

ARTICLE

Received 15 Jan 2016 | Accepted 29 Feb 2016 | Published 6 Apr 2016

DOI: 10.1038/ncomms11198

OPEN

# A squalene synthase-like enzyme initiates production of tetraterpenoid hydrocarbons in *Botryococcus braunii* Race L

Hem R. Thapa<sup>1</sup>, Mandar T. Naik<sup>1,2</sup>, Shigeru Okada<sup>3,4</sup>, Kentaro Takada<sup>3,4</sup>, István Molnár<sup>5</sup>, Yuquan Xu<sup>5,6</sup> & Timothy P. Devarene<sup>1</sup>

The green microalga *Botryococcus braunii* is considered a promising biofuel feedstock producer due to its prodigious accumulation of hydrocarbon oils that can be converted into fuels. *B. braunii* Race L produces the C<sub>40</sub> tetraterpenoid hydrocarbon lycopadiene via an uncharacterized biosynthetic pathway. Structural similarities suggest this pathway follows a biosynthetic mechanism analogous to that of C<sub>30</sub> squalene. Confirming this hypothesis, the current study identifies C<sub>20</sub> geranylgeranyl diphosphate (GGPP) as a precursor for lycopaoctane biosynthesis, the first committed intermediate in the production of lycopadiene. Two *squalene synthase* (SS)-like complementary DNAs are identified in race L with one encoding a true SS and the other encoding an enzyme with lycopaoctane synthase (LOS) activity. Interestingly, LOS uses alternative C<sub>15</sub> and C<sub>20</sub> prenyl diphosphate substrates to produce combinatorial hybrid hydrocarbons, but almost exclusively uses GGPP *in vivo*. This discovery highlights how SS enzyme diversification results in the production of specialized tetraterpenoid oils in race L of *B. braunii*.

<sup>1</sup>Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843, USA. <sup>2</sup>Biomolecular NMR Laboratory, Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843, USA. <sup>3</sup>Laboratory of Aquatic Natural Products Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Yayoi, Bunkyo, Tokyo 113-8657, Japan. <sup>4</sup>Japan Science and Technology Agency-Core Research for Evolutional Science and Technology (CREST), Gobancho, Chiyoda, Tokyo 102-0076, Japan. <sup>5</sup>Natural Products Center, School of Natural Resources and the Environment, The University of Arizona, Tucson, Arizona 85739, USA. <sup>6</sup>Biotechnology Research Institute, The Chinese Academy of Agricultural Sciences, Beijing 100081, China. Correspondence and requests for materials should be addressed to T.P.D. (email: tpd8@tamu.edu).

Microalgae are a promising next-generation source of feedstocks for biofuel production with the potential to serve as a practical alternative to petroleum-based transportation fuels<sup>1</sup>. Depending on the microalgal species, the oils produced vary from triacylglycerols to hydrocarbons<sup>2,3</sup>. Hydrocarbon-based fuels are preferred over other biofuels, as they are highly compatible with existing petroleum infrastructures and possess superior fuel properties<sup>4,5</sup>. The colony-forming green microalga *B. braunii* is an exciting candidate for biofuel feedstock production, as it produces up to 61% of its dry weight as liquid hydrocarbon oils<sup>6</sup>. These hydrocarbons are produced inside the cells of the colony, seen as intracellular oil bodies and secreted into the colony extracellular matrix where the majority of the hydrocarbons are stored<sup>6</sup> (Fig. 1a). Most importantly, catalytic hydrocracking of hydrocarbons from this alga results in petroleum-equivalent fuels of gasoline, kerosene and diesel<sup>7</sup>. Intriguingly, geologic evidence also shows a direct contribution of this alga to the formation of currently used fossil fuel deposits around the globe<sup>8–12</sup>. Despite the aforementioned advantages of *B. braunii*, its use for biofuel feedstock production is hindered by a slow growth rate and the lack of transformation systems to achieve targeted genetic modification<sup>3</sup>. Thus, the identification of *B. braunii* hydrocarbon biosynthetic pathways and associated genes/enzymes can provide options for metabolically engineering these pathways into heterologous hosts with better growth characteristics and the ability to be genetically manipulated. This would then allow the development of improved versions of hydrocarbon biosynthetic enzymes, to direct production towards the most commercially desirable products<sup>13</sup>.

There are three different races of *B. braunii* based on the hydrocarbons synthesized. Race A produces fatty acid-derived C<sub>23</sub>–C<sub>33</sub> alkadienes and alkatrienes. Races B and L produce isoprenoid-derived hydrocarbons: methylsqualenes and C<sub>30</sub>–C<sub>37</sub> botryococcene triterpenoids in race B and the C<sub>40</sub> tetraterpenoid lycopadiene, the focus of this study, in race L<sup>6</sup>. Green algae have been shown to possess only the plastid-localized methyl erythritol phosphate pathway to supply isoprenoid precursors for terpene production<sup>14,15</sup>, and thus lycopadiene is predicted to be generated from C<sub>5</sub> precursors via this route. However, the exact mechanism of lycopadiene biosynthesis from C<sub>20</sub> prenyl diphosphate intermediates has been a mystery and two possible biosynthetic routes have been suggested (Fig. 1b)<sup>16</sup>. The first entails C<sub>20</sub> geranylgeranyl diphosphate (GGPP) reduction by GGPP reductase to produce C<sub>20</sub> phytol diphosphate (PPP; Fig. 1b). Two molecules of PPP would then undergo head-to-head condensation (1-1' linkage) to produce lycopadiene (Fig. 1b). The second possibility is the head-to-head condensation of two GGPP molecules to produce lycopaoctaene, followed by stepwise enzymatic reduction to produce lycopadiene (Fig. 1b).

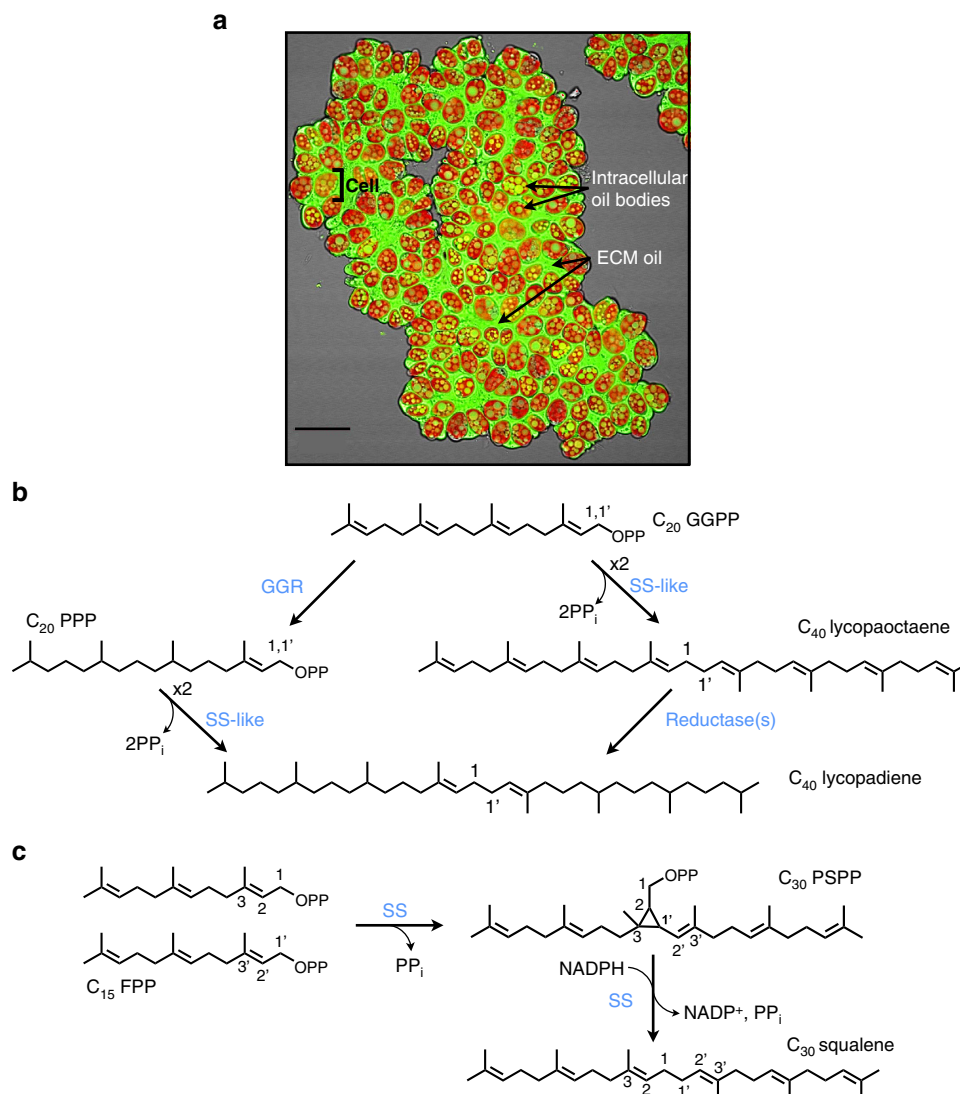
Using either proposed route, the condensation of PPP or GGPP is predicted to proceed in a reaction mechanism similar to that carried out by the enzyme squalene synthase (SS), which forms C<sub>30</sub> squalene, a precursor required for sterol biosynthesis in eukaryotes<sup>17</sup> and hopanoids in some prokaryotes<sup>18</sup>. SS enzymes catalyse a two-step reaction. First, the condensation of two C<sub>15</sub> farnesyl diphosphate (FPP) molecules yields the cyclopropyl intermediate presqualene diphosphate (PSPP; Fig. 1c). Second, PSPP undergoes NADPH-dependent reductive rearrangement to form squalene with a 1-1' linkage between the two FPP molecules (Fig. 1c)<sup>19</sup>. Herein we report the elucidation of the first committed step in the lycopadiene hydrocarbon biosynthetic pathway in *B. braunii* Race L, which is catalysed by a new SS-like (SSL) enzyme.

## Results

**L race hydrocarbon content and related enzyme activity.** Previous studies reported *trans,trans*-lycopadiene as the predominant hydrocarbon (98% of total hydrocarbons) produced by race L, with a small amount of lycopatriene also documented<sup>16,20,21</sup>. However, the lycopatriene structure was not determined. Thus, we first set out to ascertain and refine the hydrocarbon profile of *B. braunii* Race L, to provide a baseline for our biosynthetic models. Analysis of purified hydrocarbons by gas chromatography–mass spectrometry (GC–MS) showed that lycopadiene accounts for 95% of the total hydrocarbon pool (Fig. 2a (I)), with six other minor compounds constituting the remaining 5%: lycopatriene, lycopatetraene, lycopapentaene, lycopapentaene isomer, lycopahexaene and a C<sub>35</sub>H<sub>64</sub> molecule (Fig. 2a (II–VII) and Supplementary Figs 1–4). Ozonolysis experiments suggested lycopatriene and lycopapentaene share an identical reduced C<sub>20</sub> moiety with lycopadiene (Supplementary Figs 5–7). Nuclear magnetic resonance (NMR) spectroscopy was used to confirm identity and structure, including double bond positions, of each molecule (Fig. 2a, Supplementary Table 1 and Supplementary Figs 8–15). This is the first report of lycopatetraene, lycopapentaene and lycopahexaene hydrocarbons from *B. braunii* race L. The lack of detectable amounts of lycopaoctaene and lycopaeptaene in the total hydrocarbon fraction could be due to rapid conversion of these metabolites into more highly saturated homologues. Interestingly, the unique C<sub>35</sub>H<sub>64</sub> isoprenoid hydrocarbon detected contains seven isoprene units and thus is assigned to the recently named C<sub>35</sub> terpene class ‘sesquarterpenes’<sup>22</sup>.

To gain more insight into the mechanism of lycopadiene biosynthesis and assess the validity of the two possible lycopadiene biosynthetic routes (Fig. 1b), a cell-free *in vitro* enzyme assay was developed using total, soluble and microsomal protein fractions with <sup>3</sup>H-labelled PPP and GGPP substrates for potential incorporation into lycopadiene and lycopaoctaene, respectively. Squalene production by SS enzyme activity with <sup>3</sup>H-labelled FPP as substrate was used as a control<sup>23</sup>. Lycopadiene production from PPP was not detected in any protein fraction tested, whereas lycopaoctaene biosynthesis from GGPP was easily observed, mainly in the microsomal fraction (Fig. 2b). This indicates a lycopaoctaene synthase (LOS) activity is localized to a membrane system, possibly the endoplasmic reticulum (ER) as is seen for SS<sup>24</sup>. Tellingly, LOS activity in these assays was detected when using the same cofactor and divalent metal cation as those used for SS activity measurement. These results suggest direct lycopadiene biosynthesis from PPP is not a significant contributor to C<sub>40</sub> hydrocarbon production in race L, whereas conversion of GGPP into lycopaoctaene occurs readily, as proposed in the second hypothetical route (Fig. 1b). Correspondingly, LOS enzyme activity directly correlated with lycopadiene accumulation over the growth cycle, suggesting this activity is related to hydrocarbon biosynthesis (Supplementary Fig. 16). Detection of LOS activity from an algal homogenate in an assay similar to that of SS supports the notion that an LOS enzyme may be similar to a typical SS enzyme.

**Identification of an LOS enzyme.** To identify the gene(s) responsible for lycopaoctaene biosynthesis, we generated and computationally screened an L race transcriptomic database for SSL sequences. Two SSL cDNAs were identified and named based on the function of their encoded proteins as detailed below: SS from race L (*LSS*) and *LOS*. Both the *LSS* and *LOS* proteins contain all five conserved activity domains, the transmembrane domain and the NADPH-binding residues found in typical SS

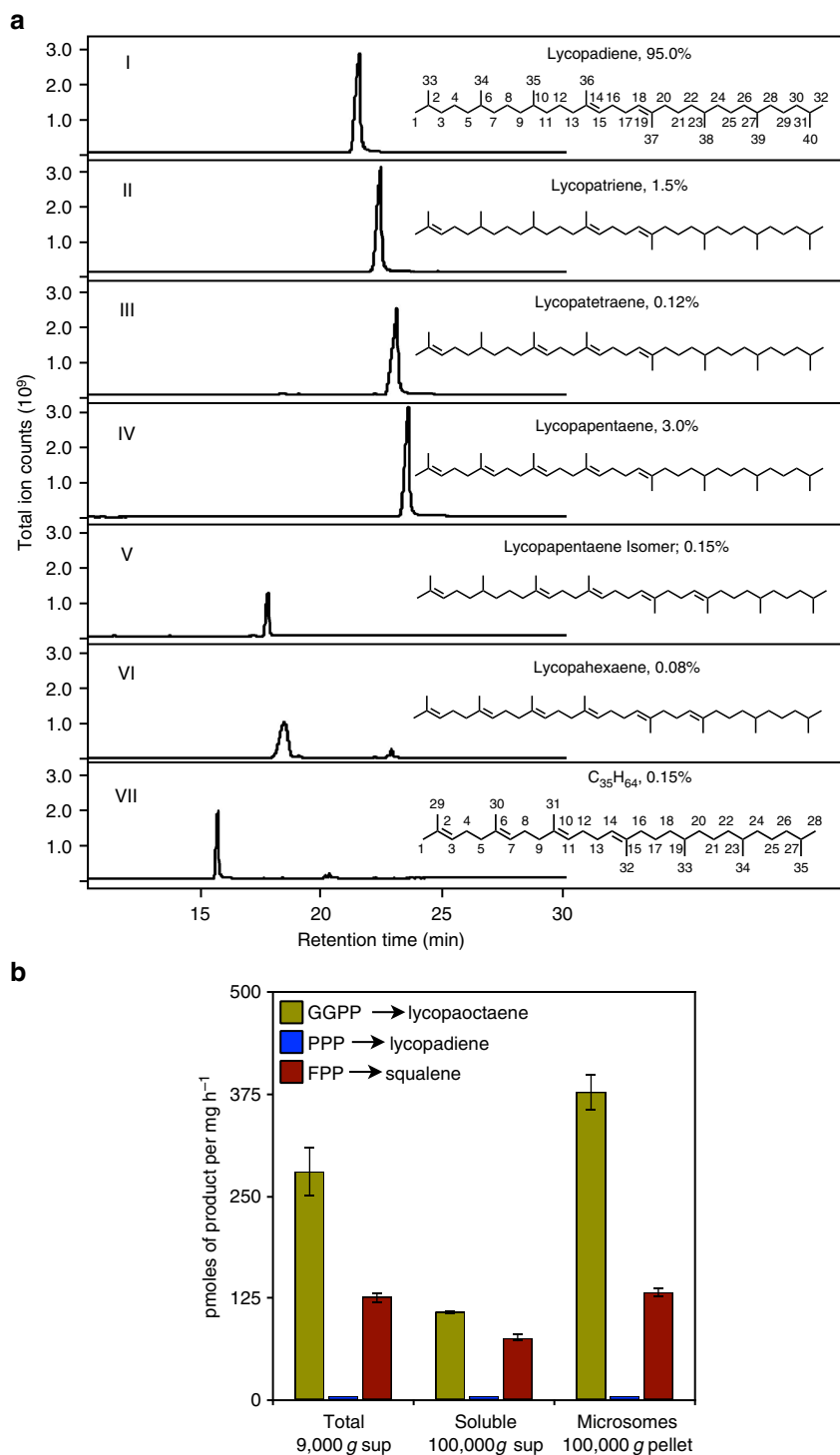


**Figure 1 | Background information on lycopadiene biosynthesis in *B. braunii* race L. (a)** Confocal microscopy image of an L race colony of *B. braunii*. Shown are hydrocarbon oils stained using the neutral lipid-binding stain Nile red (false coloured green) and chlorophyll autofluorescence (red). Indicated are intracellular hydrocarbon oil bodies, hydrocarbon oil in the extracellular matrix (ECM) and an individual cell as defined by chlorophyll autofluorescence from the single chloroplast in each cell. Image is an overlay of Nile red signal, chlorophyll autofluorescence signal and a bright-field image. Scale bar, 20  $\mu\text{m}$ . **(b)** Two possible pathways for lycopadiene biosynthesis. Possibility 1: GGPP reduction to PPP followed by condensation of two PPP molecules, to directly produce lycopadiene. Possibility 2: condensation of two molecules of GGPP to form lycopaoctaene, which would then be reduced to lycopadiene. **(c)** Two-step reaction catalysed by single SS enzyme: two molecules of FPP are condensed to squalene through the PSPP intermediate. In **b** and **c**, enzyme names are coloured blue.

enzymes (Supplementary Figs 17 and 18)<sup>17</sup>. Comparison of the encoded amino acid sequences showed 57.5 and 49.2% identity to SS from race B (BSS)<sup>25</sup> for LSS and LOS, respectively (Supplementary Fig. 17). In addition, comparison of LOS with other SS and SSL enzymes showed 52.5% identity to LSS, 47.3% identity to the *Solanum lycopersicum* (tomato) SS (SSS), 44.6, 61.7 and 44.1% identity to the SSL-1, -2 and -3 from the B race of *B. braunii*, and 33.9% identity to the human SS (HSS; Supplementary Fig. 18). Several studies<sup>17,26–28</sup>, including the recent determination of the crystal structure of HSS<sup>29</sup>, have elucidated the catalytic mechanism of the SS reaction. These studies have shown the FLAP domain, domains I–V and the JK loop of HSS form the SS active site<sup>29</sup>. Furthermore, several HSS residues (Ser51, Sr53, Tyr171 and Gln212), two DxxED motifs in domains II and IV, and the NADPH-binding residues were determined to be key for catalytic activity (see Supplementary

Fig. 18)<sup>29</sup>. All of these catalytic residues and motifs are also conserved in both the LSS and LOS enzymes (Supplementary Fig. 18), making it difficult at this time to determine which domains/residues are specific for LOS activity in a comparison of LSS and LOS.

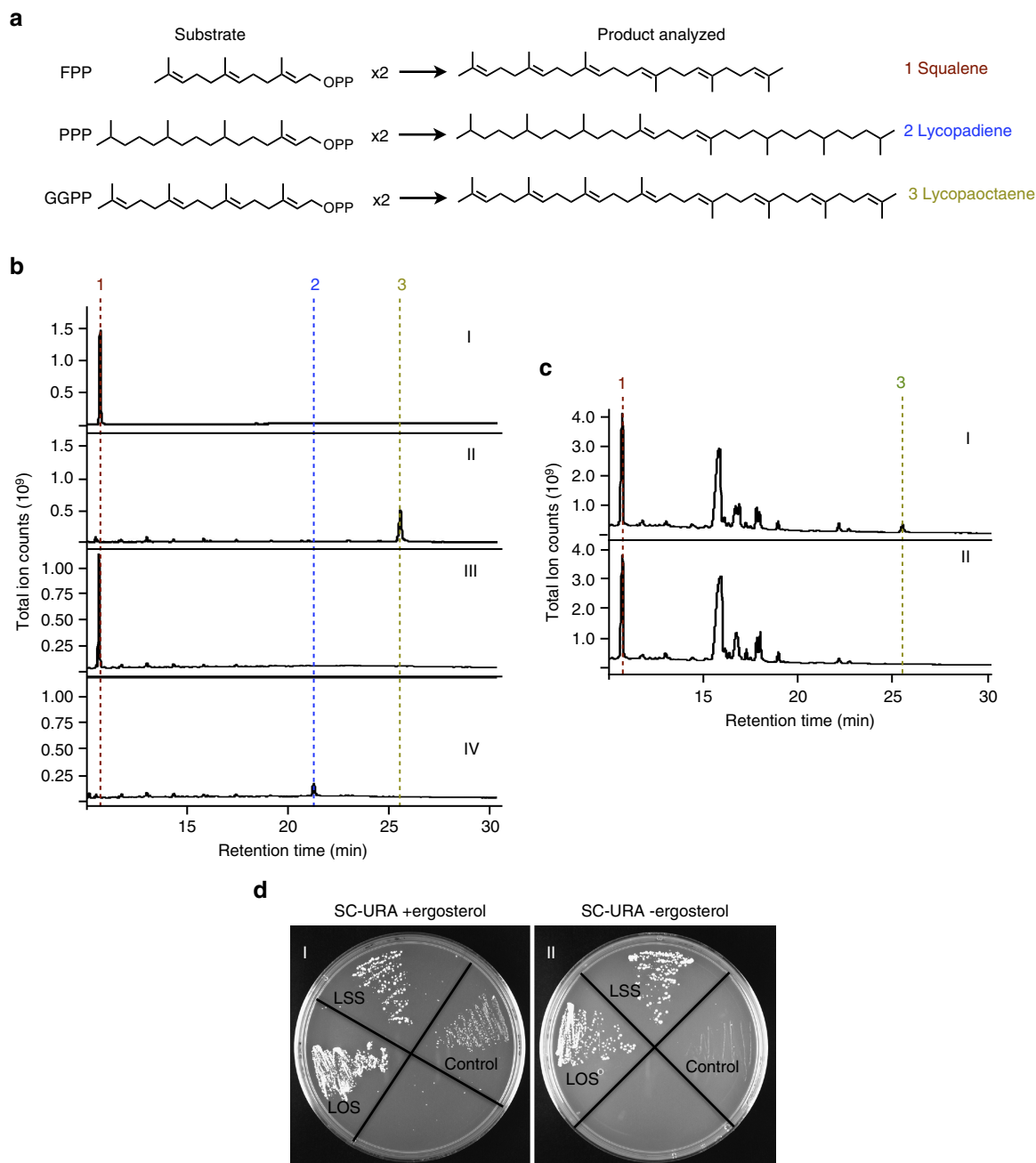
For initial characterization of the enzymatic activity of LSS and LOS, both proteins were expressed in *Escherichia coli* and the purified recombinant proteins tested for enzyme activity. LSS was shown to yield squalene as its sole product in an *in vitro* assay with FPP as the substrate, as expected (Fig. 3a,b (I)). However, purified recombinant LOS incubated with GGPP in an *in vitro* assay showed lycopaoctaene as the sole reaction product (Fig. 3a,b (II)). As LOS may have arisen from an SS paralogue that evolved to accept GGPP as substrate for lycopaoctaene production, we considered that LOS may have retained the ability to use FPP to produce squalene. Indeed, squalene production was detected



**Figure 2 | Hydrocarbon composition and hydrocarbon-related enzyme activities in *B. braunii* race L.** (a) Representative GC-MS profiles of purified hydrocarbons and their corresponding names and structures (I–VII). Structures of hydrocarbons were determined by <sup>1</sup>H- and <sup>13</sup>C-NMR. Lycopadiene carbon numbering (I) is used for numbering all C<sub>40</sub> hydrocarbons of race L. Percentage next to the name of each hydrocarbon corresponds to its average amount present in total hydrocarbon pool from three independent culture cycles (*n* = 3). (b) Radioactive enzyme assays were conducted using different protein fractions of race L cell lysates to test the two possible lycopadiene biosynthetic pathways. Enzyme activities were calculated based on the incorporation of <sup>3</sup>H-GGPP into lycopaoctaene, <sup>3</sup>H-PPP into lycopadiene and <sup>3</sup>H-FPP into squalene. SS enzyme activity is used as a positive control. Values shown are the mean ± s.e. obtained from three independent experiments (*n* = 3).

when LOS was incubated with FPP in an *in vitro* reaction (Fig. 3a,b (III)). More surprising, LOS incubation with PPP yielded lycopadiene, albeit at levels much lower than lycopaoctaene or squalene (Fig. 3a,b (IV)). Similar *in vitro*

incubations of LSS with GGPP or PPP did not result in lycopaoctaene or lycopadiene production (see below). As LOS can use three different substrates, steady-state kinetic experiments were performed with FPP, GGPP and PPP, to analyse LOS



**Figure 3 | Functional characterization of SSL enzymes from race L *in vitro* and *in vivo*.** (a) Diagram of reactions showing substrates and products for the analysis in b and c. (b) GC-MS profiles of LSS and LOS *in vitro* reaction products: I, LSS products using FPP as substrate to produce squalene (1); II, LOS products using GGPP as substrate to produce lycopaoctaene (3); III, LOS products using FPP as substrate to produce squalene (1); and IV, LOS products using PPP as substrate to produce lycopadiene (2). (c) GC-MS profiles of LOS yeast *in vivo* reaction products: I, *n*-hexane extractable metabolites of yeast expressing *AtGGPPS11* and LOS, to make lycopaoctaene (3); or II, yeast expressing *AtGGPPS11* alone. (d) Expression of LSS and LOS in the SS (*ERG9*) yeast knockout strain: I, in the presence of ergosterol, or II, in the absence of ergosterol, to test restoration of ergosterol prototrophy. The control carries the expression vector without a transgene. Data shown in b, c and d are representatives from three independent experiments ( $n = 3$ ).

substrate specificity. The LOS enzyme shows higher substrate affinity ( $K_m$ ) for GGPP compared with that of FPP and PPP, and the turnover number ( $k_{cat}$ ) and catalytic efficiency ( $k_{cat}/K_m$ ) for PPP are an order of magnitude less than for FPP and GGPP (Table 1 and Supplementary Fig. 19).

To further characterize LOS, it was coexpressed in yeast with *Arabidopsis thaliana* GGPP synthase-11 (*AtGGPPS11*)<sup>30</sup>, as unlike FPP, yeast GGPP biosynthesis is limited and is considered a bottleneck for GGPP-derived isoprenoid production<sup>31</sup>. Coexpression of LOS and *AtGGPPS11* in yeast resulted in

lycopaoctaene production (Fig. 3c (I)), which was undetectable when *AtGGPPS11* was expressed without LOS (Fig. 3c (II)). Furthermore, when expressed in a yeast SS knockout strain, LOS restored ergosterol prototrophy, indicating its ability to produce squalene *in vivo* (Fig. 3d (I,II)). LSS was also able to restore ergosterol prototrophy as expected (Fig. 3d (I,II)). Taken together, the results in Fig. 3 suggest LSS is a true SS enzyme, whereas LOS appears to be a promiscuous SSL enzyme with broader substrate chain length and saturation specificity.

**Table 1 | LOS steady-state kinetic parameters for GGPP, FPP and PPP substrates.**

	GGPP	FPP	PPP
$K_m$ (mM)	$0.07 \pm 0.02$	$0.13 \pm 0.02$	$0.11 \pm 0.01$
$k_{cat}$ ( $s^{-1}$ )	$1.14 \times 10^{-2}$	$2.05 \times 10^{-2}$	$1.16 \times 10^{-3}$
$k_{cat}/K_m$ ( $\mu M^{-1} s^{-1}$ )	$1.65 \times 10^{-4}$	$1.62 \times 10^{-4}$	$1.03 \times 10^{-5}$

FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; LOS, lycopaoctaene synthase; PPP, phytyl diphosphate.  $K_m$  values shown are the mean  $\pm$  s.e. obtained from three independent measurements ( $n=3$ ).

Previous studies have reported the production of lycopaoctaene (a.k.a. lycopersene) *in vitro* from protein extracts during studies conducted to decipher the carotenoid biosynthetic pathway<sup>32–34</sup>; however, subsequent studies determined lycopaoctaene could not be an intermediate in carotenoid biosynthesis<sup>35,36</sup>, raising the possibility that lycopaoctaene production in these studies was an *in vitro* artefact. A more relevant study reported the ability of purified yeast SS to use GGPP for lycopaoctaene production *in vitro*, at levels much lower than native squalene production<sup>32</sup>. Importantly, we did not detect any lycopaoctaene production when *AtGGPPS11* was expressed alone in wild-type yeast (Fig. 3c (II)), suggesting that yeast SS does not use GGPP *in vivo* under the conditions employed. Similarly, the ability of LOS to catalyse the conversion of PPP to lycopadiene may not have biological significance as we did not detect lycopadiene production using PPP as the substrate *ex vivo* in race L cell lysates (Fig. 2b) and the efficiency of LOS to use PPP as a substrate is quite low (Table 1). Together, these results suggest that the relevant *in vivo* route to lycopadiene production is unlikely to involve the condensation of two PPP molecules by LOS, but rather, as we contend, the condensation of two GGPP molecules to form lycopaoctaene, for eventual conversion to lycopadiene.

**LOS is promiscuous towards prenyl substrates.** The LOS enzyme was further characterized using combinations of FPP, GGPP and PPP as substrates in *in vitro* reactions. Interestingly, in the presence of FPP and GGPP, LOS produced significant amounts of squalene and a  $C_{35}H_{58}$  molecule, and lesser amounts of lycopaoctaene (Fig. 4a (I)). The  $C_{35}H_{58}$  molecule is a chimera produced from head-to-head condensation of one FPP molecule and one GGPP molecule, and its identity was confirmed by GC–MS (Supplementary Fig. 20). Next, when LOS was supplied with FPP and PPP, squalene production predominated with small amounts of  $C_{35}H_{64}$  and lycopadiene (Fig. 4a (II)).  $C_{35}H_{64}$ , also a chimeric hydrocarbon, is produced as a result of head-to-head condensation of one molecule each of FPP and PPP. Finally, LOS incubation with GGPP and PPP produced lycopadiene and lycopapentaene as minor products and lycopaoctaene as the major product (Fig. 4a (III)).

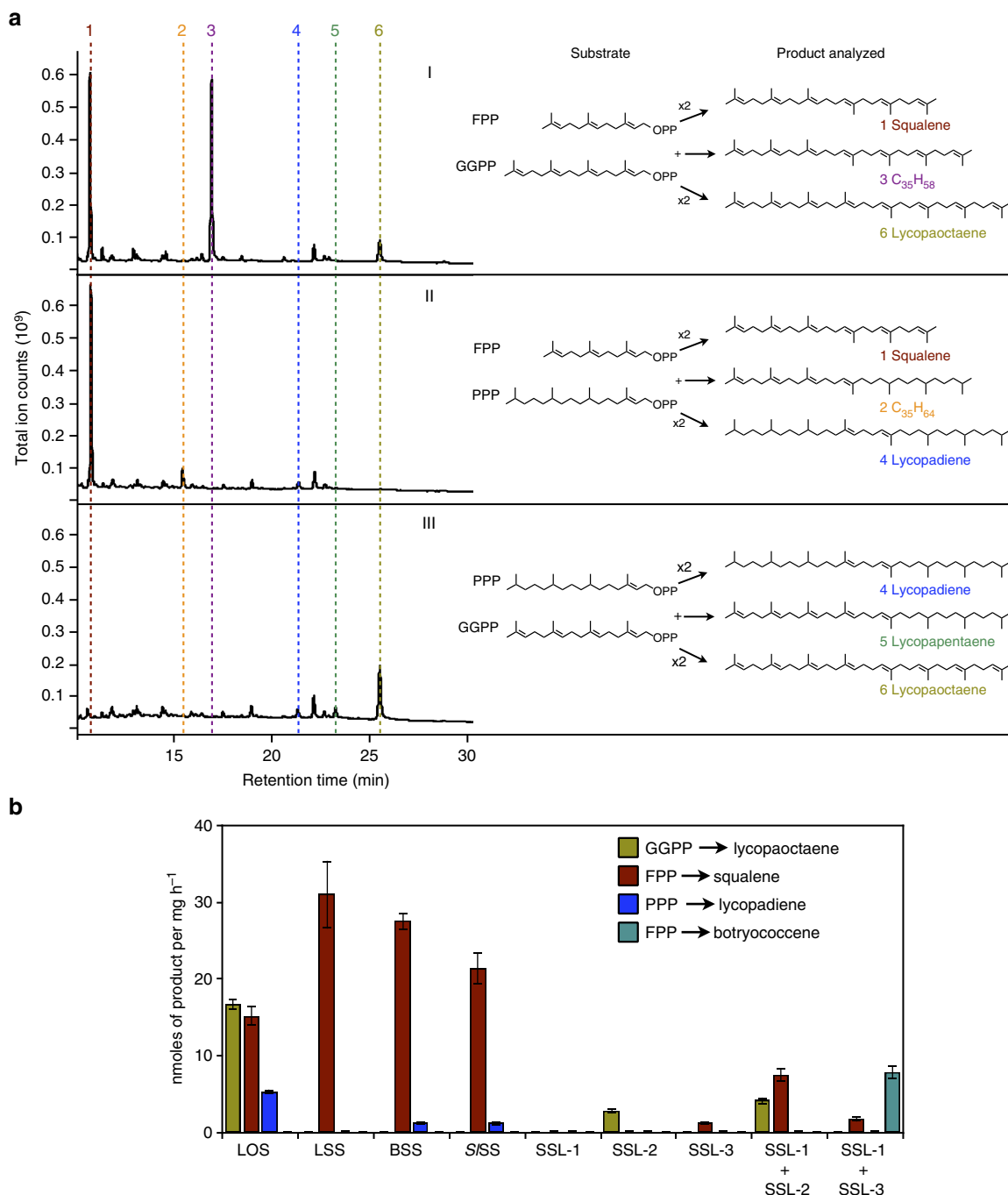
To the best of our knowledge, this is the first report where a wild-type, non-mutated eukaryotic SS or SSL enzyme has been shown to be able to use three naturally occurring prenyl diphosphate substrates to yield hydrocarbon products in all possible combinations. To support this notion, several SS and SSL enzymes were tested for their ability to use FPP, GGPP and PPP as substrates (Fig. 4b). The enzymes tested included LOS and three typical SS enzymes: LSS, BSS and S/SS. In addition, enzymes SSL-1, SSL-2 and SSL-3 from race B were chosen because of their SSL activities. SSL-1 uses FPP to produce PSPP, which is then converted by SSL-2 to squalene, or by SSL-3 to  $C_{30}$ -botryococcene as the major product and squalene as a minor product<sup>37</sup>. As shown in Fig. 4b, the typical SS enzymes display SS activity as their main catalytic function, with BSS and S/SS also using PPP to produce minute amounts of lycopadiene. In contrast, the B race SSL enzymes have limited substrate flexibility, allowing SSL-2 and

SSL-1 plus SSL-2 to use GGPP to generate minor amounts of lycopaoctaene (Fig. 4b). However, LOS is the only enzyme tested that is able to use all three substrates and to produce significant amounts of lycopaoctaene, squalene and lycopadiene (Fig. 4b).

**The LOS reaction uses a cyclopropyl intermediate.** We next conducted enzyme assays to determine whether the LOS reaction mechanism with GGPP as substrate is similar to that of SS, that is, uses a PSPP-like cyclopropyl diphosphate intermediate, which is termed prelycopaoctaene diphosphate (PLPP; Fig. 5a). First, assays were conducted with or without a dinucleotide reducing agent, which would be required to convert PLPP to lycopaoctaene (Fig. 5a). LOS successfully used both NADH and NADPH as reducing agents for lycopaoctaene production, with preference for NADPH (Fig. 5b). In the absence of a dinucleotide reducing agent, LOS activity was lost (Fig. 5b), suggesting the presence of the PLPP reaction intermediate. This result is consistent with previous reports where yeast SS used GGPP *in vitro* to produce lycopaoctaene; however, without NADH or NADPH only the reaction intermediate PLPP accumulated<sup>32</sup>. Next, we showed that LOS is strongly inhibited by squalstatin (Fig. 5c), a potent SS inhibitor that mimics PSPP binding<sup>23,38–40</sup>. Finally, we identified PLPP as a reaction intermediate by conducting a GGPP-based LOS assay in the absence of NADPH, followed by acid phosphatase treatment to convert PLPP to prelycopaoctaene alcohol (PLOH). Analysis by GC–MS showed the presence of PLOH (Fig. 5d) with a fragmentation pattern consistent with previous reports of PLOH (Supplementary Fig. 21)<sup>32</sup>. Taken together, these results are consistent with a two-step LOS reaction to produce lycopaoctaene. By analogy to SS, LOS catalyses the condensation of two GGPP units in the first half reaction to form the cyclopropylcarbinyl diphosphate intermediate PLPP, with concomitant release of one molecule of inorganic pyrophosphate (Fig. 5a). In the second half reaction, the PLPP cyclopropyl ring is cleaved and rearranged to form a 1-1' linkage and further reduction by NADPH forms lycopaoctaene (Fig. 5a).

## Discussion

Our studies identified the first committed step in lycopadiene biosynthesis in race L of *B. braunii* and describe a new SSL enzyme, LOS, which carries out this reaction. Importantly, this enzyme may be used to engineer the effective production of hydrocarbon biofuel feedstocks in other photosynthetic or heterotrophic organisms in the future. This novel enzyme produces the tetraterpenoid lycopaoctaene from GGPP in a reaction analogous to that of SS. Lycopaoctaene would then undergo sequential reduction by a yet-to-be identified reductase enzyme(s) to form lycopahexaene, lycopapentaene, lycopetraene, lycopatriene and finally lycopadiene as the major hydrocarbon product of the pathway (Fig. 6). Furthermore, we show that the wild-type LOS enzyme displays remarkable substrate flexibility and can use at least three different, naturally occurring  $C_{15}$  and  $C_{20}$  prenyl diphosphate substrates *in vivo* and *in vitro* for the combinatorial biosynthesis of hydrocarbons

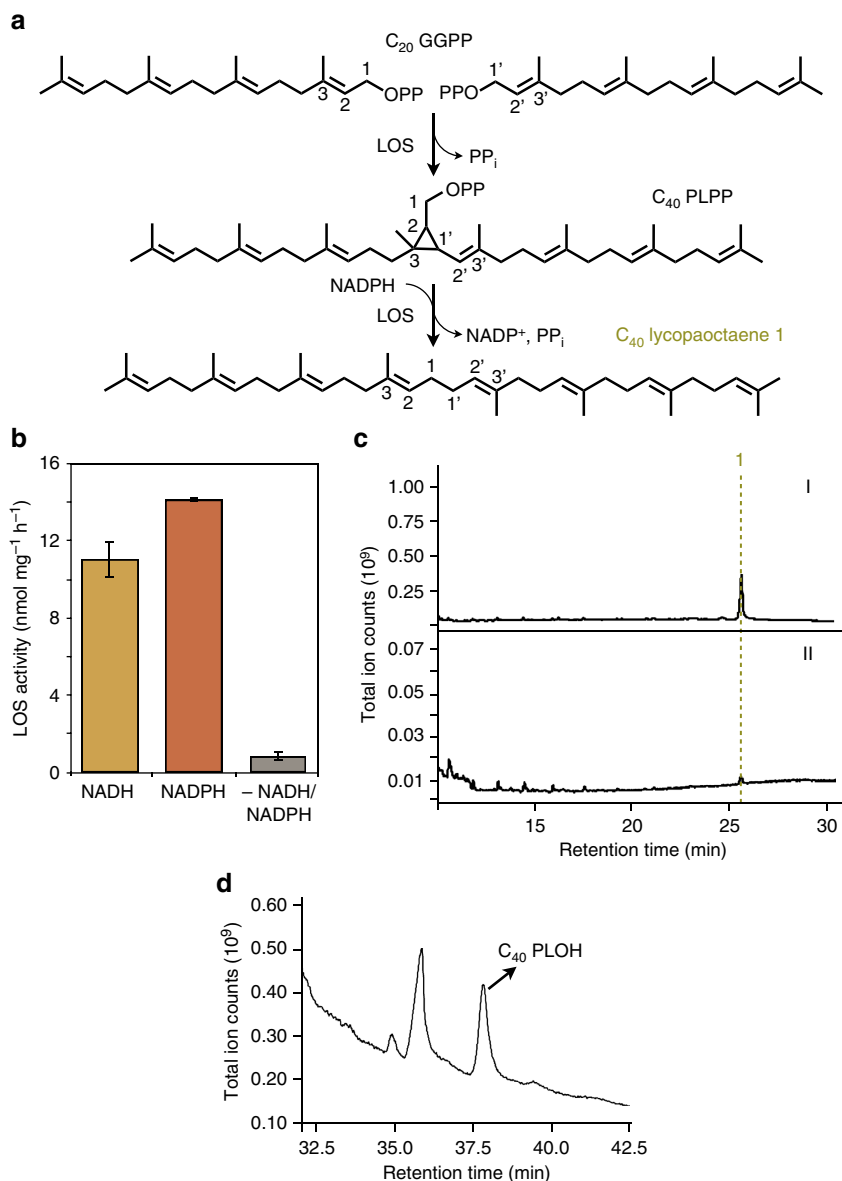


**Figure 4 | Characterization of substrate use for SS and SSL enzymes.** (a) Diagram of reactions showing substrates and products, and GC-MS profiles of LOS enzyme assay products with: I, FPP and GGPP substrates to produce squalene (1), C<sub>35</sub>H<sub>58</sub> (3) and lycopaoctaene (6); II, FPP and PPP substrates to make squalene (1), C<sub>35</sub>H<sub>64</sub> (2) and lycopadiene (4); and III, GGPP and PPP substrates to produce lycopadiene (4), lycopapentaene (5) and lycopaoctaene (6). Data shown are representatives from three independent experiments ( $n = 3$ ). (b) Radioactive enzyme assays using <sup>3</sup>H substrates GGPP, FPP and PPP with selected SS or SSL enzymes. SSL-1 produces PSPP from FPP substrate; however, PSPP was not analysed for in this assay. BSS, race B SS; LOS, lycopaoctaene synthase; LSS, race L SS; S/SS, SS from *S. lycopersicum* (tomato), and SSL-1, SSL-2, SSL-3 are SSL enzymes from the B race of *B. braunii*. Values shown are the mean  $\pm$  s.e. obtained from three independent measurements ( $n = 3$ ).

belonging to three terpene classes: triterpenes, sesquiterpenes and tetraterpenes.

When used under artificial reaction conditions, other SSL enzymes have been described to show relaxed substrate flexibility and to yield products of varying lengths. One such enzyme is CrtM, a dehydrosqualene synthase from *Staphylococcus aureus* that is involved in C<sub>30</sub> carotenoid biosynthesis<sup>41–43</sup>. In *S. aureus* and also when overexpressed in *E. coli*, CrtM uses two FPP

molecules to produce dehydrosqualene, a C<sub>30</sub> squalene-like molecule with a *cis* double bond at the 1-1' linkage of the two FPP molecules<sup>44</sup>. Expression of a GGPP synthase in *E. coli* also expressing CrtM led to the depletion of FPP and the overproduction of GGPP<sup>41</sup>. Under these conditions, CrtM was found to generate its native C<sub>30</sub> product, a C<sub>35</sub> homologue and C<sub>40</sub> phytoene, indicating the ability of CrtM to accept GGPP as a substrate when the natural FPP substrate is limiting<sup>41</sup>. CrtM



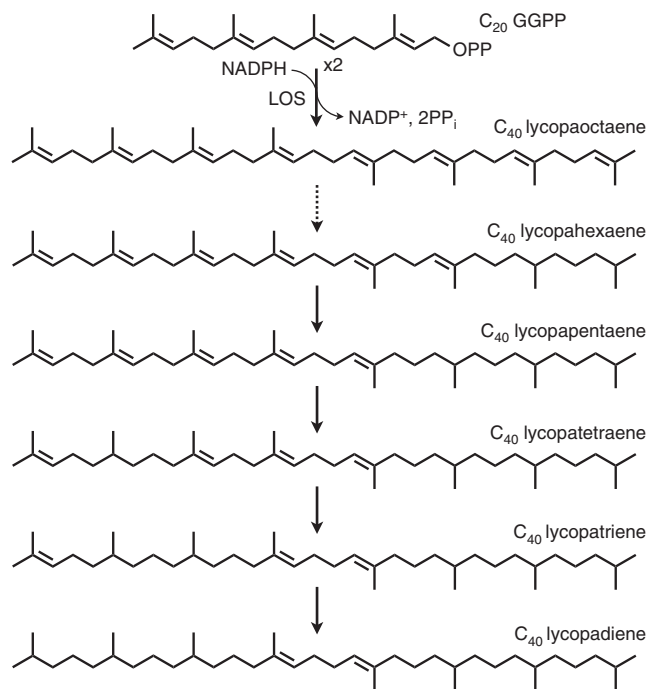
**Figure 5 | Characterization of the LOS reaction intermediate.** (a) Proposed LOS reaction mechanism to produce the PLPP intermediate and lycopaoctaene final product. (b) Radioactive LOS enzyme assays conducted with <sup>3</sup>H-GGPP substrate with or without NADH or NADPH reducing agents. Values shown are the mean  $\pm$  s.e. obtained from three independent measurements ( $n=3$ ). (c) GC-MS profiles of LOS enzyme assay with GGPP as substrate to produce lycopaoctaene (I): I, in the absence, or II, in the presence of 20 nM squalastatin. (d) GC-MS profile of products from LOS enzyme assay with GGPP substrate in the absence of NADPH to produce PLPP, which was hydrolysed by acid phosphatase to yield PLOH for GC-MS analysis. Data shown in c and d are representatives from three different experiments ( $n=3$ ).

was further engineered through mutations to accept  $C_{25}$  farnesylgeranyl diphosphate, yielding  $C_{40}$ ,  $C_{45}$  and  $C_{50}$  homologues of dehydrosqualene in the presence of FPP, farnesylgeranyl diphosphate and GGPP<sup>42,43</sup>. In another study, SSs from yeast, pig and rat have been shown to accept non-natural derivatives of FPP, to generate several unnatural terpene products *in vitro*<sup>45,46</sup>. In addition, yeast SS has been shown to yield alternative products *in vitro* from FPP as a substrate under various non-physiological reaction conditions, such as in the absence of NADPH with extended incubation times or in the presence of an unreactive NADPH analogue<sup>47–49</sup>. Irrespective of these examples, the promiscuity of LOS is still remarkable, considering that the wild-type LOS enzyme is able to accept three naturally occurring prenyl diphosphate substrates and produce the chimeric hydrocarbons described here without

the need for mutations or altered, non-physiological reaction conditions. Such remarkable intrinsic substrate promiscuity has not been documented for any other eukaryotic SS or SSL enzyme up until now.

Ours and other studies<sup>37,50</sup> raise the question about how SSL enzymes arise. LSS and LOS from *B. braunii* share >50% sequence identity and both contain the important catalytic residues and conserved domains of SS enzymes (Supplementary Fig. 18), but they still have different substrate preferences and product formation. Thus, we hypothesize that an ancient SS gene may have undergone gene duplication in race L. The paralogue that gave rise to LSS maintained the ability to produce squalene as its sole product to support sterol biosynthesis for primary metabolism. However, the paralogue that yielded LOS developed the use of alternative substrates for hydrocarbon oil





**Figure 6 | Proposed pathway for lycopadiene biosynthesis.** LOS uses GGPP in a two-step reaction to produce lycopaoctaene, which is further reduced by a yet unknown enzyme(s) to produce lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene and finally lycopadiene.

production, while also preserving SS activity. A similar scenario has been suggested as a possible mechanism for generating new secondary metabolites that may provide a fitness benefit to the host organism<sup>42,51</sup>. For *B. braunii*, it has been proposed that hydrocarbon oils and their derivatives increase the buoyancy of colonies for increased exposure to sunlight<sup>50,52,53</sup>. Similar gene diversifications have also been proposed in the B race of *B. braunii* for botryococcene production, and in some bacteria for squalene generation. For botryococcene production, SS duplications gave rise to the *SSL-1* and *SSL-3* genes/enzymes responsible for botryococcene production, while retaining a residual ability to produce squalene<sup>37</sup>. For bacteria, a three-enzyme system for squalene production arose from successive gene duplication events<sup>50</sup>.

LOS produces a variety of terpenes *in vitro* when offered a mixture of substrates and may likely also do so *in vivo*. In the *in vitro* mixed substrate assays with LOS (Fig. 4a), a  $C_{35}H_{64}$  molecule was observed from FPP and PPP condensation. This product has the exact same GC retention time and mass spectrum as the  $C_{35}H_{64}$  hydrocarbon detected in the total hydrocarbon pool (Fig. 2a (VII) and Fig. 4a (II)). This suggests that LOS is also promiscuous in *B. braunii* cells, but at best has limited access to the FPP and PPP pools, as the  $C_{35}H_{64}$  product only comprises 0.15% of the total hydrocarbons (Fig. 2a (VII)). The FPP–GGPP condensation product  $C_{35}H_{58}$  seen in these *in vitro* assays (Fig. 4a (I)) is not detected in the total hydrocarbon pool, possibly because this molecule is reduced to  $C_{35}H_{64}$  *in vivo*. It should be noted that two previous studies described several  $C_{35}$  squalene-like isoprenoids from *Methanococcus jannaschii* and *Thermococcus barophilus*, including the  $C_{35}H_{58}$  and  $C_{35}H_{64}$  molecules identified here<sup>54,55</sup>, although the enzymatic basis for the biosynthesis of these products remains uncharacterized. In addition, the ability of LOS to produce lycopapentaene as an observed *in vitro* condensation product of GGPP and PPP (Fig. 4a (III)) may indicate the existence of an alternative *in vivo*

biosynthetic route to this molecule and may be partially responsible for the relative prevalence of lycopapentaene (3% of total hydrocarbons; Fig. 2a (IV)) among the minor hydrocarbons of race L. This would suggest that LOS can also use PPP *in vivo*. However, LOS does not appear to use PPP as a single substrate for lycopadiene production *in vivo*, as we were not able to detect lycopadiene synthesis activity in cell fractions using PPP as substrate (Fig. 2b).

The promiscuous LOS described in this work is nevertheless used by *B. braunii* Race L to produce lycopaoctaene preferentially over squalene, as shown by the large amounts of hydrocarbon oils produced by this alga and the absence of detectable amounts of squalene or non-sterol squalene derivatives in the total hydrocarbon pool. In contrast, accumulation of the non-sterol squalene derivatives methylsqualenes in race B are thought to arise from squalene produced by the additional enzyme SSL-2 (ref. 37). Thus, detection of non-sterol squalene derivatives in race L would be expected if LOS was producing squalene *in vivo*.

We hypothesize that the observed preference of LOS for  $C_{40}$  lycopaoctaene biosynthesis over  $C_{30}$  squalene production *in vivo* may result from an increased flux from GGPP over FPP due to metabolic channelling, to form a biosynthetic metabolon specific for lycopadiene production. Formation of such metabolons is common in plants and is used for channelling substrates and intermediates required for the biosynthesis of different classes of plant natural products, including isoprenoids<sup>56–58</sup>. When expressed in yeast, LOS may produce squalene presumably due to a lack of this proposed metabolon, thus giving LOS access to the FPP pool. We therefore conclude that LOS is responsible for synthesizing lycopaoctaene as the first committed step towards lycopadiene hydrocarbon biosynthesis in race L of *B. braunii*.

## Methods

**Reagents.** [ $1\text{-}^3\text{H}$ ]-FPP (specific activity,  $18.2\text{ Ci mmol}^{-1}$ ) was purchased from PerkinElmer. [ $1\text{-}^3\text{H}$ ]-GGPP (specific activity,  $20.0\text{ Ci mmol}^{-1}$ ), [ $1\text{-}^3\text{H}$ ]-PPP (specific activity,  $20.0\text{ Ci mmol}^{-1}$ ) and non-radiolabelled PPP were purchased from American Radiolabeled Chemicals. When required, specific activities of radiolabelled chemicals were adjusted with non-labelled chemicals. Solvents and chemicals were purchased from VWR. All other reagents were purchased from Sigma unless otherwise noted.

**Culturing of *B. braunii*.** *B. braunii*, race L, Songkla Nakarin strain<sup>16</sup> was obtained from Algalbank-Caen Microalgal Culture Collection, University of Caen Basse-Normandie, France, and grown in a modified Chu 13 medium, pH 7.5 (ref. 59) at  $22\text{ }^{\circ}\text{C}$  under continuous aeration of filter-sterilized air enriched with 2.5%  $\text{CO}_2$ . The concentrations of chemicals in modified Chu 13 medium were as follows:  $\text{KNO}_3$  ( $0.4\text{ g l}^{-1}$ ),  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  ( $0.1\text{ g l}^{-1}$ ),  $\text{K}_2\text{HPO}_4$  ( $0.052\text{ g l}^{-1}$ ),  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  ( $0.054\text{ g l}^{-1}$ ),  $\text{FeNa EDTA}$  ( $0.01\text{ g l}^{-1}$ ),  $\text{H}_3\text{BO}_4$  ( $2.86\text{ mg l}^{-1}$ ),  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$  ( $1.54\text{ mg l}^{-1}$ ),  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  ( $0.22\text{ mg l}^{-1}$ ),  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  ( $0.08\text{ mg l}^{-1}$ ),  $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$  ( $0.06\text{ mg l}^{-1}$ ) and  $\text{CoSO}_4 \cdot 7\text{H}_2\text{O}$  ( $0.09\text{ mg l}^{-1}$ ). The cultures were grown under a 12:12 h light:dark cycle with a light intensity of  $120\text{ }\mu\text{E m}^{-2}\text{ s}^{-1}$ . Algal cells were subcultured by inoculation of 100 ml of a floating 6-week-old culture to 750 ml of fresh medium.

**Hydrocarbon staining and microscopy.** Nile Red (Sigma) staining of L race hydrocarbons was done by treating 500  $\mu\text{l}$  of *B. braunii* L race colonies in medium at early phase (5 day) density with 2.5  $\mu\text{l}$  of a stock solution of Nile Red dissolved in acetone ( $0.15\text{ mg ml}^{-1}$ ) for a final concentration of Nile red of  $0.375\text{ g ml}^{-1}$  and acetone of 0.25%. Samples were kept in the dark and incubated at room temperature for 15 min before imaging. Confocal microscopy images were obtained using an Olympus IX83-FV1200 inverted confocal microscope with spectral detection unit using an Olympus UPLSAPO  $\times 60$  oil-immersion objective (numerical aperture = 1.35). The confocal aperture diameter was set to 1 Airy disk unit and Z-series images were acquired with voxel dimensions of  $101 \times 101 \times 490$  (nm; x-y-z, respectively). Nile Red and chlorophyll were excited using a 488-nm laser and detected using 540–590 nm and 655–755 nm barrier filter settings, respectively. Bright-field images were acquired via the transmitted detector using the 488-nm laser. Maximum intensity projections and brightness/contrast adjustments were completed using the Olympus FV-ASW software version 4.

**Hydrocarbon purification.** Hydrocarbons were purified as described previously<sup>60,61</sup> and as follows. Algal cells were harvested by vacuum filtration using a 10- $\mu$ m nylon-mesh filter, freeze-dried, extracted twice with *n*-hexane for 2 h, to recover extracellular hydrocarbons, and then twice with CHCl<sub>3</sub>:MeOH (2:1) for 12 h stirring at room temperature, to recover intracellular hydrocarbons. Both extracts were concentrated separately using a rotary evaporator and resuspended in a small amount of *n*-hexane before running through separate gravity-fed silica gel columns with *n*-hexane as solvent. The *n*-hexane eluent before the pigment front was collected as the hydrocarbon fraction. Both extracellular and intracellular hydrocarbon fractions were combined and evaporated to dryness, to recover the total hydrocarbon pool of the algae. Individual hydrocarbon molecules were purified by injecting the total hydrocarbon sample onto an HPLC Develosil 60 silica column (20 mm  $\times$  250 mm) using *n*-hexane as a mobile phase at a flow rate of 6 ml min<sup>-1</sup> and detection at 210 nm. An aliquot of the purified hydrocarbons were analysed by GC-MS for purity analysis.

**GC-MS conditions.** Enzyme assay products, purified hydrocarbons and yeast extracts were analysed by GC-MS (Bruker 436-GC-SCION SQ Premium) using a 5% Phenyl BR-5 ms capillary column (30 m  $\times$  0.25 mm, film thickness: 0.25  $\mu$ m) in electron ionization (70 eV) mode. Initial oven temperature was 220 °C, held for 1 min, then increased to 280 °C at the rate of 5 °C min<sup>-1</sup> and then ramped to 300 °C at a rate of 2 °C min<sup>-1</sup> and held for 20 min. Helium was used as a carrier gas at a flow rate of 2.58 ml min<sup>-1</sup>. Temperatures of injection port, interface and ion source were 280 °C, 250 °C and 200 °C, respectively.

**Hydrocarbon NMR analysis.** One-dimensional NMR spectra of race L hydrocarbons molecules suffer from severe spectral overlap and the data interpretation is further complicated by degenerate chemical shifts as a result of symmetry planes in the lycopadiene structure. Consequently, hetero-nuclear two-dimensional (2D) experiments were used for unambiguous structural characterization of each entity. NMR experiments were performed at 25 °C on purified samples dissolved in deuterated chloroform using a 500 MHz Bruker Avance III HD spectrometer equipped with an inverse detection TXI probe. We have relied on 2D heteronuclear experiments acquired at natural abundance of <sup>13</sup>C nuclei, to deduce structures of these molecules. Multiplicity-edited 2D heteronuclear single quantum coherence spectrum was acquired for each compound and the resonances were assigned using a combination of 2D heteronuclear 2 bond correlation, 2D heteronuclear multiple bond correlation and 2D heteronuclear single quantum coherence-total correlation spectroscopy experiments. The <sup>13</sup>C chemical shifts obtained for lycopadiene match those previously obtained<sup>16</sup>.

**Spectral data for hydrocarbons analysed.** Lycopadiene: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  0.87, 0.89, 1.08, 1.09, 1.16, 1.2, 1.29, 1.39, 1.4, 1.54, 1.61, 1.96, 2.04, 2.17, 5.15; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta$  15.9, 19.7, 22.7, 24.5, 24.7, 25.4, 28.0, 28.3, 31.2, 32.8, 36.8, 37.4, 39.4, 40.0, 124.0, 135.9; MS (*m/z*): [M<sup>+</sup>] calcd. for C<sub>40</sub>H<sub>78</sub>, 558.61; found, 558.70. Lycopatriene: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  0.87, 0.89, 1.08, 1.09, 1.13, 1.16, 1.2, 1.29, 1.29, 1.39, 1.4, 1.54, 1.61, 1.63, 1.71, 1.96, 1.98, 2.04, 2.16, 5.12, 5.15; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta$  15.9, 17.6, 19.7, 22.7, 24.5, 24.7, 25.4, 25.6, 25.7, 28.0, 28.3, 30.8, 32.7, 36.8, 37.1, 37.4, 39.4, 40.0, 124.1, 125.1, 131.1, 135.9; MS (*m/z*): [M<sup>+</sup>] calcd. for C<sub>40</sub>H<sub>76</sub>, 556.59; found, 556.70. Lycopetraene: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  0.87, 0.89, 1.08, 1.13, 1.16, 1.2, 1.29, 1.39, 1.4, 1.54, 1.61, 1.63, 1.71, 1.95, 1.96, 1.99, 2.04, 2.16, 5.12, 5.15; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta$  15.9, 17.6, 19.7, 22.7, 24.5, 24.7, 25.4, 25.5, 25.7, 28.0, 28.3, 30.8, 32.7, 36.6, 37.1, 37.4, 39.3, 39.4, 40.0, 124.1, 125.2, 131.1, 135.9; MS (*m/z*): [M<sup>+</sup>] calcd. for C<sub>40</sub>H<sub>74</sub>, 554.58; found, 554.60. Lycopapentaene: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  0.87, 0.89, 1.08, 1.16, 1.2, 1.29, 1.39, 1.4, 1.55, 1.62, 1.63, 1.71, 2, 2.04, 2.09, 2.23, 5.15; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta$  16.0, 17.7, 19.8, 22.7, 24.5, 24.7, 25.4, 25.7, 26.7, 28.0, 28.3, 32.7, 36.6, 37.1, 37.4, 39.4, 39.8, 124.2, 124.8, 131.1, 135.9; MS (*m/z*): [M<sup>+</sup>] calcd. for C<sub>40</sub>H<sub>72</sub>, 552.56; found, 552.90. Lycopahexaene: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  0.87, 0.89, 1.08, 1.16, 1.2, 1.29, 1.4, 1.55, 1.61, 1.63, 1.71, 1.96, 1.99, 2.04, 2.06, 5.15; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta$  16.0, 17.7, 19.8, 22.7, 24.5, 24.7, 25.7, 26.8, 28.0, 28.3, 32.8, 36.6, 37.4, 39.4, 39.8, 124.3, 131.1, 135.9; MS (*m/z*): [M<sup>+</sup>] calcd. for C<sub>40</sub>H<sub>70</sub>, 550.55; found, 550.60. Lycopaoctaene: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  1.62, 1.63, 1.71, 2, 2.03, 2.04, 5.15; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta$  16.0, 17.7, 25.7, 27.2, 28.3, 39.8, 124.3, 131.1, 135.9; MS (*m/z*): [M<sup>+</sup>] calcd. for C<sub>40</sub>H<sub>66</sub>, 546.52; found, 546.70. C<sub>35</sub>H<sub>64</sub>: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  0.83, 0.86, 1.06, 1.13, 1.19, 1.21, 1.24, 1.32, 1.36, 1.52, 1.61, 1.68, 1.98, 2.01, 2.06, 5.13; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta$  16.3, 17.9, 20.0, 22.9, 24.4, 24.8, 25.1, 26.0, 26.3, 27.2, 28.2, 33.0, 37.7, 39.6, 40.0, 124.8, 131.4, 135.1; MS (*m/z*): [M<sup>+</sup>] calcd. for C<sub>35</sub>H<sub>64</sub>, 484.50; found, 484.70.

**Ozonolysis experiments.** Owing to the limited amount of minor hydrocarbons in the total L race hydrocarbon pool, ozonolysis experiments were conducted only on those hydrocarbons that could be purified to obtain at least 10 mg: lycopadiene,

lycopatriene and lycopapentaene. These hydrocarbons were separately dissolved in dichloromethane and submitted to ozone cleavage for 5 min at -78 °C. Each product from reductive cleavage of the resulting ozonide was directly subjected to GC-MS (electron ionization) analyses without purification. GC-MS analysis was carried out using a GCMS-QP2010 Ultra spectrometer (Shimadzu, Kyoto, Japan) equipped with a capillary column (InertCap 1MS, GL Science; 60 m  $\times$  0.25 mm, film thickness: 0.25  $\mu$ m). The column temperature was programmed as follows: 50 °C for 1 min, raised at 10 °C min<sup>-1</sup> from 50 to 220 °C, then at 2 °C min<sup>-1</sup> from 220 to 260 °C and held for 22 min at the final temperature. Helium was used as a carrier gas at a flow rate of 41.2 cm s<sup>-1</sup>. Temperatures of injection port, interface and ion source were 260, 250 and 200 °C, respectively.

**Preparation of algal cell lysate.** Algae from the rapid growth phase were harvested, snap frozen with liquid nitrogen and stored at -80 °C for future use. In a typical preparation, 200 mg of frozen algae was added to eight individual tubes containing 0.8 ml extraction buffer (50 mM MOPS pH 6.8, 20 mM MgCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol, 5 mM EGTA and 20% (v/v) glycerol) with 200  $\mu$ l of stainless steel metal beads (mixture of 0.9–2 mm diameter beads) and then homogenized using a Bullet Blender Storm 24 (Next Advance, Inc., USA) for 10 min at a speed of 10. The algal homogenates were centrifuged at 9,000 g for 10 min at 4 °C. The 9,000 g supernatants were pooled and then centrifuged at 100,000 g for 1 h to obtain the soluble fraction. The pellet was washed with 4 ml extraction buffer and centrifuged again at 100,000 g. This process was repeated twice and the pellet was resuspended in 0.5 ml extraction buffer to obtain a microsomal fraction. An aliquot of each protein fraction was used for enzyme assays.

**Enzyme assays.** Radioactive enzyme assays were conducted in a 50- $\mu$ l total reaction volume as described previously<sup>23,37</sup> and as follows. Reactions were initiated by adding 10  $\mu$ g algal lysate or 1  $\mu$ g of recombinant enzyme purified from *E. coli* to a reaction buffer containing 50 mM MOPS pH 6.8, 2.5 mM  $\beta$ -mercaptoethanol, 20 mM MgCl<sub>2</sub>, 2 mM NADPH or NADH and 10  $\mu$ M [<sup>3</sup>H]-prenyl-PP substrate (specific activity = 0.25 Ci mmol<sup>-1</sup>). Enzyme assays were incubated at 37 °C for 60 min and terminated by adding 60  $\mu$ l of *n*-hexane, followed by brief vortexing and centrifugation. Thirty microlitres of the organic layer was spotted onto silica gel 60 thin-layer chromatography plates along with authentic standards of squalene (*R<sub>f</sub>* = 0.17), C<sub>30</sub>-botryococcene (*R<sub>f</sub>* = 0.2), lycopaoctaene (*R<sub>f</sub>* = 0.09) and lycopadiene (*R<sub>f</sub>* = 0.5), and then developed with *n*-hexane. Hydrocarbon standards were visualized by iodine vapour and spots corresponding to authentic standards were scraped and analysed by liquid scintillation counter, to determine radioactive incorporation into the respective hydrocarbons. SS and botryococcene synthase assays were conducted using <sup>3</sup>H-FPP, whereas LOS and lycopadiene synthase assays were conducted using <sup>3</sup>H-GGPP and <sup>3</sup>H-PPP, respectively. For GC-MS analysis of reaction products, enzyme assays were conducted with non-radiolabelled substrates in glass vials by scaling the total reaction volume and contents to 500  $\mu$ l. Two separate 500  $\mu$ l reactions were each extracted twice with 1 ml *n*-hexane, the extracts pooled, dried under a stream of nitrogen gas, resuspended in 100  $\mu$ l *n*-hexane and a 5- $\mu$ l aliquot analysed by GC-MS.

**Hydrocarbon accumulation and LOS activity over growth cycle.** Several flasks of race L culture were grown over a 39-day culture period. Twenty-five millilitres of culture was collected at indicated times on pre-weighed Whatman GF/C filters by vacuum filtration. The filters with algae were dried in an 80 °C oven and total hydrocarbons were extracted using the protocol described above. Similarly, 500 ml of algae cells were also harvested at indicated times by filtering through a 10- $\mu$ m nylon mesh and the samples used to determine LOS enzyme activity over the culture period. LOS enzyme assays were conducted using 10  $\mu$ g of 9,000 g supernatant.

**RNA isolation.** Total RNA was isolated as previously described<sup>62</sup> and as follows. For *B. braunii*, liquid nitrogen frozen samples were pulverized using a TissueLyser II (Qiagen, Valencia, CA). Approximately 200 mg of this frozen pulverized *B. braunii* tissue, or frozen leaf tissues of *A. thaliana* and *S. lycopersicum*, were ground in a mortar and pestle with liquid nitrogen. The samples were then added to 1 ml of TRIzol (Life Technologies, Grand Island, NY) and the total RNA was isolated following the manufacturer's instructions. The RNA from *A. thaliana* and *S. lycopersicum* were used directly for gene cloning (see below). For the *B. braunii* RNA, contaminating polysaccharides were removed before the use for gene cloning as follows. The RNA pellet was resuspended in 0.5 ml 2 M LiCl, incubated for 5 min and centrifuged at 12,000 g for 15 min at 4 °C to pellet the total RNA. The polysaccharides remain in the supernatant. This process was repeated until the size of RNA pellet remained constant. The RNA pellet was then dissolved in 0.5 ml 1  $\times$  Tris-EDTA (TE), extracted with an equal volume of phenol/chloroform/isoamyl alcohol mix (25:4:1) and centrifuged at 12,000 g for 15 min at 4 °C. The aqueous supernatant phase was removed and extracted with 0.5 ml of chloroform followed by centrifugation at 12,000 g for 15 min at 4 °C. The RNA was precipitated from the aqueous supernatant by adding 0.1 volume 3 M sodium acetate and 2.5 volumes of 100% ethanol, incubated for 20 min at -20 °C and centrifuged at 12,000 g for 15 min at 4 °C. The RNA pellet was washed twice with 0.5 ml 70%

ethanol, centrifuged at 12,000 g for 15 min at 4 °C, the pellet dried using a speedvac and the RNA resuspended in 50 µl of RNAase-free water.

**LSS and LOS cloning.** RNAseq analysis on the L race of *B. braunii* was carried out as previously reported<sup>62</sup> and as follows. RNA isolated from days 0, 3, 7, 14, 21 and 28 over the 4-week culture cycle were kept separate, paired-end libraries were prepared from each RNA sample and each sample was sequenced under the Illumina platform. A transcriptome contig library was created using the Trinity software suite. SSL sequences were computationally screened using this transcriptomic database with BSS as a query. Two SSL cDNAs, LSS and LOS, were identified. Total RNA was extracted from a day-3 race L culture and first-strand cDNA was prepared using the SuperScript III first-strand synthesis kit (Invitrogen). Primers specific to each cDNA based on the transcriptome sequence were used to amplify the PCR product from first-strand cDNA using GoTaq DNA polymerase mix (Promega) followed by cloning into the pGEM-T vector (Promega). Gene-specific primers were as follows: For LSS, forward primer 5'-ATGGGAAGCTACAGGAGGTTTGAAGC-3' and reverse primer 5'-TCAGGC AAGGCCCGCGAAG-3'; for LOS, forward primer 5'-ATGAAGTACACAGAT TTCTTGCGC-3' and reverse primer 5'-TCACACAGTCTTGAGGGCGAG-3'.

**AtGGPPS11 and SISS cloning.** The cDNA sequences for *AtGGPPS11* (At4g36810)<sup>30</sup> and SISS were obtained from the NCBI nucleotide database. Total RNA isolated from the leaves of *A. thaliana* and *S. lycopersicum* were used for first-strand cDNA synthesis and PCR products corresponding to each cDNA were amplified using gene-specific primers followed by cloning into the pGEM-T vector. Gene-specific primers used for PCR amplification were as follows: for *AtGGPPS11*, forward primer 5'-ATGGCTCAGTGTCTAGGTTTC-3' and reverse primer 5'-TCAGTCTGTCTATAGGCAATG-3'; for SISS, forward primer 5'-ATGGGAA CATTGAGGGCA-3' and reverse primer 5'-CTAAGACCGCTGCCAAAAAGT TG-3'.

**Protein expression and purification.** Except for SSL-1 and SSL-3, DNA sequences encoding the predicted transmembrane domain at the carboxy terminus of each SS or SSL protein were deleted and then cloned to pET28a, to encode for amino terminal 6 × His-tagged proteins. DNA templates of BSS, SSL-1, SSL-2 and SSL-3 were received from previous studies<sup>25,37</sup>. The pET28a expression constructs were made using appropriate restriction sites; *LOS*<sup>4391-444</sup> with NheI and HindIII; *LSS*<sup>4399-462</sup>, *SISS*<sup>4387-411</sup>, *BSS*<sup>4399-461</sup>, *SSL-1* and *SSL-3* with NdeI and HindIII; and *SSL-2*<sup>4392-463</sup> with NdeI and SalI. The expression constructs were transformed into *E. coli* BL21(DE3), grown at 37 °C to OD<sub>600</sub> = 0.8 and protein expression for each gene induced by adding 1 mM isopropylthio-β-D-galactoside. The induced cultures were then grown for an additional 6 h at 25 °C.

His-tagged proteins were purified at 4 °C by standard procedures. In a typical purification, pellets from a 100-ml culture were resuspended in 10 ml extraction buffer (50 mM sodium phosphate buffer pH 7.8, 300 mM NaCl, 10 mM imidazole, 1 × general protease inhibitor cocktail (Sigma), 1 mM MgCl<sub>2</sub> and 1% glycerol (v/v)), and then sonicated four times for 15 s at 70% maximum power with 2 min interval between each sonication. The sonicated samples were centrifuged at 16,000 g for 10 min at 4 °C. The supernatants were applied to a gravity-fed column containing Ni-NTA agarose (Qiagen) and the His-tagged proteins were purified according to the manufacturer's recommendations. The wash buffer (50 mM sodium phosphate buffer pH 7.8, 300 mM NaCl, 60 mM imidazole, 1 mM MgCl<sub>2</sub> and 1% glycerol (v/v)) and the elution buffer (50 mM sodium phosphate buffer pH 7.8, 300 mM NaCl, 400 mM imidazole, 1 mM MgCl<sub>2</sub> and 1% glycerol (v/v)) were used for protein purification. The eluted fractions were dialysed using storage buffer (300 mM NaCl, 20 mM Tris-HCl pH 7.5, 5 mM dithiothreitol and 2 mM MgCl<sub>2</sub>), concentrated with an Amicon Ultra centrifugal filter (0.5 ml, 30 kDa cutoff; EMD Millipore) to desired protein concentration, an equal amount of 100% glycerol added and stored at -20 °C for 1–2 months without a loss of enzyme activity.

**LOS steady-state kinetic experiments.** Michaelis–Menten enzyme kinetics experiments with the LOS enzyme were set up using the protocol described previously<sup>63</sup> and as follows. Pilot experiments were initially conducted using radioactive enzyme assays, to determine the reaction conditions where reaction velocity is linear (<10% turnover) with respect to enzyme concentration. For steady-state kinetics experiments, enzyme assays (50 µl total reaction volume) were conducted with 100 nM of purified LOS enzyme at the indicated concentration of <sup>3</sup>H-GGPP, <sup>3</sup>H-FPP or <sup>3</sup>H-PPP for a set time interval. Kinetics for the NADPH cofactor were not determined and this factor was held constant at 2 mM in all assays. The reaction velocities were plotted against substrate concentrations to generate Michaelis–Menten curves and the kinetics parameters of LOS enzyme for GGPP, FPP and PPP were determined by analysing the data using GraphPad Prism 6 software.

**Yeast expression.** The DNA sequence encoding the 56 N-terminal amino acids for the plastid-targeting signal of *AtGGPPS11* were deleted resulting in *AtGGPPS11*<sup>Δ57</sup>, which was cloned into pESC-TRP using BamHI and SalI restriction sites for

expression under the inducible GAL1 promoter. For coexpression studies, the DNA sequences encoding the predicted C-terminal transmembrane domain of *LOS* was deleted resulting in *LOS*<sup>Δ391-444</sup>, which was cloned into the second multiple cloning site of *AtGGPPS11*<sup>Δ57</sup>:pESC-TRP using EcoRI and SpeI restriction sites for expression under the inducible GAL10 promoter. The yeast expression constructs *AtGGPPS11*<sup>Δ57</sup>:pESC-TRP and *AtGGPPS11*<sup>Δ57</sup> + *LOS*<sup>Δ391-444</sup>:pESC-TRP were introduced into yeast strain CKY457 (*MATa*, *leu2Δ1*, *ura3-52*, *trp1Δ63*, *his3Δ200* and *lys2-128Δ*) via lithium acetate transformation followed by selection on yeast synthetic dropout medium (SC-TRP). Positive transformants were grown at 30 °C in 150 ml SC-TRP media to mid-log phase, induced with 2% final galactose concentration and grown for an additional 130 h. Yeast cells were harvested, freeze-dried and extracted with *n*-hexane for 2 h by stirring at room temperature. The organic extracts were centrifuged at 1,000 g, the supernatant dried using a rotary evaporator, resuspended in 500 µl *n*-hexane and a 5-µl aliquot analysed by GC–MS to evaluate for the production of lycopaoctane.

**Purification of lycopaoctane standard.** A 3-l culture of yeast strain CKY457 expressing *AtGGPPS11*<sup>Δ57</sup> + *LOS*<sup>Δ391-444</sup>:pESC-TRP was grown and the metabolites were extracted with *n*-hexane using the protocol described above. The organic extracts were applied to a silica gel gravity-fed column and metabolites eluted sequentially with two different solvent systems, *n*-hexane and 1% diethyl ether in *n*-hexane. The 1% diethyl ether in *n*-hexane eluent fraction containing lycopaoctane and other metabolites was concentrated and resuspended in a small volume of *n*-hexane before injecting the samples onto a 10 mm × 250 mm Cosomil 5C<sub>18</sub>-AR-II HPLC column. Lycopaoctane was purified by reversed phase HPLC using methanol:acetone (60:40) as a mobile phase at a flow rate of 3 ml min<sup>-1</sup> and detection at 210 nm. Identity of lycopaoctane molecule was confirmed by GC–MS (Supplementary Fig. 3) and NMR (Supplementary Table 1 and Supplementary Fig. 15).

**LSS and LOS complementation of yeast SS knockout strain.** The yeast strain ZX 178-08 (*MATa*, *his3*, *leu2*, *met15*, *ura3*, *SUE* and *erg9Δ::HPH*), which has a knockout of the endogenous SS gene (*ERG9*), was used for complementation experiments<sup>64</sup>. The expression vector XURA used in our study was made by replacing the GAL1 and GAL10 promoters in pESC-URA vector backbone with constitutive promoters TEF1 and *GPD*, respectively<sup>65</sup>. Full-length LSS was cloned into XURA using NotI and SpeI restriction sites. Full-length LOS was cloned into EcoRI and SpeI sites of XURA vector. The expression constructs LSS:XURA, LOS:XURA and XURA were transformed into ZX 178-08 yeast strain via lithium acetate transformation followed by selection on yeast synthetic dropout medium supplemented with ergosterol synthetic complete-Uracil (SCE URA). The positive transformants were further characterized by streaking on selection media with or without ergosterol, to test the ability of individual genes to restore ergosterol prototrophy of ZX 178-08 yeast.

**GC–MS analysis of PLOH.** The enzyme assay for PLPP was conducted following the protocol described previously for analysis of PSP<sup>37,50</sup> and as follows. The PLPP assay contained 50 mM MOPS (pH 6.8), 2.5 mM β-mercaptoethanol, 20 mM MgCl<sub>2</sub>, 180 µM GGPP and 750 µg of purified LOS enzyme in a 2.5-ml total reaction volume. The reaction mixture was incubated at 37 °C for 2 h and terminated by snap freezing with liquid nitrogen followed by lyophilization. For hydrolysis of PLPP to PLOH, the white residue obtained after lyophilization was resuspended in 2 ml phosphatase solution (20% 1-propanol, 100 mM sodium acetate pH 4.7, 0.1% Triton-X-100 and 50 units of sweet potato acid phosphatase) and then incubated for 16 h in a 28 °C shaker. The dephosphorylation reaction mixture was extracted three times with 4 ml of methyl tertiary butyl ether (MTBE), the extracts pooled, dried under a stream of nitrogen gas, resuspended in 100 µl *n*-hexane and a 5-µl aliquot analysed by GC–MS as under the conditions described above with the following differences. Chemical ionization using methane gas at 20 psi was employed for analysis of the PLOH molecule, as electronic ionization did not result in a molecular ion of PLOH. Initial oven temperature was 70 °C, held for 1 min, then increased to 200 °C at the rate of 8 °C min<sup>-1</sup> and then ramped to 300 °C at a rate of 20 °C min<sup>-1</sup> and held for 20.75 min. Temperatures of injection port, interface and ion source were 260 °C, 250 °C and 200 °C, respectively.

## References

- Moody, J. W., McGinty, C. M. & Quinn, J. C. Global evaluation of biofuel potential from microalgae. *Proc. Natl Acad. Sci. USA* **111**, 8691–8696 (2014).
- Chisti, Y. Constraints to commercialization of algal fuels. *J. Biotechnol.* **167**, 201–214 (2013).
- Banerjee, A., Sharma, R., Chisti, Y. & Banerjee, U. C. *Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals. *Crit. Rev. Biotechnol.* **22**, 245–279 (2002).
- Yoshida, M., Tanabe, Y., Yonezawa, N. & Watanabe, M. M. Energy innovation potential of oleaginous microalgae. *Biofuels* **3**, 761–781 (2012).
- Lee, S. Y., Kim, H. M. & Cheon, S. Metabolic engineering for the production of hydrocarbon fuels. *Curr. Opin. Biotechnol.* **33**, 15–22 (2015).
- Metzger, P. & Largeau, C. *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. *Appl. Microbiol. Biotechnol.* **66**, 486–496 (2005).

7. Hillen, L. W., Pollard, G., Wake, L. V. & White, N. Hydrocracking of the oils of *Botryococcus braunii* to transport fuels. *Biotechnol. Bioeng.* **24**, 193–205 (1982).
8. Moldowan, J. M. & Seifert, W. K. First discovery of botryococcane in petroleum. *J. Chem. Soc. Chem. Commun.* **19**, 912–914 (1980).
9. Mckirdy, D. M., Cox, R. E., Volkman, J. K. & Howell, V. J. Botryococcane in a new class of Australian non-marine crude oils. *Nature* **320**, 57–59 (1986).
10. Lichtfouse, E., Derenne, S., Mariotti, A. & Largeau, C. Possible algal origin of long chain odd *n*-alkanes in immature sediments as revealed by distributions and carbon isotope ratios. *Org. Geochem.* **22**, 1023–1027 (1994).
11. Glikson, M., Lindsay, K. & Saxby, J. *Botryococcus*-A planktonic green alga, the source of petroleum through the ages: transmission electron microscopical studies of oil shales and petroleum source rocks. *Org. Geochem.* **14**, 14 (1989).
12. Adam, P., Schaeffer, P. & Albrecht, P. C40 monoaromatic lycopene derivatives as indicators of the contribution of the alga *Botryococcus braunii* race L to the organic matter of Messel oil shale (Eocene, Germany). *Org. Geochem.* **37**, 584–596 (2006).
13. Khan, N. E., Nybo, E. S., Chappell, J. & Curtis, W. R. Triterpene hydrocarbon production engineered into a metabolically versatile host—*Rhodobacter capsulatus*. *Biotechnol. Bioeng.* **112**, 1523–1532 (2015).
14. Rohmer, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **16**, 565–574 (1999).
15. Sato, Y., Ito, Y., Okada, S., Murakami, M. & Abe, H. Biosynthesis of the triterpenoids, botryococcenes and tetramethylsqualene in the B race of *Botryococcus braunii* via the non-mevalonate pathway. *Tetrahedron Lett.* **44**, 7035–7037 (2003).
16. Metzger, P. & Casadevall, E. Lycopadiene, a tetraterpenoid hydrocarbon from new strains of the green-alga *Botryococcus braunii*. *Tetrahedron Lett.* **28**, 3931–3934 (1987).
17. Robinson, G. W., Tsay, Y. H., Kienzle, B. K., Smith-Monroy, C. A. & Bishop, R. W. Conservation between human and fungal squalene synthetases: similarities in structure, function, and regulation. *Mol. Cell. Biol.* **13**, 12 (1993).
18. Rohmer, M. Hopanoids. *Handbook of Hydrocarbon and Lipid Microbiology* 133–142 (Springer, 2010).
19. Poulter, C. D. Biosynthesis of non-head-to-tail terpenes—formation of 1'-1 and 1'-3 linkages. *Acc. Chem. Res.* **23**, 70–77 (1990).
20. Metzger, P., Pouet, Y. & Summons, R. Chemotaxonomic evidence for the similarity between *Botryococcus braunii* L race and *Botryococcus neglectus*. *Phytochemistry* **44**, 1071–1075 (1997).
21. Zhang, Z., Metzger, P. & Sachs, J. P. Biomarker evidence for the co-occurrence of three races (A, B and L) of *Botryococcus braunii* in El Junco Lake, Galápagos. *Org. Geochem.* **38**, 20 (2007).
22. Sato, T. *et al.* Sesquiterpenes (C35 terpenes) biosynthesized via the cyclization of a linear C35 isoprenoid by a tetraprenyl-beta-curcumene synthase and a tetraprenyl-beta-curcumene cyclase: identification of a new terpene cyclase. *J. Am. Chem. Soc.* **133**, 9734–9737 (2011).
23. Okada, S., Devarenne, T. P., Murakami, M., Abe, H. & Chappell, J. Characterization of botryococcene synthase enzyme activity, a squalene synthase-like activity from the green microalga *Botryococcus braunii*, Race B. *Arch. Biochem. Biophys.* **422**, 110–118 (2004).
24. Stamellos, K. D. *et al.* Subcellular localization of squalene synthase in rat hepatic cells. Biochemical and immunochemical evidence. *J. Biol. Chem.* **268**, 12825–12836 (1993).
25. Okada, S., Devarenne, T. P. & Chappell, J. Molecular characterization of squalene synthase from the green microalga *Botryococcus braunii*, Race B. *Arch. Biochem. Biophys.* **373**, 307–317 (2000).
26. Gu, P., Ishii, Y., Spencer, T. A. & Shechter, I. Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase. *J. Biol. Chem.* **273**, 12515–12525 (1998).
27. Pandit, J. *et al.* Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis. *J. Biol. Chem.* **275**, 30610–30617 (2000).
28. Liu, C. I., Jeng, W. Y., Chang, W. J., Ko, T. P. & Wang, A. H. Binding modes of zaragozic acid A to human squalene synthase and staphylococcal dehydrosqualene synthase. *J. Biol. Chem.* **287**, 18750–18757 (2012).
29. Liu, C. I. *et al.* Structural insights into the catalytic mechanism of human squalene synthase. *Acta Crystallogr. D* **70**, 231–241 (2014).
30. Beck, G. *et al.* Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* **82**, 393–416 (2013).
31. Ignea, C. *et al.* Efficient diterpene production in yeast by engineering Erg20p into a geranylgeranyl diphosphate synthase. *Metab. Eng.* **27**, 65–75 (2015).
32. Qureshi, A. A., Barnes, F. J., Semmler, E. J. & Porter, J. W. Biosynthesis of prelycopersene pyrophosphate and lycopersene by squalene synthetase. *J. Biol. Chem.* **248**, 2755–2767 (1973).
33. Barnes, F. J., Qureshi, A. A., Semmler, E. J. & Porter, J. W. Prelycopersene pyrophosphate and lycopersene. Intermediates in carotene biosynthesis. *J. Biol. Chem.* **248**, 2768–2773 (1973).
34. Kushwaha, S. C. & Kates, M. Isolation and identification of 'bacteriorhodopsin' and minor C<sub>40</sub>-carotenoids in *Halobacterium cutirubrum*. *Biochim. Biophys. Acta* **316**, 235–243 (1973).
35. Gregonis, D. & Rilling, H. The stereochemistry of trans-phytoene synthesis. Some observations on lycopersene as a carotene precursor and a mechanism for the synthesis of *cis*- and *trans*-phytoene. *Biochemistry* **13**, 1538–1542 (1974).
36. Kushwaha, S. C., Kates, M. & Porter, J. W. Enzymatic synthesis of C<sub>40</sub> carotenes by cell-free preparation from *Halobacterium cutirubrum*. *Can. J. Biochem.* **54**, 816–823 (1976).
37. Niehaus, T. D. *et al.* Identification of unique mechanisms for triterpene biosynthesis in *Botryococcus braunii*. *Proc. Natl Acad. Sci. USA* **108**, 12260–12265 (2011).
38. Baxter, A. *et al.* Squalostatin 1, a potent inhibitor of squalene synthase, which lowers serum cholesterol *in vivo*. *J. Biol. Chem.* **267**, 11705–11708 (1992).
39. Bergstrom, J. D. *et al.* Zaragozic Acids—a family of fungal metabolites that are picomolar competitive inhibitors of squalene synthase. *Proc. Natl Acad. Sci. USA* **90**, 80–84 (1993).
40. Lindsey, S. & Harwood, H. J. Inhibition of mammalian squalene synthetase-activity by zaragozic acid is a result of competitive-inhibition followed by mechanism-based irreversible inactivation. *J. Biol. Chem.* **270**, 9083–9096 (1995).
41. Umeno, D. & Arnold, F. H. A C35 carotenoid biosynthetic pathway. *Appl. Environ. Microbiol.* **69**, 3573–3579 (2003).
42. Umeno, D. & Arnold, F. H. Evolution of a pathway to novel long-chain carotenoids. *J. Bacteriol.* **186**, 1531–1536 (2004).
43. Tobias, A. V. & Arnold, F. H. Biosynthesis of novel carotenoid families based on unnatural carbon backbones: a model for diversification of natural product pathways. *Biochim. Biophys. Acta* **1761**, 235–246 (2006).
44. Wieland, B. *et al.* Genetic and biochemical analyses of the biosynthesis of the yellow carotenoid 4,4'-diaponeurosporene of *Staphylococcus aureus*. *J. Bacteriol.* **176**, 7719–7726 (1994).
45. Ortiz de Montellano, P. R., Wei, J. S., Vinson, W. A., Castillo, R. & Boparai, A. S. Substrate selectivity of squalene synthetase. *Biochemistry* **16**, 2680–2685 (1977).
46. Koyama, T., Ogura, K. & Seto, S. Substrate specificity of squalene synthetase. *Biochim. Biophys. Acta* **617**, 218–224 (1980).
47. Jarstfer, M. B., Zhang, D. L. & Poulter, C. D. Recombinant squalene synthase. Synthesis of non-head-to-tail isoprenoids in the absence of NADPH. *J. Am. Chem. Soc.* **124**, 8834–8845 (2002).
48. Blagg, B. S., Jarstfer, M. B., Rogers, D. H. & Poulter, C. D. Recombinant squalene synthase. A mechanism for the rearrangement of presqualene diphosphate to squalene. *J. Am. Chem. Soc.* **124**, 8846–8853 (2002).
49. Zhang, D. L. & Poulter, C. D. Biosynthesis of non-head-to-tail isoprenoids—synthesis of 1'-1-structures and 1'-3-structures by recombinant yeast squalene synthase. *J. Am. Chem. Soc.* **117**, 1641–1642 (1995).
50. Pan, J. J. *et al.* Biosynthesis of squalene from farnesyl diphosphate in bacteria: three steps catalyzed by three enzymes. *ACS Cent. Sci.* **1**, 77–82 (2015).
51. Vining, L. C. Roles of secondary metabolites from microbes. *Ciba Found. Symp.* **171**, 184–194 (1992).
52. Weiss, T. L. *et al.* Colony organization in the green alga *Botryococcus braunii* (Race B) is specified by a complex extracellular matrix. *Eukaryot. Cell* **11**, 1424–1440 (2012).
53. Niehaus, T. D. *et al.* Functional identification of triterpene methyltransferases from *Botryococcus braunii* race B. *J. Biol. Chem.* **287**, 8163–8173 (2012).
54. Manquin, B. P., Morgan, J. A., Ju, J., Muller-Spath, T. & Clark, D. S. Production of C35 isoprenoids depends on H2 availability during cultivation of the hyperthermophile *Methanococcus jannaschii*. *Extremophiles* **8**, 13–21 (2004).
55. Cario, A., Grossi, V., Schaeffer, P. & Oger, P. M. Membrane homeoviscous adaptation in the piezo-hyperthermophilic archaeon *Thermococcus barophilus*. *Front. Microbiol.* **6**, 1152 (2015).
56. Jørgensen, K. *et al.* Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* **8**, 280–291 (2005).
57. Field, B. & Osbourn, A. E. Metabolic diversification—-independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543–547 (2008).
58. Nutzmann, H. W. & Osbourn, A. Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.* **26**, 91–99 (2014).
59. Chu, S. P. The influence of the mineral composition of the medium on the growth of planktonic algae: part I. Methods and culture media. *J. Ecol.* **30**, 284–325 (1942).
60. Metzger, P., Allard, B., Casadevall, E., Berkaloff, C. & Coute, A. Structure and chemistry of a new chemical race of *Botryococcus braunii* (Chlorophyceae) that produces lycopadiene, a tetraterpenoid hydrocarbon. *J. Phycol.* **26**, 258–266 (1990).

61. Weiss, T. L. *et al.* Raman spectroscopy analysis of botryococcene hydrocarbons from the green microalga *Botryococcus braunii*. *J. Biol. Chem.* **285**, 32458–32466 (2010).
62. Molnar, I. *et al.* Bio-crude transcriptomics: gene discovery and metabolic network reconstruction for the biosynthesis of the terpenome of the hydrocarbon oil-producing green alga, *Botryococcus braunii* race B (Shawa). *BMC Genomics* **13**, 576 (2012).
63. Garrett, S. R., Morris, R. J. & O'Maille, P. E. Steady-state kinetic characterization of sesquiterpene synthases by gas chromatography-mass spectroscopy. *Methods Enzymol.* **515**, 3–19 (2012).
64. Zhuang, X. & Chappell, J. Building terpene production platforms in yeast. *Biotechnol. Bioeng.* **112**, 1854–1864 (2015).
65. Zhuang, X. Engineering novel Terpene Production Platforms in the Yeast *Saccharomyces cerevisiae*. *Theses and Dissertations—Plant and Soil Sciences Paper* **17**, 1–189 (2013).

## Acknowledgements

This work was supported by grants NSF-EFRI-PSBR 1240478 and DOE-ARPA-E-PETRO DE-AR000203 to T.P.D.; by contract DE-EE0003046 awarded to the National Alliance for Advanced Biofuels and Bioproducts (NAABB) from the U.S. Department of Energy, of which T.P.D. and I.M. are members; and by a grant from the Japan Science and Technology (JST) agency, Core Research for Evolutionary Science and Technology (CREST) programme to S.O. We thank Joe Chappell, University of Kentucky, for yeast strain ZX 178-08 and XURA vector; Eric Bridenbaugh of Olympus Scientific Solutions Americas, for assistance with microscopy; and the following people at Texas A&M University: Craig Kaplan for yeast strain CKY457, Larry Dangott for HPLC use, Dan Browne for help with bioinformatics, and Indranil Malik and Ben Kaster for helpful comments.

## Author contributions

H.R.T., M.T.N., S.O. and T.P.D. conceived and designed the experiments. H.R.T., M.T.N., S.O., K.T. and I.M. performed experiments. H.R.T., M.T.N., S.O., I.M., Y.X. and T.P.D. analysed data. H.R.T., S.O., I.M. and T.P.D. co-wrote the paper.

## Additional information

**Accessions codes:** New DNA sequences described in this study have been deposited in GenBank: *LSS* accession #KT388100, *LOS* accession #KT388101 and *SISS* accession #NM\_001247787.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Thapa, H. R. *et al.* A squalene synthase-like enzyme initiates production of tetraterpenoid hydrocarbons in *Botryococcus braunii* Race L. *Nat. Commun.* **7**:11198 doi: 10.1038/ncomms11198 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>