NatureMicrobiol_AltSQE

Figure1R

Figure2R

Figure3R

NatureMicrobiol_AltSQE_SI

A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis

Jacob Pollier[1,2], Emmelien Vancaester[1,2], Unnikrishnan Kuzhiumparambil[3], Claudia E. Vickers[4,5], Klaas Vandepoele[1,2], Alain Goossens[1,2*] and Michele Fabris[3,4]

[1]Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium. [2]VIB Center for Plant Systems Biology, 9052 Ghent, Belgium. [3]Climate Change Cluster, University of Technology Sydney, 2007 Ultimo, NSW, Australia. [4]CSIRO Synthetic Biology Future Science Platform, GPO Box 2583, Brisbane 4001, Australia. [5]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St Lucia, QLD, Australia.

*Correspondence to: alain.goossens@psb-vib.ugent.be.

Steroids are essential triterpenoid molecules that are present in all eukaryotes and modulate the fluidity and flexibility of cell membranes. Steroids also serve as signalling molecules crucial for growth, development and differentiation of multicellular organisms[1-3]. The steroid biosynthetic pathway is highly conserved and is key in eukaryote evolution[4-7]. The flavoprotein squalene epoxidase (SQE) catalyses the first oxygenation reaction in this pathway and is rate-limiting. However, despite its conservation in animals, plants, and fungi, several phylogenetically widely-distributed eukaryote genomes lack an SQE-encoding gene[7,8]. Here we discovered and characterized an alternative SQE (AltSQE) belonging to the fatty acid hydroxylase superfamily. AltSQE was identified through screening of a gene library of the diatom *Phaeodactylum tricornutum* in a SQE-deficient yeast. In accordance with its divergent protein structure and need for co-factors, we found that AltSQE is insensitive to the conventional SQE inhibitor terbinafine. *AltSQE* is present in many eukaryotic lineages but is mutually exclusive with *SQE* and shows a patchy distribution within monophyletic clades. Our discovery provides an alternative element for the conserved steroid biosynthesis pathway, raises questions about eukaryote metabolic evolution, and opens routes to develop selective SQE inhibitors to control hazardous organisms.

Eukaryotes synthesize and utilize a wide variety of steroids. The principle steroid in animal cell membranes is cholesterol, which is also the precursor of steroid hormones such as testosterone, estradiol and progesterone. In fungi, ergosterol represents the main steroid, whereas plant steroids are more diverse with sitosterol, stigmasterol and the brassinosteroid phytohormones as the principle compounds[1-3]. In protists, including diatoms, the variety of steroids is wider[9], albeit less characterized. They are often involved in programmed cell death and algal bloom regulation[10].

The precursor of all steroids is the linear 30-carbon terpenoid squalene. Prior to cyclization into lanosterol (animals, fungi) or cycloartenol (plants), squalene is oxidized at one of its terminal double bonds by the flavoprotein squalene epoxidase (SQE), also known as squalene monooxygenase (SQMO) (Fig. 1a). The incorporation of molecular oxygen in the steroid backbone by this enzyme is a distinctive and universal feature of eukaryotic steroid biosynthesis. In contrast, biosynthesis of the bacterial analogues, the hopanoids, does not require oxygen. Therefore, steroids are considered a marker of eukaryotic life, and for this reason have been used to identify and date evolutionary steps. It is generally accepted that steroid biosynthesis originated in the Proterozoic, possibly emerging during atmospheric and oceanic oxygenation by cyanobacterial oxygenic photosynthesis around 2.3 billion years ago[4-6]. This also implies that the core steroid biosynthesis pathway, including SQE, was already present in the last common ancestor of eukaryotes[4,7]. Because of their essential role, steroid levels are precisely regulated in eukaryotic cells, often by conserved molecular mechanisms that control levels and activities of steroid biosynthesis enzymes in the endoplasmic reticulum (ER), including the rate-limiting enzymes 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) and SQE[11-13]. However, despite the conserved and essential role of SQE, analysis of sequenced genomes suggested its absence in several eukaryotic species distributed across the tree of life, including the model diatom species *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*[7,8].

A knockout mutation of *ERG1*, the gene encoding SQE, leads to ergosterol auxotrophy in yeast (*Saccharomyces cerevisiae*) strain JP064[14]. To identify functional SQEs from species lacking the conventional SQE, such as *P. tricornutum*, a cDNA library[15] was screened for genes capable of complementing the *erg1*-knockout in the haploid yeast strain JP064 grown on selective medium lacking ergosterol. Of the four yeast colonies surviving selection, three contained a plasmid with the entire open reading frame (ORF) of the *Phatr3_J45494* gene,

which was renamed *Alternative squalene epoxidase* (*AltSQE*; GenBank Accession Number MH422131). To confirm its role as a SQE, the *AltSQE* ORF sequence was sub-cloned and re-expressed in the JP064 strain, as well as in the yeast strain PA123[14]. The latter strain contains an additional knockout mutation in the *ERG7* gene, encoding lanosterol synthase, the enzyme catalysing the step immediately following SQE in the ergosterol biosynthesis pathway (Fig. 1a). This strain accumulates high levels of squalene but does not convert oxidosqualene to lanosterol; this allows identification of SQE activity by measurement of 2,3-oxidosqualene upon expression of a functional SQE[14]. In JP064 cells, growth on selective medium lacking ergosterol was restored by expression of either *AltSQE* or yeast *ERG1*, as a positive control (Fig. 1b). In PA123 cells, expression of either *AltSQE* or *ERG1* led to reduced squalene accumulation in parallel with the production of 2,3-oxidosqualene and minute amounts of 2,3-22,23-dioxidosqualene (Fig. 1c,d), demonstrating the SQE activity of AltSQE.

The existence of a non-conventional SQE in *P. tricornutum* had previously been suggested based on a pharmacological analysis that demonstrated insensitivity of this organism to terbinafine, the well-known inhibitor of the conventional SQE enzyme[8]. Accordingly, expression of *AltSQE* in the wild-type yeast strain PA059[16] alleviated terbinafine sensitivity of this strain, in contrast with overexpression of *ERG1* (Fig. 1e), indicating that the AltSQE enzyme is indeed insensitive to this drug.

Conventional SQE is a flavoprotein monooxygenase; in contrast, AltSQE belongs to the fatty acid hydroxylase superfamily (Supplementary Fig. 1a). This family of integral membrane proteins contains enzymes that catalyse oxygen-dependent desaturation, hydroxylation, and epoxidation reactions of primarily lipid-based substrates[17,18]. Hydropathy plot and protein modelling suggested a membranous nature of the AltSQE protein (Supplementary Fig. 1b,c). To verify the subcellular localization, stable transformed *P. tricornutum* cells producing the AltSQE C-terminally fused with the mVenus yellow fluorescent protein (YFP) were

generated and examined by confocal microscopy. The YFP signal was observed around the chloroplast (Supplementary Fig. 2a), which is characteristic for ER-localized proteins in this organism[19]. Similarly, when produced transiently in *Agrobacterium*-infiltrated *Nicotiana benthamiana* leaves, green fluorescent protein-tagged AltSQE co-localized with a red fluorescent ER-marker (Supplementary Fig. 2b). The observed ER-localization of AltSQE is consistent with the general subcellular location of its substrate squalene, and corresponds to the subcellular localization of the conventional SQE enzyme in animals, plants, and fungi[12,20,21]. To investigate the role of *AltSQE* in *P. tricornutum*, we attempted to alter its expression by overexpression or RNAi-mediated silencing. However, the transformed diatom lines showed no differences in *AltSQE* expression and no marked shifts in the accumulation of the major steroids or their precursors were observed (Supplementary Fig. 3 and Supplementary Fig. 4). Together, this points to a robust control of AltSQE levels and activity, which is not surprising given the essential role of steroids in eukaryotes.

A common characteristic of proteins belonging to the fatty acid hydroxylase superfamily is the presence of nine conserved histidine residues that coordinate a dimetal centre at the active site of the enzyme[18,22]. These residues are also conserved in AltSQE (Supplementary Fig. 1b and Supplementary Fig. 5) and substitution of any of them with an alanine residue reduced the enzymatic activity of the protein (Supplementary Fig. 6), confirming their importance for catalytic activity. The molecular oxygen required for the oxidative reaction is activated by the dimetal centre (Supplementary Fig. 7) that therefore needs to be in a reduced state[17]. The electrons needed for reduction of the dimetal centre are typically provided by cytochrome b_5, which in turn receives the electrons from NADH, usually through cytochrome b_5 reductase proteins[17,18,22]. Two candidate cytochrome b_5 encoding genes, *Phatr3_J30770* and *Phatr3_J16195* (GenBank Accession Numbers MH422132 and MH422133, respectively), were identified in the *P. tricornutum* genome and their capacity to physically interact with

AltSQE was assessed using a yeast-based split-ubiquitin interaction assay[23]. Direct interaction between AltSQE and the Phatr3_J30770 protein, but not the Phatr3_J16195 protein, was observed (Supplementary Fig. 6). This suggests that AltSQE functions by reduction of the dimetal centre using electrons provided by cytochrome b_5, followed by activation of molecular oxygen by the reduced dimetal centre and epoxidation of squalene (Supplementary Fig. 7). In contrast, the conventional SQE functions by loosely binding a flavin adenine dinucleotide (FAD) cofactor that gets reduced by a NADPH-cytochrome P450 reductase, followed by activation of molecular oxygen by the reduced FAD and epoxidation of squalene via an electrophilic addition reaction (Supplementary Fig. 8). Cytochromes b_5 can transfer electrons to a large number of different oxidases, including fatty acid hydroxylases and cytochromes P450. Phatr3_J30770 is therefore not likely to be specific to AltSQE, but rather a general co-factor interacting with oxidases that reside within the ER. Likewise, the membrane-bound cytochrome b_5 in *S. cerevisiae*, CYB5, could be capable of electron transfer to AltSQE, thus ensuring its activity in yeast.

Interestingly, the conventional *SQE* gene was not detectable in the genomes of several sequenced stramenopiles (heterokonts), including other diatoms (Bacillariophyta) besides *P. tricornutum* and the brown alga (Pelagophyceae) *Aureococcus anophagefferens*[7,8]. A protein search against the NCBI nr database was therefore performed. This revealed the presence of AltSQE homologs within several other diatom species, including *T. pseudonana* and *Fragilariopsis cylindrus*. AltSQE-like sequences could also be identified in the genomes of species other than diatoms, including the coccolithophore (Haptophyta) *Emiliania huxleyi*, the cryptomonad (Cryptophyta) *Guillardia theta* and, surprisingly, *E. huxleyi* viruses. To further assess this seemingly wider distribution of AltSQEs, a hidden Markov model (HMM) was built for both SQE and AltSQE to carry out HMM profile searches against gene catalogues from 667 different species encompassing all eukaryotic clades of the tree of life

(419 genomes and 661 transcriptomes, Supplementary Data 1). This analysis revealed that, just like the SQE proteins, the AltSQE proteins are indeed widely dispersed among the tree of life (Fig. 2), covering 235 and 190 different species, respectively. We could not identify any organism harbouring both SQE and AltSQE, suggesting they are mutually exclusive (Supplementary Data 1). This suggests an evolutionary disadvantage of having both genes present, or an advantage of having only one gene present, notwithstanding that our yeast complementation assays (Fig. 1e) indicate that concurrent presence of both SQE types is not lethal.

To confirm that the identified AltSQE sequences correspond to functional AltSQE enzymes, candidate AltSQEs from eight representative species were selected for further analysis: *F. cylindrus* (Bacillariophyta), *T. pseudonana* (Bacillariophyta), *Salpingoeca rosetta* (Choanoflagellida), *G. theta* (Cryptophyta), *Bigelowiella natans* (Cercozoa), *Symbiodinium minutum* (Dinophyta), *E. huxleyi* (Haptophyta), and *E. huxleyi* virus 84. For all, *S. cerevisiae* codon-optimized gene sequences were synthesized, which, when expressed in yeast strains JP064 and PA123, all complemented the *erg1*-knockout mutation (Fig. 3a) and led to production of 2,3-oxidosqualene (Fig. 3c), respectively. Likewise, all eight alleviated the terbinafine sensitivity of yeast strain PA059 (Fig. 3b). Together, this confirms the SQE activity of all tested enzymes and implies that insensitivity to terbinafine is common to all AltSQE proteins.

The discovery of a widespread alternative enzyme that catalyses the epoxidation of squalene has profound implications on our understanding of the evolution of eukaryotic life. The patchy distribution of the mutually exclusive conventional and alternative SQEs across the eukaryotic tree of life (Fig. 2 and Supplementary Data 2,3) seems incompatible with vertical inheritance from a last common eukaryotic ancestor and alludes either to multiple eukaryote-to-eukaryote horizontal gene transfer (HGT) events or to a long-term coexistence of both

genes followed by several independent losses of either the conventional or the alternative SQE. The presence of a functional AltSQE in *E. huxleyi* viruses may hint at HGT[24], however, this is not supported by our phylogenetic analysis. SQE and AltSQE sequences roughly follow a phylogenetic distribution corresponding to the species distribution in the tree of life, but the relationship is imperfect. Subsequently, we cannot discriminate between the coexistence-followed-by-loss-hypothesis and the more complex scenario of several independent (eukaryotic) HGT events followed by replacement (Fig. 2 and Supplementary Data 2,3). Such a scattered distribution of functional analogues has been observed only a few times before in eukaryotes, but thus far only for paralogues with a relatively high amino acid identity between the two counterparts[25,26]. In contrast, here convergent evolution gave rise to two completely unrelated enzymes carrying out the same essential metabolic reaction. As the conventional SQE was hitherto considered part of the highly conserved core steroid biosynthesis pathway in eukaryotic life[4,7], the discovery of a widespread AltSQE prompts the need to revisit currently accepted notions on the evolution of steroid biosynthesis. Our findings may also open potential routes to develop selective AltSQE inhibitors to control hazardous organisms depending on AltSQE activity, such as dinoflagellate and diatom species that produce neurotoxins leading to various seafood poisoning syndromes in humans[27,28]. Finally, it is known that *E. huxleyi* viruses hijack lipid synthesis pathways of their host for lytic infection [24]; possibly they may likewise use AltSQE to modulate the host sterol synthesis pathway to promote lytic infection. Since *E. huxleyi* infection by *E. huxleyi* viruses may be a key contributor to the formation of sea spray aerosols, cloud formation and ultimately Earth's climate system[29], AltSQE may play an important role in marine cloud formation and climate regulation.

Methods

cDNA library preparation. 239 µL of a custom-made *Phaeodactylum tricornutum* cDNA library[15] with an original titre of 4.5 x 10^6 to 7 x 10^6 cfu/mL was used to inoculate 100 mL of Lysogeny Broth (LB) medium containing 50 mg/mL kanamycin. The resulting *Escherichia coli* culture was incubated for 8 h at 37°C with shaking at 300 rpm. Plasmid DNA was isolated from the entire culture (yield: 25 µg) and used for transfer of the cDNA library to the constitutive high-copy yeast expression vector pAG423GPD-ccdB (AddGene plasmid # 14150). To this end, 150 ng of library plasmid DNA was mixed with 1.35 µg of pAG423GPD-ccdB plasmid DNA and 6 µL of LR clonase in a total volume of 20 µL. After overnight incubation at 25°C, the LR reaction was stopped by adding 2 µL of proteinase K, followed by 15 min of incubation at 37°C and 10 min at 75°C, after which the DNA was precipitated by adding 80 µL of sterile water, 1 µL of glycogen (20 µg/µL), 50 µL of 7.5 M of ammonium acetate and 375 µL of 100% ethanol. The resulting mixture was incubated overnight at -20°C. DNA was pelleted by centrifugation at 20,800 x g at 4°C for 10 min, after which the DNA pellet was washed with 150 µL of 70% ethanol. The resulting DNA was briefly dried under vacuum and dissolved in a final volume of 12 µL of H_2O. This DNA was transformed into *E. coli* ElectroMAX™ DH10B™ T1 Phage Resistant Cells (Invitrogen) via electroporation using 0.1-cm electroporation cuvettes and a Gene Pulser electroporation system (BioRad) with the following settings: 2.2 kV; 200 Ω; 25 µF. Six aliquots of 50 µL of competent cells were transformed, each with 2 µL of precipitated LR. After electroporation, 1 mL of SOC medium was added to each aliquot, after which the cells were allowed to recover for 1 h at 37°C. The resulting vials were pooled in a 15-mL Falcon tube and an equal amount of sterile freezing medium (60% SOC, 40% glycerol) was added. Finally, the cells with the transferred library were divided in 1-mL aliquots and frozen at -70°C. Prior to dividing the cells, a 200-µL aliquot was used for titre determination (obtained titre: 1.8 x 10^5 cfu/mL).

cDNA library screening. 1 x 10^6 cfu of the transferred cDNA library were used to inoculate 100 mL of LB medium containing 100 mg/mL carbenicillin. The resulting *E. coli* culture was incubated for 10 h at 37°C with shaking at 300 rpm after which plasmid DNA was isolated (yield: 227 µg) and used to transform yeast strain JP064[14] *via* electroporation. To this end, a 100-mL yeast culture grown in yeast extract peptone dextrose (YPD) medium supplemented with ergosterol (20 µg/mL), hemin (13 µg/mL) and Tween 80 (5 mg/mL) to OD 1.0 was harvested by centrifugation and the yeast cells were repeatedly washed with 25 mL of water, 2 mL of 1 M sorbitol, 2 mL of 0.1 M lithium acetate and 2 mL of 1 M sorbitol. Finally, the cells were dissolved in 500 µL of 1 M sorbitol. Aliquots of 100 µL of cells were transferred to a cooled 0.2-cm electroporation cuvette and 2.5 µg of plasmid DNA was added to each aliquot. Electroporation was carried out using a GenePulser electroporation system (Biorad) with the following settings: 1.5 kV, 600 Ω, 25 µF. Electroporated cells were allowed to recover for 1 h in 1 mL of YPD medium supplemented with ergosterol (20 µg/mL), hemin (13 µg/mL) and Tween 80 (5 mg/mL) after which they were washed with 1 mL of 1 M sorbitol and finally plated on selective plates containing Synthetic Defined (SD) medium (Clontech) supplemented with minus histidine amino acid dropout mix (SD-His) and hemin (13 µg/mL). A transformation control was plated on non-selective SD-His plates containing ergosterol (20 µg/mL), hemin (13 µg/mL) and Tween 80 (5 mg/mL). The resulting yeast plates were incubated at 30°C for seven days, after which the seven colonies growing on the selective plates were subcultured to a new selective plate. Only four of these colonies were viable and subsequently used for plasmid DNA extraction using a Zymoprep Yeast Plasmid Miniprep kit (Zymo research) according to the manufacturer's instructions. The resulting plasmids were transferred to *E. coli* for propagation, isolated and Sanger sequenced.

Gene cloning. The full-length ORF of *P. tricornutum AltSQE* was PCR-amplified with and without stop-codon from a Sanger-sequenced plasmid resulting from the library screening.

Likewise, the full-length ORFs of the *P. tricornutum* cytochrome b_5 genes were amplified with and without stop-codon from the original library prep. All primers used for gene cloning are listed in Supplementary Table 1. *S. cerevisiae* codon-optimized (https://eu.idtdna.com/CodonOpt) *AltSQE* gene sequences for *F. cylindrus* (GenBank Accession Number MH422141), *T. pseudonana* (GenBank Accession Number MH422143), *S. rosetta* (GenBank Accession Number MH422142), *G. theta* (GenBank Accession Number MH422135), *B. natans* (GenBank Accession Number MH422137), *S. minutum* (GenBank Accession Number MH422139), *E. huxleyi* (GenBank Accession Number MH422140) and its virus *E. huxleyi* virus 84 (GenBank Accession Number MH422144) were synthesized including AttB-sites using a BioXp 3200 DNA synthesis machine. The obtained PCR and synthesis fragments were Gateway-recombined into the donor vector pDONR207 and sequence-verified. Mutations in the conserved histidine residues of *P. tricornutum AltSQE* were introduced using overlap-extension PCR with the primers listed in Supplementary Table 1. For *G. theta*, the available *AltSQE* gene model was incorrect. To find the full-length ORF of the *G. theta AltSQE* gene, a *G. theta* RNA-Seq library (Illumina HiSeq200; 365,495,598 paired-end reads, 2 x 100 bp) was obtained from the NCBI Short Read Archive (Accession number SRR747855) and used for a *de novo* transcriptome assembly using the CLC Genomics Workbench 7.0.4 (CLC bio, Aarhus, Denmark) with default parameters. As such, 34,201 contigs with an average size of 1,241 bp were obtained. This assembly was used to make a BLAST database, and a TBLASTX search with the original *G. theta* AltSQE revealed only one hit, a contig encoding a protein corresponding to *G. theta* AltSQE, but which is 12 amino acids longer at its C-terminus, compared to the available GenBank sequence. The synthesized clone was corrected using two additional reverse primers (Supplementary Table 1) and the corrected sequence was deposited in the GenBank under accession number MH422134. Also for *B. natans* and *S. minutum*, the available gene models

were incorrect. Like for *G. theta*, RNA-Seq data was downloaded from the NCBI Short Read Archive (Accession numbers SRR1296871 (*B. natans*) and DRR003868 (*S. minutum*)) and used for a *de novo* transcriptome assembly using the CLC Genomics Workbench 7.0. with default parameters. TBLASTX searches in the assembled transcriptomes with the *P. tricornutum* AltSQE as query revealed the full-length open reading frames of the *AltSQE* homologues of both species. The correct sequences were synthesized and deposited in the GenBank under accession numbers MH422136 (*B. natans*) and MH422138 (*S. minutum*).

Complementation assays in strain JP064. The entry clones containing the full-length ORFs of the investigated genes were Gateway recombined with the constitutive high-copy yeast expression vector pAG423GPD-ccdB (AddGene plasmid # 14150). The resulting plasmids were transformed into strain JP064[14] *via* electroporation as described above. Transformed yeast cells were plated on non-selective SD-His plates supplemented with ergosterol (20 µg/mL), hemin (13 µg/mL) and Tween 80 (5 mg/mL). For the complementation assay, cells were precultured for two days in SD-His medium supplemented with ergosterol (20 µg/mL), hemin (13 µg/mL) and Tween 80 (5 mg/mL), after which they were serially diluted and dropped on selective SD-His plates supplemented with hemin (13 µg/mL) only. As growth control, the cells were dropped on non-selective SD-His plates supplemented with ergosterol (20 µg/mL), hemin (13 µg/mL) and Tween 80 (5 mg/mL). The resulting yeast plates were incubated at 30°C for seven days. For each tested gene, at least three individual transformants were assessed. For each assay, a positive (ERG1) and negative (empty vector) control was included.

Functional analysis in strain PA123. The full-length ORFs of the investigated genes were Gateway recombined into the constitutive high-copy yeast expression vector pAG426GPD-ccdB (AddGene plasmid # 14156) and transformed into yeast strain PA123[14]. Transformed yeast cells were plated on SD medium supplemented with minus uracil amino acid dropout

mix (SD-Ura), ergosterol (20 μg/mL), hemin (13 μg/mL) and Tween 80 (5 mg/mL). For GC-MS analysis, cells were cultured for three days in SD-Ura supplemented with ergosterol (20 μg/mL), hemin (13 μg/mL) and Tween 80 (5 mg/mL). Four mL of each culture was harvested by centrifugation. Yeast cells were broken by freezing the cells in liquid nitrogen, adding metal balls (1 ø5mm and 2 ø3mm balls) and milling using a RETCH ball mill (2 times 30 s at 30 Hz). Broken cells were extracted with 1 mL of methanol for 30 min at room temperature. After centrifugation, the methanol phase was collected and evaporated under vacuum. The resulting residue was dissolved in 200 μL of heptane, of which 1 μL was used for GC-MS analysis. GC-MS analysis was carried out using a GC model 6890 and MS model 5973 (Agilent, Santa Clara, United States) as previously described[30]. Products (squalene, 2,3-oxidosqualene, 2,3:22,23-dioxidosqualene and ergosterol) were identified using authentic standards and peak areas were determined using the ChemStation software (Agilent).

Terbinafine assays. The full-length ORFs that were Gateway recombined into the constitutive high-copy yeast expression vector pAG423GPD-ccdB (AddGene plasmid # 14150) were transformed into yeast strain PA059[16]. Transformed yeast cells were plated on SD-His medium. For the terbinafine assay, cells were precultured for one day in SD-His, after which they were serially diluted and dropped on selective SD-His plates containing 600 μM of terbinafine. As growth control, the cells were dropped on SD-His plates without terbinafine. The resulting yeast plates were incubated at 30°C for three days. For each tested gene, at least three individual transformants were assessed.

Subcellular localization in *N. benthamiana.* The *AltSQE* gene was Gateway recombined into the expression vectors pK7WGF2 and pK7FWG2 for N- and C-terminal GFP-fusion, respectively. The resulting constructs were transformed into *Agrobacterium tumefaciens* strain C58C1, carrying the pMP90 helper plasmid. To suppress gene silencing, *A. tumefaciens* carrying a *35S:p19* construct was used. The ER-rk[31] construct served as a red fluorescent ER-

localization control. For *N. benthamiana* infiltrations, *A. tumefaciens* was grown for two days in a shaking incubator (150 rpm) at 28°C in 5 mL of yeast extract broth (YEB) medium, supplemented with appropriate antibiotics (100 µg/mL spectinomycin, 300 µg/mL streptomycin and 20 µg/mL gentamycin for localization constructs and 25 µg/mL kanamycin and 20 µg/mL gentamycin for the p19 and ER-rk constructs). After incubation, 500 µL of bacterial culture was used to inoculate 9.5 mL of YEB medium supplemented with appropriate antibiotics, and containing 10 mM MES (pH 5.7) and 20 µM acetosyringone. After overnight incubation in a shaking incubator (150 rpm) at 28°C, bacteria for transient co-expression were mixed, collected via centrifugation and resuspended in 5 mL of infiltration buffer (100 µM acetosyringone, 10 mM $MgCl_2$, 10 mM MES, pH 5.7). The amount of bacteria harvested for each construct was adjusted to obtain a final OD_{600} of 1.0 in the infiltration buffer. After 2 to 3 h of incubation at room temperature, the bacteria mixtures were infiltrated to the abaxial side of fully expanded leaves of 3- to 4-week-old *N. benthamiana* plants grown at 25°C under a 14-h/10-h light/dark regime. The infiltrated plants were incubated under normal growth conditions for three days prior to confocal microscopy analysis. Fluorescence was analysed with a Zeiss LSM 710 confocal laser scanning microscope. Dual GFP and RFP fluorescence was imaged with 488- and 543-nm light for GFP and RFP excitation, respectively.

Split-ubiquitin interaction assay. Because both AltSQE and cytochrome b_5 are membrane-bound proteins, a split-ubiquitin interaction assay is required to show interaction of these proteins in yeast. To this end, an assay based on the ubiquitin-based split-protein sensor (USPS) system by Johnsson & Varshavsky[23] was used, in which the AltSQE protein served as bait and the cytochromes b_5 proteins as prey. AltSQE was fused to the C_{Ub}-URA3 fusion by Gateway recombination of the entry clone without stop codon into the bait vector pMKZ, a Gateway compatible version of the C_{Ub}-URA3 fusion vector[32]. The cytochrome b_5 ORFs were

fused to the N-terminal half of ubiquitin by Gateway recombination of the entry clones without stop codon in the prey vectors pCup-NuI-GWY-myc-CYC1 and pCup-KZ-myc-GWY-NuI-CYC1. The obtained expression clones were sequence-verified using the primers listed in Supplementary Table 1. The split-ubiquitin assay was carried out in *S. cerevisiae* strain JD53[32]. First, pMKZ-AltSQE was transformed into JD53 and transformed cells were plated on SD-His medium. To check the stability of the AltSQE-C_{Ub}-URA3 fusion protein, colonies growing on the SD-His plate were streaked on a SD-His-Ura plate. Next, yeast growing on the SD-His-Ura plate was transformed with the prey vectors containing the cytochrome b_5 genes or the empty vectors. Transformed yeast cells were selected on SD-His-Trp plates. For each construct, four colonies were selected and cultured for one day in SD-His-Trp-Met medium, after which they were serially diluted and dropped on selective SD-His-Trp-Met plates supplemented 100 µM of $CuSO_4$ and 1 mg/mL 5-fluoroorotic acid (5-FOA). As growth control, the cells were dropped on SD-His-Trp-Met plates supplemented 100 µM of $CuSO_4$ without 5-FOA. The resulting yeast plates were incubated at 30°C for two days. Yeast cells capable of growing on plates containing 5-FOA are indicative of ubiquitin-mediated URA3 degradation, and hence of interaction between AltSQE and the tested cytochrome b_5.

Generation of transgenic *P. tricornutum* lines. The inverted-repeat RNAi construct targeting *P. tricornutum AltSQE (Phatr3_J45494)* was cloned into the pKS-Sh ble-FA plasmid as described[33] using primers listed in Supplementary Table 1, yielding the construct FcpBp-*AltSQE*-FcpA3'. The pKS-Sh ble-FA GUS and pAF6 plasmids were kindly provided by Dr. Angela Falciatore (IBPS, Paris, France). *P. tricornutum* strain CCAP 1055/1 *wild-type* cells were transformed with either *AltSQE*-RNAi or empty pAF6 control vectors by biolistic transformation with a PDS-1000/He System with Hepta adapter (Bio-Rad, USA). Transformants were selected on half-strength ESAW[34]–agar (1%) medium containing

100 µg/mL Zeocin (InvivoGen, USA) and subsequently maintained in liquid ESAW medium supplemented with 50 µg/mL Zeocin. The AP1p-*AltSQE-mVenus* and AP1p-*mVenus* constructs were generated by Gibson assembly® into the pPTPBR11 episome backbone (Addgene plasmid #80386). Briefly, single fragments corresponding to the *P. tricornutum* AP1 (*Phatr3_J49678*) promoter[35], *AltSQE*, *mVenus*[36] and the *FcbpF (Phatr3_J51230)* terminator region were PCR-amplified using the primers listed in Supplementary Table 1. The obtained PCR fragments were purified and assembled on the pPTPBR11 episome backbone (Addgene plasmid #80386) upon linearization with *Sbf*I-HF (New England Biolabs, USA). The FcpBp-*AltSQE-mVenus* and FcpBp-*mVenus* constructs were generated by replacing the AP1 promoters with the FcpBp (*Phatr3_J18049*) promoter between the *SpeI* and *NdeI* sites previously inserted in the abovementioned constructs. Correct episome assemblies were confirmed by diagnostic restriction digestion and Sanger sequencing. Next, episomes were co-transformed with the pTA-MOB[37] mobility plasmid, kindly provided by Dr. Ian Monk (University of Melbourne, Australia) in *E. coli* strain Epi300 and used for *P. tricornutum* conjugation as described[38]. Exconjugants were selected on half-strength ESAW plates (1% agar) containing 100 µg/mL Zeocin and subsequently maintained in liquid ESAW medium supplemented with 50 µg/mL Zeocin. Colonies expressing the fusion proteins AltSQE-mVenus and mVENUS were screened Beckman Coulter's CytoFLEX LX Flow Cytometer (Beckman Coulter, USA).

Subcellular localization in *P. tricornutum*. Hundred µL of transgenic diatom cultures expressing the AP1p-*AltSQE-mVenus* or AP1p-*mVenus* constructs were cultivated for three days in ESAW medium, centrifuged, washed twice and finally resuspended in 100 µl of ESAW medium lacking phosphate and incubated for 20 hours. Next, cells were mounted on slides and imaged with 100x oil lens on a confocal laser scanning Nikon A1 Plus microscope

(Nikon, Japan) using the 488 and 637 nm lasers for mVENUS and chlorophyll autofluorescence, respectively.

Sterol perturbation experiments in transgenic *P. tricornutum*. Three independent transgenic diatom lines harbouring the constructs *AltSQE* knock-down, pAF6, FcpBp-*AltSQE*-*mVenus*, and FcpBp-*mVenus* were grown in triplicate in 1-L baffled shake flasks (Sigma-Aldrich) in 150 mL of ESAW medium[34] in fully controlled incubators (Kühner, Switzerland) under continuous light (150 µE m^{-2} s^{-1}) at 21˚C for four days to late exponential phase. Hundred mL of diatom culture was harvested by centrifugation at 4500 g for 5 min and washed with 2 mL of PBS. The resulting pellet was flash frozen and lyophilized. Freeze-dried algal biomass was extracted in fresh 10% KOH ethanolic (50%) solution at 80°C for 1 h in a heat block with occasional stirring. After cooling, 1 mL of saturated NaCl solution and 400 µL of hexane (Sigma-Aldrich, Castle Hill, Australia) was added to the samples. After thoroughly mixing the samples on a roller shaker for 3 min, the mixture was allowed to settle and the organic phase was collected. The extraction with hexane was repeated twice and the pooled organic layer was dried under nitrogen and derivatized using 50 µL of *N,O*-Bis(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich) by heating at 70°C for 60 min, followed by addition of 100 µL of hexane. GC-MS analysis was performed on an Agilent 7890 (Agilent Technologies, Santa Clara, California, USA) series GC equipped with a HP-5 capillary column (30 m; 0.25 mm i.d., film thickness 0.25 µm) coupled to an Agilent quadrupole MS (5975 N) instrument. The oven temperature was initially set at 50°C with a gradient from 50 to 250°C (15.0°C/min), from 250 to 310°C (8°C/min) and a 10 min hold. The injector temperature was set at 250°C, carrier gas (helium) flow at 0.9 mL/min and splitless injection was used with a purge time of 1 min. The injection volume was 2 µL. Mass spectrometer operating conditions were as follows: ion source temperature 230°C, quadrupole temperature 150°C and ionization voltage 70 eV. Mass analysis was performed with an

electron impact (EI) source, under scan mode from 50 to 700 m/z. Sterol peaks were identified using authentic standards (Sigma-Aldrich). Sterol concentrations were calculated using five-point calibration curves based on known standard concentrations. 5-α-cholestane (Sigma-Aldrich) was used as an internal standard.

***AltSQE* gene expression analysis in transgenic diatoms.** Expression of *AltSQE* was assessed by qRT-PCR analysis. Transgenic *P. tricornutum* cells from 25 mL cultures were harvested by centrifugation at 4500 g for 5 min and washed with 2 mL of PBS. The resulting pellet was flash frozen and RNA was subsequently extracted with the Illustra RNAspin Mini Kit (GE Healthcare Life Sciences, USA). 500 ng of RNA was used for cDNA synthesis using iScript (Bio-Rad). qRT-PCR was carried out with SYBR Green QPCR Master Mix (Biorad, USA) in a CFX384™ Real-Time System (BioRad, USA). The reference genes used for normalization were RPS and TubB[39]. qRT-PCR primers (Supplementary Table 1) were designed with Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/). Gene expression was analyzed in regression mode using the CFX Manager™ 3.1 Software (BioRad, USA) by interpolation with standard curves of *AltSQE*, *RPS* and *TubB* (cDNA dilution gradient of 243, 81, 27, 9 and 3 ng). *AltSQE* relative transcript quantity was normalized to the geometrical mean of the quantity of *RPS* and *TubB* transcripts.

Squalene epoxidase distribution pattern. Homologs of the *P. tricornutum* AltSQE protein sequence were retrieved via BLASTP searches against the NCBI nr database (http://www.ncbi.nlmh.nih.gov/). Together with the *P. tricornutum* protein query, the *Fistulifera solaris*, *F. cylindrus*, *A. anophagefferens*, *T. pseudonana*, and *E. huxleyi* AltSQE homologues were aligned using MAFFT L-INS-i 7.187[40]. The aligned sequences were used as a seed alignment for building a hidden Markov model (HMM) using hmmer3 v3.1b2[41]. Also for the conventional SQE, a HMM profile was built, using sequences from *Arabidopsis thaliana*, *Cyanidioschyzon merolae*, *Galdieria sulphuraria*, *Ectocarpus siliculosus*,

436 *Nannochloropsis gaditana*, *S. cerevisiae*, *Schizosaccharomyces pombe*, and *Homo sapiens*.
437 The predicted proteomes for 419 species where downloaded from http://protists.ensembl.org,
438 http://plants.ensembl.org, http://metazoa.ensembl.org and http://www.ensembl.org. For the
439 dinophyte *S. minutum*, sequences were retrieved from
440 http://marinegenomics.oist.jp/symb/viewer/download?project_id=21. Re-assembled
441 MMETSP[42] transcriptomes were retrieved from
442 https://figshare.com/articles/Marine_Microbial_Eukaryotic_Transcriptome_Sequencing_Proje
443 ct_re-assemblies/3840153/6. To complement clades not well represented in the tree of life, the
444 transcriptomes of four Centrohelida, one Apusozoa[43], and *Euglena gracilis*[44], available from
445 https://doi.org/10.5061/dryad.rj87v and GenBank accession GDJR00000000.1, were
446 downloaded and further processed into proteomes using TRAPID[45]. The aforementioned SQE
447 and AltSQE HMM profiles were used to screen the proteomes using an e-value cut-off of
448 10^{-10}. To assert that absence of a protein was not due to an annotation mistake, all raw genome
449 sequences were screened with the SQE and AltSQE profiles using AUGUSTUS-PPX[46]. SQE
450 distribution patterns were compared with the proposed tree of life topology as in[43]; the results
451 are summarized in Fig. 2 and all detected (Alt)SQE sequences are provided in Supplementary
452 Data 1. The HMM profile which was initially used to screen eukaryotic genomes and
453 MMETSP transcriptomes was used to scan the Ocean Microbial Reference Gene Catalog
454 (OM-RGC) that contains all prokaryotic assembled genes from the large-scale TARA oceans
455 project[47,48], however, no prokaryotic AltSQE homologues were identified. To identify viral
456 homologues, the *E. huxleyi* virus 86 protein EhV088[49], which was detected in the initial
457 BLASTP search, was aligned with all detected AltSQE homologues to create a novel HMM
458 profile. Subsequently, this HMM profile was screened against several viral proteome datasets:
459 The Pacific Ocean Virome[50], the Reference Viral Database[51], the assembled Osaka Bay and

TARA oceans viromes[52], and the Virus-Host database[53]. This analysis failed to identify viral AltSQE homologues.

Phylogenetic analysis. All AltSQE sequences were aligned using MAFFT L-INS-i 7.187[40] and trimmed using trimAl v1.4.1[54], thereby removing all positions in the alignment with gaps in 90% or more of the sequences (-gt 0.1). Phylogenetic tree construction was performed using the maximum likelihood method IQ-TREE v1.5.5[55] under the best amino acid substitution model selected by the built-in model-selection function ModelFinder[56] using the following set as potential models: JTT, LG, WAG, Blosum62, VT, and Dayhoff. To account for rate-heterogeneity across sites, the empirical amino acid frequencies (+F) and the FreeRate model (+R), options were set. Branch supports were estimated by 1000 ultra-fast bootstrap replicates (-bb 1000).

Data availability

Gene sequences used in this study were deposited in GenBank under accession numbers MH422131 to MH422144. All other data that support the findings of this study are available from the corresponding author upon request.

References

1. Benveniste, P. Biosynthesis and accumulation of sterols. *Annu. Rev. Plant Biol.* **55**, 429-457 (2004).
2. Payne, A. H. & Hales, D. B. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. *Endocr. Rev.* **25**, 947-970 (2004).
3. Weete, J. D., Abril, M. & Blackwell, M. Phylogenetic distribution of fungal sterols. *PLoS ONE* **5**, e10899 (2010).

4. Gold, D. A., Caron, A., Fournier, G. P. & Summons, R. E. Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* **543**, 420-423 (2017).

5. Summons, R. E., Bradley, A. S., Jahnke, L. L. & Waldbauer, J. R. Steroids, triterpenoids and molecular oxygen. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 951-968 (2006).

6. Crowe, S. A. et al., Atmospheric oxygenation three billion years ago. *Nature* **501**, 535-538 (2013).

7. Desmond, E. & Gribaldo, S. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.* **1**, 364-381 (2009).

8. Fabris, M. et al. Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*. *New Phytol.* **204**, 521-535 (2014).

9. Rampen, S. W., Abbas, B. A., Schouten, S. & Damsté, J. S. S. A comprehensive study of sterols in marine diatoms (Bacillariophyta): Implications for their use as tracers for diatom productivity. *Limnol. Oceanogr.* **55**, 91-105 (2010).

10. Gallo, C., d'Ippolito, G., Nuzzo, G., Sardo, A. & Fontana, A. Autoinhibitory sterol sulfates mediate programmed cell death in a bloom-forming marine diatom. *Nat. Commun.* **8**, 1292 (2017).

11. Goldstein, J. L., DeBose-Boyd, R. A. & Brown, M. S. Protein sensors for membrane sterols. *Cell* **124**, 35-46 (2006).

12. Gill, S., Stevenson, J., Kristiana, I. & Brown, A. J. Cholesterol-dependent degradation of squalene monooxygenase, a control point in cholesterol synthesis beyond HMG-CoA reductase. *Cell Metab.* **13**, 260-273 (2011).

13. Pollier, J. et al. The protein quality control system manages plant defence compound synthesis. *Nature* **504**, 148-152 (2013).

14. Dong, L. et al. Co-expression of squalene epoxidases with triterpene cyclases boosts production of triterpenoids in plants and yeast. *Metab. Eng.* **49**, 1-12 (2018).
15. Huysman, M. J. J. et al. AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *Plant Cell* **25**, 215-228 (2013).
16. Miettinen, K. et al. The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nat. Commun.* **8**, 14153 (2017).
17. Shanklin, J. & Cahoon, E. B. Desaturation and related modifications of fatty acids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 611-641 (1998).
18. Zhu, G., Koszelak-Rosenblum, M., Connelly, S. M., Dumont, M. E. & Malkowski, M. G. The crystal structure of an integral membrane fatty acid α-hydroxylase. *J. Biol. Chem.* **290**, 29820-29833 (2015).
19. Liu, X. et al. Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal Res.* **20**, 249-257 (2016).
20. Leber, R. et al. Dual localization of squalene epoxidase, Erg1p, in yeast reflects a relationship between the endoplasmic reticulum and lipid particles. *Mol. Biol. Cell* **9**, 375-386 (1998).
21. Laranjeira, S. et al. *Arabidopsis* squalene epoxidase 3 (SQE3) complements SQE1 and is important for embryo development and bulk squalene epoxidase activity. *Mol. Plant* **8**, 1090-1102 (2015).
22. Bai, Y. et al. X-ray structure of a mammalian stearoyl-CoA desaturase. *Nature* **524**, 252-256 (2015).
23. Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. USA* **91**, 10340-10344 (1994).

24. Monier, A. et al. Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Res.* **19**, 1441-1449 (2009).

25. Keeling, P. J. & Inagaki, Y. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1alpha. *Proc. Natl. Acad. Sci. USA* **101**, 15380-15385 (2004).

26. Szabová, J., Růžička, P., Verner, Z., Hampl, V. & Lukeš, J. Experimental examination of EFL and MATX eukaryotic horizontal gene transfers: coexistence of mutually exclusive transcripts predates functional rescue. *Mol. Biol. Evol.* **28**, 2371-2378 (2011).

27 Wang, D. Z. Neurotoxins from marine dinoflagellates: a brief review. *Mar. Drugs* **6**, 349-371 (2008).

28 Vasconcelos, V., Azevedo, J., Silva, M. & Ramos, V. Effects of marine toxins on the reproduction and early stages development of aquatic organisms. *Mar. Drugs* **8**, 59-79 (2010).

29 Trainic, M. et al. Infection dynamics of a bloom-forming alga and its virus determine airborne coccolith emission from seawater. *iScience* **6**, 327-335 (2018).

30. Moses, T. et al. Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16α hydroxylase from *Bupleurum falcatum*. *Proc. Natl. Acad. Sci. USA* **111**, 1634-1639 (2014).

31. Nelson, B. K., Cai, X. & Nebenführ, A. A multicolored set of *in vivo* organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* **51**, 1126-1136 (2007).

32. Deslandes, L. et al. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl. Acad. Sci. USA* **100**, 8024-8029 (2003).

33. De Riso, V. et al. Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **37**, e96 (2009).

34. Berges, J. A., Franklin, D. J. & Harrison, P. J. Evolution of an artificial seawater medium: improvements in enriched seawater, artificial water over the last two decades. *J. Phycol.* **37**, 1138-1145 (2001).

35. Lin, H.-Y. et al. Alkaline phosphatase promoter as an efficient driving element for exogenic recombinant in the marine diatom *Phaeodactylum tricornutum*. *Algal Res.* **23**, 58-65 (2017).

36 Nagai, T. et al. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87-90 (2002).

37 Strand, T. A., Lale, R., Degnes, K. F., Lando, M. & Valla, S. A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS ONE* **9**, e90372 (2014).

38. Diner, R. E., Bielinski, V. A., Dupont, C. L., Allen, A. E. & Weyman, P. D. Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Front. Bioeng. Biotechnol.* **4**, 65 (2016).

39. Siaut, M. et al. Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**, 23-25 (2007).

40. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).

41. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

42. Johnson, L. K., Alexander, H. B. & Brown, C. T. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. Preprint at https://www.biorxiv.org/content/early/2018/09/18/323576 (2018).

583 43. Burki, F. et al. Untangling the early diversification of eukaryotes: a phylogenomic
584 study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R.*
585 *Soc. Lond. Ser. B-Biol. Sci.* **283**, 20152802 (2016).
586 44. Yoshida, Y. et al. *De novo* assembly and comparative transcriptome analysis of
587 *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics* **17**, 182 (2016).
588 45. Van Bel, M. et al. TRAPID: an efficient online tool for the functional and comparative
589 analysis of *de novo* RNA-Seq transcriptomes. *Genome Biol.* **14**, R134 (2013).
590 46. Keller, O., Kollmar, M., Stanke, M. & Waack, S. A novel hybrid gene prediction
591 method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757-763
592 (2011).
593 47 Villar, E. et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes
594 online. *Nucleic Acids Res.* **46**, W289-W295 (2018).
595 48 Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science*
596 **348**, 1261359 (2015).
597 49. Wilson, W. H. et al. Complete genome sequence and lytic phase transcription profile
598 of a *Coccolithovirus*. *Science* **309**, 1090-1092 (2005).
599 50. Hurwitz, B. L. & Sullivan, M. B. The Pacific Ocean Virome (POV): a marine viral
600 metagenomic dataset and associated protein clusters for quantitative viral ecology.
601 *PLoS ONE* **8**, e57355 (2013).
602 51. Goodacre, N., Aljanahi, A., Nandakumar, S., Mikailov, M. & Khan, A. S. A Reference
603 Viral Database (RVDB) to enhance bioinformatics analysis of high-throughput
604 sequencing for novel virus detection. *mSphere* **3**, e00069-00018 (2018).
605 52. Nishimura, Y. et al. Environmental viral genomes shed new light on virus-host
606 interactions in the ocean. *mSphere* **2**, e00359-00316 (2017).
607 53. Mihara, T. et al. Linking Virus Genomes with Host Taxonomy. *Viruses* **8**, 66 (2016).

54. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

55. Nguyen, L.-T., Schmidt, H. A., von Haeseler, & A. Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).

56. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587-589 (2017).

Acknowledgements

We thank Rebecca De Clercq and Taya Lapshina for technical assistance, Yuechen Bai, Michael Johnson and Raffaela Abbriano-Burke for support with confocal microscopy, and Marie Huysman and Lieven De Veylder for providing the *P. tricornutum* cDNA library. J.P. is a postdoctoral fellow of the Research Foundation-Flanders. E.V. is funded by BOF project GOA01G01715. M.F. is supported by a CSIRO Synthetic Biology Future Science Fellowship, co-funded by CSIRO and the University of Technology Sydney.

Author contributions

J.P., E.V., U.K., and M.F. carried out experiments. J.P., K.V., A.G., and M.F. designed the experiments. All authors contributed to writing of the manuscript.

Competing interests

The authors declare no competing interests.

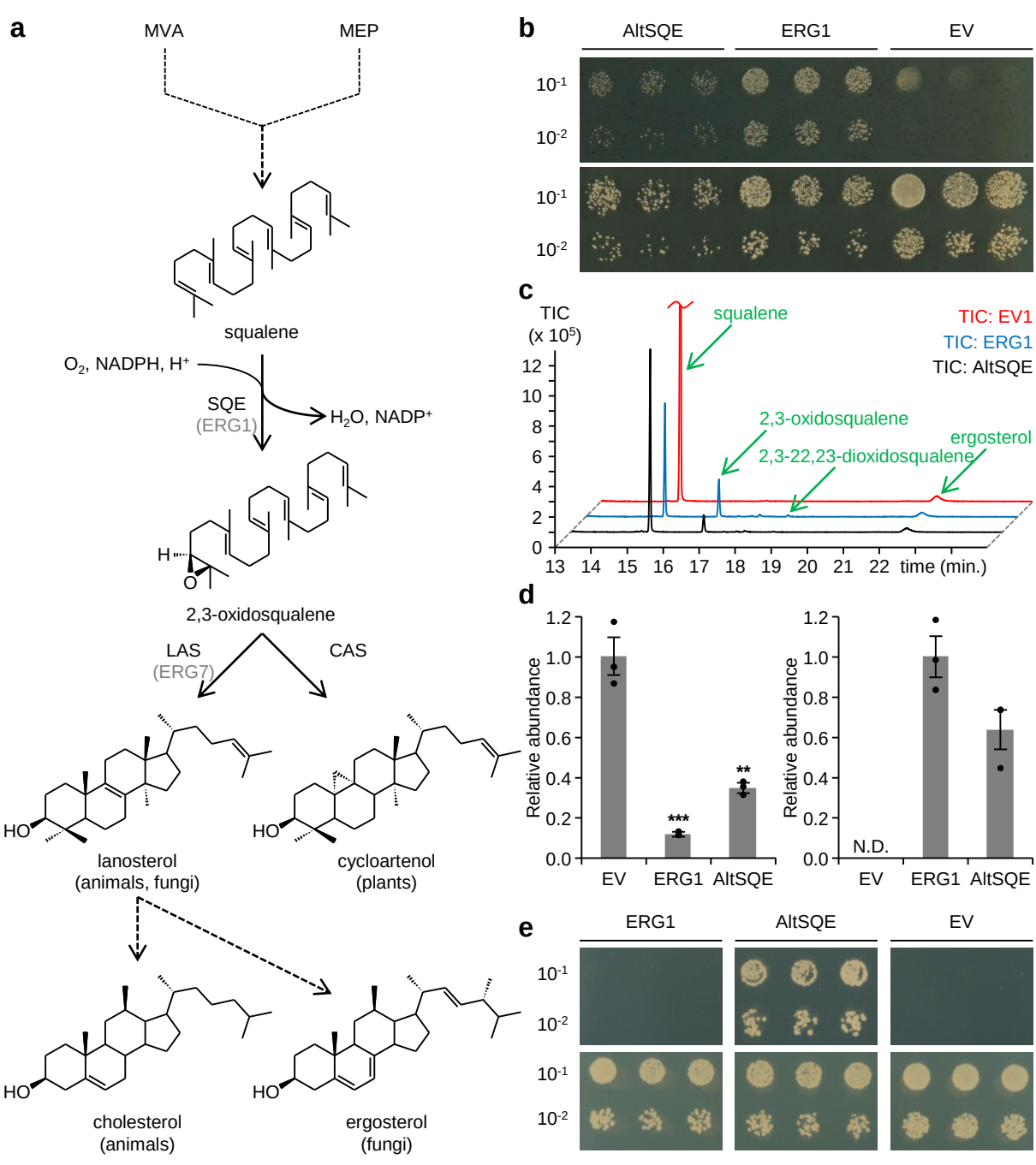
Figures Legends

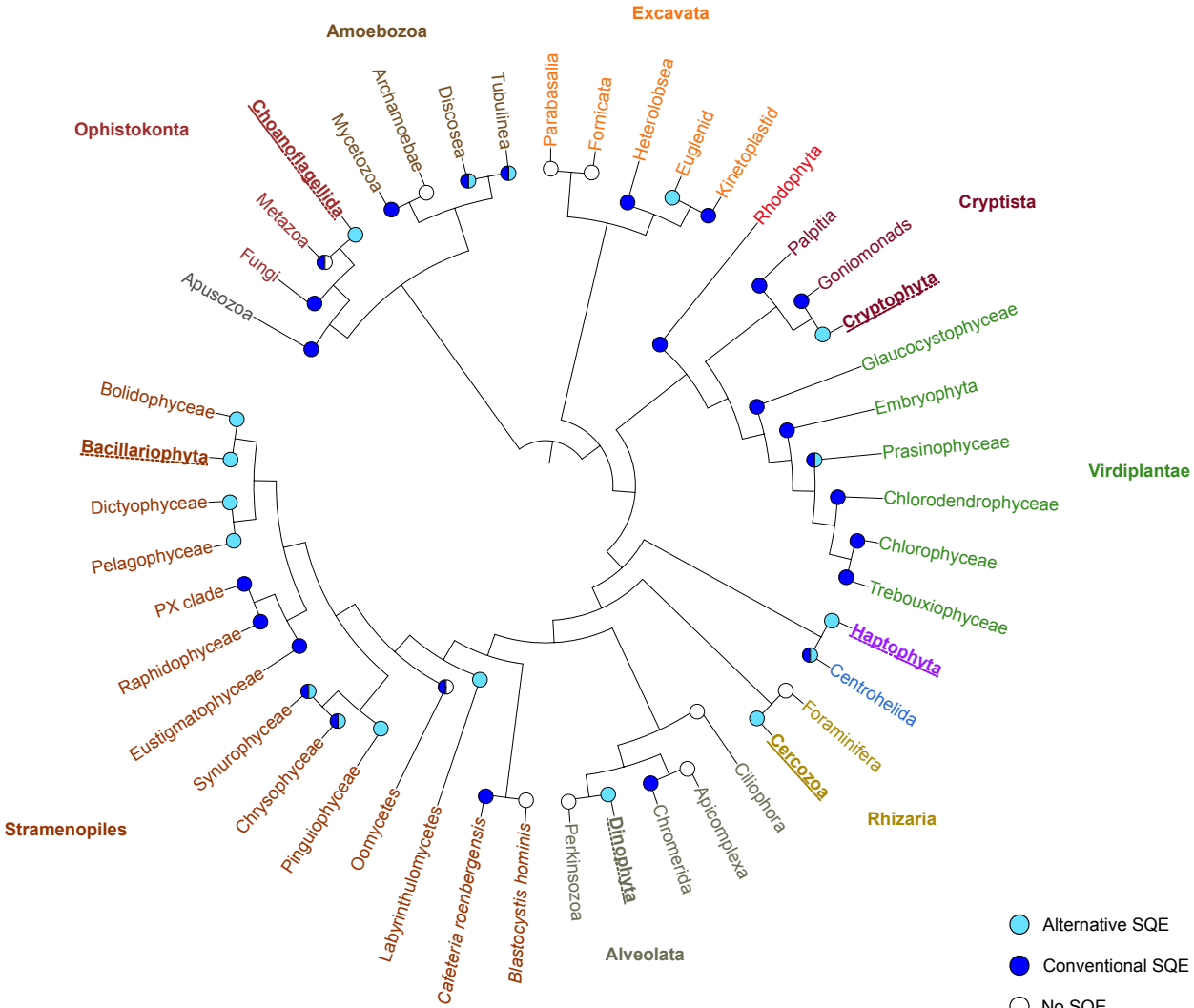
Fig. 1 | Identification of an alternative squalene epoxidase from *P. tricornutum*. a, Overview of the steroid biosynthesis pathway. Dashed arrows indicate multiple steps. MVA, mevalonate pathway; MEP, 2-*C*-methyl-D-erythritol 4-phosphate pathway; LAS, lanosterol synthase; CAS, cycloartenol synthase. **b,** A serial dilution of yeast JP064 (*erg1Δ*) cells expressing either *ERG1*, *AltSQE* or an empty vector (EV) control was dropped on plates lacking (top) or supplemented with (bottom) ergosterol. Yeast pre-cultures were diluted 10-fold (10^{-1}) or 100-fold (10^{-2}) in sterile water prior to dropping on the plates. Two independent repeats of this experiment led to similar results. **c,** Overlay of GC-MS total ion current (TIC) chromatograms showing accumulation of squalene and 2,3-oxidosqualene. Similar results were obtained for two independent experiments with three biological replicates each. **d,** Relative quantification of squalene (left) and 2,3-oxidosqualene (right) in yeast PA123 cells expressing either *ERG1*, *AltSQE* or an EV control. The mean and standard error (n=3) are shown and dot plots (black dots) are overlaid. Statistical significance was determined by a two-tailed Student's *t*-test (***P*<0.01, ****P*<0.001). **e,** A serial dilution of yeast PA059 (wild-type) cells expressing *ERG1*, *AltSQE* or an EV control was dropped on plates lacking (bottom) or supplemented with (top) 600 μM of terbinafine. Yeast pre-cultures were diluted 10-fold (10^{-1}) or 100-fold (10^{-2}) in sterile water prior to dropping on the selective plates. Two additional independent repeats of this experiment led to similar results.

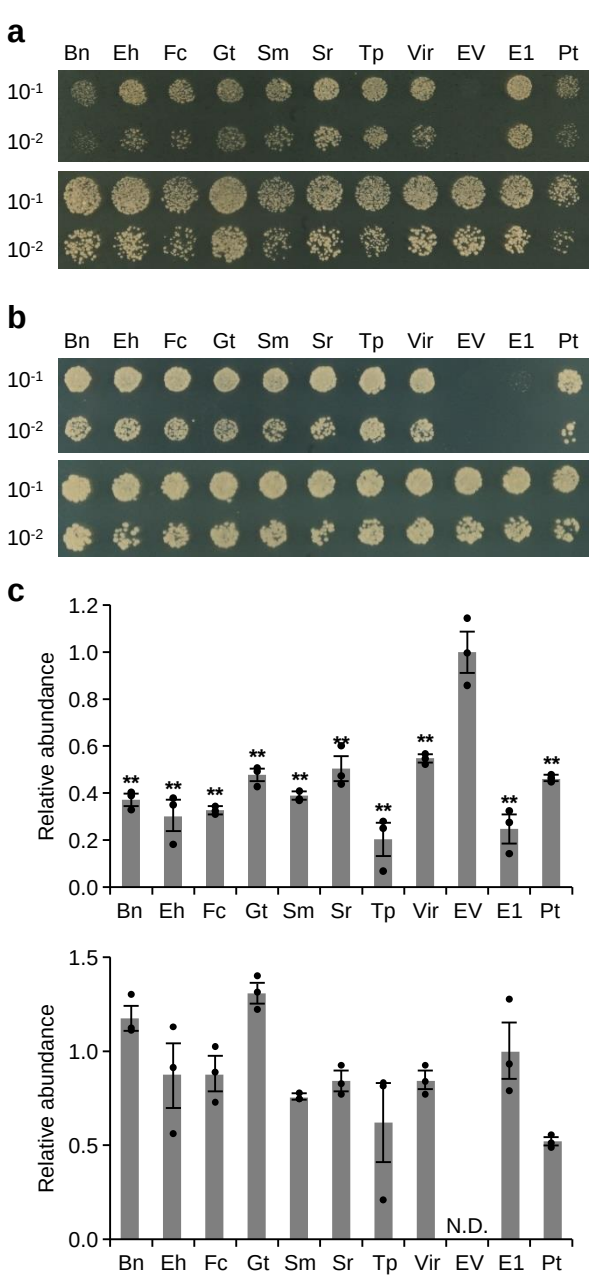
Fig. 2 | Distribution of SQE across the tree of life. Coloured circles indicate the type of SQE that can be found in every clade, white circles denote an absence of any SQE gene. Demi-filled circles show a patchy signal in the clade. Topology of the tree is drawn as

proposed[43]. For the clades in bold and underlined font, corresponding AltSQE enzymes have been functionally characterized in this study.

Fig. 3 | Characterization of AltSQEs from various species. a, Serially diluted yeast JP064 (*erg1Δ*) cells expressing *AltSQE* from various species, *ERG1* (E1) or an empty vector (EV) control were dropped on plates lacking (top) or supplemented with (bottom) ergosterol. **b,** Serially diluted yeast strain PA059 (wild-type) cells expressing *AltSQE* from various species, *ERG1* (E1) or an EV control were dropped on plates lacking (bottom) or supplemented with (top) 600 μM of terbinafine. The experiments in a and b were repeated with three additional biological replicates. Yeast pre-cultures were diluted 10-fold (10^{-1}) or 100-fold (10^{-2}) in sterile water prior to dropping on the selective plates. **c,** Relative quantification of squalene (top) and 2,3-oxidosqualene (bottom) in yeast PA123 (*erg1Δ*, *erg7Δ*) cells expressing *AltSQE* from various species, *ERG1* (E1) or an EV control. The mean and standard error (n=3) are plotted and dot plots (black dots) are overlaid. Statistical significance was determined by a two-tailed Student's *t*-test (**P<0.01) compared to the EV control. N.D.: not detected; Bn: *Bigelowiella natans*; Eh: *Emiliania huxleyi*; Fc: *Fragilariopsis cylindrus*; Gt: *Guillardia theta*; Sm: *Symbiodinium minutum*; Sr: *Salpingoeca rosetta*; Tp: *Thalassiosira pseudonana*; Vir: *E. huxleyi* virus 84; Pt: *Phaeodactylum tricornutum*.


a
MVA
MEP
squalene
O2, NADPH, H+
SQE
(ERG1)
H2O, NADP+
H
O
2,3-oxidosqualene
LAS
(ERG7)
CAS
HO
lanosterol
(animals, fungi)
HO
cycloartenol
(plants)
HO
cholesterol
(animals)
HO
ergosterol
(fungi)
b
AltSQE
ERG1
EV
10-1
10-2
10-1
10-2
c
TIC
(x 10^5)
squalene
TIC: EV1
TIC: ERG1
TIC: AltSQE
2,3-oxidosqualene
ergosterol
2,3-22,23-dioxidosqualene
0
2
4
6
8
10
12
13 14 15 16 17 18 19 20 21 22 time (min.)
d
Relative abundance
0.0
0.2
0.4
0.6
0.8
1.0
1.2
***
**
EV
ERG1
AltSQE
Relative abundance
N.D.
EV
ERG1
AltSQE
e
ERG1
AltSQE
EV
10-1
10-2
10-1
10-2


Amoebozoa
Excavata
Ophistokonta
Choanoflagellida
Myxetozoa
Archamoebae
Discosea
Tubulinea
Parabasalia
Fornicata
Heterolobosea
Euglenid
Kinetoplastid
Rhodophyta
Cryptista
Palpitia
Goniomonads
Cryptophyta
Metazoa
Fungi
Apusozoa
Glaucocystophyceae
Embryophyta
Prasinophyceae
Virdiplantae
Chlorodendrophyceae
Chlorophyceae
Trebouxiophyceae
Bolidophyceae
Bacillariophyta
Dictyophyceae
Pelagophyceae
PX clade
Raphidophyceae
Eustigmatophyceae
Synurophyceae
Chrysophyceae
Pinguiophyceae
Oomycetes
Labyrinthulomycetes
Cafeteria roenbergensis
Blastocystis hominis
Haptophyta
Centrohelida
Foraminifera
Cercozoa
Rhizaria
Ciliophora
Apicomplexa
Chromerida
Dinophyta
Perkinsozoa
Alveolata
Stramenopiles
Alternative SQE
Conventional SQE
No SQE


a
Bn Eh Fc Gt Sm Sr Tp Vir EV E1 Pt
10-1
10-2
10-1
10-2
b
Bn Eh Fc Gt Sm Sr Tp Vir EV E1 Pt
10-1
10-2
10-1
10-2
c
Relative abundance
1.2
1.0
0.8
0.6
0.4
0.2
0.0
**
Bn Eh Fc Gt Sm Sr Tp Vir EV E1 Pt
Relative abundance
1.5
1.0
0.5
0.0
N.D.
Bn Eh Fc Gt Sm Sr Tp Vir EV E1 Pt

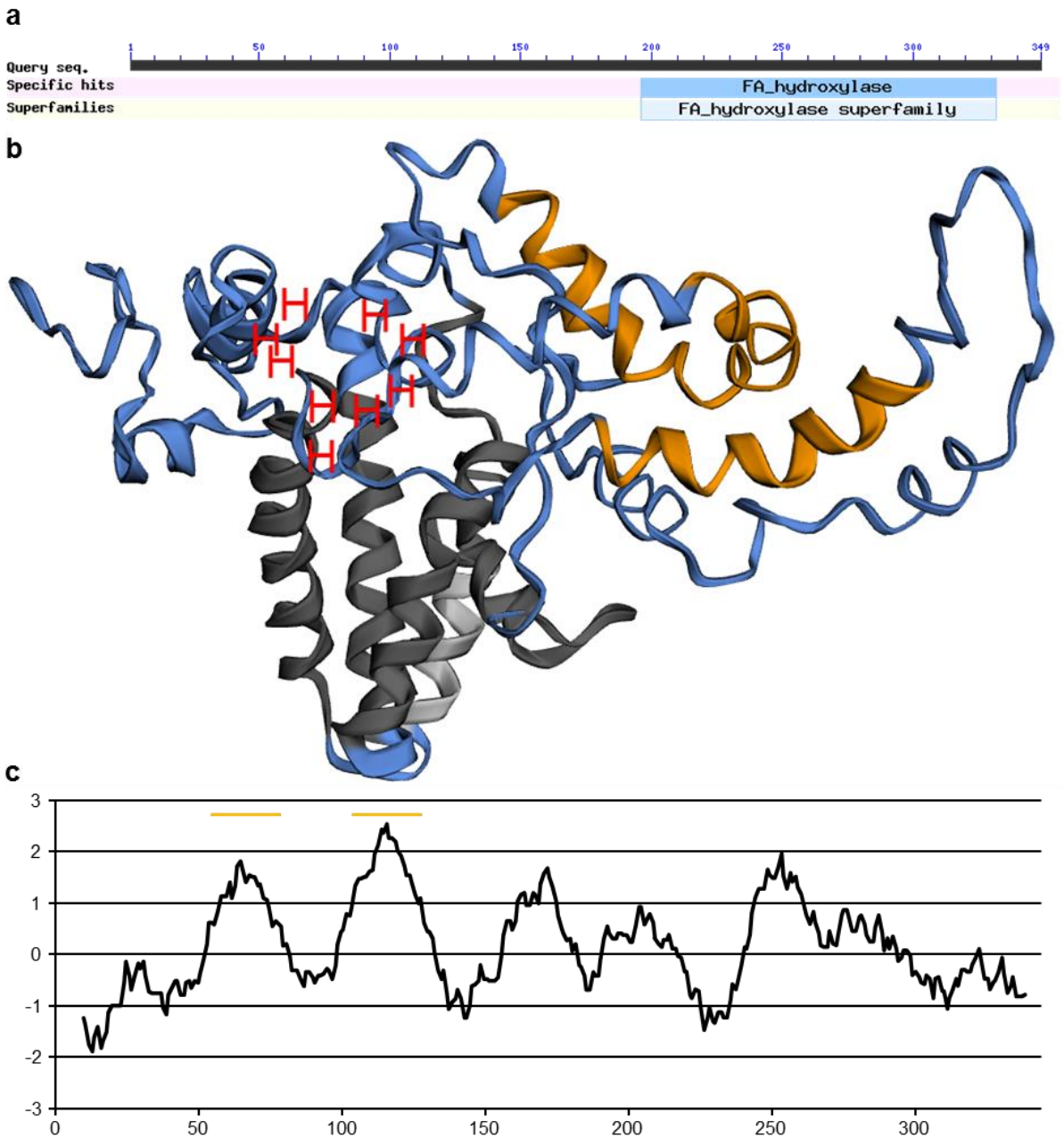
A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis

Jacob Pollier[1,2], Emmelien Vancaester[1,2], Unnikrishnan Kuzhiumparambil[3], Claudia E. Vickers[4,5], Klaas Vandepoele[1,2], Alain Goossens[1,2*] and Michele Fabris[3,4]

[1]Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium. [2]VIB Center for Plant Systems Biology, 9052 Ghent, Belgium. [3]Climate Change Cluster, University of Technology Sydney, 2007 Ultimo, NSW, Australia. [4]CSIRO Synthetic Biology Future Science Platform, GPO Box 2583, Brisbane 4001, Australia. [5]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St Lucia, QLD, Australia.

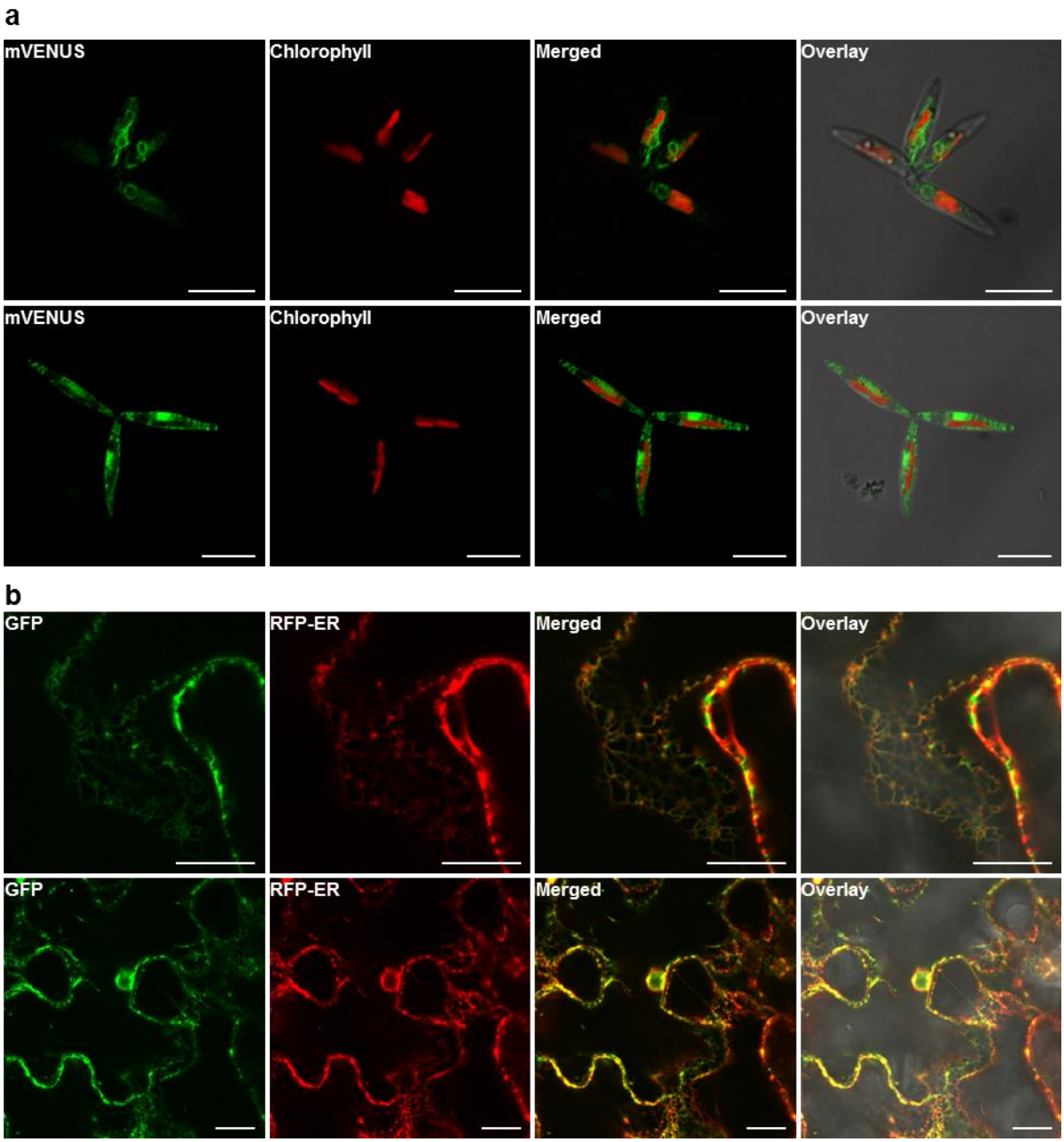
*Correspondence to: alain.goossens@psb-vib.ugent.be.

Supplementary Figures



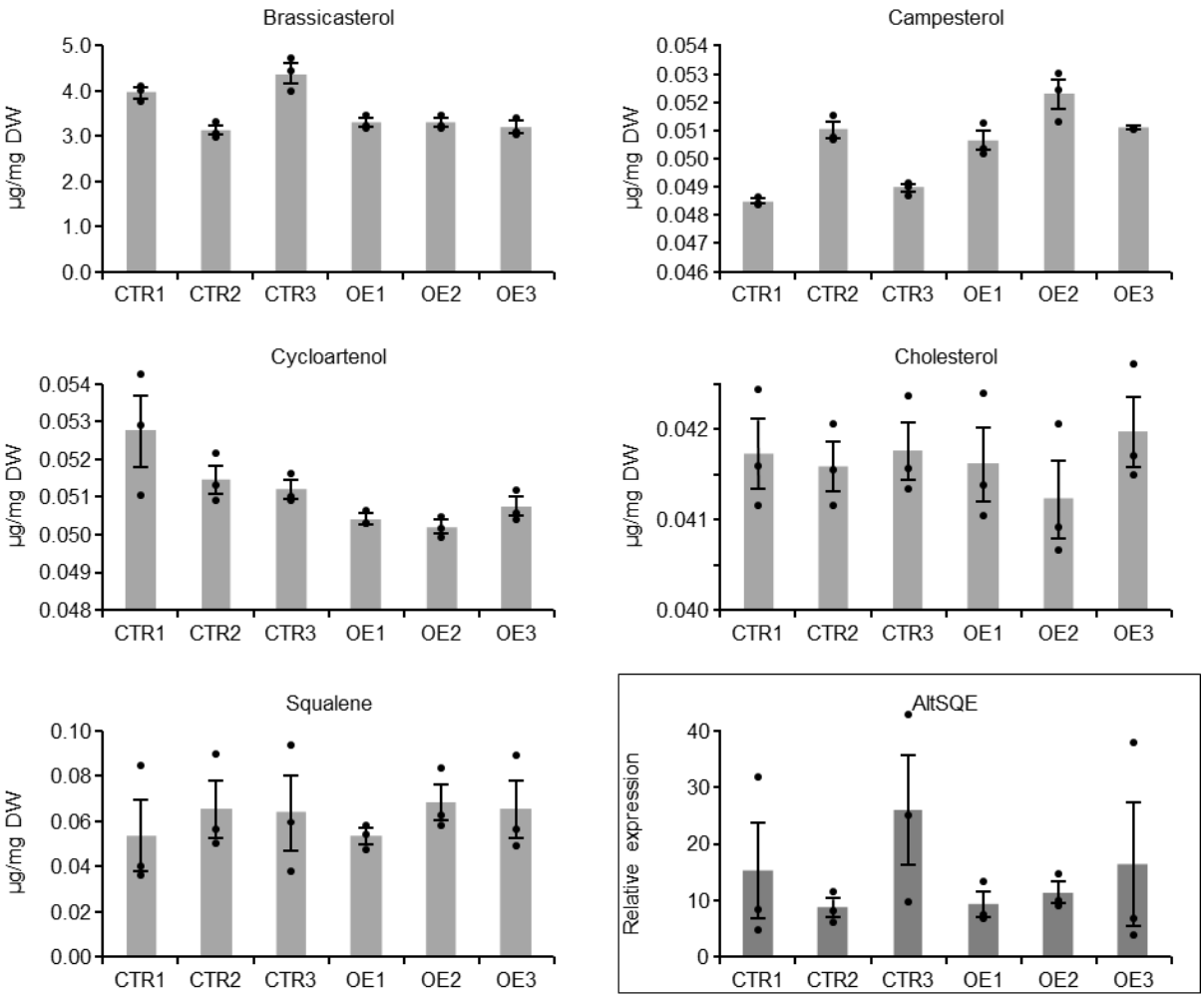
Supplementary Fig. 1 | AltSQE protein analysis. a, A protein domain search revealed that AltSQE belongs to the fatty acid hydroxylase superfamily. **b,** AltSQE protein model obtained using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The protein was modelled based on the ceramide very long chain fatty acid hydroxylase scs7p[18] template (99.7% confidence). The conserved histidine residues that coordinate the dimetal centre are indicated in red; the four transmembrane domains based on the scs7p protein template are

indicated in grey. Orange regions indicate additional potential transmembrane domains identified using Phyre2. **c,** Kyte-Doolittle hydropathy plot of AltSQE with window size 19. With this window size, potential transmembrane regions should score higher than 1.8. The orange regions correspond to the two additional potential transmembrane domains identified with Phyre2.

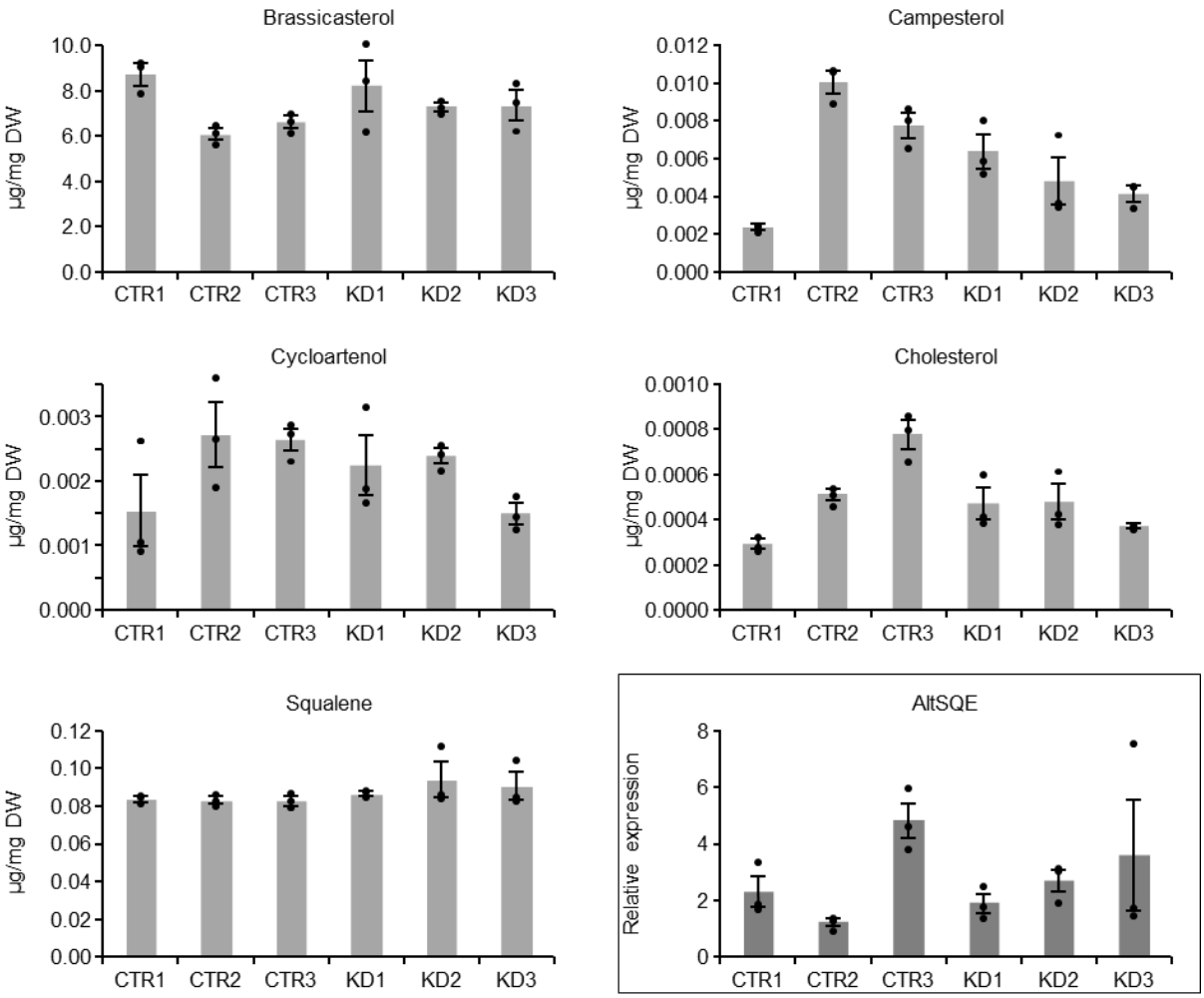


Supplementary Fig. 2 | Subcellular localization of *P. tricornutum* AltSQE. a, Confocal microscopy analysis revealing ER-localization of AltSQE in *P. tricornutum*. *P. tricornutum* cells transformed with either AltSQE-mVENUS (top) or free mVENUS (bottom) were imaged using confocal microscopy for the mVENUS (YFP; green) or chlorophyll autofluorescence (red) signals. The YFP signal observed around the chloroplast and nucleus indicates ER-localization of the AltSQE-mVENUS fusion protein. The free mVENUS protein localizes to the cytosol. Imaging was performed with three independent diatom cell lines for

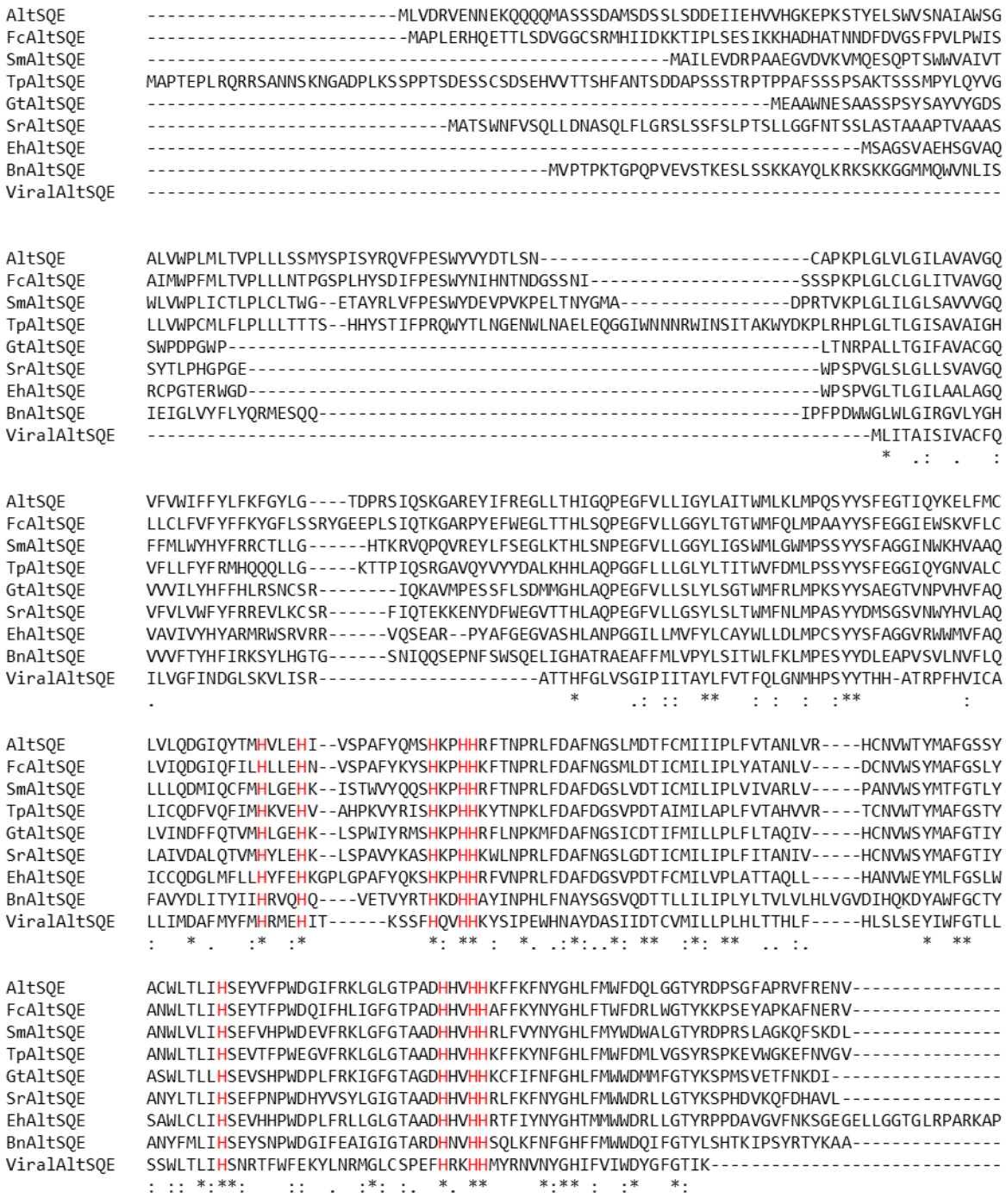
each construct, with similar results. Scale bars, 10 µm. **b,** Confocal microscopy analysis revealing ER-localization of AltSQE in *N. benthamiana* leaves *Agrobacterium*-infiltrated with both AltSQE-GFP (green) and the ER-rk red fluorescent ER-marker (red) construct[31] as localization control. Both red and green signals overlap, indicating ER-localization of the AltSQE-GFP fusion protein in *N. benthamiana* leaves. Imaging was performed with four leaves originating from two individual plants, with similar results. Scale bars, 20 µm.



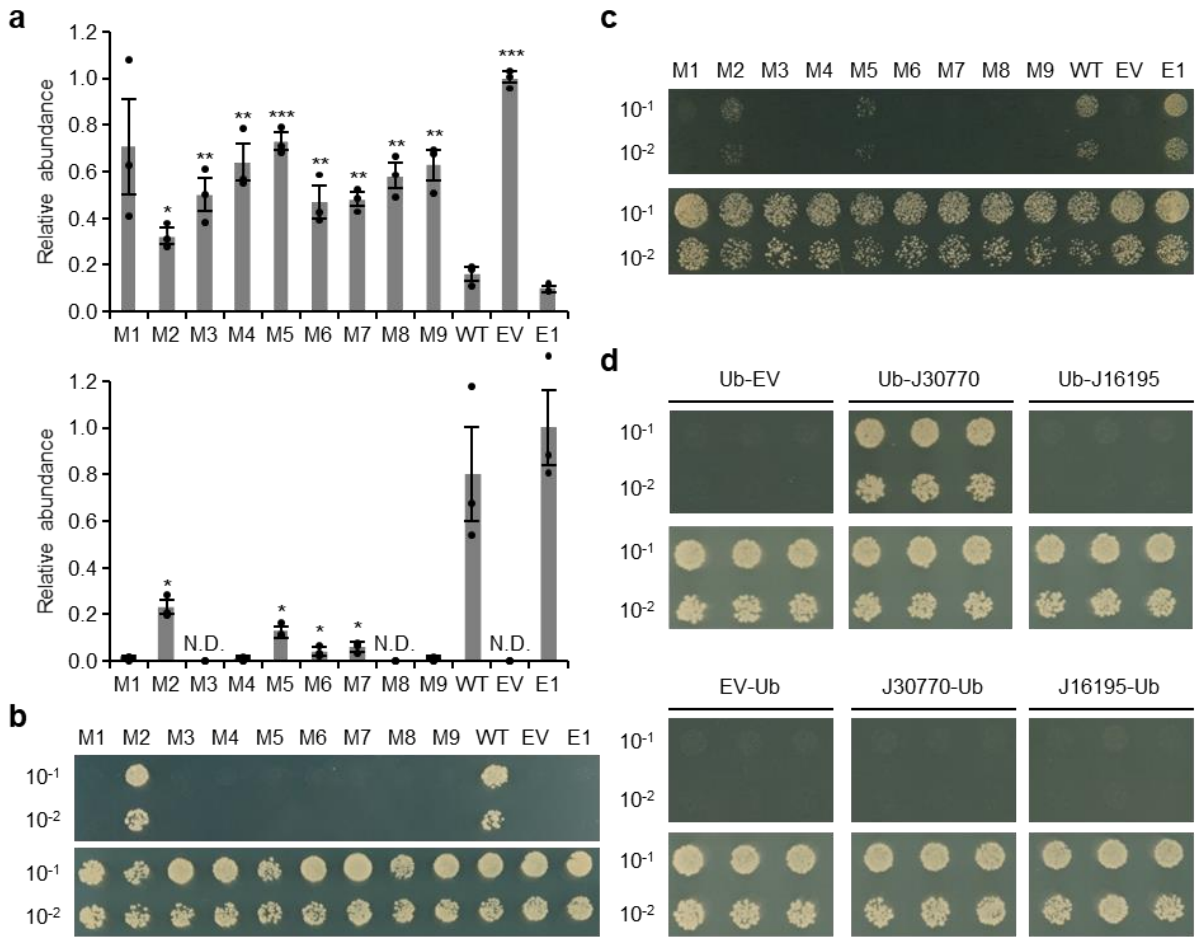
Supplementary Fig. 3 | Profiling of transgenic *P. tricornutum* strains transformed with *AltSQE* overexpression constructs. Diatoms carrying the extrachromosomal episome pPTBR11 either expressing the FcpBp-*AltSQE*-*mVenus* (OE) or the control FcpBp-*mVenus* (CTR) construct, profiled for triterpenoid accumulation (light grey bars) and for relative *AltSQE* gene expression normalized to the two reference genes used (boxed graph with dark grey bars). In accordance with previous reports in which 2,3-oxidosqualene could only be detected upon treatment of *P. tricornutum* with the CAS inhibitor Ro 48-8071[8], 2,3-oxidosqualene could not be detected in any of the samples. For each construct, three independent cell lines were generated and analysed. The mean and standard error (n=3) are plotted and dot plots (black dots) are overlaid. No statistically significant differences were identified using the Tukey method.



Supplementary Fig. 4 | Profiling of transgenic *P. tricornutum* strains transformed with *AltSQE* silencing constructs. Diatoms transformed with either the control pAF6 plasmid (CTR) or the inverted repeat RNAi construct FcpBp-*AltSQE*-FcpA3' (KD) were profiled for triterpenoid accumulation (light grey bars) and for relative *AltSQE* gene expression normalized to the two reference genes used (boxed graph with dark grey bars). In accordance with previous reports in which 2,3-oxidosqualene could only be detected upon treatment of *P. tricornutum* with the CAS inhibitor Ro 48-8071[8], 2,3-oxidosqualene could not be detected in any of the samples. For each construct, three independent cell lines were generated and analysed. The mean and standard error (n=3) are plotted and dot plots (black dots) are overlaid. No statistically significant differences were identified using the Tukey method.

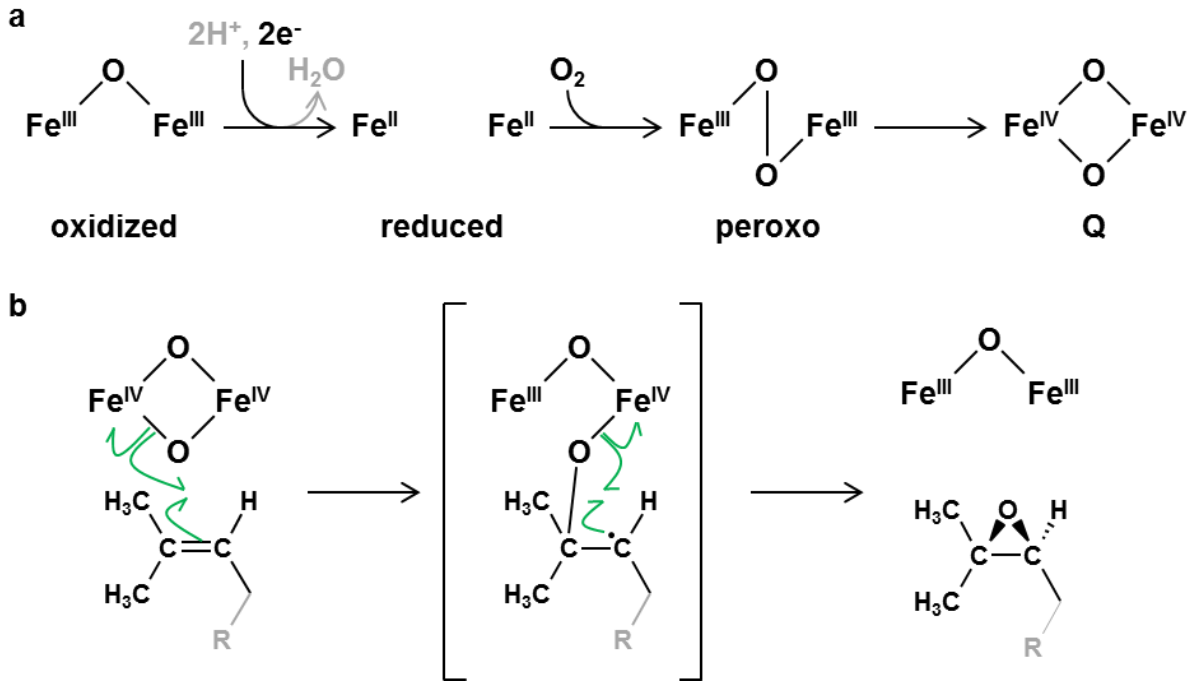


Supplementary Fig. 5 | Sequence alignment of *P. tricornutum* AltSQE with AltSQE sequences from other species. The conserved histidine residues are indicated in red. Bn: *Bigelowiella natans*; Eh: *Emiliania huxleyi*; Fc: *Fragilariopsis cylindrus*; Gt: *Guillardia theta*; Sm: *Symbiodinium minutum*; Sr: *Salpingoeca rosetta*; Tp: *Thalassiosira pseudonana*; Viral: *E. huxleyi* virus 84.

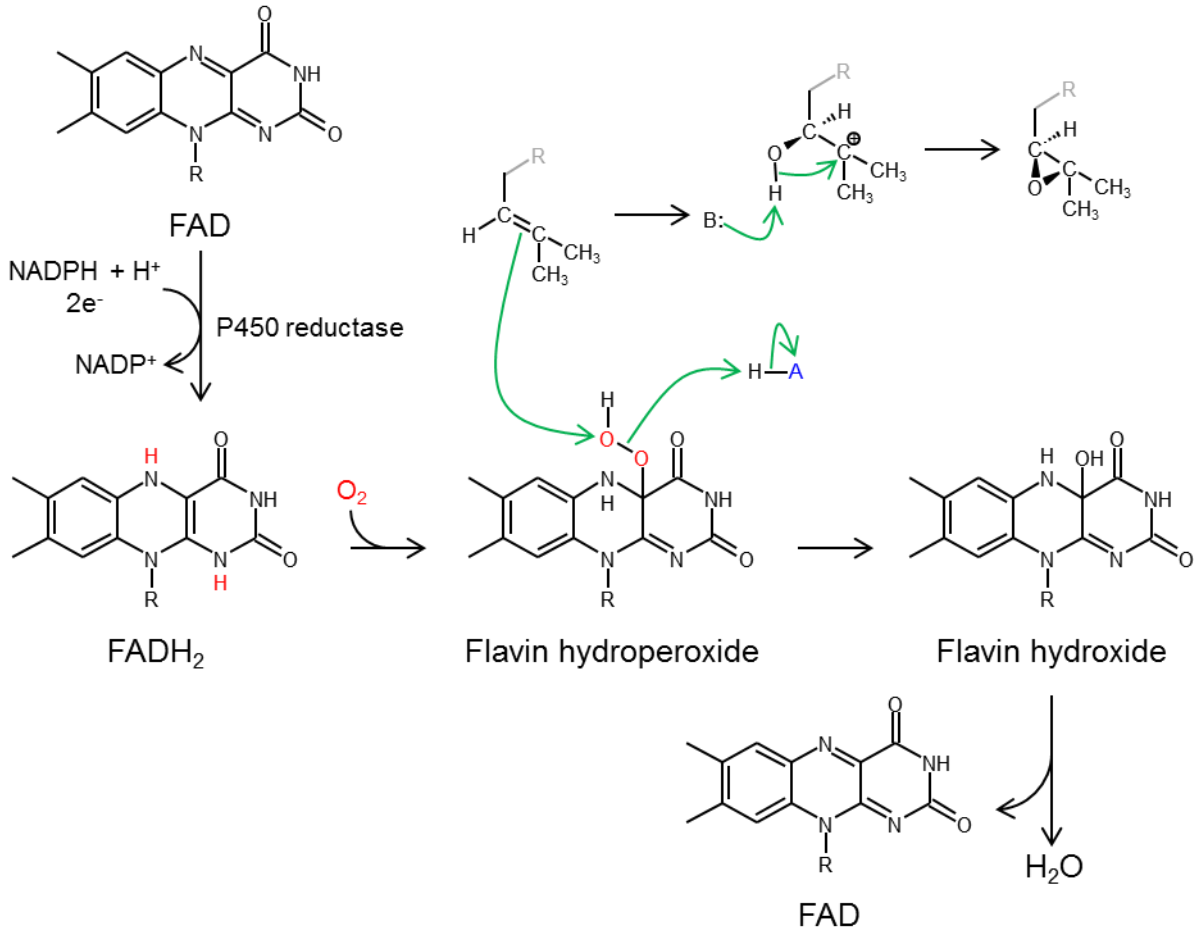


Supplementary Fig. 6 | Characterization of *P. tricornutum* AltSQE. a, Relative quantification of squalene (top) and 2,3-oxidosqualene (bottom) in yeast PA123 (*erg1Δ, erg7Δ*) cells expressing mutated versions of *AltSQE* (M1 to M9), wild-type *AltSQE* (WT), *ERG1* (E1) or an empty vector (EV) control. The mean and standard error (n=3) are shown and dot plots (black dots) are overlaid. Statistical significance was determined by a two-tailed Student's *t*-test (**P*<0.05, ***P*<0.01, ****P*<0.001) compared to the WT *AltSQE*. **b,** Serially diluted yeast PA059 (wild-type) cells expressing mutated versions of *AltSQE* (M1 to M9), WT *AltSQE*, *ERG1* (E1) or an EV control were dropped on plates lacking (bottom) or supplemented with (top) 600 μM of terbinafine. **c,** Serially diluted yeast strain JP064 (*erg1Δ*) cells expressing mutated versions of *AltSQE* (M1 to M9), WT *AltSQE*, *ERG1* (E1) or an EV control were dropped on plates lacking (top) or supplemented with (bottom) ergosterol. **d,** A split-ubiquitin assay reveals direct interaction between CytB5 and AltSQE. Serially diluted

yeast cells were dropped on plates with (top) or without (bottom) 5-fluoroorotic acid (5-FOA). Yeast cells capable of growing on plates containing 5-FOA are indicative of interaction between AltSQE and the tested cytochrome b_5. Interaction between AltSQE and Phatr3_J30770 was only observed for the N-terminally tagged cytochrome b_5. Yeast pre-cultures were diluted 10-fold (10^{-1}) or 100-fold (10^{-2}) in sterile water prior to dropping on the selective plates. For b and c, the experiments were repeated with two additional biological replicates and for d, four yeast colonies were assayed, all with similar results.



Supplementary Fig. 7 | Proposed mechanism of squalene epoxidation by AltSQE. a, Proposed O_2 activation pathway based on Shanklin & Cahoon (1998)[17]. In its resting state, the dimetal centre is in its oxidized diferric (Fe^{III}-Fe^{III}) state, coordinated by the conserved histidine residues. Using electrons provided by NADH *via* a cytochrome b_5, the dimetal centre is reduced to its diferrous (Fe^{II}-Fe^{II}) state. This is followed by binding of molecular oxygen to the reduced dimetal centre, leading to the peroxo state. Finally, scission of the peroxide bond leads to "compound Q". During this process, a water molecule is lost, however, because its precise timing is unknown, it is indicated in grey. **b,** Proposed epoxidation mechanism of squalene by AltSQE. Homolytic cleavage of the Fe^{IV}-oxygen bond of "compound Q" and the π-system from squalene results in the formation of a new C-O bond between squalene and "compound Q", generating a free radical that induces homolytic cleavage of the second Fe^{IV}-oxygen bond to form the epoxide.



Supplementary Fig. 8 | Proposed mechanism of squalene epoxidation by the conventional SQE. In a first step, the flavin adenine dinucleotide (FAD) redox cofactor, which is loosely bound to SQE, is reduced by NADPH-cytochrome P450 reductase to $FADH_2$ by addition of two protons (H^+) and two electrons (e^-). The resulting $FADH_2$ reacts with molecular oxygen (O_2), producing a flavin hydroperoxide intermediate that transfers an oxygen molecule to squalene via an electrophilic addition reaction initiated by a nucleophilic attack of the double bond on the terminal oxygen of the flavin hydroperoxide. The remaining flavin hydroxide is converted back to FAD *via* a dehydration reaction.

Supplementary Table 1 | List of oligonucleotides used in this study. Gateway™ recombination sites are underlined, additional bases to ensure the correct reading frame are italicized, lower case letters represent nucleotides used for gene mutation, lower case italicized letters represent nucleotides used for Gibson assembly and bold font indicates restriction sites used for cloning.

Name	Sequence	Description
combi6499	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGCTGGTAGATCGAGTC	AttB1-AltSQE-Fw (*P. tricornutum*)
combi6500	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CTACACGTTTTCTCGAAA	AttB2-AltSQE-Rv (*P. tricornutum*)
combi6507	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CACGTTTTCTCGAAACACGC	AttB2-AltSQE-NoStop-Rv (*P. tricornutum*)
combi6510	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGTTGATAACGGCGATC	AttB1-AltSQE-Fw (*E. huxleyi* virus 84)
combi6511	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*TTACTTGATCGTTCCGAA	AttB2-AltSQE-Rv (*E. huxleyi* virus 84)
combi6574	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGAGCGCCGGGAGCGTTGCG	AttB1-AltSQE-Fw (*E. huxleyi*)
combi6575	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CTAATCTTCATGGGCACGCGC	AttB2-AltSQE-Rv (*E. huxleyi*)
combi6576	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGGCTCCGCTTGAACGTCACC	AttB1-AltSQE-Fw (*F. cylindrus*)
combi6577	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*TCACACACGCTCGTTAAACGC	AttB2-AltSQE-Rv (*F. cylindrus*)
combi6578	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGGAAGCTGCGTGGAACGAG	AttB1-AltSQE-Fw (*G. theta*)
combi6579	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*TCATTTGTATGTGCCAAACATC	AttB2-AltSQE-Rv (*G. theta*) (**incorrect**)
combi6580	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGGCTACTTCCTGGAACTTTG	AttB1-AltSQE-Fw (*S. rosetta*)
combi6581	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CTATAACACCGCGTGATCGAATTG	AttB2-AltSQE-Rv (*S. rosetta*)
combi6582	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGGCCCCAACTGAACCCCTAAG	AttB1-AltSQE-Fw (*T. pseudonana*)
combi6583	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CTATACACCTACGTTGAACTC	AttB2-AltSQE-Rv (*T. pseudonana*)
combi6619	GTACACGATGgcTGTTCTCGAG	AltSQE Histidine mutation 1 Fw
combi6620	TGTTCTCGAGgcCATTGTATCA	AltSQE Histidine mutation 2 Fw
combi6621	TCAAATGTCGgcTAAACCGCAT	AltSQE Histidine mutation 3 Fw
combi6622	GCATAAACCGgcTCACCGCTTC	AltSQE Histidine mutation 4 Fw
combi6623	TAAACCGCATgcCCGCTTCACC	AltSQE Histidine mutation 5 Fw
combi6624	GACATTGATTgcTTCCGAATAC	AltSQE Histidine mutation 6 Fw
combi6625	TCCTGCTGACgcTCATGTTCAT	AltSQE Histidine mutation 7 Fw
combi6626	CCATCATGTTgcTCACAAGTTT	AltSQE Histidine mutation 8 Fw
combi6627	TCATGTTCATgcCAAGTTTTTC	AltSQE Histidine mutation 9 Fw
combi6628	CTCGAGAACAgcCATCGTGTAC	AltSQE Histidine mutation 1 Rv
combi6629	TGATACAATGgcCTCGAGAACA	AltSQE Histidine mutation 2 Rv

combi6630	ATGCGGTTTAgcCGACATTTGA	AltSQE Histidine mutation 3 Rv
combi6631	GAAGCGGTGAgcCGGTTTATGC	AltSQE Histidine mutation 4 Rv
combi6632	GGTGAAGCGGgcATGCGGTTTA	AltSQE Histidine mutation 5 Rv
combi6633	GTATTCGGAAgcAATCAATGTC	AltSQE Histidine mutation 6 Rv
combi6634	ATGAACATGAgcGTCAGCAGGA	AltSQE Histidine mutation 7 Rv
combi6635	AAACTTGTGAgcAACATGATGG	AltSQE Histidine mutation 8 Rv
combi6636	GAAAAACTTGgcATGAACATGA	AltSQE Histidine mutation 9 Rv
combi6637	TGCTCTCTTGGCTTTTCATC	pMKZ Split Ub vector primer Fw
combi6638	GACAGCGTTCTACCGTCTTCT	pMKZ Split Ub vector primer Rv
combi6639	AGAAGCAAAAAGAGCGATGC	pCup Split Ub vector primer Fw
combi6640	CTTTTCGGTTAGAGCGGATG	pCup Split Ub vector primer Rv
combi6655	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGTCCGCCGAAAAGGAATACATAC	AttB1-Phatr3_J30770-Fw
combi6656	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*TTACTTCATTTGCGTTTGGTAG	AttB2-Phatr3_J30770-Rv
combi6657	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGAACGATAGCACGATGCCTCCG	AttB1-Phatr3_J16195-Fw
combi6658	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CTAGAGCTTCCCCTTTCCTTCAAC	AttB2-Phatr3_J16195-Rv
combi6659	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CTTCATTTGCGTTTGGTAG	AttB2-Phatr3_J30770-NoStop-Rv
combi6660	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*GAGCTTCCCCTTTCCTTCAAC	AttB2-Phatr3_J16195-NoStop-Rv
combi6702	GATATCCTTGTTAAAAGTCTCCACTGACATTGGGCTTTTGTATGTGCCAAACATCATG	AltSQE-Rv (*G. theta*) for clone correction
combi6703	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*TTAGATATCCTTGTTAAAAGTCTCCAC	AttB2-AltSQE-Rv (*G. theta*) correct model
combi6757	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGGTGCCAACACCTAAAACAG	AttB1-AltSQE-Fw (*B. natans*)
combi6758	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*TTAGGCCGCTTTATATGTTCTG	AttB2-AltSQE-Rv (*B. natans*)
combi6759	GGGGACAAGTTTGTACAAAAAAGCAGGCT*TA*ATGGCTATCTTAGAGGTCGATC	AttB1-AltSQE-Fw (*S. minutum*)
combi6760	GGGGACCACTTTGTACAAGAAAGCTGGGT*A*CTAAAGGTCCTTGGAAAATTGC	AttB2-AltSQE-Rv (*S. minutum*)
MF574	*ctctagagtcgacctgcacatatg*CCAAAATTTCGTTCACGG	AP1 (*Phatr3_J49678*) promoter-Fw
MF575	*gctcaccatactagt*TTTGCAGGTCCGATAATG	AP1 (*Phatr3_J49678*) promoter-Rv
MF576	*acctgcaaaactagt*ATGGTGAGCAAGGGCGAG	mVenus-Fw
MF577	*tcgaggtag*TTACTTGTACAGCTCGTCCATG	mVenus-Rv
MF578	*gtacaagtaa*CTACCTCGACTTTGGCTG	FcbpF *(Phatr3_J51230)* 3'-Fw
MF579	*gcggaagagatgcctgca*TGAAGACGAGCTAGTGTTATTC	FcbpF *(Phatr3_J51230)* 3'-Rv
MF584	*taccagcatactagt*TTTGCAGGTCCGATAATG	AP1 (*Phatr3_J49678*) promoter-Rv
MF585	*acctgcaaaactagt*ATGCTGGTAGATCGAGTC	AltSQE-NoStop-Fw
MF586	*gctcaccat*CACGTTTTCTCGAAACAC	AltSQE-NoStop-Rv
MF587	*gaaaacgtg*ATGGTGAGCAAGGGCGAG	mVenus-Fw
MF596	ATATAT**CATATG**AATCTCGCCTATTCATGGTGT	FcpBp (*Phatr3_J18049*) promoter-Fw

MF597	ATATA**ACTAGT**CTGGCAACCGTGAAATATGCG	FcpBp (Phatr3_J18049) promoter-Rv
MF530	ATATAT**GAATTC**ATATTCCGCGAGGGTCTTCTTA	AltSQE_RNAi_F_1
MF531	TATATA**TCTAGA**AGGATTGGTGAAGCGGTGATG	AltSQE_RNAi_R_1
MF532	TATATA**TCTAGA**AGCCAGCATGCGTAGGACGAA	AltSQE_RNAi_R_2
MF606	CTTTCCTTGGGACGGCATTTTTCG	AltSQE qPCR (v1) Fw used for OE lines
MF607	AAACACGCGGGGAGCGAATC	AltSQE qPCR (v1) Rv used for OE lines
MF548	CGAAGTCAACCAGGAAACCAA	RPS qPCR Fw
MF549	GTGCAAGAGACCGGACATACC	RPS qPCR Rv
MF550	TAACGCGACTCTTTCCATCCA	TubB qPCR Fw
MF551	TTGGCAATCAATTGGTTCAGG	TubB qPCR Rv

121

Supplementary Data (separate pdf files)

Supplementary Data 1

Overview of all detected alternative and conventional SQE protein sequences in all queried organisms.

Supplementary Data 2

Maximum-likelihood phylogeny of eukaryotic and viral AltSQE proteins constructed from an alignment of homologs from 202 organisms and 403 informative aligned sites, rooted with ERG3. Proteins are coloured based on their phylogenetic affiliations and bootstrap values are mentioned at the right of every node.

Supplementary Data 3

Maximum-likelihood phylogeny of eukaryotic conventional SQE proteins constructed from an alignment of homologs from 235 organisms and 624 informative aligned sites, rooted with UbiH. Proteins are coloured based on their phylogenetic affiliations and bootstrap values are mentioned at the right of every node.