Supplementary Information

Origin and Early Evolution of the Plant Terpene Synthase Family

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

\textsuperscript{a}Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN 37996;

\textsuperscript{b}Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, IA 50011;

\textsuperscript{c}Department of Biochemistry, Max-Planck-Institute of Chemical Ecology, Germany

\textsuperscript{d}Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996;

\textsuperscript{e}Key Laboratory of Tea Quality and Safety Control, Ministry of Agriculture and Rural Affairs, Tea Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310008, China;

\textsuperscript{f}BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China;

\textsuperscript{g}Department of Biological Sciences, University of Alberta, Edmonton AB, T6G 2E9, Canada;

\textsuperscript{h}Department of Medicine, University of Alberta, Edmonton AB, T6G 2E1, Canada

\textsuperscript{§}Q.J. and R.B. contributed equally to this work.

*To whom correspondence should be addressed. E-mail: qjia2@vols.utk.edu; rjpeters@iastate.edu or fengc@utk.edu
SI Material and Methods

Genomic databases of green plants analyzed in this study

To mine TPS genes from a wide range of green plants, especially nonseed land plants and green algae, we compiled one genome dataset and one transcriptome dataset. The genome dataset contains the genome sequences of 31 species of green algae that include six species of charophytes and 25 species of chlorophytes (Table S1) and 14 species of land plants (Tables S2). For the transcriptome dataset, the bulk of the data came from the one thousand plant transcriptome (OneKP) project (https://sites.google.com/a/ualberta.ca/onekp/), which covers a wide taxonomic range of green plants. The OneKP transcriptome data are of generally high quality as assessed using various criteria, including BUSCO (1). It should be mentioned that OneKP data has been employed in our previous study of microbial-terpene synthase-like (MTPSL) genes, which led to the publication of PNAS article (2). Additionally, the transcriptome data for 69 species of ferns from a recent work (3) were added to the transcriptome dataset, making the total number of species in the transcriptome data 1178. For each of the sequenced genomes, the annotated proteomes were downloaded from appropriate sources (Table S2). The proteomes for the OneKP dataset were created in our previous study (2). The proteomes for the 69 species of ferns were created from their respective DNA sequences using TransDecoder v5.5.0 (4).

Enzyme assays for terpene synthases

The expression construct for each TPS in pEXP-5-CT/TOPO was transformed into the C41 OverExpress strain of *E. coli* (Lucigen), with the putative class II and bifunctional genes cotransformed with a plasmid encoding a GGPP synthase (pGG), and the putative class I diterpene synthases with the pGG plasmid also containing a CPS that produces normal, *ent*- or *syn*- CPP (pGGnC, pGGeC and pGGsC, respectively). Transformed bacteria were then grown in 50 ml TB cultures in 250 ml glass Erlenmeyer flasks, induced at OD_{600} ~0.6 with 1 mM IPTG, and incubated at 16 °C for 72 hours. The resulting terpenoid products were then extracted with 50 ml hexanes, which was then dried under a flow of nitrogen gas and resuspended in 100 μL of hexane for analysis via gas chromatography with mass spectral detection (GC-MS). The class II diterpene cyclases shown to
produce CPP were further coexpressed with the KS from *Arabidopsis thaliana* (AtKS), which is specific for *ent*-CPP (5), to investigate such stereospecific product outcome. For the bifunctional CPSKS, the first two aspartates of the DDxxD motif were mutated to alanine to knock out class I functionality and the resulting CPS then coexpressed with AtKS for the same purpose.

GC-MS analysis was carried out as previously described (2). Briefly, 1 µl of each resuspended sample was injected using an 8400 autosampler onto a Varian 3900 GC with a Saturn 2100T ion trap mass spectrometer set to electron ionization (at 70 eV) mode and run over a 30 meter Agilent HP-5MS column (Agilent, 19091S-433) with 1.2 mL/min helium flow rate. The injector port was run in splitless mode and pre-heated to 250 °C. The column oven was held at 50 °C for 3 minutes, then heated to 300 °C at 15 °C/min, with 300 °C held for 3 minutes once reached. The Saturn 2100T ion trap mass spectrometer recorded mass-to-charge (m/z) ratios from 90 to 650 after a delay of 13 minutes to allow the solvent front to pass. The results were analyzed using the Varian WorkStation software package, and figures produced using Igor Pro. In most cases, the products could be identified by comparison of retention time and mass spectra to that of known standards.

In two cases the product was not readily identified, one olefin and one dephosphorylated class II diterpene cyclase product (primary alcohol). Thus, 1L batches of the relevant culture were grown following the same protocol. These cultures were extracted with 1L hexanes, which was separated and then dried via rotary evaporation under vacuum, with the residue resuspended in 5 mL hexanes. This extract was first fractionated over 4 g-silica column on a Grace Reveletis flash chromatography system with UV detection and automated injector and fraction collector, at a flow rate of 15 mL/min. The silica column was pre-equilibrated with hexanes before sample injection and run with 100% hexanes for 4 minutes following injection. Subsequently, a gradient of 0-100% acetone over 15 mL with an additional 45 mL in 15 mL fractions collected at 100% acetone. The olefin product eluted in the hexane fraction while the alcohol product was found in the acetone fraction. These fractions were dried under N₂ and resuspended in 2 mL of an 80:20 water to methanol solution for purification via high-pressure liquid chromatography (HPLC). This was carried out with an Agilent 1200 series HPLC instrument equipped with a diode array UV detector and automated injector and fraction collector. The products were first further fractionated in 50 µL increments over a semi-preparative C-8 column (ZORBAX Eclipse XDB-C8, 25×0.94 cm) run at 4 mL/min, which had been preequilibrated with acetonitrile/water (1:1 for the olefin and 4:1 for the alcohol). The column was washed for two minutes
with the equilibration mixture, increased to 100% acetonitrile over eight minutes, and eluted for an additional 20 minutes at 100% acetonitrile, with collection of 0.5 mL fractions. The fractions containing the products were identified by GC-MS analysis, pooled, dried under N₂ and resuspended in 1 mL (80:20 water to methanol). Each product was then purified in 20 µL increments over an analytical C-18 Column (3µm 100 Angstrom, 250 x 4.6 mm) run at 0.5 mL/min with the same solvents and elution program used with the C-8 column. Fractions containing pure compounds again were identified by GC-MS analysis, pooled, dried under N₂, and then dissolved in 0.5 mL deuterated dimethyl sulfoxide (DMSO) for structural analysis by NMR. NMR spectra were acquired on a Bruker AVIII-800 spectrometer equipped with 5 mm cryogenic HCN probe, run with TopSpin 3.2 software, located in the Iowa State NMR facility. Analysis was carried out at 25 °C, chemical shifts calculated with reference to those known for deuterated DMSO (¹³C 39.52 ppm, ¹H 2.50 ppm), with acquisition of 1D ¹H, 1D ¹³C, 2d COSY, 2D HSQC, 2D HMBC, DEPT 135, 90, and 45, 2d MSQC, and 2D NOESY spectra. The olefin was determined to be isopimara-8,15-diene, while the alcohol was found to be cis-kolav-3,13-dien-15-ol, in part by matching to previously published data (6,7).
Table S1. Species with sequenced genomes of green algae analyzed in this study.

| Species                          | Reference/source |
|----------------------------------|------------------|
| **charophyta**                   |                  |
| Chara braunii                    | (8)              |
| Klebsormidium flaccidum          | (9)              |
| Chlorokybus atmophyticus         | (10)             |
| Mesostigma viride                | (10)             |
| Spirogloea muscicola             | (11)             |
| Mesotaenium endlicherianum       | (11)             |
| **chlorophyta**                  |                  |
| Asterochloris sp                 | (12)             |
| Chlorella protothecoides         | (13)             |
| Bathycoccus prasinus             | (14)             |
| Botryococcus braunii             | (15)             |
| Chlamydomonas eustigma           | (16)             |
| Chlamydomonas reinhardtii        | (17)             |
| Chlorella variabilis             | (18)             |
| Chromochloris zofingiensis       | (19)             |
| Coccomyxa subellipsoidea         | (20)             |
| Dunaliella salina                | (21)             |
| Gonium pectorale                 | (22)             |
| Helicosporidium sp               | (23)             |
| Micractinium conductrix          | (24)             |
| Micromonas commoda               | (25)             |
| Micromonas pusilla               | (25)             |
| Monoraphidium neglectum          | (26)             |
| Ostreococcus lucimarinus         | (27)             |
| Ostreococcus sp                  | (12)             |
| Ostreococcus tauri               | (28)             |
| Picoclorum soleocismus           | (29)             |
| Raphidocelis subcapitata         | (30)             |
| Symbiochloris reticulata         | (12)             |
| Tetrabaena socialis              | (31)             |
| Ulva mutabilis                   | (32)             |
| Volvox carteri                   | (33)             |
Table S2. Land plants with sequenced genomes analyzed for TPS genes in this study.

| Species                        | Source of genome |
|--------------------------------|-------------------|
| Amborella trichopoda           | (34)              |
| Arabidopsis thaliana           | (34)              |
| Oryza sativa                   | (34)              |
| Ginkgo biloba                  | (35)              |
| Pseudotsuga menziesii          | (35)              |
| Picea abies                    | (35)              |
| Pinus lambertiana              | (35)              |
| Selaginella moellendorffii     | (34)              |
| Azolla filimculoides           | (36)              |
| Salvinia cucullata             | (36)              |
| Sphagnum fallax                | (34)              |
| Anthoceros agrestis            | (37)              |
| Anthoceros punctatus           | (37)              |
| Marchantia polymorpha          | (34)              |
### Table S3. TPS genes from sequenced nonseed plants.

| Lineage     | Species                | Number of TPS | Reference  |
|-------------|------------------------|---------------|------------|
|             |                        | All           |            |
|             |                        | >= 350 aas<sup>a</sup> |            |
| Liverworts  | *Marchantia polymorpha*| 7             | (38)       |
| Hornworts   | *Anthoceros agrestis*  | 20            | This study |
|             | *Anthoceros punctatus* | 11            | This study |
| Mosses      | *Physcomitrella patens*| 1             | (39)       |
|             | *Sphagnum fallax*      | 5             | This study |
| Ferns       | *Salvinia cucullata*   | 3             | This study |
|             | *Azolla filiculoides*  | 1             | This study |
| Lycophytes  | *Selaginella moellendorffii* | 18           | (40)       |

<sup>a</sup>Protein length based on genome annotation. These genes were used for phylogenetic analysis presented in Figure 3.
Table S4. Functionally characterized TPSs included in phylogenetic analysis.

| Gene Name | TPS subfamily | Species               | Lineage    | Reference |
|-----------|---------------|-----------------------|------------|-----------|
| JsCPSKS   | TPS-c         | *Jungermannia subulata* | Liverworts | (41)      |
| HpDTC1    | TPS-c         | *Hypnum plumaeforme*  | Mosses     | (42)      |
| LjCPSKS   | TPS-c         | *Lygodium japonicum*  | Ferns      | (43)      |
| TaKSL5    | TPS-e/f       | *Triticum aestivum*   | Angiosperms| (44)      |
respectively) for phytohormone biosynthesis. For all motifs, "✓" indicates conservation, while red text indicates informative changes discussed in text. "II" and "III" depict class II, class III and class III motifs, respectively; "Nonfunctional" motifs are depicted in green.

| Gene name | Species | Class | Active motif | Nonactive motif | Products | Gene name | Species | Class | Active motif | Nonactive motif | Products |
|-----------|---------|-------|--------------|-----------------|----------|-----------|---------|-------|--------------|-----------------|----------|
| Asu_TPS1  | Fern    | II    | I             | I               | ✓        | Asu_TPS1  | Fern    | II    | I             | I               | ✓        |
| BRY_TMAJ_PTPS2 | Moss | II    | I             | I               | ✓        | BRY_TMAJ_PTPS2 | Moss | II    | I             | I               | ✓        |
| BRY_CMEQ_PTPS1 | Liverwort | II    | I             | I               | ✓        | BRY_CMEQ_PTPS1 | Liverwort | II    | I             | I               | ✓        |
| BRY_IGUH_PTPS1 | Liverwort | II    | I             | I               | ✓        | BRY_IGUH_PTPS1 | Liverwort | II    | I             | I               | ✓        |

Table S5. TPSs from nonseed plants selected for biochemical characterization.
| Gene name | Biochemical function | Species | Lineage | Reference |
|-----------|----------------------|---------|---------|-----------|
| MpDTPS1   | KS                   | Marchantia polymorpha | Liverwort | (45)      |
| MpDTPS3   | CPS                 | Marchantia polymorpha | Liverwort | (45)      |
| MpDTPS4   | KS                   | Marchantia polymorpha | Liverwort | (45)      |
| JsCPSKS   | CPS/KS               | Jungermannia subulata | Liverwort | (41)      |
| PpCPSKS   | CPS                 | Physcomitrium patens | Liverwort | (39)      |
| HpDTC1    | syn-Pimara-7,15-diene synthase | Hypnum plumbariae | Moss | (42)      |
| SmKS      | KS                   | Selaginella moellendorffii | Lycophyte | (46)      |
| SmMDS     | miltiradiene synthase | Selaginella moellendorffii | Lycophyte | (47)      |
| SmCPSKSL1 | γ-7,13E-diene-15-ol synthase | Selaginella moellendorffii | Lycophyte | (48)      |
| SmDTC3    | ent-16α-hydroxykaurene synthase | Selaginella moellendorffii | Lycophyte | (46)      |
| SmTPS10   | CPS                 | Selaginella moellendorffii | Lycophyte | (46)      |
| SmTPS9    | CPS                 | Selaginella moellendorffii | Lycophyte | (46)      |
| SmPCPS2   | CPS                 | Selaginella moellendorffii | Lycophyte | (46)      |
| LjCPSKS   | CPS/KS               | Lygodium japonicum | Fern | (43)      |

Table S6. TPSs from nonseed plants that were previously characterized.
Figure S1: Unrooted phylogenetic tree for plant TPSs with bootstrap values. The circular layout of the unrooted tree shown in Figure 3 and 4. Bootstrap values > 50% are shown. Known and putative TPSs with αβ didomain architecture are indicated by green and red asterisks, respectively.
Figure S2. Phylogenetic tree of TPSs reconstructed with proteins longer than 475 amino acids.
Figure S3. Phylogenetic tree of TPSs reconstructed with proteins longer than 500 amino acids.
Figure S4: GC/MS Analysis of LYC_GAON_PTPS1 (HsLS)

Above, a total ion chromatogram shows shared retention time of the authentic levopimaradiene standard (49) (A) relative to GC of the LYC_GAON_PTPS1 (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E).
Above, a total ion chromatogram shows shared retention time of the authentic levopimaradiene standard (49) (A) relative to GC of the LYC_ZZEI_PTPS4 (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E).
Figure S6: GC/MS Analysis of Copaly-I-PP Synthases

Above, an extracted ion chromatogram (m/z 257) shows shared retention time of the authentic ent-CPP standard (50) (A) relative to GC of the Lyc_JKAA_PTPS2, ANT_RXRQ_PTPS19, MAR_TXVB_PTPS5, and MAR_LFVP_PTPS12 genes (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E). Note, asterisk denotes derivatized copalol as described in (S1).
Figure S7: GC/MS Analysis of MON_UOMY_PTPS13 (OsCPS)

A) 8-endo Copalol

Above, a total ion chromatogram shows shared retention time of the 8-endo-copalol (52) relative to MON_UOMY_PTPS13. Note the pyrophosphate is cleaved due to endogenous bacterial phosphatases before extraction. (B). The fragmentation pattern of this peak is shown on the top right with matching fragmentation identity. Structure shown on the right as produced by the enzyme with retained diphosphate group.
Figure S8: GC/MS Class I Knock Out Analysis of MAR_YFGP_PTPS5 (PICPSKS) and BRY_IGUH_PTPS1 (LjuCPSKS)

Above, a total ion chromatogram shows shared retention time of the authentic ent kaurene standard produced with a known class II ent-CPP producing enzyme coupled to a known class I ent-kaurene producing enzyme (A) relative to GC of the MAR_YFGP_PTPS5 and BRY_IGUH_PTPS1 wild type enzymes and class I knock out enzymes B). The fragmentation pattern ent copalol standard peak (C) is shown on the top right along with the ent kaurene standard peak (E) coupled to the matching fragmentation identity of the gene product D). Structure of Ent-Kaurene shown on the right (F).
Figure S9: GC/MS Analysis of Coupled ent-CPP Synthases to an ent-Kaurene Synthase

ent-Kaurene Std
LYC_JKAA_PTPS2 + AtKS (SwCPS)
ANT_RXRQ_PTPS19 + AtKS (PcCPS)
MAR_TXVB_PTPS5 + AtKS (LcCPS)
MAR_LFVP_PTPS12 + AtKS (MpcCPS)

Above, a total ion chromatogram shows shared retention time of the authentic ent-kaurene standard (50) (A) relative to GC of the LYC_JKAA_PTPS2, ANT_RXRQ_PTPS19, MAR_TXVB_PTPS5, and MAR_LFVP_PTPS12 genes expressed with known monofunctional class I ent-kaurene producing AtKS gene (B). The fragmentation pattern of this peak of the standard (C) is shown on the top right with matching fragmentation identity (D). Structure shown on the right (E).
Figure S10: GC/MS Class I Knock Out Analysis of AsuTPS1 (AfKS)

Above, a total ion chromatogram shows shared retention time of the authentic ent-kaurene standard (50) (A) relative to GC of the AsuTPS1 gene coupled to class II ent-CPG production (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E).
Figure S11: GC/MS Analysis of MAR_YBQN_PTPS2 (OpKOS)

Above, a total ion chromatogram shows shared retention time of the authentic kolavenol standard (50) (A) relative to GC of the MAR_YBQN_PTPS2 gene (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right in the produce form with retained pyrophosphate (E).
Figure S12: GC/MS Analysis of MpDTPS5

Above, a total ion chromatogram shows shared retention time of the authentic terpentedienol standard (50) (A) relative to GC product of the MpDTPS5 (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E).
Figure S13: GC/MS Analysis of Fern_RS_38c8471 (QjMS)

Above, a total ion chromatogram shows shared retention time of the authentic syn-manool standard (S3) (A) relative to GC of the fern RS_38c8471 (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E).
Figure S14: GC/MS Analysis of MON_NDUV_PTPS4 (VaCPS)

Above, an extracted ion chromatogram (m/z 272) shows shared retention time of the authentic ent-Copalol and ent-Kaurene standards (A) relative to GC of the MON_NDUV_PTPS4 and the known stereospecific ent-Kaurene specific synthase AtKS (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure of the wild type product ent-CPP shown on the right (E).
Figure S15: GC/MS Analysis of ANT_WCZB_PTPS1(PcSS)

Above, an extracted ion chromatogram (m/z 257 and m/z 272) shows shared retention time of the authentic sandaracopimaradiene (A) relative to GC of the ANT_WCZB_PTPS1 (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E).
Figure S16 GC/MS Analysis of MpDTPS7

Above, an extracted ion chromatogram (m/z 95) shows shared retention time of the authentic terpentedienol standard (B) relative to GC of the MpDTPS7 product (A). The fragmentation pattern of these peaks are (C) for the gene product and (D) from terpentedienol standard. Structure shown on the right (E).
Figure S17 GC/MS Analysis of MpDTPS2

Above, an extracted ion chromatogram (m/z 95) shows shared retention time of the authentic kolavenol standard (50) (B) relative to the MpDTPS2 product indicated in (A) with the asterisks. The fragmentation pattern of these peaks are (C) for the gene product and (D) from the terpenedienol standard. Structure shown to the right in the ent form though result could be normal as well given the limitation of the gc/ms analysis (E).
Figure S18: GC/MS Analysis of BRY_CMEQ_PTPS1 (OIIAS)

Above, a total ion chromatogram shows shared retention time of the authentic Isobienol standard (53) (A) relative to GC of the BRY_CMEQ_PTPS1 (B). The fragmentation pattern of this peak for the standard (C) is shown on the top right with matching fragmentation identity of the gene product (D). Structure shown on the right (E).
**Figure S19: NMR Assignment of cis-kolav 3,13E-Dienyl-PP**

| Position | $\delta_C$ | $\delta_H$ |
|----------|------------|------------|
| 1 a      | 17.72      | 1.79 [1 H, m] |
| b        | 1.98 [1 H, m] |
| 2 a      | 24.07      | 2.10 [1 H, m] |
| b        | 2.00 [1 H, m] |
| 3        | 123.13     | 5.25 [1H, br s] |
| 4        | 139.9      |            |
| 5        | 36.86      |            |
| 6 a      | 32.71      | 2 [1 H, m] |
| b        | 1.06 [1H, $d_i = 4$ Hz] |
| 7 a      | 28.79      | 1.23 [1H, m] |
| b        | 1.19 [1H, m] |
| 8        | 37.79      | 1.41 [1H, m] |
| 9        | 40.08      |            |
| 10       | 44.6       | 1.34 [1H, dd, $J=7$, 2 Hz] |
| 11 a     | 36.88      | 1.57 [1H, m] |
| b        | 1.35 [1H, m] |
| 12       | 32.7       | 1.86 [2H, br t, $J=11$ Hz] |
| 13       | 141.6      |            |
| 14       | 122.8      | 5.4 [1H, tq, $J=9$, 1 Hz] |
| 15       | 59.5       | 4.13 [2H, $d$, $J=9$ Hz] |
| 16       | 16.5       | 1.68 [3H, br s] |
| 17       | 15.96      | .75 [3H, d $J=9$ Hz] |
| 18       | 19.79      | 1.66 [3H, $q$ $J=2$] |
| 19       | 33.11      | 1.02 [3H, s] |
| 20       | 17.31      | .82 [3H, s] |

**A)** Carbon and Proton Assignments of cis-kolav 3,13E-Dienyl-pp. Product previously reported as ent-neo-cis,cis-kolavenol (b). **(B)** Structure of molecule.
Figure S20: NMR Assignments of ent-iso-Pimara 8,15 Diene

| Position | $\delta_c$ | $\delta_h$ |
|----------|------------|------------|
| 1 a      | 37         | 1.732 (1H, m) |
| b        |            | 1.032 1H, (td, J=16) |
| 2 a      | 19.27      | 1.57 (1H, m) |
| b        |            | 1.436 (1H, m) |
| 3 a      | 42.11      | 1.33 (1H, m) |
| b        |            | 1.375 (1H, m) |
| 4        | 33.54      |            |
| 5        | 52.07      | 1.137 (1H, d, J=13) |
| 6 a      | 19.187     | 1.632 (1H, m), |
| b        |            | 1.436 (1H, m) |
| 7        | 32.88      | 1.886 (2H, m) |
| 8        | 124.36     |            |
| 9        | 136.82     |            |
| 10       | 37.77      |            |
| 11       | 22.88      | 1.269 (2H, m) |
| 12 a     | 34.237     | 1.387 (1H, m), |
| b        |            | 1.349 (1H, m) |
| 13       | 34.988     |            |
| 14 a     | 42.52      | 1.733 (1H, d, J=17) |
| b        |            | 1.549 (1H, d, J=23) |
| 15       | 149.65     | 5.792 (1H, dd, J = 13 Hz, J = 8 Hz) |
| 16 a     | 109.64     | 4.89 (1H, d, J=22), |
| b        |            | 4.848 (1H, J= 13) |
| 17       | 23.416     | 0.919 (3H, s) |
| 18       | 33.47      | 0.863 (3H, s) |
| 19       | 21.9       | 0.828 (3H, s) |
| 20       | 19.657     | 0.932 (3H, s) |

A) Carbon and Proton Assignments of ent-iso-Pimara 8,15 Diene. B) Structure of molecule
References

1. E. J. Carpenter et al., Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). Gigascience 8 (2019).
2. Q. Jia et al., Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. Proceedings of the National Academy of Sciences of the United States of America 113, 12328-12333 (2016).
3. H. Shen et al., Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. Gigascience 7, 1-11 (2018).
4. B. J. Haas et al., De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8, 1494-1512 (2013).
5. Y. Wu et al., Functional characterization of wheat copalyl diphosphate synthases sheds light on the early evolution of labdane-related diterpenoid metabolism in the cereals. Phytochemistry 84, 40-46 (2012).
6. K. A. Pelot, D. M. Hagelthorn, Y. J. Hong, D. J. Tantillo, P. Zerbe, Diterpene Synthase-Catalyzed Biosynthesis of Distinct Clerodane Stereoisomers. ChemBioChem 20, 111-117 (2019).
7. M. Xu et al., Characterization of an orphan diterpenoid biosynthetic operon from Salinispora arenicola. J Nat Prod 77, 2144-2147 (2014).
8. T. Nishiyama et al., The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell 174, 448-464.e424 (2018).
9. K. Hori et al., Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nat Commun 5, 3978 (2014).
10. S. Wang et al., Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. Nat Plants 6, 95-106 (2020).
11. S. Cheng et al., Genomes of Subaerial Zygmatophyceae Provide Insights into Land Plant Evolution. Cell 179, 1057-1067.e1014 (2019).
12. I. V. Grigoriev et al., MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research 42, D699-D704 (2013).
13. C. Gao et al., Oil accumulation mechanisms of the oleaginous microalga Chlorella protothecoides revealed through its genome, transcriptomes, and proteomes. BMC Genomics 15, 582 (2014).
14. H. Moreau et al., Gene functionalities and genome structure in Bathycoccus prasinus reflect cellular specializations at the base of the green lineage. Genome biology 13, R74 (2012).
15. D. R. Browne et al., Draft Nuclear Genome Sequence of the Liquid Hydrocarbon-Accumulating Green Microalga Botryococcus braunii Race B (Showa). Genome Announc 5 (2017).
16. S. Hirooka et al., Acidophilic green algal genome provides insights into adaptation to an acidic environment. Proc Natl Acad Sci U S A 114, E8304-e8313 (2017).
17. S. S. Merchant et al., The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318, 245-250 (2007).
18. G. Blanc et al., The Chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell 22, 2943-2955 (2010).
19. M. S. Roth et al., Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production. Proc Natl Acad Sci U S A 114, E4296-e4305 (2017).
20. G. Blanc et al., The genome of the polar eukaryotic microalga Coccomyxa subellipsoidea reveals traits of cold adaptation. Genome Biol 13, R39 (2012).
21. J. E. W. Polle et al., Draft Nuclear Genome Sequence of the Halophilic and Beta-Carotene-Accumulating Green Alga Dunaliella salina Strain CCAP19/18. Genome Announc 5 (2017).
22. E. R. Hanschen et al., The Gonium pectorale genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat Commun* **7**, 11370 (2016).
23. J. F. Pombert, N. A. Blouin, C. Lane, D. Boucias, P. J. Keeling, A lack of parasitic reduction in the obligate parasitic green alga Heliocorposporidium. *PLoS Genet* **10**, e1004355 (2014).
24. M. A. Arrioni et al., Genome sequences of Chlorella sorokiniana UTEX 1602 and Micractinium conductrix SAG 241.80: implications to maltose excretion by a green alga. *Plant J* **93**, 566-586 (2018).
25. A. Z. Worden et al., Green evolution and dynamic adaptations revealed by genomes of the marine picocyanurates Micromonas. *Science* **324**, 268-272 (2009).
26. C. Bogen et al., Reconstruction of the lipid metabolism for the microalgae Monoraphidium neglectum from its genome sequence reveals characteristics suitable for biofuel production. *BMC Genomics* **14**, 926 (2013).
27. B. Palenik et al., The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A* **104**, 7705-7710 (2007).
28. R. Blanc-Mathieu et al., An improved genome of the model marine alga Ostreococcus tauri unfolds by assessing Illumina de novo assemblies. *BMC Genomics* **15**, 1103 (2014).
29. C. R. Gonzalez-Esquer, S. N. Twary, B. T. Hovde, S. R. Starkenburg, Nuclear, Chloroplast, and Mitochondrial Genome Sequences of the Prospective Microalgal Biofuel Strain Picoclorhizum salocecismus. *Genome Announc* **6** (2018).
30. S. Suzuki, H. Yamaguchi, N. Nakajima, M. Kawachi, Raphidocelis subcapitata (=Pseudokirchneriellia subcapitata) provides an insight into genome evolution and environmental adaptations in the Sphaeropleales. *Sci Rep* **8**, 8058 (2018).
31. J. Featherston et al., The 4-Celled Tetraebaena socialis Nuclear Genome Reveals the Essential Components for Genetic Control of Cell Number at the Origin of Multicellularity in the Volvocale Lineage. *Mol Biol Evol* **35**, 855-870 (2018).
32. O. De Clerck et al., Insights into the Evolution of Multicellularity from the Sea Lettuce Genome. *Curr Biol* **28**, 2921-2933.e2925 (2018).
33. S. E. Prochnik et al., Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. *Science* **329**, 223-226 (2010).
34. D. M. Goodstein et al., Phytzoome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**, D1178-D1186 (2011).
35. J. L. Wegzryn et al., Cyberinfrastructure to Improve Forest Health and Productivity: The Role of Tree Databases in Connecting Genomes, Phenomes, and the Environment. *Front Plant Sci* **10**, 813 (2019).
36. F. W. Li et al., Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature plants* **4**, 460-472 (2018).
37. F. W. Li et al., Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature plants* **6**, 259-272 (2020).
38. J. L. Bowman et al., Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell* **171**, 287-304.e215 (2017).
39. K. Hayashi et al., Identification and functional analysis of bifunctional ent-kaurene synthase from the moss Physcomitrella patens. *FEBS Lett* **580**, 6175-6181 (2006).
40. G. Li et al., Nonseed plant Selaginella moellendorffii [corrected] has both seed plant and microbial types of terpene synthases. *Proc Natl Acad Sci U S A* **109**, 14711-14715 (2012).
41. H. Kawaide et al., Identification of the single amino acid involved in quenching the ent-kaurenyl cation by a water molecule in ent-kaurene synthase of Physcomitrella patens. *Febs J* **278**, 123-133 (2011).
42. K. Okada et al., HpDTC1, a Stress-Inducible Bifunctional Diterpene Cyclase Involved in Momilactone Biosynthesis, Functions in Chemical Defence in the Moss Hypnum plumaeforme. Sci Rep 6, 25316 (2016).
43. J. Tanaka et al., Antheridiogen determines sex in ferns via a spatiotemporally split gibberellin synthesis pathway. Science 346, 469-473 (2014).
44. M. L. Hillwig et al., Domain loss has independently occurred multiple times in plant terpene synthase evolution. Plant J 68, 1051-1060 (2011).
45. S. Kumar et al., Molecular Diversity of Terpene Synthases in the Liverwort Marchantia polymorpha. Plant Cell 28, 2632-2650 (2016).
46. M. Shimane et al., Molecular evolution of the substrate specificity of ent-kaurene synthases to adapt to gibberellin biosynthesis in land plants. Biochem J 462, 539-546 (2014).
47. Y. Sugai et al., Enzymatic (13)C labeling and multidimensional NMR analysis of miltiradiene synthesized by bifunctional diterpene cyclase in Selaginella moellendorffii. J Biol Chem 286, 42840-42847 (2011).
48. S. Mafu, M. L. Hillwig, R. J. Peters, A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from Selaginella moellendorffii. Chembiochem 12, 1984-1987 (2011).
49. R. J. Peters, R. B. Croteau, Abietadiene synthase catalysis: mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. Proceedings of the National Academy of Sciences of the United States of America 99, 580-584 (2002).
50. M. Jia, K. C. Potter, R. J. Peters, Extreme promiscuity of a bacterial and a plant diterpene synthase enables combinatorial biosynthesis. Metab Eng 37, 24-34 (2016).
51. C. Lemke, K. C. Potter, S. Schulte, R. J. Peters, Conserved bases for the initial cyclase in gibberellin biosynthesis: from bacteria to plants. Biochem J 476, 2607-2621 (2019).
52. K. A. Pelot et al., Functional Diversity of Diterpene Synthases in the Biofuel Crop Switchgrass. Plant Physiol 178, 54-71 (2018).
53. M. Jia, S. K. Mishra, S. Tufts, R. L. Jernigan, R. J. Peters, Combinatorial biosynthesis and the basis for substrate promiscuity in class I diterpene synthases. Metab Eng 55, 44-58 (2019).