Novel dimethylsulfoniopropionate biosynthesis enzymes in diverse marine bacteria, cyanobacteria and abundant algae

Jinyan Wang
Ocean University of China https://orcid.org/0000-0002-5358-3698
Shun Zhou
Ocean University of China https://orcid.org/0000-0002-6937-3513
Andrew Curson
University of East Anglia
Ana Vieira
i3S - Instituto de Investigação e Inovação em Saúde
Keanu Walsham
University of East Anglia
Serena Monaco
University of East Anglia
Chun-Yang Li
Ocean University of China https://orcid.org/0000-0002-1151-4897
Peter Paolo Rivera
https://orcid.org/0000-0002-8587-9178
Xiao-Di Wang
Ocean University of China
Libby Hanwell
University of East Anglia
Xiao-Yu Zhu
University of East Anglia
Pedro Leão
Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR)
David J. Lea-Smith
University of East Anglia https://orcid.org/0000-0003-2463-406X
Yuzhong Zhang
Shandong University https://orcid.org/0000-0002-2017-1005
Xiaohua Zhang
Ocean University of China https://orcid.org/0000-0002-7428-7775
Jonathan Todd (jonathan.todd@uea.ac.uk)

University of East Anglia https://orcid.org/0000-0003-0777-9312

Article

Keywords: DMSP synthesis genes, DMSP-producing algae, bacteria and cyanobacteria, osmoprotection

Posted Date: March 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2678769/v1

Additional Declarations:
There is **NO** Competing Interest.

Table 1 is available in the Supplementary Files section.

Novel dimethylsulfoniopropionate biosynthesis enzymes in diverse marine bacteria, cyanobacteria and abundant algae

Jinyan Wang[1,2], Shun Zhou[1,2], Andrew R. J. Curson[2], Ana Vieira[3], Keanu S. Walsham[2], Serena Monaco[4], Chun-Yang Li[1,5], Peter Paolo L. Rivera[2], Xiao-Di Wang[1,5], Libby Hanwell[2], Xiao Yu Zhu[2], Pedro N. Leão[3], David J. Lea-Smith[2], Yu-Zhong Zhang[1,5], Xiao-Hua Zhang[1,6*] and Jonathan D. Todd[1,2*]

[1] Frontiers Science Center for Deep Ocean Multispheres and Earth System, and College of Marine Life Sciences, Ocean University of China, Qingdao, China.

[2] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, UK.

[3] Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Portugal.

[4] School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich, UK.

[5] State Key Lab of Microbial Technology, Marine Biotechnology Research Center, Shandong University, Qingdao, China.

[6] Laboratory for Marine Ecology and Environmental Science, Laoshan Laboratory, Qingdao, China.

***Correspondence:**

Jonathan D. Todd, Email: jonathan.todd@uea.ac.uk, Tel. +44-1603-592264

Xiao-Hua Zhang, Email: xhzhang@ouc.edu.cn, Tel: +86-532-82032767

Running title: The novel DMSP biosynthesis genes *dsyGD, dsyG* and *DSYE*

Keywords: DMSP synthesis genes, DMSP-producing algae, bacteria and cyanobacteria, osmoprotection

Abstract:

Dimethylsulfoniopropionate (DMSP) is an abundant marine organosulfur compound[1] with roles in stress protection[2, 3], chemotaxis[4], nutrient and sulfur cycling[5] and, potentially, climate regulation[6, 7]. Marine algae and bacteria are considered significant DMSP producers, but many diverse representatives lack known DMSP synthesis genes/enzymes[8, 9]. Here, new DMSP biosynthesis enzymes were identified that considerably increase the number and diversity of potential DMSP-producing organisms, inferring new and significant global DMSP producers. A novel bifunctional DMSP biosynthesis enzyme, DsyGD, identified in the rhizobacterium *Gynuella sunshinyii*, produces DMSP at levels higher than any other bacterium from methylthiohydroxybutyrate (MTHB) via an N-terminal MTHB *S*-methyltransferase domain (termed DsyG) and a C-terminal dimethylsulfoniohydroxybutyrate (DMSHB) decarboxylase domain (termed DsyD, which is the first reported enzyme with this activity). DsyGD is also found in some filamentous cyanobacteria, not previously known to produce DMSP. Regulation of DMSP production and *dsyGD* transcription was consistent with their role in osmoprotection. Indeed, cloned *dsyGD* conferred osmotolerance to bacteria deficient in osmolyte production, something not previously demonstrated for any known DMSP synthesis gene, and which could be exploited for biotechnology e.g., engineering salt tolerance. DsyGD characterisation led to identification of phylogenetically distinct DsyG-like proteins, termed DSYE, with MTHB *S*-methyltransferase activity, in diverse and environmentally abundant *Chlorophyta, Chlorachniophyta, Ochraphyta, Haptophyta* and *Bacillariophyta* algae. These algae comprise a mix of low, high and previously unknown DMSP producers[10]. Algae containing *DSYE*, particularly bloom-forming *Pelagophyceae* species, which we showed to accumulate medium-high intracellular DMSP levels, were globally more abundant DMSP producers than *Haptophyta*, *Dinophyta* and *Bacillariophyta* with *DSYB* and/or *TpMMT*. This highlights the potential importance of *Pelagophyceae* and other DSYE containing algae in global DMSP production and sulfur cycling.

Main

Isolated from the rhizosphere of the saltmarsh plant *Carex scabrifolia*, *Gynuella sunshinyii* YC6258 is an unusual *Gammaproteobacterium* with antifungal activity[11] and the potential for diverse natural product synthesis[12, 13]. *G. sunshinyii* is also relatively abundant (~0.21%) in the rhizosphere of *Spartina alterniflora*[14], an environment known to be rich in DMSP, due to the high amounts this cordgrass produces[15-18], and microbial DMSP cycling[19-22]. Given *G. sunshinyii* contains a BCCT family transporter likely to import DMSP (with 40% amino acid identity to the *Pseudomonas* sp. J465 DMSP transporter DddT[23], Supplementary Table 1), we proposed that *C. scabrifolia* (not previously suspected to produce DMSP) and/or *Spartina* produced DMSP and fed this to *G. sunshinyii* in return for favorable bacterial traits/metabolites, e.g., activity against major fungal pathogens[12, 24-26]. Indeed, DMSP was detected in *C. scabrifolia* leaves (6.9±0.49 nmol DMSP g^{-1} fresh weight) and its rhizosphere (5.5±0.49 nmol DMSP g^{-1}) (Supplementary Fig. 1). However, *G. sunshinyii* could not use DMSP as a carbon source, catabolise DMSP, nor could it liberate DMS or methanethiol from DMSP, consistent with its genome lacking all known DMSP lyase genes[27-35] and the DMSP demethylation gene *dmd*A[36]. Interestingly, *G. sunshinyii* was found to produce DMSP when grown in the absence of added organosulfur compounds and at levels (89.4 ± 9.43 pmol µg protein^{-1} h^{-1}) higher than any other reported bacterium. This finding quashed the hypothesis that *C. scabrifolia* was feeding *G. sunshinyii* DMSP and we moved to investigate how this *Gammaproteobacterium* produced DMSP, since its genome lacked all known DMSP synthesis genes.

To predict the DMSP synthetic pathway, *G. sunshinyii* was grown with intermediates from the three known DMSP synthesis pathways (Fig. 1a) and its DMSP levels were monitored. Intermediates of the Met transamination pathway[37, 38], 4-methylthio-2-hydroxybutyrate (MTHB) and 4-dimethylsulfonio-2-hydroxybutyrate (DMSHB), significantly enhanced *G. sunshinyii* DMSP production (2 to 30 fold) compared to controls with no added intermediates (Fig. 1b). In contrast, addition of the

Met methylation and decarboxylation pathway intermediates had no significant effect on *G. sunshinyii* DMSP production (Fig. 1b). Furthermore, *G. sunshinyii* cell extracts displayed *in vitro* MTHB *S*-methylation and DMSHB decarboxylation activities (6.1 and 68.1 pmol DMSP μg protein^{-1} h^{-1}, respectively). These data indicated that *G. sunshinyii* likely synthesises DMSP via the Met transamination pathway (Fig. 1a).

To identify *G. sunshinyii* DMSP synthesis genes, a genomic library of this Gammaproteobacterium was constructed and screened in *Rhizobium leguminosarum* J391 for MTHB *S*-methyltransferase (MSM) activity. Since *R. leguminosarum* has DMSHB decarboxylase (DDC) activity, any DMSHB produced through MSM activity would lead to DMSP production[39]. One clone from 3,000 screened (termed pBIO2208) conferred MSM activity and thus, DMSP production (0.58 ± 0.02 pmol DMSP μg protein^{-1} h^{-1}). pBIO2208 conferred MSM activity (unlike *dsyB/DSYB* clones[9, 39]) and intriguingly, also DMSHB decarboxylase activity to *E. coli* (0.74 ± 0.08 pmol DMSP μg protein^{-1} h^{-1}), implying that *G. sunshinyii* contains a gene cluster for DMSP synthesis. The ~30 kb insert in pBIO2208 was sequenced and found to contain only one methyltransferase gene, termed *dsyGD*, adjacent to a reductase gene, predicted to encode a 4-methylthio-2-oxobutyrate (MTOB) reductase enzyme (Fig. 1a and Fig. 1c). DsyGD is a 494 amino acid protein with two domains. The N-terminal methyltransferase (PF08241.15) domain, termed DsyG, phylogenetically clusters away from all known *S*-methyltransferases involved in DMSP synthesis (Fig. 2) and had < 33% amino acid identity to TpMMT from *Thalassiosira pseudonana*[40]. The DsyGD C-terminus contained a predicted ureidoglycolate lyase domain (PF04115.15), termed DsyD, likely acting as a DMSHB decarboxylase.

Cloned *G. sunshinyii dsyGD* conferred *in vivo* MSM (177.42 ± 3.23 pmol DMSHB μg protein^{-1} h^{-1}) and DDC activity (13.81 ± 0.97 pmol DMSP μg protein^{-1} h^{-1}) when expressed in *E. coli*, and fully restored DMSP production to a *Labrenzia aggregata* LZB033 *dsyB*$^{-}$ mutant [39] which does not produce DMSP (Table 1). Furthermore, purified recombinant DsyGD (Supplementary Fig. 2a) exhibited *in vitro S*-adenosyl-Met (SAM)-dependent MSM and DDC activity with an optimal temperature of 25°C (Supplementary Fig. 3a) and pH of 7.0 (Supplementary Fig. 3b). DsyGD had K_m values of 22.6 μM and 96.6 μM, and K_{cat} values of 5.0 s^{-1} and 5.4 s^{-1}, for MTHB and SAM, respectively (Supplementary Fig. 3c, d). DsyGD also had a K_m value of 0.91 mM and a

K_{cat} value of 2.09 s^{-1} for DMSHB at pH 7 and 25°C (Supplementary Fig. 3e). The individual *G. sunshinyii* DsyG (methyltransferase) and DsyD (decarboxylase) domains and the predicted MTOB reductase enzyme did not have the expected activities on MTHB, DMSHB or MTOB, respectively (Fig. 1a), when expressed in *E. coli* or as purified proteins under the conditions tested here. It is possible these specific *G. sunshinyii* DsyG and DsyD domains evolved to require each other. Unfortunately, we could not transform or conjugate plasmids into *G. sunshinyii*, preventing gene mutagenesis. These data support DsyGD as the first bifunctional DMSP synthesis enzyme with two DMSP synthesis specific and sequential enzyme activities in the Met transamination pathway[38], and the first reported enzyme with DMSHB decarboxylase activity. We hypothesise that fusion of these *S*-methyltransferase and decarboxylase domains into one translational unit allows coordinated expression of enzyme activities to produce DMSP from MTHB, possibly explaining why *G. sunshinyii* accumulated the highest DMSP concentration of any reported DMSP-producing bacteria.

DsyGD was not predicted from any other sequenced genomes, MAGs or transcriptomes at high amino acid identity, thus, its origin is unclear. However, DsyGD proteins with MSM and DDC activity (see Table 1) and only 46% amino acid identity to ^{Gs}DsyGD (Supplementary Table 2) were encoded from the MAGs of two filamentous cyanobacteria (*Symploca* sp. SIO3E6 and *Oscillatoria* sp. SIO1A7) of the *Oscillatoriales* order (Fig. 1c, Fig. 2; Supplementary Table 2). Interestingly, a single-domain DsyG with MSM activity and ~50% amino acid identity to this domain of ^{Gs}DsyGD was identified in an *Oscillatoriales* isolate *Zarconia navalis* LEGE 11467[41, 42] (Fig. 1c, Fig. 2; Supplementary Table 2). It is unknown why this single-domain ^{Zn}DsyG had MSM activity but the truncated ^{Gs}DsyG did not. *Z. navalis* LEGE 11467, isolated from a subtidal epilithic marine sample [41, 42], lacked *dsyD* but did produce DMSP (0.385 ± 0.069 nmol DMSP mg fresh weight). These DMSP levels were much lower than for *G. sunshinyii* supporting our hypothesis that the double domain DsyGD enzyme was responsible for higher level production. Unlike DsyG, a single domain DsyD was not identified from any sequenced genomes, MAGs or transcriptomes. The proteins most similar to the ^{Gs}DsyD domain were from *Prymnesium parvum* Texoma1 and *Alexandrium monilatum* CCMP3105, showed only 29.8% amino acid identity to DsyD, and lacked DDC activity (Table 1, Supplementary Table 2, Supplementary Table 3 and Supplementary Fig. 4).

After ^{Zn}DsyG, the next most homologous proteins to ^{Gs}DsyG, sharing ~39% amino acid identity to the ^{Gs}DsyG domain (Supplementary Table 2), were from a *Planctomycetales* bacterium MAG and the red alga *Porphyra umbilicalis*. These DsyG-like proteins either phylogenetically clustered more closely to TpMMT than ^{Gs}DsyG (*P. umbilicalis*) or were positioned in-between TpMMT and ^{Gs}DsyG (*Planctomycetales* bacterium) (Fig. 2). Note, the *P. umbilicalis* protein, like ^{Gs}DsyGD, contained two domains, but its C-terminal domain belonged to the aspartate decarboxylase protein family (pfam02261), which seemed a good candidate DMSHB decarboxylase as a DsyD isoform enzyme. Despite these facts both the *Planctomycetales* and the *P. umbilicalis* DsyG-like proteins lacked MSM and DDC activity (Table 1). Note, there were no proteins with strong homology (>38% identity) to DsyG predicted from the genomes and/or transcriptomes of eukaryotic algae. Overall, these data support *dsyGD*/*dsyG* as being a reporter gene for DMSP synthesis in bacteria and filamentous cyanobacteria, not previously suspected to produce DMSP, and that *Z. navalis* likely contained a novel DMSHB decarboxylase. The data also provide further warning of the need for careful functional analysis of DMSP synthesis genes/enzymes before predicting DMSP synthesis in organisms based on their presence. This is particularly relevant for TpMMT which has only been ratified from *Thalassiosira pseudonana*[40].

To infer the role of DMSP in *G. sunshinyii* and *Z. navalis* with DsyGD/DsyG we studied the production of DMSP and ^{Gs}*dsyGD*/^{Zn}*dsyG* transcription under different growth conditions (Fig. 3a, Supplementary Fig. 5, Supplementary Fig. 6a, c, d). DMSP may not be an important cryoprotectant in *G. sunshinyii* since DMSP production and *dsyGD* transcription were not upregulated by growth under low temperature. However, these bacteria also produced the nitrogenous osmolyte glycine betaine (GB), which was more likely used as a cryoprotectant in *G. sunshinyii* since its production was upregulated by cold temperatures (Supplementary Fig. 5a, b, Supplementary Fig. 6a, b).

Organisms with DsyGD/DsyG likely produce DMSP as an osmolyte, especially when nitrogen is in short supply, as it is in most surface marine waters[43]. DMSP production and/or ^{Gs}*dsyGD*/^{Zn}*dsyG* gene transcription was significantly upregulated by growth under increased salinity and low nitrogen conditions in both *G. sunshinyii* (Fig. 3a and Supplementary Fig. 5) and *Z. navalis* (Supplementary Fig. 6). These data are

consistent with findings on other DMSP-producing organisms[8, 9], where sulfur osmolyte production over nitrogen-containing equivalents was proposed to be advantageous in sulfur rich but low nitrogen marine settings. Note, DMSP production also releases nitrogen from the Met transamination of methionine (Fig. 1). GB production in *G. sunshinyii* and *Z. navalis* was also enhanced by increased salinity (Supplementary Fig. 5b, Supplementary Fig. 6a, b) and was likely a major osmolyte, since it was found at higher concentrations than DMSP. An important exception to this rule was in low nitrogen conditions where GB and DMSP levels were more similar and the potential importance of DMSP was enhanced (Supplementary Fig. 5b, Supplementary Fig. 6a, b, c).

Further supporting the role of DMSP and DsyGD in osmoprotection, cloned ^{Gs}*dsyGD* significantly enhanced the growth of an osmosensitive *E. coli* strain FF4169[44] under increased salinity in the presence of MTHB (which has limited osmoprotective properties[38]) or, especially, DMSHB, compared to control strains lacking cloned *dsyGD* (Fig. 3b). This osmoprotection phenotype was likely due to the DMSP produced from MTHB and DMSHB (5.49 ± 0.99 and 10.13 ± 0.63 pmol DMSP µg protein^{-1} h^{-1}, respectively), since *E. coli* strain FF4169 lacking cloned ^{Gs}*dsyGD* produced no DMSP from MTHB or DMSHB (Fig. 3c). This is the first demonstration of any known DMSP synthesis genes conferring osmoprotection.

Although no DsyGD proteins were predicted in eukaryotic algae, we did identify single domain DsyG-like proteins with < 38% amino acid identity to ^{Gs}DsyG from sequenced algal genomes (*Fragilariopsis cylindrus* CCMP1102 and *Nitzschia inconspicua* strain hildebrandi) and the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) database[45]. Furthermore, 61 DsyG-like proteins were predicted from the transcriptomes of 397 different marine eukaryotes in the MMETSP (Supplementary Table 4). These algal proteins, termed DSYE for eukaryotes, were phylogenetically distinct to DsyG and were themselves separated into five separate clades (termed DSYE clade A-E) (Fig. 2). Multiple representative DSYE proteins of the five clades were expressed in *E. coli* and all showed MSM activity (Table 1; Fig. 2).

Clade A DSYE proteins were predicted in *Chloroarachniophyta*, notably

Bigelowiella natans, known to produce high levels of DMSP[10], and *Norrisiella spp.*, not previously known to produce DMSP, but whose Clade A DSYE from *N. sphaerica* BC52 was shown to have MSM activity (Fig. 2; Supplementary Fig. 7, Table 1).

Clade B DSYE proteins were identified from many diverse and highly abundant *Chlorophyta* algae, comprising a mix of high and low DMSP producers, including *Tetraselmis* sp.[46], *Pyraminonas* sp.[46], *Bathy*coccus sp.[47] and *Mantoniella* sp.[46], known to accumulate low levels of DMSP, while *Micromonas* sp. contain both high and low DMSP-producing representatives[46, 47] (Fig. 2; Supplementary Fig. 7; Supplementary Table 4). Clade B DSYE proteins were ratified from *Tetraselmis striata* (Fig. 2) and also from *Ostreococcus* sp. (*O. prasinos* and *O. tauri*), a highly abundant and widely distributed genus in Earth's oceans[48] not previously shown to produce DMSP (Fig. 2; Table 1; Supplementary Table 4). Consistent with this work, *O. tauri* cells were found to contain DMSP (0.34 nmol DMSP ug protein^{-1}) (Table 1; Supplementary Table 5).

Clade C DSYE proteins were mostly identified in *Pelagophyte* algae, e.g. *Pelagococcus* sp., such as *P. subviridis* CCMP1429, which had a functional DSYE and also contained DSYB[9], and *Pelagomonas* spp., both thought to produce low levels of DMSP[10, 46, 47] (Fig. 2; Supplementary Fig. 7). Since very few *Pelagophyte* algae have been tested for DMSP production, we examined diverse axenic cultures of *Chrysocystis*, *Aureococcus*, *Pelagococcus*, *Chrysoreinhardia* and *Pelagomonas* for this ability. These abundant picoeukaryotes, which are bloom-forming and sometimes toxin-producing [49-51], accumulated cellular DMSP to an average concentration of 85 mM (13.79 – 233.81 mM, Supplementary Table 5). Thus, it is possible that these picoeukaryotes, e.g. *P. calceolate*, amongst the most abundant eukaryotic species in the oceans[52], are important sources of DMSP in Earth's oceans.

Haptophyta are generally thought to produce high levels of DMSP and contain *DSYB*[9, 53]. *Pavlova* spp. and *Exanthemachysis* spp. are exceptions that lack *DSYB* but which contain a functional Clade D *DSYE* (Fig. 2; Supplementary Fig.7, Supplementary

Table 4). Most *Pavlova* spp. are also high DMSP producers but some, e.g., *P. lutheri*, are considered low DMSP producers, as are all tested *Exanthemachysis* spp. [10].

Clade E DSYE proteins were exclusively in diatoms, none of which contained TpMMT, although some did contain *DSYB*, e.g. *Fragilariopsis cylindrus* CCMP1102 and *Pseudonitzschia fraudulenta* WWA7, whilst others, e.g., *Nitzschia inconspicua*[9], contained only *DSYE*. DSYE from *F. cylindrus* and *N. inconspicua* were both shown to be functional. Most diatoms produce low intracellular levels of DMSP[46, 47].

The identification of *DSYE* with that of *DSYB*, *TpMMT* in algae and *dsyGD*, *dsyG*, *dsyB* and *mmtN* in diverse bacteria has greatly expanded our ability to predict which organisms, particularly algae, can produce DMSP (Fig. 2; Supplementary Fig.7, Supplementary Table 4). With inclusion of *DSYE*, 66% of the predicted 162 DMSP-producing eukaryotes[10] studied within MMETSP expressed a known *S*-methyltransferase gene involved in DMSP synthesis, an increase from 44% when considering only DSYB and TpMMT (Supplementary Table 4). Most of the remaining candidate DMSP producers on MMETSP which lacked *DSYE*, *DSYB* or *TpMMT* were predicted to be low DMSP producers or had not been tested (Supplementary Table 4). Outside of MMETSP data, there are still known DMSP-producing organisms which lack these *S*-methyltransferase genes, but their numbers are now reduced and are mainly confined to plants such as *Spartina* spp. and *Melanthera biflora* that utilize the Met methylation pathway for DMSP synthesis[21, 54, 55], macroalgae, such as *Ulva* spp., and cyanobacteria such as *Trichodesmium* that produce low levels of DMSP[56] (Supplementary Fig. 7).

A significant question left unanswered was whether the presence of known DMSP synthesis genes in an organism can imply more than just their potential to produce DMSP, for example, can they be used to predict how much DMSP they make? McParland *et al*., suggested that the presence of *DSYB* or *TpMMT* in algae was an indicator of high or low level DMSP production, respectively[53]. Certainly, this is an appealing hypothesis and there was a strong correlation of *DSYB* in high DMSP

producers (Supplementary Table 4)[9, 10, 53]. However, it is difficult to infer any such reverse scenario with TpMMT because this protein has only been studied in *Thalassiosira pseudonana*[40]. Further work on the TpMMT family is necessary to inform such hypotheses especially considering the functional data presented here (Fig. 2). However, all proteins with high homology to *T. pseudonana* TpMMT were from diatoms, predicted to be low DMSP producers (Supplementary Table 4), consistent with [10]. Considering *DSYE* was found in a mix of organisms predicted to be both low and high DMSP producers[10], it would be difficult to predict an organism's DMSP production level based on the occurrence of this gene (Supplementary Table 4), as substantiated by the varied DMSP levels seen in *Pelagophyte* algae with *DSYE*. Furthermore, representative Clade E DSYE from organisms producing lower DMSP levels (e.g. *Nitzschia inconspicua* and *Fragilariopsis cylindrus*) showed similar levels of MSM activity to Clade A and D DSYE from predicted high DMSP producers (e.g., *Bigelowiella natans* and *Pavlova* sp. CCMP459) (Supplementary Table 4). It is also worth noting that bacterial DsyB is as efficient an enzyme as algal DSYB, despite bacteria generally producing low intracellular levels of DMSP[9, 39, 57], and that there are many examples of organisms with *DSYB* that are considered low DMSP producers (e.g., *F. cylindrus, Chrysochromulina tobin*). For these reasons and as [8, 9, 39, 40] shows, we propose that DMSP synthesis gene transcript and to a greater extent, protein levels, are more robust indicators of an organism's potential DMSP levels than which DMSP synthesis gene/s it contains, since it is these expression levels that are guided by varying environmental conditions, e.g., nitrogen and salinity levels, and govern DMSP synthesis potential along with substrate availability.

Irrespective of whether one can infer high or low DMSP production according to an organism's genotype, it was possible to obtain a better understanding of the global distribution and significance of *DsyG/GD* and *DSYE* compared to *DsyB*, *DSYB*, *MmtN* and *TpMMT* genes in marine waters by examining their abundance and expression profiles in Tara Oceans datasets[58]. Initially, the ocean microbial reference gene catalogue (OM-RGC) metagenomic dataset, generated from samples fractionated to < 3 μm[59] and apportioned to bacterioplankton was analyzed. As previously described[8],

alphaproteobacterial *dysB* and its transcripts were far more abundant than those for *mmtN* in Earth's oceans, and these *dsyB* genes/transcripts were >2-fold more abundant in the surface (SRF) and deep chlorophyll maximum (DCM) than in mesopelagic (MES) waters (Fig. 4; Supplementary Table 6). *dsyGD/dsyG* genes and transcripts were not detected in any OMRGC sample, consistent with this system being largely irrelevant to marine DMSP cycling or that some species, notably filamentous cyanobacteria containing these genes, aggregated and were not captured by the bacterioplankton sampling methods. However, eukaryotic clade B *DSYE* genes and their transcripts from *Chlorophyta* algae (picoeukaryotes including *Pyramimonas*, *Pterosperma, Ostreococcus*, *Micromonas* and *Tetraselmis*), small enough to be in the bacterioplankton samples, were present in almost all stations, at similar to 2-fold lower levels than *dsyB* (normalized to *recA* for comparison) in SRF and DCM samples (Supplementary Fig. 8). Here, we estimated that ~6% of the picoeukaryotes in these SRF and DCM samples contained *DSYE*. *DSYE* and its transcripts were barely detected in MES samples consistent with the phototrophic lifestyle of their algal hosts (Supplementary Table 6; Fig. 4). The *dsyB* and *DSYE* genes and transcripts in OMRGC were most abundant in high latitude polar samples, with a few exceptions, notably, maximal *dsyB* abundance was seen in a mid-latitude DCM sample (Supplementary Fig. 8).

Within the eukaryotic Marine Atlas of Tara Ocean Unigenes (MATOU), algal DMSP synthesis genes and transcripts were also barely detected in data from MES but were much better represented in the SRF and DCM samples, consistent with their being in phototrophs (Supplementary Table 7). Although *DSYB* genes, mostly from *Haptophyta* and *Dinophyta* were detected in all stations, *DSYE* genes, predominantly from *Pelagophyceae* (clade C) and to a lesser extent, *Chlorophyta* (clade B), were marginally and ~2-fold more abundant in the photic SRF and DCM samples, respectively (Fig. 4 and Supplementary Table 6). The *DSYB* and *DSYE* genes showed similar biogeographical distribution patterns in MATOU stations, being concentrated in non-polar sites between -50 to 50 latitude (Fig. 4). Unlike *Haptophyta* and *Dinophyta*, few *Pelagophyte* algae had been studied for DMSP production despite their general ability to form large blooms and significant global abundance[60, 61], thus, they were not generally thought to be globally important DMSP producers. We now know these DSYE-containing algae can produce high intracellular DMSP levels.

In contrast to the metagenomic data, *DSYB* transcripts were ~2-fold higher than those for *DSYE* in SRF and DCM MATOU data (Fig. 4 and Supplementary Table 6) and this may be a better indication of DMSP production than gene abundance. Diatom *TpMMT* and their transcripts were generally 1-2 orders of magnitude less abundant than those for algal *DSYB* or *DSYE* (Fig. 4 and Supplementary Table 6). This bioinformatic analysis is consistent with previous reports of *Haptophyta* and *Dinophyta*[9], but also now, of *Pelagophyte* algae being important global producers of DMSP, with most diatoms having a less prominent role. Further environmental sampling work on *Pelagophyte* algae is required to explore their importance in global DMSP cycling, especially during blooms where they are likely to have a more considerable impact.

Discussion

DMSP is an abundant and ecologically important marine organosulfur compound. This study identified the novel and unusual DMSP synthesis genes *dsyGD*/*dsyG* in the rhizobacterium *G. sunshinyii* and filamentous cyanobacteria, never suspected to produce DMSP, and provided evidence for DMSP being an osmolyte in these organisms. It also facilitated identification of the *DsyG*-like and environmentally important *DSYE* gene in diverse eukaryotic algae, which together greatly enhanced our understanding of which organisms produced DMSP in Earth's oceans and how they did so.

DsyGD is the first reported bifunctional DMSP synthesis enzyme with two distinct domains that sequentially catalyse the last two DMSP synthesis-specific steps of the Met transamination pathway. The origin and transfer of *dsyGD* and *dsyG* between organisms was potentially interesting but difficult to address because these genes were rare in sequenced organisms and environmental samples. The DsyG domain was most closely related to the diatom TpMMT MTHB *S*-methyltransferase but was phylogenetically distinct to this and any other isoform enzymes (Fig. 2). Furthermore, there were proteins lacking MSM activity phylogenetically placed in between TpMMT and DsyG, indicating it is a new family of MTHB *S*-methyltransferases.

Functional genomics identified DSYE, a new family of eukaryotic enzymes with MSM activity which were phylogenetically distinct from DsyG and the other known MTHB *S*-methyltransferases. The DSYE family was diverse and separated into five clades (A-E), each comprising taxonomically distinct DMSP-producing eukaryotic

algae. Algae with *DSYE* comprise an eclectic mix of low, medium and high DMSP producers, and algae that had not previously been reported to produce DMSP, e.g. *Ostreococcus tauri* and multiple *Pelagophyte algae*. *DSYE*, with *DSYB* and *TpMMT*, serve as reporter genes of DMSP synthesis, and their combined presence in most known DMSP-producing algae with available transcriptomic/genomic data, allows more comprehensive predictions of key algal producers in marine environments with available 'omics data. Given the level of conflicting high and low DMSP levels reported for algae even from the same genera (Supplementary Table 4), it may be inappropriate to use the presence of *DSYE*, and perhaps *DSYB* and *TpMMT*, as anything more than an indicator of DMSP production. However, the DMSP synthesis gene transcript and/or protein levels from organisms and environments are likely a more robust indicator of the process, as highlighted in [9].

The *dsyGD/dsyG* and *DSYE* genes were at the different ends of the spectrum for their perceived importance in marine environments. Bacteria with *dsyGD/dsyG* were not detected in any TARA metagenomic or metatranscriptomic data, consistent with them having a negligible role in marine DMSP production. Furthermore, *dsyGD/dsyG* could not be detected in metagenomic data from *Spartina* rhizosphere samples in which *G. sunshinyii* was present[14], suggesting that *dsyGD* may not even be universal in this species. In contrast, *DSYE* genes, particularly from *Pelagophyte* and *Chlorophyta* algae were more abundant than *DSYB* from *Haptophyta* and *Dinophyta* and orders of magnitude more abundant than *TpMMT* from diatoms in Earth's surface waters. However, *DSYE* transcripts were ~2-fold less abundant than for *DSYB* in these samples, which is likely a better indicator of DMSP production. Even with these reduced transcript levels, *Pelagophyte* and *Chlorophyta* algae with *DSYE* should be considered as potentially globally important DMSP producers, especially given many of these algae are known to form large blooms and were shown here to produce medium to high levels of DMSP. Further environmental focused work on these algae is vital because they have not received the same attention from DMSP biologists as e.g., *Haptophyta* and *Dinophyta* algae[62, 63].

Assuming that the known *S*-methyltransferase genes in microbial DMSP synthesis pathways were the major isoforms, which our analysis of algal transcriptomes implied they were, it was puzzling as to why these genes and their transcripts were not more

abundant in marine systems. This is an especially relevant question considering the ubiquity of DMSP and DMSP[57] catabolic genes/transcripts, e.g., *dddP*, predicted to be in 17.5% and 2.0% of SRF marine bacteria, respectively[64]. There are still many DMSP-producers that lack known DMSP synthesis genes, e.g., DMSP-producing plants, macroalgae such as *Ulva* spp., cyanobacteria such as *Trichodesmium* and *Synechococcus*[64], and other bacteria, like *Marinobacter* sp.[8], but these are not expected to be major DMSP producers on the same scale as *Haptophyta* and *Dinophyta* algae for example. It is possible that these phototrophs contain other unidentified isoform MTHB *S*-methyltransferases or DMSP synthesis pathways with novel enzymes. This has been proposed for the *Dinophyta Crypthecodinium cohnii*, which has multiple *DSYB* copies[9] and is thought to utilize a Met decarboxylation pathway for which no genes/enzymes are known. Finally, it is also possible that the DMSP synthesis gene products are more abundant and active than their gene and transcript abundance implies. Further molecular work is required on model marine organisms to address these important questions, combined with more comprehensive environmental quantification of DMSP stocks and synthesis rates, and of DMSP biosynthetic enzyme abundance.

Materials and Methods

Strains，plasmids and culture conditions

Strains, plasmids and primers used in this study are detailed in Supplementary Tables 8, 9 and 10 respectively. *G. sunshinyii* YC6258, the *L. aggregata dsyB*$^-$ mutant strain and *Ruegeria pomeroyi* DSS-3 were grown in yeast tryptone sea salt (YTSS)[65] media or MBM minimal medium (with 10 mM succinate as carbon source and 10 mM NH_4Cl as nitrogen source) at 28°C. Where indicated, the salinity of MBM was adjusted by altering the amount of sea salts (Sigma-Aldrich), and nitrogen levels were altered by adding varying amounts of NH_4Cl. *Z. navalis* LEGE 11467 was grown in BG-11 medium[66] at 25°C. *E. coli* strains were grown in Luria-Bertani (LB) complete medium or M9 minimal medium at 37°C. *R. leguminosarum* J391 was grown in tryptone yeast (TY) [67] or RM media[67] at 28°C. All liquid cultures were grown with shaking at 180-200 rpm unless otherwise specified. Where necessary, the following antibiotics were added to media at the final concentrations: ampicillin (100 μg mL^{-1}), streptomycin (400 μg ml^{-1}), kanamycin (20 μg mL^{-1}), rifampicin (20 μg mL^{-1}), tetracycline (10 μg mL^{-1}) and gentamicin (20 μg mL^{-1} in general or 80 μg mL^{-1} for pLMB509 complementation

in *L. aggregata dsyB*$^{-}$ mutants).

Analysis of DMSP in *C. scabrifolia*

Plant and rhizosphere soil were obtained in a saltern area in Shandong Province, China (120.7°E, 36.5°N). *C. scabrifolia* plants were carefully uprooted and placed into sterile plastic bags. Plant material was washed to remove sediment and then separated into different tissue types (roots and leaves) using ethanol sterilized scissors and tweezers. The rhizosphere of *C. scabrifolia* were sampled by washing 5 g of roots as in [67]. To measure DMSP content, different samples were prepared as in [67]. DMSP was assayed by gas chromatography (GC) as described below.

***G. sunshinyii* YC6258 DMSP cycling gene analysis**

The *G. sunshinyii* YC6258 genome sequence and protein annotation data were downloaded from NCBI (PRJNA233633) and interrogated for the presence of DMSP synthesis cycling gene/proteins. *G. sunshinyii* DMSP cycling genes/proteins were predicted using local BLAST, with ratified DMSP cycling genes/proteins as probes (Supplementary table 11) and thresholds of E-value $\leq 1\times10^{-5}$, amino acid identity ≥40% and coverage ≥70%.

Quantification of DMS, DMSHB and DMSP by GC

All GC assays involved measurement of headspace DMS, either directly produced or produced through alkaline lysis of DMSP and/or DMSHB using a flame photometric detector (Agilent 7890A GC fitted with a 7693 autosampler) and a HP-INNOWax 30 m×0.320 mm capillary column (Agilent Technologies, J&W Scientific) as in Curson *et al*.[39]. Unless otherwise stated, all GC measurements for DMSP and/or DMSHB were performed using 2 mL glass vials containing 200 µL liquid samples and 100 µL of 10M NaOH sealed with PTFE/rubber crimp caps, followed by incubation overnight at 28 ℃, allowing DMS to accumulate in the headspace. DMS chemically released from DMSP for use in a calibration curve[39], and DMSHB was synthesized as described in Curson *et al*.[39]. The detection limit for headspace DMS from DMSP was 0.015 nmol in water/media and 0.15 nmol in methanol, and from DMSHB was 0.3 nmol in water/media.

DMSP synthesis in *G. sunshinyii*

To study the DMSP synthesis pathway in *G. sunshinyii* YC6258 cells, cultures were incubated overnight in YTSS liquid medium, adjusted to equal optical densities (OD_{600} = 0.3), washed three times with 35 PSU MBM minimal medium, then diluted 1:100 into 5 mL 35 PSU MBM medium (with 10 mM succinate as carbon source, 10 mM NH_4Cl as nitrogen source) with or without (control) different DMSP synthesis intermediates (0.5 mM Met, MTOB, MTHB, DMSHB, 3-methylthiopropylamine (MTPA), methylmercaptopropionate (MMPA)) and incubated for 24 h at 30°C. Three biological replicates were prepared for each condition and DMSP production activities were normalized to protein concentrations determined using the Bradford method (Quick Start™ Bradford, USA, BioRad), as in Curson *et al*.[39]. A student's t-test ($p < 0.05$) was used to identify significant differences in DMSP production. Error bars represent standard deviation (n = 3).

To quantify *in vitro* MSM and DMSHB decarboxylation (DDC) activities in *G. sunshinyii* YC6258, this bacterium was first cultured in triplicate overnight in 5 mL YTSS medium at 30°C, harvested by centrifugation at 20,000 *g* for 5 min, washed 3 times with 1 ml 50 mM Tris-HCl buffer (pH 7.5), and then resuspended in 1 ml 50 mM Tris-HCl buffer. To generate cell lysates, cells were sonicated (3×10 s) on ice using a Markson GE50 Ultrasonic Processor set to an output of 70, then centrifuged at 20,000 *g* for 5 min to pellet the debris, and the lysate was removed and used. The lysates were dialysed to remove any pre-existing metabolites, using dialysis tubing (3,500 Da molecular weight cut-off; Spectrum Labs) in 2 L of dialysis buffer (20 mM HEPES, 150 mM NaCl, pH 7.5) at 4°C overnight [8]. 200 µl of cell-free extracts alone (control), or extract supplemented with 1 mM MTHB (Sigma) and 1 mM SAM (New England Biolabs) or just 1 mM DMSHB (synthesized as in [39]), were added to GC vials and incubated at 28°C for 30 min. After incubation, assays were immediately mixed with 100 µL 10 M NaOH and assayed for DMSHB and/or DMSP by GC, as above.

Identification of *G. sunshinyii* YC6258 *dsyGD*

A genomic library of *G. sunshinyii* YC6258 was constructed as described in Curson *et al*.[33]. Briefly, high quality *G. sunshinyii* YC6258 genomic DNA was partially digested with *Eco*RI, and then ligated into *Eco*RI-digested and dephosphorylated pLAFR3[68]. The ligation mix was packaged and transfected into *E. coli* strain 803. The library

comprising 90,000 clones was transferred *en masse* into the heterologous host *R. leguminosarum* J391 by conjugation using the helper plasmid pRK2013 (in *E. coli* 803) and transconjugants were selected on TY media containing streptomycin and tetracycline. Transconjugants were picked to 200 µL RM medium containing 0.5 mM MTHB in 2 mL GC vials and incubated at 30°C for 48 h. 100 µL of 10 M NaOH was then added to each of the vials, which were sealed and assayed for DMSHB and DMSP by GC analysis, as above. DMSHB plus DMSP levels in the headspace were normalised to protein levels, as above. *R. leguminosarum* J391 with empty pLAFR3 vector and media only, with and without MTHB substrate, were used as controls.

Heterologous expression, *in vivo* assays and purification of *G. sunshinyii* YC6258 proteins.

Full-length *G. sunshinyii* YC6258 *dsyGD* (including the *dsyG* methyltransferase and *dsyD* decarboxylase domain), the separate *dsyG* methyltransferase and *dsyD* decarboxylase gene domains, and the putative reductase gene were PCR-amplified and cloned into pET-22b (Supplementary table 10). All clones were ratified by sequencing. The pET-22b clones were transformed into *E. coli* BL21 (DE3), cultured in LB medium containing ampicillin at 37°C to an OD_{600} of 0.8–1.0, and then induced at 18°C for 14 h with 0.4 mM IPTG. These cells were either incubated with 0.5 mM MTHB (for cloned *dsyGD* and *dsyG*), DMSHB (for cloned *dsyGD* and *dsyD*) or 0.5 mM MTOB (for the cloned reductase) and assayed for *in vivo* MSM and DDC activity assays by GC, as above, and MR activity as in [8], or without for control experiments and for protein purification work. For the latter, cells were harvested by centrifugation (20 min, 7,500 *g*, 4°C), washed, and resuspended in 25 mM Tris-HCl (pH 8.0), 150 mM NaCl. The overexpressed recombinant proteins were initially purified by Ni^{2+}-NTA affinity chromatography (GE healthcare, America), and then further isolated by gel filtration on a Superdex200 column (Cytiva), as described in [57]. Purified protein (Supplementary Fig. 2) was flash-frozen in liquid nitrogen and stored at −80°C until required.

***in vitro* MSM, DDC and MTOB reductase enzyme assays**

Where appropriate, recombinant *G. sunshinyii* YC6258 DsyGD and individual DsyG, DsyD and reductase domain proteins were assayed for MSM, DDC and MR activity as in [8]. For *in vitro* MSM activity, 5-1000 µM MTHB and 10-1000 µM SAM and 0.1 µM

purified DsyGD/DsyG were mixed in a total volume of 100 µL reaction buffer containing 100 mM Tris-HCl (pH 7.0) and incubated at 25℃ for 10 min in triplicate. 15 µL of 20% HCl was added to stop the reactions. Reaction buffers with no enzymes added were used as negative controls in triplicate. MSM activity was measured by detecting production of *S*-adenosyl-homocysteine (SAH) from demethylation of SAM via ultraviolet absorbance measurements by HPLC (Ultimate 3000, Dionex and LC-20AT, Shimadzu) on a SunFire C18 column (Waters) with a linear gradient of 1–20% acetonitrile in 50 mM ammonium acetate (pH 5.5) over 24 min at 260 nm, as described in [32].

For *in vitro* DDC activity, 0.5-3 mM DMSHB and 0.1 µM purified DsyGD and DsyD proteins were mixed in a total volume of 100 µL with reaction buffer (100 mM Tris-HCl (pH 7.0)), before incubation at 25℃ for 10 min in triplicate. 15 µL of 20% HCl was added to stop the reaction. *in vitro* DDC activity of DsyGD and DsyD was monitored via the HPLC detection of acrylate (as for SAH, above) produced from the alkaline hydrolysis of the DMSP reaction product[69, 70].

The optimal temperature of DsyGD was determined by incubating the reaction mixture at 10, 15, 20, 25, 30, 40, 50 and 60℃ with MTHB and monitoring MSM activity. The optimal pH of DsyGD was determined by incubating the reaction mixture with MTHB in Britton-Robinson buffer at 25℃ at pH 4, 5, 6, 7, 8, 9, and 10, as performed in [71] and assaying MSM activity. The kinetic parameters of DsyGD for MSM and DDC activity were determined by nonlinear analysis based on the initial rates determined with 5-1000 µM MTHB, 0-250 µM SAM or 500-3000 µM DMSHB at the optimal temperature and pH, as described in [8].

For *in vitro* MR activity 1 mM MTOB was added to 0.25 mM NADPH and incubated at 30 ℃ in triplicate. The reaction was started by addition of 1 µM purified reductase enzyme in a total volume of 2 mL reaction buffer (10 mM Tris-HCl, pH 8.0) and the reaction mixture without reductase was used as negative controls. MR activity was monitored by NADPH reduction at 340 nm using a V550 UV/VIS spectrophotometer (Jasco, Japan) at 0, 15 and 180 minutes after enzyme addition.

RNA isolation and reverse transcription quantitative polymerase chain reaction (RT– qPCR) work with *G. sunshinyii*

G. sunshinyii YC6258 was inoculated in 5 mL YTSS media incubated with shaking at

30°C for overnight, then adjusted to equal optical densities (OD_{600} = 0.8), as described in [39]. Three biological replicates were then grown under standard conditions (salinity at 35 PSU, 10 mM NH_4Cl, 30°C) or a range of stress conditions including: low salt/low nitrogen (5 PSU, 0.5 mM NH_4Cl), low nitrogen (35 PSU, 0.5 mM NH_4Cl), low temperature (35 PSU,10 mM NH_4Cl, 16°C), high salt/low nitrogen (50 PSU, 0.5 mM NH_4Cl) and high salt (50 PSU, 10 mM NH_4Cl). Sampling was performed in exponential-phase (OD_{600} of ~0.5). RNA was isolated according to the RNeasy Mini Kit protocol (Qiagen, Germany) and quantified using a Qubit 3.0 Fluorometer and Qubit RNA HS Assay Kit (Thermo Fisher Scientific). 1 µg of DNA-free RNA was used for reverse transcription using a QuantiTect Reverse Transcription Kit (Qiagen, Germany). No reverse transcriptase and no template controls were performed.

Primers for RT-qPCR for *G. sunshinyii* YC6258 *dsyGD* and control housekeeping genes *recA* and *rpoD* were designed by primer premier 6 and synthesized by Eurofins Genomics (Supplementary Table 10). The optimum primer melting temperature was 60°C. Primer GC content was between 40% and 60% and primer efficiencies were all 90–110% and within recommended limits. Three technical replicates were performed for each sample. RT-qPCR was performed as in Curson *et al*.[39]. Reactions (20 µL) were performed with an annealing/elongation temperature of 60°C. Standard curves were included in each run to calculate the reaction efficiency (five points in 1:10 dilutions starting from 100 ng gDNA and water only as negative control). Analysis of the post-run melt curves was also performed. For each condition and gene, the cycle threshold (Ct) values of the technical and biological replicates were averaged. The *rpoD* and *recA* Ct values and efficiencies were then averaged and the relative expression ratio was calculated[72] and expressed as normalized fold change relative to the standard conditions.

Quantification of DMSP and glycine betaine by Nuclear Magnetic Resonance (NMR)

Triplicate 5 mL cultures of *G. sunshinyii* YC6258 grown in MBM media under different stress conditions (see above), were pelleted at 12,000 *g* for 10 min and resuspended in 445 µL of deuterium oxide (D_2O, Sigma-Aldrich). The cells were then lysed by 3 × 20

sec sonication rounds using a Markson GE50 Ultrasonic Processor, followed by centrifugation at 12,000 *g* for 10 min. 5 µL of pyrazine (Sigma-Aldrich; 50 mM final concentration) was added to the clarified samples as an internal NMR standard, mixed, and 445 µL was transferred to 5 mm NMR tubes. NMR experiments were performed at 298 K on a Bruker 500 MHz spectrometer with an auto-sampler. The pulse sequence used incorporated a double echo excitation sculpting component for water suppression (Bruker library zgesgp) to remove the residual water coming from the original culture. Each sample was run with the number of scans at 128 and the relaxation delay d1 was 1 s.

All spectra were phased, base-corrected and calibrated for the pyrazine peak at 8.63975 ppm. The chemical shift of the methyl groups of GB ($(CH_3)_3N$) and DMSP ($(CH_3)_2S$) were, respectively, 3.256 ppm and 2.913 ppm at 298 K. The final concentration of the analytes GB and DMSP was obtained by calculating the ratio of the absolute integral of pyrazine (accounting for 4 protons) with the methyl peaks of GB and DMSP (accounting for 9 protons and 6 protons respectively). These ratios were then multiplied by i) the correction factor derived from the calibration curves and ii) the dilution factor of the samples from the original culture to the final NMR sample. The calibration curve correction factor was 2.963 and 2.719 for GB and DMSP, respectively. The dilution factor was 0.1125, accounting for the dilution from 4 mL to 0.45 mL. Calibration curves were obtained for GB and DMSP at 0.2 mM, 0.4 mM, 0.8 mM and 1.6 mM in the presence of 1 mM pyrazine. For each sample, a zgesgp at d1=1 s was recorded, and the data were plotted to obtain the correction factor. For both curves, the R^2 was 0.99.

Salinity tolerance experiments in *E. coli* strains

E. coli strain MC4100 and its salt sensitive derivative FF4169 with a $\triangle otsA^-$ mutation in the trehalose-6-phosphate synthase gene that renders it unable to produce the osmolyte trehalose[44], were used to study salt tolerance conferred by cloned $^{Gs}dsyGD$. The $^{Gs}dsyGD$ gene with promoter region was PCR amplified and cloned into the pUCm-T vector and transformed into *E. coli* FF4169. Starter cultures of MC4100, FF4169 and

FF4169 with the *dsyGD* clone plasmid (FF4169: pUCm-T: *dsyGD*) were grown in LB medium overnight (in triplicate). All starter cultures were adjusted to OD_{600} = 0.3 and washed twice with M63 medium lacking NaCl and sulfur followed by resuspension in 1 mL M63 medium as described in [38]. The suspensions were diluted 1:100 in new M63 medium (22 mM D-glucose as carbon source, 1 mM $MgSO_4$ as sulfur source) with high salinity (0.5 M NaCl) and either DMSP, GB, MTHB or DMSHB (all at 1 mM final concentration) as substrates. IPTG was added at a final concentration of 0.1 mM to induce the expression of pUCm-T: *dsyGD* in FF4169. All cultures were grown at 37°C with continuous shaking and OD_{600nm} was monitored using a plate reader (Thermo Scientific, Multiskan GO) every 1 h until stationary phase. DMSP production was assayed at the end of each experiment to confirm production of DMSP.

Identification and characterization of DsyGD, DsyG and DsyD homologues

^{Gs}DsyGD was used in a BLASTp search (cutoff e-value at $1e^{-55}$) against the NCBI database to identify homologous proteins with 38-50% amino acid identity to full length ^{Gs}DsyGD or either of the individual ^{Gs}DsyG and 29.8% to the ^{Gs}DsyD domains. *E. coli* codon-optimized genes, corresponding to: DsyGD homologues in *Symploca* sp. SIO3E6, *Oscillatoria* sp. SIO1A7; a DsyG homologue in *Z. navalis* LEGE 11467; DsyG-like proteins in *P. umbilicalis* and *Planctomycetales* bacterium; and DsyD-like proteins from *P. parvum* Texoma1 and *A. monilatum* CCMP3105, were synthesized (Sangon Biotech, Shanghai Co., Ltd.), cloned into the T7 expression plasmid pET-16b and transformed into *E. coli* BL21(DE3).

To measure MSM and/or DDC activity of the DsyGD, DsyG and DsyG-like enzymes, *E. coli* BL21(DE3) expressing these recombinant proteins (Supplementary Table 2) were grown in triplicate, induced by IPTG and assayed for *in vivo* MSM and/or DDC activity in *E. coli*. *E. coli* BL21(DE3) containing empty pET-16b vector was used as a control.

To further study the *in vivo* MSM and DDC activity of the above genes, where indicated (Supplementary Table 2), the *dsyGD*, *dsyG*, and *dsyD* gene homologues were

cloned into the *Nde*I and *Eco*RI sites of the wide host range taurine inducible expression plasmid pLMB509[8]. Plasmids were conjugated into the *L. aggregata dsyB*$^-$ mutant, that makes no DMSP[39] and/or *R. pomeroyi* DSS-3 strain (for *dsyD* clones as it cannot produce DMSP from DMSHB,) using the helper plasmid pRK2013[73], as described in Curson *et al*.[39]. For DMSP production assays, cultures were grown in YTSS complete medium (in triplicate), at 30°C for 24 h. Cultures were then adjusted to an OD_{600} of 0.3, washed three times with 35 PSU MBM medium, before being diluted 1:100 into 5 mL MBM medium with 5 mM taurine (Sigma-Aldrich). Where indicated, 0.5 mM MTHB or 0.5 mM DMSHB were added as substrate and samples were incubated at 28°C for 24 h before DMSP production was monitored by GC.

Growth of *Zarconia navalis* LEGE 11467 under different conditions

Z. navalis LEGE 11467 was obtained from the Blue Biotechnology and Ecotoxicology Culture Collection (LEGE-CC) from CIIMAR in Portugal[41] and grown with shaking at 25°C in BG-11 medium at 25 PSU (with 17.65 mM $NaNO_3$ as a nitrogen source) as described in [66] for 30 days. Triplicate samples were then set up with different salt and nitrate concentrations: standard conditions (25 PSU, 17.65 mM $NaNO_3$); low salt (5 PSU, 17.65 mM $NaNO_3$); high salt (50 PSU, 17.65 mM $NaNO_3$); and low nitrogen (25 PSU, 0.5 mM $NaNO_3$). Sample were collected at 0, 7 and 18 days after inoculation by centrifuging 50 mL of the culture at 5,000 *g* at 4°C for 10 min. Samples were washed with distilled water, split and used respectively for GC, NMR and RNA isolation.

RNA isolation from *Z. navalis* LEGE 11467 and cDNA synthesis

Z. navalis LEGE 11467 culture pellets were resuspended in 1 mL of RNAlater RNA Stabilization Reagent (Qiagen). The suspension was incubated at room temperature for 5 min, followed by centrifugation at 2,000 *g* for 5 min at 4°C. The supernatant was removed and the cell pellet stored at -80ºC for RNA extraction. To extract RNA, the PureLink® RNA Mini Kit (Invitrogen) was used. Cells were first disrupted using liquid nitrogen followed by addition of 600 µL of lysis buffer and 6 µL of 2-mercaptoethanol

683 to homogenize cells. The lysate was transferred into a clean tube and centrifuged at
684 12,000 *g* for 5 min. Isolated RNA was stored on ice and incubated with DNase I to
685 remove DNA contamination, using the Rapid Out DNA Removal Kit (Thermo
686 Scientific™). RNA samples were quantified in a DS-11 Series
687 Spectrophotometer/Fluorimeter (DeNovix). cDNA was synthesized using NZY First-
688 Strand cDNA Synthesis Kit (Nzytech) according to the manufacturer's instructions.
689 Samples were kept at -20°C until RT-qPCR analysis.
690

691 **Identification and characterization of eukaryotic DSYE enzymes**

692 BLASTP searches (with a raised E value of $1e^{-55}$ and ≥ 70% ^{Gs}DsyG sequence coverage)
693 were performed against the predicted proteomes of genomes on NCBI and the 678
694 transcriptomes on the Marine Microbial Eukaryote Transcriptome Sequencing Project
695 [45] (MMETSP, downloaded from iMicrobe (https://imicrobe.us/#/projects/104). Local
696 BLASTP (E value of $1e^{-5}$) analysis was also performed against NCBI and MMETSP
697 for the DsyD domains (Supplementary Table 3).
698 All prokaryotic DsyB, MmtN, DsyGD, DsyG, DsyG-like (lacking MSM function)
699 and DsyD, and eukaryotic DSYB, TpMMT and DSYE sequences were aligned in
700 MAFFT version 7[74] using default settings, then visually checked. The *S*-
701 methyltransferase or the decarboxylase domains of these enzyme sequences were then
702 collected for construction of a Maximum Likelihood phylogenetic trees using MEGA
703 version X, as in [75] (Fig 2 and Supplementary Fig. 4). The maximum likelihood
704 phylogenetic trees were visualized and annotated using the Interactive Tree Of Life
705 (iTOL) version 6.6[76]. Two DsyD-like proteins were uncovered in *P. parvum* Texoma1
706 and *A. monilatum* CCMP3105 and were analysed for DDC activity, see above. For the
707 *S*-methyltransferase tree (Fig. 2), it was apparent that TpMMT, DsyG and DSYE
708 represented phylogenetically distinct protein families, with the former two being more
709 similar but separated by the two DsyG-like proteins lacking MSM function. The DSYE
710 family of proteins was separated into five distinct taxonomic clades (Clade A-E).
711 Multiple representative *DSYE* sequences from each of the five clades were codon

optimized for expression in *E. coli*, synthesized (Sangon Biotech, Shanghai Co., Ltd.), cloned into pET-16b (incorporating *Nde*I and *Bam*HI sites for subcloning) and transformed into *E. coli* BL21(DE3) (Supplementary Table 2, Supplementary Table 4). These recombinant DSYE proteins were overexpressed and assayed for in vivo MSM activity (Table 1). Where indicated, these genes were subcloned into pLMB509 using primers listed in Supplementary Table 10, conjugated into the *L. aggregata* LZB033 *dsyB*$^-$ mutant and tested for their ability to complement DMSP production.

Quantification of DMSP and glycine betaine in *Pelagophyceae* algae

Axenic *Chrysocystis fragilis* RCC 6172, *Aureococcus anophagefferens* RCC 4094, *Pelagococcus subviridis* RCC 4422, *Chrysoreinhardia* sp. RCC 2956, and *Pelagomonas calceolata* RCC 100 were purchased from the Roscoff Culture Collection (RCC). *Ostreococcus tauri* was kindly provided by V. Jackson and A. Monier at the University of Exeter (Supplementary Table 5). Cultures were acclimated at 22°C under16 h light (120 µmol photons m^{-2} s^{-1})/ 8 h dark prior to DMSP measurements. To obtain samples for DMSP quantification by GC, 4 mL of culture was centrifuged at 6,000 *g* for 10 min and the pellet resuspended in 200 µL methanol. Samples were stored at −20°C for 24 h to allow for extraction of cellular metabolites. The methanol extracts were transferred to GC vials, and 100 µL 10 M NaOH was added. Vials were crimped immediately, incubated at 22°C for 24 h in the dark prior to DMSP measurements by GC. All measurements for DMSP production in *Pelagophyte* strains are based on the mean of at least three biological replicates per strain tested. For algal cell enumeration, aliquot samples were sampled and diluted then quantified using a CASY model TT cell counter (Sedna Scientific).

Cyanobacteria and *Pelagophyceae* algae sample preparation for NMR

6 mL of *Z. navalis* LEGE 11467 or *Pelagophyceae* algae culture were pelleted at 5,000 *g* for 10 min in triplicate samples, the supernatant was discarded, and the pellets were resuspended in 500 µL of deuterium oxide (D_2O, Sigma-Aldrich). These samples were transferred to 2 mL tubes containing 0.25 g 1.4 mm ceramic spheres, 0.1 mm silica spheres and one 4 mm glass bead. Samples were homogenized for 3 x 40 s with 2 min interval at speed of 6.0 m/s using the FastPrep-24 5G (FP5G, FastPrep™ system, MP

Biomedicals™). Pyrazine (Sigma-Aldrich) was added to a 50 mM final concentration and 500 μL samples analysed by NMR for DMSP and GB (Supplementary Fig. 9), as described above.

Metagenomic analysis of *dsyGD* presence

Metagenomes data of *S. alterniflora*, *R. stylosa* and mangrove sediment samples were download from the Chinese National Genomics Data Center GSA database PRJCA002729. Relative abundance of *dsyGD* was analysis as in [77]. Only homologues with ≥40% amino acid identity and ≥70% coverage to ratified sequences were counted and used for analysis.

Statistical methods

Statistical methods for RT–qPCR are described in the relevant section above. All measurements for DMSP and DMS production (in bacterial strains or enzyme assays) are based on the mean of three biological replicates per strain/condition tested, and error bars are shown from calculations of standard deviations, with all experiments performed at least twice. To identify statistically significant differences between standard and experimental conditions in Figs 1b, 3a and Supplementary Fig. 5b, a single-tailed independent Student's t-test ($P < 0.05$) was applied to the data.

Bioinformatics

To search for gene homologs in the Tara Ocean metagenome/metatranscriptome datasets, a Hidden Markov Model (HMM) profile was generated based on the amino acid sequences of reported DMSP synthesis enzymes and ratified DSYE proteins (Supplementary Data 1) by the HMMER tools (v.3.3, http://hmmer.janelia.org/)[78]. Hmmer method searching was performed under default settings with a threshold of $1e^{-30}$. HMM databases were then submitted to online webserver Ocean Gene Atlas[58] to search against the prokaryote-enriched Oceans Microbiome Reference Gene Catalog (OM-RGC_v2) and eukaryote-enriched Marine Atlas of Tara Oceans Unigenes

(MATOU) dataset using a cutoff e-value of $1e^{-30}$[77]. The homolog sequences files, normalized abundance, and environmental data were obtained. Only homologues with ≥40% amino acid identity and ≥70% coverage to ratified sequences were counted and used for analysis. The different methyltransferase protein biogeographic distribution was plotted by R (v. 4.0.3) using scatterpie and ggplot2[79]. Eukaryotic methyltransferase protein abundance was normalized by *β-Actin* and prokaryotic methyltransferase protein abundance was normalized by *recA*[77].

Availability of data and materials

All sequence data are archived in the NCBI database

Acknowledgements

We acknowledge Young Ryun Chung from Gyeongsang National University for kindly providing G. sunshinyii YC6258 strain, Dr. Ji Liu, Dr. Yanfen Zheng and Dr. Jingli Liu for experimental advice. We also thank Dr. Ornella Carrión for advice on genomic library construction and Guang Yang from the Ocean University of China for help with the sampling and identification of *Carex scabrifolia*.

This work was funded by the National Natural Science Foundation of China (91751202 and 41730530), the National Key Research and Development Program of China (2018YFE0124100), the Fundamental Research Funds for the Central Universities (202172002) in X-HZ's lab, the Fundação para a Ciência e a Tecnologia, Portugal (grants UIDB/04423/2020 and UIDP/04423/2020) in PNL's lab, Biotechnology and Biological Sciences Research Council, UK (BB/X005968), Natural Environmental Research Council, UK, (NE/P012671, NE/S001352, NE/X000990 and NE/X014428 with DLS) and the Leverhulme trust (RPG-2020-413) grants in JDT's lab. KSW and LH acknowledges funding support from the BBSRC Norwich Research Park Doctoral Training Partnership programme (BB/2244848 and BB/M011216/1, respectively).

Conflicts of interest

The authors declare that they have no conflict of interest.

Authors' contributions

J.D.T. and **X.-H.Z.** conceived **and** designed all of the experiments, analyzed the data

and wrote the paper with **J.Y. W.** wrote the paper, designed all of the experiments and performed or contributed to all of the experiments, analysed all the data, and prepared figures and tables. **S. Z.** performed *G. sunshinyii* RNA isolation and *dsyGD* RT - qPCR. **A.R.J** performed the following experiments: *dsyGD* cloning into pET-16b and assays of DMSP production by *G. sunshinyii* with different synthesis intermediates) and provided advice of genomic library construction. **A.V.** and **P.N.L** performed *Z. navalis* LEGE 11467 growth experiments and RNA isolation. **K.S.W** performed the *G. sunshinyii* and *Z. navalis* LEGE 11467 protein assays for normalizing DMSP production. **P.P.L.R.** performed the phytoplankton growth experiments and DMSP production assays. **S.M.** performed the NMR detection. **L.H.** performed the *Ostreococcus tauri* DMSP production measurements. **X.Y. Z.** performed the *Z. navalis* LEGE 11467 *dsyG* RT-qPCR. **C.Y. L.** and **Y.-Z. Z** performed protein purification and activity assays. **X.D. W.** performed experiments (purified protein and activity assay). **D.L.S.** performed critical revision of the manuscript. All authors edited and approved the manuscript.

References

1. Ksionzek, K. B. et al. Dissolved organic sulfur in the ocean: Biogeochemistry of a petagram inventory. *Science* **354**, 456-459 (2016).
2. Sunda, W., Kieber, D. J., Kiene, R. P. & Huntsman S. An antioxidant function for DMSP and DMS in marine algae. *Nature* **418,** 317-320 (2002).
3. Zheng Y. F. et al. Bacteria are important dimethylsulfoniopropionate producers in marine aphotic and high-pressure environments. *Nat. Commun.* **11,** 1-12 (2020).
4. Seymour, J. R., Simo, R., Ahmed, T & Stocker, R. Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. *Science* **329,** 342-345 (2010).
5. Sievert, S. M., Kiene, R. P. & Schulz-Vogt, H. N. The sulfur cycle. *Oceanography* **20,** 117-123 (2007).
6. Stefels, J., Steinke, M., Turner, S., Malin, G. & Belviso, S. Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling.

Biogeochemistry **83**, 245-275 (2007).

7. Vallina S. M. & Simo R. Strong relationship between DMS and the solar radiation dose over the global surface ocean. *Science* **315**, 506-508 (2007).
8. Williams, B. T. et al. Bacteria are important dimethylsulfoniopropionate producers in coastal sediments. *Nat. Microbiol.* **4**, 1815-1825 (2019).
9. Curson, A. R. J. et al. DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. *Nat. Microbiol.* **3**, 430-439 (2018).
10. McParland E. L. et al. The role of differential DMSP production and community compositionin predicting variability of global surface DMSP concentrations. *Limnol. Oceanogr*. 2018: **9999**, 1-17 (2018).
11. Chung, E. J., Park, J. A., Jeon, C. O. & Chung, Y. R. *Gynuella sunshinyii* gen. nov., sp. nov., an antifungal rhizobacterium isolated from a halophyte, *Carex scabrifolia* Steud. *Int. J. Syst. Evol. Microbiol.* **65**, 1038-1043 (2015).
12. Helfrich, E. J. N. et al. Evolution of combinatorial diversity in *trans*-acyltransferase polyketide synthase assembly lines across bacteria. *Nat. Commun.* **12**, 1422 (2021).
13. Ueoka, R. et al. Genome-based identification of a plant-associated marine bacterium as a rich natural product source. *Angew. Chem. Int. Ed. Engl.* **57**, 14519-14523 (2018).
14. Mo, S. et al., Integration of Microbial Transformation Mechanism of Polyphosphate Accumulation and Sulfur Cycle in Subtropical Marine Mangrove Ecosystems with *Spartina alterniflora* Invasion. *Microb. Ecol.* **85**, 478-494 (2023).
15. Dacey, J. W. H., King, G. M. & Wakeham, S. G. Factors controlling emission of dimethylsulfide from salt marshes. *Nature* **330**, 643-645 (1987).
16. Steudler, P. A. & Peterson, B. J. Contribution of gaseous sulphur from salt marshes to the global sulphur cycle. *Nature* **311**, 455-457 (1984).
17. Kocsis, M. G. et al. Dimethylsulfoniopropionate biosynthesis in *Spartina alterniflora1*. Evidence that *S*-methylmethionine and dimethylsulfoniopropylamine are intermediates. *Plant Physiol.* **117**, 273-281 (1998).
18. Challenger, F. & Simpson, M. I. A precursor of the dimethyl sulphide evolved by Polysiphonia fastigiata. dimethyl-2-carboxyethylsulphonium hydroxide and its

868 salts. *Studies on biological methylation. Part XII*. **43**, 1591-1597 **(**1948).
869 19. Rolando, J. L., Kolton, M., Song, T. & Kostka, J. E. The core root microbiome of
870 *Spartina alterniflora* is predominated by sulfur-oxidizing and sulfate-reducing
871 bacteria in Georgia salt marshes, *Microbiome* **10**, 1-17 (2022).
872 20. McNeil, S. D., Nuccio, M. L. & Hanson, A. D. Betaines and related osmoprotectants.
873 Targets for metabolic engineering of stress resistance. *Plant Physiol.* **120**, 945-950
874 (1999).
875 21. Otte, M. L. & Morris,J. T. Dimethylsulphoniopropionate (DMSP) in *Spartina*
876 *alterniflora* Loisel. *Aquat. Bot.* **48**, 239-259 (1994).
877 22. Kiehn, W. M. & Morris, J. T. Variability in dimethylsulfoniopropionate (DMSP)
878 concentrations in *Spartina alterniflora* and the effect on *Littoraria irrorata*. *Mar.*
879 *Ecol. Prog. Ser.* **406**, 47-55 (2010).
880 23. Todd, J. D. et al. Molecular dissection of bacterial acrylate catabolism-unexpected
881 links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production.
882 *Environ. Microbiol.* **12**, 327-343 (2010).
883 24. Ueoka, R. et al. Genome mining of oxidation modules in *trans*-acyltransferase
884 polyketide synthases reveals a culturable source for lobatamides. *Angew. Chem. Int.*
885 *Ed. Engl.* **59**, 7761-7765 (2020).
886 25. Wang, Y., Qin, Z., Fan, L. & Zhao, L. Structure-function analysis of *Gynuella*
887 *sunshinyii* chitosanase uncovers the mechanism of substrate binding in GH family
888 46 members. *Int. J. Biol. Macromol.* **165**, 2038-2048 (2020).
889 26. Mavrodi, O. V. et al. Rhizosphere Microbial Communities of *Spartina alterniflora*
890 *and Juncus roemerianus* From Restored and Natural Tidal Marshes on Deer Island,
891 Mississippi. *Front Microbiol*. **9**, 1-13 (2018).
892 27. Todd, J.D., et al. The *dddP* gene, encoding a novel enzyme that converts
893 dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean
894 metagenomes and marine bacteria and also occurs in some Ascomycete fungi.
895 *Environ Microbiol*, **11**, 1376-1385 (2009).
896 28. Todd, J. D. et al. DddQ, a novel, cupin-containing, dimethylsulfoniopropionate
897 lyase in marine roseobacters and in uncultured marine bacteria. *Environ. Microbiol.*

898 **13**, 427-438 (2011).
899 29. Todd, J. D., Kirkwood, M., Newton-Payne, S & Johnston A. W. B. DddW, a third
900 DMSP lyase in a model Roseobacter marine bacterium, *Ruegeria pomeroyi* DSS-3.
901 *ISME J.* **6**, 223-226 (2012).
902 30. Curson, A. R. J., Sullivan, M. J., Todd, J. D. & Johnston, A. W. B. DddY, a
903 periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse
904 species of Proteobacteria. *ISME J.* **5**, 1191-1200 (2011).
905 31. Curson, A. R., et al. Identification of genes for dimethyl sulfide production in
906 bacteria in the gut of Atlantic Herring (*Clupea harengus*). *ISME J*, **4**, 144-146
907 (2010).
908 32. Li, C.Y. et al. A novel ATP dependent dimethylsulfoniopropionate lyase in bacteria
909 that releases dimethyl sulfide and acryloyl- CoA. *elife* **10**, 1-22 (2021).
910 33. Curson, A. R. J., Rogers, R., Todd, J. D., Brearley, C. A. & Johnston, A. W. B.
911 Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the
912 climate-changing gas dimethylsulfide in several marine *alpha-proteobacteria* and
913 rhodobacter sphaeroides. *Environ. Microbiol.* **10**, 757-767 (2008).
914 34. Uria, A., Shifra, B.-D., Ester, F., Yishai, L., Dan, S. T. & Assaf, V. Identification of
915 the algal dimethyl sulfide–releasing enzyme: A missing link in the marine sulfur
916 cycle. *Science* **348**, 1466-1469 (2015).
917 35. Todd, J. D. et al. Structural and regulatory genes required to make the gas dimethyl
918 sulfide in bacteria. *Science* **315**, 666-669 (2007).
919 36. Erinn, C. H. et al. Bacterial Taxa That Limit Sulfur Flux from the Ocean. *Science*
920 **314**, 649-651 (2006).
921 37. Gage, D. A. et al. A new route for synthesis of dimethylsulphoniopropionate in
922 marine algae. *Nature* **387**, 891-894 (1997).
923 38. Summers, P. S. et al. Identification and stereospecificity of the first three enzymes
924 of 3-dimethylsulfoniopropionate biosynthesis in a chlorophyte algal. *Plant Physiol.*
925 **116**, 369-378 (1998).
926 39. Curson, A. R. J. et al. Dimethylsulfoniopropionate biosynthesis in marine bacteria
927 and identification of the key gene in this process. *Nat. Microbiol.* **2**, 17009 (2017).

928 40. Kageyama, H, Tanaka, Y., Shibata, A., Waditee-Sirisattha, R. & Takabe, T.
929 Dimethylsulfoniopropionate biosynthesis in a diatom *Thalassiosira pseudonana*:
930 identification of a gene encoding MTHB-methyltransferase. *Arch. Biochem.*
931 *Biophys.* **645**, 100-106 (2018).
932 41. Hentschke, G.S., et al. *Zarconia navalis* gen. nov., sp. nov., *Romeriopsis navalis*
933 gen. nov., sp. nov. and *Romeriopsis marina* sp. nov., isolated from inter- and subtidal
934 environments from northern Portugal. *Int J Syst Evol Microbiol*. **72**, 1-12 (2022).
935 42. Eusebio, N. et al. Distribution and diversity of dimetal-carboxylate halogenases in
936 cyanobacteria. *BMC Genomics.* **22**, 633 (2021).
937 43. Voss, M., et al. The marine nitrogen cycle: recent discoveries, uncertainties and the
938 potential relevance of climate change. *Philos Trans R Soc Lond B Biol Sci*, **368**, 1-
939 10 (2013).
940 44. Giaever, H. M., Styrvold, O. B., Kassen, I. & Strøm, A. R. Biochemical and genetic
941 characterization of osmoregulatory trehalose synthesis in *Escherichia coli*. *J.*
942 *Bacteriol.* **170**, 2841-2849 (1988).
943 45. Keeling, P. J. et al. The Marine Microbial Eukaryote Transcriptome Sequencing
944 Project (MMETSP): illuminating the functional diversity of eukaryotic life in the
945 oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).
946 46. Keller, M. D. Dimethyl sulfide production and marine phytoplankton: the
947 importance of species composition and cell size. *Biol. Oceanogr.* **6**, 375-382 (1989).
948 47. Corn, M., Belviso, S., Partensky, F., Simon, N., & Christaki, U. Origin and
949 importance of picoplanktionic DMSP. 1996.
950 48. Demir-Hilton, E. et al. Global distribution patterns of distinct clades of the
951 photosynthetic picoeukaryote *Ostreococcus*. *ISME J.* **5**, 1095-1107 (2011).
952 49. Schaffelke, B. et al. Blooms of *Chrysocystis fragilis* on the Great Barrier Reef.
953 *Coral Reefs*, 2004.
954 50. Gobler, C.J. et al. Niche of harmful alga *Aureococcus anophagefferens* revealed
955 through ecogenomics. *Proc Natl Acad Sci U S A*. **108**, 4352-4357(2011).
956 51. Gobler, C.J. & Sunda, W.G. Ecosystem disruptive algal blooms of the brown tide
957 species, *Aureococcus anophagefferens* and *Aureoumbra lagunensis*. *Harmful Algae.*

14, 36-45 (2012).

52. Guerin, N. et al. Genomic adaptation of the picoeukaryote *Pelagomonas calceolata* to iron-poor oceans revealed by a chromosome-scale genome sequence. *Commun Biol.* **5**, 1-14 (2022).
53. McParland, E. L. et al. DMSP synthesis genes distinguish two types of DMSP producer phenotypes. *Environ Microbiol.* **23**, 1656-1669 (2021).
54. Hanson, A. D., Rivoal, J., Paquet, L. & Gage, D. A. Biosynthesis of 3-dimethylsulfoniopropionate in *Wollastonia biflora* (L.) DC. (Evidence that S-methylmethionine is an intermediate). *Plant Physiol.* **105**, 103-110 (1994).
55. James, F., Paquet, L., Sparace, S. A., Gage, D. A. & Hanson, A. D. Evidence implicating dimethylsulfoniopropionaldehyde as an intermediate in dimethylsulfoniopropionate biosynthesis. *Plant Physiol.* **108**, 1439-1448 (1995).
56. Keller, M. D., Bellows, W. K., & Guillard, R. R. L. Dimethyl Sulfide Production in Marine Phytoplankton. *American Chemical Society*. 167-182 (1989).
57. Li, C.Y., et al. Mechanistic insights into the key marine dimethylsulfoniopropionate synthesis enzyme DsyB/DSYB. *mLife*. **1**,114-130 (2022).
58. Villar, E. et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res.* **46**, 289-295 (2018).
59. Sunagawa, S. et. al. Structure and function of the global ocean microbiome. *Science* **348**, 1-9 (2015).
60. Zhao, J.-Y. et al. Green Tides in the Yellow Sea Promoted the Proliferation of Pelagophyte *Aureococcus anophagefferens*. *Environmental Science & Technology.* **56**, 3056-3064 (2022).
61. Wang, Y. et al. Coastal eutrophication in China: Trend, sources, and ecological effects. *Harmful Algae.* **107**, 1-13 (2021).
62. Van Alstyne, K. L., Koellermeier, L. & Nelson, T. A. Spatial variation in dimethylsulfoniopropionate (DMSP) production in *Ulva lactuca (*Chlorophyta*)* from the Northeast Pacific. *Marine Biology.* **150**, 1127-1135 (2006).
63. Kessler, R. W. et al. Macroalgal-bacterial interactions: Role of dimethylsulfoniopropionate in microbial gardening by *Ulva* (Chlorophyta). *Mol*

988 *Ecol.* **27**, 1808-1819 (2018).
989 64. Carrion, O. et al. Function and wide distribution of DMSOP cleaving enzymes in
990 marine organisms. https://doi.org/10.21203/rs.3.rs-2412526/v1(2022).
991 65. González, J. M., Whitman, W. B., Hodson, R. E. & Moran, M. A. Identifying
992 numerically abundant culturable bacteria from complex communities: an example
993 from a lignin enrichment culture. *Appl. Environ. Microbiol.* **62**, 4433-4440 (1996).
994 66. Rippka, R. Isolation and purification of cyanobacteria. *Methods in Enzymology* **167**,
995 3-27 (1988).
996 67. Beringer, J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.*
997 **84**, 188-198 (1974).
998 68. Staskawicz, B., Dahlbeck, D., Keen, N. & Napoli, C. Molecular characterization of
999 cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv.
1000 *glycinea*. *J. Bacteriol.* **169**, 5789-5794 (1987).
1001 69. Kuek, F. W. I. et al. DMSP Production by Coral-Associated Bacteria. *Frontiers in*
1002 *Marine Science*, **9**, 1-12, 2022..
1003 70. Peng, M. et al. Structure-Function Analysis Indicates that an Active-Site Water
1004 Molecule Participates in Dimethylsulfoniopropionate Cleavage by DddK. *Appl*
1005 *Environ Microbiol*. **85**, 1-11, (2019).
1006 71. Peng, M. et al. Insights into methionine S-methylation in diverse organisms. *Nat.*
1007 *Commun.* **13**, 1-15, (2022).
1008 72. Michael, W. P. A new mathematical model for relative quantification in real-time
1009 RT–PCR. *Nucleic Acids Research.* **29**, 2005-2007, (2001).
1010 73. Figurski, D. H. & Helimshi, H. D. Replication of an origin-containing derivative of
1011 plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Nati. Acad*.
1012 *Sc*. **76**, 1648-1652 (1979).
1013 74. Sudina, A. et al. Detection of glycosylase, endonuclease and methyltransferase
1014 enzyme activities using immobilized oligonucleotides. *IUBMB Life*. **56**, 139-143
1015 (2004).
1016 75. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular
1017 evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**,

1018 1547-1549 (2018).
1019 76. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for
1020 phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, 293-296 (2021).
1021 77. Liu, J. et al. Oceanospirillales containing the DMSP lyase DddD are key utilisers of
1022 carbon from DMSP in coastal seawater. *Microbiome* **10**, 110 (2022).
1023 78. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs.
1024 *Bioinformatics* **29**, 2487-2489 (2013).
1025 79. Team, R. C. R: a language and environment for statistical computing. *MSOR*
1026 *connections.* **1**, **(**2014).

Figures

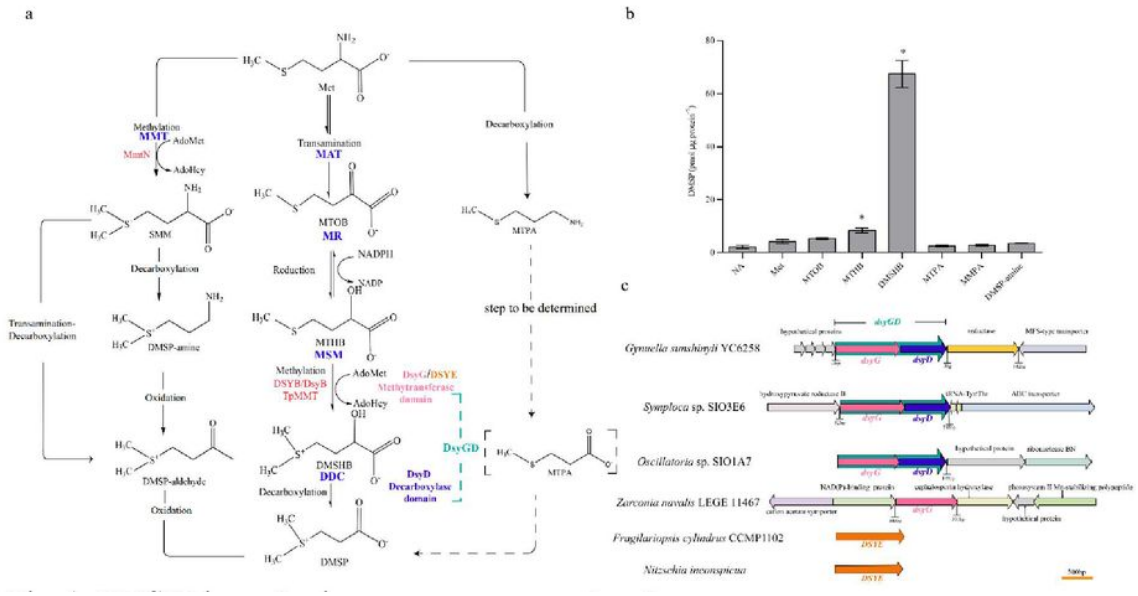

Figure 1

DMSP biosynthesis genes, enzymes and pathways

a, predicted pathways for DMSP biosynthesis in some higher plants with MMT and bacteria containing MmtN (left); algae, bacteria and corals with DSYB/DsyB, DsyGD/DsyG, DSYE and/or TpMMT2 (middle); and the dinoflagellate Crypthecodinium (right). b, DMSP production by G. sunshinyii YC6258 when incubated with or without (control) different DMSP synthesis intermediates (0.5 mM) in MBM medium (with 10 mM succinate as carbon source, 10 mM NH4Cl as nitrogen source). Three biological replicates were used for each condition. Error bars represent standard deviation (n = 3). Asterisks denote when DMSP production differences were significant ($p<0.05$) compared with controls (without intermediates). c, Genomic location of dsyGD/dsyG from bacteria and DSYE from algae. dsyG (S-methyltransferase) is indicated by a pink arrow, dsyD (decarboxylase) by a blue arrow, the translation fusion of dsyG to dsyD by a turquoise arrow, and DSYE by an orange arrow. For Oscillatoria sp. SIO1A7, dsyGD is at the start of the contig. Example single domain DSYE genes from diatoms are included for size comparison.

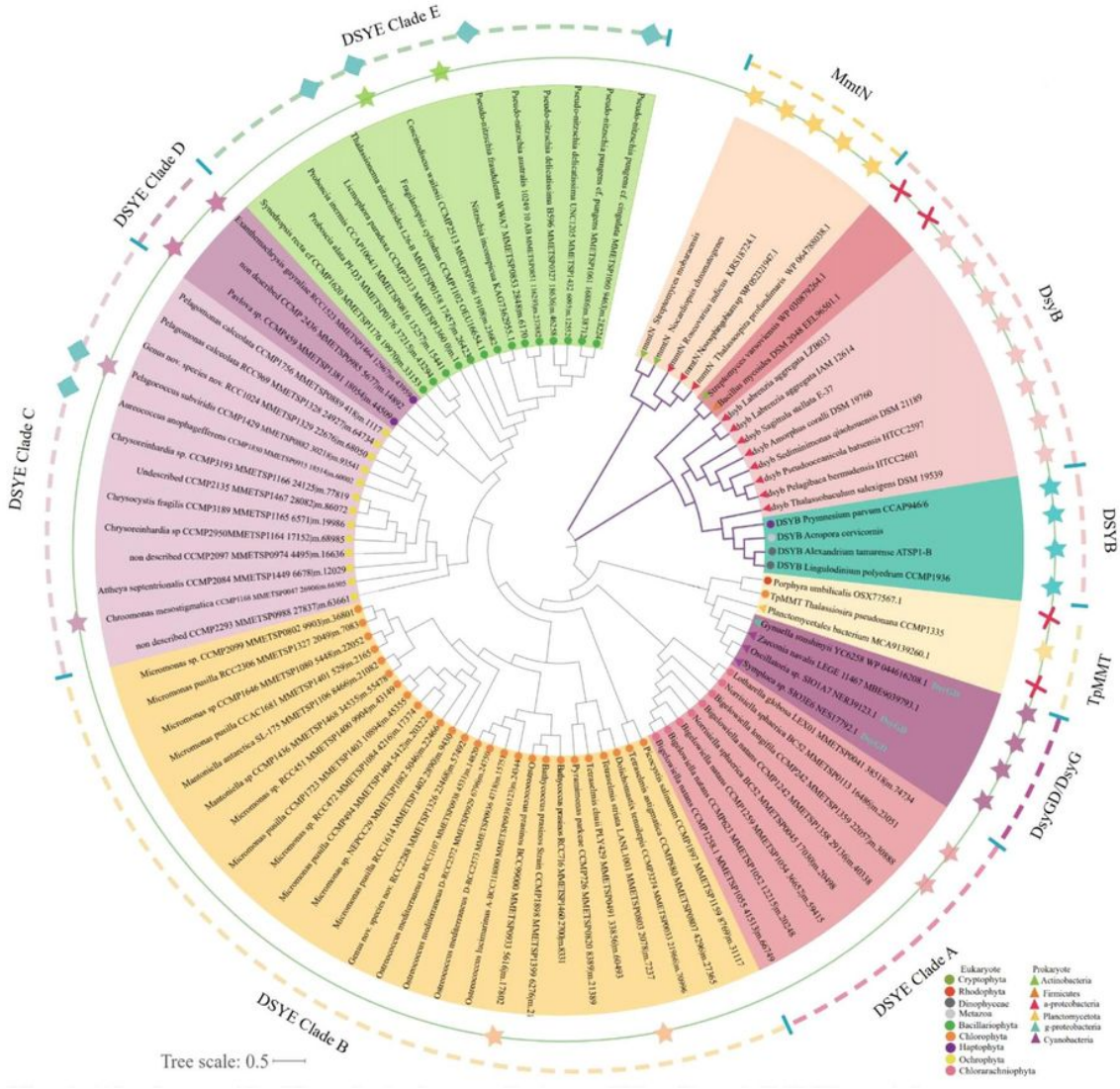

Figure 2

Maximum-likelihood phylogenetic tree of DsyG and DSYE proteins.

The tree was constructed in Mega X using proteins previously shown to have the expected S-methyltransferase enzyme activity in DMSP synthesis pathways or not[1-3], and from this study, together with those retrieved from The Marine Microbial Eukaryote Transcriptome Sequencing Project. Where proteins were multidomain (DsyGD, indicated on the tree), only the DsyG S-methyltransferase domain was analysed here. Ratified proteins from this and previous studies are labeled as ( ), and those which lacked the expected enzyme activity are labeled as (××). Proteins from eukaryotes ( ) and prokaryotes ( ) are indicated with their taxonomy by colour as specified. Organisms with DSYE which also contain DSYB are labeled as ( ). Proteins identified and discussed from previous studies are marked with lavender branches.

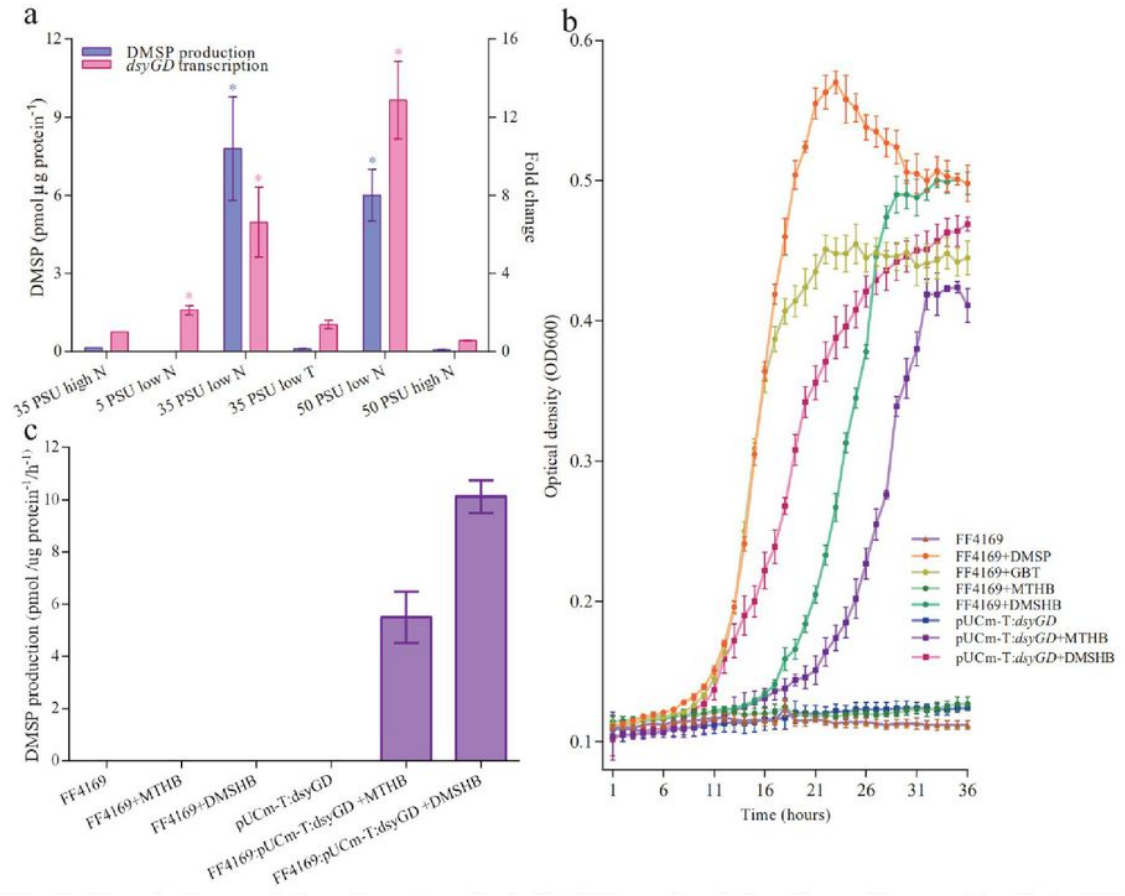

Figure 3

Regulation and functional analysis DMSP synthesis by Gynuella sunshinyii dsyGD. a, G. sunshinyii DMSP production and dysGD transcription from cultures grown under different conditions. Standard conditions were MBM minimal medium at 35 PSU and 10 mM NH4Cl, with incubation at 30°C and sampling at exponential growth phase. Three biological replicates and three technical replicates were used for each condition. Error bars indicate standard deviations (n =3). Significance was determined using a student's t-test ($P < 0.05$). b, Growth of the E. coli otsA- mutant strain FF4169 (deficient in trehalose production) and FF4169 strains containing cloned dsyGD were monitored in media containing 0.5 M NaCl and 1 mM GB, and DMSP or DMSP synthesis intermediates (MTHB and DMSHB) where indicated (n=3). 0.1 mM IPTG was also added to induce the expression of cloned dsyGD. c, DMSP production was assayed in selected cells after 36 hours incubation (n=3).

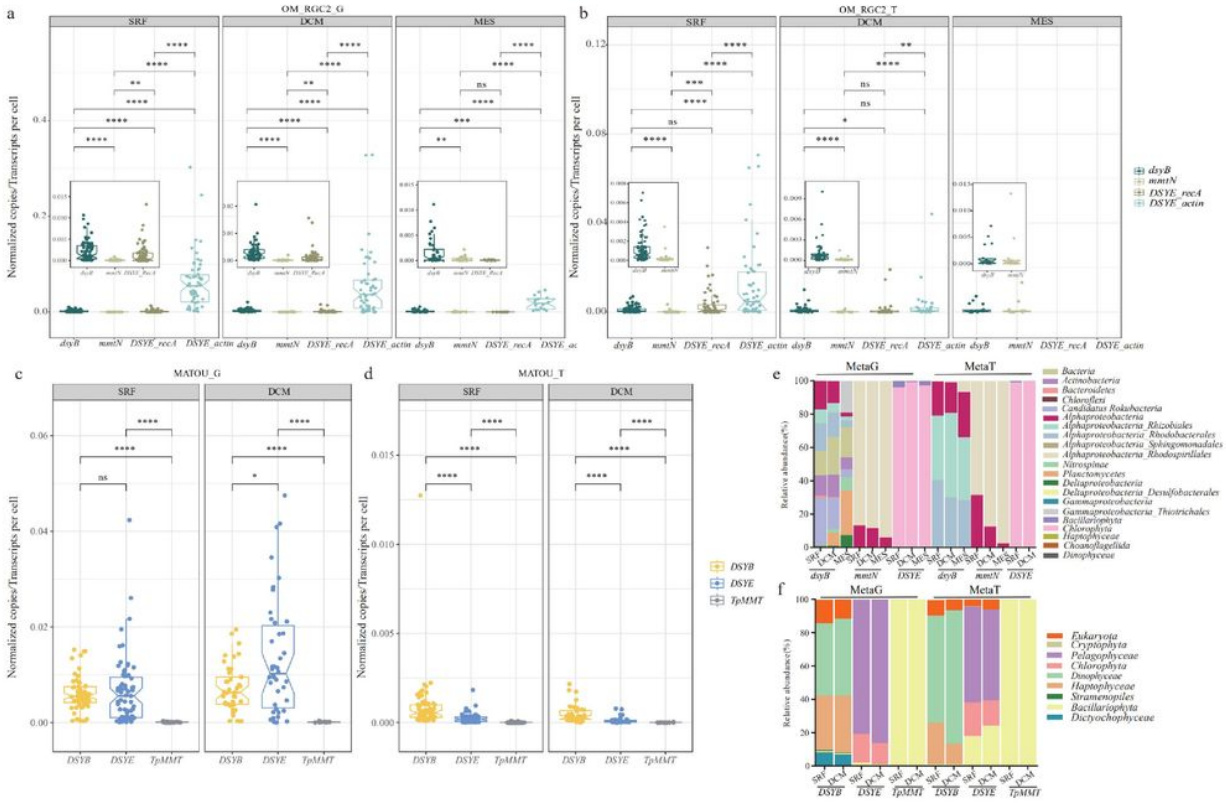

Figure 4

Distribution of DMSP synthesis genes and transcripts in the Tara Ocean dataset. (a, b), Distribution of the DMSP synthesis S-methyltransferase genes dsyB, dsyGD/dsyG, DSYE and mmtN in the Ocean Microbial Reference Gene Catalog (OM-RGC_V2 dataset) metagenomes and metatranscriptomes apportioned to bacterioplankton. (c, d), Distribution of eukaryotic DMSP synthesis S-methyltransferase genes DSYB, DSYE and TpMMT in the MATOU metagenomes and metatranscriptomes. All data were divided into surface water layer (SRF) and deep chlorophyll maximum layer (DCM) for MATOU dataset and also mesopelagic water layer (MES) for OM-RGC_V2 dataset. Significant differences ($p<0.05$) between different water layers were determined by a Wilcoxon test. MetaG, metagenomes data; MetaT, metatranscriptomes data. All prokaryotic genes are normalized to recA gene, DSYE from OM-RGC_V2 dataset were normalized to both recA and ββ-actin. All eukaryotic genes form MATOU dataset were normalized to ββ-actin. e, Taxonomic assignment of methyltransferase genes dsyB, DSYE and mmtN in the OM-RGC_V2 dataset. F, Taxonomic assignment of eukaryotic DMSP synthesis S-methyltransferase genes DSYB, DSYE and TpMMT in the MATOU dataset. metaG, metagenome; metaT, metatranscriptome. Note, no dsyG/dsyGD sequences were detected.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureandtable.pdf
- SupplementaryData1.txt
- SupplementaryTable3.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable6.xlsx
- SupplementaryTable7.xlsx
- SupplementaryFig.8.pdf
- Table.pdf