New Phytologist

Research

Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*

Michele Fabris[1,2,3], Michiel Matthijs[1,2,3], Sophie Carbonelle[1,2], Tessa Moses[1,2], Jacob Pollier[1,2], Renaat Dasseville[3], Gino J. E. Baart[1,2,3], Wim Vyverman[3] and Alain Goossens[1,2]

[1]Department of Plant Systems Biology, VIB Technologiepark 927, B-9052 Gent, Belgium; [2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052, Gent, Belgium; [3]Department of Biology, Laboratory of Protistology and Aquatic Ecology, Ghent University, Krijgslaan 281 (S8), B-9000, Gent, Belgium

Author for correspondence:
Alain Goossens
Tel: +32 9 3313851
Email: alain.goossens@psb.vib-ugent.be

Received: *10 December 2013*
Accepted: *2 June 2014*

New Phytologist (2014) **204:** 521–535
doi: 10.1111/nph.12917

Key words: chimeric pathway, diatom, fusion enzymes, isopentenyl diphosphate isomerase, oxidosqualene cyclase, *Phaeodactylum tricornutum*, squalene epoxidase, sterol biosynthesis.

Summary

- Diatoms are unicellular photosynthetic microalgae that play a major role in global primary production and aquatic biogeochemical cycling. Endosymbiotic events and recurrent gene transfers uniquely shaped the genome of diatoms, which contains features from several domains of life. The biosynthesis pathways of sterols, essential compounds in all eukaryotic cells, and many of the enzymes involved are evolutionarily conserved in eukaryotes. Although well characterized in most eukaryotes, the pathway leading to sterol biosynthesis in diatoms has remained hitherto unidentified.
- Through the DiatomCyc database we reconstructed the mevalonate and sterol biosynthetic pathways of the model diatom *Phaeodactylum tricornutum in silico*. We experimentally verified the predicted pathways using enzyme inhibitor, gene silencing and heterologous gene expression approaches.
- Our analysis revealed a peculiar, chimeric organization of the diatom sterol biosynthesis pathway, which possesses features of both plant and fungal pathways. Strikingly, it lacks a conventional squalene epoxidase and utilizes an extended oxidosqualene cyclase and a multifunctional isopentenyl diphosphate isomerase/squalene synthase enzyme.
- The reconstruction of the *P. tricornutum* sterol pathway underscores the metabolic plasticity of diatoms and offers important insights for the engineering of diatoms for sustainable production of biofuels and high-value chemicals.

Introduction

Sterols are fundamental terpenoids in eukaryotes. They are important components of the plasma membrane, play relevant roles in cellular defence and signalling, and are precursors of several hormones and bioactive secondary metabolites (Adolph *et al.*, 2004; Benveniste, 2004; Dufourc, 2008; Vinci *et al.*, 2008; Galea & Brown, 2009; Tomazic *et al.*, 2011). The ability to synthesize sterols is a common feature of eukaryotes, with rare exceptions represented by some insects, nematodes and oomycete plant pathogens, such as *Phytophthora* spp. (Desmond & Gribaldo, 2009; Gaulin *et al.*, 2010). A few examples of bacteria with a minimal sterol pathway have been reported, although most prokaryotes synthesize hopanoids, structurally similar compounds that do not incorporate oxygen in position C-3 (Pearson *et al.*, 2003; Lamb *et al.*, 2007). Therefore, being deeply rooted in the early history of eukaryotic life, the sterol biosynthesis pathway can be considered as a prime example of metabolic evolutionary conservation. It is believed that the Last Eukaryotic Common Ancestor (LECA) already possessed some of the metabolic enzymes of the present sterol pathway (Desmond & Gribaldo, 2009). The presumed presence of a primitive form of this pathway in the LECA is reflected by the conservation of many aspects of the pathway in all sterol-producing organisms.

Originally classified into three main variants yielding cholesterol in animals, ergosterol in fungi and diverse phytosterols in land plants (Fig. 1), it is now becoming evident that this subdivision oversimplifies the actual organization of the sterol biosynthesis pathways. The recent increase in the number of sequenced nonmodel organism genomes, and technical advances in genomics and metabolomics triggered a re-evaluation of the organization of sterol biochemistry in fungi (Weete *et al.*, 2010), green algae (Massé *et al.*, 2004; Miller *et al.*, 2012), choanozoa (Kodner & Summons, 2008), kinetoplastids (Nes *et al.*, 2012), dinoflagellates (Leblond & Lasiter, 2012) and even land plants, in which an alternative branch of the pathway has been discovered in *Arabidopsis thaliana* (Ohyama *et al.*, 2009).

Despite the diversity in the end products of the sterol biosynthesis pathway, many upstream reactions and intermediates are

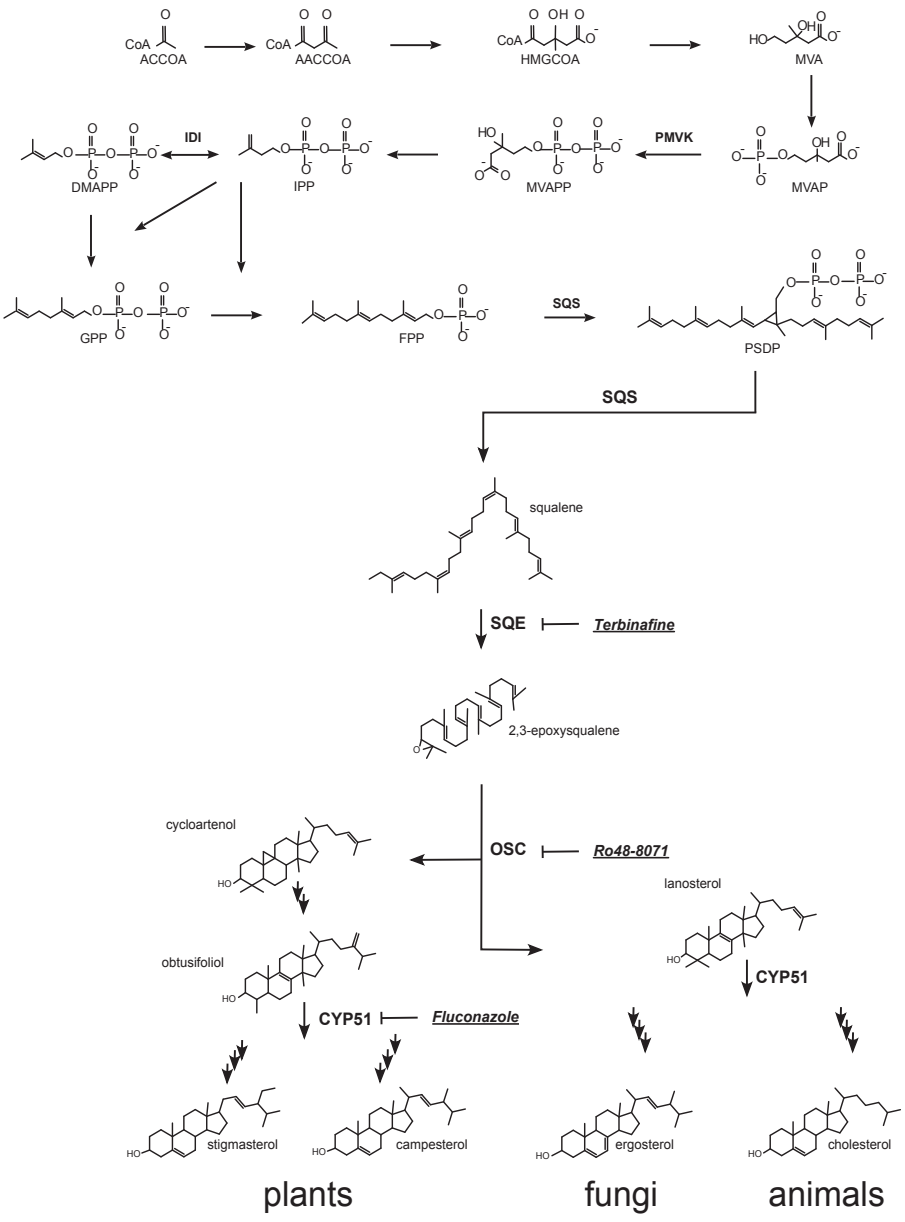

Fig. 1 Conserved reactions of the mevalonate (MVA) and sterol biosynthesis pathways. Enzymes described in the text are highlighted in bold. Chemical enzyme inhibitors are indicated in italics, bold and underlined font. IDI, isopentenyl diphosphate isomerase; PMVK, phosphomevalonate kinase; SQS, squalene synthase; SQE, squalene epoxidase; OSC, oxidosqualene cyclase; CYP51, cytochrome P450 sterol-14-demethylase; ACCOA, acetyl-CoA; AACCOA, aceto-acetyl-CoA; HMGCOA, 3-hydroxy-3-methylglutaryl-coenzyme A; MVA, mevalonate; MVAP, mevalonate phosphate; MVAPP, mevalonate diphosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; PSDP, presqualene diphosphate.

ubiquitously conserved in the different taxonomical groups (Fig. 1). Land plants and several photosynthetic organisms use two distinct parallel pathways for the synthesis of terpenoid precursors. Generally, the sterol precursors are produced through the cytosolic mevalonate (MVA) pathway, whereas precursors of, for example, carotenoids are synthesized through the plastidic methylerythritol phosphate (MEP) pathway. In green algae and some members of red algae, however, the sterol building blocks are provided by the MEP pathway because the MVA pathway was lost in these organisms (Massé *et al.*, 2004; Lohr *et al.*, 2012). Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are isomeric intermediates of both pathways. In the MVA pathway, IPP is formed from mevalonate diphosphate (MVAPP) and isomerized to DMAPP by isopentenyl diphosphate isomerase (IDI). The condensation of IPP and DMAPP produces geranyl diphosphate (GPP), which forms farnesyl diphosphate (FPP) by condensing with another molecule of IPP. FPP is the substrate of squalene synthase, which initiates sterol biosynthesis. The squalene in prokaryotes is cyclized to hopanoids by squalene-hopanoid cyclases (SHCs), whereas in eukaryotes it is first epoxidized by squalene epoxidase (SQE) before cyclization by an oxidosqualene cyclase (OSC). The cyclization of 2,3-epoxysqualene to lanosterol in animals and fungi, and cycloartenol in plants and green algae is catalysed by distinct OSCs: lanosterol synthase (LAS) and cycloartenol synthase (CAS), respectively. The next conserved reaction is catalysed by a cytochrome P450 (P450), sterol-14-demethylase (CYP51), which removes a methyl group from lanosterol in animals and fungi, and from obtusifoliol, a product of cycloartenol conversions, in the green lineage (i.e. the Viridiplantae that include land plants and green algae). After this reaction, the synthesis of sterols becomes more specific and varies depending on the phylogeny.

Contrary to the well understood biochemistry of sterols in animals, plants and fungi, our knowledge of other organisms, and in particular the many groups of unicellular eukaryotes, remains fragmentary. Although efforts to chemically characterize the sterol composition of several diatom species have revealed a marked diversity in products (Volkman, 2003; Rampen *et al.*, 2010; Giner & Wikfors, 2011), the diatom sterol biosynthesis pathway remains unknown. Some diatom sterols seem to derive from cycloartenol, like phytosterols, whereas others originate from lanosterol, making the collocation of the pathway in the tripartite subdivision difficult (Rampen *et al.*, 2010). Diatoms belong to the same kingdom as the oomycetes, one of the rare eukaryotic groups that have lost the ability to synthesize sterols during their evolution (Gaulin *et al.*, 2010). Therefore, the enzymatic and genetic study of sterol biosynthesis in diatoms is fascinating, from both evolutionary and ecological perspectives.

Here, we report the reconstruction of the MVA and sterol biosynthesis pathways of the model pennate diatom *Phaeodactylum tricornutum*. Using DiatomCyc (www.diatomcyc.org; Fabris *et al.*, 2012) to identify the set of genes putatively involved in diatom sterol synthesis, we mapped and experimentally validated the main biochemical steps of the pathway. The proposed biochemical route is a hybrid pathway that shares elements with fungal and plant pathways. We report that diatoms, and many other groups of marine organisms, surprisingly lack the ubiquitously conserved SQE and harbour uncommon enzymes, including an extended OSC and an IDI-SQS multifunctional fusion enzyme.

Materials and Methods

Chemicals

Squalene, 2,3-epoxysqualene, cycloartenol, lanosterol, ergosterol, fenpropimorph, fluconazole, imidazole and terbinafine were purchased from Sigma-Aldrich, and Ro 48-8071 from Cayman Chemicals (Ann Arbor, MI, USA).

Generation of plasmid vectors

Full-length (FL) genes were amplified from a cDNA library of *Phaeodactylum tricornutum* (Huysman *et al.*, 2013) with PrimeSTAR® HS DNA Polymerase (Takara Bio, Ōtsu, Japan), Gateway™ cloned into pDONR221 (Invitrogen) and sequence verified. PCR primers (Table S1) were designed with Vector NTI (Invitrogen). Destination vectors were generated with the Gateway™ cloning technology, unless specified otherwise.

For heterologous expression in *Escherichia coli*, pDEST17 (Invitrogen) was used. *PtOSC* was fused to Maltose Binding Protein (MBP) at its N-terminus by cloning into an in-house modified pDEST17 (pDEST17-MBP). For the co-expression of *PtIDISQS* and *SQE* candidate genes, the *PtIDISQS* cassette flanked by the T7 promoter and terminator was amplified with *Sma*I and *Spe*I containing primers and cloned into the backbone of pDONR223, *de novo* amplified by PCR to introduce compatible restriction sites and remove the Gateway cassette.

For heterologous expression in *Saccharomyces cerevisiae* W303, *PHATRDRAFT_51757* and *PHATRDRAFT_30461* were cloned into pAG426GAL-ccdB and pAG423GAL-ccdB (Alberti *et al.*, 2007), respectively. For heterologous expression in *S. cerevisiae* strain TM5 (Moses *et al.*, 2014), *PtOSC* was cloned into pAG423GAL-ccdB (Alberti *et al.*, 2007).

Bacterial and yeast culturing

E. coli BL21 (DE3) (Invitrogen) transformants were grown in Luria-Bertani (LB) broth supplemented with 50 mg l^{-1} chloramphenicol and 100 mg l^{-1} carbenicillin at 37°C until A_{600nm} of 0.6. Gene expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), followed by a 48-h incubation at 20°C in darkness. Cells were harvested and lyophilized for HPLC analysis.

S. cerevisiae transformants were selected on SD medium (Clontech) containing appropriate drop out supplements (Clontech Laboratories, Mountain View, CA, USA). W303 derived strains were grown overnight in liquid medium at 30°C with shaking. Gene expression was induced by inoculating washed precultures in SD GAL/RAF medium (Clontech) containing appropriate drop out supplements and incubated for 2 d at 30°C. TM5 derived strains were cultured as described (Moses *et al.*, 2014).

Protein extraction and immunoblot analysis

Cells from a 200-ml *E. coli* culture were resuspended in phosphate buffered saline and lysed by sonication for 1 min with 10-s pulses using a Heat Systems Ultrasonics sonicator (Heat Systems Inc., Newtown, CT, USA). Cell lysates were directly used for *in vitro* enzymatic assays or separated from cell debris by centrifugation at 10 000 ***g*** for 20 min. Protein purification was performed using Ni-NTA Superflow resin (Qiagen). Immunoblot analysis was carried out with Mini-PROTEAN® Precast Gels and related equipment (Bio-Rad), anti-His Antibody Selector Kit (Qiagen), HRP-linked anti-mouse antibody (GE Healthcare, Dienen, Belgium) and the Clarity™ Western ECL Substrate (Bio-Rad).

Treatments with chemical inhibitors

Three-day-old *P. tricornutum* CCAP 1055/1 cultures grown in ESAW medium (Berges *et al.*, 2001) at 21°C in continuous light (average intensity 75 µmol photons m^{-2} s^{-1}), were treated in triplicate with terbinafine, fluconazole, Ro 48-8071, fenpropimorph or imidazole for 48 h. Samples were harvested 1, 2, 4, 6, 8, 12, 24 and/or 48 h after treatment. Mock treatments were performed with the corresponding solvents: dimethyl sulfoxide for terbinafine and fenpropimorph, methyl acetate for Ro 48-8071, 10% ethanol for fluconazole, and water for imidazole.

Sterol extraction and gas chromatography – mass spectrometry (GC-MS) analysis

Cells from 50 ml of *P. tricornutum*, 2 ml of *E. coli*, 1 ml of *S. cerevisiae* W303 or 12 ml of *S. cerevisiae* TM5 cultures were collected and snap frozen. Cells were lysed by incubation for

10 min at 95°C in equal volumes (250 µl) of 40% KOH and 50% ethanol. Extraction was achieved by adding 900 µl hexane to the lysate and collecting the organic phase. This procedure was repeated two times. The pooled organic fractions were evaporated and derivatised with 20 µl pyridine (Sigma-Aldrich) and 100 µl *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich). Authentic standards were dissolved in hexane, evaporated and derivatised similarly.

GC-MS analysis was performed with the GC model 6890 and MS model 5973 (Agilent, Santa Clara, CA, USA). One microlitre of sample was injected in splitless mode into a VF-5ms capillary column (Varian CP9013; Agilent). Helium carrier gas was set at a constant flow of 1 ml min^{-1}. The injector temperature was set to 280°C and the oven temperature was programmed as follows: 80°C for 1 min post injection; ramped to 280°C at 20°C min^{-1}, held at 280°C for 45 min, ramped to 320°C at 20°C min^{-1}, held at 320°C for 1 min and cooled to 80°C at 50°C min^{-1} at the end of the run. The MS transfer line was set to 250°C, the MS ion source to 230°C, and the quadrupole to 150°C, throughout. Full EI-MS spectra were generated by scanning the *m/z* range of 60–800 with a solvent delay of 7.8 min.

Extraction and quantification of lycopene

Aliquots of lyophilized *E. coli* samples were weighed (0.1 mg accuracy) and extracted with acetone : water, 90 : 10 (v/v) and sonicated with a tip sonicator at 40 W for 30 s, 2-s pulses. Extracts were filtered through a 0.2-µm Alltech® nylon syringe filter (Thermo Fisher Scientific, Waltham, MA, USA) to remove cell debris and injected into an Agilent 1100 series HPLC system equipped with a Grace® reverse phase Eclipse XDB C_{18} column (150 × 4.6 mm; 3.5 µm). Lycopene was analysed as described (Van Heukelem & Thomas, 2001) using two solvents: solvent A, 70 : 30 (v/v) methanol, 28 mM aqueous *N*-tert-butylacrylamide (TBAA), pH 6.5; solvent B, methanol. Lycopene was identified by comparing retention times and absorption spectra, and quantified by calculating response factors, using pure lycopene standards (DHI, Hørsholm, Denmark).

Gene silencing in *P. tricornutum*

The *PtOSC* RNA interference (RNAi) construct was prepared as described elsewhere (De Riso *et al.*, 2009) using the primers listed in Supporting Information Table S1. *P. tricornutum* cells were transformed with *PtOSC*-RNAi or empty pAF6 vectors by biolistic transformation as reported (Falciatore *et al.*, 1999) with a PDS-1000/He™ System (Bio-Rad). Transformants were selected and cultivated on ESAW medium containing 100 µg ml^{-1} phleomycin.

Nile Red staining and fluorescence microscopy

One millilitre of *P. tricornutum* culture was stained with 5 µl Nile Red (9-(diethylamino)-5H benzo [α] phenoxazin-5-one) (from a stock solution of 2 mg ml^{-1} in acetone) and incubated for 15 min in the dark. Fluorescence microscopy images were acquired with a Zeiss Axio Imager.M2m microscope (Carl Zeiss, Germany) as described previously (Greenspan *et al.*, 1985).

Quantitative real-time PCR (qRT-PCR) analysis

cDNA was generated with iScript (Bio-Rad) from RNA extracted with RNeasy (Qiagen) from *P. tricornutum* cultures. qRT-PCR was carried out with a Lightcycler 480 (Roche) and SYBR Green QPCR Master Mix (Stratagene, La Jolla, CA, USA). *Histone H4* (*H4*) and *Tubulin* β *chain* (*TubB*) were used as the reference genes for normalisation (Siaut *et al.*, 2007). Primers for amplification of *PtOSC* (Table S1) were designed with Beacon Designer (Premier Biosoft; www.premierbiosoft.com).

Results

P. tricornutum cultures accumulate the C-28 sterols brassicasterol and campesterol

In order to profile the pool of sterols produced by *P. tricornutum*, we analysed 3-d-old cultures with GC-MS. *P. tricornutum* only accumulated C-28 sterols, mainly brassicasterol (24-methylcholest-5,22-dien-3β-ol) and campesterol (24-methylcholest-5-en-3β-ol) (Fig. 2), in agreement with earlier reports (Rampen *et al.*, 2010).

In silico reconstruction of the *P. tricornutum* sterol pathway

We mined DiatomCyc (www.diatomcyc.org; Fabris *et al.*, 2012) to retrieve the set of genes putatively encoding the biosynthetic enzymes involved in MVA and sterol biosynthesis in *P. tricornutum*. The identified genes, listed in Table 1, generally show homology with orthologues from plants, yeast, algae and animals, with high orthology scores (Table S2). Most genes lacked either a correct start or stop codon or both, thus their gene models were manually corrected with the aid of in-house available RNA-Seq data and uploaded in DiatomCyc v2.0 (Fabris *et al.*, 2012).

We identified the whole set of genes encoding the enzymes involved in the MVA pathway (Fig. 3). Many of these genes show pronounced similarity with plant orthologues (Tables 1, S2), in accordance with the fact that this pathway also provides the sterol precursors in plants, but not in green algae, in which it is lost. Initially, a phosphomevalonate kinase (PMVK) could not be identified in *P. tricornutum* by the orthology prediction method (Fabris *et al.*, 2012), or by BLAST searches on the NCBI and JGI databases (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html), whereas we could find corresponding orthologues in all other sequenced diatom genomes. By performing BLAST searches with Arabidopsis *PMVK* (*At1g31910*) as the query sequence and public and in-house generated *P. tricornutum* transcript sequences, we identified a 1395-nt transcript encoding the missing PMVK. The novel *PtPMVK*

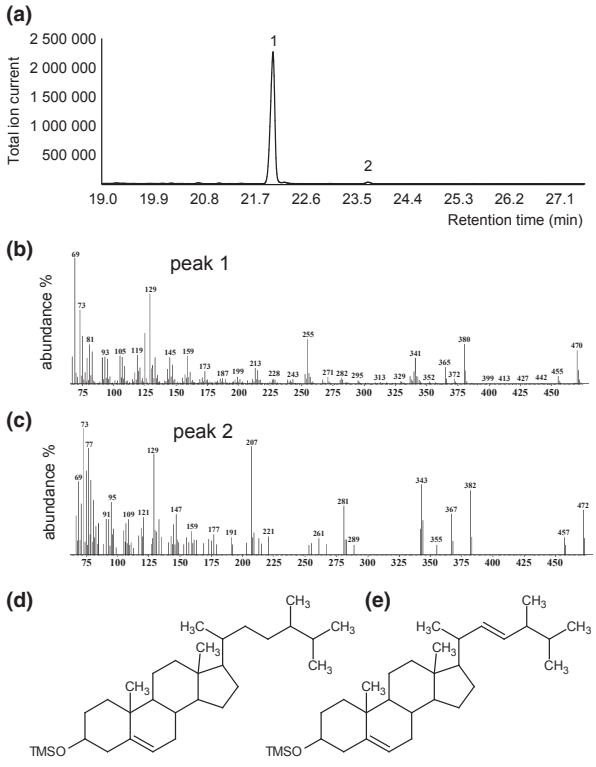

Fig. 2 *Phaeodactylum tricornutum* sterol composition. (a) GC-MS chromatogram of a TMS-derivatised extract from a 3-d-old *P. tricornutum* culture. (b, c) EI-MS spectra of TMS-derivatised brassicasterol (b) and campesterol (c) standards. (d, e) Structures of TMS-campesterol (d) and TMS-brassicasterol (e).

transcript could be mapped, in reverse orientation, onto a specific region on chromosome 4 (chr_4:691145-692539), as part of the locus *PHATRDRAFT_44478*, erroneously predicted as an intron (Fig. S1).

P. tricornutum possesses two isoforms of IDI, encoded by *PHATRDRAFT_12533* and *PHATRDRAFT_12972* (Fig. 3), and localized on chromosome 8 and 10, respectively. Although the first exists as an independent ORF, the latter is part of an elongated gene model, not predicted by the original genome annotation (Bowler *et al.*, 2008). Remarkably, the manual refinement of the gene model revealed an in-frame fusion with the only locus encoding the SQS enzyme (*PHATRDRAFT_13160*). GPP is generated by the GPP synthases *PHATRDRAFT_47271* and *PHATRDRAFT_49325*. Likely, the latter also catalyses the subsequent conversion to FPP, as it shows high similarity with ERG20, which in *S. cerevisiae* catalyses both reactions (Tables 1, S2).

Beyond FPP, the *P. tricornutum* sterol biosynthesis pathway is predicted to involve 11 genes (Table 1) that encode enzymes necessary for the conversion of squalene to the sterols brassicasterol and campesterol. These enzymes allow different hypothetical pathway reconstructions; both of the synthetic routes used by fungi and plants are theoretically possible (Figs 1, 3). Among the identified genes we found a single OSC, encoded by a noticeably extended ORF. Another fact that emerged from the *in silico* analysis of the pathway is the lack of a conventional SQE. The two latter observations, the occurrence of the *IDI-SQS* fusion gene, and the overall pathway structure were investigated further.

Identification of the *IDI-SQS* fusion gene in *P. tricornutum*

The gene model of the ORF that putatively encodes a fusion enzyme with predicted IDI and SQS activities was manually adjusted, resulting in a 2335-nt long ORF. Originally predicted as independent adjacent ORFs, *PHATRDRAFT_12972* (*PtIDI*) and *PHATDRAFT_13160* (*PtSQS*) are actually spaced by a short coding sequence that does not include stop codons. The encoded protein consists of 763 amino acids (AAs) subdivided in a NUDIX hydroxylase/isopentenyl diphosphate delta-isomerase domain at the N-terminus and a squalene/phytoene synthase domain at the C-terminus (Fig. 4a), both sharing high similarity with the IDI and SQS enzymes of brown and green algae, respectively (Table S2). The same *IDI-SQS* fusion gene is present in all sequenced diatoms, as well as in *Aureococcus anophagefferens* and *Ectocarpus siliculosus* (Fig. S2), suggesting that it is conserved among Stramenopiles. A notable exception is the parasitic heterokont oomycete *Phytophthora*, which possesses the *IDI* gene, but lacks the *SQS* gene, as well as the whole sterol biosynthesis pathway (Gaulin *et al.*, 2010).

The *in silico* predicted *IDI-SQS* gene structure was validated at the transcript level by mining public and in-house generated *P. tricornutum* transcript sequences and by performing RT-PCR (Fig. 4b), and at the protein level by producing recombinant PtIDISQS proteins in *E. coli* (Fig. 4c).

Functional characterization of *PtIDISQS*

The isomerization of IPP to DMAPP is a key step in the biosynthesis of triterpenoids through the MVA pathway. Also for IPP synthesized through the MEP pathway, the equilibrium of the conversion favours the formation of DMAPP, which is considered a rate-limiting step for the synthesis of other isoprenoids such as the carotenoids (Sun *et al.*, 1998). Therefore, we evaluated the functionality of the IDI domain of PtIDISQS in *E. coli* transformed with the pAC-LYC plasmid that carries a set of *Erwinia herbicola* genes that enable the synthesis of the carotenoid lycopene (Cunningham & Gantt, 2007). *E. coli* transformed with pAC-LYC yield pink colonies, due to the accumulation of lycopene made from DMAPP and IPP precursors provided by the endogenous MEP pathway of the bacteria. Enhancing IDI activity results in increased lycopene accumulation, producing a visible change in the colour of *E. coli* cultures (Cunningham & Gantt, 2007). Accordingly, expression of *PtIDISQS* produced darker pink coloured cells resulting from an increased accumulation of lycopene (Fig. 5a), demonstrating the functionality of the IDI domain in the fusion protein.

In order to determine whether *PtIDISQS* also encodes a functional SQS, its ability to convert FPP to squalene was tested in *E. coli* as well. Wild-type *E. coli* does not synthesize squalene or sterols, but naturally produces FPP. Plasmids carrying *PtIDISQS* were transformed into *E. coli* and 48 h after induction of

Table 1 List of genes putatively involved in the mevalonate (MVA) and sterol biosynthesis pathway of *Phaeodactylum tricornutum*

Gene ID	EC #	Predicted function	Closest orthologue in orthology-prediction (DiatomCyc)[1] (score) *GeneID* (*organism*)	
			1st InParanoid hit	*2nd InParanoid hit*
PHATRDRAFT_23913	2.3.1.9	Acetyl-coa c-acetyltransferase	(423) AT5G47720 (*A. thaliana*)	(399) YPL028W (*S. cerevisiae*)
PHATRDRAFT_16649	2.3.3.10	Hydroxymethylglutaryl-CoA (HMG-CoA) synthase	(385) AT4G11820 (*A. thaliana*)	(355) 3157 (*H. sapiens*)
PHATRDRAFT_16870	1.1.1.34	Hydroxymethylglutaryl-CoA (HMG-CoA) reductase	(410) AT1G76490 (*A. thaliana*)	(270) MJ0705 (*M. jannaschii*)
PHATRDRAFT_53929	2.7.1.36	Mevalonate kinase	(164) YMR208W (*S. cerevisiae*)	(140) 4598 (*H. sapiens*)
PHATRDRAFT_44478	2.7.4.8	Phosphomevalonate kinase	–	–
PHATRDRAFT_BD1325	4.1.1.33	Diphosphomevalonate decarboxylase	(318) AT2G38700 (*A. thaliana*)	(315) 4597 (*H. sapiens*)
PHATRDRAFT_12972	5.3.3.2	Isopentenyl-diphosphate δ-isomerase	(265) CHLREDRAFT_24471 (*C. reinhardtii*)	(237) OSTLU_13493 (*O. lucimarinus*)
PHATRDRAFT_12533	5.3.3.2	Isopentenyl-diphosphate δ-isomerase	(265) CHLREDRAFT_24471 (*C. reinhardtii*)	(156) 3422 (*H. sapiens*)
PHATRDRAFT_13160[2]	2.5.1.21	Squalene synthase	(279) OSTLU_31144 (*O. lucimarinus*)	(274) 2222 (*H. sapiens*)
PHATRADRFT_47271	2.5.1.1	Geranyl-diphosphate synthase	(309) 9453(*H. sapiens*)	(202) YPL069C (*S. cerevisiae*)
PHATRDRAFT_49325	2.5.1.1	Geranyl-diphosphate synthase	(343) Ot03g02400 (*O. tauri*)	(340) YJL167W (*S. cerevisiae*)
PHATRDRAFT_49325	2.5.1.10	Farnesyl-diphosphate synthase	(343) Ot03g02400 (*O. tauri*)	(340) YJL167W (*S. cerevisiae*)
not found	1.14.99.7	Squalene epoxidase	–	–
PHATR_645[2]	5.4.99.8	Oxidosqualene cyclase	(602) Ot03g00850 (*O. tauri*)	(596) AT2G07050 (*A. thaliana*)
PHATRDRAFT_10824	2.1.1.41	24-methylenesterol C-methyltransferase	(301) AT5G13710 (*A. thaliana*)	(296) OSTLU_30710 (*O. lucimarinus*)
PHATRDRAFT_49447	5.5.1.9	Cycloeucalenol cycloisomerase	(250) Ot04g04910 (*O. tauri*)	(231) OSTLU_6401 (*O. lucimarinus*)
PHATRDRAFT_31339	1.14.13.70	14-α-demethylase	(435) CMS319C (*C. merolae*)	(427) OSTLU_43938 (*O. lucimarinus*)
PHATRDRAFT_10852	1.14.13.72	Methylsterol monooxygenase	(82) AT1G07420 (*A. thaliana*)	–
PHATRDRAFT_48864	1.1.1.170	3-β-hydroxysteroid-4-α-carboxylate -3-dehydrogenase	(318) Ot04g04390 (*O. tauri*)	(305) OSTLU_87094 (*O. lucimarinus*)
PHATRDRAFT_5780	1.1.1.270	putative SDR oxidoreductase	(103) AT5G65205 (*A. thaliana*)	(97) 3248 (*H. sapiens*)
PHATR_36801	5.3.3.5	3-Beta-hydroxysteroid-delta(8), delta(7)-isomerase	–	–
PHATRDRAFT_14208	1.14.21.6	δ-7-sterol δ-5-dehydrogenase	(193) AT3G02580 (*A. thaliana*)	(178) CHLREDRAFT_59933 (*C. reinhardtii*)
PHATRDRAFT_30461	1.3.1.21	Δ7-sterol reductase	(426) AT1G50430 (*A. thaliana*)	–
PHATRDRAFT_51757	1.3.1.-.	sterol C-22 desaturase	(230) CHLREDRAFT_196874 (*C. reinhardtii*)	(219) CMJ284C (*C. merolae*)

Genes retrieved from DiatomCyc are listed with their predicted function and first and second InParanoid hits determined in (Fabris *et al.*, 2012) and listed in Supporting Information Table S2.
[1]Stramenopiles are excluded.
[2]Reconstructed gene models are discussed in the text.

expression, the nonsaponifiable lipid fraction was extracted and analysed with GC-MS. The resulting chromatograms showed a recurrent small peak corresponding to squalene in strains expressing *PtIDISQS*, but not in control strains (Fig. 5b,c), confirming the predicted SQS activity of the fusion enzyme, as well as supporting the occurrence of squalene as a sterol pathway intermediate in *P. tricornutum*.

Many Stramenopiles lack a conventional *SQE* gene

Because we could not identify *SQE* orthologues in the *P. tricornutum* genome in the dataset generated for the reconstruction of DiatomCyc, we performed an extensive BLASTP search in eukaryotes, excluding the group of fungi, animals and land plants, and using the *SQE* sequences from *S. cerevisiae, Homo sapiens* and Arabidopsis as queries. Surprisingly, *SQE* seemed to be widely absent in the analysed groups, as already noticed previously for some of the species therein (Desmond & Gribaldo, 2009), particularly in most of the analysed Stramenopiles, with the exception of *E. siliculosus* and the oomycete pathogens *Aphanomyces eusteiches* and *Saprolegnia diclina*. No hits were obtained in any Alveolata and Choanoflagellida either, but SQE seemed to be conserved in the red and green algal lineages, with the exception of *Chlamydomonas reinhardtii* (Fig. 6).

Chemical inhibitor treatments indicate that 2,3-epoxysqualene is a sterol pathway intermediate in *P. tricornutum*

The absence of a conventional SQE raised the question of whether diatoms would use 2,3-epoxysqualene as the precursor for the cyclization step. To determine this, we treated diatom cultures with terbinafine and Ro 48-8071, specific inhibitors of conventional SQE and OSC enzymes, respectively. Pilot experiments

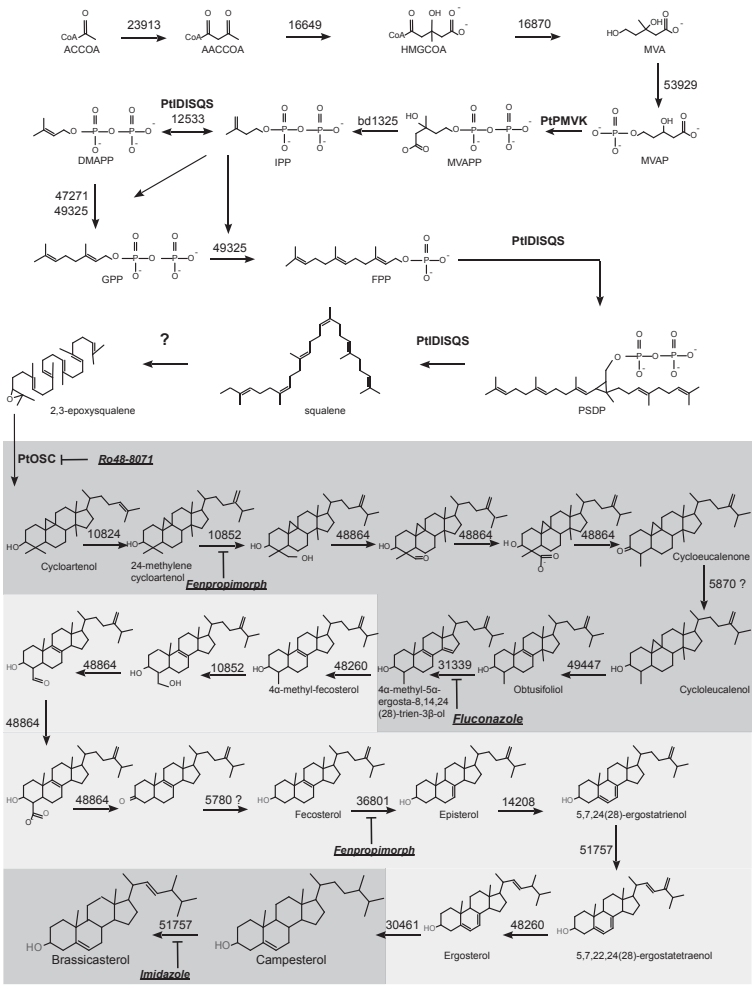

Fig. 3 *In silico* reconstruction of the mevalonate (MVA) and sterol biosynthesis pathways of *Phaeodactylum tricornutum*. The numbers above the arrows indicate the *PHATRDRAFT* accession number of the corresponding enzyme. Dark and light grey shading indicate similarities with the sterol biosynthesis of plants and fungi, respectively. Chemical enzyme inhibitors are indicated in italics, bold and underlined font. See Fig. 1 for explanation of abbreviations.

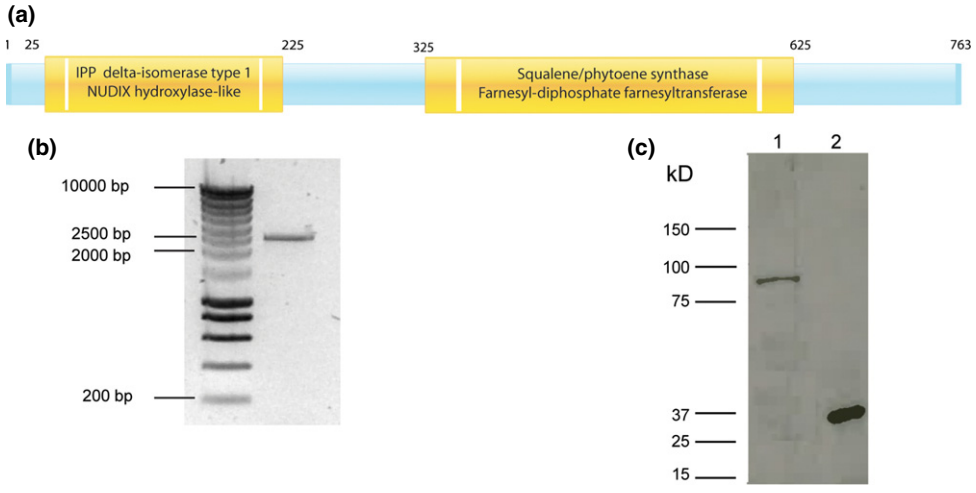

Fig. 4 The *Phaeodactylum tricornutum PtIDISQS* fusion gene. (a) Schematic organization of the IDI and SQS domains in *PtIDISQS* according to InterProScan predictions (Hunter *et al.*, 2009). Numbers refer to the approximate amino acid (AA) residue position. (b) RT-PCR amplification of *PtIDISQS* from cDNA of *P. tricornutum*. (c) Immunoblot analysis with anti-His antibodies of *Escherichia coli* BL21 (DE3) samples expressing *6xHis-PtIDISQS* (lane 1) or *6xHis-PtEDA* (Fabris *et al.*, 2012) as a control (lane 2). According to the *in silico* predictions, the 6xHis-PtIDISQS protein should have a mass of 87.7 kDa, which is in agreement with the size of the product found.

revealed that *P. tricornutum* cultures could be treated with 40 µM terbinafine without major growth impairments, but that Ro 48-8071 at a concentration of 30 µM killed them within 48 h. Therefore, time course experiments were initiated, both with terbinafine and Ro 48-8071, to allow timely detection of sterol pathway intermediates.

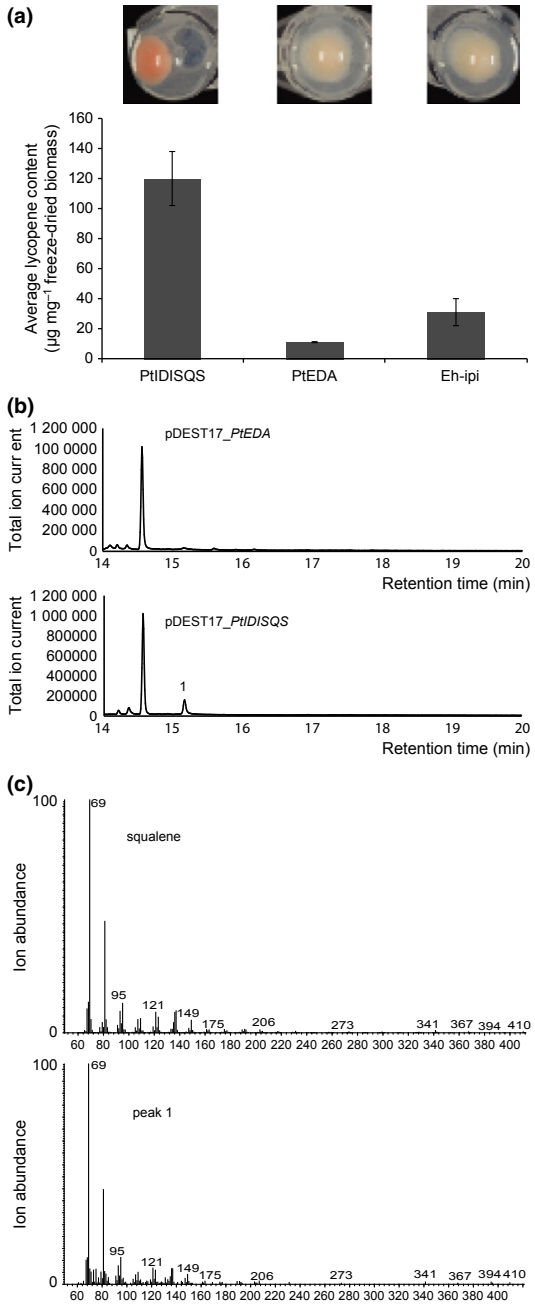

Fig. 5 Enzymatic activity of PtIDISQS. (a) Screening of lycopene accumulation in *Escherichia coli* cells transformed with the pAC-LYC plasmid. Quantification by HPLC of the lycopene content of *E. coli* colonies transformed with the pAC-LYC and *IDI* expression plasmids. Bacterial pellets showing a darker pink colour indicate an increased IDI activity caused by *PtIDISQS* overexpression (see insets on top). As a positive control, *E. coli* cells co-transformed with the *IDI* gene of *E. herbicola* (Eh-ipi) (Cunningham & Gantt, 2007) were used. As a negative control, *E. coli* cells transformed with the *PtEDA* gene were used. The latter strain indicates the basal amounts of lycopene accumulation caused by the endogenous *idi* gene of *E. coli*. Error bars, ± SE of the mean ($n = 3$). (b) GC-MS chromatogram of *E. coli* BL21 (DE3) cells expressing *PtIDISQS* and showing accumulation of squalene (peak1), compared to *E. coli* cultures expressing *PtEDA* as negative control. (c) Comparison of the MS of an authentic squalene standard with that of peak 1 from panel (b).

In accordance with the predicted absence of a conventional SQE, treatments with 40 μM terbinafine had no effect on growth, appearance or squalene accumulation in *P. tricornutum* cultures, not even at higher concentrations (up to 340 mM). Terbinafine treatment triggered some accumulation of ergosterol (Fig. 7b), perhaps because of interference with another oxidoreductase in the pathway, such as *PHATRDRAFT_30461* or *PHATRDRAFT_51757*.

By contrast, *P. tricornutum* cells treated with 10 μM Ro 48-8071 displayed pronounced growth effects 8–12 h after treatment, and accumulated both squalene and 2,3-epoxysqualene (Fig. 7a). Furthermore, Ro 48-8071-treated cells accumulated intracellular lipid droplets, as became apparent after staining with Nile Red (Fig. 7e). Overall, these findings confirm that diatoms produce sterols through the cyclization of the conventional precursor 2,3-epoxysqualene by a conventional OSC but using a nonconventional, terbinafine-insensitive SQE enzyme to generate the 2,3-epoxysqualene precursor.

Screening for the *P. tricornutum SQE*

We mined the *P. tricornutum* genome for genes possibly encoding enzymes catalysing epoxidation of squalene to 2,3-epoxysqualene. This led to a list of candidate SQEs that included P450s, FAD-dependent monooxygenases, carotenoid epoxidases and hydroxylases (Table S3). Given its extended structure (see further below), the *P. tricornutum* OSC was included in this list as well, as a possible multifunctional SQE-OSC enzyme.

First, we attempted to detect SQE activity in an *in vitro* assay, using lysates of *E. coli* cells transformed with pDEST17 plasmids carrying the candidate genes. Established protocols were followed for *in vitro* SQE activity determination (Nagumo *et al.*, 1995; Laden *et al.*, 2000; Germann *et al.*, 2005) using nonlabelled squalene, with or without the Arabidopsis Cytochrome P450 reductase, an enzyme required for functional reconstitution of most P450 and FAD monooxygenases. However, no 2,3-epoxysqualene could be detected by GC-MS in any of the samples. Deletion of the putative targeting peptide of *PtOSC* (see next paragraph) or the transmembrane domains of *PHATRDRAFT_46438*, *PHATRDRAFT_26422* and *PHATRDRAFT_45845* in an attempt to optimize did not lead to any detectable SQE activity, either. Second, we exploited the ability of *E. coli* cells expressing *PtIDISQS* to accumulate low amounts of squalene (Fig. 5b) to detect potential SQE activity of the candidate genes *in vivo*. This assay also did not reveal any accumulation of 2,3-epoxysqualene. Because immunoblot analysis of tagged versions of the candidate SQEs indicated that most of them were not or hardly expressed in the transformed *E. coli* cells (data not shown), we did not pursue SQE discovery any further.

Diatoms possess an extended *OSC* gene

We identified the locus *PHATRDRAFT_645* as part of an ORF putatively encoding an OSC. The gene model provided by the

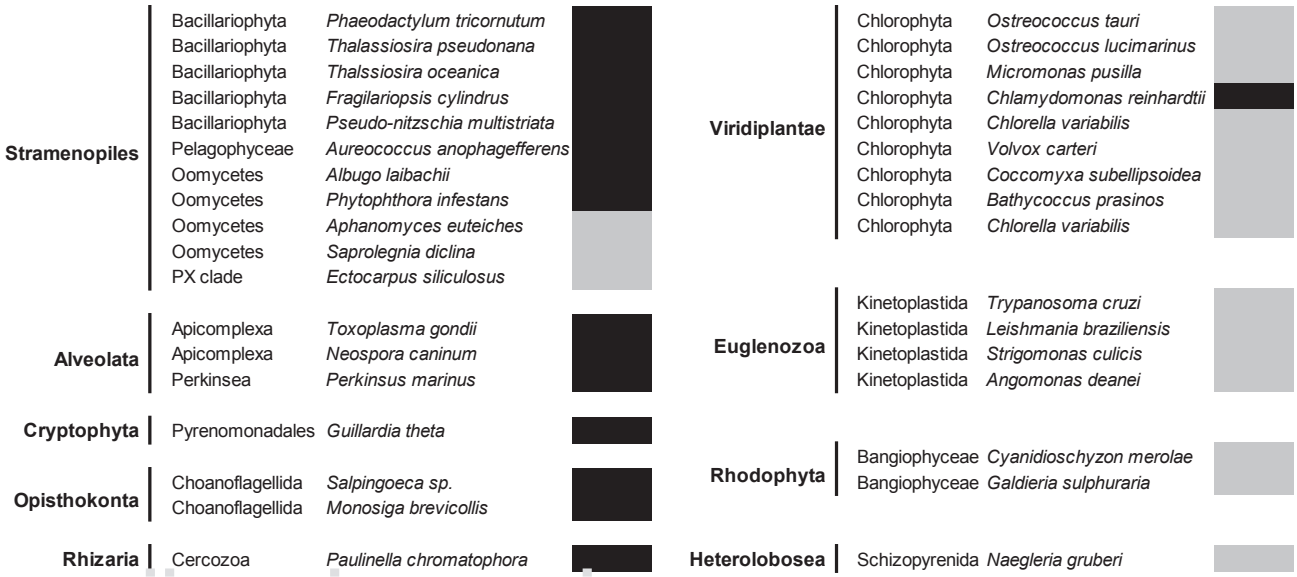

Fig. 6 Conservation (grey) and loss (black) of the *SQE* gene in representative organisms belonging to the groups of Stramenopiles, green algae, Alveolata, Choanoflagellida, Kinetoplastids, Cryptophyta, Heterolobosea and Rizharia, according to BLASTP searches.

JGI assembly (Bowler *et al.*, 2008) was incomplete and wrongly annotated as acetyl coenzyme A synthase. Therefore, it was manually adjusted, resulting in an ORF of 2931 nt, denominated *PtOSC* and spanning the coding regions of both the preceding and the succeeding gene on chromosome 11, *PHATRDRAFT_46724* and *PHATRDRAFT_46726*. The reconstructed *PtOSC* model was confirmed at the transcript level by mining of public and in-house generated *P. tricornutum* transcript sequences and by RT-PCR analysis, producing a single band of the predicted size (Fig. S3a). Notably, as for the *PtIDISQS* gene, a similarly extended *OSC* locus could be identified in all annotated diatom genomes (Fig. S4). In the predicted PtOSC protein, the sequence corresponding to the first 46 N-terminal AAs is possibly part of an ER-targeting peptide (Emanuelsson *et al.*, 2007), the following 103 AAs potentially correspond to three transmembrane domain repeats (Emanuelsson *et al.*, 2007), and the C-terminus shows similarity to common CAS and LAS enzymes. Although all diatom OSC sequences share an N-terminal extension, its sequence is apparently species-specific and not conserved among diatoms (Fig. S4). The existence of the FL protein could not be verified yet as tagged versions could not be expressed in *E. coli* or *S. cerevisiae*, perhaps due to differences in codon usage, the considerable size of the polypeptide and/or the presence of potential membrane targeting domains at the N-terminus. Therefore, a truncated version of *PtOSC,* in which the N-terminal 46 AAs were removed, was fused to MBP and expressed in *E. coli*. Immunoblot analysis revealed the existence of a large protein of *c.* 104 kDa (Fig. S3b), matching the *in silico* predicted protein model.

PtOSC encodes a cycloartenol synthase (CAS)

Being a highly conserved enzyme, the specificity of an OSC can be inferred from the AA residues at positions 381, 449 and 453 (numbering relative to the OSC of *H. sapiens*) that are specific for the formation of cycloartenol (Y381,H449,I453) or lanosterol (T381,C/Q449,V453) (Fig. S4) (Summons *et al.*, 2006), which would simultaneously also allow location of this portion of the diatom pathway within the photosynthetic or the nonphotosynthetic lineage, respectively. Analysis of these active site residues in PtOSC suggested that *P. tricornutum*, as well as other diatoms, cyclizes 2,3-epoxysqualene to cycloartenol (Fig. S4), like plants and in agreement with previous postulations for the *Thalassiosira pseudonana* OSC (Desmond & Gribaldo, 2009).

We were not able to detect any enzymatic activity with the truncated, recombinant PtOSC protein produced in *E. coli* (Fig. S3b). Considering the inefficient expression of tagged *PtOSC*, either in *E. coli* or *S. cerevisiae*, we assessed PtOSC activity in a yeast strain (TM5) that we have engineered for high, inducible, accumulation of 2,3-epoxysqualene (Moses *et al.*, 2014). Expression of full-length, nontagged *PtOSC* in this strain led to the accumulation of cycloartenol, indicating that *PtOSC* encodes a CAS (Fig. 8a).

Additionally, we employed a pharmacological approach and determined the accumulation of sterol pathway intermediates in *P. tricornutum* cultures treated with different concentrations of fluconazole, a specific inhibitor of CYP51 (*PHATRDRAFT_31339*), another conserved enzyme in the sterol pathway acting downstream of the OSC (Fig. 3). The strongest effect of fluconazole on *P. tricornutum* cultures was observed to occur 48 h after treatment at a concentration of 653 mM (200 mg ml^{-1}), but lower concentrations (tested down to 30 μM) produced similar results (data not shown). No visible effects on growth were apparent, even at the highest concentration tested, but GC-MS analysis demonstrated the presence of two distinct peaks, as compared to the mock-treated samples (Fig. 7c). The first corresponds to obtusifoliol (Figs 7c, S5), a known derivative of cycloartenol in plants, thus confirming the specificity of PtOSC. The second peak is an unknown steroid

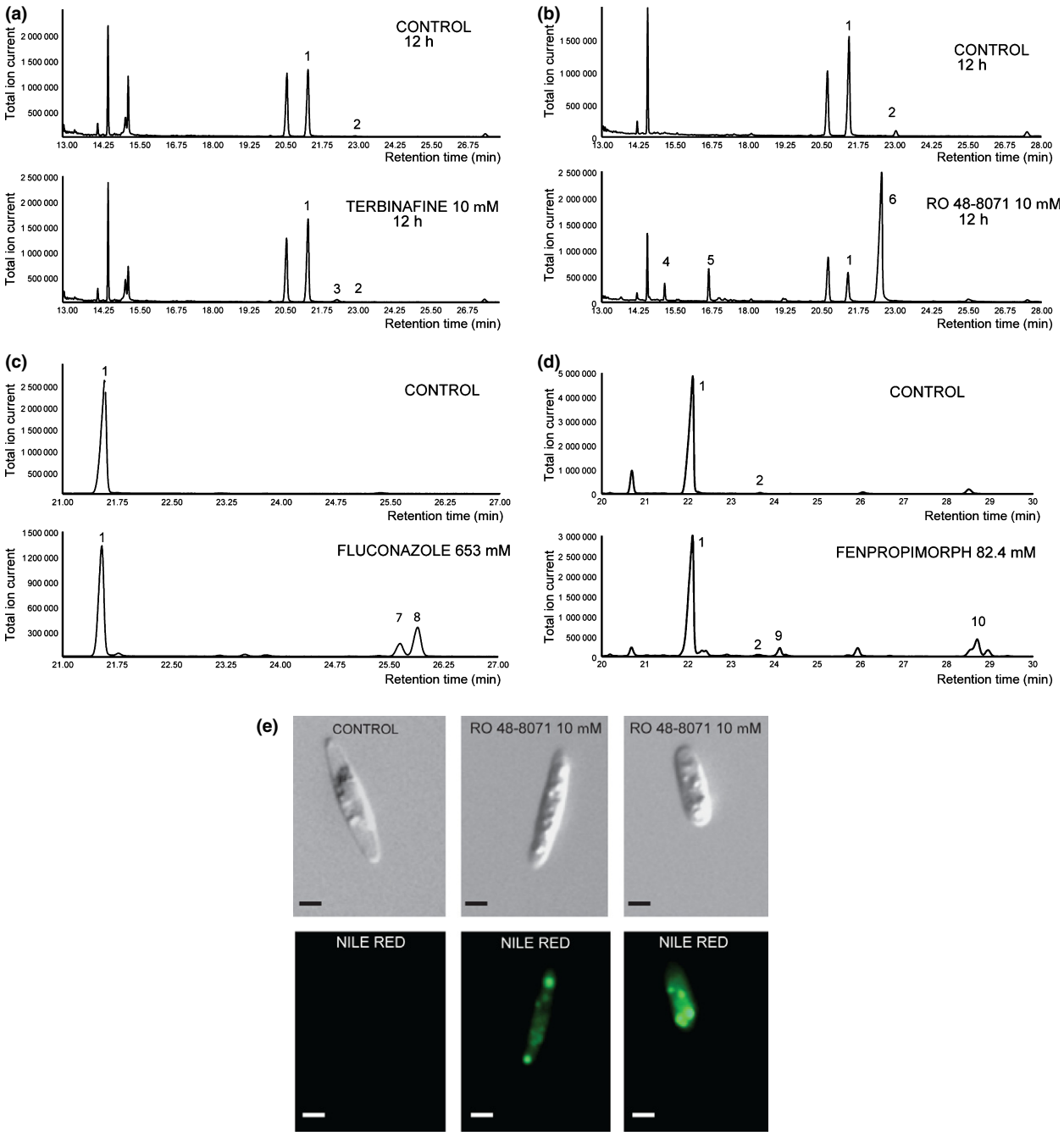

Fig. 7 Chemical perturbation of *Phaeodactylum tricornutum* sterol synthesis. (a–d) GC-MS chromatograms of TMS-derivatised nonsaponifiable lipid extracts of *P. tricornutum* cells treated for 12 h with 10 μM Ro 48-8071 (a), 40 μM terbinafine (b), 653 mM fluconazole (c) and 82.4 mM fenpropimorph (d) and compared to the respective mock treatments. Peak numbering: 1, TMS-brassicasterol; 2, TMS-campesterol; 3, TMS-ergosterol; 4, squalene; 5, 2, 3-epoxysqualene; 6, Ro 48-8071; 7, TMS-obtusifoliol; 8, TMS derivative of unknown steroid compound of FW 500; 9, putative TMS-fecosterol; 10, TMS-cycloartenol. EI-MS spectra of these peaks are given in Supporting Information Fig. S5. (e) Effects of Ro 48-8071 on lipid accumulation in *P. tricornutum* cells. Upper and lower panels show differential interference contrast (DIC) images of *P. tricornutum* cells and epifluorescence images of the same cells upon staining with Nile Red, to visualise intracellular lipids. Bars, 3 μm.

compound with the formula weight 428, which possibly corresponds to a reduced form of obtusifoliol (Fig. S5).

In order to further support the role of PtOSC in diatom sterol biosynthesis, *P. tricornutum* cells were transformed with RNAi constructs targeting this gene within the locus *PHATRDRA .FT_645*. No detectable amounts of OSC precursors accumulated in the *PtOSC* RNAi lines but they exhibited a striking phenotype characterized by dramatically impaired growth, reduced sterol content, and a significant increase in lipid accumulation (Fig. 8b–d). These phenotypes are typical for unicellular organisms with impaired sterol biosynthesis (Wentzinger *et al.*, 2002; Ta *et al.*, 2012). Also in plant and yeast cells, the

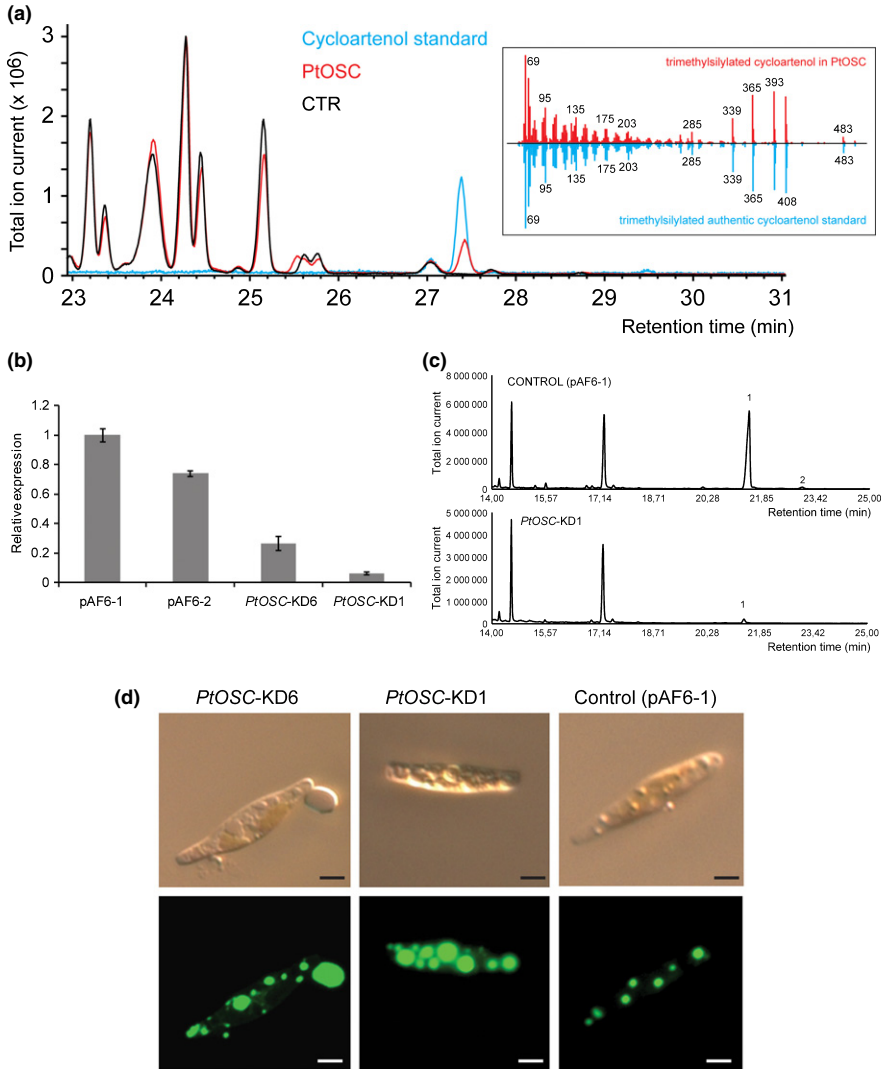

Fig. 8 PtOSC is a cycloartenol synthase. (a) Cycloartenol synthase activity of PtOSC in transformed yeast cells. Overlay of GC chromatograms showing accumulation of cycloartenol in cells transformed with *PtOSC* but not in control cells (CTR). Blue, Cycloartenol standard; red, *PtOSC*; black, CTR. The inset shows the EI-MS spectrum of trimethylsilylated cycloartenol. The GC retention time and EI-MS spectra of cycloartenol produced in yeast match those of the authentic standard. (b–d) Effects of *PtOSC* silencing in transformed *Phaeodactylum tricornutum* cells. (b) qRT-PCR analysis of *PtOSC* transcript levels in *PtOSC* RNAi lines (KD) relative to control lines transformed with the pAF6 vector. The expression ratio was normalized to the pAF6_1 control line. Error bars, ± SE of the mean ($n = 3$). (c) GC-MS chromatograms of the TMS-derivatised nonsaponifiable lipid fraction of *PtOSC-KD1*, showing a dramatic decrease in the accumulation of the main steroid compound compared to the pAF6 control line. Peak numbering: 1, TMS-brassicasterol; 2, TMS-campesterol. (d) Effects of *PtOSC* silencing on lipid accumulation in cells of 3-d-old *Phaeodactylum tricornutum* cultures. Upper and lower panels show differential interference contrast (DIC) images of *P. tricornutum* cells and epifluorence images of intracellular lipid accumulation, in cells stained with Nile Red. Bars, 3 μm.

perturbation of the sterol metabolism through chemical inhibitors causes the appearance of lipid droplets, whereas in human cells, the blockage of HMGR, catalysing the rate-limiting step of the MVA pathway, triggers accumulation of polyunsaturated fatty acids and upregulation of fatty acid biosynthesis genes (Wentzinger *et al.*, 2002; Plée-Gautier *et al.*, 2012; Ta *et al.*, 2012). Hence, these findings confirm the involvement of PtOSC in the sterol pathway and support the link between sterol biosynthesis and lipid accumulation, already observed after treatment with Ro 48-8071, the specific inhibitor of conventional OSC enzymes (Fig. 7e).

P. tricornutum employs a chimeric sterol synthesis route

The chemical fenpropimorph has strong effects and possibly multiple targets in sterol and other synthesis pathways in eukaryotes. Although drug concentrations between 50 and 100 μg ml^{-1} rapidly killed *P. tricornutum* cells, treatments at concentrations of 12.5 and 25 μg ml^{-1} caused an altered sterol profile. GC-MS analysis of sterols accumulated by fenpropimorph-treated diatom cultures revealed the presence of cycloartenol and another peak, presumably fecosterol (Figs 7d, S5), a typical fungal sterol. The accumulation of cycloartenol, which further confirms the predicted activity of PtOSC, is likely caused by the inhibition of the methylsterol monooxygenase encoded by *PHATRDRAFT_10852*, as this enzyme type is also a known target of fenpropimorph (Burden *et al.*, 1989). The presence of fecosterol may be supported by the fact that the enzyme Δ7-sterol isomerase, likely encoded by *PHATRDRAFT_36801*, that converts this compound to episterol, is a known target of fenpropimorph (Campagnac *et al.*, 2009). Based on the outcomes of the inhibitor treatments, the full *P. tricornutum* sterol pathway could be tentatively reconstructed and is proposed to be a hybrid of the plant and fungal sterol synthesis routes (Fig. 3).

The enzymes Δ7-sterol reductase and sterol C-22 desaturase catalyse the final steps in the synthesis of the main *P. tricornutum* sterols

The genes *PHATRDRAFT_30461* and *PHATRDRAFT_51757* putatively encode a sterol Δ7-reductase and a sterol C-22

desaturase, respectively (Table 1). Both show high similarity to reported Arabidopsis sterol enzymes (Table S2). *PHATRDRAFT_30461* is the putative orthologue of *DWARF5* (*At1g50430*) that encodes a sterol Δ7-reductase, whereas *PHATRDRAFT_51757* is the putative orthologue of *At2g34490*, which encodes a P450 of the CYP710 subfamily with sterol 22-desaturase activity (Morikawa *et al.*, 2006). In our proposed pathway reconstruction (Fig. 3), these enzymes are associated with the last two reactions, theoretically converting ergosterol to campesterol (*PHATRDRAFT_ 30461*) and subsequently to brassicasterol (*PHATRDRAFT_ 51757*), the two main sterols found in *P. tricornutum* (Fig. 2).

In order to confirm their predicted activity, both genes were cloned and expressed in *S. cerevisiae*, which naturally accumulates ergosterol as the main sterol compound. GC-MS analysis of yeast expressing *PHATRDRAFT_30461* confirmed its Δ7-reductase activity and its capacity to convert ergosterol to campesterol (Fig. 9a). Additionally, in the same samples, brassicasterol was also detected, presumably as a result of a desaturation of campesterol in position C-22 by ERG5, the endogenous C-22 sterol reductase. Both sterols were also detected in yeast cells expressing both diatom genes (Fig. 9b). In support of the possible activity of *PHATRDRAFT_51757* in converting campesterol to brassicasterol, we observed that in the presence of increasing concentrations of imidazole, which inhibits P450s of the CYP710 subfamily with minor specificity, *P. tricornutum* cultures increasingly accumulate campesterol in a concentration-dependent manner (Fig. 9c).

Discussion

Reconstruction of the *Phaeodactylum tricornutum* sterol synthesis pathway

Through computational analysis, we reconstructed the sterol biosynthesis pathway of *P. tricornutum in silico*. Several elements of the predicted pathway were experimentally supported with data from pharmacological assays, metabolite profiling of wild-type and transgenic *P. tricornutum* lines, and activity assays using recombinant *P. tricornutum* enzymes produced in bacteria and yeast. Together these data suggest that *P. tricornutum* utilizes a chimeric pathway that leads to the main sterols brassicasterol and campesterol, and has the initial and final parts in common with the pathway from plants, whereas the central part appears to be similar to that of fungi (Fig. 3). This supports the postulation, based on phylogenomic analysis only, that another model diatom, *T. pseudonana*, also has a sterol pathway that displays a mixture of features from fungi, animals and land plants (Desmond & Gribaldo, 2009).

Based on our analysis, the following reaction scheme is postulated. Despite the absence of a conventional SQE, *P. tricornutum* also generates 2,3-epoxysqualene as the precursor for the cyclization to cycloartenol, like plants and green algae. Cycloartenol is then methylated at C-24 by *PHATRDRAFT_10824*, a sterol methyltransferase yielding 24-methylene-cycloartenol. One of the two methyl groups at C-4 is then removed by the subsequent actions of three enzymes. The first is the methylsterol monooxygenase encoded by *PHATRDRAFT_10852*, which shows

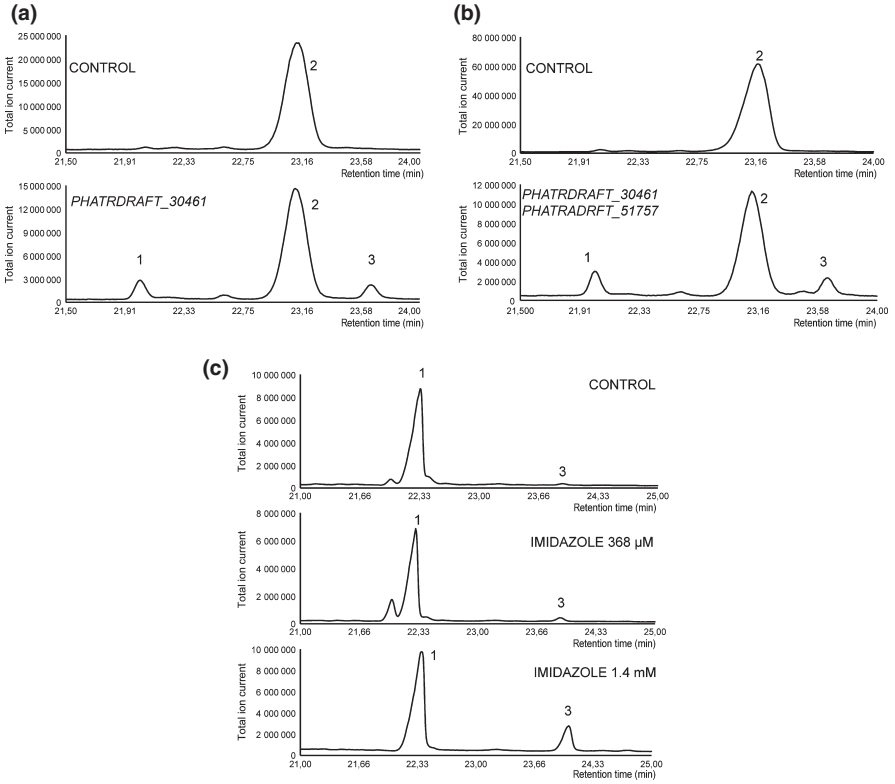

Fig. 9 Functional characterization of *PHATRDRAFT_30461* and *PHATRDRAFT_51757*. (a, b) GC-MS chromatograms of TMS-derivatised nonsaponifiable lipid extracts of *Saccharomyces cerevisiae* cells expressing *PHATRDRAFT_30461* alone (a) or in combination with *PHATRDRAFT_51757* (b) compared to controls transformed with corresponding empty vector(s). (c) GC-MS chromatograms showing the perturbed sterol composition of *Phaeodactylum tricornutum* cultures treated for 48 h with 368 µM or 1.4 mM imidazole, and compared to mock-treated cultures. Peak numbering: 1, TMS-brassicasterol; 2, TMS-ergosterol; 3, TMS-campesterol.

significant similarity with plant orthologues. The resulting hydroxysterol is decarboxylated on the C-4 hydroxy-methyl group and at the same time dehydrogenated on the hydroxyl group at the C-3 carbon by an NADPH-dependent reaction, catalysed by a 3β-hydroxysteroid-4α-carboxylate-3-dehydrogenase encoded by *PHATRDRAFT_48864,* with highest orthology scores to the *Ostreococcus* enzymes postulated previously as the homologues of the yeast C-3 dehydrogenase/C-4 decarboxylase ERG26 (Desmond & Gribaldo, 2009). The resulting keto-sterol requires a reduction of the oxygen at C-3 in order to be further converted. This step requires a 3-keto-steroid reductase, which is currently unidentified in plants and algae. We postulate that a reductase involved in other reactions might have acquired specificity for keto-sterols. For example, using Pathologic (Karp *et al.*, 2002), we identified *PHATRDRAFT_5870*, encoding a short-chain dehydrogenase/reductase as the best candidate for this reaction, potentially yielding the molecule cycloleucalenol, the typical precursor of obtusifoliol in plants. Similar to land plants, a cycloleucalenol cycloisomerase breaks the cyclopropane ring present between C-19 and C-9 with a consequent desaturation at position C-9/C-8. This enzyme is possibly encoded by *PHATRDRAFT_49447*, which shares highest orthology with the corresponding enzymes in the Viridiplantae lineage. The resulting product, obtusifoliol, a common intermediate in the sterol biosynthesis of plants and green algae, is the substrate of the P450 sterol 14α-demethylase (PtCYP51), likely encoded by *PHATRDAFT_31339*. This enzyme catalyses the removal of the methyl group at C-14 and the formation of the double bond between C-14 and C-15, yielding (4α)-methyl-(5α)-ergosta-8,14,24(28)-trien-3β-ol. The presence of a sterol C14-24 reductase, corresponding to the locus *PHATRDRAFT_48260* (PtSC14-24R), allows the conversion of this compound to methyl-fecosterol, which in turn, through the re-iteration of the oxidative demethylation at C-4 catalysed by the same enzymes as described above, is converted to fecosterol, a typical intermediate of ergosterol biosynthesis in yeast and fungi. *PHATRDR AFT_36801* encodes a Δ8-Δ7 sterol isomerase that can possibly catalyse the shift of the double bond from C-8/C-9 to C-8/C-7. The resulting episterol is first desaturated by a Δ(7)-sterol 5-desaturase, possibly encoded by *PHATRDR AFT_14208*, subsequently subjected to a second desaturation reaction at position C-22 by the P450 C-22 sterol desaturase (PtCYP710) encoded by *PHATRDRA FT_51757*, and lastly reduced at position C-24 by PtSC14-24R (*PHATRDRAFT_48260*), to yield the typical fungal sterol ergosterol. Despite being the main sterol of fungi, ergosterol is often found in algae as well (Miller *et al.*, 2012). Finally, in two steps the ergosterol is converted to the phytosterols campesterol and brassicasterol, the final products of sterol biosynthesis in *P. tricornutum*, by the Δ7-sterol reductase encoded by *PHATRDRAFT_30461* and the C-22 desaturase encoded by *PHATRDRAFT_51757*, respectively. Interestingly, orthologues of *PHATRDRAFT_51757* have not been found in *T. pseudonana* (Table S2), indicating a possible difference in the pathways of the two diatoms and in agreement with the fact that the sterols produced by *T. pseudonana* are not desaturated at position C-22 (Rampen *et al.*, 2010).

Peculiar enzymes in the diatom sterol pathway

Besides its chimeric nature, the sterol biosynthesis pathway of *P. tricornutum* harbours several other peculiarities that make it unique among the known variants of this pathway across the kingdoms of life. Particularly, the pathway is characterized by: the fusion of the IDI and SQS activities in a single multifunctional enzyme; the lack of a conventional SQE; and the presence of an exotic CAS that is composed of a conserved C-terminal domain and a large, less conserved N-terminal region. Remarkably, these features recur in all sequenced diatoms.

Joining different enzymatic activities in a single fusion protein is a frequent event in protein evolution and often implicates enzymes subjected to co-regulation or involved in the same pathway (Hwang *et al.*, 2011). In diatoms, the presence of fusion enzymes that catalyse subsequent reactions appears relatively frequent. Examples are found in carbohydrate metabolism, where a triosephosphate-isomerase/glyceraldehyde-3-phosphate dehydrogenase (*PHATRDRAFT_25308*), a UDP-glucose-pyrophosphorylase/phosphoglucomutase (*PHATRDRAFT_50444*) and a glucose-6-phosphate-dehydrogenase/6-phosphogluconate-dehydrogenase (*PHATRDRAFT_54663*) were predicted to exist as fusion proteins in *P. tricornutum* (Kroth *et al.*, 2008). In contrast to the former fusion enzymes, IDI and SQS activities do not occupy a consecutive position within the sterol pathway. DMAPP, the product of IDI, is converted to FPP before being converted by SQS to squalene. This intermediate conversion occurs through two additional reactions encoded by the loci *PHATRDRAFT_49325* and *PHATRDRAFT_47271*, located on different chromosomes than *PtIDISQS*. The IDI-SQS fusion gene is conserved in diatoms, and presumably even in the Stramenopiles, suggesting that its origin might be considerably ancient and confer a selective advantage. Although the occurrence of protein–protein interactions with other enzymes of the pathway cannot be excluded, it is plausible that a potential selective advantage of the IDI-SQS protein resides in a more efficient co-regulation, rather than in metabolic channelling.

The epoxidation of squalene by SQE, requiring FAD as cofactor and molecular oxygen, is believed to be an ubiquitous reaction, occurring in aerobic conditions in every sterol-producing organism through an identical mechanism. As for most of the other proteins working upstream of the OSC, the degree of conservation of SQE is so high that no sterol-producing organisms with alternative SQE enzymes have been reported yet. Our and previous (Desmond & Gribaldo, 2009) surveys clearly indicate that the *SQE* gene seems to be lost in several groups. The existence of alternative biochemical mechanisms for the epoxidation of squalene thus seems plausible and diatoms might use one of them.

For example, diatoms might have evolved a particular P450 that acquired a novel activity or specificity. In rat tumour cells it has been demonstrated that a P450 17-alpha hydroxylase-17,20 lyase (CYP17) has a secondary SQE activity (Liu *et al.*, 2005). Alternatively, among the substantial number of genes encoding proteins with unknown function (Maheswari *et al.*, 2010; Fabris *et al.*, 2012) diatoms might harbour an enzyme with

unprecedented SQE activity. Theoretically, squalene epoxidation could occur anaerobically using water as source of oxygen instead of O_2 (Raymond & Blankenship, 2004), although the thermodynamics of such a reaction would be significantly less favourable (Summons *et al.*, 2006). Because the synthesis of sterols presumably evolved in anaerobic eukaryotes facing an increasingly oxidative environment, it is possible that the primordial SQE was anaerobic (Raymond & Blankenship, 2004) and substantially different. Before being replaced by the 'modern' SQE, this hypothetical primitive enzyme might have persisted in some groups. Based on some of these assumptions, we compiled a list of possible alternative enzymes that might have replaced SQE in diatoms (Table S3) and we tested their activity in a series of preliminary assays in *E. coli*; these have been unsuccessful thus far.

The utility of DiatomCyc

The reconstruction of the complex diatom sterol pathway starting from the DiatomCyc database highlights the value and reliability of this tool for research on diatom metabolism. The proposed sterol biosynthesis pathway follows a chimeric fungal/plant route and introduces novel multifunctional enzymes and novel, yet unknown, enzymatic alternatives to a highly conserved biochemical event. This further underscores the prominent metabolic plasticity of diatoms, suggests that sterol biosynthesis in the marine environment might have evolved differently, and requires a general reconsideration of the sterol biosynthetic pathway, currently considered as a highly conserved pathway and subdivided into rigid phylogenetic variants. Further efforts will be required to identify the alternative SQE and confirm the role of some of the sterol enzymes postulated in this study. Similarly, the existence of a conserved link between the regulation of sterol biosynthesis and lipid accumulation, suggested by the increased lipid accumulation in diatoms with impaired sterol biosynthesis, might warrant further investigation at a physiological, metabolic and enzymatic level. Such multi-level analysis might benefit from the information made available through DiatomCyc and other, analogous resources.

Acknowledgements

We thank Dr Hubert Schaller (Institut de Biologie Moléculaire des Plantes University of Strasbourg, France) for the obtusifoliol standard, Professor Francis X. Cunningham (University of Maryland, USA) for the pAC LYC and pAC LYCipi plasmids, Wilson Ardiles-Diaz for sequencing and Annick Bleys for help in preparing the manuscript. This work was supported by the Agency for Innovation by Science and Technology in Flanders ('Strategisch Basisonderzoek' grant no. 80031 and by a predoctoral fellowship to M.M.). J.P. is a Postdoctoral Fellow of the Research Foundation Flanders.

References

Adolph S, Bach S, Blondel M, Cueff A, Moreau M, Pohnert G, Poulet SA, Wichard T, Zuccaro A. 2004. Cytotoxicity of diatom-derived oxylipins in organisms belonging to different phyla. *The Journal of Experimental Biology* **207**: 2935–2946.

Alberti S, Gitler AD, Lindquist S. 2007. A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**: 913–919.

Benveniste P. 2004. Biosynthesis and accumulation of sterols. *Annual Review of Plant Biology* **55**: 429–457.

Berges JA, Franklin DJ, Harrison PJ. 2001. Evolution of an artificial seawater medium: improvements in enriched seawater, artificial water over the last two decades. *Journal of Phycology* **37**: 1138–1145.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP *et al.* 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**: 239–244.

Burden RS, Cooke DT, Carter GA. 1989. Inhibitors of sterol biosynthesis and growth in plants and fungi. *Phytochemistry* **28**: 1791–1804.

Campagnac E, Fontaine J, Lounès-Hadj Sahraoui A, Laruelle F, Durand R, Grandmougin-Ferjani A. 2009. Fenpropimorph slows down the sterol pathway and the development of the arbuscular mycorrhizal fungus *Glomus intraradices*. *Mycorrhiza* **19**: 365–374.

Cunningham FX, Gantt E. 2007. A portfolio of plasmids for identification and analysis of carotenoid pathway enzymes: *Adonis aestivalis* as a case study. *Photosynthesis Research* **92**: 245–259.

De Riso V, Raniello R, Maumus F, Rogato A, Bowler C, Falciatore A. 2009. Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Research* **37**: e96.

Desmond E, Gribaldo S. 2009. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biology and Evolution* **1**: 364–381.

Dufourc EJ. 2008. Sterols and membrane dynamics. *Journal of Chemical Biology* **1**: 63–77.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953–971.

Fabris M, Matthijs M, Rombauts S, Vyverman W, Goossens A, Baart GJE. 2012. The metabolic blueprint of Phaeodactylum tricornutum reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant Journal* **70**: 1004–1014.

Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C. 1999. Transformation of nonselectable reporter genes in marine diatoms. *Marine Biotechnology* **1**: 239–251.

Galea AM, Brown AJ. 2009. Special relationship between sterols and oxygen: were sterols an adaptation to aerobic life? *Free Radical Biology & Medicine* **47**: 880–889.

Gaulin E, Bottin A, Dumas B. 2010. Sterol biosynthesis in oomycete pathogens. *Plant Signaling & Behavior* **5**: 258–260.

Germann M, Gallo C, Donahue T, Shirzadi R, Stukey J, Lang S, Ruckenstuhl C, Oliaro-Bosso S, McDonough V, Turnowsky F *et al.* 2005. Characterizing sterol defect suppressors uncovers a novel transcriptional signaling pathway regulating zymosterol biosynthesis. *The Journal of Biological Chemistry* **280**: 35 904–35 913.

Giner J-L, Wikfors GH. 2011. "Dinoflagellate Sterols" in marine diatoms. *Phytochemistry* **72**: 1896–1901.

Greenspan P, Mayer EP, Fowler SD. 1985. Nile red: a selective fluorescent stain for intracellular lipid droplets. *The Journal of Cell Biology* **100**: 965–973.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L *et al.* 2009. InterPro: the integrative protein signature database. *Nucleic Acids Research* **37**: D211–D215.

Huysman MJJ, Fortunato AE, Matthijs M, Costa BS, Vanderhaeghen R, Van den Daele H, Sachse M, Inzé D, Bowler C, Kroth PG *et al.* 2013. AUREOCHROME1a-mediated induction of the diatom-specific cyclin *dsCYC2* controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *The Plant Cell* **25**: 215–228.

Hwang S, Rhee SY, Marcotte EM, Lee I. 2011. Systematic prediction of gene function in *Arabidopsis thaliana* using a probabilistic functional gene network. *Nature Protocols* **6**: 1429–1442.

Karp PD, Paley S, Romero P. 2002. The Pathway Tools software. *Bioinformatics* **18**: S225–S232.

Kodner R, Summons R. 2008. Sterols in a unicellular relative of the metazoans. *Proceedings of the National Academy of Sciences, USA* **105**: 9897–9902.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, Stanley MS, Kaplan A, Caron L, Weber T *et al.* 2008. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* **3**: e1426.

Laden BP, Tang Y, Porter TD. 2000. Cloning, heterologous expression, and enzymological characterization of human squalene monooxygenase. *Archives of Biochemistry and Biophysics* **374**: 381–388.

Lamb DC, Jackson CJ, Warrilow AGS, Manning NJ, Kelly DE, Kelly SL. 2007. Lanosterol biosynthesis in the prokaryote *Methylococcus capsulatus*: insight into the evolution of sterol biosynthesis. *Molecular Biology and Evolution* **24**: 1714–1721.

Leblond JD, Lasiter AD. 2012. Sterols of the green-pigmented, aberrant plastid dinoflagellate, *Lepidodinium chlorophorum* (Dinophyceae). *Protist* **163**: 38–46.

Liu Y, Yao Z-X, Papadopoulos V. 2005. Cytochrome P450 17α hydroxylase/17,20 lyase (CYP17) function in cholesterol biosynthesis: identification of squalene monooxygenase (epoxidase) activity associated with CYP17 in Leydig cells. *Molecular Endocrinology* **19**: 1918–1931.

Lohr M, Schwender J, Polle JEW. 2012. Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Science* **185–186**: 9–22.

Maheswari U, Jabbari K, Petit J-L, Porcel BM, Allen AE, Cadoret J-P, De Martino A, Heijde M, Kaas R, La Roche J *et al.* 2010. Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biology* **11**: R85.

Massé G, Belt ST, Rowland SJ, Rohmer M. 2004. Isoprenoid biosynthesis in the diatoms *Rhizosolenia setigera* (Brightwell) and *Haslea ostrearia* (Simonsen). *Proceedings of the National Academy of Sciences, USA* **101**: 4413–4418.

Miller MB, Haubrich B, Wang Q, Snell WJ, Nes WD 2012. Evolutionarily conserved $\Delta^{25(27)}$-olefin ergosterol biosynthesis pathway in the alga *Chlamydomonas reinhardtii*. *Journal of Lipid Research* **53**: 1636–1645.

Morikawa T, Mizutani M, Aoki N, Watanabe B, Saga H, Saito S, Oikawa A, Suzuki H, Sakurai N, Shibata D *et al.* 2006. Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato. *The Plant Cell* **18**: 1008–1022.

Moses T, Pollier J, Almagro L, Buyst D, Van Montagu M, Pedreño MA, Martins JC, Thevelein JM, Goossens A. 2014. Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16α hydroxylase from *Bupleurum falcatum*. *Proceedings of the National Academy of Sciences, USA* **111**: 1634–1639.

Nagumo A, Kamei T, Sakakibara J, Ono T. 1995. Purification and characterization of recombinant squalene epoxidase. *Journal of Lipid Research* **36**: 1489–1497.

Nes CR, Singha UK, Liu J, Ganapathy K, Villalta F, Waterman MR, Lepesheva GI, Chaudhuri M, Nes WD. 2012. Novel sterol metabolic network of *Trypanosoma brucei* procyclic and bloodstream forms. *The Biochemical Journal* **443**: 267–277.

Ohyama K, Suzuki M, Kikuchi J, Saito K, Muranaka T. 2009. Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **106**: 725–730.

Pearson A, Budin M, Brocks JJ. 2003. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proceedings of the National Academy of Sciences, USA* **100**: 15 352–15 357.

Plée-Gautier E, Antoun J, Goulitquer S, Le Jossic-Corcos C, Simon B, Amet Y, Salaün J-P, Corcos L. 2012. Statins increase cytochrome P450 4F3-mediated eicosanoids production in human liver cells: a PXR dependent mechanism. *Biochemical Pharmacology* **84**: 571–579.

Rampen SW, Abbas BA, Schouten S, Sinninghe Damsté J. 2010. A comprehensive study of sterols in marine diatoms (Bacillariophyta): implications for their use as tracers for diatom productivity. *Limnology and Oceanography* **55**: 91–105.

Raymond J, Blankenship RRE. 2004. Biosynthetic pathways, gene replacement and the antiquity of life. *Geobiology* **2**: 199–203.

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C. 2007. Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**: 23–35.

Summons RE, Bradley AS, Jahnke LL, Waldbauer JR. 2006. Steroids, triterpenoids and molecular oxygen. *Philosophical transactions of the Royal Society B* **361**: 951–968.

Sun Z, Cunningham FX, Gantt E. 1998. Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte. *Proceedings of the National Academy of Sciences, USA* **95**: 11 482–11 488.

Ta MT, Kapterian TS, Fei W, Du X, Brown AJ, Dawes IW, Yang H. 2012. Accumulation of squalene is associated with the clustering of lipid droplets. *The FEBS Journal* **279**: 4231–4244.

Tomazic ML, Najle SR, Nusblat AD, Uttaro AD, Nudel CB. 2011. A novel sterol desaturase-like protein promoting dealkylation of phytosterols in *Tetrahymena thermophila*. *Eukaryotic Cell* **10**: 423–434.

Van Heukelem L, Thomas CS. 2001. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *Journal of Chromatography* **910**: 31–49.

Vinci G, Xia X, Veitia RA. 2008. Preservation of genes involved in sterol metabolism in cholesterol auxotrophs: facts and hypotheses. *PLoS ONE* **3**: e2883.

Volkman JK. 2003. Sterols in microorganisms. *Applied Microbiology and Biotechnology* **60**: 495–506.

Weete J, Abril M, Blackwell M. 2010. Phylogenetic distribution of fungal sterols. *PLoS ONE* **5**: e10899.

Wentzinger LF, Bach TJ, Hartmann M-A. 2002. Inhibition of squalene synthase and squalene epoxidase in tobacco cells triggers an up-regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Plant Physiology* **130**: 334–346.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Screenshot of the JGI genome browser of *P. tricornutum* (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) corresponding to the erroneously predicted locus *PHATRDRAFT_44478* (in dark blue) on chromosome 4.

Fig. S2 Alignment of the AA sequences of the reconstructed IDI-SQS fusion enzyme of diatoms with the corresponding orthologues of *Aureococcus anophagefferens* and *Ectocarpus siliculosus*.

Fig. S3 The *PtOSC* gene.

Fig. S4 Alignment of the AA sequences of the reconstructed *OSC* gene models of diatoms with those of plants (*A. thaliana*), animals (*H. sapiens*) and fungi (*S. cerevisiae*).

Fig. S5 EI-MS spectra of the main peaks of the chromatograms shown in Fig. 7 and that of the corresponding authentic standards.

Table S1 Primers used

Table S3 List of candidate genes screened for SQE activity

Table S2 Orthology scores of *P. tricornutum* MVA and sterol pathway enzymes

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.