

Article

Monogalactosyldiacylglycerol and Sulfolipid Synthesis in Microalgae

Gennaro Riccio ¹, Daniele De Luca ²  and Chiara Lauritano ^{1,*} 

¹ Department of Marine Biotechnology, Stazione Zoologica Anton Dohrn, CAP80121 Naples, Italy; gennaro.riccio@szn.it

² Department of Humanities, Università degli Studi Suor Orsola Benincasa, CAP80135 Naples, Italy; daniela.deluca088@gmail.com

* Correspondence: chiara.lauritano@szn.it; Tel.: +39-081-5833-221

Received: 12 April 2020; Accepted: 27 April 2020; Published: 1 May 2020



Abstract: Microalgae, due to their huge taxonomic and metabolic diversity, have been shown to be a valuable and eco-friendly source of bioactive natural products. The increasing number of genomic and transcriptomic data will give a great boost for the study of metabolic pathways involved in the synthesis of bioactive compounds. In this study, we analyzed the presence of the enzymes involved in the synthesis of monogalactosyldiacylglycerols (MGDGs) and sulfoquinovosyldiacylglycerols (SQDG). Both compounds have important biological properties. MGDGs present both anti-inflammatory and anti-cancer activities while SQDGs present immunostimulatory activities and inhibit the enzyme glutaminyl cyclase, which is involved in Alzheimer's disease. The Ocean Global Atlas (OGA) database and the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) were used to search MGDG synthase (MGD), UDP-sulfoquinovose synthase (SQD1), and sulfoquinovosyltransferase (SQD2) sequences along microalgal taxa. *In silico* 3D prediction analyses for the three enzymes were performed by Phyre2 server, while binding site predictions were performed by the COACH server. The analyzed enzymes are distributed across different taxa, which confirms the importance for microalgae of these two pathways for thylakoid physiology. MGD genes have been found across almost all analyzed taxa and can be separated in two different groups, similarly to terrestrial plant MGD. SQD1 and SQD2 genes are widely distributed along the analyzed taxa in a similar way to MGD genes with some exceptions. For Pinguicophyceae, Raphidophyceae, and Synurophyceae, only sequences coding for MGDG were found. On the contrary, sequences assigned to Ciliophora and Eustigmatophyceae were exclusively corresponding to SQD1 and SQD2. This study reports, for the first time, the presence/absence of these enzymes in available microalgal transcriptomes, which gives new insights on microalgal physiology and possible biotechnological applications for the production of bioactive lipids.

Keywords: microalgae; monogalactosyldiacylglycerol synthase; UDP-sulfoquinovose synthase; sulfoquinovosyltransferase; monogalactosyldiacylglycerols; sulfoquinovosyldiacylglycerols; transcriptome analysis

1. Introduction

Microalgae are eukaryotic photosynthetic microorganisms that are adapted to live in ecologically different habitats, which results in a wide diversity of species and natural products [1,2] with pharmaceutical [3], nutraceutical [4], and cosmeceutical [5] interests. Microalgae can be cultivated in huge quantities and this advantage overcomes the bottleneck of drug discovery from marine macro-organisms and destructive collection practices. In addition, many studies have focused on optimizing the culturing conditions in order to obtain the metabolites of interest or produce them

in large amount. Regarding microalgal molecular resources, various microalgal genomes have been sequenced [6]. However, there are still few genomes available compared to the huge number of existing microalgae [6]. This is due to the fact that microalgae, especially dinoflagellates, can have very large genomes, up to 112 Gbp [7]. On the contrary, several transcriptomes are available for microalgae [6]. Many of these have been sequenced thanks to the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP), the TARA ocean, and the Global Ocean Sampling expeditions, while others are on the way. This will also help discover the enzymes involved in the synthesis of bioactive metabolites and suggest which are the most promising species for genetic engineering manipulations in order to increase the production of specific metabolites.

At the moment, there are only a few studies on metabolic pathways involved in the synthesis of bioactive compounds from microalgae [7–11]. The aim of this paper was to investigate the presence in microalgae of enzymes involved in the synthesis of specific functional lipids with huge interest for pharmaceutical, nutraceutical, and cosmeceutical applications. In particular, microalgae have been found to contain high contents of monogalactosyldiacylglycerols (MGDGs) [12]. MGDGs are compounds of biotechnological interest because they have been investigated for the potential treatment of human pathologies such as cancer (e.g., human pancreatic cancer cell lines) [13] and inflammation [14]. A purified MGDG, extracted from spinach, is a strong DNA polymerase inhibitor and it enhances chemotherapeutic efficiency in human pancreatic cancer cell lines (BxPC-3, MIAPaCa2, and PANC-1) [15]. Regarding marine examples, two MGDGs from the diatom *Phaeodactylum tricornerutum* showed *in vitro* pro-apoptotic activity on immortal mouse epithelial cell lines (W2 cells) [16].

Two different MGDGs from the green microalga (Chlorophyta) *Tetraselmis chuii* [17] and four MGDGs from *Nannochloropsis granulata* (Ochrophyta, Eustigmatophyceae) [18] showed anti-inflammatory properties. These MGDGs, extracted from both *T. chuii* and *N. granulata*, were able to reduce NO production and inducible nitric oxide synthase (iNOS) protein levels in lipopolysaccharide (LPS)-stimulated RAW264.7 macrophage cells. In addition, sulfoglycolipids, which constitute the anionic fraction of MGDGs [19], also present interesting biological properties, such as glutaminyl cyclase (QC) inhibitory activity [20] and immuno-stimulatory activity [21]. Sulfolipids extracted from the green microalgae *Tetrademus lagerheimii* (formerly *Scenedesmus acuminatus*), *Scenedesmus producto-capitatus*, *Pectinodesmus pectinatus* (formerly *Scenedesmus pectinatus*), and *Tetrademus wisconsinensis* are able to inhibit QC [20] (an enzyme involved in Alzheimer's disease progression [22]) and, thus, have potential as lead compounds against Alzheimer's disease. Furthermore, a synthetic sulfolipid derived from *Thalassiosira weissflogii* CCMP1336 (renamed *Conticribra weissflogii*, Bacillariophyta), named β -SQDG18, is a potent vaccine adjuvant [21]. β -SQDG18 is able to trigger a more effective immune response against cancer cells to improve dendritic cell (DC) maturation, to increase CD83-positive DC, and to stimulate the production of pro-inflammatory cytokines (IL-12 and INF- γ) [21].

The aim of the present work was to investigate the presence of genes involved in MGDG and sulfoglycolipid synthesis in microalgae (Bacillariophyta, Cercozoa, Chlorophyta, Chrysophyceae, Coccolithophyceae, Cryptophyta, Dictyochophyceae, Dinophyceae, Euglenophyceae, Eustigmatophyceae, Glaucophyceae, Pavlovophyceae, Pelagophyceae, Pinguicophyceae, Raphidophyceae, Rhodophyta, Synurophyceae, and Xanthophyceae) because little information is available. Our analyses were performed considering all the microalgal transcriptomes and metatranscriptomes available from the MMETSP project and the recent Tara Oceans and Global Ocean sampling expeditions, respectively.

Galactolipids represent up to 80% of the total lipids of the plastid membranes [23]. They can contain one or two galactose (Gal) molecules, which bond to the glycerol backbone at sn-3 position, and are called, respectively, MGDGs or digalactosyldiacylglycerols (DGDGs). The reaction required for MGDG biosynthesis is catalyzed by a MGDG synthase (MGD), which transfers the Gal moiety from UDP-Gal to diacylglycerol [24] (Figure 1a). MGDG is required in plastids biogenesis and integrity and in photosynthesis [24]. In terrestrial plants, the MGDG synthase is encoded by two types of genes, namely type-A (AtMGD1) and type-B (AtMGD2 and AtMGD3) isoforms, and enzymes are well characterized [25–30]. In microalgae, there is no information available yet.

Sulfolipids are constituent of the thylakoids in plant and algal chloroplasts [31]. Sulfoquinovose, which is the building block of sulfolipids, is also the major component of the biological sulfur cycle [32] and it is produced by photosynthetic organisms at a rate of 1010 tons per year [33]. The biosynthesis of these lipids proceeds in two reactions. The first reaction is catalysed by the UDP-sulfoquinovose synthase (SQD1), that assembles the UDP-sulfoquinovose from UDP-glucose and sulphite [34]. The second reaction is catalysed by sulfoquinovosyltransferase (SQD2), which transfers the sulfoquinovose to diacylglycerol [35]. This produces sulfoquinovosyldiacylglycerol (SQDG) (Figure 1b).

Different classes of microalgal-derived compounds have been identified and several have shown specific biological activities, such as anti-cancer [13], anti-inflammatory [14], anti-tuberculosis [36], anti-epilepsy [37], anti-microbial [38], immune-regulatory [39], anti-hypertensive [40], anti-atherosclerosis [40], and anti-osteoporosis [40] activity. The systematic studies of the genes involved in the synthesis of the bioactive compounds of interest are increasing [7,9,10] and can open new perspectives for gene-editing and boost the use of microalgae as a source of new marine natural products.

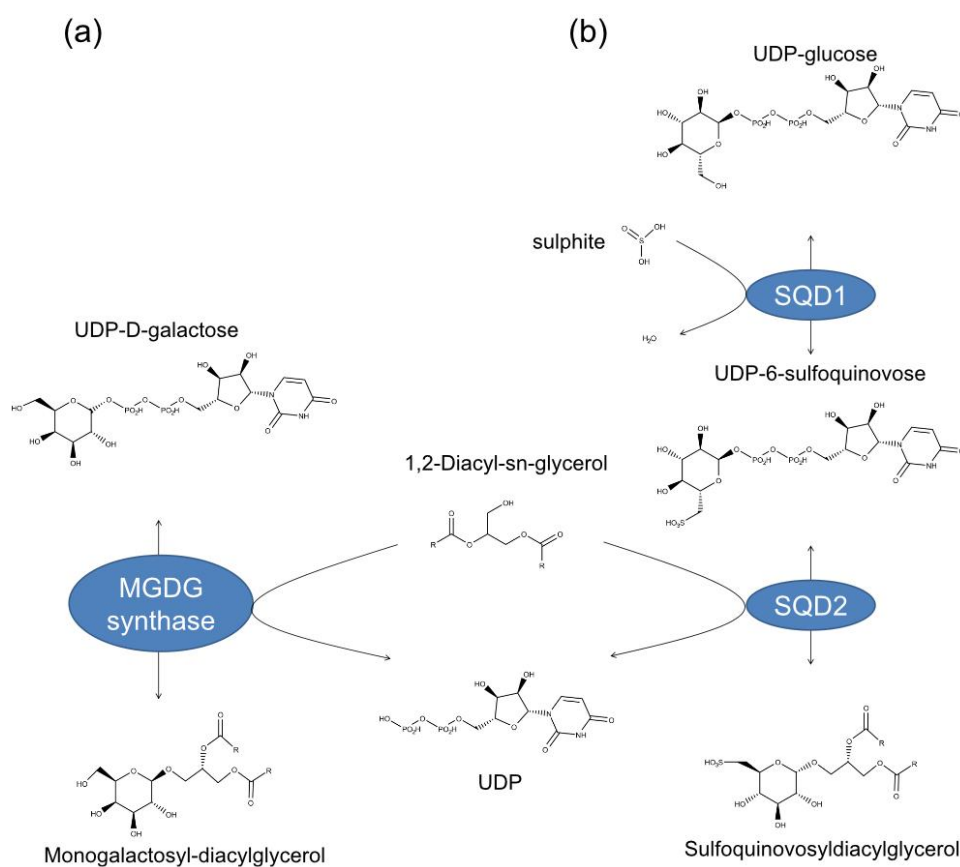


Figure 1. Enzymes responsible for (a) monogalactosyldiacylglycerol (i.e., MGDG synthase) and (b) sulfoquinovosyldiacylglycerol (i.e., UDP-sulfoquinovose synthase or SQD1 and sulfoquinovosyltransferase or SQD2) biosynthesis.

2. Results and Discussion

2.1. Identification and Taxonomic Assignment of MGD, SQD1, and SQD2 Homologous Sequences

The BLASTp search conducted in the OGA database returned 1,614 annotated eukaryotic sequences putatively attributable to MGD, 1,422 to SQD1 and 1,340 to SQD2. Of these sequences, the following were unambiguously identified as the genes of interest after Blast2GO analysis: 1154 to MGDG, 817 to SQD1, and 273 to SQD2 (Supplementary Data S1–S3, respectively).

The taxonomic pattern of these sequences, excluding the ones annotated as only as “Eukaryota” and “Stramenopiles” is illustrated in Figure 2. Of the 15 taxonomic divisions considered, 13 were found in the sequences assigned to MGD, 11 to SQD1, and 7 to SQD2 (Figure 2). In this database (OGA), we found homologs for all three genes investigated in the following microalgal taxa: Bacillariophyta, Chlorophyta, Dictyochophyceae, Coccolithophyceae, Pavlovophyceae, and Pelagophyceae. In particular, homolog sequences were particularly abundant in Bacillariophyta, Coccolithophyceae, and Pelagophyceae. For Pinguioophyceae, Raphidophyceae, and Synurophyceae, only sequences coding for MGDG were found. On the contrary, sequences assigned to Ciliophora and Eustigmatophyceae were exclusively corresponding to SQD1 and SQD2. Considering that the results are based on transcriptomic data, they can suggest the absence of a specific gene in certain taxa or simply the analyzed organisms were not expressing the specific gene at the time of sampling and fixing for the RNA-sequencing.

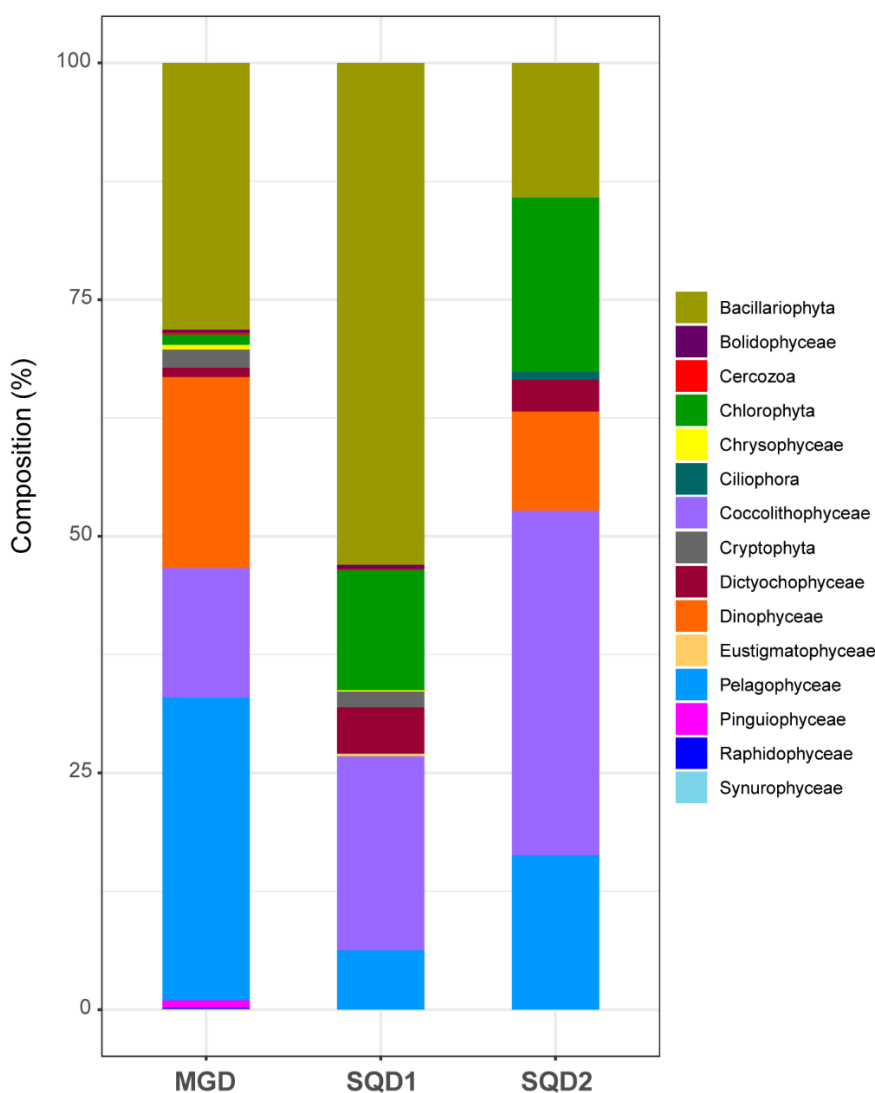


Figure 2. Taxonomic characterization and composition of the analyzed genes. MGD, SQD1, and SQD2 homologs were retrieved from Tara Oceans meta-transcriptomes. Graphical view of taxonomic composition for each gene is reported.

For MMETSP transcriptomes, we retrieved 313 sequences for MGD, 84 for SQD1, and 266 for SQD2 after performing a blastp search. After functional annotation (and removal of sequences considered not valid), 267 sequences were left for MGD, 80 for SQD1, and 121 for SQD2 (Supplementary Data S4–S6, respectively).

We reported in Table 1 the species and the strains in which MGD, SQD1, and SQD2 were found. The analyzed genes were found in 96 different species or strains across 14 different taxonomic categories. All the three genes were found in the reported taxa, with the exception of the two species of Chrysophyceae (*Dinobryon* sp. UTEXLB2267 and *Paraphysomonas imperforata* PA2, Supplementary Data S7). In addition, considering that the results are based on transcriptomic data, they can be interpreted as either the absence of that specific gene or lack of its expression in that particular condition.

Table 1. Species and strains from MMETSP transcriptomes in which MGD, SQD1, and SQD2 homologs were found. Colors refer to taxonomic ranks as in Figure 2. P = presence of validated gene homologs.

Color Legend	Taxonomic Group	Species/Strain Surveyed	Accepted Synonym	MGD	SQD1	SQD2
	Dinophyceae	<i>Alexandrium monilatum</i> CCMP3105		P	P	P
	Dinophyceae	<i>Alexandrium tamarense</i> CCMP1771		P	P	
	Dinophyceae	<i>Amphidinium carterae</i> CCMP1314		P		
	Bacillariophyta	<i>Amphiprora</i> sp.		P	P	P
	Bacillariophyta	<i>Amphora coffeaeformis</i> CCMP127	<i>Halamphora coffeiformis</i>	P	P	P
	Bacillariophyta	<i>Asterionellopsis glacialis</i> CCMP134		P	P	P
	Pelagophyceae	<i>Aureococcus anophagefferens</i> CCMP1850		P		P
	Pelagophyceae	<i>Aureoumbra lagunensis</i> CCMP1510		P	P	P
	Dinophyceae	<i>Azadinium spinosum</i> 3D9		P		
	Bacillariophyta	<i>Ceratium fusus</i> PA161109	<i>Tripos fusus</i>	P		
	Bacillariophyta	<i>Chaetoceros affinis</i> CCMP159		P	P	P
	Bacillariophyta	<i>Chaetoceros curvisetus</i>		P	P	P
	Bacillariophyta	<i>Chaetoceros debilis</i> MM31A 1		P	P	P
	Bacillariophyta	<i>Chaetoceros neogracile</i> CCMP1317		P	P	P
	Raphidophyceae	<i>Chattonella subsalsa</i> CCMP2191		P	P	P
	Coccolithophyceae	<i>Chrysochromulina polylepis</i> CCMP1757	<i>Prymnesium polylepis</i>	P	P	P
	Bacillariophyta	<i>Corethron pennatum</i> L29A3		P	P	P
	Dinophyceae	<i>Cryptocodinium cohnii</i> Seligo		P		
	Bacillariophyta	<i>Ditylum brightwellii</i> GSO103		P	P	P
	Bacillariophyta	<i>Ditylum brightwellii</i> GSO104		P	P	P
	Bacillariophyta	<i>Ditylum brightwellii</i> GSO105		P	P	P
	Chlorophyta	<i>Dunaliella tertiolecta</i> CCMP1320		P	P	P
	Dinophyceae	<i>Durinskia baltica</i> CSIRO CS 38	<i>Durinskia dybowskii</i>	P	P	P
	Coccolithophyceae	<i>Emiliana huxleyi</i> 374		P		P
	Coccolithophyceae	<i>Emiliana huxleyi</i> 379		P	P	P
	Coccolithophyceae	<i>Emiliana huxleyi</i> CCMP370		P	P	P
	Coccolithophyceae	<i>Emiliana huxleyi</i> PLYM219		P	P	P
	Euglenophyceae	<i>Eutreptiella gymnastica</i> like CCMP1594		P		P
	Bacillariophyta	<i>Extubocellulus spinifer</i> CCMP396		P	P	
	Bacillariophyta	<i>Fragilariopsis kerguelensis</i> L26 C5		P	P	P
	Bacillariophyta	<i>Fragilariopsis kerguelensis</i> L2 C3		P	P	P
	Coccolithophyceae	<i>Gephyrocapsa oceanica</i> RCC1303		P	P	P
	Dinophyceae	<i>Glenodinium foliaceum</i> CCAP1116 3	<i>Kryptoperidinium foliaceum</i>	P	P	P
	Glaucoephyceae	<i>Gloeochaete wittrockiana</i> SAG46 84		P	P	P
	Cryptophyta	<i>Goniomonas pacifica</i> CCMP1869			P	
	Raphidophyceae	<i>Heterosigma akashiwo</i> CCMP2393		P	P	P
	Raphidophyceae	<i>Heterosigma akashiwo</i> CCMP3107		P	P	P
	Raphidophyceae	<i>Heterosigma akashiwo</i> CCMP452		P	P	P
	Raphidophyceae	<i>Heterosigma akashiwo</i> NB		P	P	P
	Coccolithophyceae	<i>Isochrysis galbana</i> CCMP1323		P	P	P
	Coccolithophyceae	<i>Isochrysis</i> sp. CCMP1244		P	P	P
	Coccolithophyceae	<i>Isochrysis</i> sp. CCMP1324		P	P	P
	Dinophyceae	<i>Karenia brevis</i> CCMP2229		P		

Table 1. Cont.

Color Legend	Taxonomic Group	Species/Strain Surveyed	Accepted Synonym	MGD	SQD1	SQD2
	Dinophyceae	<i>Karenia brevis</i> SP1		P		
	Dinophyceae	<i>Karenia brevis</i> SP3		P		
	Dinophyceae	<i>Karenia brevis</i> Wilson		P		P
	Dinophyceae	<i>Karlodinium micrum</i> CCMP2283	<i>Karlodinium veneficum</i>	P		P
	Dinophyceae	<i>Kryptoperidinium foliaceum</i> CCMP1326		P	P	P
	Dinophyceae	<i>Lingulodinium polyedra</i> CCMP1738		P	P	
	Cercozoa	<i>Lotharella globosa</i> CCCM811		P	P	P
	Chlorophyta	<i>Micromonas</i> sp. CCMP2099		P		
	Chlorophyta	<i>Micromonas</i> sp. NEPCC29		P	P	
	Chlorophyta	<i>Micromonas</i> sp. RCC472		P	P	P
	Bacillariophyta	<i>Nitzschia punctata</i> CCMP561	<i>Tryblionella punctata</i>	P	P	P
	Chrysophyceae	<i>Ochromonas</i> sp. CCMP1393		P	P	P
	Dinophyceae	<i>Oxyrrhis marina</i>		P		P
	Dinophyceae	<i>Oxyrrhis marina</i> LB1974				P
	Pavlovophyceae	<i>Pavlova</i> sp. CCMP459		P	P	P
	Pelagophyceae	<i>Pelagococcus subviridis</i> CCMP1429		P		
	Pelagophyceae	<i>Pelagomonas calceolata</i> CCMP1756		P	P	P
	Dinophyceae	<i>Peridinium aciculiferum</i> PAER 2	<i>Apocalathium aciculiferum</i>	P		
	Chlorophyta	<i>Picocystis salinarum</i> CCMP1897			P	
	Coccolithophyceae	<i>Pleurochrysis carterae</i> CCMP645	<i>Chrysotila carterae</i>	P		P
	Bacillariophyta	<i>Proboscia alata</i> PI D3		P	P	P
	Dinophyceae	<i>Prorocentrum minimum</i> CCMP1329	<i>Prorocentrum cordatum</i>	P	P	
	Dinophyceae	<i>Prorocentrum minimum</i> CCMP2233	<i>Prorocentrum cordatum</i>	P	P	
	Coccolithophyceae	<i>Prymnesium parvum</i> Texoma1		P	P	P
	Bacillariophyta	<i>Pseudo-nitzschia australis</i> 10249 10 AB		P	P	P
	Bacillariophyta	<i>Pseudo-nitzschia fradulenta</i> WWA7		P	P	P
	Dictyochophyceae	<i>Pseudopedinella elastica</i> CCMP716		P	P	P
	Dictyochophyceae	<i>Pteridomonas danica</i> PT		P	P	P
	Chlorophyta	<i>Pyramimonas parkeae</i> CCMP726		P	P	P
	Rhodophyta	<i>Rhodella maculata</i> CCMP736	<i>Rhodella violacea</i>	P	P	
	Cryptophyta	<i>Rhodomonas</i> sp. CCMP768		P	P	P
	Dinophyceae	<i>Scrippsiella hangoei</i> SHTV5	<i>Apocalathium malmogiense</i>	P		
	Dinophyceae	<i>Scrippsiella hangoei</i> like SHHI 4		P		
	Dinophyceae	<i>Scrippsiella trochoidea</i> CCMP3099		P		
	Bacillariophyta	<i>Skeletonema dohrnii</i> SkelB		P	P	P
	Bacillariophyta	<i>Skeletonema marinoi</i> SkelA		P	P	P
	Bacillariophyta	<i>Skeletonema menzelii</i> CCMP793		P	P	P
	Dinophyceae	<i>Symbiodinium</i> sp. C1		P		
	Dinophyceae	<i>Symbiodinium</i> sp. CCMP2430		P		
	Dinophyceae	<i>Symbiodinium</i> sp. Mp		P		
	Chlorophyta	<i>Tetraselmis striata</i> LANL1001		P	P	P
	Bacillariophyta	<i>Thalassionema nitzschioides</i> L26 B		P		P
	Bacillariophyta	<i>Thalassiosira antarctica</i> CCMP982		P	P	P
	Bacillariophyta	<i>Thalassiosira gravida</i> Gmp14c1		P	P	P
	Bacillariophyta	<i>Thalassiosira miniscula</i> CCMP1093		P	P	P
	Bacillariophyta	<i>Thalassiosira oceanica</i> CCMP1005		P	P	P
	Bacillariophyta	<i>Thalassiosira rotula</i> CCMP3096	<i>Thalassiosira gravida</i>	P	P	P
	Bacillariophyta	<i>Thalassiosira rotula</i> GSO102	<i>Thalassiosira gravida</i>	P	P	P
	Bacillariophyta	<i>Thalassiosira weissflogii</i> CCMP1010	<i>Conticribra weissflogii</i>	P	P	
	Bacillariophyta	<i>Thalassiosira weissflogii</i> CCMP1336	<i>Conticribra weissflogii</i>	P	P	P
	Bacillariophyta	<i>Thalassiothrix antarctica</i> L6 D1		P	P	
	Xanthophyceae	<i>Vaucheria litorea</i> CCMP2940		P	P	P

2.2. Final Dataset and Phylogenetic Inferences

The final alignments consisted of 649 sequences for MGD, 521 for SQD1, and 244 for SQD2 (Supplementary Data S8–S10, respectively), from both the OGA and MMETSP database. The length of each dataset after trimming of poorly aligned regions was as follows: 377 bp for MGD, 198 bp for SQD1, and 295 bp for SQD2.

The MGD phylogenetic tree showed that most of the taxa here investigated contained paralog copies (homolog copies resulting from duplication events) of monogalactosyldiacylglycerol synthase gene (Figure 3). These paralogs generally occurred in two copies, resulting in two distinct and highly supported clades for most taxa (red circles, Figure 3, Supplementary File S1). A few taxa (e.g., Coccolithophyceae, Cryptophyta, and Xantophyceae) presented only one MGD copy (and, therefore, a single group of sequences), while others (e.g., Dinophyceae) presented several MGD paralogs (Figure 3 or Supplementary File S1).

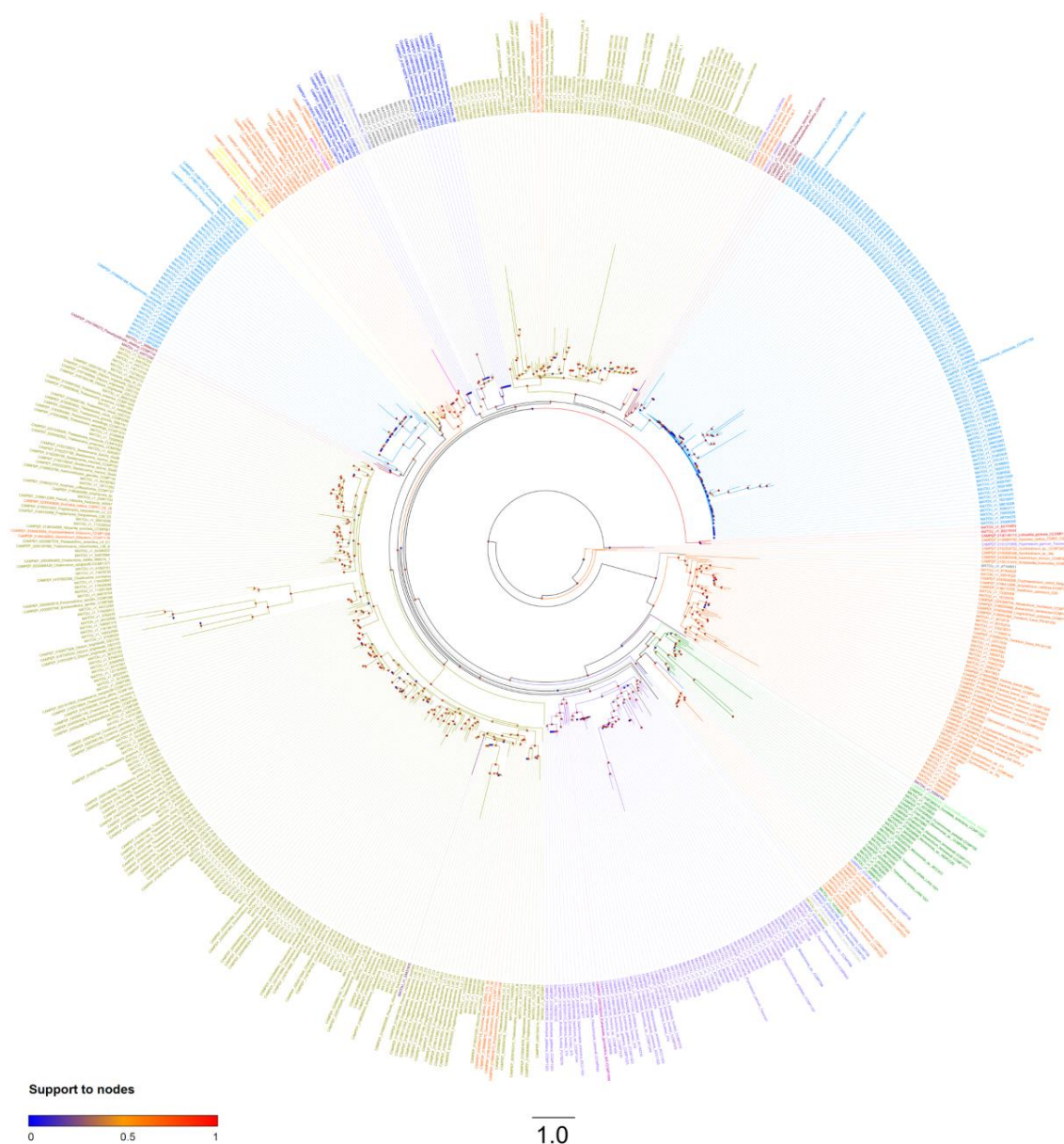


Figure 3. MGD unrooted phylogenetic tree. Colored circles at the base of each node refer to branch support after aLRT SH-like test. Colors of taxa refer to taxonomic groups as in Figure 2 and Table 1.

For the SQD1 gene, the phylogenetic tree showed that all the sequences belonging to the same taxon formed a highly supported monophyletic group in most species (Figure 4, Supplementary File S2). Among the exceptions, there are the diatoms (Bacillariophyta) where, beside a clade containing the most diatom sequences retrieved from OGA and MMETSP, there are a few others interspersed across the tree (Figure 4). One of these only contains sequences of *Thalassiosira rotula* (synonym of *T. gravida*) strain CCMP1093 from MMETSP. The others contain small groups of sequences from OGA (Figure 4, Supplementary File S2). Similarly, SQD1 dinoflagellates (Dinophyceae) homologs do not form a monophyletic group in the tree. This could be due to the high complexity of the plastid evolution in Dinophyceae. Dinophyceae acquired plastid with four different endosymbiotic events. [41]. In a secondary endosymbiotic event, Dinophyceae acquired plastid by serial endosymbiosis with green microalgae (Chlorophyta) [42]. In the tertiary endosymbiotic event, Dinophyceae acquired tertiary plastids by endosymbiosis with Cryptophyceae-Cryptomonads, Haptophyta, and Bacillariophyta [41].

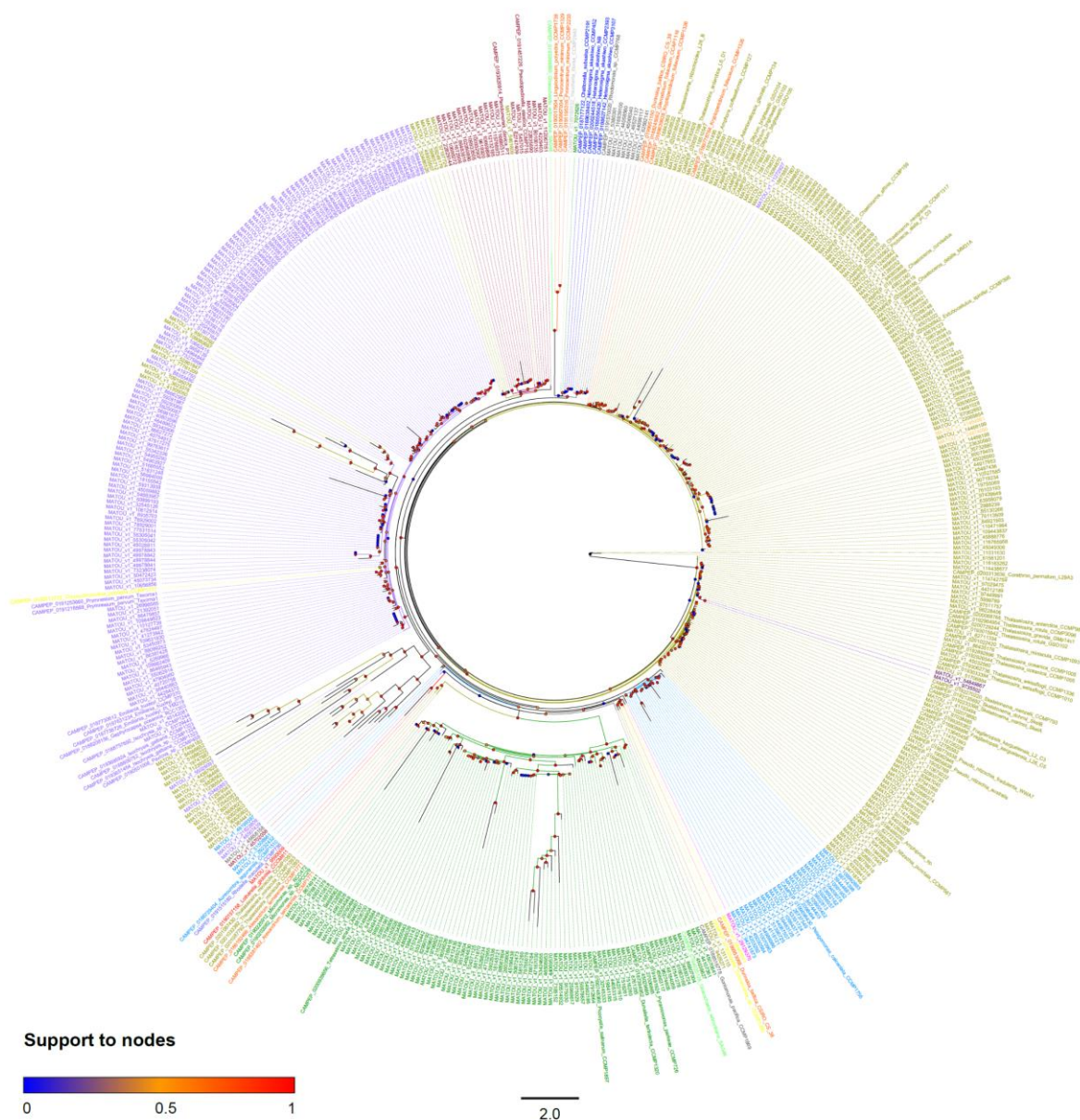


Figure 4. SQD1 unrooted the phylogenetic tree. Colored circles at the base of each node refer to branch support after an aLRT SH-like test. Colors of taxa refer to taxonomic groups in Figure 2 and Table 1.

For the SQD2 gene, all the sequences from the same taxonomic group formed monophyletic groups except for the Dinophyceae (Figure 5, Supplementary File S3). All the dinoflagellate sequences from OGA and MMETSP were interspersed across the tree with high support, close to Bacillariophyta, Chlorophyta, Dictyocophyceae, and Coccolithophyceae.

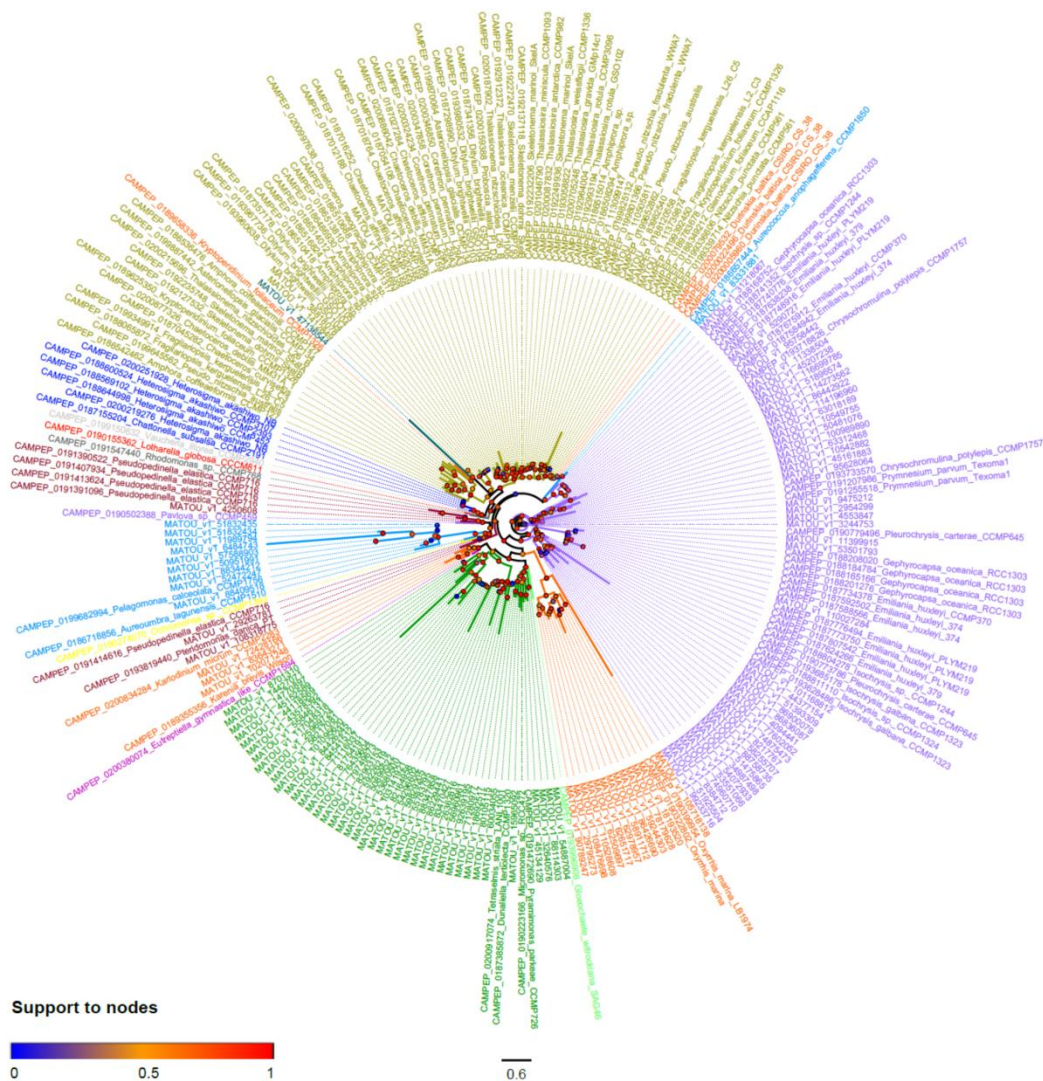


Figure 5. SQD2 unrooted phylogenetic tree. Colored circles at the base of each node refer to branch support after the aLRT SH-like test. Colors of taxa refer to taxonomic groups in Figure 2 and Table 1.

2.3. Structural Details of MDGs, SQD1, and SQD2 from *Thalassiosira weissflogii* CCMP1336 (*Conticribrilla weissflogii*)

We built *in silico* models for MGDs, SQD1, and SQD2 proteins from the diatom *T. weissflogii* CCMP1336 (synonym of *Conticribrilla weissflogii*). *T. weissflogii* (*C. weissflogii*) CCMP1336 has been selected because it is known to have immunostimulatory activity, and both galactolipids and sulfolipids have been found to play a key role in this bioactivity [39,43]. 3D *in silico* models (Figure 6) were generated using the amino acid sequences of MGD (CAMPEP_0193073380, CAMPEP_0193064960 and CAMPEP_0193062160) (Figure 6a–c), SQD1 (CAMPEP_0193062736) (Figure 6d), and SQD2 (CAMPEP_0193058822) (Figure 6e), obtained from the MMETSP database. Phyre2 results were summarized in Table 2. The analyses pointed out the homology between *T. weissflogii* (*C. weissflogii*) and *Arabidopsis thaliana* proteins. Structural prediction of MGDs (CAMPEP_0193073380, Supplementary Data S11, CAMPEP_0193064960 Supplementary Data S12, and CAMPEP_0193062160, Supplementary

Data S13) found high similarity with the 3D structure of MGD1 from *A. thaliana*. The percentage of identity between *T. weissflogii* (*C. weissflogii*) and *A. thaliana* MGD1 is higher than 40%. Similar results were obtained for structural prediction of SQD1 from *T. weissflogii* (*C. weissflogii*) (Supplementary Data S14) with 45% identity with the 3D structure of SQD1 from *A. thaliana*.

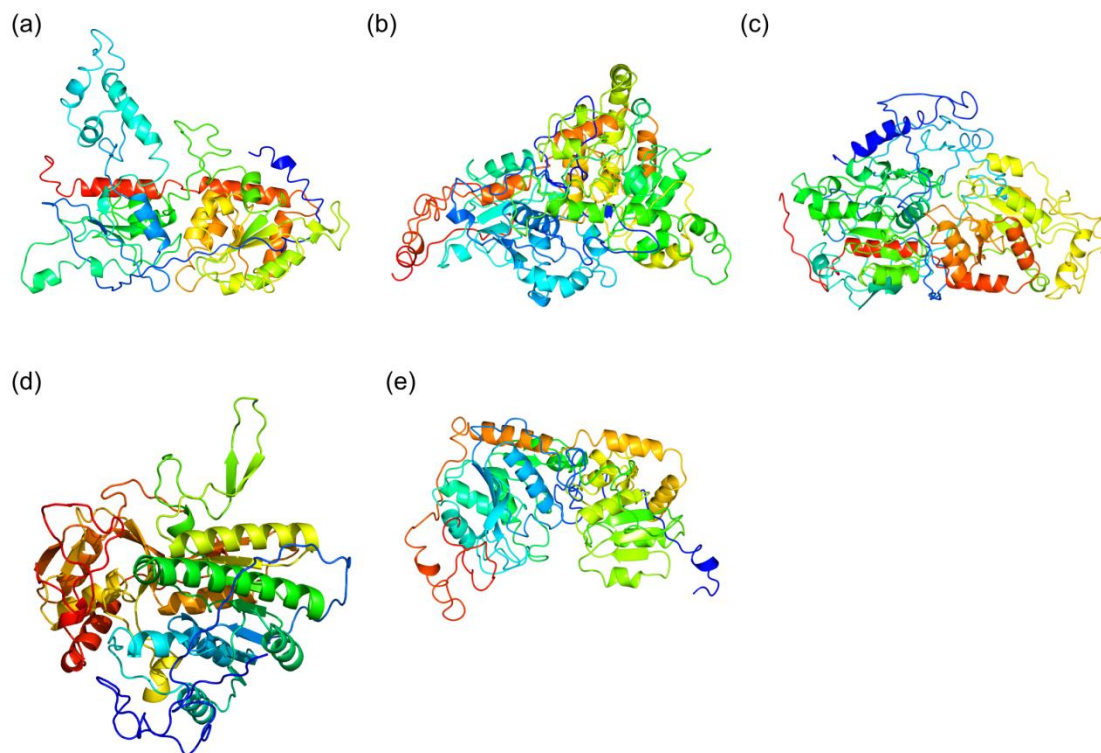


Figure 6. *In silico* model generated by Phyre2 for: (a) MGD CAMPEP_0193073380, (b) MGD CAMPEP_0193064960, (c) MGD CAMPEP_0193062160, (d) SQD1 CAMPEP_0193062736, and (e) SQD2 CAMPEP_0193058822.

Table 2. Report of Phyre2 analysis. We report the template (protein of known structure used for the prediction analysis) and its protein data bank (PDB) code, confidence (probability that the sequence and template are homologous), and percent id (percent of identity).

<i>Thalassiosira</i> (<i>Conticribra</i>) <i>weissflogii</i> CCMP1336 Proteins	Template (PDB Code)	Confidence	% id
MGD CAMPEP 0193073380	MGD1 from <i>A. thaliana</i> (4X1T)	100	46
MGD CAMPEP 0193064960	MGD1 from <i>A. thaliana</i> (4X1T)	100	47
MGD CAMPEP 0193062160	MGD1 from <i>A. thaliana</i> (4X1T)	100	40
SQD1 CAMPEP_0193062736	SQD1 from <i>A. thaliana</i> (1I24)	100	45
SQD2 CAMPEP_0193058822	SUS1 from <i>A. thaliana</i> (3S29)	100	18

Structural prediction for SQD2 (Supplementary Data S15) found a similarity with the 3D structure of sucrose synthase-1 (SUS1) from *A. thaliana*. The percentage of identity between SQD2 and SUS1 is 18%, which is lower than the percent of identity of SQD1 and MDGs. SUS1 from *A. thaliana* catalyzed a reversible sucrose synthesis. It transfers glucose moiety from UDP to fructose and it has been found to form a complex with both UDP–glucose and UDP–fructose [44]. These findings may indicate the presence of a low specificity ligand-binding site in the SQD2 enzyme.

In order to evaluate whether predicted structures preserved the functional active sites of MGDs, SQD1, and SQD2, they were also analyzed by using the COACH server. MGD analysis was performed for all the structures obtained from Phyre2. The analysis pointed out the presence of a UDP binding

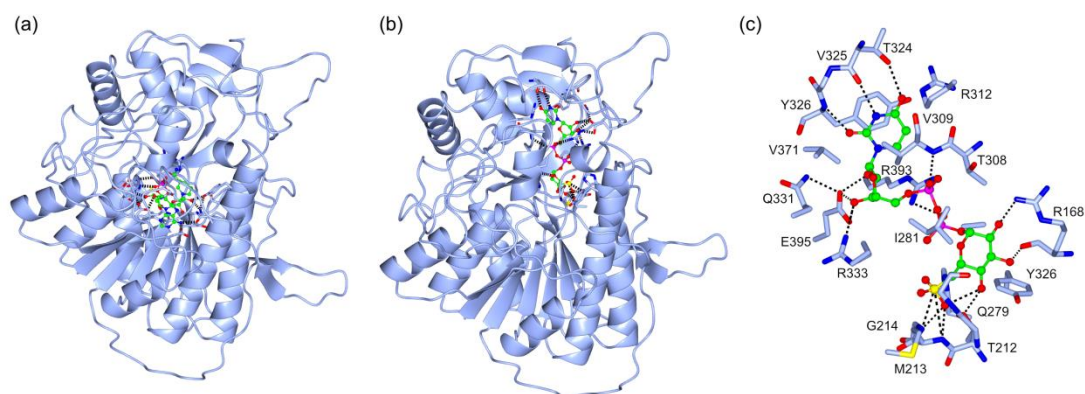


Figure 8. Protein-ligand binding site prediction by COACH server of SQD1 from *Thalassiosira (Conticribra) weissflogii* CCMP1336. (a) Prediction of binding site of the complex SQD1–NAD. (b) Prediction of binding site of the complex SQD1–USQ. (c) Structure of USQ binding-pocket and the specific interaction between USQ and highlighted amino acid residues of SQD1.

SQD2 analysis indicated the presence of a UDP-binding pocket (Figure 9a) (C-score 0.28), which involved the residues S82, G83, N86, F256, V282, G283, R284, K289, E309, Q332, L333, L338, E355, G358, F359, V360, and E363. Moreover, COACH analyses pointed out the presence of a N-Acetylglucosamine (NAG)-binding pocket (Figure 9b) (C-score 0.33) that involved the residues G83, Y84, R87, H190, T191, K249, S354, E355, T356, L357, and G358. These data confirmed the presence of a low specificity ligand binding pocket, and suggest that SQD2 from diatoms could be involved in the synthesis of other diacylglycerols such as glucuronosyldiacylglycerol (GlcADG) in *Arabidopsis* [47] or flavonoid glycosylation in *Oryza sativa* [48,49].

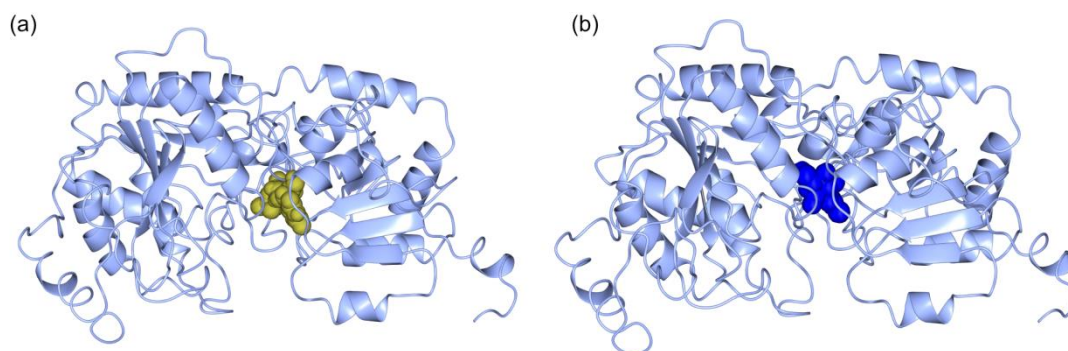


Figure 9. Protein-ligand binding site prediction by COACH server of SQD2 from *Thalassiosira (Conticribra) weissflogii* CCMP1336. (a) Prediction of binding site of the complex SQD2–UDP. (b) prediction of binding site of the complex SQD2–NAG.

3. Materials and Methods

3.1. Identification of MGD, SQD1, and SQD2 Homologous Sequences

Since our primary interest was to ascertain the occurrence of MGD, SQD1, and SQD2 in microalgae, we used the sequences of MGD (accession number XP002181685), SQD1 (accession number XP002185968), and SQD2 (accession number XP002185276) from the diatom *Phaeodactylum tricornutum* as queries for a BLAST [50] search against the Ocean Global Atlas (OGA, [51]) database (<http://tara-oceans.mio.osupytheas.fr/oceangene-atlas/>), and several protist transcriptomes from the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETS, [52]) available at <https://zenodo.org/record/12125852585>. The OGA database contains a collection of more than 116 million eukaryote and 40 million prokaryotic genes gathered during the *Tara Oceans* [53,54]

and the Global Ocean Sampling [55] expeditions. Instead, MMETSP contains the transcriptomes of some of the most abundant and ecologically significant microbial eukaryotes in the oceans. To retrieve homologs, we used the blastp algorithm against the metagenome/metatranscriptome and transcriptome databases contained in OGA and the MMETSP, setting the expect threshold to 1E-10. For OGA, after ascertaining that the sequences obtained from metagenomes and metatranscriptomes were identical, we only used one dataset (metatranscriptomes) for further analyses. Since a sequence-based homology search could have recovered different genes with similar functions, we used Blast2GO [56] to obtain a functional annotation of the homologs retrieved. We used the default settings (i.e., blastx program, using the nr BLAST database and with a BLAST expectation value of 1.0E-3) for the analysis. We considered valid Blast2GO annotations containing the following names: 1,2-diacylglycerol 3-beta-galactosyltransferase and monogalactosyldiacylglycerol synthase for MGD, sulfolipid biosynthesis protein, sulfoquinovosyldiacylglycerol synthesis protein, UDP sulfoquinovose synthase and uridine 5'-diphosphate-sulfoquinovose synthase for SQD1, sulfoquinovosyl transferase SQD2, sulfoquinovosyldiacylglycerol 2, and UDP-sulfoquinovose: DAG sulfoquinovosyltransferase for SQD2. All the sequences identified as only “predicted or hypothetical protein” without specification were *a priori* discarded.

3.2. Taxonomic Overview

All the sequences from OGA that passed the Blast2GO analysis with the criteria specified above were annotated using the annotation file generated during homolog retrieval. We removed all the sequences without a taxonomic annotation, annotated only as “Eukaryota” and, whenever applicable, with a generic annotation that did not allow to discriminate such sequences from others of lower taxonomic rank (e.g., “Stramenopiles”). All the taxa in which the MGD, SQD1, and SQD2 genes were found and were organized into a table. The abundance of such sequences was plotted as histograms using the R [57] working packages *scales* [58] and *ggplot2* [59]. For the homologs retrieved from the MMETSP transcriptomes, we generated a table illustrating the species surveyed and the genes of interest that were found in each of them.

3.3. Sequence Alignment and Phylogenetic Inference

The sequences of each gene from OGA and MMETSP that passed Blast2GO annotation and with a length of at least 200 aa were aligned using COBALT [60] available at <https://www.ncbi.nlm.nih.gov/tools/cobalt/>). Unlike other common software used for protein alignments (e.g., ClustalW, MAFFT, MUSCLE, ProbCons) that only use sequence information, COBALT also integrates the information of protein-motif regular expressions (PROSITE database) and of conserved protein domains (NCBI CDD database). In doing so, COBALT has a better chance of producing a biologically meaningful multiple alignment compared to tools that do not utilize this information [60]. Poorly aligned regions from each alignment were removed with trimAl v1.2 [61] in order to increase the quality of subsequent phylogenetic analyses. We used the *automated1* option to find the most appropriate mode to trim the alignments (use of gaps or similarity scores) depending on the alignment characteristics.

In order to infer reliable phylogenetic trees, all the sequences that, after trimming, were shorter than one quarter (around 60–70 aa) of the final length of the alignment were removed, unless longer sequences were not available for that particular taxon. Maximum likelihood phylogenetic trees were inferred using PhyML [62] using the LG substitution model [63], which turned out to be the best evolution model for the three genes investigated according to the Akaike Information Criterion implemented in SMS [64]. Support to nodes was calculated using the Shimodaira-Hasegawa-like (aLRT SH-like) procedure [65]. Trees were visualised and graphically edited in FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

3.4. In Silico Protein Model

The three-dimensional (3D) *in silico* models of MGD, SQD1, and SQD2 proteins from *Thalassiosira weissflogii* CCMP1336 (currently regarded as synonym of *Conticribra weissflogii*, Bacillariophyta) were

generated using Phyre2 server (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) [66]. Protein ligand-binding site predictions were performed using COACH analyses from the Zhang Lab server. COACH generated ligand-binding site predictions and a confidence score (C-score) of the prediction. The C-score ranged from 0 to 1. A higher score indicates a reliable prediction (<https://zhanglab.ccmb.med.umich.edu/COACH/>) [67,68]. Pictures were obtained using CCP4MG, version 2.10.11 obtained by <http://www.ccp4.ac.uk/MG/download/> [69].

4. Conclusions

This study analyzes the presence of MGD, SQD1, and SQD2 in microalgae and gives a broad overview of their presence among different microalgal classes. It has been shown that, within each taxa, some species do not express all the sequences, which suggests the absence of a specific gene. Despite the fact that several species have been found to express MGDs, SQD1, and SQD2 (Table 1), only a relatively small number of active microalgae have been studied and reported to possess anti-inflammatory/immunomodulatory activities, including the diatoms *Attheya longicornis*, *Cylindrotheca closterium*, *Trieres mobiliensis* (formerly *Odontella mobiliensis*), *Phaeodactylum tricorutum*, *Porosira glacialis*, *Pseudo-nitzschia pseudodelicatissima*, and *Thalassiosira (Conticribra) weissflogii*, and the flagellates *Amphidinium carterae* (Dinophyceae), *Edaphochlamys debaryana* (formerly *Chlamydomonas debaryana*), *Chloroidium saccharophilum* (formerly *Chlorella ovalis*), *Dunaliella salina* (formerly *Dunaliella bardawil*) (Chlorophyceae), *Nannochloropsis granulata*, *Nannochloropsis oculata* (Eustigmatophyceae), *Diacronema lutheri* (formerly *Pavlova lutheri*, Pavlovophyceae), *Tetraselmis chuii*, and *Tetraselmis suecica* (Chlorophyta, Chlorodendrophyceae) [39]. Our study can direct and accelerate the discovery of new bioactive species and give new insights on enzyme discovery from microalgae with biotechnological applications.

3D *in silico* prediction analyses indicated that the enzymes, maintaining conserved domains, could be effectively involved in the synthesis of compounds with known anticancer and immune-modulatory activities, such as MGDs and SQDs. This approach can give preliminary information for the selection of specific microalgal species for drug discovery and for genetic engineering approaches in order to produce huge amounts of bioactive compounds of pharmaceutical interest.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1660-3397/18/5/237/s1>. **Data S1:** Blast2GO analysis of MGD genes. BLASTp search conducted in the OGA database highlighted 1614 annotated eukaryotic sequences putatively attributable to MGD. Of these sequences, 1154 were identified as the gene of interest. **Data S2:** Blast2GO analysis of SQD1 genes. BLASTp search conducted in the OGA database highlighted 1422 annotated eukaryotic sequences putatively attributable to SQD1. Of these sequences, 817 were identified as the gene as the gene of interest. **Data S3:** Blast2GO analysis of SQD2 genes. BLASTp search conducted in the OGA database highlighted 1340 annotated eukaryotic sequences putatively attributable to SQD2. Of these sequences, 273 were identified as the gene of interest. **Data S4:** MGD amino acid sequences from MMETSP transcriptomes. After functional annotation, 267 sequences were left for MGD. **Data S5:** SQD1 amino acid sequences from MMETSP transcriptomes. After functional annotation, 84 sequences were left for MGD. **Data S6:** SQD2 amino acid sequences from MMETSP transcriptomes. After functional annotation, 121 sequences were left for SQD2. **Data S7:** Original data for Table 1. All the species and the strains were analyzed from MMETSP. **Data S8:** Alignments for MGD amino acid sequences. The alignments were performed using 649 amino acid sequences for MGD. **Data S9:** Alignments for SQD1 amino acid sequences. The alignments were performed using 521 amino acid sequences for SQD1. **Data S10:** Alignments for SQD2 amino acid sequences. The alignments were performed using 295 amino acid sequences for SQD2. **Data S11:** PDB file of MGD CAMPEP_0193073380 predicted structure. *In silico* model generated by Phyre2 for MGD CAMPEP_0193073380. **Data S12:** PDB file of MGD CAMPEP_0193064960 predicted structure. *In silico* model generated by Phyre2 for MGD CAMPEP_0193064960. **Data S13:** PDB file of MGD CAMPEP_0193062160 predicted structure. *In silico* model generated by Phyre2 for MGD CAMPEP_0193062160. **Data S14:** PDB file of SQD1 CAMPEP_0193062736 predicted structure. *In silico* model generated by Phyre2 for SQD1 CAMPEP_0193062736. Additional file 15: Data S15: PDB File of SQD2 CAMPEP_0193058822 predicted structure. *In silico* model generated by Phyre2 for SQD2 CAMPEP_0193058822. **File S1:** PDF file of MGD unrooted phylogenetic tree. Colored circles at the base of each node refer to branch support after aLRT SH-like test. Colors of taxa refer to taxonomic groups as in Table 1. **File S2:** PDF file of SQD1 unrooted phylogenetic tree. Colored circles at the base of each node refer to branch support after aLRT SH-like test. Colors of taxa refer to taxonomic groups in Table 1. **File S3:** PDF file of SQD2 unrooted phylogenetic tree. Colored circles at the base of each node refer to branch support after the aLRT SH-like test. Colors of taxa refer to taxonomic groups as in Table 1.

Author Contributions: Conceptualization, G.R. and C.L. Methodology and formal analysis, G.R., D.D.L., and C.L. Supervision, C.L., G.R., D.D.L., and C.L. co-wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: The “Antitumor Drugs and Vaccines from the Sea (ADVISE)” project (PG/2018/0494374) funded this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Plaza, M.; Herrero, M.; Cifuentes, A.; Ibanez, E. Innovative natural functional ingredients from microalgae. *J. Agric. Food Chem.* **2009**, *57*, 7159–7170. [[CrossRef](#)] [[PubMed](#)]
2. Jaspars, M.; De Pascale, D.; Andersen, J.H.; Reyes, F.; Crawford, A.D.; Ianora, A. The marine biodiscovery pipeline and ocean medicines of tomorrow. *J. Mar. Biol. Assoc. UK* **2016**, *96*, 151–158. [[CrossRef](#)]
3. Mimouni, V.; Ulmann, L.; Pasquet, V.; Mathieu, M.; Picot, L.; Bougaran, G.; Cadoret, J.P.; Morant-Manceau, A.; Schoefs, B. The potential of microalgae for the production of bioactive molecules of pharmaceutical interest. *Curr. Pharm. Biotechnol.* **2012**, *13*, 2733–2750. [[CrossRef](#)] [[PubMed](#)]
4. Koyande, A.K.; Chew, K.W.; Rambabu, K.; Tao, Y.; Chu, D.-T.; Show, P.-L. Microalgae: A potential alternative to health supplementation for humans. *Food Sci. Hum. Wellness* **2019**, *8*, 16–24. [[CrossRef](#)]
5. Mourelle, M.L.; Gómez, C.P.; Legido, J.L. The potential use of marine microalgae and cyanobacteria in cosmetics and thalassotherapy. *Cosmetics* **2017**, *4*, 46. [[CrossRef](#)]
6. Lauritano, C.; Ferrante, M.I.; Rogato, A. Marine natural products from microalgae: An -omics overview. *Mar. Drugs* **2019**, *17*, 269. [[CrossRef](#)]
7. Lauritano, C.; De Luca, D.; Ferrarini, A.; Avanzato, C.; Minio, A.; Esposito, F.; Ianora, A. De novo transcriptome of the cosmopolitan dinoflagellate *Amphidinium carterae* to identify enzymes with biotechnological potential. *Sci. Rep.* **2017**, *7*, 11701. [[CrossRef](#)]
8. Di Dato, V.; Di Costanzo, F.; Barbarinaldi, R.; Perna, A.; Ianora, A.; Romano, G. Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Sci. Rep.* **2019**, *9*, 9893. [[CrossRef](#)]
9. Lauritano, C.; De Luca, D.; Amoroso, M.; Benfatto, S.; Maestri, S.; Racioppi, C.; Esposito, F.; Ianora, A. New molecular insights on the response of the green alga *Tetraselmis suecica* to nitrogen starvation. *Sci. Rep.* **2019**, *9*, 3336. [[CrossRef](#)]
10. Elagoz, A.M.; Ambrosino, L.; Lauritano, C. De novo transcriptome of the diatom *Cylindrotheca closterium* identifies genes involved in the metabolism of anti-inflammatory compounds. *Sci. Rep.* **2020**, *10*, 4138. [[CrossRef](#)]
11. Vingiani, G.M.; De Luca, P.; Ianora, A.; Dobson, A.D.W.; Lauritano, C. Microalgal enzymes with biotechnological applications. *Mar. Drugs* **2019**, *17*, 459. [[CrossRef](#)] [[PubMed](#)]
12. Junpeng, J.; Xupeng, C. Monogalactosyldiacylglycerols with High PUFA Content From Microalgae for Value-Added Products. *Appl. Biochem. Biotechnol.* **2019**. [[CrossRef](#)] [[PubMed](#)]
13. Martinez Andrade, K.A.; Lauritano, C.; Romano, G.; Ianora, A. Marine Microalgae with Anti-Cancer Properties. *Mar. Drugs* **2018**, *16*, 165. [[CrossRef](#)] [[PubMed](#)]
14. Da Costa, E.; Silva, J.; Mendonça, S.H.; Abreu, M.H.; Domingues, M.R. Lipidomic approaches towards deciphering glycolipids from microalgae as a reservoir of bioactive lipids. *Mar. Drugs* **2016**, *14*, 101. [[CrossRef](#)] [[PubMed](#)]
15. Akasaka, H.; Sasaki, R.; Yoshida, K.; Takayama, I.; Yamaguchi, T.; Yoshida, H.; Mizushima, Y. Monogalactosyl diacylglycerol, a replicative DNA polymerase inhibitor, from spinach enhances the anti-cell proliferation effect of gemcitabine in human pancreatic cancer cells. *Biochim. Biophys. Acta Gen. Subj.* **2013**, *1830*, 2517–2525. [[CrossRef](#)]
16. Andrianasolo, E.H.; Haramaty, L.; Vardi, A.; White, E.; Lutz, R.; Falkowski, P. Apoptosis-inducing galactolipids from a cultured marine diatom, *Phaeodactylum tricornutum*. *J. Nat. Prod.* **2008**, *71*, 1197–1201. [[CrossRef](#)]
17. Banskota, A.H.; Gallant, P.; Stefanova, R.; Melanson, R.; Oleary, S.J.B. Monogalactosyldiacylglycerols, potent nitric oxide inhibitors from the marine microalga *Tetraselmis chui*. *Nat. Prod. Res.* **2013**, *23*, 1084–1090. [[CrossRef](#)]

18. Banskota, A.H.; Stefanova, R.; Gallant, P.; McGinn, P.J. Mono- and digalactosyldiacylglycerols: Potent nitric oxide inhibitors from the marine microalga *Nannochloropsis granulata*. *J. Appl. Phycol.* **2013**, *27*, 349–357. [[CrossRef](#)]
19. Kobayashi, K. Role of membrane glycerolipids in photosynthesis, thylakoid biogenesis and chloroplast development. *J. Plant. Res.* **2016**, *129*, 565–580. [[CrossRef](#)]
20. Hielscher-Michael, S.; Griehl, C.; Buchholz, M.; Demuth, H.U.; Arnold, N.; Wessjohann, L.A. Natural Products from Microalgae with Potential against Alzheimer's Disease: Sulfolipids Are Potent Glutaminy Cyclase Inhibitors. *Mar. Drugs* **2016**, *14*, 203. [[CrossRef](#)]
21. Manzo, E.; Cutignano, A.; Pagano, D.; Gallo, C.; Barra, G.; Nuzzo, G.; Sansone, C.; Ianora, A.; Urbanek, K.; Fenoglio, D.; et al. A new marine-derived sulfoglycolipid triggers dendritic cell activation and immune adjuvant response. *Sci. Rep.* **2017**, *7*, 6286. [[CrossRef](#)] [[PubMed](#)]
22. Morawski, M.; Schilling, S.; Kreuzberger, M.; Waniek, A.; Jäger, C.; Koch, B.; Cynis, H.; Kehlen, A.; Arendt, T.; Hartlage-Rübsamen, M.; et al. Glutaminy cyclase in human cortex: Correlation with (pGlu)-amyloid- β load and cognitive decline in Alzheimer's disease. *J. Alzheimer's Dis.* **2014**, *39*, 385–400. [[CrossRef](#)] [[PubMed](#)]
23. Joyard, J.; Maréchal, E.; Miège, C.; Block, M.A.; Dorne, A.-J.; Douce, R. Structure, Distribution and Biosynthesis of Glycerolipids from Higher Plant Chloroplasts. In *Lipids in Photosynthesis: Structure, Function and Genetics*; Springer: Dordrecht, The Netherlands, 2006; pp. 21–52.
24. Benning, C.; Ohta, H. Three enzyme systems for galactoglycerolipid biosynthesis are coordinately regulated in plants. *J. Biol. Chem.* **2005**, *280*, 2397–2400. [[CrossRef](#)] [[PubMed](#)]
25. Kobayashi, K.; Nakamura, Y.; Ohta, H. Type A and type B monogalactosyldiacylglycerol synthases are spatially and functionally separated in the plastids of higher plants. *Plant. Physiol. Biochem.* **2009**, *47*, 518–525. [[CrossRef](#)]
26. Qi, Y.; Yamauchi, Y.; Ling, J.; Kawano, N.; Li, D.; Tanaka, K. Cloning of a putative monogalactosyldiacylglycerol synthase gene from rice (*Oryza sativa* L.) plants and its expression in response to submergence and other stresses. *Planta* **2004**, *219*, 450–458. [[CrossRef](#)]
27. Myers, A.M.; James, M.G.; Lin, Q.; Yi, G.; Stinard, P.S.; Hennen-Bierwagen, T.A.; Becraft, P.W. Maize opaque5 encodes monogalactosyldiacylglycerol synthase and specifically affects galactolipids necessary for amyloplast and chloroplast function. *Plant. Cell* **2011**, *23*, 2331–2347. [[CrossRef](#)]
28. Dubots, E.; Audry, M.; Yamaryo, Y.; Bastien, O.; Ohta, H.; Breton, C.; Maréchal, E.; Block, M.A. Activation of the chloroplast monogalactosyldiacylglycerol synthase MGD1 by phosphatidic acid and phosphatidylglycerol. *J. Biol. Chem.* **2010**, *285*, 6003–6011. [[CrossRef](#)]
29. Basnet, R.; Zhang, J.; Hussain, N.; Shu, Q. Characterization and Mutational Analysis of a Monogalactosyldiacylglycerol Synthase Gene OsMGD2 in Rice. *Front. Plant. Sci.* **2019**, *10*, 992. [[CrossRef](#)]
30. Awai, K.; Maréchal, E.; Block, M.A.; Brun, D.; Masuda, T.; Shimada, H.; Takamiya, K.I.; Ohta, H.; Joyard, J. Two types of MGDG synthase genes, found widely in both 16:3 and 18:3 plants, differentially mediate galactolipid syntheses in photosynthetic and nonphotosynthetic tissues in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10960–10965. [[CrossRef](#)]
31. Frentzen, M. Phosphatidylglycerol and sulfoquinovosyldiacylglycerol: Anionic membrane lipids and phosphate regulation. *Curr. Opin. Plant. Biol.* **2004**, *7*, 270–276. [[CrossRef](#)]
32. Roy, A.B.; Hewlins, M.J.; Ellis, A.J.; Harwood, J.L.; White, G.F. Glycolytic breakdown of sulfoquinovose in bacteria: A missing link in the sulfur cycle. *Appl. Environ. Microbiol.* **2003**, *69*, 6434–6441. [[CrossRef](#)] [[PubMed](#)]
33. Speciale, G.; Jin, Y.; Davies, G.J.; Williams, S.J.; Goddard-Borger, E.D. YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat. Chem. Biol.* **2016**, *12*, 215–217. [[CrossRef](#)] [[PubMed](#)]
34. Sanda, S.; Leustek, T.; Theisen, M.J.; Garavito, R.M.; Benning, C. Recombinant *Arabidopsis* SQD1 converts UDP-glucose and sulfite to the sulfolipid head group precursor UDP-sulfoquinovose in vitro. *J. Biol. Chem.* **2001**, *276*, 3941–3946. [[CrossRef](#)] [[PubMed](#)]
35. Yu, B.; Xu, C.C.; Benning, C. *Arabidopsis* disrupted in SQD2 encoding sulfolipid synthase is impaired in phosphate-limited growth. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5732–5737. [[CrossRef](#)] [[PubMed](#)]
36. Lauritano, C.; Martín, J.; De La Cruz, M.; Reyes, F.; Romano, G.; Ianora, A. First identification of marine diatoms with anti-tuberculosis activity. *Sci. Rep.* **2018**, *8*, 2284. [[CrossRef](#)]

37. Brillatz, T.; Lauritano, C.; Jacmin, M.; Khamma, S.; Marcourt, L.; Righi, D.; Romano, G.; Esposito, F.; Ianora, A.; Queiroz, E.F.; et al. Zebrafish-based identification of the antiseizure nucleoside inosine from the marine diatom *Skeletonema marinoi*. *PLoS ONE* **2018**, *13*, e0196195. [[CrossRef](#)]
38. Lauritano, C.; Andersen, J.H.; Hansen, E.; Albrigtsen, M.; Escalera, L.; Esposito, F.; Helland, K.; Hanssen, K.O.; Romano, G.; Ianora, A. Bioactivity Screening of Microalgae for Antioxidant, Anti-Inflammatory, Anticancer, Anti-Diabetes, and Antibacterial Activities. *Front. Mar. Sci.* **2016**, *3*. [[CrossRef](#)]
39. Riccio, G.; Lauritano, C. Microalgae with Immunomodulatory Activities. *Mar. Drugs* **2020**, *18*, 2. [[CrossRef](#)]
40. Giordano, D.; Costantini, M.; Coppola, D.; Lauritano, C.; Núñez Pons, L.; Ruocco, N.; di Prisco, G.; Ianora, A.; Verde, C. Biotechnological Applications of Bioactive Peptides From Marine Sources. In *Advances in Microbial Physiology*; Elsevier: Cambridge, UK, 2018; pp. 171–220. ISBN 9780128151907.
41. Keeling, P.J. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 729–748. [[CrossRef](#)]
42. Falkowski, P.G.; Katz, M.E.; Knoll, A.H.; Quigg, A.; Raven, J.A.; Schofield, O.; Taylor, F.J.R. The evolution of modern eukaryotic phytoplankton. *Science* **2004**, *305*, 354–360. [[CrossRef](#)]
43. Manzo, E.; Gallo, C.; Sartorius, R.; Nuzzo, G.; Sardo, A.; De Berardinis, P.; Fontana, A.; Cutignano, A. Immunostimulatory Phosphatidylmonogalactosyldiacylglycerols (PGDG) from the marine diatom *Thalassiosira weissflogii*: Inspiration for a novel synthetic toll-like receptor 4 agonist. *Mar. Drugs* **2019**, *17*, 103. [[CrossRef](#)] [[PubMed](#)]
44. Zheng, Y.; Anderson, S.; Zhang, Y.; Garavito, R.M. The structure of sucrose synthase-1 from *Arabidopsis thaliana* and its functional implications. *J. Biol. Chem.* **2011**, *286*, 36108–36118. [[CrossRef](#)] [[PubMed](#)]
45. Rocha, J.; Sarkis, J.; Thomas, A.; Pitou, L.; Radzimanowski, J.; Audry, M.; Chazalet, V.; de Sanctis, D.; Palcic, M.M.; Block, M.A.; et al. Structural insights and membrane binding properties of MGD1, the major galactolipid synthase in plants. *Plant. J.* **2016**, *85*, 622–633. [[CrossRef](#)] [[PubMed](#)]
46. Mulichak, A.M.; Theisen, M.J.; Essigmann, B.; Benning, C.; Garavito, R.M. Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13097–13102. [[CrossRef](#)]
47. Okazaki, Y.; Otsuki, H.; Narisawa, T.; Kobayashi, M.; Sawai, S.; Kamide, Y.; Kusano, M.; Aoki, T.; Hirai, M.Y.; Saito, K. A new class of plant lipid is essential for protection against phosphorus depletion. *Nat. Commun.* **2013**, *4*, 1510. [[CrossRef](#)]
48. Zhan, X.; Shen, Q.; Wang, X.; Hong, Y. The Sulfoquinovosyltransferase-like Enzyme SQD2.2 is Involved in Flavonoid Glycosylation, Regulating Sugar Metabolism and Seed Setting in Rice. *Sci. Rep.* **2017**, *7*, 4685. [[CrossRef](#)]
49. Zhan, X.; Shen, Q.; Chen, J.; Yang, P.; Wang, X.; Hong, Y. Rice sulfoquinovosyltransferase SQD2.1 mediates flavonoid glycosylation and enhances tolerance to osmotic stress. *Plant. Cell Environ.* **2019**, *42*, 2215–2230. [[CrossRef](#)]
50. Altschul, S. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
51. Villar, E.; Vannier, T.; Vernet, C.; Lescot, M.; Cuenca, M.; Alexandre, A.; Bachelerie, P.; Rosnet, T.; Pelletier, E.; Sunagawa, S.; et al. The Ocean Gene Atlas: Exploring the biogeography of plankton genes online. *Nucleic Acids Res.* **2018**, *46*, W289–W295. [[CrossRef](#)]
52. Keeling, P.J.; Burki, F.; Wilcox, H.M.; Allam, B.; Allen, E.E.; Amaral-Zettler, L.A.; Armbrust, E.V.; Archibald, J.M.; Bharti, A.K.; Bell, C.J.; et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol.* **2014**, *12*, e1001889. [[CrossRef](#)]
53. Carradec, Q.; Pelletier, E.; Da Silva, C.; Alberti, A.; Seeleuthner, Y.; Blanc-Mathieu, R.; Lima-Mendez, G.; Rocha, F.; Tirichine, L.; Labadie, K.; et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **2018**, *9*, 373. [[CrossRef](#)] [[PubMed](#)]
54. Salazar, G.; Paoli, L.; Alberti, A.; Huerta-Cepas, J.; Ruscheweyh, H.J.; Cuenca, M.; Field, C.M.; Coelho, L.P.; Cruaud, C.; Engelen, S.; et al. Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell* **2019**, *179*, 1068–1083. [[CrossRef](#)] [[PubMed](#)]
55. Rusch, D.B.; Halpern, A.L.; Sutton, G.; Heidelberg, K.B.; Williamson, S.; Yooseph, S.; Wu, D.; Eisen, J.A.; Hoffman, J.M.; Remington, K.; et al. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **2007**, *5*, e77. [[CrossRef](#)] [[PubMed](#)]

56. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)] [[PubMed](#)]
57. R Core Team. 2019: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria. Available online: <http://r-project.org/> (accessed on 18 November 2019).
58. Wickham, H. Package ‘scales’—Scale Functions for Visualization. *CRAN Repos.* **2018**.
59. Wickham, H. *ggplot2 Elegant Graphics for Data Analysis*. Springer: Dordrecht, The Netherlands, 2016; ISBN 978-3-319-24275-0.
60. Papadopoulos, J.S.; Agarwala, R. COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **2007**, *23*, 1073–1079. [[CrossRef](#)]
61. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)]
62. Guindon, S.; Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **2003**, *52*, 596–704. [[CrossRef](#)]
63. Le, S.Q.; Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320. [[CrossRef](#)]
64. Lefort, V.; Longueville, J.E.; Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol. Biol. Evol.* **2017**, *34*, 2422–2424. [[CrossRef](#)]
65. Anisimova, M.; Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **2006**, *55*, 539–552. [[CrossRef](#)] [[PubMed](#)]
66. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)] [[PubMed](#)]
67. Yang, J.; Roy, A.; Zhang, Y. BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* **2013**, *41*, D1096–D1103. [[CrossRef](#)] [[PubMed](#)]
68. Yang, J.; Roy, A.; Zhang, Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **2013**, *29*, 2588–2595. [[CrossRef](#)] [[PubMed](#)]
69. McNicholas, S.; Potterton, E.; Wilson, K.S.; Noble, M.E.M. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 386–394. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).