Royal Netherlands Institute for Sea Research

This is a postprint of:

Villanueva, L., Bale, N., Hopmans, E.C., Schouten, S., Sinninghe Damsté, J.S. (2014). Diversity and distribution of a key sulfolipid biosynthetic gene in marine microbial assemblages. Environmental Microbiology, 16(3), 774-787

Published version: http://dx.doi.org/ 10.1111/1462-2920.12202

Link NIOZ Repository: www.vliz.be/nl/imis?module=ref&refid=239856

[Article begins on next page]

The NIOZ Repository gives free access to the digital collection of the work of the Royal Netherlands Institute for Sea Research. This archive is managed according to the principles of the Open Access Movement, and the Open Archive Initiative. Each publication should be cited to its original source - please use the reference as presented.
When using parts of, or whole publications in your own work, permission from the author(s) or copyright holder(s) is always needed.

1 # Diversity and distribution of a key
2 # sulfolipid biosynthetic gene in marine
3 # microbial assemblages
4

5

6 Laura Villanueva*, Nicole Bale, Ellen C. Hopmans, Stefan Schouten,

7 and Jaap S. Sinninghe Damsté

8

9

10

11 NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Organic

12 Biogeochemistry, PO Box 59, 179AB Den Burg, The Netherlands

13

14 Resubmitted to Environmental Microbiology Reports April 12th 2013

15

16 *To whom correspondence should be addressed. Royal Netherlands Institute for Sea Research

17 P.O. Box 59, NL-1790 AB Den Burg, The Netherlands. E-mail: laura.villanueva@nioz.nl, phone

18 number: +31 (0)222-369-428, fax number: +31 (0)222-319-674.

19

20 Running title: A key sulfolipid biosynthetic gene in marine environments

21 Keywords: Sulfolipid, *sqd*B gene, sulfoquinovosyldiacylglycerol, SQDG, microbial mat, North

22 Sea water

23

Summary

Sulfoquinovosyldiacylglycerols (SQDG) are polar sulfur-containing membrane lipids, whose presence has been related to a microbial strategy to adapt to phosphate deprivation. In this study, we have targeted the *sqd*B gene coding the uridine 5'-diphosphate-sulfoquinovose (UDP-SQ) synthase involved in the SQDG biosynthetic pathway to assess potential microbial sources of SQDGs in the marine environment. The phylogeny of the *sqd*B-coding protein reveals two distinct clusters: one including green algae, higher plants, and cyanobacteria, and another one comprising mainly non-photosynthetic bacteria, as well as other cyanobacteria and algal groups. Evolutionary analysis suggests that the appearance of UDP-SQ synthase occurred twice in cyanobacterial evolution and one of those branches led to the diversification of the protein in members of the phylum Proteobacteria. A search of homologues of *sqd*B-proteins in marine metagenomes strongly suggested the presence of heterotrophic bacteria potential SQDG producers. Application of newly developed *sqd*B-primers in the marine environment revealed a high diversity of sequences affiliated to cyanobacteria and Proteobacteria in microbial mats, while in North Sea surface water most of the detected *sqd*B genes were attributed to the cyanobacterium *Synechococcus* sp. Lipid analysis revealed that specific SQDGs were characteristic of microbial mat depth, suggesting that SQDG lipids are associated with specific producers.

Introduction

Sulfoquinovosyldiacylglycerols (SQDG) are anionic glycoglycerolipids found in thylakoid membranes of photosynthetic eukaryotes (Benson *et al*., 1959) but also in some cyanobacterial genera (Aoki *et al*., 2004), and bacteria from the α and γ-Proteobacteria phyla (Imhoff, 1991; Benning, 1998a). Among the proteobacteria, SQDGs have been reported in some Caulobacteria (Abraham *et al*., 1997), Rhizobiales (Cedergren and Hollingsworth, 1994), and in Gram-positive bacteria (Sprott *et al*., 2006) (see Table S1 for an overview). However, the presence of SQDGs is not universal in all members of these bacterial groups (Selstam and Campbell, 1996; Benning, 1998b). The SQDG lipids have received a lot of attention due to their role in the function and evolution of photosynthetic membranes, and also because of their significance in compensating for nutrient limitation (Benning *et al*., 1993). SQDG lipids were initially thought to be important for photosynthesis due to their correlation with chlorophyll content and being associated with photosynthetic membranes (Barber and Gournaris, 1986; Sakurai *et al*., 2006). However, in other cases no requirement of SQDGs for photosynthesis was found, for example in the anoxygenic phototrophic bacterium *Rhodobacter sphaeroides* or in the cyanobacterium *Synechococcus sp*. PCC7942 (Benning *et al*., 1993; Güler *et al*., 1996). Therefore, there seems to be no strict association between photosynthetic capacity and SQDG presence (Abraham *et al*., 1997; Cedergren and Holingsworth, 1994). A recent study (Aoki *et al*., 2012) has suggested that SQDGs might be involved in DNA replication and eventually in the progression of the cell cycle, which may explain why in certain organisms (e.g. *Synechocystis* sp.) SQDG presence is required for growth.

The first report of the presence of SQDG outside plants and photosynthetic bacteria was by Cedergren and Hollingsworth (1994), when they described the presence of this lipid class in

65 members of the Rhizobiaceae family. The presence of SQDGs was suggested to be related to a
66 functional role of symbiosis of this bacterial group in their association with plants, maybe
67 associated with the transference of the capacity to produce SQDGs from the plant to the
68 bacterium and thus indicative of a long symbiotic association of the partners.

69 Another role invoked for SQDGs is as a substitute for phospholipids, an important
70 biochemical mechanism for cyanobacteria and eukaryotic phytoplankton to maintain
71 photosynthesis in environments where phosphorous is scarce. It has been found that SQDG and
72 phosphatidylglycerol (PG) lipids substitute for each other when their producers grow under
73 phosphate limitation (i.e. increasing SQDG and decreasing PG content) (Benning *et al*., 1993;
74 Güler *et al*., 1996; Sato *et al*., 2000). This can be interpreted as an increase of the phosphorous
75 flow to more critical compounds (e.g. DNA) and compensation for a charge balance in
76 membranes of the photosynthetic apparatus as both PG and SQDGs are anionic lipids. In the
77 algae *Chlamydomonas reinhardtii* a sulfur limitation has been associated to a decrease in the
78 SQDG pool to ensure a sulfur source for protein synthesis (Sugimoto *et al*., 2007). In some
79 Gram-positive microorganisms (e.g. *Marinococcus*) grown under phosphate limitation, salinity
80 in the culture media seems to also influence the abundance of SQDGs (Sprott *et al*., 2006). Van
81 Mooy *et al.* (2006, 2009) suggested that the strategy of picocyanobacteria of minimizing their
82 phosphorus requirements through the synthesis of SQDG could be beneficial to compete against
83 phospholipid-rich heterotrophic bacteria, and thus it could explain the dominance of
84 picocyanobacteria in oligotrophic oceanic regions.

85 Several studies have suggested that SQDG lipids derive from photoautotrophic organisms
86 in surface waters of the North Atlantic, South Pacific, Sargasso Sea and Mediterranean Sea
87 (Popendorf *et al*., 2011a, 2011b; Van Mooy et al., 2006, 2009; Van Mooy and Fredricks, 2010).

88 Schubotz et al., (2009) also detected SQDGs, among other intact polar lipids, in surface waters

89 of the Black Sea and suggested that they represented a mixed community of eukaryotic algae,

90 cyanobacteria and heterotrophic bacteria. In addition, Brandsma *et al*. (2012a, 2012b) studied the

91 abundance and distribution of SQDGs, and other intact polar lipids (IPLs), in the North Sea and

92 did not detect a direct relationship between SQDGs and microbial groups, as identified by flow

93 cytometry, suggesting that they were not derived from a single microbial group. A study of IPL

94 composition and abundance in hypersaline microbial mats (Villanueva *et al*., 2010) observed

95 highest high concentration of SQDGs with saturated and monounsaturated acyl moieties

96 containing 16 and 18 carbon atoms in microbial mat layers underlying the photosynthetic

97 community of the mat and attributed them to heterotrophic microorganisms.

98 The biosynthetic pathway of SQDG lipids is characterized by two specific enzymes: (1)

99 uridine 5'-diphosphate (UDP)-sulfoquinovose (UDP-SQ) synthase, which is responsible for the

100 synthesis of the polar head group, and (2) SQDG synthase that catalyzes the assembly of the

101 sulfolipid. UDP-glucose is condensed with sulfite (SO_3^-) by UDP-SQ synthase, giving UDP-

102 sulfoquinovopyranose (UDP-SQ) as a product that is then transferred to diacylglycerol by SQDG

103 synthase (Figure S1). Several genes related to the SQDG synthetic pathway have been identified,

104 i.e. the *sqd*A gene (Benning and Somerville, 1992b), and the operon containing the *sqd*BCD

105 open reading frames (Benning and Somerville, 1992a). The *sqd*A gene encodes an

106 acyltransferase-like protein (Benning and Somerville, 1992b) similar to bacterial

107 acyltransferases, while *sqd*C-coding protein is a reductase and *sqd*D protein is a predicted

108 glycosyltransferase (Benning and Somerville, 1992a; Rossak *et al*., 1997). These latter genes are

109 essential for SQDG biosynthesis (Benning and Somerville, 1992a, 1992b), but seem to be

110 involved in the assembly of the SQDG lipid rather than in the biosynthesis of its precursor, UDP-

111 SQ, although the exact route is still unknown (Figure S1). The *sqd*B gene coding the UDP-SQ
112 synthase enzyme is the most conserved and more studied of these genes. A survey of *sqd*B genes
113 in the Sargasso Sea metagenome (Van Mooy *et al*., 2006) concluded that the synthesis of
114 sulfolipids in subtropical gyres was confined primarily to picocyanobacteria due to the fact that
115 no *sqd*B gene sequences related to known heterotrophic bacterial SQDG lineages were found.
116 Another study by Popendorf *et al*., (2011a) focused on the microbial sources of intact polar
117 membranes lipids in surface waters from the North Atlantic subtropical gyre, a highly
118 oligotrophic, phosphorus-depleted environment, and concluded that photoautotrophs are the
119 almost exclusive source of SQDGs.

120 SQDG lipids are lipid biomarkers of the abundance of primary producers (algae and
121 cyanobacteria), as well as of heterotrophic SQDG-producers, and thus their abundance in the
122 environment cannot be used to represent one group or the other, but may also indicate specific
123 environmental conditions under which these lipids are produced. In addition, the taxonomic
124 distribution of SQDG producers is still unclear as it is mostly based on culture studies. Here, we
125 investigated the possibility to constrain sources of microbial SQDGs in the environment by
126 examining a functional gene involved in their biosynthesis. The environmental genomic
127 characterization of enzymes involved in lipid biosynthetic pathways provides an independent
128 assessment of an organism's ability to produce a lipid molecule of interest independent of culture
129 conditions. This approach was pioneered by Pearson *et al*. (2007) who investigated potential
130 producers of hopanoids in the environment by analyzing the sequence diversity and distribution
131 of the squalene-hopane cyclase, a key enzyme of the hopanoid synthetic pathway. For our study,
132 the specific aims were to (i) unravel the evolutionary history of the SQDG-lipid biosynthetic
133 enzyme, (ii) survey the occurrence of *sqd*B gene-harboring microorganisms in microbial mats

134 and marine suspended particulate matter, and (iii) compare the gene diversity with the presence
135 and distribution of SQDG lipids in these environments.

136 **Results and Discussion**

137 *UDP-SQ synthase protein phylogeny and sqd*B *gene primer design*

138 The coding sequence of UDP-SQ synthase protein (coded by the *sqd*B gene) annotated in
139 the genomes of members of the phylum Cyanobacteria, Proteobacteria (α and γ-Proteobacteria),
140 Actinobacteria, Crenarchaeota, Euryarchaeota, and eukaryotic photosynthetic phyla
141 Bacillariophyta (diatoms; *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*),
142 Pelagophyceae class (e.g. *Aureococcus anophagefferens*), Phaeophyceae class (brown algae; e.g.
143 *Ectocarpus siliculosus*), Rhodophyta (red algae; e.g. *Cyanidioschyzon merolae*), and phyla
144 Chlorophyta (green algae) were obtained from genomic databases. UDP-SQ synthase protein
145 sequences of other organisms were obtained by annotation based on protein blast (pBLAST)
146 with close homologues in published whole genomes (see supporting information for details). The
147 phylogeny of the UDP-SQ synthase protein sequences (Figure 1) shows a clear separation in
148 three clusters. Cluster 1 comprises the phylum Chlorophyta (e.g. *Ostreococcus* sp.), a higher
149 plant (*Arabidopsis thaliana*) and cyanobacterial genera (e.g. *Synechocystis* sp., and filamentous
150 cyanobacteria such as *Microcoleus* sp.). Cluster 2 includes some unicellular cyanobacterial
151 groups (*Synechococcus* sp., *Prochlorococcus* sp. and *Cyanobium* sp.), other bacteria belonging to
152 the phyla of α- and γ-Proteobacteria (e.g. *Brevundimona*s sp., *Alteromonas* sp.), δ-Proteobacteria
153 (sulfate reducer *Desulfarculus baarsii*), Actinobacteria (*Nocardia*, *Rhodococcus*,
154 *Corynebacterium*), and algae (diatoms, Phaeophyceae, Pelagophyceae, and Rhodophyta).

Finally, cluster 3 contains putative UDP-SQ synthase annotated in archaeal genomes from both Euyarchaeota (e.g. *Haloarcula* sp.) as well as Crenarchaeota (e.g. *Sulfolobu*s sp.).

The presence of the UDP-SQ synthase protein in algae and higher plants is consistent with their ability to produce SQDGs (Benson *et al*., 1959; Sato, 2004), including the algae falling in cluster 2 (Yongmanitchai and Ward, 1993; Yang *et al*., 2011; Martin *et al*., 2011; Wurch *et al*., 2011). There are also numerous reports of SQDG production in the bacterial domain (see Table S1 for an overview but note that many of the bacterial species in cluster 2 have not been tested in culture for their ability to produce SQDGs; all known SQDG-producing organisms are indicated with a star in Figure 1). However, the putative annotation of the UDP-SQ synthase protein in the Archaea (cluster 3) is not supported by any reported occurrence of SQDGs in Archaea. UDP-SQ synthases are closely related to UDP-glucose epimerases and similar sugar nucleotide-modifying enzymes (Essigmann *et al*., 1999). It is possible that the putative archaeal UDP-SQ synthases are not correctly annotated (putative UDP-SQ synthase proteins in the superphylum Archaea were annotated based on pBLAST with query sequences NP_440474.1 *Synechocystis* sp. PCC 6803; YP_399597.1 *Synechococcus* sp. PCC 7942; ABA78726.1 *Rhodobacter sphaeroides* with protein identity $\geq$ 35% and e-value $\leq 1e^{-40}$) or that Archaea are actually able to synthesize UDP-sulfoquinovose but lack the latter enzymes of the SQDG synthetic pathway. Indeed, a study by Meyer *et al.* (2011) has confirmed the presence of UDP-SQ synthase function involved in the N-glycosylation of surface layer glycoproteins in *Sulfolobus acidocaldarius,* and stressed its importance in the maintenance of the cell membrane in extreme environments.

The UDP-SQ synthase protein divergence of the eukaryotes and bacteria in two main clusters (1 and 2) has been observed before (Benning *et al*., 2008) and is thought to be related to

178 the role of SQDGs, or the lack of it, for the maintenance of the photosystem II (Sato *et al*., 2003;
179 Sato, 2004; Sakurai *et al*., 2006); all species falling in cluster 1 are photosynthetic, whereas
180 many of the bacterial species in cluster 2 are not photosynthetic and lack photosystem II.
181 Surprisingly, cluster 2 also includes UDP-SQ synthase protein sequences annotated in genomes
182 from diatoms, brown, and red algae, falling in a separate sub-cluster (Figure 1). This separation
183 of algal UDP-SQ synthase proteins can be attributed to the fact that the chromists (brown algae,
184 diatoms, cryptophytes, etc.) are believed to originate from secondary endosymbiosis by an
185 ancestral red algal cell while green algae and land plants would originate from an ancestral green
186 algal cell (Sato, 2006; Curtis *et al*., 2012). It has also been demonstrated for *Synechococcus* sp.,
187 *Rhodobacter* sp., and *Sinorhizobium meliloti* (bacteria falling in cluster 2) that the absence of
188 SQDGs does not induce any obvious impairment of function (Benning *et al*., 1993; Güler *et al*.,
189 1996; Weissenmayer *et al*., 2000), but their presence is important under conditions of phosphate
190 deprivation (Benning *et al*., 1993). This could imply that in the rest of the SQDG bacterial
191 producers included in cluster 2 (or potential producers based on the presence of *sqd*B gene), as
192 well as the algae included in this cluster, SQDGs are not essential for the photosynthetic
193 apparatus. Further studies involving mutation of the SQDG synthetic pathway in microalgae and
194 bacteria contained in cluster 2 should be performed to confirm this hypothesis.
195 Strikingly, UDP-SQ synthases of cyanobacterial origin are separated over cluster 1 and 2
196 (Figure 1). This separation suggests that the appearance of UDP-SQ synthase may have occurred
197 twice in cyanobacterial evolution. It is likely that all cyanobacteria falling in cluster 1 require
198 SQDG lipids for the maintenance of their photosynthetic apparatus as was described for
199 *Synechocystis* sp. (Sato *et al.*, 2003). Two UDP-SQ synthase proteins of cyanobacterial
200 representatives (i.e., *Lyngbya majuscula* and *Acaryochloris marina*) falling in cluster 2 are

201 closely affiliated to members of the γ-Proteobacteria (e.g. *Thioalkavibrio*). Considering the close
202 spatial proximity of filamentous cyanobacteria with γ-proteobacterial heterotrophic bacteria in
203 marine environments and their metabolic associations (Paerl and Pinckley, 1996), a possible
204 event of horizontal gene transfer (HGT) followed by diversification of the protein in members of
205 the Proteobacteria phylum may be hypothesized. Our phylogenetic analysis does not actually
206 prove HGT events but merely suggests it, and further testing would be required. Alternatively, it
207 is also possible that the ability to synthesize SQDG lipids would have arisen in evolution to
208 secure the function of the photosynthetic apparatus. This function was probably lost over time in
209 some photosynthetic organisms (cluster 2), but the ability was still conserved because of the
210 advantage of regulating the lipid membrane composition to overcome phosphate limitation, and
211 subsequently spread in other bacterial phyla (α-, γ, and δ-Proteobacteria, and Actinobacteria).

212 In summary, the phylogenetic analysis of UDP-SQ synthase proteins presented here
213 emphasizes the diverse evolutionary history of photosynthetic membranes. The SQDG
214 biosynthetic pathway of those organisms in which SQDG lipids seem essential for the
215 maintenance of the photosynthetic function (Chlorophyta green algae, land plants, and
216 cyanobacteria such as *Synechocystis*, *Cyanothece* as well as filamentous cyanobacteria) remained
217 conserved, while UDP-SQ synthase proteins of the cluster 2 diversified and extended to other
218 phyla (α- and γ-Proteobacteria) apparently with no implications in photosynthetic activity. It is
219 remarkable that the UDP-SQ synthase protein displays a high protein sequence divergence even
220 between members of the same bacterial order known to produce SQDG lipids (e.g. 14% protein
221 divergence between *Rhizobium leguminosarum* and *Hoeflea phototrophica*, members of the
222 Rhizobiales order). It seems that the UDP-SQ synthase diversity can probably be attributed to

223 phylogenetic diversification and also be a reflection of an enzyme which conserves its function
224 in spite of its sequence plasticity.

225 *Potential SQDG producers in marine environment microbiomes*

226 As mentioned in the introduction, a previous survey of the *sqd*B gene in the metagenome of
227 surface waters of the Sargasso Sea (Van Mooy *et al*., 2006) revealed an almost exclusive
228 contribution of sequences attributed to picocyanobacteria, suggesting a major role of this group
229 in the SQDG production in marine surface water. Considering the diversification in heterotrophic
230 bacteria of cluster 2 of *sqd*B-coding proteins shown in Figure 1, we performed a more extensive
231 survey of the presence of *sqd*B-coding genes related to this cluster 2 in marine metagenomes
232 deposited in the Integrated Microbial Genomes (IMG) system with microbiome samples
233 (IMG/M) of the Joint Genome Institute (JGI) (see supplementary methods for details). In order
234 to direct our search towards *sqd*B-coding sequences related to cluster 2, we performed a
235 pBLAST search using the UDP-SQ synthase sequence of *Rhodobacter sphaeroides* as a query
236 sequence. pBLAST was restricted to search homologues in marine environmental microbiome
237 metagenomes. The marine microbiomes included: Artic bacterioplankton, (sub-) surface
238 sediments, planktonic communities, hypersaline mats, salt marshes, whale fall, extreme
239 environments, oxygen minimum zones (OMZs), and anoxic basin (see IMG/M marine samples
240 for details and Table S2 for the codes of microbiomes mentioned in Figure S2). Protein
241 sequences were recovered and aligned with the sequences included in Figure 1 to generate the
242 tree in Figure S2.

243 Putative UDP-SQ synthase proteins recovered from coastal waters, intertidal sediments,
244 anoxic basins, whale fall, and OMZs clustered with *sqd*B-coding proteins of heterotrophic
245 bacteria of the α-Proteobacteria Caulobacterales and Parvularculares (*Phenylobacterium,*

246 *Brevundimonas, Oceanicaulis, Parvularcula*) and γ-Proteobacteria (*Alteromonas*,

247 *Thiorhodococcus*, *Nitrosococcus* and *Thioalkalivibrio*) of cluster 2 (Figure S2). In addition,

248 sequences from OMZs were closely related to putative UDP-SQ synthases of Actinomycetales

249 (*Nocardia, Rhodococcus*) (Figure S2). Sequences obtained from a metagenome of a salt marsh

250 microbial mat were related to the *sqd*B-proteins of cluster *sqd*B2B (*Rhodobacter* among others)

251 (Figure S2). Our analysis of *sqd*B-protein sequences in marine metagenomes strongly suggests

252 the presence of heterotrophic organisms with the potential to produce SQDGs in these

253 ecosystems. In addition, the pBLAST search with the UDP-SQ synthase sequence of

254 *Rhodobacter sphaeroides* as a query sequence returned some metagenomic sequences closely

255 related to cyanobacteria of the cluster 2 and 1. Sequences from hydrothermal plumes and OMZs

256 were closely related to the *sqd*B-proteins of the cyanobacterial group in cluster 2 (*Synechococcus*

257 and *Prochlorococcus*) (Figure S2). Interestingly, our metagenomic sequence analysis recovered

258 sequences from *Lyngbya* mats of Guerrero Negro (Lyngb mats; Figure S2) closely related to the

259 *sqd*B-proteins of the *Lyngbya* sp. PCC8106 sequence of cluster 1 as well as sequences related to

260 the *Lyngbya majuscula* sequence of cluster 2, which supports the observation made above about

261 the presence of putative cyanobacterial UDP-SQ synthases from *Lyngbya* sp. associated with

262 both cluster 1 and 2 (Figure 1).

263 *Detection of* *sqd*B *gene sequences and sulfolipids in the marine environment*

264 Based on the homology of UDP-SQ synthase proteins (Figure 1), we designed primers

265 for the detection of *sqd*B gene fragments of some cyanobacterial genera in cluster 1 (cluster

266 *sqd*B1A and *sqd*B1B). In cluster 2, only two groups included sequences related closely enough

267 to be able to design primers for their amplification: (i) the cyanobacterial *sqd*B sequences

268 included in the *sqd*B2A cluster, and (ii) the *sqd*B2B cluster, including *sqd*B sequences of the α-

269 Rhodobacterales and α-Rhizobiales groups. These primer pairs were first tested with pure
270 cultures to optimize PCR conditions (Table S3).

271 In order to test the *sqd*B gene designed primers, we applied them in microbial mats from
272 Dutch barrier island Schiermonnikoog (Stal *et al*., 1985) and the salterns of Salins-de-Giraud,
273 Camargue (Caumette *et al.*, 1994), and surface water suspended particulate matter (SPM) from
274 the North Sea. In addition, we also analyzed the SQDG lipid composition in the same samples.
275 Lipid analysis showed the presence of SQDGs in all samples analyzed (Table 1), indicating the
276 presence of SQDG-synthesizing microbes. It is important to note that the diversity of UDP-SQ
277 synthase (*sqd*B coding gene) does not cause SQDG lipid diversity (differences in acyl chains) as
278 UDP-SQ synthase is involved in the condensation of sulfite with UDP-glucose in the formation
279 of the polar head group of the SQDG lipid rather than in the formation and attachment of the acyl
280 chains (Figure S1). Thus, the study presented here analyzes the *sqd*B gene as a marker of the
281 presence of potential SQDG producers and compares it with the diversity of SQDG lipids that is
282 also indicative of different producers. However, as observed in Table S1, in general most of the
283 microbial groups synthesize a wide diversity of SQDGs in different relative abundance which
284 makes the interpretation of sulfolipid producers based on the SQDGs detected complicated
285 without the taxonomic affiliation provided by the analysis of the *sqd*B-gene sequence presented
286 in this study.

287 **Microbial mats.** Microbial mat samples were chosen as they are well-known sources of bacterial
288 diversity and lipid biomarkers (e.g. SQDGs, Villanueva *et al*., 2010), and also to compare the
289 SQDG diversity found in microbial mats from two different location/physicochemical
290 characteristics in contrast to the diversity in the North Sea SPM. Sequences of *sqd*B1A and
291 *sqd*B1B clusters, targeting cyanobacterial SQDG producers, were amplified from both Camargue

292 and Schiermonnikoog microbial mats and were clearly clustered according to the microbial mat
293 location (Figure 2). *sqd*B1A cluster sequences from the Camargue microbial mat were more
294 diverse, but all sequences from both mat systems were closely related to the *sqd*B gene of
295 *Microcoleus chthonoplastes* (Figure 2A). *sqd*B1B gene sequences (Figure 2B) recovered from
296 microbial mats were also clearly clustered according to their location and closely related to the
297 *M. chthonoplastes sqd*B gene sequence, as seen for *sqd*B1A cluster sequences (Figure 2A), but
298 were also related to *sqd*B sequences of the *Synechocystis* sp., *Cylindrospermopsis* and
299 *Raphidiopsis* cluster (Figure 2B).
300 Previous studies have described a high diversity of filamentous cyanobacteria in these
301 microbial mat systems. Microscopic and genetic studies in the hypersaline mats of Camargue
302 reported the presence of *M. chthonoplastes*, *Halomicronema excentricum*, *Leptolyngbya* sp.,
303 *Limnothrix* sp., and *Pseudoanabaena* sp., as filamentous cyanobacteria in decreasing order of
304 importance, while unicellular cyanobacteria *Chroococcus* sp., *Microcystis* sp., and members of
305 *Synechocystis*, and *Gloeocapsa* accounted for 24% of total cyanobacteria (Fourçans *et al*., 2004).
306 In Schiermonnikoog mats, cyanobacterial population is normally dominated by *Oscillatoria* sp.,
307 *Spirulina* sp., *M. chthonoplastes*, and members of the LPP-B group (*Lyngbya*, *Plectonema* and
308 *Phormidium*) (Stal *et al*., 1985; Bolhuis & Stal, 2011). The fact that the *sqd*B1 cluster sequences
309 reported here could not be clearly assigned to known cyanobacterial SQDG-producers (with the
310 exception of the *sqd*B sequence of *M. chthonoplastes*) can be due to the lack of *sqd*B gene
311 sequences derived from filamentous cyanobacteria in whole genome databases or absence of
312 SQDG-production capacity in this group of cyanobacteria.
313 The primers developed in this study for the *sqd*B2A cluster, targeting other
314 cyanobacteria, failed to amplify DNA from microbial mats, probably due to the fact that the

315 targeted cyanobacteria of this cluster (*Synechococcus* sp.*, Prochlorococcus* sp.*, and Cyanobium*
316 sp.) are not abundant in these microbial mat systems (Stal *et al*., 1985; Bolhuis & Stal, 2011;
317 Caumette *et al*., 1994). However, the *sqd*B2B cluster primer, targeting the α-Rhodobacterales
318 and α-Rhizobiales groups, gave positive amplification and the diversity of the recovered *sqd*B2B
319 cluster gene sequences was large at the DNA level. Most *sqd*B2B gene sequences from microbial
320 mats showed, as expected, a close affiliation to either α-Proteobacteria of the order
321 Rhodobacterales or Rhizobiales (Figure 3). Some of the Camargue mat *sqd*B2B cluster gene
322 sequences were also related to *Paracoccus* sp., and *Rhodobacter* sp. In addition, one of the
323 *sqd*B2B cluster gene sequences of the Schiermonnikoog mat was close to the *sqd*B gene of
324 *Oceanicaulis alexandrii*. The *Oceanicaulis* genus belongs to the Caulobacteriales order that
325 includes known SQDG lipid producers (Table S1). Caulobacteria are ubiquitous, thrive in
326 oligotrophic conditions, and presumed to be important in the mineralization of dissolved organic
327 material in aquatic environments (Staley *et al*., 1987). In fact, the lifestyle of Caulobacteria,
328 based on a mode of reproduction with a motile cell to disperse and avoid competition for
329 resources, is consistent with a high tolerance to nutrient starvation (Poindexter, 1981), and might
330 explain why this microbial group displays the ability to produce SQDG in their membranes.
331 The presence and diversity of *sqd*B gene sequences related to cyanobacterial genera and
332 also to α-Proteobacteria recovered from microbial mats indicates a high diversity of potential
333 SQDG-producers. The high diversity of *sqd*B gene sequences belonging to the Rhodobacterales
334 and Rhizobiales orders is especially remarkable, and suggests that these microbial mat
335 communities could be used in future studies to screen the presence of active SQDG lipid
336 production by targeting the gene expression of the detected *sqd*B gene sequences under different
337 nutrient limitations.

338 SQDG lipids detected in both microbial mats contained acyl moieties with 14 to 19
339 carbon chains (Table 1). In the case of the Camargue mat, lipid analysis was performed in higher
340 resolution by slicing the mat in three layers: upper (1 mm); intermediate (5 mm); lower (1 cm),
341 which visually separated upper diatom-cyanobacterial, purple sulfur bacteria-rich, and black
342 sulfate reducing-bacteria layers (Caumette *et al*., 1994). The most abundant SQDGs in the upper
343 layer of the Camargue mat were SQDG with 16:0/18:1 and 16:0/18:2 fatty acid moieties,
344 respectively, that were also present in the Schiermonnikoog mat (Table 1). In the middle layer of
345 the Camargue mat, SQDG (16:0/18:1) and (16:0/18:2) had also high relative abundance as in the
346 upper layer, but also the relative abundances of SQDG (16:0/16:1) and (16:0/16:0) increased. In
347 addition, SQDG (16:0/18:3) was only found in the Schiermonnikoog mat and in the middle layer
348 of the Camargue mat. In the deeper layer of the Camargue microbial mat, the relative abundance
349 of SQDG (16:0/18:2) and (16:0/18:1) decreased dramatically. Some cyanobacterial groups can
350 synthesize acyl side chains with several unsaturations (18:2, 18:3) because they possess
351 desaturases (Murata *et al*., 1992), while members of the α-Proteobacteria phylum seem to mainly
352 display SQDGs with saturated (16:0, 18:0) or monounsaturated (18:1) acyl moieties (Table S1).
353 Considering the ability of cyanobacteria to synthesize fatty acid chains with multiple
354 unsaturations (Murata *et al*., 1992), it is then likely that SQDG lipids such as SQDG (16:0/18:2)
355 are associated with cyanobacterial photosynthetic biomass due to their high relative abundance
356 and position in the upper most layers of the microbial mat samples.
357 Some SQDG lipids were present in all microbial mat samples independent of location
358 and depth, i.e. SQDG (16:0/16:1) and (16:0/16:0) that had a higher relative abundance in the
359 middle layer of the Camargue mat and in the Schiermonnikoog mat. This may be attributed to
360 the fact that the Schiermonnikoog mat comprised topmost 3 cm of the mat and is thus

361 comparable to the SQDG profile of the combined three layers of the Camargue mat. Other
362 SQDG lipids (e.g. SQDG 18:0/19:1) were not detected in upper layers of the latter mat but
363 increased their relative percentages with depth (Table 1). SQDGs containing C19 fatty acids
364 were restricted to the deep black anaerobic layer of the Camargue mats which is dominated by
365 sulfate reducing bacteria (Fourçans *et al*., 2004), and thus we may speculate that anaerobic or
366 facultative anaerobic microorganisms produce this type of SQDG lipids. In fact, the synthesis of
367 SQDGs by anaerobic microorganisms is possible as revealed by the presence of an annotated
368 UDP-SQ synthase in the strictly anaerobic sulfate reducer *Desulfarculus baarsii* (protein
369 accession number YP_003805997) (see Figure 1, cluster 2).
370 **Suspended particular matter (SPM) of the North Sea**. Amplification of *sqd*B gene fragments
371 using *sqdB*2A and *sqd*B2B cluster primer pairs was positive in the North Sea SPM, while
372 *sqd*B1A and *sqd*B1B cluster primers gave no amplification. The *sqd*B2A cluster sequences
373 amplified from the North Sea SPM displayed low sequence diversity and were closely related to
374 the cyanobacterium *Synechococcus* (Figure 4). Sequences of the *sqd*B2B gene cluster (Figure 3)
375 recovered from the North Sea SPM clustered with the α-Rhizobiales but other sequences were
376 closely related to *sqd*B gene sequences of Actinomycetales (Actinobacteria). Despite of the fact
377 that *sqd*B2B cluster gene sequences were retrieved, the contribution of α-Proteobacteria to the
378 pool of SQDG lipids is probably minor as SQDGs with 18:0 and 18:1 acyl chains characteristic
379 of this group (Table S1) had a low relative abundance in the North Sea SPM (Table 1).
380 The analysis of SQDG lipids showed predominantly SQDGs with acyl moieties with 14
381 and, to a lesser extent, 16 carbon atoms (Table 1). High relative abundances of SQDGs with 14:0
382 and 16:0 acyl moieties have been previously described in the water column of the eastern
383 subtropical South Pacific (Van Mooy and Fredricks, 2010), and in North Sea surface waters

384 (Brandsma *et al*., 2012a). SQDGs with 14:0 acyl chains only represent less than 0.5% of the
385 mole percent of SQDG fatty acids in cyanobacterial cultures analyzed up to date (Murata *et al*.,
386 1992) but they have been detected in high relative abundance in the marine diatoms *Skeletonema*
387 sp. (Yan *et al*., 2011), and *Phaeodactylum tricornutum* (Yongmanitchai and Ward, 1993). Thus,
388 the high relative abundance of SQDGs with 14 carbon acyl chains in the North Sea water could
389 be attributed to the presence of eukaryotic microalgae (Brandsma *et al*., 2012a), while SQDGs
390 with 18:2 and 18:3 acyl moieties could be more characteristic of cyanobacteria, and a higher
391 presence of SQDGs with 18:1 (and probably 17 and 19 carbon atom acyl chains) could be an
392 indicator of a more predominant SQDG-producing Proteobacterial population (Table S1).

393 **Conclusions**

394 In this study we have constructed an extended phylogeny of UDP-SQ synthase proteins
395 and revealed the presence of two distinct clusters comprising: (1) green algae, higher plants, and
396 cyanobacteria, and (2) mainly non-photosynthetic bacteria, as well as cyanobacteria and other
397 algal groups. As previously suggested, the diversification of the UDP-SQ synthase might be
398 related to the role of SQDG lipids in the maintenance of photosynthetic function, and also reflect
399 the evolution of algal groups. The evolutionary analysis also suggested that the appearance of
400 UDP-SQ synthase occurred twice in cyanobacterial evolution and one of those branches led to
401 the acquisition of SQDG lipid synthesis ability in members of the Proteobacteria phylum. A
402 search of homologues of cluster 2 *sqd*B-proteins in marine metagenomes strongly suggested the
403 presence of heterotrophic organisms with potential to produce SQDGs in marine ecosystems.
404 By using specific primers we assessed the diversity of *sqd*B genes in marine microbial
405 mats and North Sea SPM which showed that the diversity of potential SQDG-producers in
406 microbial mats seems to be higher than in marine surface waters, and widespread in the

407 cyanobacteria phylum, but also in the Proteobacteria. Lipid analysis also showed a high diversity
408 of SQDG lipids. Some of the detected SQDGs were specific of sample location and depth and it
409 was possible to link them to potential microbial sources, which demonstrates the potential of
410 combining SQDG lipid and *sqd*B gene profiling as a tool to track diversity and potentiality of
411 SQDG lipid production in the environment. Future studies will need to assess the ability of the
412 *sqd*B-gene harboring bacteria to produce SQDG lipids and under which conditions or nutrient
413 limitation this production occurs. In addition, the newly developed *sqd*B gene primers tested in
414 this study can be applied in future studies to determine if a higher expression of the *sqd*B gene is
415 correlated to a higher abundance of specific SQDG lipids.

416

417 Acknowledgements

418 This study was partly supported by the Division of Earth and Life Sciences of the Dutch
419 Organization for Scientific Research (NWO) by a grant (839.08.331) to JSSD. We thank Prof.
420 Lucas Stal (NIOZ) and Prof. Ricardo Guerrero for help in obtaining the microbial mat samples
421 used in this study.

422

423

424

425

426

427 ## References

428 Abraham, W.R., Meyer, H., Lindholst, S., Vancanneyt, M., and Smit, J. (1997) Phospho- and
429 sulfolipids as biomarkers of Caulobacter sensu lato, *Brevundimonas* and *Hyphomonas*. *Syst Appl*
430 *Microbiol* **20:** 522–539.

431 Aoki, M., Sato, N., Meguro, A., and Tsuzuki, M. (2004) Differing involvement of
432 sulfoquinovosyl diacylglycerol in photosystem II in two species of unicellular cyanobacteria.
433 *Eur J Biochem 271:* 685–693.

434 Aoki, M., Tsuzuki, M., and Sato, N. (2012) Involvement of sulfoquinovosyl diacylglycerol in
435 DNA synthesis in *Synechocystis* sp. PCC 6803. *BMC Res Notes* **5:** 98.

436 Barber, J., and Gounaris, K. (1986) What role does sulfolipid sulfoquinovosyl diacylglycerol
437 play within the thylakoid membrane? *Photosynthes Res* **9:** 239–249.

438 Benning, C., Garavito, R.M., and Shimojima, M. (2008) Sulfolipid biosynthesis and function in
439 plants. In Sulfur Metabolism in Phototrophic Organisms. Advances in Photosynthesis and
440 Respiration (Rüdiger Hell *et al*. eds. Springer) Volume 27, pp 185–200.

441 Benning, C. (1998a) Biosynthesis and function of the sulfolipid sulfoquinovosyl diacylglycerol.
442 *Annu Rev Plant Physiol Plant Mol Biol* **49:** 53–75.

443 Benning, C. (1998b) Membrane lipids in anoxygenic bacteria. In Siegenthaler, P.A., Murata, N.
444 (eds) Lipids in photosynthesis. Kluwer Academic. Dordrecht. pp 83–101.

445 Benning, C., Beatty, J.T., Prince, R.C., and Somerville, C.R. (1993) The sulfolipid
446 sulfoquinovosyl diacylglycerol is not required for photosynthetic electron transport in

447 *Rhodobacter sphaeroides* but enhances growth under phosphate limitation. *Proc Natl Acad Sci*
448 *USA* **90:** 1561–1565.

449 Benning, C., and Somerville, C.R. (1992a) Identification of an operon involved in sulfolipid
450 biosynthesis in *Rhodobacter sphaeroides*. *J Bacteriol* **174:** 6479–6487.

451 Benning, C., and Somerville, C.R. (1992b) Isolation and genetic complementation of a
452 sulfolipid-deficient mutant of *Rhodobacter sphaeroides*. *J Bacteriol* **174:** 2352–2360.

453 Benson, A.A., Daniel, H., and Wiser, R. (1959) A sulfolipid in plants. *Proc Natl Acad Sci USA*
454 **45:** 1582–1587.

455 Bolhuis, H., and Stal, L.J. (2011) Analysis of bacterial and archaeal diversity in coastal microbial
456 mats using massive parallel 16S rRNA gene tag sequencing. *ISME J* **5:** 1701–1712.

457 Brandsma, J., Hopmans, E.C., Phillippart, K.J.M., Vedhuis, M.J.W., Schouten, S., and Sinninghe
458 Damsté, J.S. (2012a) Temporal variations in abundance and composition of intact polar lipids in
459 North Sea coastal marine water. *Biogeosciences* **9:** 1073–1084.

460 Brandsma, J., Hopmans, E.C., Brussaard, C., Witte, H., Schouten, S., and Sinninghe Damsté, J.S.
461 (2012b) Spatial distribution of intact polar lipids in North Sea surface waters: Relationship with
462 environmental conditions and microbial community composition. *Limnol Oceanogr* **57:** 959–
463 973.

464 Caumette, P., Matheron, R., Raymond, N., and Relexans, J.C. (1994) Microbial mats in the
465 hypersaline ponds of Mediterranean salterns (Salins-de-Giraud, France). *FEMS Microbiol Ecol*
466 **13:** 273–286.

467 Cedergren, R.A., and Hollingsworth, R.I. (1994) Occurrence of sulfoquinovosyl diacylglycerol
468 in some members of the family Rhizobiaceae. *J Lipid Res* **35:** 1452–1461.

469 Curtis, B.A., Tanifuji, G., Brurki, F., et al. (2012) Algal genomes reveal evolutionary mosaicism
470 and the fate of nucleomorphs. *Nature* **492:** 59–65.

471 Essigmann, B., Hespenheide, B. H., Kuhn, L.A., and Benning, C. (1999) Prediction of the
472 Active-Site Structure and NAD^+ Binding in SQD1, a Protein Essential for Sulfolipid
473 Biosynthesis in *Arabidopsis*. *Archiv Biochem Biophys* **369:** 30–41.

474 Fourçans, A., García de Oteyza, T., Wieland, A., Solé, A., Diestra, E., van Bleijswijk, J.,
475 Grimalt, J.O., Kühl, M., Esteve, I., Muyzer, G., Caumette, P., and Duran, R. (2004)
476 Characterization of functional bacterial groups in a hypersaline microbial mat community
477 (Salins-de-Giraud, Camargue, France). *FEMS Microb Ecol* **51:** 55–70.

478 Güler, S., Seeliger, A., Härtel, H., Ranger, G., and Benning, C. (1996) A null mutant of
479 *Synechococcus* sp. PCC7942 deficient in the sulfolipid sulfoquinovosyl diacylglycerol. *J Biol.*
480 *Chem* **271:** 7501–7507.

481 Imhoff, J.F. (1991) Polar lipids and fatty acids in the genus *Rhodobacter*. *Syst Appl Microbiol*
482 **14:** 228–234.

483 Martin, P., Van Mooy, B.A.S., Heihoff, A., and Dyhrman, S. T. (2011) Phosphorous supply
484 drives rapid turnover of membrane phopsholipids in the diatom *Thalassiosira pseudonana*. *ISME*
485 *J* **5:** 1057–1060.

486 Meyer, B.H., Zolghadr, B., Peyfoon, E., Pabst, M., Panico, M., Morris, H.R., Haslam, S.M.,
487 Messner, P., Schäffer, C., Dell, A., and Albers, S.V. (2011) Sulfoquinovose synthase - an
488 important enzyme in the N-glycosylation pathway of *Sulfolobus acidocaldarius*. *Mol Microbiol*
489 **82:** 1150–1163.

490 Murata, N., Wada, H., and Gombos, Z. (1992) Modes of fatty-acid desaturation in cyanobacteria.

491 *Plant Cell Physiol* **33:** 933–941.

492 Paerl, H.W., and Pinckley J.L. (1996) A mini-review of microbial consortia: Their roles in
493 aquatic production and biochemical cycling. *Microb Ecol* **31:** 225–247.

494 Pearson, A., Flood Page, S.R., Jorgenson, T.L., Fischer, W.W., and Higgins, M.B. (2007) Novel
495 hopanoid cyclases from the environment. Environ Microbiol **9:** 2175–2188.

496 Poindexter, J.S. (1981) Oligotrophy. Fast and famine existence. *In* Microbial Ecology, vol. 5, pp.
497 63–89. Edited by M. Alexander. New York: Plenum.

498 Popendorf, K.J., Lomas, M.W., and Van Mooy, B.A.S. (2011a) Microbial sources of intact polar
499 diacylglycerolipids in the Western North Atlantic Ocean. *Org. Geochem.* **42:** 803–811.

500 Popendorf, K.J., Tanaka, T., Pujo-Pay, M., Lagaria, A., Courties, C., Conan, P., Orial, L., Sofen,
501 L.E., Mountin, T., and Van Mooy, B.A.S. (2011b) Gradients in intact polar diacylglycerolipids
502 across the Mediterranean Sea are related to phosphate availability. *Biogeosciences* **8:** 3733–
503 3745.

504 Rossak, M., Schäfer, A., Xu, N., Gage, D.A., and Benning, C. (1997) Accumulation of
505 sulfoquinovosyl-1-O-dihydroxyacetone in a sulfolipid-deficient mutant of *Rhodobacter*
506 *sphaeroides* inactivated in *sqd*C. *Arch Biochem Biophys* **340:** 219–230.

507 Sakurai, I., Shen, J.R., Leng, J., Ohashi, S., Kobayashi, M., and Wada, H. (2006) Lipids in
508 oxygen-evolving photosystem II complexes of cyanobacteria and higher plants. *J. Biochem.*
509 *(Tokyo)* **140:** 201–209.

510 Sato, N. (2006) Origin and evolution of plastids: genomic view on the unification and diversity
511 of plastids, p. 75–102. *In* R. R. Wise and J. K. Hoober (ed.), The structure and function of
512 plastids. Springer, Berlin, Germany.

513 Sato, N. (2004) Roles of the acidic lipids sulfoquinovosyldiacylglycerol and
514 phosphatidylglycerol in photosynthesis: their specificity and evolution. *J Plant Res* **117:** 495–
515 505.

516 Sato, N., Aoki, M., Maru, Y., Sonoike, K., Minoda, A., and Tsuzuki, M. (2003) Involvement of
517 sulfoquinovosyl diacylglycerol in the structural integrity and heat tolerance of photosystem II.
518 *Planta* **217:** 245–251.

519 Sato, N., Hagio, M., Wada, H., and Tsuzuki, M. (2000) Environmental effects on acidic lipids of
520 thylakoid membranes. *Biochem Soc Trans* **28:** 912–914.

521 Schubotz, F., Wakeham, S.G., Lipp, J.S., Fredricks, H.F., and Hinrichs, K.-U. (2009) Detection
522 of microbial biomass by intact membrane lipid analysis in the water column and surface
523 sediments of the Black Sea. *Environ Microbiol* **11:** 2720–2734.

524 Selstam, E., and Campbell, D. (1996) Membrane lipid composition of the unusual
525 cyanobacterium *Gloeobacter violaceus* sp. PCC7421,which lacks sulfoquinovosyl
526 diacylglycerol. *Arch Microbiol* **166:** 132–135.

527 Sprott, G.D., Bakouche, L., and Rajagopal, K. (2006) identification of sulfoquinovosyl
528 diacylglycerol as a major polar lipid in *Marinococcus halophilus* and *Salinicoccus hispanicus*
529 and substitution with phosphatidylglycerol. *Can J Microbiol* **52:** 209–219.

530 Stal, L.J., van Gemerden, H., and Krumbein, W.E. (1985) Structure and development of a
531 benthic marine microbial mat. *FEMS Microbiol Ecol* **31:** 111–125.

532 Staley JT, Konopka AE, Dalmasso JP. (1987) Spatial and temporal distribution of Caulobacter
533 spp. in two mesotrophic lakes. *FEMS Microbiol Ecol* ***45:*** 1–6.

534 Sugimoto, K., Sato, N., and Tsuzuki, M. (2007) Utilization of a chloroplast membrane sulfolipid
535 as a major internal sulfur source for protein synthesis in the early phase of sulfur starvation.
536 *FEBS Lett* **581:** 4519–4522.

537 Van Mooy, B.A.S., Fredricks, H.F. (2010) Bacterial and eukaryotic intact polar lipids in the
538 eastern subtropical South Pacific: Water-column distribution, planktonic sources, and fatty acid
539 composition. *Geochim Cosmochim Acta* **74:** 6499–6516.

540 Van Mooy, B.A.S., Fredricks, H.F., Pedler, B.E., Dyhrman, S.T., Karl, D.M., Koblížek, M., *et*
541 *al*. (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus
542 scarcity. *Nature* **458:** 69–72.

543 Van Mooy, B.A.S., Rocap, G., Fredricks, H.F., Evans, C.T., and Devol, A.H. (2006) Sulfolipids
544 dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine
545 environments. *Proc Natl Acad Sci USA* **103:** 8607–8612.

546 Villanueva, L., del Campo, J., Guerrero, R., and Geyer, R. (2010) Intact phospholipid and
547 quinone biomarkers to assess microbial diversity and redox state in microbial mats. *Microb Ecol*
548 **60:** 226–238.

549 Weissenmayer, B., Geiger, O., and Benning, C. (2000) Disruption of a gene essential for
550 sulfoquinovosyldiacilglycerol biosynthesis in *Sinorhizobium meliloti* has no detectable effect on
551 rot nodule symbiosis. *Mol Plant Microb Inter* **13:** 666–672.

552 Wurch, L.L., Bertrand, E.M., Saito, M.A., Van Mooy, B.A.S., and Dyhrman, S. (2011) proteome
553 changes driven by phosphorous deficiency and recovery in the brown tide-forming alga
554 *Aureococcus anophagefferens*. *Plos One* **6:** e28949.

555 Yang, X., Chen, D., Xu, J., and Zhou, C. (2011) Profiles of photosynthetic glycerolipids in three
556 strains of Skeletonema determined by UPLC-Q-TOF-MS. *J Appl Phycol* **23:** 271–282.

557 Yongmanitchai, W., and Ward, O.P. (1993) Positional distribution of fatty acids, and molecular
558 species of polar lipids in the diatom *Phaeodactylum tricornutum*. *J Gen Microbiol* **139:** 465–472.

559
560
561
562
563
564
565
566
567
568
569
570
571

572 # Supporting information

573 **Appendix S1.** Experimental procedures.

574 *Sampling*

575 North Sea suspended particulate matter (SPM) from the upper 2 m was sampled at a jetty
576 platform on the 28th July 2010 in high tide at the NIOZ at the western entrance of the North Sea
577 into the Wadden Sea at the Island Texel (53°0’2’’N, 4°7’2’’E) by using a bucket. Mat samples
578 from Schiermonnikoog (53°29´N and 6°08´E, Stal *et al*., 1985) were treated as a whole core
579 while the Camargue microbial mats (44°40′N, 4°51′E, Caumette *et al*., 1994) were sliced in three
580 layers in depth: upper (1 mm); intermediate (5 mm); lower (1 cm), which visually separated
581 upper diatom-cyanobacterial, purple sulfur bacteria-rich, and black sulfate reducing-bacteria
582 layers. For genetic analysis, the three layers of the Camargue mat were combined before DNA
583 extraction.

584 *DNA extractions*

585 Approximately 1 L of North Sea water suspended particulate matter (SPM) sample was filtered
586 through a 142 mm diameter, 0.2 µm pore size polycarbonate filter (Millipore, Billerica, MA) and
587 stored at –80°C until extraction. Filter was cut in sections and extracted by bead-beating with 1.5
588 g of sterile 0.1-mm zirconium beads (Biospec, Bartlesville, OK) in a extraction buffer containing
589 10 mM Tris-HCl pH 8, 25 mM Na_2EDTA pH 8, 1% (v/v) sodium dodecyl sulfate (SDS), 100
590 mM NaCl, and molecular biology grade water. The filter sample was extracted with phenol-
591 chloroform (Sambrook et al., 1989). After extraction, DNA was precipitated using ice-cold
592 ethanol, dried, and re-dissolved in 100 µl of 10 mM Tris-HCl, pH 8. The DNA was further
593 purified by using DNeasy columns from Qiagen (Qiagen Inc., Valencia, CA) following

594 manufacturer's instructions. Microbial mat samples were homogenized and extracted with

595 PowerBiofilm™ DNA Isolation Kit from Mo-bio (Mo-Bio Lab Inc., Carlsbad, CA). Total

596 nucleic acid concentrations were quantified spectrophotometrically (Nanodrop, Thermo

597 Scientific, Wilmington, DE, USA) and checked by agarose gel electrophoresis for quality.

598 Extracts were kept frozen at -80°C.

599 *PCR and cloning conditions*

600 Primer pairs for the different *sqd*B gene clusters were designed manually and check for

601 secondary structures and % G + C the Sigma DNA calculator (http://www.sigma-

602 genosys.com/calc/DNACalc.asp) and Primer3web (http://primer3.wi.mit.edu/) (see Table S3).

603 PCR reaction mixture was the following (final concentration): Q-solution (PCR additive) 1×;

604 PCR buffer 1×; BSA (200 µg/ml); dNTPs (20 µM); primers (0.2 pmol/µl); $MgCl_2$ (1.5 mM);

605 1.25 U Taq polymerase (Qiagen, Valencia, CA, USA) and/or BioThermD Taq DNA polymerase

606 (Semiramis Genetics Ltd., Manchester, UK). PCR conditions for these amplifications were the

607 following: 95°C, 5 min; 40× [95°C, 1 min; Tm, 1 min; 72°C, 1 min]; final extension 72°C, 5

608 min. A gradient PCR cycle was performed for each set of primers and samples and representative

609 pure cultures (see Table S2 for details) from 48 to 61°C melting temperature. Positive

610 amplification bands were excised from agarose gel and gel or PCR purified (QIAquick gel/PCR

611 purification kit, Qiagen) and cloned in the TOPO-TA cloning® kit from Invitrogen (Carlsbad,

612 CA, USA) and transformed in *E. coli* TOP10 cells following the manufacturer's

613 recommendations. Recombinant clones plasmid DNAs were purified by Qiagen Miniprep kit and

614 screening by sequencing using M13F (-20) (5'-GTA AAA CGA CGG CCA G-3') and M13R

615 (5'-CAG GAA ACA GCT ATG AC-3') primers with BigDye® v1.1 sequencing kit in house on

616 a ABI PRISM® 310 Genetic analyzer (Applied Biosystems, Foster city, CA, USA) or sequenced

617 in Macrogen Europe Inc.

618 *Phylogenetic reconstruction*

619 The coding sequence of UDP-SQ synthase proteins (coded by the *sqd*B gene) annotated in the

620 genomes of members of the phylum Cyanobacteria, Proteobacteria (α and γ-Proteobacteria),

621 Actinobacteria, Crenarchaeota, Euryarchaeota, and eukaryotic phyla Bacillariophyta (diatoms;

622 *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*), Pelagophyceae class (*Aureococcus*

623 *anophagefferens*), Phaeophyceae class (*Ectocarpus siliculosus*), and phyla Chlorophyta (green

624 algae) were obtained from genomic databases or obtained by protein blast (pBLAST) using

625 annotated UDP-SQ synthase proteins as query (NP_440474.1 *Synechocystis* sp. PCC 6803;

626 NP_195029.1 *Arabidopsis thaliana*; YP_399597.1 *Synechococcus* sp. PCC 7942; ABA78726.1

627 *Rhodobacter sphaeroides,* accession numbers). The protein identity and e-value of members of

628 the same cluster was higher or equal of 65% and 0.0, respectively. The protein identity and e-

629 value of the pBLAST between a representative member of cluster 1 (*Synechocystis* sp. PCC

630 6803) and of cluster 2 (*Rhodobacter sphaeroides*) was 44% and $2e^{-93}$. Putative UDP-SQ synthase

631 proteins in the superphylum Archaea were annotated based on pBLAST with query sequences

632 NP_440474.1 *Synechocystis* sp. PCC 6803; YP_399597.1 *Synechococcus* sp. PCC 7942;

633 ABA78726.1 *Rhodobacter sphaeroides* with protein identity ≥ 35% and e-value ≤ $1e^{-40}$. Putative

634 and annotated UDP-SQ synthase sequences were aligned by ClustalW and Muscle (Edgar, 2004)

635 in Mega5 software (Tamura *et al*., 2011) and edited manually. Phylogenetic reconstruction of

636 putative UDP-SQ synthase proteins (Figure 1) was performed by maximum likelihood in PhyML

637 v3.0 (Guindon and Gascuel, 2003) using the LG model plus gamma distribution and invariant

638 site (LG+G+I) indicated by ProtTest 2.4 (Abascal *et al*., 2005). Branch support was calculated

639 with the approximate likelihood ratio test (aLRT) and indicated on the branches (Figure 1).

640 Putative *sqd*B gene partial sequences obtained from environmental samples were translated to

641 protein by submitting them as query sequences in translated blast (xblast: Find similar proteins to

642 translated query in a protein database) and reviewed by manual annotation. DNA alignments

643 were performed by ClustalW (Thompson *et al*., 1994) and Muscle (Edgar, 2004) in the Mega5

644 software (Tamura *et al*., 2011). Phylogenetic trees including environmental sequences (DNA

645 sequences) were constructed with the Neighbor-Joining method and evolutionary distances by

646 using the Jukes-Cantor method. Bootstrap values of 1,000 replicates were also estimated.

647 *Metagenomic search*

648 UDP-SQ synthase protein sequence of *Rhodobacter sphaeroides* (ABA78726.1) has a protein

649 identity of 64–68% and e-value of zero with members of the cluster 2 and a protein identity of

650 approximate 44% and e-value 3e^{-104} to 1e^{-95} with cyanobacteria from cluster 1 (*Synechocystis*

651 sp.) by protein blast (pBLAST). In order to bias our search towards *sqd*B-coding sequences

652 related to cluster 2 we performed a pBLAST search using the UDP-SQ synthase sequence of

653 *Rhodobacter sphaeroides* (ABA78726.1) as a query sequence with an e-value of 1e^{-50} in the *find*

654 *genes* option of the Integrated Microbial Genomes (IMG) system with microbiome samples

655 (IMG/M) of the DOE's Joint Genome Institute (http://img.jgi.doe.gov). The pBLAST was

656 restricted to search homologues in marine environmental microbiome metagenomes. 130

657 sequences were recovered and aligned with the sequences included in Figure 1 with the Muscle

658 alignment application (Edgar, 2004) included in the Mega5 software (Tamura *et al*., 2011).

659 Phylogenetic reconstruction was performed by maximum likelihood in PhyML v3.0 (Guindon

660 and Gascuel, 2003) using the LG model plus gamma distribution and invariant site (LG+G+I)

661 indicated by ProtTest 2.4 (Abascal *et al*., 2005). Branch support was calculated with the
662 approximate likelihood ratio test (aLRT) and indicated on the branches (Figure S2).

663 *Data submission*

664 *sqd*B partial gene sequences were deposited in GenBank under the accession numbers: *sqd*B1A
665 (KC193329–KC193401); *sqd*B1B (KC193493–KC193567); *sqd*B2A (KC193402–KC193422);
666 *sqd*B2B (KC193423–KC193492).

667 *Extraction and analysis of Intact Polar Lipids*

668 For lipid analyses, a measured volume (ca. 20 L) of water was filtered through a pre-ashed 3 µm
669 glass fiber filter GF/F (Pall, 142 mm filter diameter). GF/F filters were stored at –40°C before
670 freeze drying. Microbial mat samples were freeze dried after being frozen at –80°C. Intact polar
671 lipids were extracted from freeze-dried biomass using a modified Bligh and Dyer technique
672 (Bligh and Dyer, 1959). A known volume of single-phase solvent mixture of methanol
673 (MeOH):dichloromethane (DCM):phosphate buffer (2:1:0.8, v/v/v) was added to the sample in a
674 glass centrifuge tube and placed in an ultrasonic bath for 10 min. The extract and residue were
675 separated by centrifuging at 2500 rpm for 5 min and the solvent mixture collected in a separate
676 flask (repeated 3 times). The DCM and phosphate buffer were added to the single-phase extract
677 to give a new ratio of MeOH:DCM:phosphate buffer (1:1:0.9, v/v/v), and to induce phase
678 separation. The extract was centrifuged at 2500 rpm for 5 min. The DCM phase was collected in
679 a round-bottom flask and the MeOH:phosphate buffer phase was washed 2 additional times with
680 DCM. The combined DCM phases were reduced under rotary vacuum and evaporated to dryness
681 under a stream of N_2. The residue was dissolved in a mixture of hexane, isopropanol, and water

682 (72:27:1, vol/vol/vol), ultrasonicated, and filtered using a regenerated cellulose 0.45 µm filter

683 (Alltech, Deerfield, IL) prior to analysis by HPLC–electrospray ionization (ESI)-MS^2.

684 *SQDG analysis*

685 SQDG analysis of the lipid extracts was performed using an Agilent 1100 LC system with

686 thermostatted auto injector and column compartment, coupled to a Quantum TSQ Ultra EM

687 triple quadrupole mass spectrometer equipped with an Ion max source with ESI probe (Thermo,

688 San Jose, CA, USA). Separation was achieved on a LiChrospher diol column (2.1 x 250 mm, 5

689 µm particles) (Grace Alltech, Deerfield, IL, USA), Chromatographic and mass spectrometric

690 conditions were according to Brandsma *et al.* (2012). Data presented is the average of three

691 different runs. SQDG identification was achieved by a mass spectrometric routine where a

692 positive ion neutral loss scan (neutral loss of 261 Da, representing the sulfoquinosovyl moiety)

693 was followed by a data dependent MS^2 experiment in which the base peak of the generated mass

694 spectrum was fragmented (collision energy 25 V; 1.5 mTorr Ar collision gas) to establish the

695 fatty acid composition of the dominant SQDGs present. For quantification, the base peak area of

696 the mass chromatogram of each SQDG was compared with that of a known quantity of a mixture

697 of SQDGs, which contained predominately SQDG with 16:1 and 18:2 acyl moieties (Lipid

698 Products, Redhill, Surrey, UK).

699

700

701

702

703 *References*

704 Abascal, F., Zardoya R., and Posada D. (2005) ProtTest: Selection of best-fit models of protein
705 evolution. *Bioinformatics* **21:** 2104–2105.

706 Bligh, E.G., and Dyer, W.J. (1959). A rapid method of total lipid extraction and purification. *Can*
707 *J Biochem Physiol* **37:** 911–917.

708 Brandsma, J., Hopmans, E.C., Phillippart, K.J.M., Vedhuis, M.J.W., Schouten, S., and Sinninghe
709 Damsté, J.S. (2012) Temporal variations in abundance and composition of intact polar lipids in
710 North Sea coastal marine water. *Biogeosciences* **9:** 1073–1084.

711 Caumette, P., Matheron, R., Raymond, N., and Relexans, J.C. (1994) Microbial mats in the
712 hypersaline ponds of Mediterranean salterns (Salins-de-Giraud, France). *FEMS Microbiol Ecol*
713 **13:** 273–286.

714 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
715 throughput. *Nucleic Acids Res.* **32:** 1792–1797.

716 Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large
717 phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

718 Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual.
719 Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.

720 Stal, L.J., van Gemerden, H., and Krumbein, W.E. (1985) Structure and development of a
721 benthic marine microbial mat. *FEMS Microbiol Ecol* **31:** 111–125.

722 Tamura, K., Peterson, P., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5:
723 Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
724 maximum parsimony methods. *Mol Biol Evol* **28:** 2731–2739.

725 Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity
726 of progressive multiple sequence alignment through sequence weighting, position-specific gap
727 penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673–4680.

728

729

730

731

732

733

734

735

736

737

738

739

740 ## Figure legends

741 Figure 1. Phylogenetic clustering of putative UDP-SQ synthase (*sqd*B gene coding protein)
742 sequences annotated in available genomes. The analysis involved 101 amino acid sequences. The
743 phylogenetic tree was inferred by maximum likelihood with the LG+G+I model of protein
744 evolution. Branch support was calculated with the approximate likelihood ratio test (aLRT) and
745 indicated on the branches (red dot ≥ 90%; blue dot [≥70% and <90%], green dot [≥50% and
746 <70%], no dot <50%). The scale bar indicates evolutionary distance of 0.5 substitutions per site.
747 Color code of the clusters: *sqd*B1A (red), *sqd*B1B (pink), *sqd*B2A (green), *sqd*B2B (blue).
748 Genera in which SQDGs have been reported (Table S1) are indicated with a black star. See
749 Figure S3 for details of aLRT of cluster 1 cyanobacteria.

750 Figure 2. Phylogenetic tree of *sqd*B gene sequences obtained by applying the *sqd*B1A cluster (A)
751 and *sqd*B1B cluster (B) primers in microbial mats and *sqd*B gene sequences of closest relatives
752 microbial species (black). *sqd*B gene fragment sequences obtained from Camargue microbial
753 mats are colored in green and the ones from the Schiermonnikoog mat in blue. Neighbor-Joining
754 was used and evolutionary distances were calculated by using the Jukes-Cantor method.
755 Bootstrap values (1,000 replicates) higher than 50 are shown next to the branches.

756 Figure 3. Phylogenetic tree of *sqd*B gene sequences obtained by applying the *sqd*B2B cluster
757 primers in environmental samples and *sqd*B gene sequences of closest relatives microbial species
758 (black). *sqd*B gene fragment sequences obtained from Camargue microbial mats are colored in
759 green, from the Schiermonnikoog mat in blue, and from the North Sea suspended particulate
760 matter in brown (Neighbor-Joining, and Jukes-Cantor method. Bootstrap values 1,000
761 replicates).

762 Figure 4. Phylogenetic tree of *sqd*B2A cluster DNA sequences amplified from the North Sea

763 suspended particulate matter and *sqd*B gene sequences of closest relatives microbial species

764 (black). Phylogeny reconstruction by the Neighbor-Joining method and evolutionary distances

765 by using the Jukes-Cantor method. Bootstrap values (1,000 replicates) higher than 50 are shown

766 next to the branches. *sqd*B gene fragment sequences recovered from the tested sample are

767 indicated in brown.

768

769 **Supplementary Figures**

770 Figure S1. Details of the SQDG biosynthetic pathway.

771 Figure S2. Phylogenetic clustering of putative UDP-SQ synthase (*sqd*B gene coding protein)

772 included in Figure 1 and the closely related metagenomic sequences (pBLAST e-value $1e^{-50}$ with

773 marine microbiome metagenomes included in IMG/M). Number code (1–18) corresponds to the

774 metagenome project as indicated in Table S2. The phylogenetic tree was inferred by maximum

775 likelihood with the LG+G+I model of protein evolution. Branch support was calculated with the

776 approximate likelihood ratio test (aLRT) and indicated on the branches if >50%. This is a

777 topographic representation and the branches are not proportional to the evolutionary distance

778 between sequences.

779 Figure S3. Details of aLRT branch support of cluster 1 cyanobacteria in Figure 1.

Table 1. Relative abundance (%)[a] of SQDGs in the microbial mats and North Sea SPM studied[b].

Molecular weight[c]	732	752	762	764	766	780	790	792	794	816	818	820	822	834	846	848	860	862	874	876
Acyl moieties[d]	**14:0 14:0**	**14:0 15:0**	**14:0 16:2**	**14:0 16:1**	**14:0 16:0**	**15:0 16:0**	**16:0 16:2**	**16:0 16:1**	**16:0 16:0**	**16:0 18:3**	**16:0 18:2**	**16:0 18:1**	**16:0 18:0**	**16:0 19:1**	**18:1 18:1**	**18:1 18:0**	**18:1 19:1**	**18:0 19:1**	**19:1 19:1**	**19:0 19:1**
North Sea SPM	40	1	3	14	16		2	10	6			1	3							
Schierm	1			2	3	1	2	21	12	10	11	21	4	1	5	3				
Camarg up					1			6	8		20	44	6	1	7	2	2			
Camarg mid	1			1				19	11	9	22	30	2		2	1				
Camarg low	1	1			1	1		3	6			11	4	3	11	8	8	37	3	1

[a] Percentages lower than 1% relative abundance are not included.

[b] North Sea SPM sampled 28th July 2010. Schierm: Schiermonnikoog mat sampled in November 2009. Camarg: Camargue mat sampled in December 2009 and sliced in three layers (upper, middle and lower section in depth, see Results section for details).

[c] SQDGs were detected as the molecular ions, the molecule with an ammonium adduct (+18)

[d] X:Y, number of carbons atoms in the acyl side chain : number of unsaturations.

Table S1. Bacterial and microalgae species in which sulfoquinovosyl diacylglycerols (SQDGs) have been reported.

Species	Taxonomy (phylum, {class}, order)	Main SQDGs*	Reference
Navicula gelida	Bacillariophyta; Bacillariophycidae; Naviculales	✓	14
Phaeodactylum tricornutum	Bacillariophyta; Bacillariophycidae; Naviculales	14:0, 16:1, 16:0, 22:6	27
Fragilariopsis curta	Bacillariophyta; Bacillariophycidae; Bacillariales	✓	14
Nitzschia medioconstricta	Bacillariophyta; Bacillariophycidae; Bacillariales	✓	14
Thalassiosira pseudonana	Bacillariophyta; Bacillariophycidae; Thalassiosirales	✓	10
Stephanodiscus sp.	Bacillariophyta; Coscinodiscophyceae; Thalassiosirales	14:0, 16:0, 16:1, 16:2, 20:5, etc	25
Cyclotella meneghiniana	Bacillariophyta; Coscinodiscophyceae; Thalassiosirales	✓	23
Skeletonema sp.	Bacillariophyta; Coscinodiscophyceae; Thalassiosirales	14:0, 16:0, 16:1, 16:3	26
Chaetoceros affinis	Bacillariophyta; Coscinodiscophyceae; Chaetocerotales	✓	21
Aureococcus anophagefferens	Pelagophyceae; Pelagomonadales	✓	24
Ectocarpus siliculosus	Phaeophyceae; Ectocarpales	16:0, 18:1, 18:3, 18:2, 14:0	13
Laminaria japonica	Phaeophyceae; Laminariales	16:0, 18:1, 18:3, 20:4, 18:0, 16:1	17
Sargassum pallidum	Phaeophyceae; Fucales	16:0, 17:0, 16:4, 18:1	17
Ahnfeltia tobuchiensis	Rhodophyta; Florideophyceae; Ahnfeltiales	16:0, 20:4, 20:5, 18:1, 18; and minor	17
Cyanidioschyzon merolae	Rhodophyta; Bangiophyceae; Cyanidiales	16:0, 18:2, 18:0, 18:1	19
Ulva fenestrate	Chlorophyta; Ulvales	16:0, 18:1, 18:3	17

Chlorella kessleri	Chlorophyta; Trebouxiophyceae; Prasiolales	✓	18
Chlamydomonas reinhardtii	Chlorophyta; Chlamydomonadales	✓	23
Phaeocystis sp.	Haptophyceae; Phaeocystales; Phaeocystaceae	16:0, 16:2, 18:1, 18:0, 18:2	4
Synechococcus sp. PCC7942	Cyanobacteria; Chroococcales	16:0, 16:1	15
Synechococcus sp. PCC7002	Cyanobacteria; Chroococcales	16:0, 18:1, 16:1	15
Prochlorococcus sp.	Cyanobacteria; Prochlorales	✓	22†
Synechocystis sp. PCC6714	Cyanobacteria; Chroococcales	16:0, 18:2, 18:1	15, 22†
Synechocystis sp. PCC6803	Cyanobacteria; Chroococcales	16:0, 18:2	15
Nostoc muscorum	Cyanobacteria; Nostocales	16:0, 18:3	15
Trichodesmium erythraeum	Cyanobacteria; Oscillatoriales	✓	21
Plectonema boryanum	Cyanobacteria; Oscillatoriales	16:0, 18:3, 18:2	15
Spirulina platensis	Cyanobacteria; Oscillatoriales	16:0, 18:2	15
Anabaena variabilis	Cyanobacteria; Nostocales;	16:0, 18:3, 18:2	15
Mastigocladus laminosus	Cyanobacteria; Stigonematales	16:0, 18:1	15
Sinorhizobium meliloti	Proteobacteria; α-Proteobacteria; Rhizobiales	16:0, 18:1, 18:0	7
Rhizobium leguminosarum	Proteobacteria; α-Proteobacteria; Rhizobiales	16:0, 18:1, 18:0	7
Rhodopseudomonas sulfoviridis	Proteobacteria; α-Proteobacteria; Rhizobiales	✓	11
Hoeflea phototrophica	Proteobacteria; α-Proteobacteria; Rhizobiales	✓	6
Rhodobacter sphaeroides	Proteobacteria; α-Proteobacteria; Rhodobacterales	16:0, 18:1, 18:0	28
Rhodobacter sulfidophilus, adriaticus, euryhalinus	Proteobacteria; α-Proteobacteria; Rhodobacterales	✓	12
Maricaulis sp.	Proteobacteria; α-Proteobacteria; Rhodobacterales	18:1, 16:0, 17:0, 19:0, 19:1	2
Oceanicaulis sp.	Proteobacteria; α-Proteobacteria; Rhodobacterales	✓	3
Roseivivax isoporae	Proteobacteria; α-Proteobacteria; Rhodobacterales	✓	8
Marinicauda pacifica	Proteobacteria; α-Proteobacteria; Rhodobacterales	✓	29
Woodsholea sp.	Proteobacteria; α-Proteobacteria; Caulobacterales	18:0; 18:1	3
Phenylobacterium sp.	Proteobacteria; α-Proteobacteria; Caulobacterales	✓	1
Rhodospirillum rubrum	Proteobacteria; α-Proteobacteria; Rhodospirillales	✓	5
Arcicella aquatica	Bacteroidetes; Cytophaga; Cytophagales;	✓	16

Alteromonas sp.	Proteobacteria; γ-Proteobacteria; Alteromonadales	✓	9
Marinococcus halophilus	Firmicutes; Bacilli; Bacillales	✓	20
Salinococcus hispanicus	Firmicutes; Bacilli; Bacillales	✓	20

*✓ positive detection of SQDGs in cultures. SQDG fatty acid composition is indicated when known in decreasing order of relative abundance. In green, genera of organisms whose *sqd*B-protein sequences in included in cluster 1 of Figure 1 phylogenetic tree. In purple, genera of organisms included in cluster 2.

†Several strains of *Prochlorococcus* sp. and *Synechococcus* sp. also tested positive for SQDG production by Van Mooy *et al*., 2006 (22).

References

(1) Abraham, W.R., Macedo, A.J., Lünsdorf, H., Fischer, R., Pawelczyk, S., Smit, J., and Vancanneyt, M. (2008) Phylogeny by a polyphasic approach of the order Caulobacterales, proposal of *Caulobacter mirabilis* sp. nov., *Phenylobacterium haematophilum* sp. nov. and *Phenylobacterium conjunctum* sp. nov., and emendation of the genus *Phenylobacterium*. Int J Syst Evol Microbiol **58:** 1939–1949.

(2) Abraham, W.R., Strömpl, C., Bennasar, A., Vancanneyt, M., Swings, J., Smit, J., and Moore, E.R.B. (2002) Phylogeny of *Maricaulis* Abraham et al., 1999 and proposal of *Maricaulis virginensis* sp. nov., *M. parjimensis* sp., nov., *M. washingtonensis* sp., nov. and *M. salignorans* sp., nov. *Int J Syst Evol Microbiol* **52:** 2191–2201.

(3) Abraham, W.R., Strömpl, C., Vancanneyt, M., Bennasar, A., Swings, J., Lünsdorf, H., Smit, J., and Moore, E.R.B. (2004) *Woodsholea maritima* gen. nov., sp. nov., a marine bacterium with low diversity of polar lipids. *Int J Syst Evol Microbiol* **54:** 1227–1234.

(4) Al-Hasan, R.H., Ali, A.M., Radwan, S.S. (1990) Lipids, and their constituent fatty acids, of *Phaeocystis* sp. from the Arabian Gulf. *Mar Biol* **105:** 9–14.

(5) Benson, A.A., Daniel, H., and Wiser, R. (1959) A sulfolipid in plants. *Proc Natl Acad Sci USA* **45:** 1582–1587.

(6) Biebl, H., Tindall, B.J., Pukall, R., Lünsdorf, H., Allgaier, M., and Wagner- Dobler, I. (2006) *Hoeflea phototrophica* sp., nov., a novel marine aerobic alphaproteobacterium that forms bacteriochlorophyll a. *Int J Syst Evol Microbiol* **56:** 821–826.

(7) Cedergren, R.A., and Hollingsworth, R.I. (1994) Occurrence of sulfoquinovosyl diacylglycerol in some members of the family Rhizobiaceae. *J Lipid Res* **35:** 1452–1461.

(8) Chen, M.H., Sheu, S.Y., Chen, C.A., Wang, J.T., and Chen, W.M. (2012) *Roseivivax isoporae* sp. nov., isolated from a reef-building coral, and emended description of the genus *Roseivivax*. *Int J Syst Evol Microbiol* **62:** 1259–1264.

(9) Chiu, H.-H., Shieh, W. Y., Lin, S. Y., Tseng, C.H., Chiang, P.-W., and Wagner-Dobler, I. (2007) Alteromonas tagae sp. Nov. and Alteromonas simiduii sp. nov., mercury-resistant bacteria isolated from a Taiwanese estuary. *Int J Syst Evol Microbiol* **57:** 1209–1216.

(10) Dyhrman ST, Jenkins BD, Rynearson TA, Saito MA, Mercier ML, Alexander H, Whitney LP, Drzewianowski A, Bulygin VV, Bertrand EM, Wu Z, Benitez-Nelson C, Heithoff A. (2012) The transcriptome and proteome of the diatom Thalassiosira pseudonana reveal a diverse phosphorus stress response. *PLoS One* **7:** e33768.

(11) Imhoff, J.F. (1984) Sulfolipids in phototrophic purple non-sulfur bacteria. In Siegenthaler PA & Eichenberger W (eds) Structure, function and metabolism of plant lipids pp. 175–178. Elsevier Science Publishers, Amsterdam.

(12) Imhoff, J.F. (1991) Polar lipids and fatty acids in the genus *Rhodobacter*. *Syst Appl Microbiol* **14:** 228–234.

(13) Makewicz, A., Gribi, C., and Eichenberger, W. (1997) Lipids of *Ectocarpus fasciculatus* (Phaeophyceae). Incorporation of [l-^{14}C]Oleate and the Role of TAG and MGDG in Lipid Metabolism. *Plant Cell Physiol* **38:** 952–962.
(14) Mock, T., and Kroon, B.M. (2002) Photosynthetic energy conversion under extreme conditions-II: the significance of lipids under light limited growth in Antarctic sea ice diatoms. *Phytochemistry* **61:** 53–60.
(15) Murata, N., Wada, H., and Gombos, Z. (1992) Modes of fatty-acid desaturation in cyanobacteria. *Plant Cell Physiol* **33:** 933–941.
(16) Nikitin, D., I., Strompl, C., Oranskaya, M.S., and Abraham W.-R. (2004) Phylogeny of the ring-forming bacterium *Arcicella aquatica* gen. nov., sp. nov. (ex Nikitin et al., 1994) from a freshwater neuston biofilm. *Int J Syst Evol Microbiol* **54:** 681–684.
(17) Sanina, N.M., Goncharova, S.N., and Kostetsky, E.Y. (2004) Fatty acid composition of individual polar lipid classes from marine macrophytes. *Phytochemistry* 65:721–730.
(18) Sato, N., Kamimura, R., Sugimoto, K., and Tsuzuki, M. (2008) Difference in SQDG Metabolism Between Green Algae and Cyanobacteria Under the Sulfur-Starved Condition. In Photosynthesis. Energy from the Sun. chapter 11. pp 795–798.
(19) Sato, N., and Moriyama, T. (2007) Genomic and Biochemical Analysis of Lipid Biosynthesis in the Unicellular Rhodophyte *Cyanidioschyzon merolae*: Lack of a Plastidic Desaturation Pathway Results in the Coupled Pathway of Galactolipid Synthesis. Euk Cell **6:** 1006–1017.
(20) Sprott, G.D., Bakouche, L., and Rajagopal, K. (2006) Identification of sulfoquinovosyl diacylglycerol as a major polar lipid in *Marinococcus halophilus* and *Salinicoccus hispanicus* and substitution with phosphatidylglycerol. *Can J Microbiol* **52:** 209–219.
(21)Van Mooy, B.A.S., Fredricks, H.F., Pedler, B.E., Dyhrman, S.T., Karl, D.M., Koblížek, M., *et al*. (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* **458:** 69–72.
(22) Van Mooy, B.A.S., Rocap, G., Fredricks, H.F., Evans, C.T., and Devol, A.H. (2006) Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc Natl Acad Sci USA* **103:** 8607–8612.
(23) Vieler, A., Wilhelm, C., Goss, R., Sussb, R., and Schillerb, J. (2007) The lipid composition of the unicellular green alga *Chlamydomonas reinhardtii* and the diatom *Cyclotella meneghiniana* investigated by MALDI-TOF MS and TLC. *Chem Phys Lipids* **150:** 143–155.
(24) Wurch, L.L., Bertrand, E.M., Saito, M.A., Van Mooy, B.A.S., and Dyhrman, S. (2011) proteome changes driven by phosphorous deficiency and recovery in the brown tide-forming alga Aureococcus anophagefferens. *Plos One* **6:** e28949.
(25) Xu, J., Chen, D., Yan, X., Chen, J., Zhou, C. (2012) Global characterization of the photosynthetic glycerolipids from a marine diatom Stephanodiscus sp. by ultra performance liquid chromatography coupled with electrospray ionization-quadrupole-time of flight mass spectrometry. *Anal chimica Acta* **663:** 60–68.
(26) Yang, X., Chen, D., Xu, J., and Zhou, C. (2011) Profiles of photosynthetic glycerolipids in three strains of Skeletonema determined by UPLC-Q-TOF-MS. *J Appl Phycol* **23:** 271–282.
(27) Yongmanitchai, W., and Ward, O.P. (1993) Positional distribution of fatty acids, and molecular species of polar lipids in the diatom *Phaeodactylum tricornutum*. *J Gen Microbiol* **139:** 465–472.
(28) Zhang, X, Fhaner, C.J., Ferguson-Miller, S.M. and Reid, G.E. (2012a) Evaluation of ion activation strategies and mechanisms for the gas-phase fragmentation of sulfoquinovosyldiacylglycerol lipids from Rhodobacter sphaeroides. *Int J Mass Spectrom* **316–318**: 100–107.
(29) Zhang XY, Li GW, Wang CS, Zhang YJ, Xu XW, Li H, Liu A, Liu C, Xie BB, Qin QL, Xu Z, Chen XL, Zhou BC, Zhang YZ. (2012b) *Marinicauda pacifica* gen. nov., sp. nov., a prosthecate alphaproteobacterium of the family Hyphomonadaceae isolated from deep seawater of the Pacific. *Int J Syst Evol Microbiol* doi:10.1099/ijs.0.046656-0.

Table S2. Metagenome projects from Marine microbiomes in IMG/M used in this study.

Descriptor in tree	IMG ID number	Project	Details
ESNP OMZ	2189573016	Marine microbial communities from the Eastern Subtropical North Pacific Ocean, Expanding Oxygen minimum zones	Saanich Inlet 60 08/10/11 200m (Saanich Inlet 60 08/10/11 200m, April 2012 Assem)
Trichod bloom	2156126005	Marine Trichodesmium cyanobacterial communities from the Bermuda Atlantic Time-Series	Marine Trichodesmium cyanobacterial communities from the North Pacific Subtropical Gyre outside Oahu, HI, sample from new species B colonies
Baltic Sea	3300000129	Marine microbial communities from chronically polluted sediments in four geographic locations	Baltic Sea site KBA sample SWE 12_21m (Baltic Sea site KBA sample SWE 12_21m, Oct 2011 Assem)
Delaware Coast	3300000115	Marine microbial communities from Delaware Coast	Marine microbial communities from Delaware Coast, sample from Delaware MO Early Summer May 2010 (Delaware MO Early Summer May 2010, Feb 2012 assem)
Lyng mats	3300000354	Hypersaline water microbial communities from Lyngbya mats, Guerrero Negro, Mexico and Elkhorn Slough mats, California, USA	Elkhorn Slough mat CD2A Metagenome (Elkhorn Slough mat CD2A, July 2012 Assem)
Antartic intertidal	3300000129	Marine microbial communities from chronically polluted sediments in four geographic locations	King George Island site S1 sample

Figure 1

Cluster 3
Halorubrum lacusprofundi
Haladaptatus paucihalophilus
Haloarcula marismortui
Haloarcula hispanica
Haloquadratum walsbyi
Halopiger xanaduensis
Haloterrigena turkmenica
Halorhabdus tiamatea
Halorhabdus utahensis
Thermoplasma acidophilum
Thermoplasma volcanicum
Picrophilus torridus
Ferroplasma acidarmanus
Caldivirga maquilingensis
Vulcanisaeta distributa
Vulcanisaeta moutnovskia
Sulfolobus islandicus
Sulfolobus tokodaii
Sulfolobus acidocaldarius
Acidianus hospitalis
Metallosphaera yellowstonensis
Metallosphaera cuprina
Metallosphaera sedula
Micrarchaeum acidiphilum

Cluster 1
Arabidopsis thaliana
Ostreococcus tauri
Ostreococcus lucimarinus
Micromonas sp. RCC299
Micromonas pusilla
Chlamydomonas reinhardtii
Volvox carteri f. nagariensis
Chlorella variabilis
Cyanobacterium UCYN-A
Crocosphaera watsonii
Spirulina sp. EEW8
Cyanothece sp. PCC7424
Microcoleus chthonoplastes
Arthrospira maxima
Lyngbya sp. PCC8106
Oscillatoria sp. PCC6506
Trichodesmium erythraeum
Microcystis aeruginosa
Synechocystis sp. PCC6803
Anabaena variabilis
Fischerella sp. JSC-11
Nodularia spumigena
Rhaphidiopsis brookii
Cylindrospermopsis raciborskii
Nostoc punctiforme
sqdB1A
sqdB1B

Cluster 2
0.5
Pseudovibrio sp. JE062
Oceanicaulis alexandrii
Thiorhodococcus drewsii
Nitrosococcus oceani
Thioalkalivibrio sulfidophilus
Desulfarculus baarsii
Lyngbya majuscula
Acaryochloris marina
Parvularcula bermudensis
Alteromonas sp. S89
Phenylobacterium zucineum
Brevundimonas subvibriodes
Aureococcus anophagefferens
Thalassiosira pseudonana
Phaeodactylum tricornutum
Ectocarpus siliculosus
Cyanidioschyzon merolae
Catenulispora acidiphila
Oceanicola granulosus
Sagittula stellata
Paracoccus denitrificans
Rhodobacter sphaeroides ATCC17025
Rhodobacter sphaeroides ATCC17029
Rhodobacter sphaeroides KD131
Rhodobacter sphaeroides 2.4.1
Labrenzia alexandrii
Labrenzia aggregata
Sinorhizobium meliloti
Rhizobium leguminosarum
Agrobacterium radiobacter
Hoeflea phototrophica
Chelativorans sp. BNC1
*sqd*B2B
Prochlorococcus marinus str. MIT 9312
Prochlorococcus marinus str. MIT 9202
Prochlorococcus marinus str. NATL1A
Synechococcus sp. RS9916
Synechococcus sp. WH 7803
Synechococcus sp. WH 8102
Synechococcus sp. CC9605
Synechococcus sp. WH 8109
Synechococcus sp. CC9311
Synechococcus sp. WH 8016
Cyanobium sp. PCC7001
Synechococcus elongatus
*sqd*B2A
Rhodospirillum centenum
Serinicoccus profundis
Corynebacterium ammoniagenes
Citricoccus sp. CH26A
Propionibacterium freudenreichii
Citreicella sp. SE45
Pelagibacterium halotolerans
Ketogulonigenium vulgarum
Nocardia cyriacigeorgica
Rhodococcus opacus

Figure 2

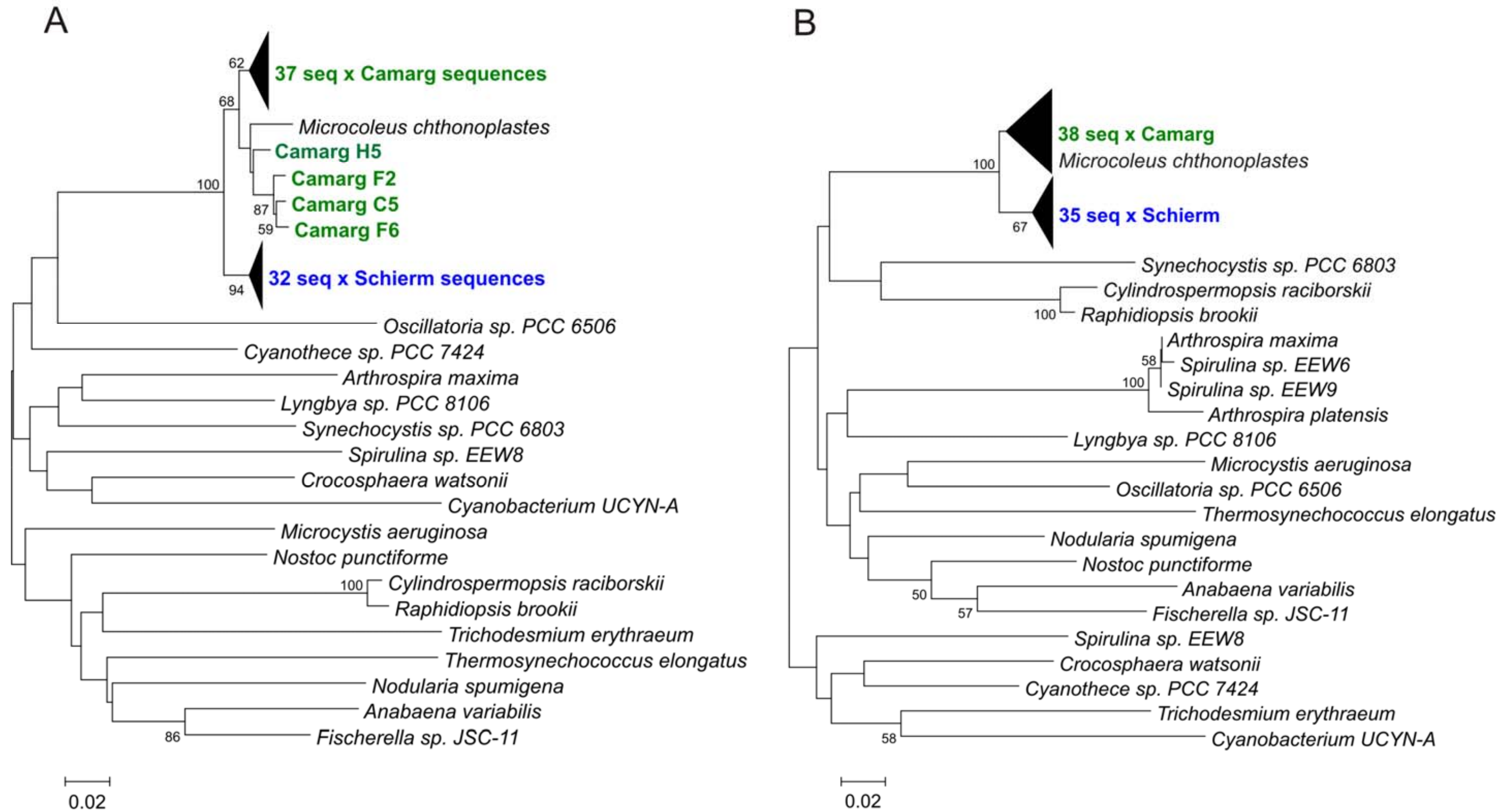


Figure 3

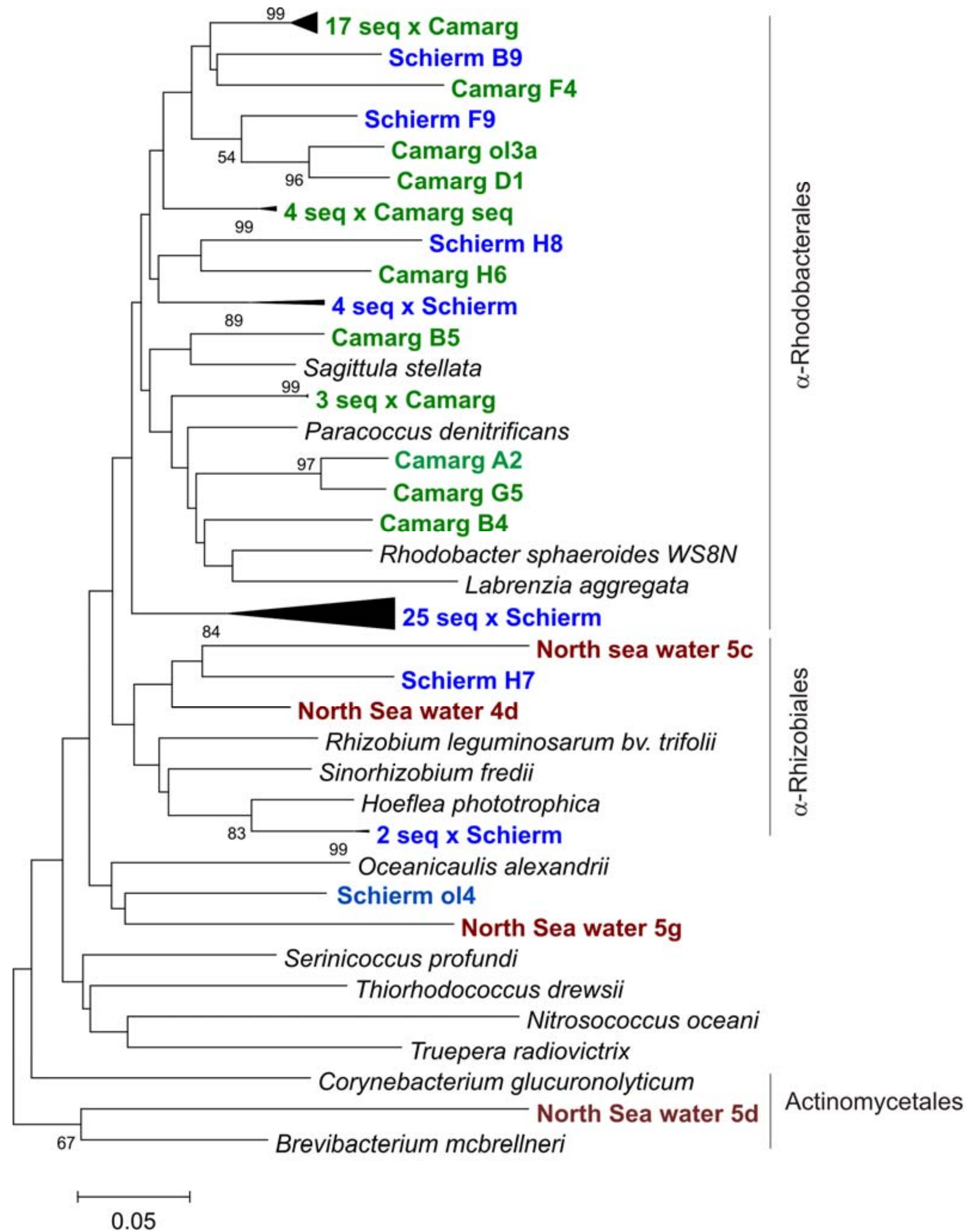


Figure 4

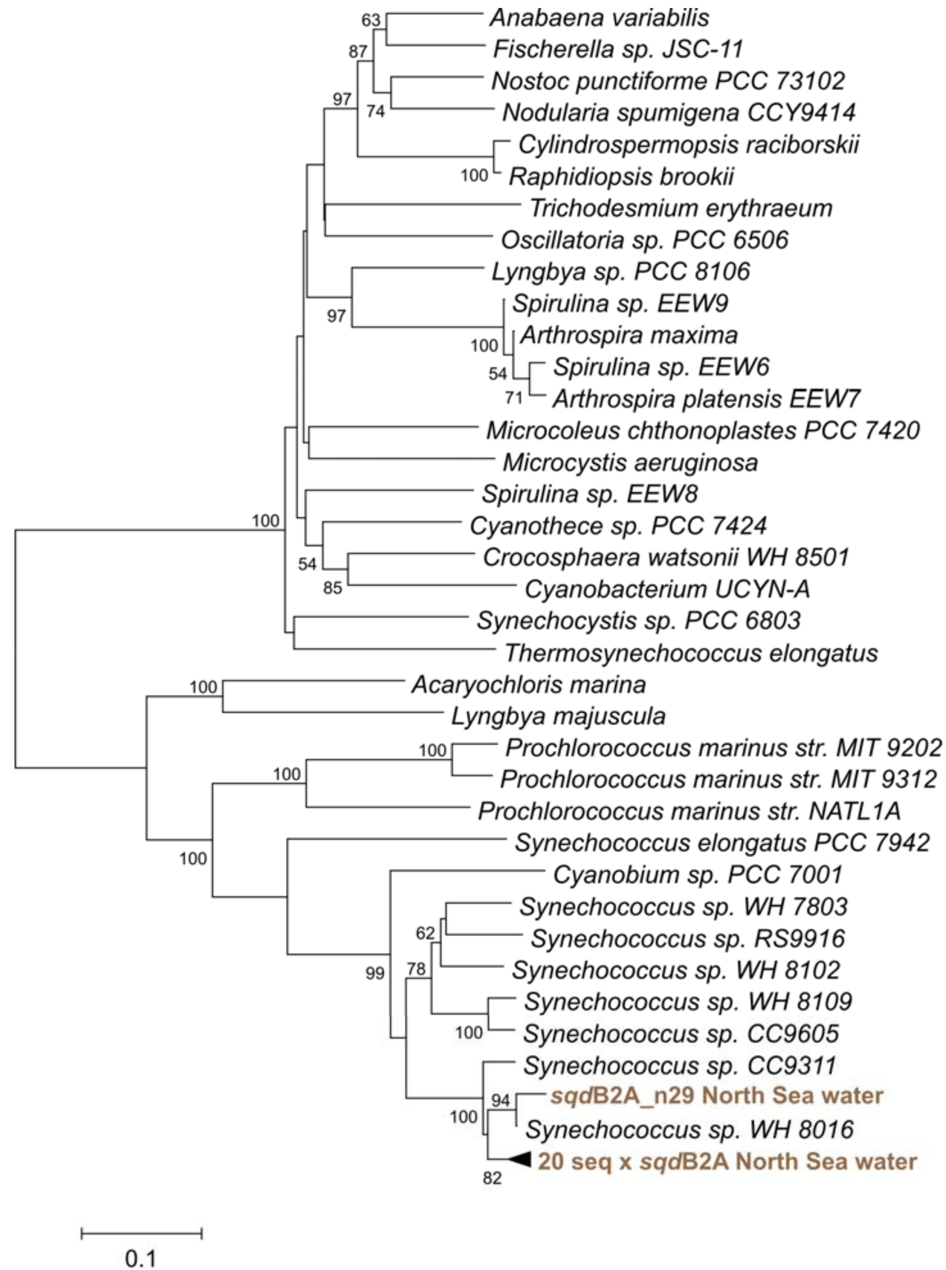


Figure S1. Sulfoquinovosyldiacylglycerol synthetic pathway (*The exact mechanism of the SQDG synthase step and the genes involved is unknown).

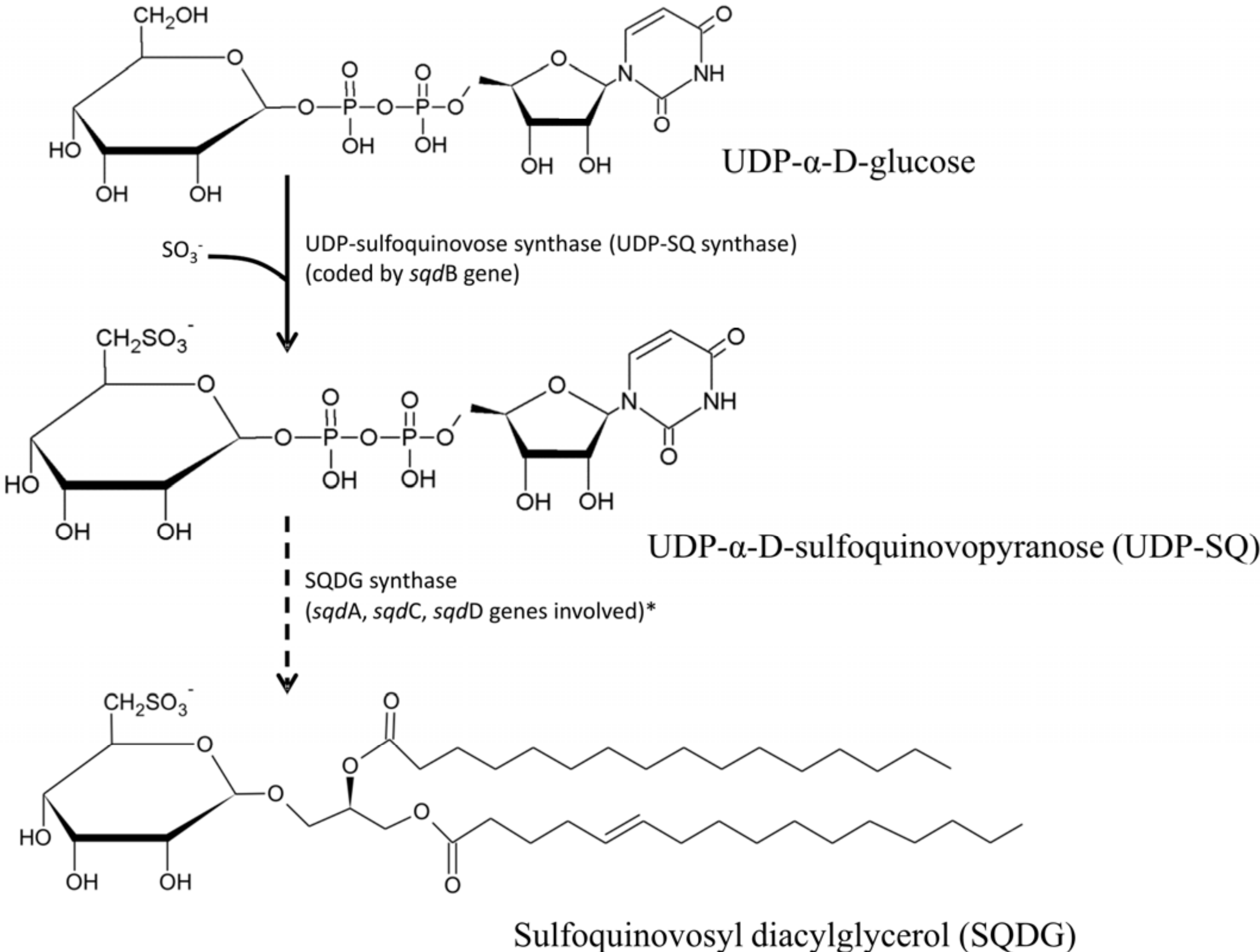


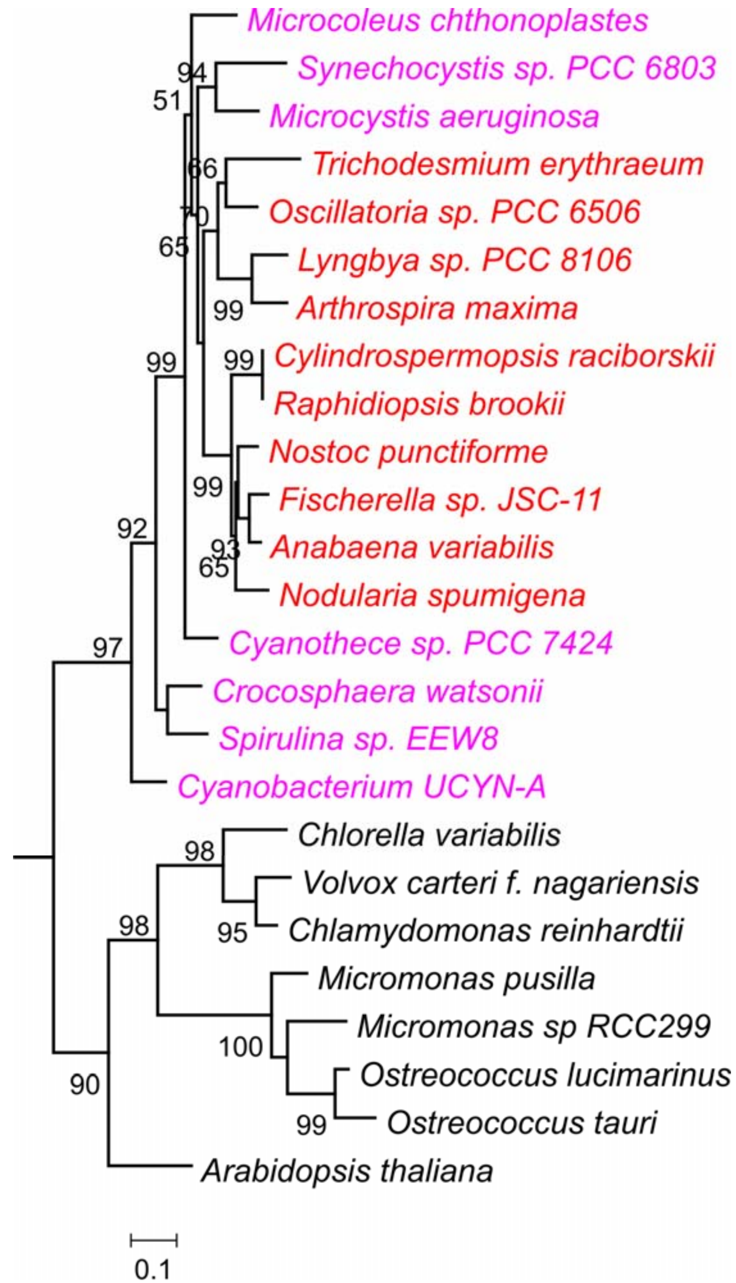
Figure S2

Cluster 1

Cluster 2

Oscillatoria sp. PCC6506
Trichodesmium erythraeum
Antarctic intertidal
Cluster 3
Cluster 3
Trichodesmium bloom
Arthrospira maxima
Lyngbya sp. PCC8106
Lyngbya mats
Microcoleus chthonoplastes
Lyngbya mats
Microcystis aeruginosa
Synechocystis sp. PCC6803
Anabaena variabilis
Fischerella sp. JSC-11
Nodularia spumigea
Cylindrospermopsis raciborskii
Nostoc punctiforme
Cyanothece sp. PCC7424
Spirulina sp. EEW8
Crocosphaera watsonii
Cyanobacterium UCYN-A
Arabidopsis thaliana
Chlamydomonas reinhardtii
Volvox carteri
Chlorella variabilis
Ostreococcus tauri
Ostreococcus lucimarinus
ESNP OMZ, Delaware coast
Micromonas pusilla
ESNP OMZ
Micromonas sp. RCC299
Delaware coast
Phaeodactylum tricornutum
Thalassiosira pseudonana
Aureococcus anophagefferens
Ectocarpus siliculosus
Cyanidioschyzon merolae
Catenulispora acidiphila
Wetland Island
Labrenzia alexandrii
Labrenzia aggregata
Hoeflea phototrophica
Agrobacterium radiobacter
Rhizobium leguminosarum
Sinorhizobium meliloti
Chelativorans sp. BNC1
Rhodobacter sphaeroides 2.4.1
Rhodobacter sphaeroides KD131
Rhodobacter sphaeroides ATCC 17029
Rhodobacter sphaeroides ATCC 17025
Paracoccus denitrificans PD1222
Sagittula stellata
Oceanicola granulosus
Salt marsh mat
Propionibacterium freudenreichii
Serinicoccus profundis
Citricoccus sp. CH26A
Corynebacterium ammoniagenes
Ketogulonigenium vulgarum
Pelagibacterium halotolerans
Citreicella sp. SE45
Whale Fall
Delaware coast
Rhodococcus opacus
Nocardia cyriacigeorgica
ESNP OMZ
Synechococcus elongatus
Cyanobium sp. PCC 7001
Synechococcus sp. WH7803
Synechococcus sp. RS9916
Synechococcus sp. WH8016
Synechococcus sp. CC9311
Synechococcus sp. WH8102
Synechococcus sp. WH8109
Synechococcus sp. CC9605
Hydrothermal
ESNP OMZ
Prochlorococcus marinus str. NATL1A
Prochlorococcus marinus str. 9202
Prochlorococcus marinus str. MIT 9312
Baltic sea
Alteromonas sp. S89
ESNP OMZ
Acaryochloris marina
Lyngbya majuscula
Lyng mats
ESNP OMZ, Baltic sea
Svalbard sed, Arctic coast, St Barbara Basin
ESNP OMZ, Baltic sea, Delaware coast, Annelida
ESNP OMZ
Thioalkalivibrio sulfodophilus
Nitrosococcus oceani
Thiorhodococcus drewsii
ESNP OMZ
Whale Fall
Parvularcula bermudensis
Antarctic intertidal, St Barbara Basin
ESNP OMZ, Delaware coast,
Brevundimonas subvibrioides
Phenylobacterium zucineum
Arctic coast
Antarctic intertidal
Oceanicaulis alexandrii
Whale Fall
Pseudovibrio sp. JE062
Rhodospirillum centenum
Desulfarculus baarsii

Figure S3



Salt Marsh mat	3300000504	Marine photosynthetic community that grows at 940nm	Microbial Communities from Little Sippewissett Salt Marsh, Woods Hole, MA that are anoxygenic and photosynthetic, Marine photosynthetic community that grows at 940nm (Marine_940nm_cellulose)
Svalbard sediment	3300000130	Svalbard Archipelago station 2 sample NOR 15_50m (Svalbard Archipelago station 2 sample NOR 15_50m, Dec 2011 Assem)	Svalbard Archipelago station 2 sample NOR 15_50m (Svalbard Archipelago station 2 sample NOR 15_50m, Dec 2011 Assem)
Wetland island	3300000313	Soil microbial communities from Twitchell Island in the Sacramento Delta	Wetland microbial communities from Twitchell Island in the Sacramento Delta, sample from surface sediment Feb2011 Site B1 Cattail (Wetland Surface Sediment Feb2011 Site B1 Cattail, Assem Ctgs Sep 2011 assem)
Artic coast	2140918005	Coastal water and sediment microbial communities from Arctic	Sediment microbial communities from Arctic Ocean, off the coast from Alaska, sample from high methane PC12-225-485cm (High methane PC12-225-485cm Jan 2011 assembly)
Whale fall	2001200003	Fossil microbial community from Whale Fall, Santa Cruz Basin of the Pacific Ocean	Fossil microbial community from Whale Fall at Santa Cruz Basin of the Pacific Ocean Sample #1 Halophile 1674 m below sea level
St Barbara Basin	2084038021	Methane oxidizing archaeal communities in the Santa Barbara Basin	Marine sediment archaeal communities from Santa Barbara Basin, CA, that are methane-oxidizing, sample 0-3 cm (ANME Sed A12 0-3 cm)
Sediment intertidal	3300000242	Tierra del Fuego site OR sample ARG 05_12.3m (Tierra del Fuego site OR sample ARG 05_12.3m, Oct 2011 Assem)	Tierra del fuego, Ushuaia, Argentina
Annelida	2004178002	Olavius algarvensis microbiome from Mediterranean sea	Olavius algarvensis endosymbiont metagenome Delta4
Hydrothermal	2061766003	Guaymas Basin hydrothermal plume	Hydrothermal vent microbial communities from Guaymas and Carmen Basins, Gulf of California

Table S3. Primers for *sqd*B gene sequences of the phylum Cyanobacteria (clusters *sqd*B1A–B, and *sqd*B2A), and α-Rhodobacterales and α-Rhizobiales groups (cluster *sqd*B2B).

Cluster	Primers*	Genera
*sqd*B1A	*sqd*B1A_559F (5’-CACGAYAGYCATAAYATYCA-3’) *sqd*B1A_877R (5’-ATTGRTTAAAVACNCGGAATT-3’) **HDSHNIH/QFRVFNQY†** (187–293, *Synechocystis* sp. PCC6803)	***Synechocystis* sp.** *Crocosphaera* sp. ***Cyanothece* sp.** *Microcystis* sp. *Cyanobacterium UCYN-A* ***Microcoleus* sp.**
*sqd*B1B	*sqd*B1B_316F (5’-TCNATGATWGAYCGNGAACA-3’) *sqd*B1B_533R (5’-CCRGGYTGYTTBGGATAVGG-3’) **SMIDREH/PYPKQP** (105–176, *Anabaena variabilis*)	***Anabaena* sp.** *Arthospira* sp. *Nostoc* sp. ***Spirulina* sp.** *Fischerella* sp. *Nodularia* sp. *Raphidiopsis* sp. *Trichodesmium* sp. *Oscillatoria* sp.
*sqd*B2A	*sqd*B2A_28F (5’-GGYTTCTGCGGVTGGCCYTG-3’) *sqd*B2A_591R (5’-ATCSAGCGTCTTSGTCATGT-3’) **GFCGWPC/HMTKTLD** (10–197, *Synechococcus* sp.)	***Synechococcus* sp.** ***Prochlorococcus* sp.** ***Cyanobium* sp.**
*sqd*B2B	*sqd*B2B_693F (5’-ATCARCCGSTTCGAYTAYGA-3’) *sqd*B2B_905R (5’-TCBGTCATCTGGTTGAASA-3’) **INRFDYD/IFNQMTE** (232–302, *Rhodobacter sphaeroides*)	***Rhodobacter* sp.** ***Paracoccus* sp.** *Sagitulla* sp. *Oceanicola* sp. *Labrenzia* sp. *Chelativorans* sp. *Hoeflea* sp. *Agrobacterium* sp. *Sinorhizobium* sp. *Rhizobium* sp.

*Degenerancies: B (T/C/G); Y (T/C); V (A/G/C); R (A/G); S (C/G); W (A/T); N (A/T/C/G). †Amino acid moieties corresponding to the primer sequences and their amino acid position in the specified *sqd*B gene coding protein. In red the pure bacterial cultures used as a positive control in the PCR reaction. Species in black were not tested.