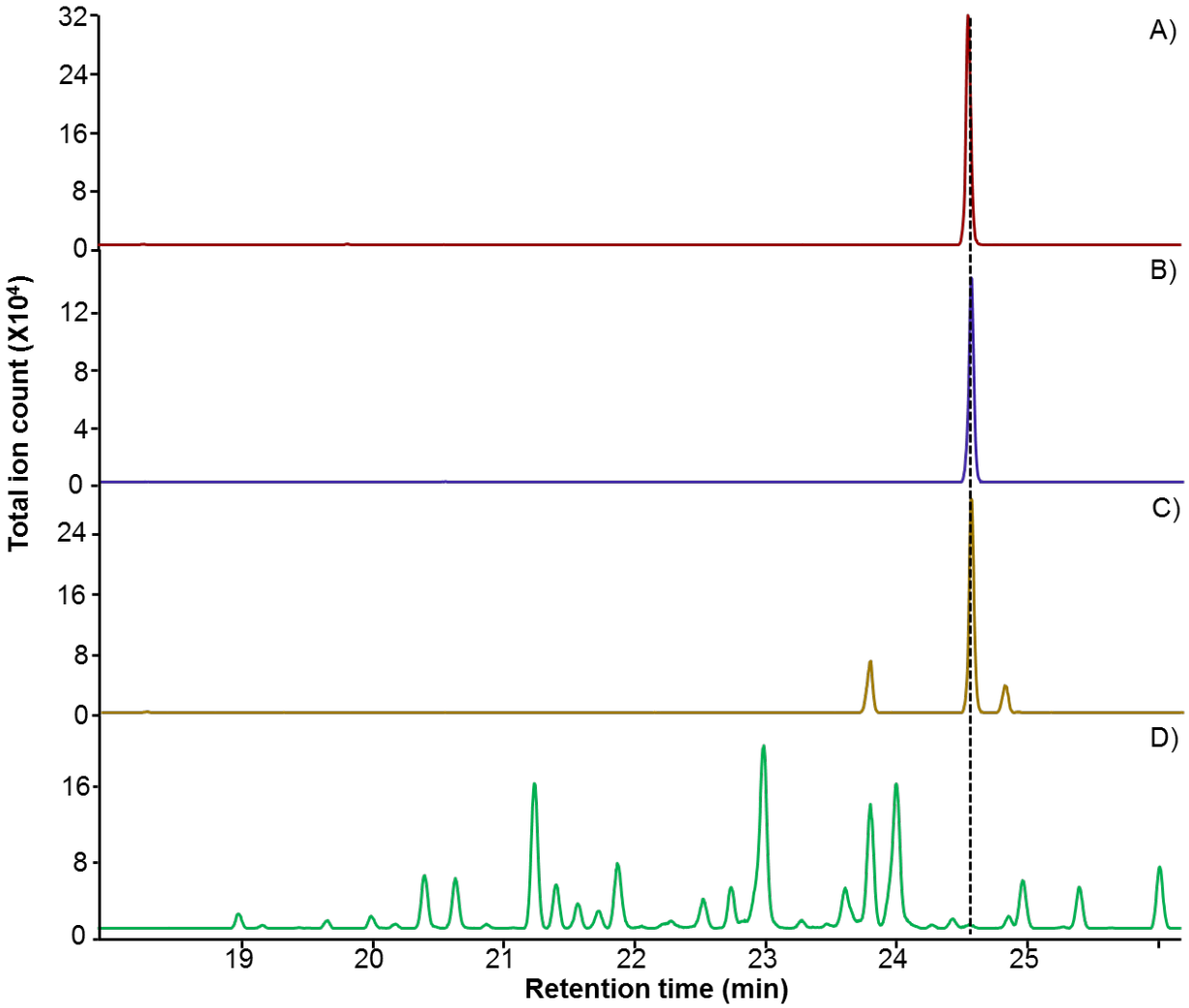
Supplemental Figure 7B. Enzyme kinetic determinations for MpMTPSL4 and MpMTPSL9 for total (non-scrubbed) and all hydrocarbon (scrubbed) reaction products.

Enzyme assays (50 µl) were set up with purified MpMTPSL4 or MpMTPSL9 at 100 nM and the indicated concentration of ^{3}H-FPP. Assays were incubated for 5 min at 37°C and stopped by addition of 50 µl stop buffer. The reactions were then extracted with 200 µl of hexane, aliquots which were subject to silica-scrub or not prior to scintillation counting. The data was analyzed using the Prism Graphpad 6.0. Data represents mean of triplicate assays.



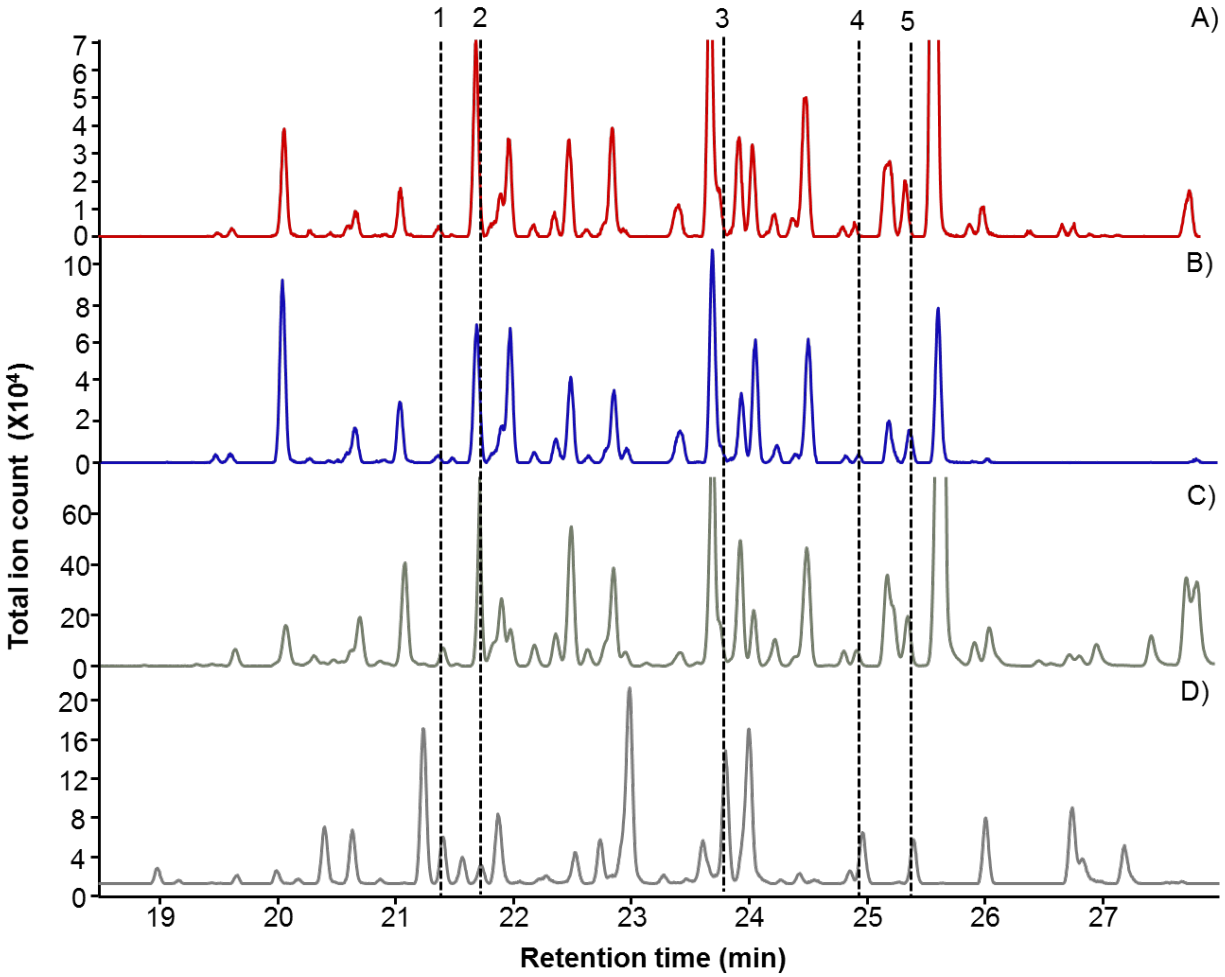
Supplemental Figure 8. GC chromatograms of terpene reaction product(s) generated by MpMTPSL3 *in vitro* and *in vivo.*

GC chromatograms of the in vitro products formed by MpMTPSL3 (~100 nM) incubated with 100 µM FPP (A) and the *in vivo* products generated by *E. coli* (B) or yeast (C) cultures expressing the Mp*MTPS3* gene. GC chromatogram of extractable terpenes from *M. polymorpha* (D) and annotated for the overlapping products (dashed black line) present in multiple samples.



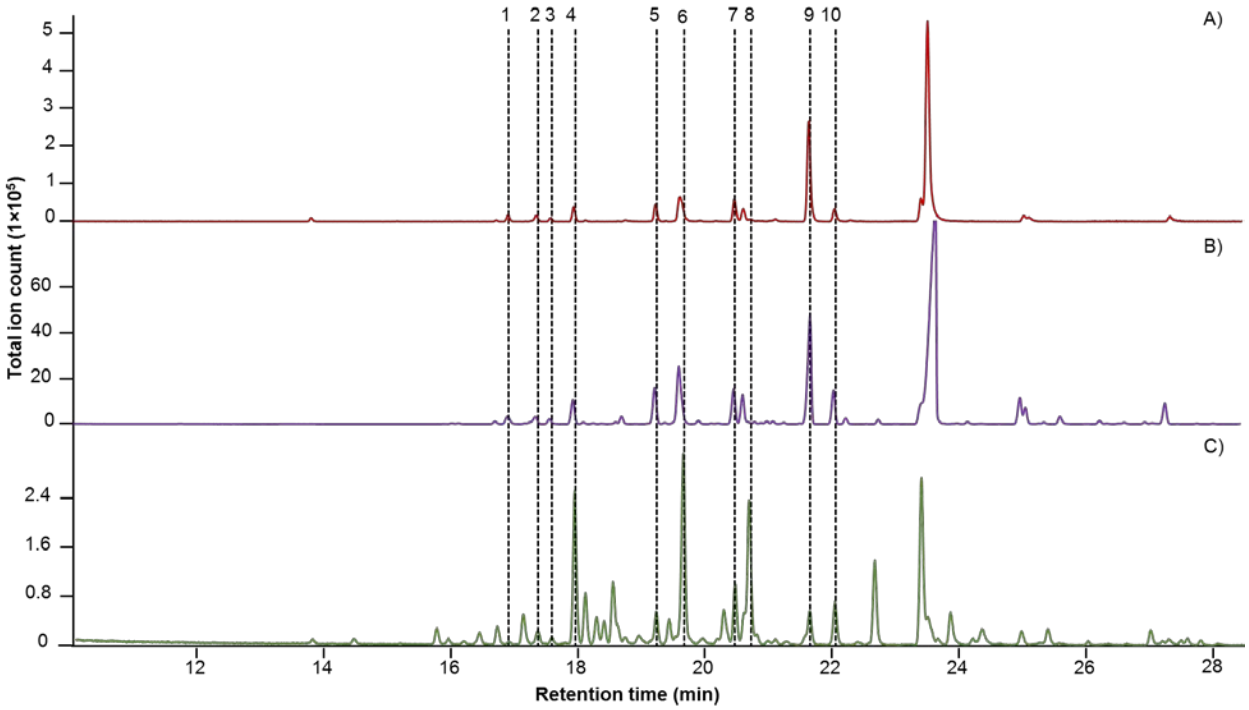
Supplemental Figure 9. Gas chromatogram of terpene reaction product(s) generated by MpMTPSL5 in vitro and in vivo.

GC chromatograms of the in vitro products formed by MpMTPSL5 (~100 nM) incubated with 100 µM FPP (A) and the in vivo products generated by *E. coli* (B) or yeast (C) cultures expressing the Mp*MTPSL5* gene. GC chromatogram of extractable terpenes from *M. polymorpha* (D) and annotated for the overlapping products (dashed black line) present in multiple samples.



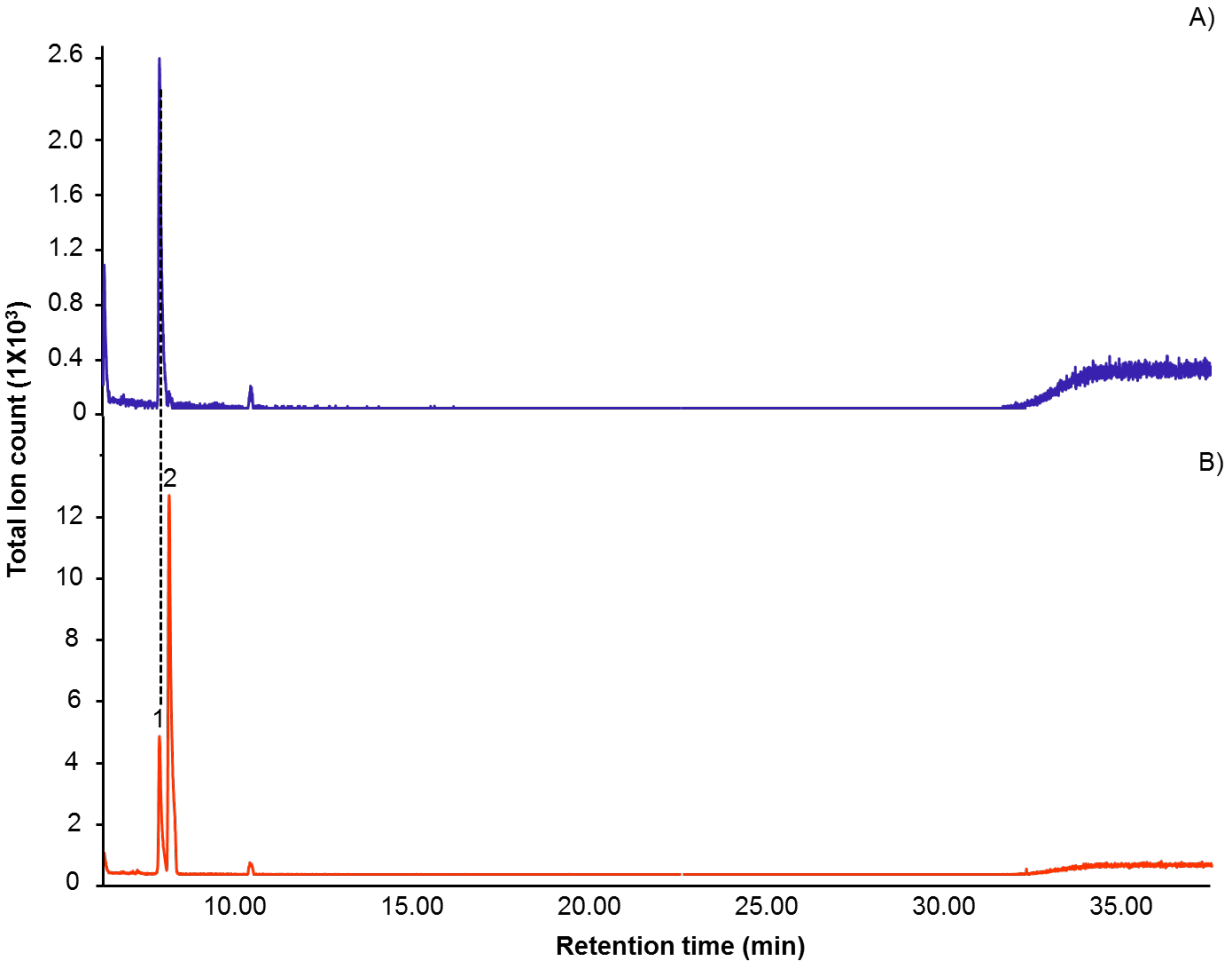
Supplemental Figure 10. Gas chromatogram of terpenes generated by MpMTPSL7 in vitro and in vivo.

GC chromatograms of the in vitro products formed by MpMTPSL7 (~100 nM) incubated with 100 µM FPP (A) and the *in vivo* products generated by *E. coli* (B) or yeast (C) cultures expressing the Mp*MTPSL7* gene. GC chromatogram of extractable terpenes from *M. polymorpha* (D) and annotated for the overlapping products (dashed black line) present in multiple samples.



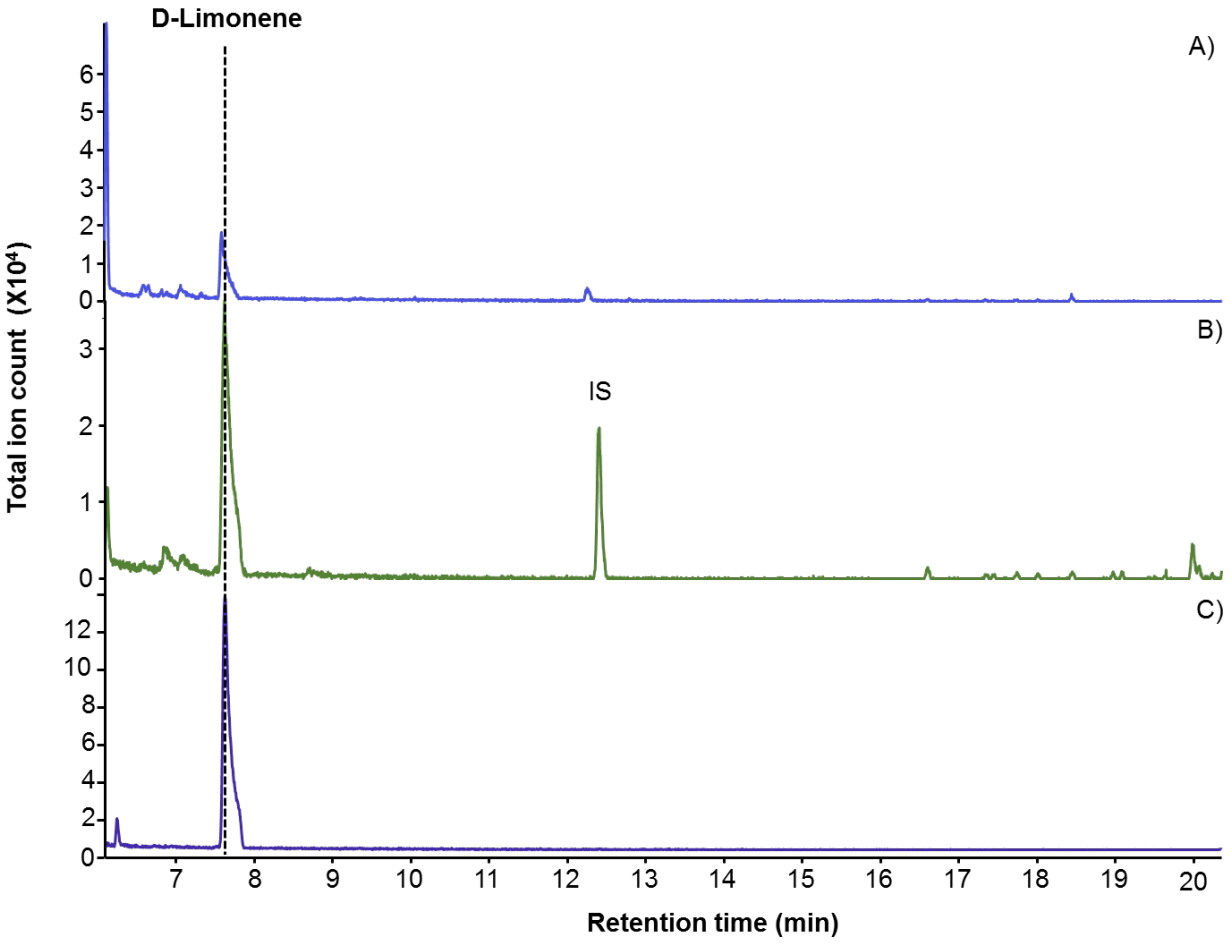
Supplemental Figure 11. Gas chromatogram of terpenes generated by MpMTPSL9 in vitro and in vivo.

GC chromatograms of the in vitro products formed by MpMTPSL9 (~100 nM) incubated with 100 µM FPP (A) and the in vivo products generated by *E. coli* (B) and GC chromatogram of extractable terpenes from *M. polymorpha* (C) and annotated for the overlapping products (dashed black line) present in multiple samples.



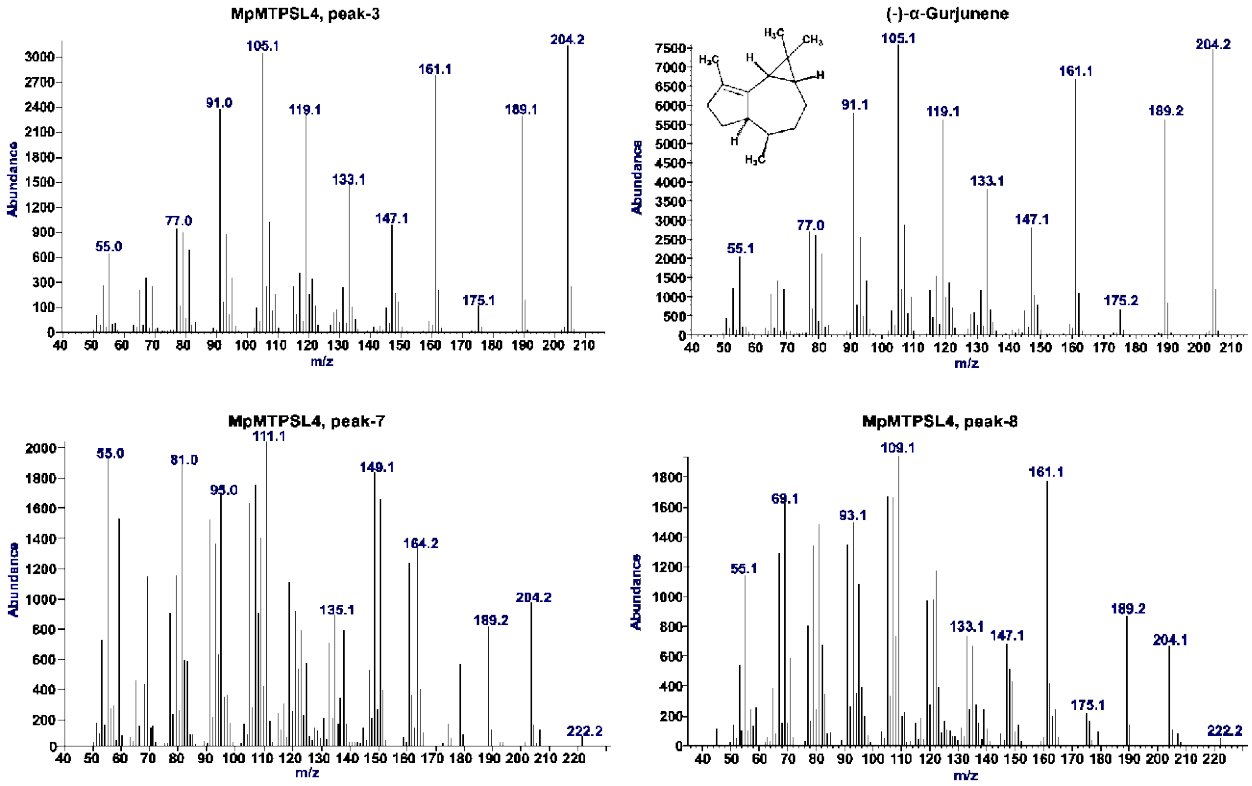
Supplemental Figure 12. GC chromatogram of terpene reaction product(s) generated in vitro by MpMTPSL6.

GC chromatogram of the in vitro products formed by MpMTPSL6 (~100 nM) incubated with 100 µM GPP (A) in comparison to a chromatogram for ocimene (B) where the two peaks in chromatogram represents cis-β-ocimene (1) and trans-β-ocimene (2).

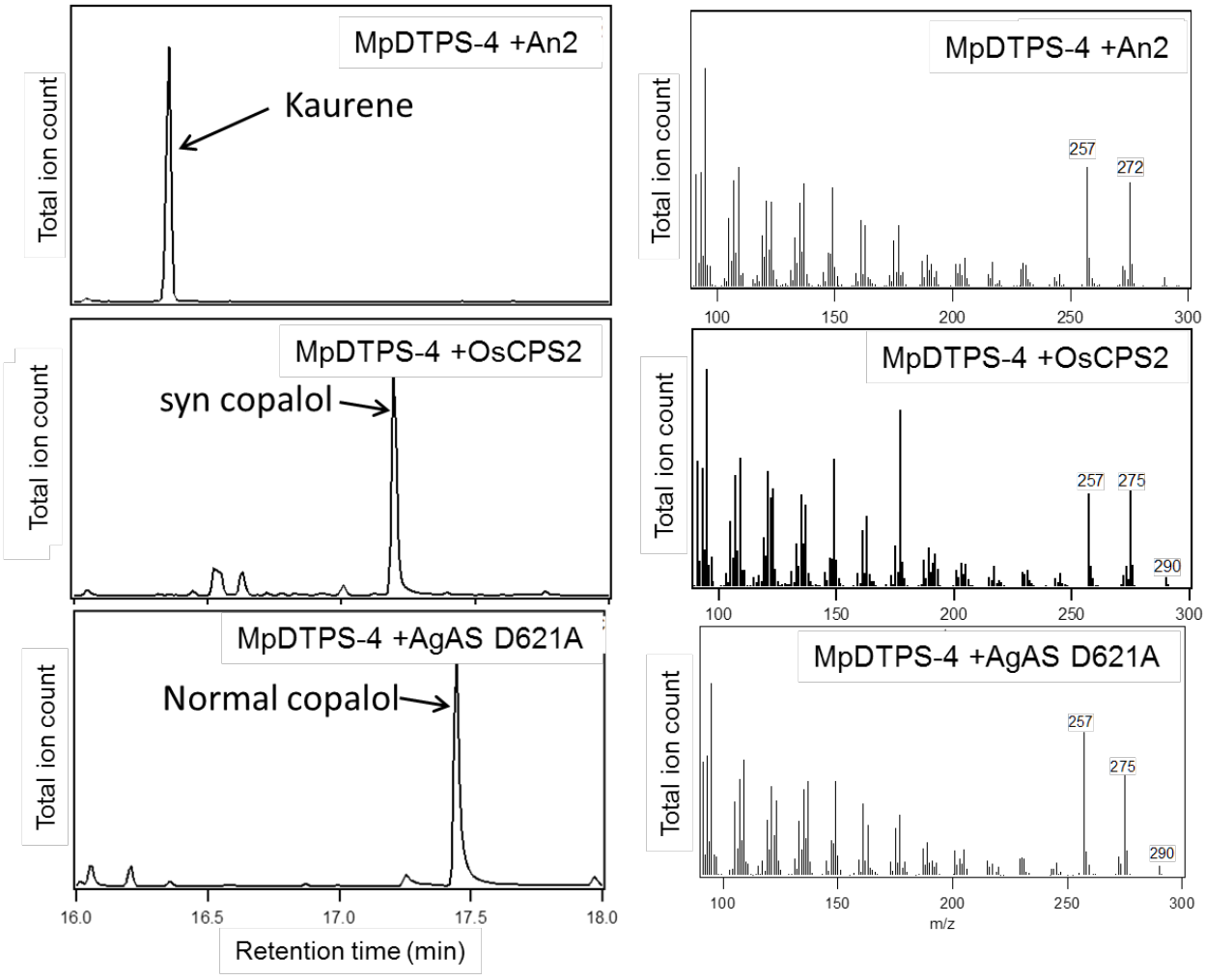


Supplemental Figure 13. GC chromatograms of terpene reaction product(s) generated by MpMTPSL2 in vitro using NPP as substrate.

GC chromatograms of the in vitro products formed by MpMTPSL2 (100 nM) incubated with 100 µM NPP (A) in comparison to the D-limonene in *M. polymorpha* (0 month axenic culture, prior to developmental accumulation of most sesquiterpenes), and an authentic D-limonene standard.

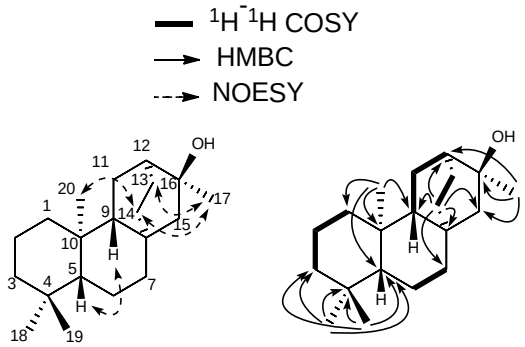


Supplemental Figure 14. Mass spectra of selected compounds produced by MpMTPSL4 as shown in Figure 6 and for (-) alpha-gurjunene standard.

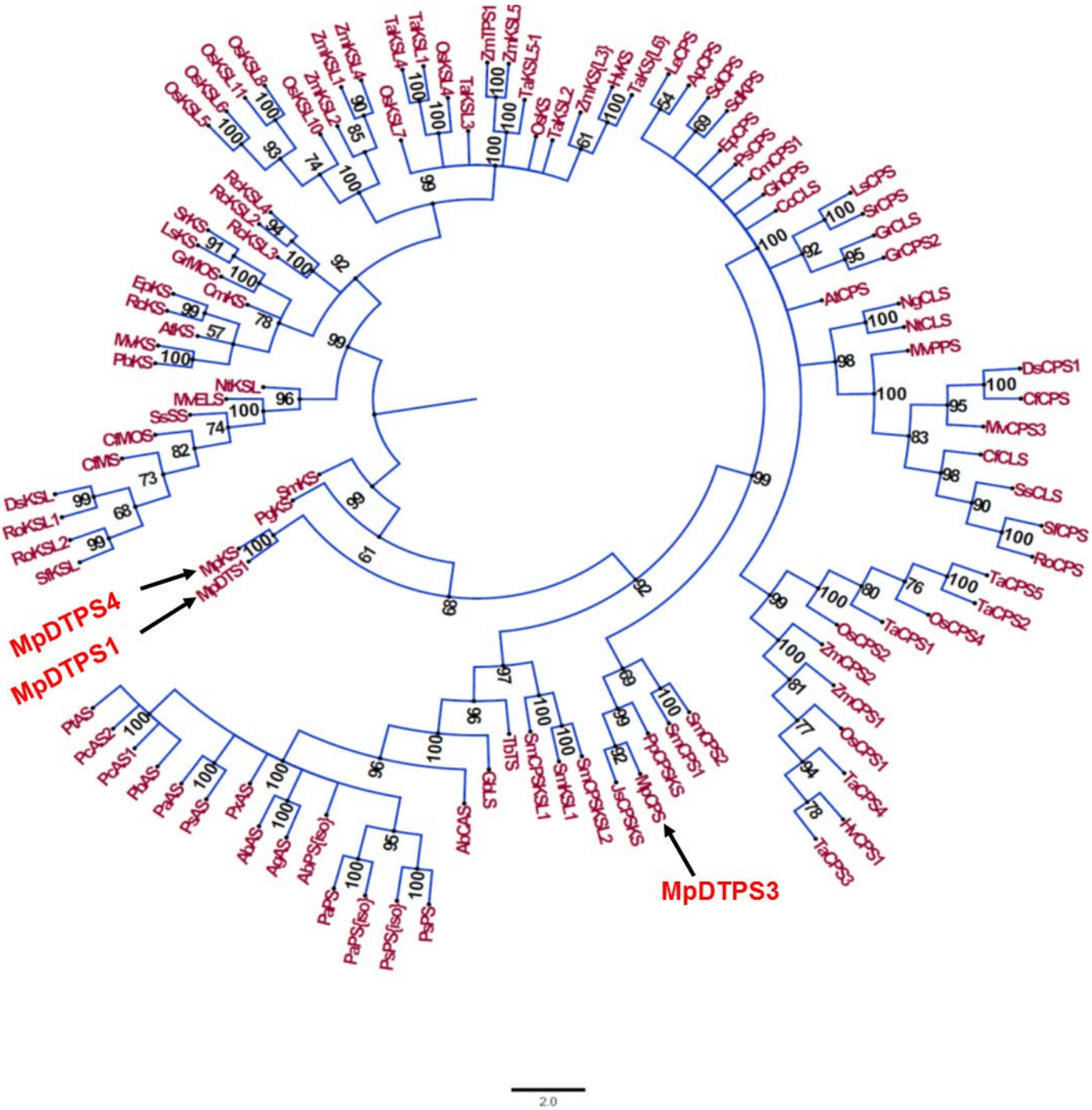


Supplemental Figure 15. GC-MS analysis of diterpenes products generated by co-expression of Mp*DTPS4*, GGPP synthase plus An2 or OsCPS2 or AgAS D621A, which are copalyl diphosphate synthases (CPS) that produce ent, syn and normal CPP from GGPP, respectively.

MpDTPS4 reacts with *ent*-CPP to form *ent*-kaurene when co-expressed with An2 from maize, while it remains largely unreactive with *syn* and normal CPP as indicated by unreacted *syn* and normal copalol, respectively when co-expressed with OsCPS2 or AgAS D621.



Supplemental Figure 16. Numbering and selected ^{1}H-^{1}H COSY, HMBC and NOESY correlations for atiseranol.



Supplemental Figure 17. Phylogenetic relationships of Mp*DTPS1, 3 and 4* to other plant mono-functional CPSs and KSs and a bifunctional KS from *Physcomitrella* as inferred using the Neighbor-Joining method (Saitou and Nei, 1987).

The *Marchantia* diterpene synthases (MpDTPS) phylogenetic tree was constructed based on the amino acid alignment presented in Supplemental Data Set 3. These terpene synthase sequences were downloaded from BLASTP search analysis tool at NCBI (Altschul et. al. 1997). Sequences for mono-functional CPSs and KSs and a bi-functional KS from *Physcomitrella* were selected based on sequence similarities. The multiple

sequence alignment was performed with the amino acid sequences of the 107 selected genes (including *M. polymorpha* diterpene synthase genes) using a commercially available MacVector program with default parameters (Rastogi, 2000) (Supplemental Data Set 3), and the phylogenetic tree was built using MEGA 6.0 (Tamura et al., 2013). The parameters used were Poisson model as substitution model with uniform rates and complete deletion for gaps or missing data. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Branches corresponding to partitions reproduced in less than 50% of the bootstrap replicates were collapsed via the program algorithm, and the tree was visualized with FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

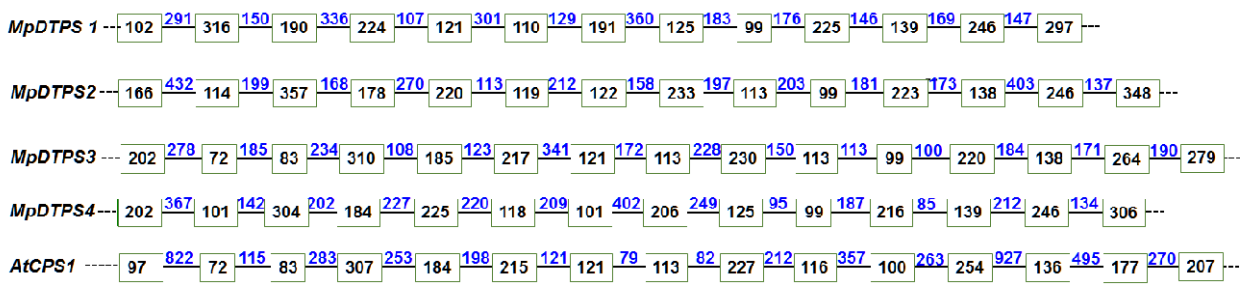
Altschul, S.F. Madden, T.L. Schäffer, A.A. Zhang, J. Zhang, Z. Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

Rastogi, P.A. (2000). MacVector. Integrated sequence analysis for the Macintosh. Methods Mol Biol. 132: 47-69.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

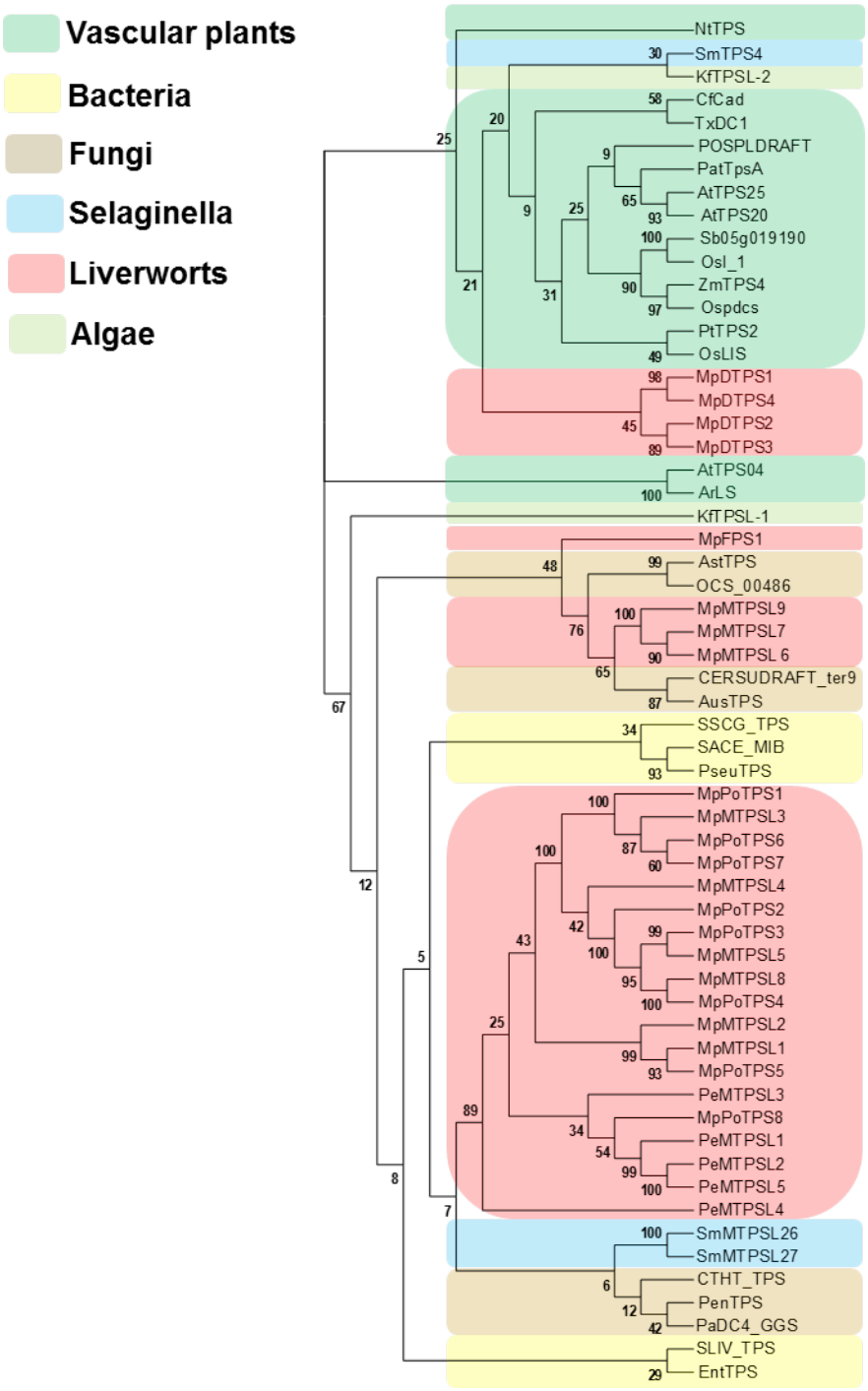
Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.



Supplemental Figure 18. Intron-exon organization of Mp*DTPS1* to *4* in comparison to a typical mono-functional diterpene synthase (CPS) found in *Arabidopsis* (AT4g02780) (Sun and Kamiya, 1994).

The data is based on in silico analysis of *M. polymorpha* genomic sequences available in the NCBI SRA database in comparison to the assembled transcriptome of *M. polymorpha.* The SRA data was downloaded and assembled in the CLC work bench ver. 4.7 as discussed earlier in case of our transcriptome assembly.

Sun, T.P. and Kamiya, Y. (1994) The Arabidopsis *GA1* locus encodes the cyclase *ent*-kaurene synthetase A of gibberellin biosynthesis. Plant Cell **6**: 1509–1518.



Supplemental Figure 19. Phylogenetic analysis of the terpene synthase-like and diterpene synthase-like proteins from *M. polymorpha* in relationship to bacterial, fungal and plant terpene synthase proteins (see Supplemental Data Sets 2 and 4).

Maximum likelihood phylogenetic tree analyses were performed with 59 amino acid sequences (see Supplementary Data Set 4) aligned across 285 positions. The terpene synthase sequences used include some sequences obtained from another liverwort (*Pellia endiviifolia*) transcriptome (Alaba et al., 2015), which have not been functionally

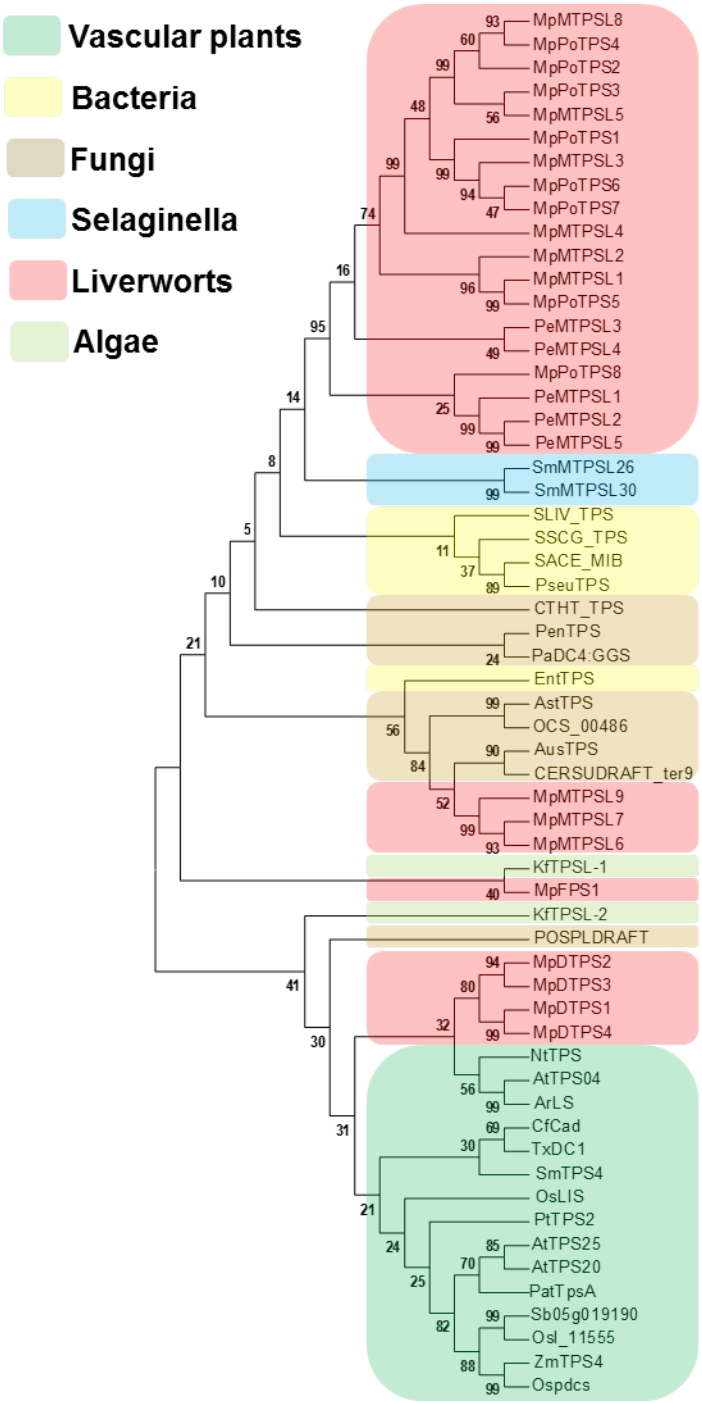
verified, plus validated terpene synthases from select bacteria and fungi, and examples across the evolutionary spectrum of plants based on the sequence similarity network (Figure 9). The maximum likelihood tree was generated using PhyML program interface in Seaview 4.0 (Gouy et al. 2010). Selection of best fit model was based on results provided by the Prottest server (Abascal et al 2005). The parameters used to generate the consensus phylogenetic tree were BIONJ as the starting tree (Gascuel 1997), using the LG substitution model with the four rate of substitution categories, estimated gamma distribution parameter and 1000 bootstrap repetitions. Bootstrap values are shown on each branches as percentage of replicates associated with end-point sequence clusters. The annotation of genes isolated from different organisms is shaded by different colors.

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105.

Alaba, S., et al. (2015). The liverwort *Pellia endiviifolia* shares microtranscriptomic traits that are common to green algae and land plants. New Phytol. 206: 352–367.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14: 685–695.

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27: 221–224.



Supplemental Figure 20. Evolutionary relationships of *Marchantia* terpenes synthase proteins to other terpene synthase proteins presented in Supplemental Data Sets 2 and 4.

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 39.98543631 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (5000 replicates) are shown next to the branches (Felsenstein 1985). The

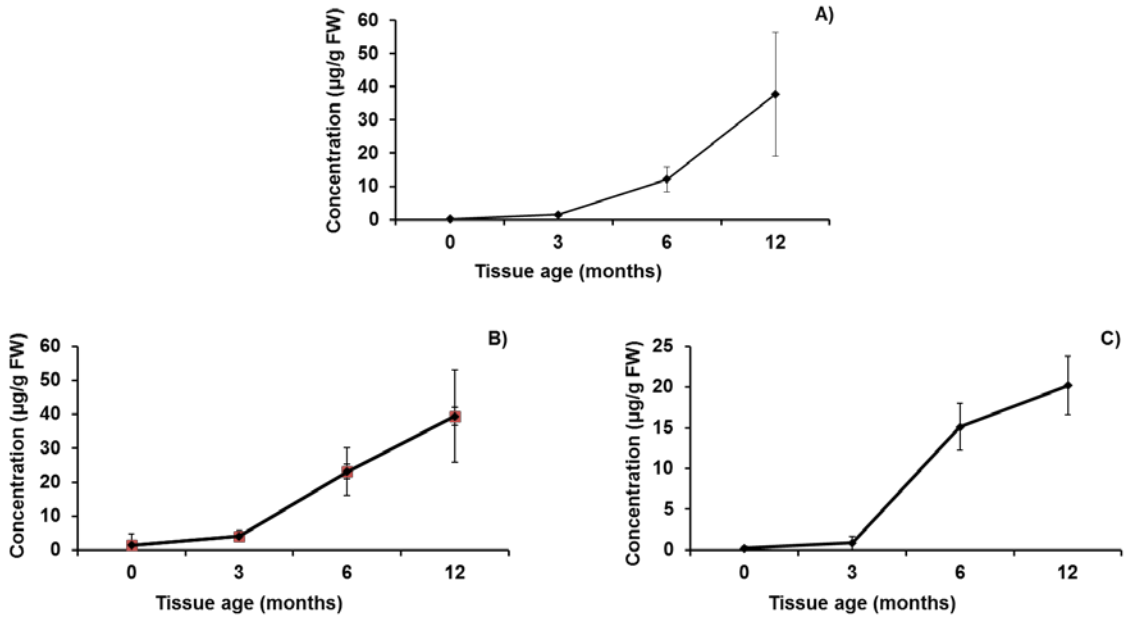
evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). The analysis involved 59 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 285 positions in the final Supplemental Data Set 4. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Kumar S., Stecher G., and Tamura K. (2015). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870–1874.

Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Zuckerkandl E. and Pauling L. (1965). Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.



Supplemental Figure 21. Developmental time course for the amounts of the major MpMTPSL4 products found in *M. polymorpha*.

Major MpMTPSL4 products measured during the development of *M. polymorpha*. (−)-α-Gurjunene is the major hydrocarbon produced by MpMTPSL4 (A). The time courses for the changes in the abundance of two suspected sesquiterpene alcohols found in *M. polymoprha* and produced by Mp*MTPSL*4 are also shown: 2.097169583 (relative retention time (RRT) to dodecane standard) (B) and 2.192242561 (RRT to dodecane standard) (C).

SUPPLEMENTAL TABLES

Supplemental Table 1. TBLASTN search of *Marchantia* transcriptome database (42,617 contigs) using archetypical mono- and sesqui-terpene synthases. Hits to *M. polymorpha* genes are noted in bold.

Query	Hits in *M. polymorpha* transcriptomic database	E-value	Score	Gaps (%)
Beta-cubebene synthase [*Magnolia grandiflora*] gb|ACC66281.1|	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig26612DeNovoAssembly (**Mp*DTPS3***)	8.35E-36	362	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig9617DeNovoAssembly (**Mp*DTPS4***)	8.47E-34	345	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig7218DeNovoAssembly (**Mp*DTPS1*)**	1.46E-30	319	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27100DeNovoAssembly (contig shows similarity to diterpene synthase from *Abies balsamea*)	7.92E-25	270	11
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	1.33E-22	254	12
AtTPS11 (Thujopsene synthase or Alpha-barbatene synthase [*Arabidopsis thaliana*] gb|Q4KSH9.2|	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27100DeNovoAssembly(contig shows similarity to diterpene synthase from *Abies balsamea*)	5.84E-20	231	9
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig7218DeNovoAssembly (**Mp*DTPS1***)	4.59E-12	168	13
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig9617DeNovoAssembly (**Mp*DTPS4***)	7.25E-11	157	12

	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig26612DeNovoAssembly (**Mp*DTPS3***)	4.38E-08	134	12
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	5.77E-04	99	14

Santalene and bergamotene synthase [*Solanum habrochaites*] gb|ACJ38409.1|	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig26612DeNovoAssembly (**Mp*DTPS3***)	7.48E-81	720	6
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	7.51E-68	616	7
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig7218DeNovoAssembly (**Mp*DTPS1***)	4.04E-62	578	9
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig9617DeNovoAssembly (**Mp*DTPS4***)	3.69E-55	521	9
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27745DeNovoAssembly (contig similar to copalyl diphosphate synthase [*Taiwania cryptomerioides*] gb|AFE61356.1|)	1.15E-23	258	13
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27100DeNovoAssembly (contig similar to levopimaradiene/abietadiene synthase [*Picea abies*] gb|AAS47691.1|)	5.00E-21	241	8
4S-limonene synthase [*Mentha spicata*] gb|AAC37366.1|	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig26612DeNovoAssembly (**Mp*DTPS3***)	1.58E-38	385	7
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig9617DeNovoAssembly (**Mp*DTPS4***)	5.64E-35	356	8

	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig7218DeNovoAssembly (**Mp*DTPS1***)	3.20E-33	342	9
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27100DeNovoAssembly (contig similar to levopimaradiene/abietadiene synthase [*Picea abies*] gb|AAS47691.1|)	4.20E-31	320	9
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	6.08E-23	258	10
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	6.08E-23	258	10
(+)-germacrene D synthase [*Solidago canadensis*] gb|AAR31144.1|	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig26612DeNovoAssembly (**Mp*DTPS3***)	4.11E-37	372	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig9617DeNovoAssembly (**Mp*DTPS4***)	1.10E-24	271	10
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27100DeNovoAssembly (contig similar to levopimaradiene/abietadiene synthase [*Picea abies*] gb|AAS47691.1|)	2.31E-23	259	10
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig7218DeNovoAssembly (**Mp*DTPS1***)	2.43E-20	236	10
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	7.38E-18	215	13
5EAS_TOBAC 5-epi-aristolochene	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig26612DeNovoAssembly (MpDTPS3)	3.62E-39	389	12

synthase sp|Q40577|	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig9617DeNovoAssembly (**Mp*DTPS4***)	6.33E-30	313	10
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27100DeNovoAssembly (contig similar to levopimaradiene/abietadiene synthase [*Picea abies*] gb|AAS47691.1|)	1.95E-24	267	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig7218DeNovoAssembly **(Mp*DTPS1***)	4.73E-23	258	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	5.14E-17	208	13
VTSS1_HYOMU Vetispiradiene synthase 1 (HVS1) sp|Q39978.2|	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig26612DeNovoAssembly	1.21E-42	418	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig9617DeNovoAssembly (**Mp*DTPS4***)	2.29E-33	341	7
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig7218DeNovoAssembly (**Mp*DTPS1***)	4.77E-28	299	8
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig27100DeNovoAssembly (contig similar to levopimaradiene/abietadiene synthase [*Picea abies*] gb|AAS47691.1|)	7.74E-28	294	7
	MHA_AA_NoIndex_L006_R1_001_pf(paired)contig6926DeNovoAssembly (**Mp*DTPS2***)	2.37E-17	211	12

Supplemental Table 2. Pfam domain search of the *M. polymorpha* assembled contigs with PF01397.

The assembled contigs for *M. polymorpha* were translated using the six frame translation module within Geneious (Geneious Pro v5.5 created by Biomatters; available from http://www.geneious.com). The six frame translated sequences were then screened for the Pfam PF01397 domain using HMMER3.0 (www.hmmer.janelia.org).

No.	Contig identified	PFAM Domain Used	HMM cut off E-value	Top hit BLATSX (NCBI)	E-value	Gene Cloned
1	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_26612_De_Novo_Assembly	PF01397	10.0	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length: 886	0	MpDTPS3
2	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_7218_De_Novo_Assembly	PF01397	10.0	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length: 886	5E-133	MpDTPS1
3	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_9617_De_Novo_Assembly	PF01397	10.0	ADB55710.1 (-)-ent-kaurene synthase [*Picea sitchensis*] Length: 757	8E-116	MpDTPS4
4	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_6926_De_Novo_Assembly	PF01397	10.0	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length: 886	0	MpDTPS2
5	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_27100_De_Novo_Assembly	PF01397	10.0	AEL99951.1 diterpene synthase TPS2, partial [*Abies balsamea*] Length=852	3.00E-92	Partial (5' missing)
6	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_27745_De_Novo_Assembly	PF01397	10.0	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length= 886	1.00E-74	Partial (5' and 3'missing)

Supplemental Table 3. Pfam domain search of the *M. polymorpha* assembled contigs with PF03936.

The assembled contigs for *M. polymorpha* were translated using the six frame translation module within Geneious (Geneious Pro v5.5 created by Biomatters; available from http://www.geneious.com). The six frame translated sequences were then screened for the Pfam PF03936 domain using HMMER3.0 (www.hmmer.janelia.org).

No.	Contig identified	Pfam domain used	HMM cut off E-value	Top hit BLATSX (NCBI)	E-value	Gene status
1	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_9617_De_Novo_Assembly	PF03936	10	ADB55710.1 (-)-ent-kaurene synthase [*Picea sitchensis*] Length: 757	8E-116	MpDTPS4
2	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_26612_De_Novo_Assembly	PF03936	10	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length: 886	0	MpDTPS3
3	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_7218_De_Novo_Assembly	PF03936	10	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length: 886	5E-133	MpDTPS1
4	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_27100_De_Novo_Assembly	PF03936	10	AEL99951.1 diterpene synthase TPS2, partial [*Abies balsamea*]= length 852	3.00E-92	Partial (5' missing)
5	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_6926_De_Novo_Assembly	PF03936	10	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length: 886	0	MpDTPS2
6	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_27182_De_Novo_Assembly	PF03936	10	WP_019432975.1 hypothetical protein [*Streptomyces* sp. AA0539] Length: 351	2.00E-19	MpMTPSL1
7	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_9911_De_Novo_Assembly	PF03936	10	WP_006971626.1 Terpene synthase, metal-binding protein [*Plesiocystis*	8.00E-11	MpMTPSL2

				pacifica] Length: 355		
8	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_9134_De_Novo_Assembly	PF03936	10	YP_006807783.1 terpene synthase metal-binding domain-containing protein [*Nocardia brasiliensis* ATCC 700358] Length: 755	2.00E-16	MpMTPSL4
9	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_29158_De_Novo_Assembly	PF03936	10	gb|ACU62166.1| Terpene synthase metal-binding domain protein [*Chitinophaga pinensis*DSM 2588]Length=321	1.00E-10	MpMTPSL5
10	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_14438_De_Novo_Assembly	PF03936	10	gb|EGO28715.1| putative terpene cyclase [*Serpula lacrymans* var. lacrymans S7.9]Length=342	5.00E-06	MpMTPSL8
11	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_27591_De_Novo_Assembly	PF03936	10	Geosmin synthase [*Streptomyces* sp. LaPpAH-95] WP_018102933.1Length: 737	7.00E-13	MpMTPSL3

Supplemental Table 4. Pfam domain search of the *M. polymorpha* assembled contigs with PF06330.

The assembled contigs for *M. polymorpha* were translated using the six frame translation module within Geneious (Geneious Pro v5.5 created by Biomatters; available from http://www.geneious.com). The six frame translated sequences were then screened for the Pfam PF06330 domain using HMMER3.0 (www.hmmer.janelia.org).

No.	Contig identified	Pfam domain used	HMM cut off E-value	Top hit BLATSX (NCBI)	E-value	*Putative gene name assigned*
1	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_13913_De_Novo_Assembly	PF06330	10	trichodiene synthase [*Fusarium sambucinum*] ACZ56400.1 Length: 247	7.00E-15	Mp*MTPSL7*
2	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_17864_De_Novo_Assembly	PF06330	10	XP_002474149.1 predicted protein [*Postia placenta* Mad-698-R] Length: 305	XP_002474149.1	Mp*MTPSL6*
3	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_9134_De_Novo_Assembly	PF06330	10	YP_006807783.1 terpene synthase metal-binding domain-containing protein [*Nocardia brasiliensis* ATCC 700358] Length: 755	2.00E-16	Mp*MTPSL4*
4	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_27182_De_Novo_Assembly	PF06330	10	WP_019432975.1 hypothetical protein [*Streptomyces* sp. AA0539] Length: 351	2.00E-19	Mp*MTPSL1*
5	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_27591_De_Novo_Assembly	PF06330	10	Geosmin synthase [*Streptomyces* sp. LaPpAH-95] WP_018102933.1Length: 737	7.00E-13	Mp*MTPSL3*
6	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_9911_De_Novo_Assembly	PF06330	10	WP_006971626.1 Terpene synthase, metal-binding protein [*Plesiocystis pacifica*] Length: 355	8.00E-11	Mp*MTPSL2*

7	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_29158_De_Novo_Assembly	PF06330	10	gb|ACU62166.1| Terpene synthase metal-binding domain protein [*Chitinophaga pinensis*DSM 2588]Length=321	1.00E-10	Mp*MTPSL5*
8	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_14438_De_Novo_Assembly	PF06330	10	gb|EGO28715.1| putative terpene cyclase [*Serpula lacrymans* var. lacrymans S7.9]Length=342	5.00E-06	Mp*MTPSL8*
9	MHA_AA_NoIndex_L006_R1_001_pf_(paired)_contig_6926_De_Novo_Assembly	PF06330	10	BAJ39816.1 ent-kaurene synthase [*Jungermannia subulata*] Length: 886	0	Mp*DTPS2*

Supplemental Table 5. Summary of Pfam domain searches of the *M. polymorpha* transcriptome (45,309 contigs, assembled from the NCBI SRA database SRP029610 according to Sharma et al. 2013) for PF01397, PF03936, and PF06330 motifs.

Contig number	**Top hit BLATSX (NCBI)**	**E-value**	**Assigned gene name**
12477	WP_019432975.1 hypothetical protein [*Streptomyces* sp. AA0539]	2E-19	MpMTPSL1
9798	WP_006971626.1 Terpene synthase, metal-binding protein [*Plesiocystis pacifica*]	8E-11	MpMTPSL2
4252, 4254, 30824	Geosmin synthase [*Streptomyces* sp. LaPpAH-95] WP_018102933.1	7E-13	MpMTPSL3
2193, 18889	YP_006807783.1 terpene synthase metal-binding domain-containing protein [*Nocardia brasiliensis* ATCC 700358]	2E-16	MpMTPSL4
9297, 16673	gb|ACU62166.1| Terpene synthase metal-binding domain protein [*Chitinophaga pinensis* DSM 2588]	1E-10	MpMTPSL5
15918	XP_002474149.1 predicted protein [*Postia placenta* Mad-698-R]	7E-11	MpMTPSL6
8770, 16809	trichodiene synthase [*Fusarium sambucinum*] ACZ56400.1	7E-15	MpMTPSL7
9297	gb|EGO28715.1| putative terpene cyclase [*Serpula lacrymans var. lacrymans* S7.9]	0.000005	MpMTPSL8
18059	Select seq gb|ACU62166.1| Terpene synthase metal-binding domain protein [*Chitinophaga pinensis* DSM 2588] ACU62166.1	7E-14	Partial

18987	Trichodiene synthase [*Penicillium expansum*] KGO36747.1	3E-10	MpMTPSL9
15161	Similar to Trichodiene synthase; acc. no. Q6A1B7 [*Pyronema omphalodes* CBS 100304] CCX30236.1	0.00001	Partial

Supplemental Table 6. References sequences used for similarity and identity comparisons.

Symbol	Gene	Accession	References
Sm*MTPSL1*	hypothetical protein SELMODRAFT_402353 [*Selaginella moellendorffii*]	XP_002960898	Li et al. 2012
Sm*MTPSL17*	hypothetical protein SELMODRAFT_412756 [*Selaginella moellendorffii*]	XP_002971982	Li et al. 2012
Sm*MTPSL26*	hypothetical protein SELMODRAFT_414571 [*Selaginella moellendorffii*]	XP_002974409	Li et al. 2012
Sm*MTPSL22*	hypothetical protein SELMODRAFT_413294 [*Selaginella moellendorffii*]	XP_002972952	Li et al. 2012
Sm*TPS9*	kaurene synthase [*Selaginella moellendorffii*]	XP_002960350	Li et al. 2012
Sm*TPS10*	terpene synthase, partial [*Selaginella moellendorffii*]	AFR34003	Li et al. 2012
PhyTPS	ent-kaurene synthase [*Physcomitrella patens*]	BAF61135	Hayashi et al. 2006
TRI5_FUSSP	sesquiterpene cyclase gene from the trichothecene-producing fungus *Fusarium sporotrichioides*	AAN05035	Hohn and Beremand 1989
TASY_TAXBR	taxadiene synthase [*Taxus brevifolia*]	AAC49310.1	Wildung and Croteau, 1996
TEAS	5-epi-aristolochene synthase	Q40577	Facchini and Chappell 1992
PrAS	Aristolochene synthase	Q03471.1	Proctor and Hohn, 1993
ScGS	cyclase [*Streptomyces coelicolor* A3(2)]	NP_630182	Hsiao and Kirby, 2008
Mp*MTPSL1*	*Marchantia* Terpene synthase like gene	KU664188	This study
Mp*MTPSL2*	*Marchantia* Terpene synthase like gene	KU664189	This study
Mp*MTPSL3*	*Marchantia* Terpene synthase like gene	KU664190	This study
Mp*MTPSL4*	*Marchantia* Terpene synthase like gene	KU664191	This study
Mp*MTPSL5*	*Marchantia* Terpene synthase like gene	KU664192	This study

Mp*MTPSL6*	*Marchantia* Terpene synthase like gene	KU664193	This study
Mp*MTPSL7*	*Marchantia* Terpene synthase like gene	KU664194	This study
Mp*MTPSL8*	*Marchantia* Terpene synthase like gene	KU664195	This study
Mp*MTPSL9*	*Marchantia* Terpene synthase like gene	KU886240	This study
Mp*DTPS1*	*Marchantia* Diterpene synthase like gene	KU664196	This study
Mp*DTPS3*	*Marchantia* Diterpene synthase like gene	KU664197	This study
Mp*DTPS4*	*Marchantia* Diterpene synthase like gene	KU664198	This study
4S-LS	4S-limonene synthase [*Mentha spicata*]	AAC37366	Colby et al 1993

Li, G. et al. (2012). Nonseed plant Selaginella moellendorffii has both seed plant and microbial types of terpene synthases. Proc. Natl. Acad. Sci. 109: 20774–20774.

Hayashi, K.I., Kawaide, H., Notomi, M., Sakigi, Y., Matsuo, A., and Nozaki, H. (2006). Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*. FEBS Lett. 580: 6175–6181.

Hohn, T.M. and Beremand, P.D. (1989). Isolation and nucleotide sequence of a sesquiterpene cyclase gene from the trichothecene-producing fungus Fusarium sporotrichioides. Gene 79: 131–138.

Wildung, M.R. and Croteau, R. (1996). A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. J. Biol. Chem. 271: 9201–9204.

Facchini, P.J. and Chappell, J. (1992). Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. Proc. Natl. Acad. Sci. U. S. A. 89: 11088–11092.

Proctor, R.H. and Hohns, T.M. (1993). Aristolochene Synthase isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (Aril) from *Penicillium roqueforti*. J. Biol. Chem. 268: 4543–4546.

Hsiao, N.-H. and Kirby, R. (2008). Comparative genomics of *Streptomyces avermitilis*, *Streptomyces cattleya*, *Streptomyces maritimus* and *Kitasatospora aureofaciens* using a *Streptomyces coelicolor* microarray system. Antonie Van Leeuwenhoek 93: 1–25.

Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993). 4S-limonene synthase from the oil glands of spearmint (*Mentha spicata*). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. J. Biol. Chem. 268: 23016–23024.

Supplemental Table 7. Kinetic constants for the terpene synthase-like enzymes from *M. polymorpha.*

Kinetic analysis was performed using purified His-tagged MpMTPSL proteins with preferred substrates from Table 1 (ND-No activity detected).

Enzyme	Preferred substrate	K_M (µM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
MpMTPSL1	ND	-	-	-
MpMTPSL2	NPP	36.15 ± 4.56	9.9×10^{-2} ± 0.004	2.76
MpMTPSL3	FPP	4.71 ± 0.70	5.2×10^{-2} ± 0.001	11.04
MpMTPSL4	FPP	17.98 ± 2.24	5.8×10^{-2} ± 0.001	3.21
MpMTPSL5	FPP	8.93 ± 1.29	14.5×10^{-2} ± 0.004	16.22
MpMTPSL6	NPP	1.42 ± 0.60	0.4×10^{-2} ± 0.001	2.98
MpMTPSL7	FPP	23.04 ± 1.44	1.65×10^{-2} ± 0.001	0.72
MpMTPSL8	ND	-	-	-
MpMTPSL9	FPP	101.5 ± 26.30	7.13×10^{-1} ± 0.107	70.2

Supplemental Table 8. ^{1}H-NMR for compound **7** (from Figure 6) in $CDCl_3$.

The structure of peak **7** was determined based on ^{1}H, ^{13}C, ^{1}H-^{1}H-gCOSY, and ^{1}H-^{13}C-HSQC experiments and on comparison with ^{1}H and ^{13}C data of other known aromadendrene alcohols. The ^{1}H-NMR demonstrated 2 singlet methyl groups at δ_H 1.25 and δ_H 1.04 and two methyl doublets at δ_H 0.86 and δ_H 0.91 (d, *J*= 6.5 Hz.) corresponding to H_3-14 and H_3-15. The ^{13}C-NMR revealed an absence of olefinic carbons in the δ_C 130-150 ppm range, and revealed the presence of a quarternary alcohol at δ_C 83.4. Because (+)-ledol, (+)-globulol, and (-)-viridiflorol all exhibit a C-10 secondary alcohol at δ_C 75.0 ppm, the downfield shift of the δ_C 83.4 alcohol indicates that it is likely at the C-1 or C-5 bridgehead carbon (Figure 7). The ^{1}H-^{1}H-gCOSY revealed all of the expected couplings for a gurjunene skeleton, except that the H-6 cyclopropane ring proton was present as a doublet in *cis* configuration with H-7 (H-6= d, *J*=10 Hz, H-7= td, *J*=10, 6, 1). In the other gurjunene alcohols, H-6 appears as a triplet coupling with H-5 and H-7 (Kaplan et al., 2000). This suggests that the position of the quarternary alcohol to be at the C-5 bridgehead carbon

Position	**δ_H**
H-1	1.66-1.78 (1H, m, complex)
H-2	1.80-1.95 (1H, *m*, complex)
H-3_{eq}	1.80-1.95 (1H, *m*, complex)
H-3_{ax}	1.08-1.18 (1H, *m*, complex)
H-4	1.66-1.78 (1H, *m*, complex)
H-5	**---**
H-6	0.18 (1H, *d*, *J*= 10)
H-7	0.67 (1H, *ddd*, *J*= 10, 9, 6)
H-8_{eq}	1.80-1.95 (1H, *m*, complex)
H-8_{ax}	1.08-1.24 (1H, *m*, complex)
H-9_{eq}	1.30-1.48 (1H, *m*, complex)
H-9_{ax}	1.30-1.48 (1H, *m*, complex)
H-10	1.80-1.95 (1H, *m*, complex)
H-11	---
H-12	1.04 (3H, *s*)
H-13	1.25 (3H, *s*)
H-14	0.92 (1H, *d*, *J*= 6)
H-15	0.86 (1H, *d*, *J*= 6).

Supplemental Table 9. ^{13}C-NMR data for compound 7 (from Figure 6), (+)-globulol, and (+)-ledol in $CDCl_3$.

δ_C position	***7***[ǂ]	*(+)-globulol*[a]	*ledol*[b]
C-1	57.3	57.0	53.8
C-2	25.0	26.1	24.6
C-3	27.7	34.6	30.8
C-4	47.3	36.3	38.4
C-5	83.4	39.7	40.8
C-6	25.2	28.3	23.4
C-7	29.3	26.7	25.0
C-8	21.1	20.2	20.3
C-9	32.0	44.6	39.2
C-10	34.5	75.3	74.6
C-11	19.0	19.4	19.2
C-12	17.2	15.8	15.4
C-13	30.7	28.7	28.7
C-14	13.6	20.1	30.5
C-15	22.5	16.0	16.0

a) Toyota, M., Tanaka, M., and Asakawa, Y. (1999). A revision of the 13C NMR spectral assignment of globulol. Spectroscopy 14: 61–66.

b) Kaplan, M. a, Pugialli, H.R.L., Lopes, D., Gottlieb, H.E., Auxiliadora, M., and Kaplan, C. (2000). The stereochemistry of ledol from Renealmia chrysotrycha: an NMR study. Phytochemistry 55: 749–53.

ǂ recorded on a Varian JNMR 400 MHz spectrometer at 100 MHz.

Supplemental Table 10. ^{1}H and ^{13}C NMR data for atiseranol.

NMR experiments were conducted on a Bruker Avance 700 spectrometer equipped with a 5-mm HCN cryogenic probe. Structural analysis was performed using one-dimensional ^{1}H, and two-dimensional DQF-COSY, HSQC, HMQC, HMBC, and NOESY experiment spectra acquired at 700 MHz, and one-dimensional ^{13}C spectrum (174 MHz) using standard experiments from the Bruker TopSpin version 1.4 software. An analysis of the DQF-COSY and HSQC spectra led to the unambiguous assignment of the protons and corresponding carbon signals. Correlations from the HMBC spectra were used to build the planar structure, and the stereochemistry was determined by the NOESY experiment. In the NOESY spectrum, the correlations of H_3-17/H-13a, H_3-17/H-14b, H_3-20/H-13b and H_3-20/H-14a indicated that 16-OH possessed a β-configuration. Chemical shifts were referenced using known chloroform (^{13}C 77.23, ^{1}H 7.24 ppm) signals offset from TMS, and compared to those previously reported.

Position	MPld40	
	δ_H	δ_C
1 a	1.45 (1H, m)	39.9
b	0.73 (1H, m)	
2 a	1.48 (1H, m)	18.8
b	1.28 (1H, m)	
3 a	1.29 (1H, m)	42.8
b	1.05 (1H, m)	
4		33.7
5	0.72 (1H, m)	57.0
6 a	1.36 (1H, m)	19.3
b	1.20 (1H, m)	
7	1.26 (1H, m)	40.2
	1.03 (1H, m)	
8		34.4
9	1.12 (1H, m)	51.9
10		38.3
11	1.87 (1H, m)	23.8
	1.07 (1H, m)	
12	1.44 (1H, m)	38.5
13 a	1.56 (1H, m)	24.7
b	1.38 (1H, m)	
14 a	1.73 (1H, m)	27.9
b	0.70 (1H, m)	
15	1.26 (1H, m)	58.3
	1.11 (1H, m)	
16		72.8
17	1.19 (3H, s)	31.0
18	0.76 (3H, s)	34.0
19	0.73 (3H, s)	22.3
20	0.85 (3H, s)	14.5

Supplemental Table 11. Primers used for cloning full-length Mp*MTPSL* genes.

Gene	Primer	Sequences (5' → 3')
Mp***MTPSL1***	MpMTPSL-1_F1	ATGTCTGCGAGGGACAATGG
	MpMTPSL-1_R1	TTAGATGATTCTGTGGTCTT
Mp***MTPSL2***	MpMTPSL-2_F1	ATGGCGGCAAAAGCTCACAGGAGTTT
	MpMTPSL-2_R1	TTATCTCAAGTAGTGGTCCT
Mp***MTPSL2***	SK299	ATGATCGGACTGGTGAGCAGTATTTATG
	MpMTPSL-2_R1	TTATCTCAAGTAGTGGTCCT
Mp***MTPSL3***	MpMTPSL-3F	ATGGCTTCACAACACTCGGCATCTACC
	MpMTPSL-3R	CTACATTTTCAGATTTTCTTCTATC
Mp***MTPSL4***	MpMTPSL-4F	ATGGCACCAACTTTAGACTCGG
	MpMTPSL-4R	CTAGACACTGTTGATGTGTTTTAC
Mp***MTPSL5***	SK202F	ATGGCCCCAAGTTTAGACTCGGATTCTAC
	K199R	GTAGAACAAATCATCAAGCGTCAGTAA
Mp***MTPSL6***	MpMTPSL-6_F1	ATGAAGCCCATCATGGTGAGCTCTG
	MpMTPSL-6_R1	TTAAACGAAATCACCTTGGAACCAATCGTC
Mp***MTPSL7***	MpMTPSL-7_F1	ATGTCGAGCATGGCGAGCTGTG
	MpMTPSL-7_R1	TTAAATGAGATTACCTTGGAACCAGCCG
Mp***MTPSL8***	SK261F	ATGGGCCCAAGTTTAAACTCG
	SK262R	CTACCTTTTTAGATTTCCTGTGTCTG
Mp***MTPSL9***	SK417F	CATGCCATGGATGACGAAGACGCTTCCGGCT
	SK418R	CCCAAGCTTCTACACACAGTCACCCGCGAACC

Supplemental Table 12. Primers used for cloning full length Mp*DTPS* genes.

Gene	Primer	Sequences (5' → 3')
Mp***DTPS1***	SK005F	GGGAATTCCATATGATGCTAGCCGTCGATGAACCGAC
	SK006R	CCCAAGCTTCTACTCGCTGGGGAAGCGTTTG
Mp***DTPS2***	SK007F	GGGAATTCCATATGATGGCGAGCTCGACTGCC
	SK008R	GGAATTCTCAAGAGAGCACGGGTTCG
Mp***DTPS3***	SK009F	GGAATTCCATATGATGGCATTCTCGTTAGCAGG
	SK010R	ATAAGAATGCGGCCGCTCAGGCCACAGGCTCGAAGAGTAG
Mp***DTPS4***	SK341F	AGCCATGATTTCGAATGATGAGG
	SK343R	CTAGGCCTGTTCACTTTCGATGG

Supplemental Table 13. Primers used for DNA sequencing.

Gene	Primer	Sequence (5' → 3')
Mp*MTPSL1*	SK187R	GGAAACCGAAAATGTCATTATGCCAACC
	SK161F	GTGACAGATCTCTTCGTCAAAGCTC
	SK162F	ACTTGTGGCTCGAGTACTGTGAAAG
	SK163F	ATCTCGGCTCTTACGTCTCGAATAC
	SK164F	CAACAAGGTTCTCATGTGGTTCTTC
	MpMTPSL-1_sF2	GATGTTGATAAATCACGAGG
	MpMTPSL-1_sF3	CAGAACTACTTACTGGGATGT
	MpMTPSL-1_sR2	TTGTTCCATGCTTGGGCGCC
	MpMTPSL-1_sR3	GCCATTGTCGAGCTTTGACG
Mp*MTPSL2*	SK167F	ATCAATGAGTTCGAGACAAGAGTGG
	SK168F	TCACAGGAGTTTCTCCAAGCTTATG
	SK169F	TTAGATTTGCAGCCTTTAGCATCAG
	SK170F	TCTGATAGCCGAGTTCAATGAAAAG
	MpMTPSL-2_sF2	GAGTTCAATGAAAAGGCCCA
	MpMTPSL-2_sR2	ACCACAAGATAACCTTCTGC
	MpMTPSL-2_sR3	TGATACTCTGCGGATAATCT
Mp*MTPSL3*	SK228F	TCCTGAGCTTATGGGCTATGTGTTC
	SK229F	GTGGAATGACCCTGAGAACAAGAG
	SK230R	CTCTTCAGCTGCCTACGAAACTGTC
	SK282F	ACAAGGACAGTTTCGTAGGC
	SK283R	CAATGTCTGTGTACTCGGGAG
	SK194R	TTGGGGTAACGGTGCTGGAAATGTG

	SK195R	GTAGATGCCGAGTGTTGTGAAGCCA
Mp*MTPSL4*	SK196R	TGCCCATAGAACTCAACCAGTCAAAAG
	SK197R	GCATTCGGTAGTCTTGGGGTTCAGTC
	SK221R	CTGGGCGATTCTAGTCGTATACTC
	SK222F	CTCCGATTTCACATATTCCCAACAC
	SK223F	GAGTATACGACTAGAATCGCCCAG
	SK240R	CGAAGGGTGACGTAATCAGCCACATTC
	SK286F	ATTTCAGAGCCCCATACATCC
	SK287R	ATGCCCATAGAACTCAACCAG
Mp*MTPSL5*	SK284F	TTGGTTCTCTTGTGGACTGAC
	SK285R	AGCCTCAATCTCCATCAACG
	SK185R	GAAAACAAGAAACGGCATCACACCTCC
	SK186R	CTTGTTCCGAGAGACAGCCCGTTATGT
	SK198F	CCTGTTGAGTGTATATCTAGACAGGCTG
	SK199R	TTACTGACGCTTGATGATTTGTTCTAC
	SK202F	ATGGCCCCAAGTTTAGACTCGGATTCTAC
Mp*MTPSL6*	SK121F	CATTCTTTTCATCGTCACTGCG
	SK122R	TCGTTGATGTGGCTTGAGG
	SK295F	GAACATCACCGCTTTGGAAG
Mp*MTPSL7*	SK038F	ATGTCGAGCATGGCGAGCTGTGGTGC
	SK039R	CTAAATGAGATTACCTTGGAACCAGC
	SK119F	GTCAGTTAGAGCACGAGTACAG
	SK120R	GAAATAGAGGAAGGTGAGGCG

	SK180R	CGAGTCGGATGTAGTTTCTTGTAC
	SK076F	GATTTAGAAATGCCTGCATGGACTATC
Mp*MTPSL8*	SK200F	CCTGTTGAGTGTATATCTAGACAAGCTG
	SK201R	CTACCTTTTTAGATTTCCTGTGTCTG
	SK211F	ATGGCGGCCAAATTCTCTAAGCTTATTG
	SK212R	CTACCTTTTTAGATTTCCTGTGTCTG
	SK251F	CCTGACAACGAGCGATTACTGGAA
	SK252F	CTGATTGCATCAGAGTTCGACGAC
	SK253F	TTCATGAATACGACTGCCAATCTTAC
	SK254F	AGACTGGATTCCTGGAACGCACGAG
	SK261F	ATGGGCCCAAGTTTAAACTCG
	SK262R	CTACCTTTTTAGATTTCCTGTGTCTG
	SK187R	GGAAACCGAAAATGTCATTATGCCAACC
	SK188R	GATGTCAGTACAAGCGGCCAGCATATTC
	SK288R	GGTCGCTCTCCGACTGCCTTAG
	SK289R	CTCGTTGTCAGGGTCATTCCAC
Mp*MTPSL9*	SK419F	GATGCCTGATGGTGGAGC
	SK420F	TACAACAAGATATATCCACTGATCCC
	SK421R	CGTCCTCGAACTCCTTGTAA
	SK422R	CACGTGAGTGATAATGTTTCTCG

Supplemental Table 14. Primers used for cloning Mp*MTPSL* genes into bacterial (*E. coli*) protein expression vectors and their restriction site.

Gene	Primer	Sequence (5' → 3')	Restriction Site	Vector
Mp*MTPSL1*	SK159F	CATATGATGTCTGCGAGGGACAATG GGGCTATC	*Nde*I	pET28a
	SK160R	AAGCTTTTTGATGATTCTGTGGTCTT CGTTGTATC	*Hind*III	
Mp*MTPSL2*	SK227F	TATACCATGGGCATGGCGGCAAAAG CTCACAG	*Nco*I	pET28b
	SK292R	CGCGGATCCCTGTCTCAAGTAGTGG TCCTCTTTTTGGTATC	*BamH*I	
Mp*MTPSL3*	SK171F	CATATGATGGCTTCACAACACTCGG CATCTACCG	*Nde*I	pET28a
	SK172R	AAGCTTCTTCATTTTCAGATTTTCTT CTATCTCAG	*Hind*III	
Mp*MTPSL4*	SK173F	GAGCTCATGGCACCAACTTTAGACT CGGATTCTAC	*Sac*I	pET28a
	SK174R	AAGCTTCTTGACACTGTTGATGTGTT TTAC	*Hind*III	
Mp*MTPSL5*	SK207F	CGGGATCCATGGCCCCAAGTTTAGA CTCG	*BamH*I	pET28a
	SK208R	ATAAGAATGCGGCCGCTTACTGACG CTTGATGATTTG	*Not*I	
Mp*MTPSL6*	SK273F	TCTAGA GAAGGAGAATGAAGCCCATC	*Xba*I	pET28b
	SK275R	CCCAAGCTTGTAAACGAAATCACCT TG	*Hind*III	
Mp*MTPSL7*	SK077F	GGAATTCCATATGTCGAGCATGGCG AGCTGTGGTGC	*Nde*I	pET28a
	SK078R	GGAATTCCTAAATGAGATTACCTTG GAACCAGC	*Eco*RI	
Mp*MTPSL8*	SK259F	CGCGGATCCATGGGCCCAAGTTTAA ACTC	*Bam*HI	pET28a
	SK260R	ACGCGTCGACCTACCTTTTTAGATTT CC	*Sal*I	
Mp*MTPSL9*	SK423F	GGGAATTCCATATGACGAAGACGCT TCCGGCT	*Nde*I	pET28a
	SK424R	CCGGAATTCCTACACACAGTCACCC GCGAACC	*Eco*RI	

Supplemental Table 15. Primers used for cloning Mp*MTPSL* genes into specific restriction sites within yeast expression vectors.

Gene	Primer	Sequences (5' → 3')	Restriction Site	Vector
Mp***MTPSL1***	mmTPS1NotIF	GGGGCGGCCGCAAAACAATGTCTGCGAGGGACAATGGGGCT	*Not*I	X-HIS
	mmTPS1SpeIR	GACTAGTTTAGATGATTCTGTGGTTTCGTTG	*Spe*I	
Mp***MTPSL3***	mmTPS3NotIF	GGGGCGGCCGCAAAACAATGGCTTCACAACACTCGGCATC	*Not*I	X-HIS
	mmTPS3SpeIR	GACTAGTCTACATTTTCAGATTTTCTTCTATC	*Spe*I	
Mp***MTPSL4***	mmTPS4SpeIF	GACTAGTAAAACAATGGCACCAACTTTAGACTCGG	*Spe*I	X-HIS
	mmTPS4SacIR	GGGGAGCTCCTAGACACTGTTGATGTGTTTTAC	*Sac*I	
Mp***MTPSL5***	mmTPS5NotIF	GGGGCGGCCGCAAAACAATGCAGGAGATTCTTTACTTCC	*Not*I	X-HIS
	mmTPS5SpeIR	GACTAGTTTAAACGAAATCACCTTGGAACC	*Spe*I	
Mp***MTPSL6***	mmTPS6NotIF	GGGGCGGCCGCAAAACAATGAAGCCCATCATGGTGAGCTC	*Not*I	X-HIS
	mmTPS6SpeIR	GACTAGTTTAAACGAAATCACCTTGGAACC	*Spe*I	
Mp***MTPSL7***	mmTPS7SacIF	GGGGAGCTCAAAACAATGGCGAGCTGTGGTGCCGGGAAC	*Sac*I	X-HIS
	mmTPS7PacIR	CCTTAATTAATTACGCCAGCTATTTAGGTGACAC	*Pac*I	
Mp***MTPSL9***	SK445F	GGGGCGGCCGCAAAACAATGACGAAGACGCTTCCG	*Not*I	X-HIS
	SK446R	GACTAGTTCTACACACAGTCACCCGCGAACC	*Spe*I	

Supplemental Table 16. Primers used for qualitative RT-PCR.

Gene	Primer	Sequence (5' → 3')
Mp*MTPSL1*	SK309F	CCTTTGTTCAATTGTGTCTCCG
	SK310R	ACAAGTCCCTGAAAGCTGATG
Mp*MTPSL2*	SK302F	CTTGAAATCTGGAGCGACATG
	SK303R	AGAGCAGAACCACAAGATAACC
Mp*MTPSL3*	SK282F	ACAAGGACAGTTTCGTAGGC
	SK283R	CAATGTCTGTGTACTCGGGAG
Mp*MTPSL4*	SK286F	ATTTCAGAGCCCCATACATCC
	SK287R	ATGCCCATAGAACTCAACCAG
Mp*MTPSL5*	SK284F	TTGGTTCTCTTGTGGACTGAC
	SK285R	AGCCTCAATCTCCATCAACG
Mp*MTPSL6*	SK121F	CATTCTTTTCATCGTCACTGCG
	SK122R	TCGTTGATGTGGCTTGAGG
Mp*MTPSL7*	SK119F	GTCAGTTAGAGCACGAGTACAG
	SK120R	GAAATAGAGGAAGGTGAGGCG
Mp*MTPSL8*	SK311F	GAAATAGAGGAAGGTGAGGCG
	SK312R	TGCCCAAAGAAACGATCCAG
Mp*MTPSL9*	SK419F	GATGCCTGATGGTGGAGC

	SK421R	CGTCCTCGAACTCCTTGTAA
Actin	SK101F	ATGAAGATTCTGACCGAGCG
	SK102R	GAAGTCCAGGGCAATGTAGG
GAPDH	SK280F	GTCATTCAGAGTACCCACCG
	SK281R	TCCCTTCATTTCGCCTTCAG
Mp*DTPS1*	SK314F	CTACTTCTCAACTCTGTCGTGC
	SK313R	GGACACATACTTCAGCTCATCG
Mp*DTPS2*	SK316F	TGGAGGTTGCAGGAATTACAG
	SK315R	CTTGGTCAGTCTCTTCGTGTG
Mp*DTPS3*	SK318F	CTCGCTCTGCCCTACAAAG
	SK317R	TCCCAGTCCAAAATCTCGTG
Mp*DTPS4*	SK320F	CAGCAAAATACAAGGGCGTG
	SK321R	AAACTAACCCAGCTCGTGTC