

ORIGINAL ARTICLE

Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes

Gurjeet S Kohli^{1,2}, Uwe John³, Frances M Van Dolah⁴ and Shauna A Murray^{1,2}

¹Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Sydney, New South Wales, Australia; ²Sydney Institute of Marine Sciences, Mosman, New South Wales, Australia; ³Alfred-Wegener-Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany and ⁴Marine Biotoxins Program, National Oceanic and Atmospheric Administration Center for Coastal and Environmental Health and Biomolecular Research, Charleston, SC, USA

Fatty acids, which are essential cell membrane constituents and fuel storage molecules, are thought to share a common evolutionary origin with polyketide toxins in eukaryotes. While fatty acids are primary metabolic products, polyketide toxins are secondary metabolites that are involved in ecologically relevant processes, such as chemical defence, and produce the adverse effects of harmful algal blooms. Selection pressures on such compounds may be different, resulting in differing evolutionary histories. Surprisingly, some studies of dinoflagellates have suggested that the same enzymes may catalyse these processes. Here we show the presence and evolutionary distinctiveness of genes encoding six key enzymes essential for fatty acid production in 13 eukaryotic lineages for which no previous sequence data were available (alveolates: dinoflagellates, *Vitrella*, *Chromera*; stramenopiles: bolidophytes, chrysophytes, pelagophytes, raphidophytes, dictyochophytes, pinguiphytes, xanthophytes; Rhizaria: chlorarachniophytes, haplosporida; euglenids) and 8 other lineages (apicomplexans, bacillariophytes, synurophytes, cryptophytes, haptophytes, chlorophyceans, prasinophytes, trebouxioophytes). The phylogeny of fatty acid synthase genes reflects the evolutionary history of the organism, indicating selection to maintain conserved functionality. In contrast, polyketide synthase gene families are highly expanded in dinoflagellates and haptophytes, suggesting relaxed constraints in their evolutionary history, while completely absent from some protist lineages. This demonstrates a vast potential for the production of bioactive polyketide compounds in some lineages of microbial eukaryotes, indicating that the evolution of these compounds may have played an important role in their ecological success.

The ISME Journal (2016) 10, 1877–1890; doi:10.1038/ismej.2015.263; published online 19 January 2016

Introduction

Marine microbial eukaryotes are major contributors to nutrient cycling and photosynthesis, responsible for a sizable proportion of the global primary production (Field *et al.*, 1998; Worden *et al.*, 2015). A subset of these organisms produce toxins involved in harmful algal blooms with major impacts on ecosystem functioning and economic impacts on aquaculture and fisheries industries (Hallegraeff, 1993, 2010 and references therein). Despite their importance, comparatively little is known regarding key biosynthetic pathways in protists (Kalaitzidis *et al.*, 2010). In all organisms studied to date, fatty

acid synthases (FASs) and polyketide synthases (PKSs) are closely related and have a common evolutionary history (Jenke-Kodama *et al.*, 2005). FASs and PKSs share a similar enzymatic domain structure in which acyl transferase (AT), ketosynthase (KS) and an acyl carrier protein (ACP) form the core structure for condensation of acyl units, and are essential for both PKSs and FASs. The other domains, ketoreductases (KR), enoyl reductase (ER) and dehydratase (DH) modify the acyl units after condensation, which is essential for FASs, but selectively present/absent in PKS (Cane *et al.*, 1998; Khosla *et al.*, 1999; Jenke-Kodama *et al.*, 2005).

Type I PKS (modular) have a single protein consisting of all catalytic domains, which are used in a progressive fashion for chain elongation until the thioesterase domain releases the finished polyketide, analogous to FASs in animals and fungi (Khosla *et al.*, 1999; Jenke-Kodama *et al.*, 2005). Type II PKSs carry each catalytic domain on separate polypeptides (mono-functional proteins) that form

Correspondence: GS Kohli or SA Murray, Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, CB07.06.039, Sydney, New South Wales 2007, Australia.

E-mail: gurjeet.kohli@uts.edu.au, gurukohli@gmail.com or shauna.murray@uts.edu.au

Received 30 June 2015; revised 18 November 2015; accepted 7 December 2015; published online 19 January 2016

multiprotein complexes, analogous to type II FASs in bacteria and plants (McFadden 1999; Jenke-Kodama *et al.*, 2005). Type III PKSs are self-contained homodimeric enzymes where each monomer performs a specific function and are found in plants, brown alga, bacteria and fungi (Khosla *et al.*, 1999; Jenke-Kodama *et al.*, 2005; Cock *et al.*, 2010). In plants, FAS genes are encoded in the nucleus and proteins are targeted towards the chloroplast, where fatty acid synthesis occurs (McFadden, 1999). Gene identification/cloning and functional characterisation of all the FAS enzymes has been carried out in higher plants and bacteria (White *et al.*, 2005; Brown *et al.*, 2010 and references therein).

In the unicellular chlorophyte *Chlamydomonas*, fatty acid synthesis is thought to be carried out in the chloroplast stroma via a type II FAS, which was characterised by identifying the genes encoding type II FAS enzymes (Riekhof and Benning, 2009 and references therein). Fatty acid desaturation takes place via multiple desaturases, as majority of fatty acids in *Chlamydomonas* are unsaturated (Riekhof and Benning, 2009). In Apicomplexa, an exceptionally large type I FAS has been characterised from *Cryptosporidium parvum* (Zhu *et al.*, 2000). In contrast, genes encoding some type II FAS enzymes have been identified in *Toxoplasma gondii* and *Plasmodium falciparum* (Waller *et al.*, 1998), with no type I FAS genes found in these organisms (Gardner *et al.*, 2002). The mechanism and genetic basis for fatty acid synthesis remain largely unknown in many eukaryote lineages (Ryall *et al.*, 2003; Armbrust *et al.*, 2004).

Type I PKS genes are known from only a handful of Apicomplexa, haptophytes, chlorophytes and dinoflagellates (Bachvaroff and Place, 2008; John *et al.*, 2008; Monroe and Van Dolah, 2008; Place, 2008; Eichholz *et al.*, 2012; Murray *et al.*, 2012; Salcedo *et al.*, 2012; Pawlowicz *et al.*, 2014; Meyer *et al.*, 2015). Recent transcriptome surveys have demonstrated the possibility that protists may encode a massive diversity of PKS genes (Pawlowicz *et al.*, 2014; Kohli *et al.*, 2015; Meyer *et al.*, 2015). Studies of the natural products produced by some protists continue to identify new polyketide compounds at a rapid rate. This indicates that our current knowledge about the genetic basis of PKS in protists is highly incomplete (Pawlowicz *et al.*, 2014; Meyer *et al.*, 2015).

The first comprehensive genetic information on 210 marine microbial genera (305 unique species, 396 unique strains, 678 transcriptomes) encompassing most of the major lineages of eukaryotes has recently been undertaken by the Marine Microbial Eukaryote Sequencing Project (MMETSP) project (Keeling *et al.*, 2014). Using this vast data resource, the major aims of this study were to (i) identify the genetic basis of FAS and PKS synthesis in the major lineages of eukaryotes and (ii) infer the constraints and processes in the evolutionary history of polyketide and fatty acid synthesis in eukaryotes.

Materials and methods

RNA extraction and construction of transcriptomic libraries

Alexandrium margalefi CS322 and *Gambierdiscus australes* CAWD149 were cultured at 18 and 25 °C, respectively, in f/2 medium, under cool white fluorescent light at a light intensity of 60 $\mu\text{mol m}^{-2}\text{s}^{-1}$ and a 12:12 light:dark cycle. RNA was first extracted via TriReagent (Life Technologies, Carlsbad, CA, USA), then purified using the RNeasy Plant mini kit (Qiagen, Limberg, Netherlands) and residual DNA removed via the TURBO DNA-free Kit (Life Technologies) according to the manufacturer's protocols. Total RNA was submitted to MMETSP for sequencing. Procedures used by MMETSP to generate transcriptomic libraries have been described in detail in Keeling *et al.* (2014). Underlying culturing conditions, environmental and experimental metadata for all the other MMETSP libraries used in this study are described in Supplementary Table S5.

Identification of FAS and PKS genes

Transcriptomic libraries representing 213 strains and 152 genera were obtained from MMETSP and other studies (Supplementary Table S1). For strains where multiple transcriptomic libraries were constructed (e.g. *Alexandrium monilatum* was grown under different physiological stresses and four transcriptomic libraries each grown under a different physiological stress were constructed), a combined assembly of data generated from all the transcriptomic libraries was provided by MMETSP to maximise transcriptomic coverage. In this study we used combined libraries where available. All the MMETSP assembled transcriptomes can be accessed from <http://data.imicrobe.us/project/view/104>. Transcriptomic libraries for *Alexandrium fundayense* (Wisecaver *et al.*, 2013), *A. pacificum* (Hackett *et al.*, 2013), *A. tamarense* (Hackett *et al.*, 2013), *Gambierdiscus belizeanus* (Kohli *et al.*, 2015) and *Symbiodinium* sp. CassKB8 (Bayer *et al.*, 2012) were obtained from Genbank and culturing conditions and metadata can be obtained from their respective references. Sequences for *T. gondii*, *Neospora caninum*, *Eimeria falciformis* and *Eimeria tenella* were obtained from ToxoDB (Gajria *et al.*, 2008) using the BLAST tool and *Azadinium spinosum* sequences as a query. *Emiliania huxleyi* CCMP1516 reference genome and transcriptome were obtained from JGI genome portal (Read *et al.*, 2013). Genes encoding type I PKSs and type II FASs were identified using HMMER (Finn *et al.*, 2011; Supplementary Tables S1 and S3) where in-house HMM databases were developed for each enzyme investigated in this study. Separate HMM profiles for each domain involved in type I PKSs and type II FASs were developed so that the profiles could recognise multiple domains on a single transcript. The presence of a type I FAS enzyme in all the

transcriptomic libraries was recorded only when the transcript encoded the full domain. For type I PKS enzymes, the presence of both partial and full domains were recorded (Supplementary Table S3). However, the completeness of the transcripts encoding full/partial domains could not be determined. Functional prediction and identification of conserved active site amino acid residues in the transcripts were identified and screened using CDD (Marchler-Bauer *et al.*, 2015) and Pfam (Punta *et al.*, 2012). Identification of transit peptide targeted towards chloroplast was detected using ChloroP (Emanuelsson *et al.*, 1999). Geneious software was used for sorting all these sequences (Kearse *et al.*, 2012).

Phylogenetic analysis

MAFFT (Katoh *et al.*, 2002) and ClustalW (Thompson *et al.*, 1994) were used to align the protein sequences from different data sets. The alignments were manually trimmed to ensure they spanned the same coding region of each enzyme and maximum likelihood phylogenetic analysis was carried out using RAxML with 1000 bootstraps using the GAMMA and LG model of rate heterogeneity (Stamatakis, 2006). Details of each alignment and phylogenetic tree (newick format) used in this study are listed in Supplementary Table S6 and sequences used to generate a concatenated alignment of type II FAS genes for Figure 3 are listed in Supplementary Table S7. Phylogenetic trees were visualised using iTOL (Letunic and Bork, 2011) and MEGA:Version6 (Tamura *et al.*, 2013).

Results and discussion

Fatty acid biosynthesis in protists

Transcriptomic libraries representing 213 strains and 152 genera were screened for seven key enzymes, that is, 3-ketoacyl ACP synthase I, II and III (KASI-FabB, KASII-FabF, KASIII-FabH), ACP *S*-malonyl-transacylase (AT-FabD), trans-3-ketoacyl ACP reductase (KR-FabG), 3-hydroxyacyl-ACP dehydratase (DH-FabZ) and enoyl-ACP reductase (ER-FabI), involved in type II fatty acid synthesis. The presence of six of the seven genes (except KASI-FabB) were confirmed in all phototrophic lineages of alveolates (dinoflagellates, apicomplexa, *Vitrella*, *Chromera*), stramenopiles (diatoms, bolidophytes, chrysophytes, pelagophytes, raphidophytes, synurophytes, dictyochophytes, pinguiphytes, xanthophytes), Rhizaria (chlorarachniophytes and Haplosporidia), Viridiplantae (chlorophytes, prasinophytes, trebouxiphytes), excavates (euglenids, only three enzymes in *Eutreptiella*), cryptophytes and haptophytes (Table 1 and Supplementary Table S1). KASI-FabB was completely absent in Rhizaria and Viridiplantae, and selectively present/absent in all other phototrophic lineages examined

(Supplementary Table S1). The presence of all seven type II FAS enzymes was also confirmed in the reference genome of *E. huxleyi* (Supplementary Table S2).

Unique among the phototrophic lineages, the glaucophytes did not possess any of the type II FAS enzymes we screened for. It is noteworthy that the two glaucophyte genera screened here belonged to freshwater habitats. The absence of FAS genes may indicate insufficient depth of sequencing or that the FAS genes were not being expressed at the time of analysis. Detailed screening of other genera would shed more light on fatty acid synthesis in glaucophytes.

Previously, only two genes involved in type II FAS synthesis were known from a limited number of phototrophic lineages: haptophyte *Prymnesium parvum*, synurophyte *Mallomonas rasilis*, bacillariophyte *Phaeodactylum tricorutum* and *Thalassiosira pseudonana*, oomycete heterokont *Thraustotheca clavata* and cryptophytes *Guillardia theta* and *Hemiselmis virescens* (Ryall *et al.*, 2003; Armbrust *et al.*, 2004).

Among the heterotrophic lineages, we detected the presence of genes coding type II FAS enzymes in the dinoflagellate *Oxyrrhis marina* (DH-FabZ and KR-FabG) and opisthokont choanoflagellate *Acanthoecalia* sp. (KASI-FabB, KASIII-FabH and KR-FabG), possibly related to secondary acquisition of these genes from a prey item (*Heterosigma akashiwo* in case of *O. marina*), as supported by their phylogenetic position (Supplementary Table S1 and Supplementary Figures S1D and S2). The heterotrophic lineages of alveolates (ciliates), Rhizaria (Foraminifera), stramenopiles (bicosoecids, labyrinthulids-thraustochytrids and chrysophytes), Amoebozoa (tubulinids and dactylopodids), excavates (kinetoplastids) and *Palpitomonas* did not possess any genes coding type II FAS enzymes (Table 1 and Supplementary Table S1). This suggests that these organisms either obtain fatty acids from their diet and/or have a different FAS pathway.

The amino acid residues comprising the active sites of all seven enzymes have been elucidated previously in plants and bacteria and mutations at these sites abolishes function of the respective enzymes (White *et al.*, 2005; Brown *et al.*, 2010 and references therein). We found these active site residues highly conserved in protists (Figure 1), indicating that these FAS genes are fully functional.

In type II FAS, KASIII initiates the condensation reaction, while fatty acid chain elongation is carried out by KAS I or II. Depending on the length of fatty acid being produced and varying substrate specificities, different types of KAS I and II are present in plant type II FAS systems (Kunst *et al.*, 1992; Millar and Kunst 1997; Millar *et al.*, 1999; Fiebig *et al.*, 2000; Dunn *et al.*, 2004), and are encoded by different gene families (Kunst *et al.*, 1992; Millar and Kunst 1997; Millar *et al.*, 1999; Fiebig *et al.*, 2000; Dunn *et al.*, 2004; White *et al.*, 2005; Brown

Table 1 List of organisms screened for fatty acid synthesis enzymes in marine microbial eukaryotes

Organism	Total number screened (libraries, strains, genera)
Positive libraries	
Alveolates	
<u>Dinoflagellates</u> : <i>Alexandrium andersonii</i> (CCMP2222-1); <i>Alexandrium fundayense</i> (38/3-1); <i>Alexandrium margalefi</i> (CS322-1); <i>Alexandrium monilatum</i> (CCMP3105-4 ^a); <i>Alexandrium ostenfeldii</i> (OF101-1); <i>Alexandrium pacificum</i> (CCMP1598-1); <i>Alexandrium tamarense</i> (CCMP1771-4 ^a); <i>Alexandrium tamarense</i> (ATSP1B-1); <i>Amphidinium carterae</i> (CCMP1314-3 ^a); <i>Azadinium spinosum</i> (3D9-3 ^a); <i>Brandtodinium nutriculum</i> (RCC3387-1); <i>Ceratium fusus</i> (PA161109-2 ^a); <i>Durinskia baltica</i> (CS/38-2 ^a); <i>Gambierdiscus australes</i> (CAWD149-1); <i>Gambierdiscus belizeanus</i> (CCMP401-1); <i>Glenodinium foliaceum</i> (CCAP1116-2 ^a); <i>Gonyaulax spinifera</i> (CCMP409-1); <i>Gymnodinium catenatum</i> (GC744-1); <i>Heterocapsa arctica</i> (CCMP445-1); <i>Heterocapsa rotundata</i> (SCCAPK/0483-1); <i>Heterocapsa triquetra</i> (CCMP448-1); <i>Karenia brevis</i> (CCMP2229-4 ^a ; Wilson-4 ^a ; SP3-2 ^a ; SP1-2 ^a); <i>Karlodinium micrum</i> (CCMP2283-3 ^a); <i>Kruptoperidinium foliaceum</i> (CCMP1326-2 ^a); <i>Lingulodinium polyerda</i> (CCMP1738-4 ^a); <i>Pelagodinium beii</i> (RCC1491-1); <i>Peridinium aciculiferum</i> (PEAR2-2 ^a); <i>Polarella glacialis</i> (CCMP1383-1); <i>Prorocentrum minimum</i> (CCMP2233-3 ^a); <i>Protoceratium reticulatum</i> (CCMP535-1); <i>Pyrodinium bahamense</i> (PBAHA01-1); <i>Scrippsiella hangoei</i> (SHTV5-3 ^a); <i>Scrippsiella hangoei</i> like (SHH14-3 ^a); <i>Scrippsiella trochoidea</i> (CCMP3099-3 ^a); <i>Symbiodinium spp</i> (CladeA-1; Ccmp421-4 ^a ; CassKb8-1; C1-2 ^a ; C15-2 ^a ; CCMP2430-3 ^a ; D1A-1; MP-4 ^a); <i>Togula jolla</i> (CCCC725-1)	Dinoflagellates (92, 46, 24); Apicomplexa (4, 4, 3); Others (2, 2, 2)
<u>Apicomplexa</u> : <i>Eimeria falciformis</i> (Bayer/Haberkorn/970-1); <i>Eimeria tenella</i> (Houghton-1); <i>Neospora caninum</i> (Liverpool-1); <i>Toxoplasma gondii</i> (GT1-1)	
<u>Others</u> : <i>Vitrella brassicaformis</i> -CCMP3346-1); <i>Chromera velia</i> (CCMP2878-1)	
Stramenopiles	
<u>Bacillariophytes</u> : <i>Amphiprora</i> sp. (CCMP467-4 ^a); <i>Amphora coffeaeformis</i> (CCMP127-3 ^a); <i>Asterionellopsis glacialis</i> (CCMP134-4 ^a); <i>Attheya septentrionalis</i> (CCMP2084-1); <i>Aulacoseira subarctica</i> (CCAP1002/5-1); <i>Chaetoceros affinis</i> (CCMP159-4 ^a); <i>Chaetoceros curvisetus</i> (NK^4-4 ^a); <i>Chaetoceros debilis</i> (MM31A/1-2 ^a); <i>Chaetoceros neogravile</i> (CCMP1317-4 ^a); <i>Corethron pennatum</i> (L29A3-2 ^a); <i>Coscinodiscus wailiesii</i> (CCMP2513-1); <i>Cyclophora tenuis</i> (ECT3854-1); <i>Cyclotella meneghiniana</i> (CCMP338-1); <i>Cylindrotheca closterium</i> (KMMCC/B181-1); <i>Dactyliosolen fragilissimus</i> (NK^1-1); <i>Detonula confervacea</i> (CCMP353-1); <i>Ditylum brightwellii</i> (GSO103-2 ^a ; GSO105-2 ^a ; GSO104-3 ^a); <i>Entomoneis</i> sp (CCMP2396-1); <i>Eucampia antarctica</i> (CCMP1452-1); <i>Extubocellulus spinifer</i> (CCMP396-4 ^a); <i>Fragilaria kerguelensis</i> (L26/C3-4 ^a ; L26/C5-4 ^a); <i>Grammatophora oceanica</i> (CCMP410-1); <i>Helicotheca tamensis</i> (CCMP826-1); <i>Leptocylindrus danicus</i> (B650-1); <i>Licmophora paradoxa</i> (CCMP2313-1); <i>Minutocellus polymorphus</i> (RCC2270-1); <i>Nitzschia punctata</i> (CCMP561-4 ^a); <i>Odontella aurita</i> (1302/5-1); <i>Proboscia alata</i> (P1/D3-1); <i>Pseudo-nitzschia australis</i> (10249/10/AB-4 ^a); <i>Rhizosolenia setigera</i> (CCMP1694-1); <i>Skeletonema dohrnii</i> (SkelB-2 ^a); <i>Skeletonema marinoi</i> (skelA-2 ^a); <i>Skeletonema menzeli</i> (CCMP793-2 ^a); <i>Stephanopyxis turris</i> (CCMP815-1); <i>Striatella unipunctata</i> (CCMP2910-1); <i>Synedropsis recta</i> (CCMP1620-1); <i>Thalassionema nitzschioides</i> (L26/B-2 ^a); <i>Thalassiosira antarctica</i> (CCMP982-4 ^a); <i>Thalassiosira gravida</i> (GMp14c1-3 ^a); <i>Thalassiosira miniscula</i> (CCMP1093-4 ^a); <i>Thalassiosira oceanica</i> (CCMP1005-4 ^a); <i>Thalassiosira rotula</i> (GSO102-4 ^a); <i>Thalassiosira rotula</i> (CCMP3096-2 ^a); <i>Thalassiosira weissflogii</i> (CCMP1336-4 ^a); <i>Thalassiothrix antarctica</i> (L6/D1-2 ^a); <i>Triceratium dubium</i> (CCMP147-1)	Bacillariophytes (111, 50, 36); Bolidophytes (2, 2, 1); Chrysophytes (1, 1, 1); Pelagophytes (12, 3, 3); Raphidophytes (18, 5, 2); Synurophytes (5, 2, 2); Dictyochophytes (8, 7, 6); Pinguiphytes (1, 1, 1); Xanthophytes (2, 1, 1)
<u>Bolidophytes</u> : <i>Bolidomonas pacifica</i> (CCMP1866-1); <i>Bolidomonas</i> sp. (RCC2347-1)	
<u>Chrysophytes</u> : <i>Chromulina nebulosa</i> (UTEXLB2642-1)	
<u>Pelagophytes</u> : <i>Aureococcus anophagefferens</i> (CCMP1850-4 ^a); <i>Aureoumbra lagunensis</i> (CCMP1510-4 ^a); <i>Pelagomonas calceolata</i> (CCMP1756-4 ^a)	
<u>Raphidophytes</u> : <i>Chattonella subsalsa</i> (CCMP2191-4 ^a); <i>Heterosigma akashiwo</i> (CCMP452-4 ^a ; CCMP2393-4 ^a ; CCMP3107-3 ^a ; NB-3 ^a)	
<u>Synurophytes</u> : <i>Dinobryon</i> sp (UTEXLB2267-3 ^a); <i>Ochromonas</i> sp (CCMP1393-2 ^a)	
<u>Dictyochophytes</u> : <i>Pseudopedinella elastica</i> (CCMP716-2 ^a); <i>Chrysoctysis fragilis</i> (CCMP3189-1); <i>Dictyocha speculum</i> (CCMP1381-1); <i>Florenciella parvula</i> (CCMP2471-1); <i>Florenciella</i> sp (RCC1587-1); <i>Pinguicoccus pyrenoidosus</i> (CCMP2078-1); <i>Rhizochromulina marina</i> (CCMP1243-1)	
<u>Pinguiphytes</u> : <i>Phaeomonas parva</i> (CCMP2877-1)	
<u>Xanthophytes</u> : <i>Vaucheria litorea</i> (CCMP2940-2 ^a)	
Rhizaria	
<u>Chlorarachniophyta</u> : <i>Bigelowiella longifila</i> (CCMP242-1); <i>Bigelowiella natans</i> (CCMP623-1)	Chlorarachniophyta (2, 2, 1);
<u>Haplosporidia</u> : <i>Chlorarachnion reptans</i> (CCC449-1); <i>Gymnochloa</i> sp (CCMP2014-1); <i>Lotharella amoebiformis</i> (CCMP2058-1); <i>Lotharella oceanica</i> (CCMP622-1); <i>Lotharella globosa</i> (CCC811-2 ^a); <i>Partenskyella glossopodia</i> (RCC365-1)	Haplosporidia (7, 6, 4)
Cryptophytes	
<i>Cryptomonas curvata</i> (CCAP979:52-1); <i>Geminigera cryophila</i> (CCMP2564-1); <i>Guillardia theta</i> (CCMP2712-1); <i>Hanusia phi</i> (CCMP325-1); <i>Hemiselmis andersenii</i> (CCMP644-1); <i>Hemiselmis rufescens</i> (PCC563-1); <i>Proteomonas sulcata</i> (CCMP704-1); <i>Rhodomonas</i> sp (CCMP768-2 ^a)	Cryptophytes (9, 8, 7)
Haptophytes	
<i>Calcidiscus leptoporus</i> (RCC1130-1); <i>Chrysochromulina brevifilum</i> (UTEXLB985-1); <i>Chrysochromulina polylepsis</i> (CCMP1757-4 ^a); <i>Chrysochromulina rotalis</i> (UIO044-1); <i>Chrysoculter rhomboideus</i> (RCC1486-1); <i>Coccolithus pelagicus</i> (PLY182g-1); <i>Emiliania huxleyi</i> (PLYM219-4 ^a ; CCMP370-4 ^a); <i>Exanthemachrysis gayraliae</i> (RCC1523-1); <i>Gephyrocapsa oceanica</i> (RCC1303-4 ^a); <i>Imantonia</i> sp. (RCC918-1); <i>Isochrysis</i> sp. (CCMP1244-2 ^a); <i>Isochrysis galbana</i> (CCMP1323-3 ^a); <i>Pavlova</i> sp. (CCMP459-3 ^a); <i>Pavlova gyrans</i> (CCMP608-1); <i>Pavlova lutheri</i> (RCC1537-1); <i>Phaeocystis antarctica</i> (CCMP1374-1); <i>Phaeocystis cordata</i> (RCC1383-1); <i>Pleurochrysis carterae</i> (CCMP645-3 ^a); <i>Prymnesium parvum</i> (Texoma1-4 ^a); <i>Scyphosphaera apsteinii</i> (RCC1455-1)	Haptophytes (43, 21, 14)
Viridiplantae	
<u>Chlorophyceans</u> : <i>Chlamydomonas euryale</i> (CCMP219-1); <i>Dunaliella tertiolecta</i> (CCMP1320-3 ^a); <i>Polytomella parva</i> (SAG63/3-1)	Chlorophyceans (5, 3, 3) Prasinophytes (9, 4, 3) Trebouxiophytes (2, 1, 1)
<u>Prasinophytes</u> : <i>Micromonas</i> sp. (NEPCC29-1); <i>Micromonas</i> sp. (CCMP2099-2 ^a); <i>Pyramimonas parkeae</i> (CCMP726-2 ^a); <i>Tetraselmis striata</i> (LANL1001-4 ^a)	
<u>Trebouxiophytes</u> : <i>Picocystis salinarum</i> (CCMP1897-2 ^a)	
Excavates	
<u>Euglenids</u> : <i>Eutreptiella gymnastica</i> (CCMP1594-3 ^a)	Euglenids (3, 1, 1)

Table 1 (Continued)

Organism	Total number screened (libraries, strains, genera)
<i>Negative libraries</i>	
<i>Alveolates</i>	
<i>Ciliates</i> : <i>Anophryoides haemophila</i> (AH6-1); <i>Aristerostoma</i> sp (atcc50986-1); <i>Blepharisma japonicum</i> (StockR1072-1); <i>Climacostomum virens</i> (StockW/24-1); <i>Condylostoma magnum</i> (COL2-1); <i>Fabrea salina</i> (NK [^] -1); <i>Favella ehrenbergii</i> (fehre1-1); <i>Favella taraikaensis</i> (FeNarragansettB-2 ^a); <i>Litonotus pictus</i> (P1-1); <i>Platyophrya macrostoma</i> (wh-1); <i>Protocruzia adherens</i> (Boccale-1); <i>Pseudokeronopsis</i> sp (Brazil-1; OXSARD2-1); <i>Strombidium inclinatum</i> (s3-1); <i>Uronema</i> sp (bbcil-1); <i>Strombidinopsis</i> sp (SopsisLIS2011-1)	Ciliates (17, 16, 14); Dinoflagellates (7, 2, 1)
<i>Dinoflagellates</i> : <i>Oxyrrhis marina</i> (LB1974-3 ^a), <i>Oxyrrhis marina</i> (NK [^] -4 ^a)	
<i>Stramenopiles</i>	
<i>Bicosoecids</i> : <i>Bicosoecid</i> sp (ms1-1); <i>Cafeteria roenbergensis</i> (E410-1); <i>Cafeteria</i> sp (Caron-1)	Bicosoecids (3, 3, 2); Labyrinthulids & Thraustochytrids (18, 5, 3); Chrysophytes (3, 3, 2)
<i>Labyrinthulids & Thraustochytrids</i> : <i>Aplanochytrium</i> sp (PBS07-4 ^a); <i>Aplanochytrium stocchinoi</i> (GSBS06-4 ^a); <i>Aurantiochytrium limacinum</i> (ATCCMYA1381-4 ^a); <i>Schizochytrium aggregatum</i> (ATCC28209-4 ^a); <i>Thraustochytrium</i> sp (LLF1b-2 ^a)	
<i>Chrysophytes</i> : <i>Paraphysomonas vestita</i> (GFlagA-1); <i>Spumella elongata</i> (CCAP955-1); <i>Paraphysomonas bandaiensis</i> (Caron-1)	
<i>Rhizaria</i>	
<i>Foraminifera</i> : <i>Ammonia</i> sp (NK [^] -1); <i>Elphidium margaritaceum</i> (NK [^] -1); <i>Rosalina</i> sp (NK [^] -1); <i>Sorites</i> sp (NK [^] -1)	Foraminifera (4, 4, 4)
<i>Amoebozoa</i>	
<i>Tubulinids</i> : <i>Filamoeba nolandii</i> (NCAS231-1)	Tubulinids: (1, 1, 1);
<i>Dactylopodids</i> : <i>Mayorella</i> sp (BSH0219019-1); <i>Neoparamoeba aestuarina</i> (SoJaBio-1); <i>Paramoeba atlantica</i> (CCAP15609-1); <i>Vexillifera</i> sp (DIVA3/564-1)	Dactylopodids (4, 4, 4)
<i>Excavates</i>	
<i>Kinetoplastids</i> : <i>Neobodo designis</i> (CCAP1951-1)	Kinetoplastids (1, 1, 1)
<i>Opisthokonts</i>	
<i>Choanoflagellates</i> : <i>Acanthoeca</i> like sp. (10tr-2 ^a)	Choanoflagellates (2, 1, 1)
<i>Others</i>	
<i>Glaucophytes</i> : <i>Cyanoptycha gloeocystis</i> (SAG497-1); <i>Gloeochaete witrockiana</i> (SAG4684-2 ^a)	Glaucophytes (3, 2, 2); Palpitomonas (1, 1, 1)
<i>Palpitomonas</i> : <i>Palpitomonas bilix</i> (NIES2562-1)	
Total	(402, 213, 152)

Species names are followed by the strain number and then number of transcriptomic libraries screened for each strain. In strains where more than one library was available, a combined assembly of all the libraries constructed was used to search for type II fatty acid synthase genes. NK[^]- strain identity not known.

^aA combined assembly of all the libraries constructed was used to search for fatty acid synthase genes.

et al., 2010). Here, KASII was confirmed in all phototrophic lineages of protists (Supplementary Table S1); however, KASI was absent in Rhizaria, Viridiplantae, raphidophytes, synurophytes, pinguiphytes, xanthophytes, *Vitrella* and *Chromera* (Supplementary Table S1). Our results show the presence of at least six different gene families that encode KAS II (Supplementary Figure S1C) suggesting the production of different types of fatty acids. The active site residues Cys-His-His in KASI and KASII and Cys-His-Asn in KASIII found in higher plants and bacteria (White *et al.*, 2005 and references therein) are conserved in protists (Figure 1).

There are two types of dehydratases, that is, DH-FabA and DH-FabZ, described in bacteria and higher plants (White *et al.*, 2005 and references therein). DH-FabA has the additional function of performing isomerisation (in addition to dehydration) essential for formation of unsaturated fatty acids and normally co-occurs with KASI-FabB (White *et al.*, 2005 and references therein). However, in protists only genes encoding the DH-FabZ enzyme were found. DH-FabA was absent from all lineages (Supplementary Table S1).

Evolution of type II FAS in protists

Several lines of evidence from our analysis support the notion that type II FAS genes are nuclear encoded, and that the initial steps of fatty acid synthesis take place in the chloroplast: (1) the presence of all seven type II FAS enzymes in the reference genome of *E. huxleyi* (Supplementary Table S2); (2) the detection of transit peptides targeted towards the chloroplast (chloroP; Emanuelsson *et al.*, 1999) in ~80% of the sequences (Supplementary Table S1); (3) the presence of eukaryotic polyA tails; and (4) 5' trans-spliced leader sequences (Zhang *et al.*, 2007) on dinoflagellate type II FAS gene transcripts. These features suggest that type II FAS genes were transferred from the plastid to the host genome at some point during their evolutionary history. Transfer of plastidial genes to the nucleus in the host organism allows for selection processes to act on genes according to their functional advantage to the host and restricts the accumulation of deleterious mutations (Muller's ratchet; Felsenstein, 1974) in the endosymbiotic genome, which cannot recombine. The nuclear location also provides protection from reactive

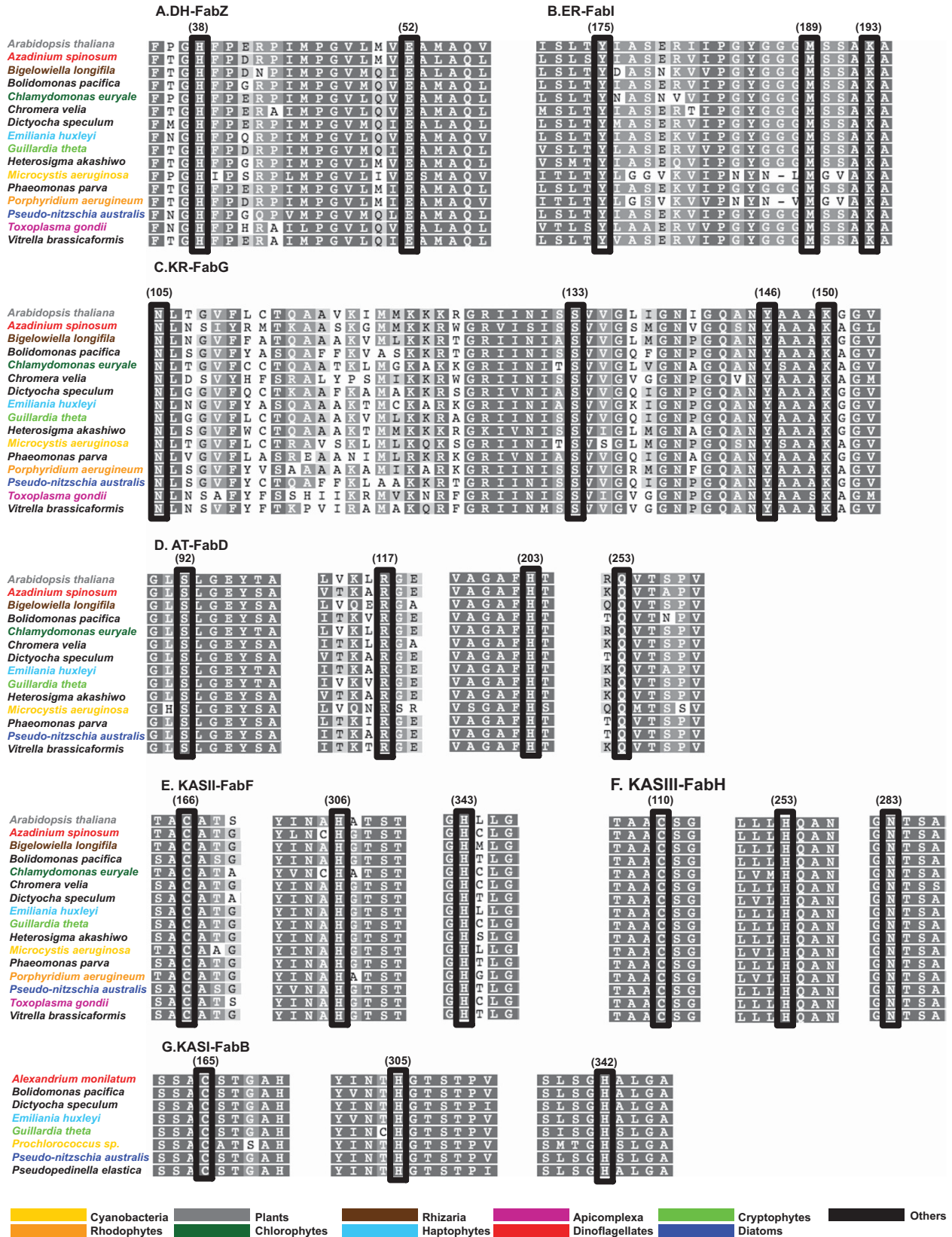


Figure 1 Conserved active sites in key fatty acid synthase enzymes in eukaryotes: 3-hydroxyacyl-ACP dehydratase (a), enoyl-ACP reductase (b), 3-ketoacyl ACP reductase (c), S-malonyltransacylase (d), 3-ketoacyl ACP synthase II (e), 3-ketoacyl ACP synthase III (f) and 3-ketoacyl ACP synthase I (g). Active site residues are highlighted in black boxes and numbers above residues are according to the *Azadinium spinosum* sequences except for *Alexandrium monilatum* sequence in (g).

oxygen species generated during the process of photosynthesis (Martin *et al.*, 1998; Martin 2003; Deusch *et al.*, 2008; McFadden 2014). As fatty acid synthesis is essential for survival, these genes were likely retained by protists in the nucleus due to strong selective pressure.

The origin of protistan plastids has been traced to either an ancestral rhodophyte or chlorophyte, with a clear distinction between the two clades (Janouškovec *et al.*, 2010). Evidence presented here indicates two possible scenarios in which either (i) type II FAS genes may have been transferred from the plastid to the nuclear genome in an early ancestral protist, before the initial split of the rhodophyte and chlorophyte lineages (Figures 2a and b) 1750–2000 million years ago (Parfrey *et al.*, 2011) or (ii) there has been a more recent transfer event from the ancestor of chlorophytes/rhodophytes to the ancestors of stramenopiles, alveolates,

Rhizaria, haptophytes and cryptophytes (Figures 2a and b), which happened 1250–1500 million years ago (Parfrey *et al.*, 2011). Within the type II FAS clade, the evolution of these genes broadly follows the trend of microbial eukaryotic evolution (Keeling, 2013; Keeling *et al.*, 2014) in which chlorophytes, higher plants, Rhizaria, haptophytes and alveolates form separate monophyletic clades, with the exception of cryptophytes, which form a monophyletic clade placed within the polyphyletic stramenopile clade (Figure 3). In addition, evidence presented here supports the hypothesis that some type II FAS genes were transferred more recently from the plastid to the nucleus of its tertiary hosts, specifically from haptophytes to their dinoflagellate hosts *Karenia* spp. and *Karlodinium* spp. (Yoon *et al.*, 2002; Figure 3 and Supplementary Figure S1) and from diatoms to their dinoflagellate hosts *Glenodinium* spp, *Durinskia* spp and *Kryptoperidium* spp

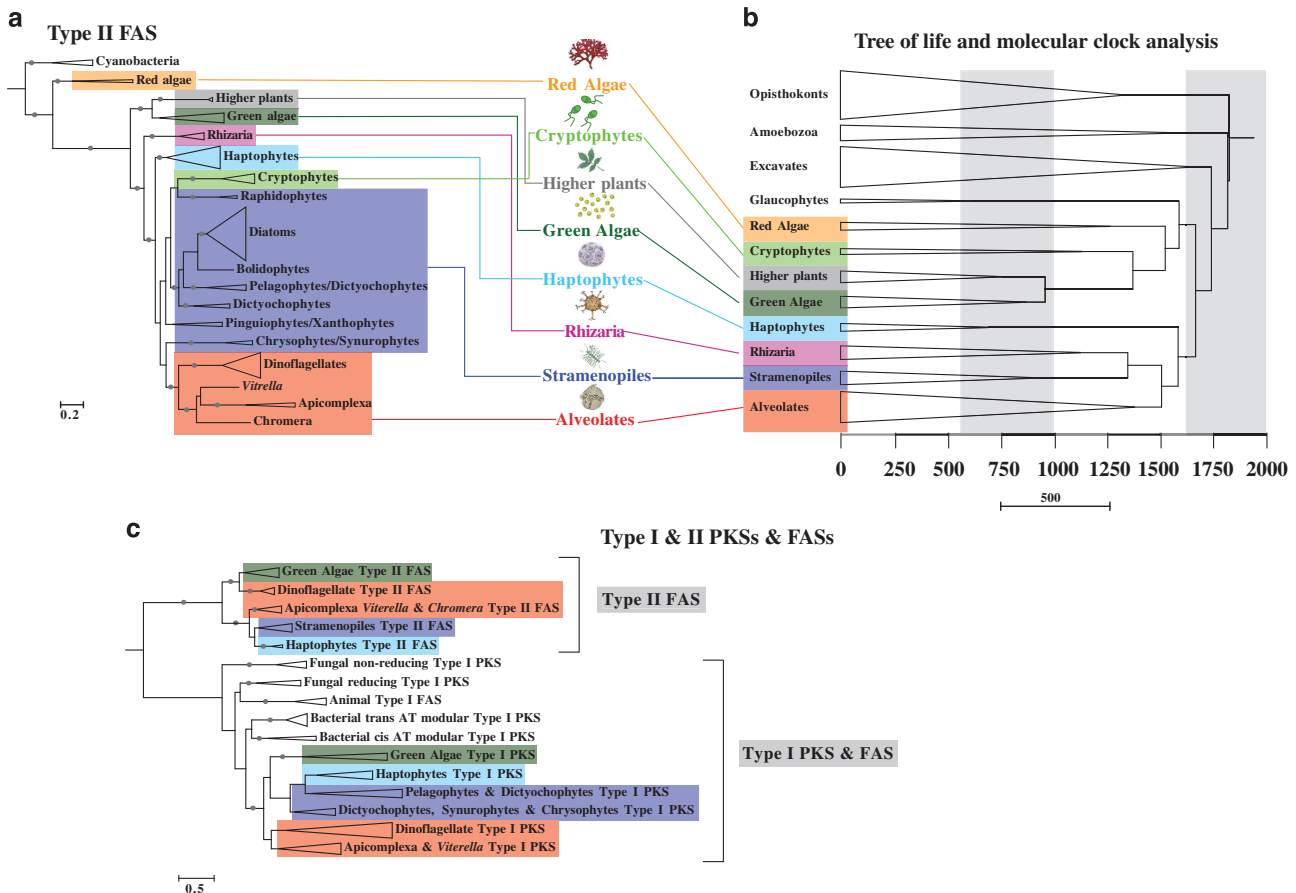


Figure 2 Comparative evolution of fatty acid and polyketide synthase. (a) Concatenated phylogeny, inferred from protein sequences of five enzymes (3-ketoacyl ACP synthase III; S-malonyltransacylase; 3-hydroxyacyl-ACP dehydratase; enoyl-ACP reductase; trans3-ketoacyl ACP reductase, 1431 characters) involved in type II fatty acid synthesis (inferred using RAxML, GAMMA model of rate heterogeneity, 1000 bootstraps). Solid circles indicate bootstrap values above ≥ 90 . (b) For comparison, a dated molecular clock phylogeny of the eukaryotic tree of life, showing absolute time scale (million years) (from Parfrey *et al.*, 2011). These phylogenetic analyses show that evolution of fatty acid synthase genes broadly follow the evolution pattern of the organism. (c) Phylogenetic analysis of 25 type II 3-ketoacyl ACP synthase II and 67 type I ketosynthase domains from prokaryotic and eukaryotic polyketide synthases and fatty acid synthases, showing the position of each major group, inferred in RAxML using GAMMA model of rate heterogeneity and 1000 bootstraps. Solid circles indicate bootstrap values above ≥ 90 . Owing to relaxed selection pressure, polyketide synthase genes were retained/lost by protists based on the functionality their polyketide product provided the organism.

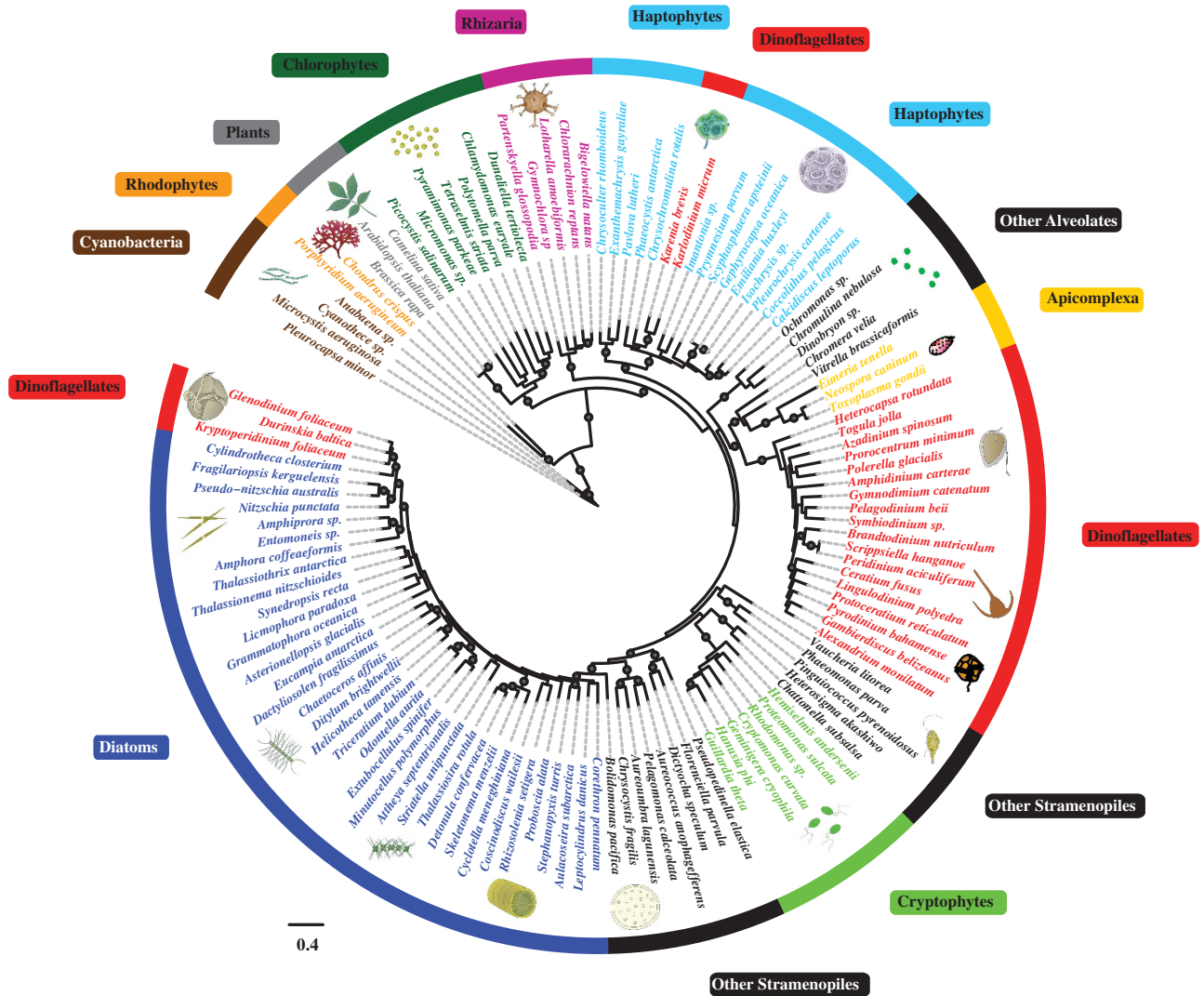


Figure 3 Fatty acid synthase gene phylogeny in eukaryotes: concatenated phylogeny of five enzymes involved in type II fatty acid synthesis: 3-ketoacyl ACP synthase III; S-malonyltransacylase; 3-hydroxyacyl-ACP dehydratase; enoyl-ACP reductase; trans3-ketoacyl ACP reductase (1431 characters), inferred using RAXML, GAMMA model of rate heterogeneity, 1000 bootstraps. Solid circles indicate bootstrap values above ≥ 90 .

(Imanian *et al.*, 2012; Figure 3 and Supplementary Figure S1), some 250–750 million years ago (Figure 2b; Parfrey *et al.*, 2011).

Polyketide synthesis in protists

We found an enormous diversity of Type I PKS genes in selected alveolates (dinoflagellates: 46 strains, 24 genera; *Vitrella*: 1 genus), stramenopiles (labyrinthulids, thraustochytrids, chrysophytes, pelagophytes, synurophytes, dictyochophytes: 15 genera), haptophytes (12 strains, 9 genera) and chlorophytes (6 strains, 5 genera) (Figure 4 and Supplementary Table S3). We confirm the absence of expressed Type I PKS genes in other alveolates (chromera, ciliates: 17 strains, 15 genera), stramenopiles (bacillariophytes, bicosoecids, bolidophytes, raphidophytes, pinguiophytes, xanthophytes: 62 strains,

43 genera), Rhizaria (chlorarachniophytes, haplosporidia, foraminifera: 12 strains, 9 genera), cryptophytes (8 strains, 7 genera), glaucophytes (2 genera) and *Palpitomonas bilix* (Figure 4 and Supplementary Table S3). In the Type I PKSs of stramenopiles, haptophytes, chlorophytes and *Vitrella*, each transcript encoded multiple PKS domains, and 286 contigs encoding multiple type I PKS domains were found (Supplementary Table S3). Increased expression of certain PKS genes has been indirectly linked to higher toxin production in the haptophyte *Prymnesium parvum* (Freitag *et al.*, 2011); therefore, the presence of these genes in the haptophyte *Chrysochromulina polylepis* and the stramenopile *Aureococcus anophagefferens* is intriguing, as they produce polyketide toxins that cause fish kills (John *et al.*, 2010; Freitag *et al.*, 2011; Gobler *et al.*, 2011).

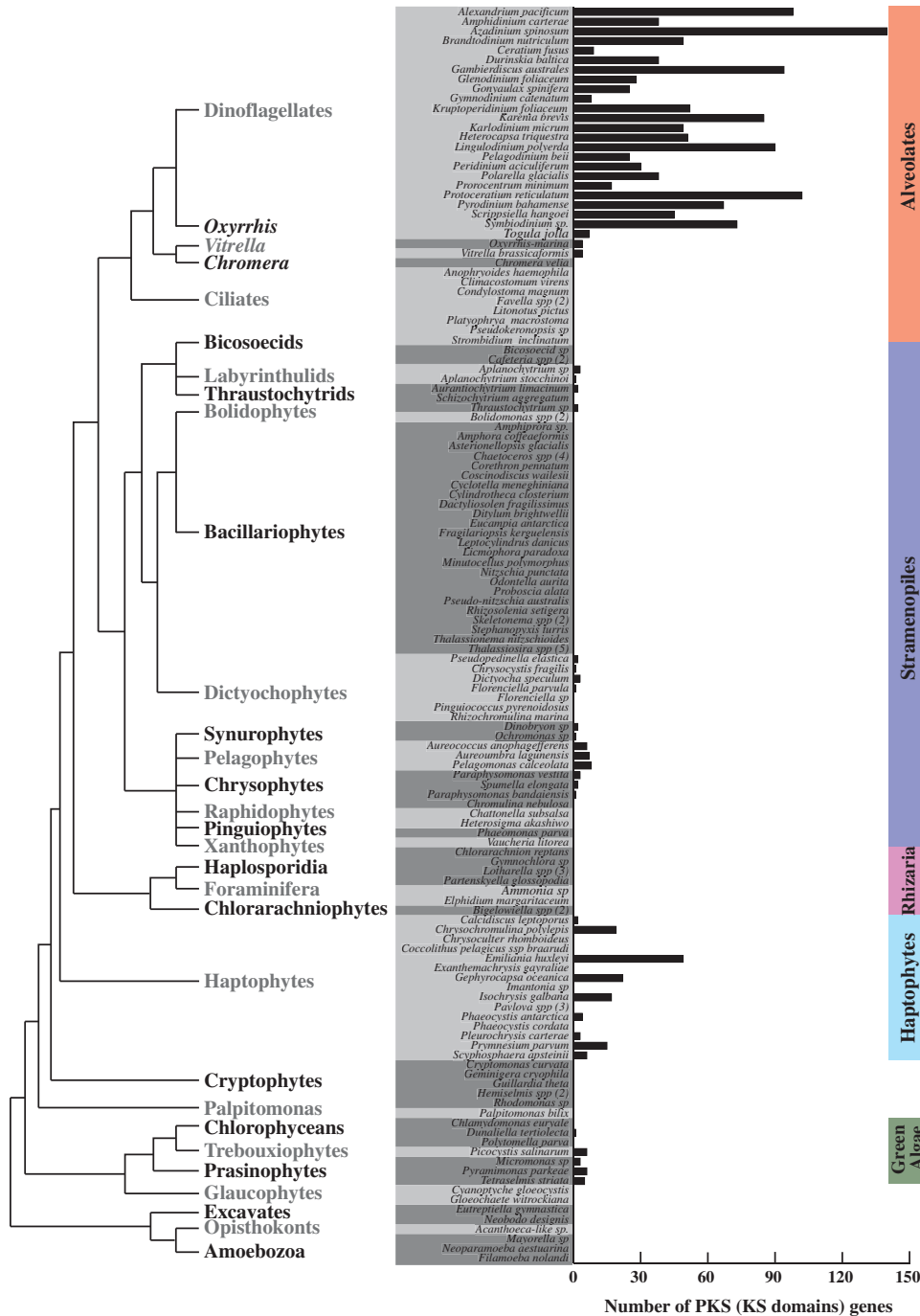


Figure 4 Survey of polyketide synthase genes in eukaryotes: The figure shows the abundance of expressed type I polyketide synthases (PKS)-ketoacyl synthase (KS) domains from various eukaryotic lineages. The KS domain gene family is highly expanded in dinoflagellates and haptophytes, and also present in *Vitrella*, labyrinthulids, thraustochytrids, chrysophytes, pelagophytes, synurophytes, dictyochophytes, chlorophyceans, trebouxiophytes and prasinophytes. The KS domains were absent in *Chromera*, ciliates, bacillariophytes, bicosoecids, bolidophytes, raphidophytes, pinguiphytes, xanthophytes, chlorarachniophytes, haplosporidia, foraminifera, cryptophytes, glaucophytes and *Palpitomonas*.

In the reference genome of *E. huxleyi* CCMP1516 (Read *et al.*, 2013), a total of 30 contigs encoding multiple type I PKS genes were found (Supplementary Table S2). A comprehensive expressed sequenced tag (EST) library containing sequences from 14 isolates of *E. huxleyi* (Read *et al.*, 2013) was used to study the expression of type I PKS genes

found in the *E. huxleyi* genome. Transcripts corresponding to 26 contigs found in the *E. huxleyi* genome were observed in the EST library of *E. huxleyi* (Supplementary Table S2). It is possible that type I PKS genes on the other four contigs were not being expressed at the time of analysis, or not present in the transcriptome due to an insufficient

depth of sequencing. Interestingly, seven sequences encoding partial type I PKS genes (KS domains) from the EST library were not found in the reference genome of *E. huxleyi* CCMP1516 (Supplementary Table S2). Read *et al.* (2013) found that *E. huxleyi* has a pan genome where certain genes are variably distributed between different strains. This might explain the absence of seven sequence encoding type I PKS genes from the reference genome of *E. huxleyi* CCMP1516.

To date, no gene has been definitively linked to the synthesis of a particular polyketide toxin produced by a eukaryotic harmful algal bloom species. This is in part due to the difficulty in producing genetically transformable protists. Therefore, genetic screening methods for detecting patterns of genes expressed by toxin-producing protists have been the most fruitful approach to date (Bachvaroff and Place, 2008; John *et al.*, 2008, 2010; Monroe and Van Dolah, 2008; Stüken *et al.*, 2011; Eichholz *et al.*, 2012; Murray *et al.*, 2012; Pawlowicz *et al.*, 2014; Meyer *et al.*, 2015). Through transcriptomic analysis, PKS genes in dinoflagellates were found to be evolutionarily related to type I PKS, but expressed as monofunctional proteins, a feature synonymous to type II PKSs (Millar *et al.*, 1999; Monroe and Van Dolah 2008; Eichholz *et al.*, 2012; Salcedo *et al.*, 2012; Pawlowicz *et al.*, 2014; Kohli *et al.*, 2015; Meyer *et al.*, 2015). In this study, we identified a much larger range of unique KS domains than expected or previously found: an average of 56/strain, comprising a total of 2577 unique KS domains (1976 full and 601 partial) and 234 KR domains (190 full and 44 partial) in 24 genera and 46 strains of dinoflagellates (Figure 4 and Supplementary Tables S3 and S4). *Azadinium spinosum*, which produces the polyketide toxin azaspiracid and its analogues, had the largest number of KS domains: 140 (Meyer *et al.*, 2015), while the non-toxic dinoflagellate species *Togula jolla* encoded seven KS domains (Supplementary Table S4). Like type I and II FASs, KS domains in type I and type II PKS have very conserved active site residues, Cys-His-His, which are essential for their functionality (Kwon *et al.*, 2002), and their presence was confirmed in 66% of our sequences (Supplementary Table S4). Thus, PKS gene families appear to have expanded dramatically within the dinoflagellates, suggesting numerous duplications and the evolution of novel functions.

Previously, no KS domains resembling type I or type II FAS have been found in dinoflagellates, possibly because of low sequencing depth, which led to the hypothesis that dinoflagellate fatty acid synthesis is carried out by enzymes resembling PKS enzymes (Pawlowicz *et al.*, 2014). However, our results suggest fatty acid synthesis is likely carried out by type II FAS in dinoflagellates. These findings are important, as the differentiation of PKS and FAS will facilitate approaches to investigating harmful algal toxin biosynthesis pathways in

dinoflagellates. The inferred distinction between type II FAS and type I PKS genes in dinoflagellates are based on sequence analysis, and a functional proof remains to be done.

The phylogeny of PKS KS and KR domains shows that protistan KS and KR domains form a monophyletic group within which dinoflagellate, chlorophyte, haptophyte and apicomplexan KS and KR domains form monophyletic clades (John *et al.*, 2008; Monroe and Van Dolah 2008) (KS—Figure 2c, KR—Supplementary Figure S5). KS domains from stramenopiles also form two well-supported clades within the protistan clade (Figure 2c). The phylogeny of 1591 KS domains within the dinoflagellate clade further shows that their KS domains form three distinct clades (Figure 5), each of which includes sequences from numerous species of multiple dinoflagellate orders, clearly not related to the species phylogeny (Orr *et al.*, 2012). No clear pattern could be established between these three clades and the chemical structure of the compounds known to be produced by these organisms. For example, *Karenia brevis*, which produces over 15 polyketide compounds (Baden *et al.*, 2005), had sequences in each of the three clades of dinoflagellate PKS (Figure 5), as did the species *Togula jolla* (not shown), which is not known to produce any toxins. It is likely that dinoflagellates produce many polyketide compounds that are as yet undetected and uncharacterised. Polyketide compounds produced by dinoflagellates that do not clearly impact fisheries or aquaculture industries are likely to have been unnoticed by researchers.

There is a growing body of evidence for the ecological benefits of some marine microbial eukaryotic toxins in the form of antipredator or allelopathic impacts (Cembella, 2003; Ianora *et al.*, 2011 and references therein), given the importance of grazing as a selective force in the marine planktonic ecosystem (Smetacek, 2001). While only one or few compounds may be necessary to produce these ecological impacts, the presence of the genetic basis for the production of a vast number of distinctive polyketide compounds within a species may be related to the Screening Hypothesis (Jones *et al.*, 1991; Firn and Jones 2003), based on the principle that 'potent biological activity is a rare property for any one molecule to possess' (Jones *et al.*, 1991; Firn and Jones 2003). This would predict that organisms that produce and screen a larger variety of chemical compounds have an increased likelihood of enhanced fitness, as the chance of producing a rare chemical with a useful biological activity will be increased. An example of this may be the production of many different congeners of brevetoxins and ciguatoxins by *Karenia brevis* and *Gambierdiscus* spp. respectively (Kalaitzis *et al.*, 2010 and references therein), which differ from one another in biological activity (Chinain *et al.*, 2010). In dinoflagellates the lack of correlation with the species phylogeny (Figure 5), and the large intraspecific

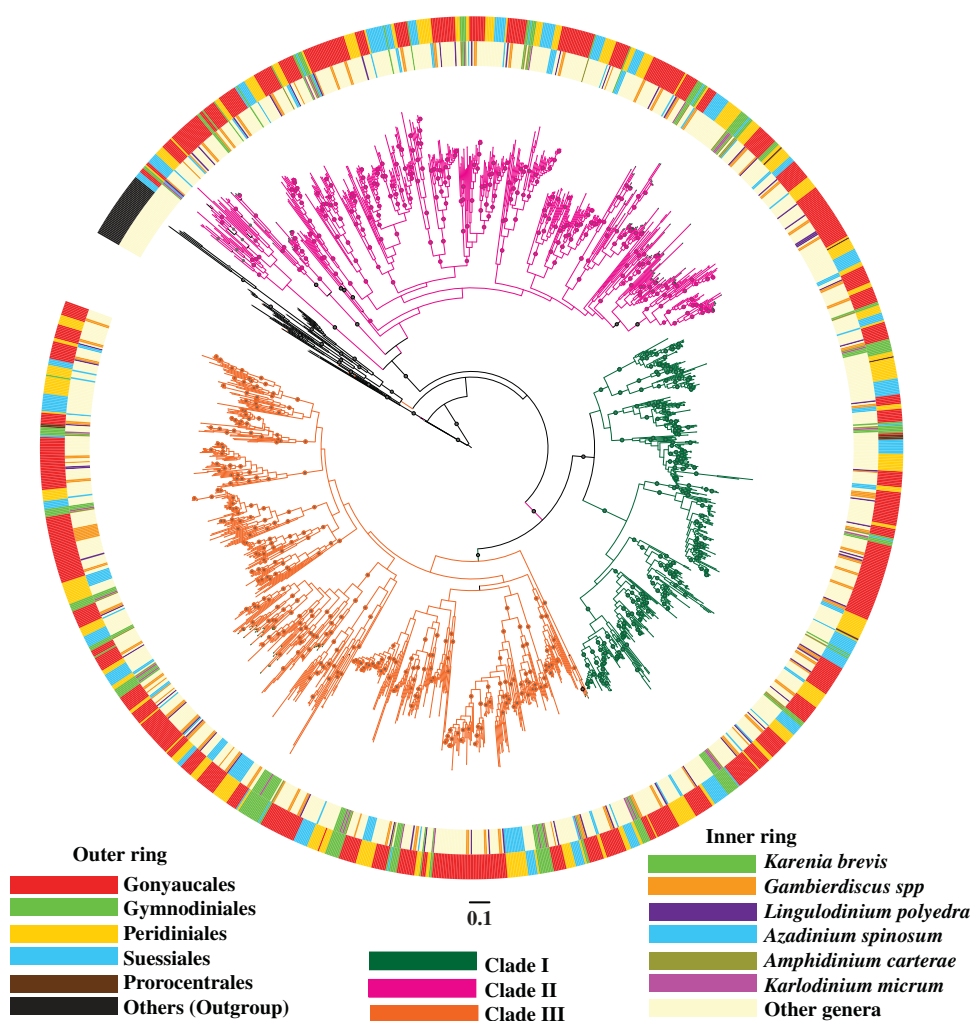


Figure 5 Polyketide synthase gene phylogeny in dinoflagellates: Phylogenetic analysis of type I ketoacyl synthase (KS) domains from prokaryotic and eukaryotic polyketide synthases (PKS) and fatty acid synthases (FAS). In total, 1633 KS domains representing 43 dinoflagellate and 30 other prokaryotic and eukaryotic taxa were inferred using RAxML, GAMMA model of rate heterogeneity and 1000 bootstraps (653 characters). PKS gene families are highly expanded in dinoflagellates, forming three distinct clades (clades I–III coloured in green, pink and orange colours respectively), where the pattern of distribution is not related to the species phylogeny and/or the chemical structure these organisms produce. Solid circles indicate bootstrap values above ≥ 80 . Clade labelled as outgroup/others consist of type I PKS-KS domains from fungi (reducing/non-reducing) and bacteria (*cis* and *trans* AT modular), type I FAS-KS domains from animals and type II PKS-KS from bacteria.

diversity in KS domains, suggests that multiple gene duplication events, domain shuffling and losses have occurred. This suggests that relaxed selection pressures have acted on the evolution of these secondary metabolite genes (Kroymann 2011; Weng *et al.*, 2012), which may have been acquired or lost based on the functionality they provided to the organism (Murray *et al.*, 2015).

Ecological experiments have been used to determine the predicted function of some polyketide compounds for the producing marine microbial eukaryotes (Cembella 2003; Ianora *et al.*, 2011 and references therein). The elucidation of genes involved in polyketide synthesis in these organisms opens up the possibility that these ecological roles can be further investigated by examining factors effecting gene regulation, and by producing

genetically transformed knockouts. This information will be crucial in revealing the biochemical and molecular basis of marine microbial eukaryotic community interactions.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dr Gaurav Sablok for help in running ChloroP analysis and Dr Fabien Burki for providing advice on protist phylogeny. This research was supported by Australian Research Council grants (FT120100704, DP120103199) to SM and Australian Academy of Science Germany-Australia mobility call 2013–2014

and BMBF PT/DLR IB FKZ 01DR14006 awarded to SM and UJ. Financial support to UJ was provided by the PACES research program of the Alfred-Wegener-Institute Helmholtz-Zentrum für Polar- und Meeresforschung. Financial support to FV was provided by National Oceanic and Atmospheric Administration NCCOS project 02E00025. This is contribution number 170 from the Sydney Institute of Marine Science.

References

- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH *et al.* (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**: 79–86.
- Bachvaroff TR, Place AR. (2008). From stop to start: tandem gene arrangement, copy number and trans splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS One* **3**: e2929.
- Baden DG, Bourdelais AJ, Jacocks H, Michelliza S, Naar J. (2005). Natural and derivative brevetoxins: historical background, multiplicity, and effects. *Environ Health Perspect* **113**: 621–625.
- Bayer T, Aranda M, Sunagawa S, Yum LK, DeSalvo MK, Lindquist E *et al.* (2012). *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS One* **7**: e35269.
- Brown AP, Slabas AR, Rafferty JB. (2010). Fatty acid biosynthesis in plants—metabolic pathways, structure and organization. In: Wada H, Murata N (eds), *Lipids in Photosynthesis*. Springer: Dordrecht, The Netherlands, pp 11–34.
- Cane DE, Walsh CT, Khosla C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**: 63–68.
- Cembella AD. (2003). Chemical ecology of eukaryotic microalgae in marine ecosystems. *Phycologia* **42**: 420–447.
- Chinain M, Darius HT, Ung A, Cruchet P, Wang Z, Ponton D *et al.* (2010). Growth and toxin production in the ciguatera-causing dinoflagellate *Gambierdiscus polynesiensis* (Dinophyceae) in culture. *Toxicon* **56**: 739–750.
- Cock JM, Sterck L, Rouze P, Scornet D, Allen AE, Amoutzias G *et al.* (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**: 617–621.
- Deusch O, Landan G, Roettger M, Gruenheit N, Kowallik KV, Allen JF *et al.* (2008). Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol* **25**: 748–761.
- Dunn TM, Lynch DV, Michaelson LV, Napier JA. (2004). A post-genomic approach to understanding uphingo-lipid metabolism in *Arabidopsis thaliana*. *Ann Bot* **93**: 483–497.
- Eichholz K, Beszteri B, John U. (2012). Putative monofunctional type I polyketide synthase units: a dinoflagellate-specific feature? *PLoS One* **7**: e48624.
- Emanuelsson O, Nielsen H, Von Heijne G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* **8**: 978–984.
- Felsenstein J. (1974). The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- Fiebig A, Mayfield JA, Miley NL, Chau S, Fischer RL, Preuss D. (2000). Alterations in CER6, a gene identical to CUT1, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant Cell* **12**: 2001–2008.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237–240.
- Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.
- Finn RD, Jones CG. (2003). Natural products—a simple model to explain chemical diversity. *Nat Prod Rep* **20**: 382–391.
- Freitag M, Beszteri S, Vogel H, John U. (2011). Effects of physiological shock treatments on toxicity and polyketide synthase gene expression in *Prymnesium parvum* (Prymnesiophyceae). *Eur J Phycol* **46**: 193–201.
- Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X *et al.* (2008). ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Res* **36**: D553–D556.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW *et al.* (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498–511.
- Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, Lobanov AV *et al.* (2011). Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc Natl Acad Sci USA* **108**: 4352–4357.
- Hackett JD, Wisecaver JH, Brosnahan ML, Kulis DM, Anderson DM, Bhattacharya D *et al.* (2013). Evolution of saxitoxin synthesis in cyanobacteria and dinoflagellates. *Mol Biol Evol* **30**: 70–78.
- Hallegraeff GM. (1993). A review of harmful algal blooms and their apparent global increase. *Phycologia* **32**: 79–99.
- Hallegraeff GM. (2010). Ocean climate change, phytoplankton community responses, and harmful algal blooms: a formidable predictive challenge. *J Phycol* **46**: 220–235.
- Ianora A, Bentley MG, Caldwell GS, Casotti R, Cembella AD, Engström-Öst J *et al.* (2011). The relevance of marine chemical ecology to plankton and ecosystem function: an emerging field. *Mar Drugs* **9**: 1625–1648.
- Imanian B, Pombert J-F, Dorrell RG, Burki F, Keeling PJ. (2012). Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. *PLoS One* **7**: e43763.
- Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci USA* **107**: 10949–10954.
- Jenke-Kodama H, Sandmann A, Müller R, Dittmann E. (2005). Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* **22**: 2027–2039.
- John U, Beszteri B, Derelle E, Van de Peer Y, Read B, Moreau H *et al.* (2008). Novel insights into evolution of protistan polyketide synthases through phylogenomic analysis. *Protist* **159**: 21–30.
- John U, Beszteri S, Glöckner G, Singh R, Medlin L, Cembella AD. (2010). Genomic characterisation of the ichthyotoxic prymnesiophyte *Chrysochromulina polylepsis*, and the expression of polyketide synthase genes in synchronized cultures. *Eur J Phycol* **45**: 215–229.

- Jones CG, Firm RD, Malcolm SB. (1991). On the evolution of plant secondary chemical diversity [and Discussion]. *Philos Trans R Soc Lond, Ser B: Biol Sci* **333**: 273–280.
- Kalaitzis JA, Chau R, Kohli GS, Murray SA, Neilan BA. (2010). Biosynthesis of toxic naturally-occurring seafood contaminants. *Toxicon* **56**: 244–258.
- Katoh K, Misawa K, Ki Kuma, Miyata T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S *et al*. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.
- Keeling PJ. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Biol* **64**: 583–607.
- Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA *et al*. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* **12**: e1001889.
- Khosla C, Gokhale RS, Jacobsen JR, Cane DE. (1999). Tolerance and specificity of polyketide synthases. *Annu Rev Biochem* **68**: 219–253.
- Kohli GS, John U, Figueroa RI, Rhodes LL, Harwood DT, Groth M *et al*. (2015). Polyketide synthesis genes associated with toxin production in two species of *Gambierdiscus* (Dinophyceae). *BMC Genomics* **16**: 410.
- Kroymann J. (2011). Natural diversity and adaptation in plant secondary metabolism. *Curr Opin Plant Biol* **14**: 246–251.
- Kunst L, Taylor DC, Underhill EW. (1992). Fatty acid elongation in developing seeds of *Arabidopsis thaliana*. *Plant Physiol Biochem* **30**: 425–434.
- Kwon H-J, Smith WC, Scharon AJ, Hwang SH, Kurth MJ, Shen B. (2002). CO bond formation by polyketide synthases. *Science* **297**: 1327–1330.
- Letunic I, Bork P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475–W478.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY *et al*. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**: D222–D226.
- Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**: 162–165.
- Martin W. (2003). Gene transfer from organelles to the nucleus: frequent and in big chunks. *Proc Natl Acad Sci USA* **100**: 8612–8614.
- McFadden GL. (1999). Plastids and protein targeting. *J Eukaryot Microbiol* **46**: 339–346.
- McFadden GL. (2014). Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb Perspect Biol* **6**: a016105.
- Meyer JM, Rödelsperger C, Eichholz K, Tillmann U, Cembella A, McGaughran A *et al*. (2015). Transcriptomic characterisation and genomic glimpse into the toxigenic dinoflagellate *Azadinium spinosum*, with emphasis on polyketide synthase genes. *BMC Genomics* **16**: 27.
- Millar AA, Kunst L. (1997). Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. *Plant J* **12**: 121–131.
- Millar AA, Clemens S, Zachgo S, Giblin EM, Taylor DC, Kunst L. (1999). CUT1, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell* **11**: 825–838.
- Monroe EA, Van Dolah FM. (2008). The toxic dinoflagellate *Karenia brevis* encodes novel type I-like polyketide synthases containing discrete catalytic domains. *Protist* **159**: 471–482.
- Murray SA, Garby T, Hoppenrath M, Neilan BA. (2012). Genetic diversity, morphological uniformity and polyketide production in dinoflagellates (*Amphidinium*, Dinoflagellata). *PLoS One* **7**: e38253.
- Murray S A, Diwan R, Orr R J, Kohli G S, John U. (2015). Gene duplication, loss and selection in the evolution of saxitoxin biosynthesis in alveolates. *Mol Phylogenet Evol* **92**: 165–180.
- Orr RJS, Murray SA, Stüken A, Rhodes L, Jakobsen KS. (2012). When naked became armored: An eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS One* **7**: e50004.
- Parfrey LW, Lahr DJG, Knoll AH, Katz LA. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci USA* **108**: 13624–13629.
- Pawlowicz R, Morey JS, Darius HT, Chinain M, Van Dolah FM. (2014). Transcriptome sequencing reveals single domain Type I-like polyketide synthases in the toxic dinoflagellate *Gambierdiscus polynesiensis*. *Harmful Algae* **36**: 29–37.
- Place A. (2008). Sorting the fatty acid chaff from the toxin wheat, or is it all wheat?—assigning dinoflagellate PKS genes to toxin synthesis. In: Kim HG, Reguera B, Hallegraeff GM, Lee CK, Han MS, Choi JK (eds). *Proceedings of the 15th International Conference on Harmful Algae*. International Society for the Study of Harmful Algae 2014, ISBN 978-87-990827-4-2, pp 170–176.
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C *et al*. (2012). The Pfam protein families database. *Nucleic Acids Res* **40**: D290–D301.
- Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F *et al*. (2013). Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* **499**: 209–213.
- Riekhof WR, Benning C (2009). Glycerolipid biosynthesis. In: Stern D (ed.), *The Chlamydomonas Sourcebook: Organellar and Metabolic Processes*. Academic Press: San Diego, CA, USA, pp 41–68.
- Ryall K, Harper JT, Keeling PJ. (2003). Plastid-derived Type II fatty acid biosynthetic enzymes in chromists. *Gene* **313**: 139–148.
- Salcedo T, Upadhyay RJ, Nagasaki K, Bhattacharya D. (2012). Dozens of toxin-related genes are expressed in a nontoxic strain of the dinoflagellate *Heterocapsa circularisquama*. *Mol Biol Evol* **29**: 1503–1506.
- Smetacek V. (2001). A watery arms race. *Nature* **411**: 745–745.
- Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stüken A, Orr RJS, Kellmann R, Murray SA, Neilan BA, Jakobsen KS. (2011). Discovery of nuclear-encoded genes for the neurotoxin saxitoxin in dinoflagellates. *PLoS One* **6**: e20096.

- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.
- Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Waller RF, Keeling PJ, Donald RG, Striepen B, Handman E, Lang-Unnasch N *et al.* (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **95**: 12352–12357.
- Weng J-K, Philippe RN, Noel JP. (2012). The rise of chemodiversity in plants. *Science* **336**: 1667–1670.
- White SW, Zheng J, Zhang Y-M, Rock CO. (2005). The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* **74**: 791–831.
- Wisecaver JH, Brosnahan ML, Hackett JD. (2013). Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. *Genome Biol Evol* **5**: 2368–2381.
- Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.
- Yoon HS, Hackett JD, Bhattacharya D. (2002). A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci USA* **99**: 11724–11729.
- Zhang H, Hou Y, Miranda L, Campbell DA, Sturm NR, Gaasterland T *et al.* (2007). Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci USA* **104**: 4618–4623.
- Zhu G, Marchewka MJ, Woods KM, Upton SJ, Keithly JS. (2000). Molecular analysis of a Type I fatty acid synthase in *Cryptosporidium parvum*. *Mol Biochem Parasitol* **105**: 253–260.



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

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)