environmental microbiology

Environmental Microbiology (2021) **00**(00), 00–00 doi:10.1111/1462-2920.15393

# DMSP synthesis genes distinguish two types of DMSP producer phenotypes

**Erin L. McParland [1,2*] Michael D. Lee [3,4] Eric A. Webb [2] Harriet Alexander [5] and Naomi M. Levine [2]**

[1]*Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts.*

[2]*Department of Biological Sciences, University of Southern California, Los Angeles, California.*

[3]*Exobiology Branch, NASA Ames Research Center, Mountain View, California.*

[4]*Blue Marble Space Institute of Science, Seattle, Washington.*

[5]*Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts.*

## Summary

**Dimethylsulfoniopropionate (DMSP) is an important organic carbon and sulfur source in the surface ocean that fuels microbial activity and significantly impacts Earth's climate. After three decades of research, the cellular role(s) of DMSP and environmental drivers of production remain enigmatic. Recent work suggests that cellular DMSP concentrations, and changes in these concentrations in response to environmental stressors, define two major groups of DMSP producers: high DMSP producers that contain ≥ 50 mM intracellular DMSP and low DMSP producers that contain < 50 mM. Here we show that two recently described DMSP synthesis genes (*DSYB* and *TpMT2*) may differentiate these two DMSP phenotypes. A survey of prokaryotic and eukaryotic isolates found a significant correlation between the presence of *DSYB* and *TpMT2* genes and previous measurements of high and low DMSP concentrations, respectively. Phylogenetic analysis demonstrated that *DSYB* and *TpMT2* form two distinct clades. *DSYB* and *TpMT2* were also found to be globally abundant in *in situ* surface communities, and their taxonomic annotations were similar to those observed for isolates. The strong correlation of the *DSYB* and *TpMT2* synthesis genes with high and low producer phenotypes establishes a foundation for direct quantification of DMSP producers, enabling significantly improved predictions of DMSP *in situ*.**

Received 25 September, 2020; revised 10 December, 2020; accepted 4 January, 2021. *For correspondence. E-mail emcparland@whoi.edu; Tel.: (+1) 508 289 2926; Fax. (+1) 508 289 2834.

## Introduction

Dimethylsulfoniopropionate (DMSP) is a labile organic sulfur and carbon compound abundant throughout the global ocean (Lana *et al*., 2011). DMSP was first studied for its role in a climate feedback loop, where DMSP is cleaved to the gaseous volatile dimethylsulfide (DMS) which is oxidized into cloud condensation nuclei in the atmosphere (Charlson *et al*., 1987). DMSP is produced in the surface ocean as a metabolite by marine microbes, and is then released as dissolved organic matter and subsequently cycled through many biotic and abiotic degradation processes (Kiene *et al*., 2000; Stefels, 2000). DMSP is estimated to account for up to 13% of bacterial carbon demand, and 100% of bacterial sulfur demand (Kiene *et al*., 2000; Tripp *et al*., 2008). Substantial progress has been made in recent years to understand the degradation of DMSP, but relatively less is known about the synthesis of DMSP (Reisch *et al*., 2011; Bullock *et al*., 2017).

A diverse array of marine microbes are capable of synthesizing DMSP, including both protists and bacteria (Keller *et al*., 1989; McParland and Levine, 2019). While the algal biosynthesis pathway for DMSP production was biochemically characterized 20 years ago (Gage *et al*., 1997), functionally ratified DMSP synthesis genes were only very recently identified (Curson *et al*., 2017; Curson *et al*., 2018; Kageyama *et al*., 2018). Hence, knowledge of DMSP producers has been limited to screening isolates for cellular DMSP. Previous monoculture studies have proposed many hypotheses for the cellular role of DMSP in producers, including: compatible solute, cryoprotectant, ballasting mechanism, signalling molecule, overflow mechanism, and antioxidant (Karsten *et al*., 1996; Stefels and Van Leeuwe, 1998; Stefels, 2000; Sunda *et al*., 2002; Seymour *et al*., 2010; Lavoie *et al*., 2015; Johnson *et al*., 2016). Although evidence for

each proposed hypothesis exists, the cellular role of DMSP has not been resolved, most likely because it cannot be assumed that all producers regulate DMSP synthesis for a single cellular role (Stefels *et al*., 2007; Archer *et al*., 2010; Bucciarelli *et al*., 2013).

Recent work based on measurements of DMSP in many different isolates demonstrated that DMSP producers can be divided into two groups: high DMSP producers (HiDPs) and low DMSP producers (LoDPs). The two groups are defined by both the intracellular DMSP concentrations of producers, and changes in DMSP concentrations in response to different environmental conditions (McParland and Levine, 2019; McParland *et al*., 2020) (Fig. 1). Most well-known DMSP producers, such as dinoflagellates and haptophytes, are HiDPs with ≥ 50 mM intracellular DMSP concentrations. HiDPs do not change intracellular DMSP concentrations in response to different environmental stressors in a predictable manner, and often do not respond at all (McParland and Levine, 2019). This suggests that HiDPs constitutively produce high DMSP concentrations, independent of the environment. Significantly less research has been conducted on LoDPs due to their relatively low intracellular DMSP concentrations of < 50 mM (Keller *et al*., 1989). However, it is now known that LoDPs are phylogenetically widespread, and include many diatoms, as well as some Cyanobacteria, and heterotrophic Alphaproteobacteria (Bucciarelli *et al*., 2013; Curson *et al*., 2017; McParland and Levine, 2019). In contrast to HiDPs, LoDPs modify intracellular DMSP concentrations as a predictable function of stressed growth across many different environmental conditions (McParland *et al*., 2020). The significant phenotypic differences between HiDPs and LoDPs are predicted to result in a general dominance of HiDP contribution to *in situ* DMSP production throughout the global open ocean, and a limited contribution by LoDPs (McParland and Levine, 2019).

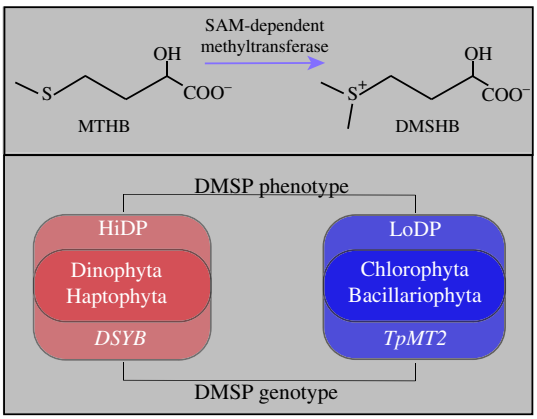


**Fig 1.** Both the *DSYB* and *TpMT2* genes are confirmed SAM-dependent methyltransferases that catalyze the third step of DMSP synthesis by methylating 4-methylthio-2-hydroxybutyrate (MTHB) to produce 4-dimethylsulfonio-2-hydroxybutyrate (DMSHB). We explored our hypothesis that the high DMSP producer (HiDP) phenotype is correlated with the *DSYB* genotype and the low DMSP producer (LoDP) phenotype is correlated with the *TpMT2* genotype.

In the few marine microbial isolates for which DMSP synthesis pathways have been investigated, most synthesize DMSP with the same methionine transamination pathway (Fig. S1A) (Gage *et al*., 1997; Lyon *et al*., 2011; Curson *et al*., 2018; Kageyama *et al*., 2018). Therefore, the observed phenotypic differences suggest that the four-step reactions of this DMSP synthesis pathway may have different regulatory mechanisms or enzymatic efficiencies in HiDPs and LoDPs. The recently described DMSP synthesis genes (the prokaryotic *dsyB*, and the eukaryotic *DSYB*, *TpMT1*, and *TpMT2*) are all S-adenosylmethionine (SAM) dependent methyltransferases that perform the third step in DMSP synthesis of adding a second methyl group to the sulfur group (Fig. 1, Fig. S1A) (Curson *et al*., 2017; Curson *et al*., 2018; Kageyama *et al*., 2018). Critically, only DMSP producers are thought to contain this methylation enzyme, whereas many non-DMSP producers harbor enzymes to perform the first two steps of the methionine transamination pathway (Ito *et al*., 2011; Curson *et al*., 2018). *dsyB* is a prokaryotic DMSP synthesis gene discovered in Alphaproteobacteria (Curson *et al*., 2017). *DSYB* is a eukaryotic homologue of *dsyB* and was found primarily in dinoflagellates and haptophytes (Curson *et al*., 2018). In contrast, *TpMT1* and *TpMT2* have only been reported in a few diatoms (Kageyama *et al*., 2018). Until now, phylogenetic exploration of DMSP synthesis was limited to knowledge of the evolutionary history of the microbes producing DMSP (Bullock *et al*., 2017; McParland et al. 2019). The phylogeny of the DMSP synthesis genes themselves could provide new insight into the evolutionary history of DMSP synthesis and improve our understanding of how the cellular role(s) of DMSP differ across taxa.

Based on the taxonomy of the isolates that the DMSP synthesis genes were discovered in, we hypothesized that *DSYB* and *TpMT1* and/or *TpMT2* may be marker genes for the HiDP and LoDP groups, respectively (Fig. 1). In the proceeding analyses, only three genes (*dsyB*, *DSYB*, and *TpMT2*) are considered as purified recombinant TpMT1 protein did not exhibit detectable methyltransferase activity (Kageyama *et al*., 2018). In this study, we conducted a survey of the presence of DMSP synthesis genes in genomes and transcriptomes of hundreds of marine microbial isolates. We compared the taxonomy of all the isolates containing at least one of the SAM-dependent methyltransferases with previous measurements of cellular DMSP and demonstrate that the two eukaryotic synthesis genes (*DSYB* and *TpMT2*) correlate with the HiDP and LoDP concentrations. Furthermore, we show that *TpMT2* amino-acid sequences

are distinct from *dsyB* and *DSYB*, with low identity (< 25%) and similarity (< 40%) (Fig. S1B). Phylogenetic analysis of *DSYB* and *TpMT2* supports our hypothesis that the two different responses of the HiDP and LoDP phenotypes to environmental stressors are associated with presence of the two different DMSP synthesis genes. Finally, we demonstrate that *DSYB* and *TpMT2* are globally abundant in *in situ* eukaryotic metatranscriptomes and could potentially serve as marker genes for DMSP production.

## Results

### *Phenotypic and genotypic surveys*

Prokaryotic and eukaryotic isolates previously classified as HiDPs or LoDPs were surveyed for the presence or absence of DMSP synthesis genes (Fig. 2, Table S1). We first describe the distribution of DMSP phenotypes across major taxonomic groups, where DMSP phenotype is defined based on previously reported intracellular DMSP concentrations (cellular concentrations normalized to cell volume). The dataset surveyed for DMSP phenotypes represents the most recent, comprehensive list of all isolates previously assayed for detectable cellular DMSP ($n = 271$) (compiled by McParland and Levine, 2019 as an update to Keller *et al*., 1989). We then present the distribution of DMSP genotypes, defined as the putative presence (or absence) of DMSP synthesis genes in transcriptomes and genomes of isolates. The DMSP genotype was primarily surveyed in the NCBI prokaryotic RefSeq genomes ($n = 1221$) (O'Leary *et al*., 2016) and the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) transcriptomes ($n = 395$) (Keeling *et al*., 2014; Johnson *et al*., 2018). If one of these databases was missing a known DMSP producer, we either searched the isolate's genome if available, or the isolate's sequences in the non-redundant NCBI nucleotide database (see Methods).

### *DMSP phenotype*

The DMSP phenotypes were previously defined as the LoDP phenotype (< 50 mM DMSP) and the HiDP phenotype (≥ 50 mM DMSP) based on measurements of intracellular DMSP in a diverse range of prokaryotes and eukaryotes (McParland and Levine, 2019). For completeness, we describe all assays of detectable DMSP, both those normalized to cell volume (intracellular DMSP) and those only normalized to cell number (cellular DMSP) (Fig. 2A, Table S1). Previously assayed prokaryotes ($n = 60$) primarily belong to the groups Alphaproteobacteria and Cyanobacteria (Fig. 2B). The majority of previously assayed isolates are eukaryotic ($n = 211$), and primarily belong to the groups Dinophyta, Haptophyta, Chlorophyta, and Bacillariophyta, but also include Pelagophyta, other Stramenopiles, Cryptophyta, and Rhizaria (Fig. 2B).

All prokaryotes with detectable intracellular DMSP concentrations exhibited a LoDP phenotype ($n = 18$), including Alphaproteobacteria, and nitrogen-fixing and non-nitrogen-fixing Cyanobacteria (Fig. 2C, Table S1). Gammaproteobacteria, specifically purple sulfur bacteria, also produced significant cellular DMSP concentrations. In general, the eukaryotic phenotypes were differentiated across major classes (Fig. 2C, Table S1). All Rhizaria exhibited a LoDP phenotype ($n = 7$). Most Stramenopiles exhibited a LoDP phenotype ($n = 21$), including all but one Bacillariophyta, and the under-sampled Pelagophyta. A small number of other Stramenopile groups exhibited a HiDP phenotype. Chlorophyta exhibited both HiDP and LoDP phenotypes ($n = 7$ HiDP, and $n = 12$ LoDP). The majority of Haptophyta exhibited a HiDP phenotype ($n = 41$). Unexpectedly, a relatively smaller proportion of Dinophyta exhibited a HiDP phenotype ($n = 19$ HiDP, and $n = 10$ LoDP). However, all Dinophyta had very high cellular concentrations of DMSP (45–1192 fmole per cell compared to the typical < 5 fmole per cell in LoDPs) (Keller *et al*., 1989).

While cellular DMSP content is most accurate when normalized to cell volume, cell volume is notoriously difficult to measure accurately, especially for the unique shapes of marine protists (Harrison *et al*., 2015). The reported cell volumes of Dinophyta LoDP phenotypes ($5 \times 10^5 - 3 \times 10^7$ $\mu m^3$) are as much as two orders of magnitude higher than other reported cell volume measurements for the same genus (Menden-Deuer and Lessard, 2000; Harrison *et al*., 2015). We therefore believe Dinophyta primarily exhibited a HiDP phenotype, and that the LoDP phenotypes were errors resulting from the cell volume measurements. This highlights the complications of accurately quantifying the normalization factor, and its influence on identifying the DMSP phenotype.

### *DMSP genotype*

We defined three different DMSP genotypes: the *dsyB*/*DSYB* (prokaryotic/eukaryotic), *TpMT2*, and *TpMT2* + *DSYB* genotypes (Fig. 2D). The *dsyB*/*DSYB* and *TpMT2* genotypes were defined by the presence of the respective gene in a surveyed isolate. The *TpMT2* + *DSYB* genotype was defined by the presence of both the *TpMT2* and *DSYB* genes in a surveyed isolate. Genotypes were only assigned if results passed the filter criteria (see Methods). Isolates were reported as not containing any of the targeted DMSP synthesis genes if no copies of the *dsyB*, *DSYB*, and *TpMT2* genes were identified that passed the filter criteria. Additionally, when apparent paralogs of the same gene (e.g. multiple copies

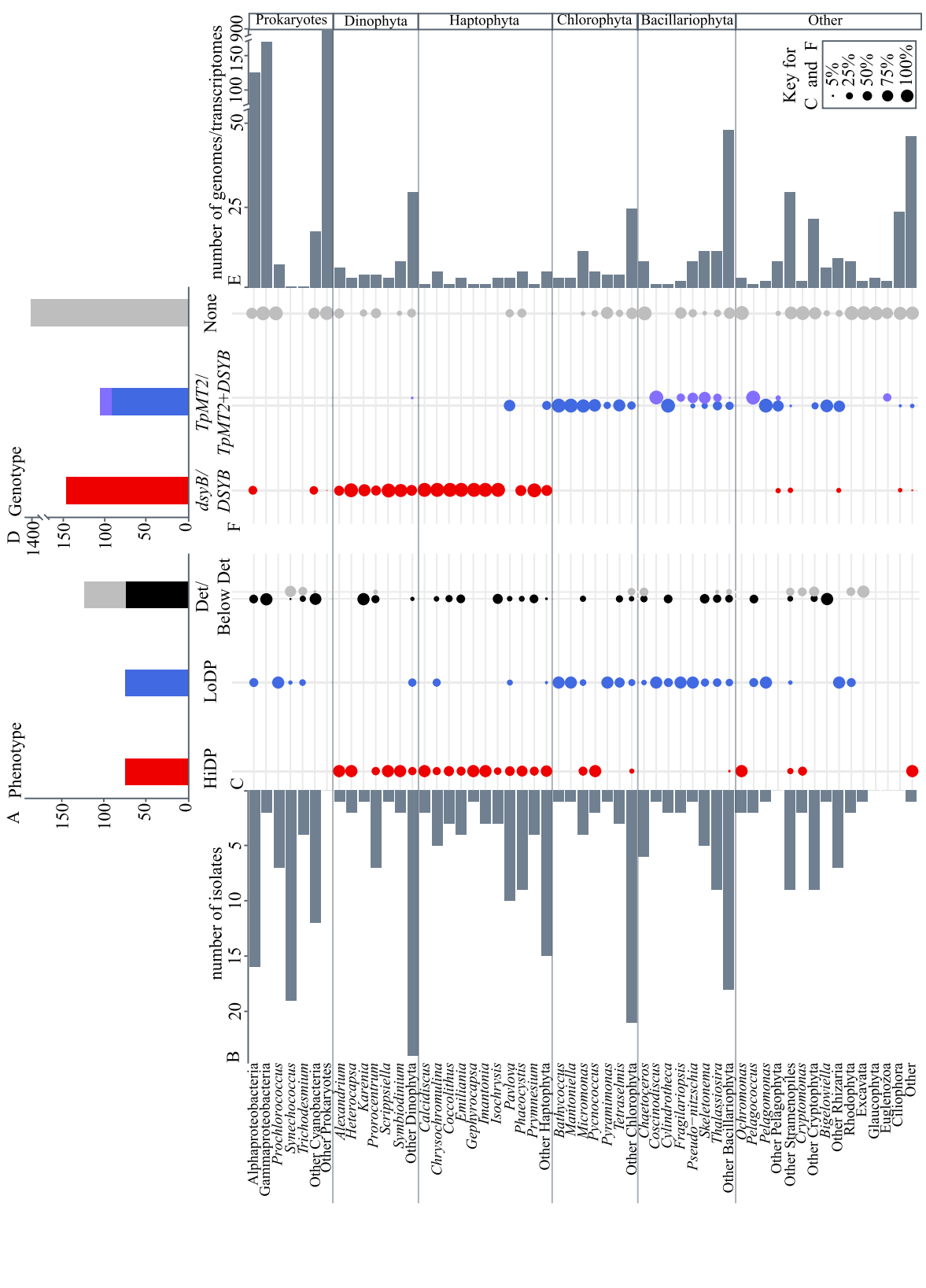


**Fig 2.** Comparison of DMSP synthesis phenotypes and genotypes. Top panels represent (A) the total assays of intracellular DMSP in HiDPs (red) or LoDPs (blue), and detected cellular DMSP (black) or below detection (grey), and (D) the total genotypes identified for *dsyB/DSYB* (red), *TpMT2* (blue) or *TpMT2 + DSYB* (purple), and no identified genotypes (grey). Side panels represent the total number of isolates in each group or genus surveyed for (B) phenotype or (E) genotype. Middle panel circles represent the percent of each group or genus assayed with (C) a HiDP, LoDP, or detected/below detection phenotype, or (F) a *dsyB/DSYB*, *TpMT2*, *TpMT2 + DSYB*, or no identified genotype.

of *DSYB*) were present within an isolate, only the sequence with the lowest e-value was retained. The genotypes were surveyed across a wide diversity of prokaryotic and eukaryotic taxa (Fig. 2E). The surveyed eukaryotic isolates contained multiple representatives of every major protistan supergroup (Burki *et al*., 2019), but the majority of isolates surveyed belonged to the Dinophyta, Haptophyta, Chlorophyta, and Bacillariophyta groups.

In an effort to limit any potential prokaryotic contamination from the eukaryotic references (Johnson *et al*., 2018), we additionally performed a recursive BLASTP search of each identified eukaryotic *TpMT2* and *DSYB* amino-acid sequence against the NCBI nonredundant database (see Methods). The returned alignments ranged widely from 0%–100% eukaryotic. A median of 13% and 6% of returned alignments for *TpMT2* and *DSYB*, respectively, were eukaryotic. This variability was not surprising given the dominance of prokaryotes in the NCBI database and that *DSYB* is considered to be a homologue of the prokaryotic *dsyB* sequence (Curson *et al*., 2018). However, the percent identity of the prokaryotic alignments (median = 36%, range = 28%–54%) was lower than the percent identity of the eukaryotic alignments (median = 48%, range = 28%–100%) (Fig. S2). Based on this analysis, only 11 sequences ($n = 10$ *DSYB*, and $n = 1$ *TpMT2*) were removed. Rather than indicating significant prokaryotic contamination, this recursive search primarily demonstrates the similarity of *dsyB*, *DSYB*, and *TpMT2* with other methyltransferases.

Across the diverse collection of prokaryotic genomes surveyed, *dsyB* was the only DMSP synthesis gene identified ($n = 69$) (Fig. 2F). *dsyB* genotypes were primarily found in known DMSP producing Alphaproteobacteria, as well as the Cyanobacteria *Nostoc* and *Anabaena*. Neither the nitrogen-fixing Cyanobacteria *Trichodesmium*, nor any of the surveyed Gammaproteobacteria contained any significant alignments to DMSP synthesis genes. A small number of *dsyB* genotypes ($n = 16$) were also found in prokaryotic groups that, to our knowledge, have not been assayed for cellular DMSP (Acidobacteria, Actinobacteria, Deltaproteobacteria, Firmicutes).

Across all eukaryotic isolates surveyed, 61% of the major taxonomic groups, and 48% overall, exhibited either a *DSYB*, *TpMT2*, or *TpMT2* + *DSYB* genotype ($n = 192$) (Table S1). The prevalence of the three genotypes varied within each major eukaryotic group: 69% of Dinophyta, 90% of Haptophyta, 59% of Chlorophyta, 48% of Bacillariophyta, 91% of Pelagophyta, and 80% of Rhizaria isolates exhibited a DMSP genotype. The DMSP genotypes were taxonomically differentiated across the major eukaryotic groups (Fig. 2F). Dinophyta exclusively exhibited a *DSYB* genotype ($n = 38$), except one *TpMT2* + *DSYB* genotype in *Pyrodinium bahamense*. The majority of Haptophyta exhibited a *DSYB* genotype ($n = 23$), though the small number of Haptophyta exhibiting a *TpMT2* genotype ($n = 4$) primarily belonged to the Pavlovaceae family. Chlorophyta exclusively exhibited a *TpMT2* genotype ($n = 32$). Bacillariophytaexhibited either a *TpMT2* or a *TpMT2* + *DSYB* genotype (n = 45). Except for one *DSYB* genotype in each group, both Pelagophyta and Rhizaria exclusively exhibited a *TpMT2* genotype (Fig. 2F). The *TpMT2* + *DSYB* genotype was most prevalent in Bacillariophyta. We confirmed that the similarity of the corresponding *DSYB* and *TpMT2* sequences in each isolate with a *TpMT2* + *DSYB* genotype was low by comparing each pair with BLASTP. All sequence pairs of *TpMT2* + *DSYB* genotypes exhibited < 40% sequence similarity.

The majority of isolates (> 78%) in other eukaryotic groups surveyed (including Cryptophyta, and other Stramenopiles) did not contain the targeted DMSP synthesis genes (Fig. 2F). Although never assayed for DMSP phenotypes, a small number of Ciliophora ($n = 3$) unexpectedly exhibited a *DSYB* or *TpMT2* genotype (Fig. 2F). However, these isolates were fed eukaryotic prey with *DSYB* or *TpMT2* genotypes (Keeling *et al*., 2014), suggesting the genotype may belong to the prey rather than Ciliophora.

### *Comparison of phenotype and genotype*

All previously confirmed prokaryotic DMSP producers exhibited a LoDP phenotype and a *dsyB* genotype (Fig. 2). In eukaryotic isolates, phenotypes and genotypes were correlated within taxonomic groups (Fig. 2). Dinophyta and Haptophyta primarily exhibited a HiDP phenotype and a *DSYB* genotype. The majority of Chlorophyta exhibited a LoDP phenotype and exclusively exhibited a *TpMT2* genotype. All Bacillariophyta exhibited a LoDP phenotype (except $n = 1$; *Melosira nummuloides*) and either a *TpMT2* or *TpMT2* + *DSYB* genotype. Although relatively fewer phenotypes have been assayed for Pelagophyta ($n = 2$) and Rhizaria ($n = 7$), both groups primarily exhibited a LoDP phenotype and a *TpMT2* genotype (except $n = 1$ *DSYB* genotype in each group) (-Table S1). Across all isolates with both a previously reported phenotype and an identified genotype ($n = 62$), the HiDP phenotype and *DSYB* genotype were significantly correlated, and the LoDP phenotype and *TpMT2* or *TpMT2* + *DSYB* genotype were significantly correlated (Pearson's $r$ rank correlation, $r = 0.64$, $p < 0.05$).

Although the DMSP phenotype and genotype were significantly correlated in isolates with identified DMSP synthesis genes, many known DMSP producers did not contain any of the targeted DMSP synthesis genes. Specifically, across all eukaryotic isolates surveyed for a DMSP genotype, 32% of Dinophyta, 10% of Haptophyta,

41% of Chlorophyta, and 52% of Bacillariophyta isolates lacked a significant alignment to *dsyB*, *DSYB*, and *TpMT2* (Table S1). Some of these isolates may truly be non-DMSP producers, such as the heterotrophic dinoflagellate *Oxyrrhis marina*, which has no detectable cellular DMSP (Keller *et al*., 1989; Curson *et al*., 2018). However, some isolates for which cellular DMSP has been detected appeared to be missing all currently identified synthesis genes, such as the HiDP autotrophic dinoflagellate *Prorocentrum micans* and the LoDP diatom *Ditylum brightwelli* (Keller *et al*., 1989).

### *Phylogeny of DMSP synthesis genes*

All identified prokaryotic *dsyB*, and eukaryotic *DSYB* and *TpMT2* amino-acid sequences were used to build a phylogeny of DMSP synthesis genes. While Curson *et al*. (2017, 2018) previously presented phylogenetic trees of *dsyB* and *DSYB*, the phylogeny of *TpMT2* has not been explored in detail (Kageyama *et al*., 2018). Additionally, the phylogenies of *DSYB* and *TpMT2* have not been previously compared. Two midpoint-rooted phylogenies are presented: one with only sequences from isolates with a corresponding DMSP phenotype (Fig. 3), and one with all sequences (Fig. S3). Bootstrap values in text refer to the phylogeny with all sequences (Fig. S3). The *dsyB*/*DSYB* and *TpMT2* sequences formed two distinct clades and, in general, the *DSYB* cluster contained only HiDP phenotypes, and the *TpMT2* cluster contained only LoDP phenotypes (Fig. 3).

The clustering of *TpMT2* was well-supported (bootstrap = 100%) and was divided into two sub-clades dominated by Bacillariophyta and Chlorophyta (Fig. 3, Fig. S3). The *TpMT2* sequences of Bacillariophyta associated with a *TpMT2* + *DSYB* genotype were found in both sub-clades (purple stars, Fig. S3). Sub-clade T1 (bootstrap = 56%) included all isolates of the Bacillariophyta genus *Skeletonema* and *Thalassiosira*, and the Chlorophyta genus *Pycnococcus*. Sub-clade T2 (bootstrap = 83%) included all isolates of the Bacillariophyta genus *Pseudo-nitzschia*, as well as all of the other Chlorophyta, and all Pelagophyta and Rhizaria. The Ciliophora isolate (*Myrionecta rubra*) clustered with their Cryptophyta prey, indicating the Ciliophora *TpMT2* may belong to the prey, not the ciliate (Fig. S3).

The *dsyB*/*DSYB* cluster formed three major sub-clades each comprised of similar taxonomies and phenotypes (Fig. 3, Fig. S3). The Alphaproteobacteria *dsyB* sub-clade and LoDP *DSYB* sub-clade only contained LoDP phenotypes (Fig. 3). The Alphaproteobacteria *dsyB* genotypes clustered together (bootstrap = 58%) separate from the other prokaryotic *dsyB* genotypes (bootstrap = 61%) (Fig. S3). The LoDP *DSYB* sub-clade was well-supported (bootstrap = 81%) (Fig. 3) and included known LoDP Bacillariophyta as well as more distantly related isolates of Dictyochophycea (sister group of Bacillariophyta (Derelle *et al*., 2016)), an unidentified Chlorophyta, and a Prymnesiophyceae (Fig. S3). All *DSYB* sequences of Bacillariophyta with a *TpMT2* + *DSYB* genotype clustered together in the LoDP *DSYB* sub-clade (bootstrap = 100%) (purple stars Fig. 3, Fig. S3).

The HiDP *DSYB* sub-clade is well-supported (bootstrap = 89%) and contained primarily HiDP phenotypes, except for one known LoDP isolate of *Chrysochromulina tobin* (Curson *et al*., 2018) (Fig. 3, Fig. S3). The HiDP *DSYB* sub-clade is dominated by Dinophyta and Haptophyta (Fig. 3), and, in general, genera of these groups cluster in accordance with current understanding of these groups' phylogeny (Edvardsen *et al*., 2011; Janouškovec *et al*., 2017). The HiDP *DSYB* sub-clade also includes a few isolates from the Pelagophyta, Dictyochophycea, Ciliophora, and Rhizaria groups (Fig. S3). Interestingly, the corresponding *DSYB* sequence for the only *TpMT2* + *DSYB* genotype in the Pelagophyta (*Pelagococcus subviridis*), clustered with the Haptophyta, rather than the LoDP *DSYB* sub-clade (Fig. 3). In addition, as observed for *TpMT2*, the Ciliophora *DSYB* clustered with the *DSYB* of their prey (Fig. S3).

### *DMSP synthesis genes* in situ

The global abundance and distribution of *DSYB* and *TpMT2* in natural surface communities were assessed using the Ocean Gene Atlas (Villar *et al*., 2018) to search the Marine Atlas of *Tara* Ocean Unigenes (MATOU), which contains millions of representative eukaryotic transcribed sequences that were assembled from metatranscriptomes and clustered by similarity (Carradec *et al*., 2018). MATOU taxonomy annotations primarily referenced the MMETSP transcriptomes described above (Keeling *et al*., 2014), as well as the UniRef90 database of protein sequences (Suzek *et al*., 2015), and the *Tara* Oceans Single-cell Amplified Genomes (Seeleuthner *et al*., 2018). The abundance and distribution of the prokaryotic *dsyB* gene were previously described (Curson *et al*., 2017). Here, we describe the prevalence and taxonomic annotation of *DSYB* and *TpMT2* unigenes in *in situ* eukaryotic microbial communities.

The *DSYB* and *TpMT2* unigenes were present throughout the global surface ocean and the combined abundance (sum of *DSYB* and *TpMT2* unigene abundances) was relatively constant ($6 \times 10^{-5} \pm 4 \times 10^{-5}$ RPKM) (mean ± 1 SD) (Fig. 4). Globally, *DSYB* unigenes were most abundant (68 ± 13% of combined abundance) and were significantly greater than *TpMT2* unigenes across all stations (paired samples Wilcoxon test, $p < 0.05$) (Fig. 4). The majority of *in situ DSYB* and

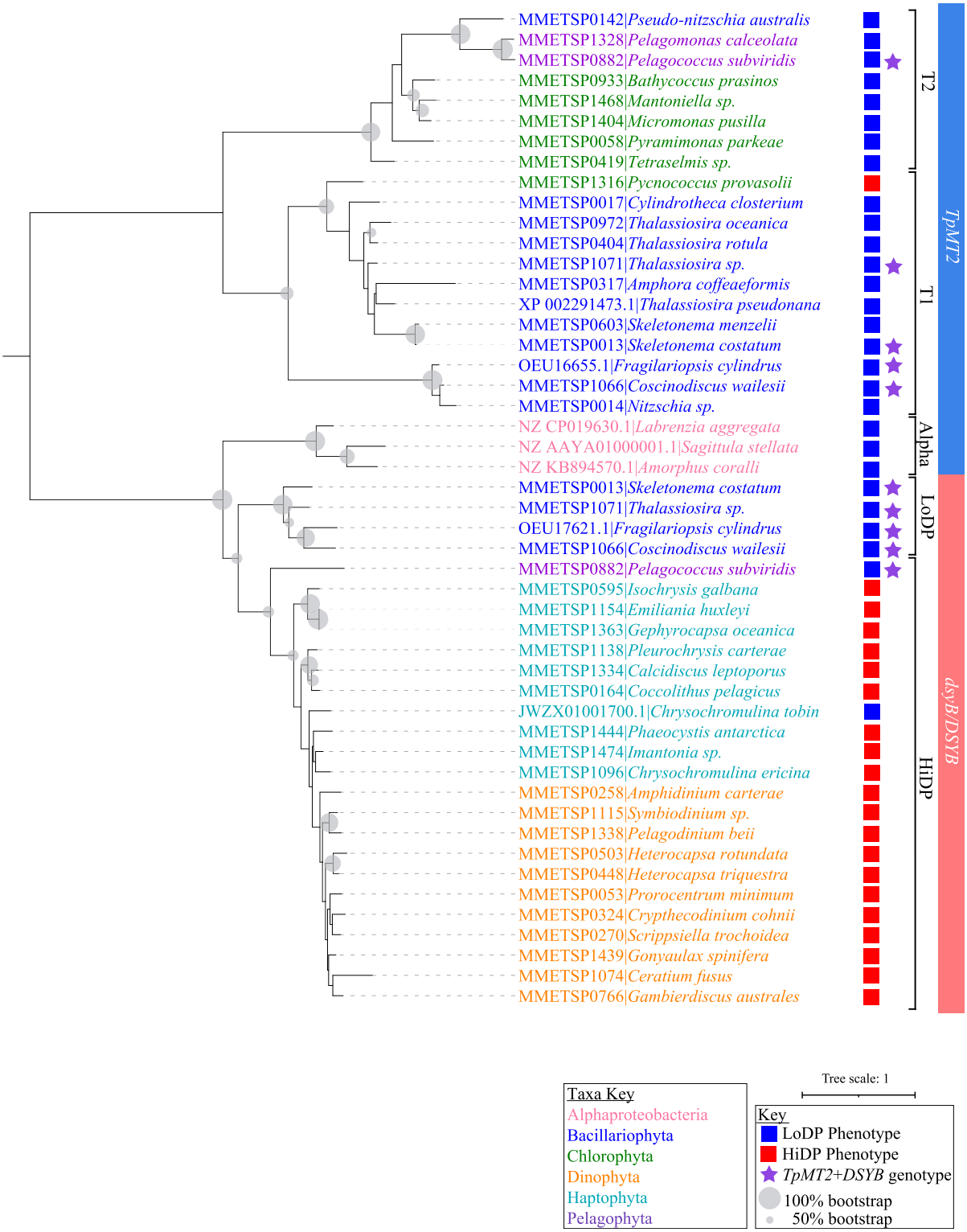


**Fig 3.** Phylogeny of *dsyB*, *DSYB*, and *TpMT2* built with RAxML. The phylogeny was midpoint-rooted. Only sequences associated with a previously assayed phenotype are presented. (The phylogeny with all sequences is presented in Supp Fig. 3). Grey circles reflect any bootstrap values >50%. Blue and red squares represent LoDP and HiDP phenotypes, respectively. Purple stars denote isolates with a *TpMT2* + *DSYB* genotype. Note that *Gonyaulax spinifera* and *Ceratium fusus* are labelled by their assumed HiDP phenotype (see main text).

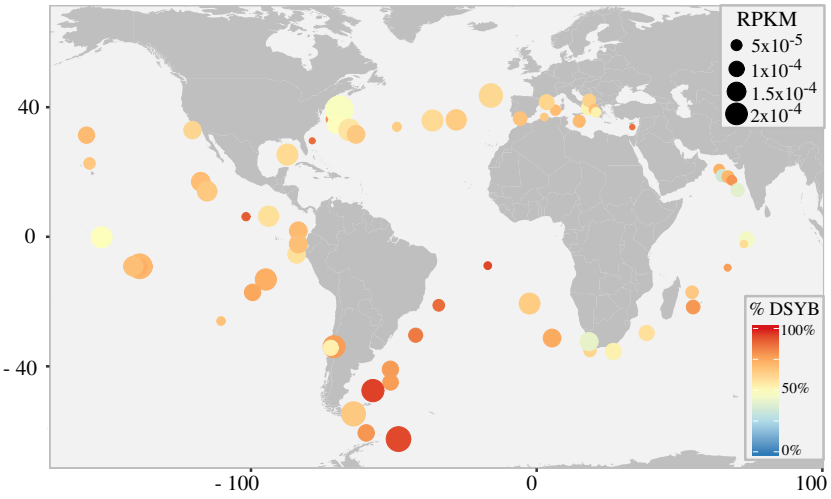


**Fig 4.** The combined normalized abundance of *DSYB* and *TpMT2* (reads per kilobase covered per million of mapped read). Circle colors represent the relative contribution of *DSYB* or *TpMT2* to the sum (100% *DSYB* = red, 100% *TpMT2* = blue).

*TpMT2* unigenes across all sites (79 ± 9% of combined abundance) were attributed to the groups Dinophyta, Haptophyta, Bacillariophyta, Chlorophyta, and Pelagophyta (Fig. 5). Other taxonomic groups annotated to *DSYB* or *TpMT2* unigenes included Cryptophyta, Euglenozoa, other Stramenopiles, and unclassified Eukaryota.

The annotated taxonomy of *DSYB* and *TpMT2* unigenes in natural communities corresponded with the taxonomy of known HiDP and LoDP phenotypes (Fig. 5). The DMSP synthesis unigenes annotated as Dinophyta and Haptophyta were almost exclusively *DSYB* (99 ± 1%, 96 ± 7%, respectively) (Fig. 5A and B). The unigenes annotated as Chlorophyta and Pelagophyta were exclusively *TpMT2* (100 ± 0%) (Fig. 5C). The unigenes annotated as Bacillariophyta were primarily *TpMT2* (85 ± 13%) (Fig. 5D). While 21% of the Bacillariophyta isolates exhibited a *TpMT2* + *DSYB* genotype (Fig. 2F), the *in situ* taxonomy was based on that of unigenes recovered from mixed communities, not individual transcriptomes. Therefore, it was not possible to determine if *DSYB* and *TpMT2* unigenes annotated as Bacillariophyta belonged to a common Bacillariophyta transcriptome containing both genes (i.e. a *TpMT2* + *DSYB* genotype).

*DSYB* unigenes annotated as Dinophyta dominated the combined unigene abundance (Fig. 5A), and were significantly greater than *DSYB* unigenes annotated as Haptophyta globally (paired samples Wilcoxon test, $p < 0.05$). However, in the southern Atlantic and Southern Ocean, as well as one North Atlantic station, *DSYB* unigenes annotated as Haptophyta surpassed those of Dinophyta (Fig. 5A and B) (paired samples Wilcoxon test, $p = 0.06$). The mean contribution of *TpMT2* unigenes annotated as Chlorophyta/Pelagophyta and Bacillariophyta to combined unigene abundance at each station was similar (7 ± 9% and 10 ± 6%, respectively) and not significantly different between the two groups (paired samples Wilcoxon test, $p = 0.08$). The contribution of *TpMT2* unigenes to the combined unigene abundance was always less than that of *DSYB*, except for two stations in the Arabian Sea where *TpMT2* unigenes annotated as Chlorophyta/Pelagophyta comprised 44 ± 5% of the combined unigene abundance (Fig. 5C) and surpassed the *DSYB* Dinophyta contribution.

## Discussion

Prediction of *in situ* DMSP production is challenging, and the cellular roles of DMSP in producers remain unknown. However, recent work suggests that the two groups of DMSP producers, HiDPs and LoDPs, regulate DMSP synthesis for different cellular roles (McParland *et al*., 2020). Accurate quantification of these two groups, particularly the HiDPs which appear to produce DMSP constitutively, could result in more accurate predictions of *in situ* DMSP production (Lyon *et al*., 2011; McParland and Levine, 2019). Here, surveys of both isolates and *in situ* communities show that the presence of two recently described DMSP synthesis genes, *DSYB* and *TpMT2*, correlates with the HiDP and LoDP phenotypes (Fig. 2). While *DSYB* and *TpMT2* are both SAM dependent methyltransferases, phylogenetic analyses revealed two distinct clades (Fig. 3, Fig. S3). We propose that *DSYB* and *TpMT2* can be used to identify and distinguish the two phenotypic groups *in situ*, and further hypothesize that these genes are in-part responsible for the observed differential responses in DMSP synthesis exhibited by HiDPs and LoDPs under environmental stress.

In general, eukaryotic *DSYB* genotypes correlated with HiDP phenotypes, and *TpMT2* genotypes with LoDP phenotypes (Fig. 2). Dinophyta and Haptophyta isolates primarily exhibited *DSYB* genotypes. Chlorophyta exclusively exhibited *TpMT2* genotypes, and Bacillariophyta isolates only exhibited *TpMT2* or *TpMT2* + *DSYB* genotypes.

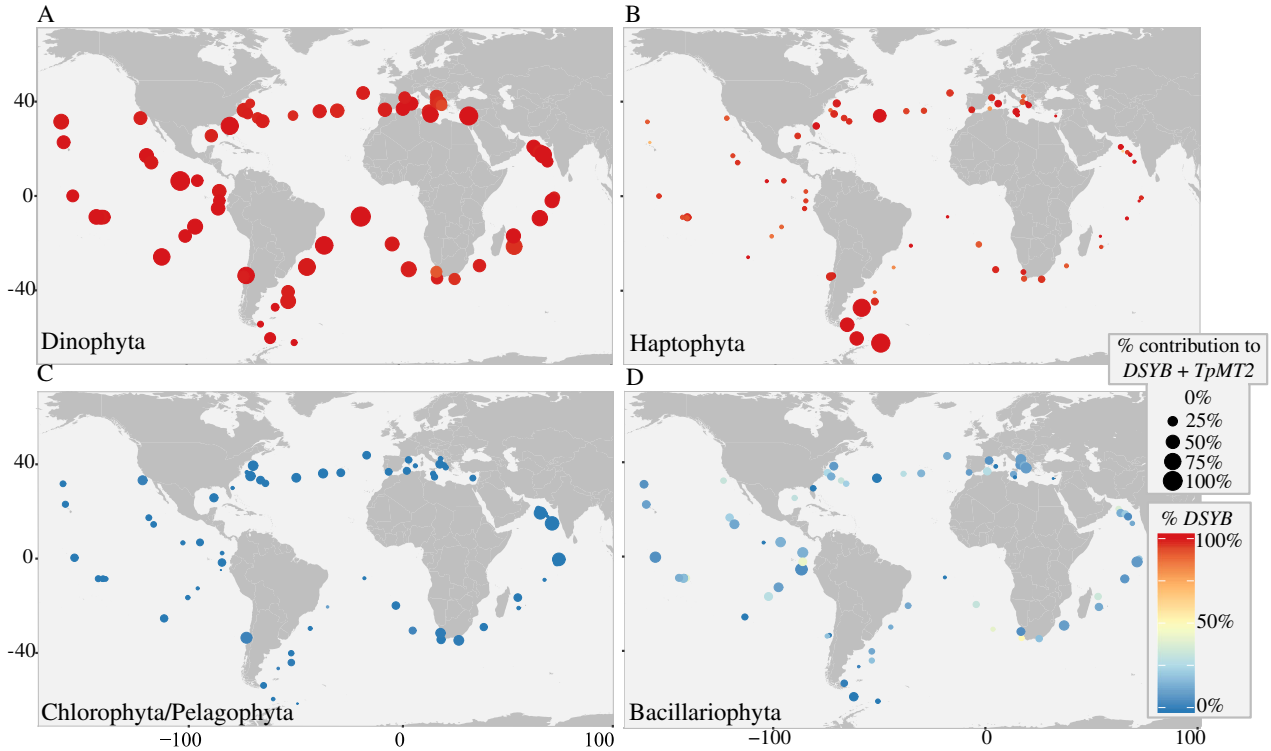


**Fig 5.** The percent contribution of (A) Dinophyta, (B) Haptophyta, (C) Chlorophyta/Pelagophyta, and (D) Bacillariophyta to the combined normalized abundance of *DSYB* and *TpMT2*. Circle colors represent the percent of *DSYB* or *TpMT2* annotated to each group (100% *DSYB* = red, 100% *TpMT2* = blue).

Furthermore, *in situ DSYB* unigenes were most often annotated as Dinophyta and Haptophyta, and *in situ TpMT2* unigenes were most often annotated as Chlorophyta/Pelagophyta and Bacillariophyta (Fig. 5).

The majority of isolates with no identified copies of *dsyB*, *DSYB*, or *TpMT2* were taxa typically assumed to not be capable of producing DMSP (e.g. Glaucophyta). However, many known DMSP producers also did not contain the targeted DMSP synthesis genes (Fig. 2). The missing synthesis genes in many confirmed DMSP producers suggest: (i) the ability to synthesize DMSP was incorrectly assumed based on indirect DMSP measurements (Spielmeyer *et al*., 2011), or (ii) the DMSP genotype could not be identified either as a result of the targeted DMSP synthesis genes not being assembled (e.g. due to insufficient relative expression levels and/or read depth), or that additional, currently unidentified DMSP synthesis genes exist. The potential failure of targeted DMSP synthesis genes to assemble due to low gene expression in certain conditions is most probable for the LoDP groups, which dynamically change intracellular DMSP concentrations in response to the environment (McParland *et al*., 2020). In contrast, HiDP groups constitutively maintain intracellular DMSP concentrations, and therefore the absence of targeted DMSP synthesis genes in these groups suggests that novel DMSP synthesis genes remain to be identified.

*DSYB* and *TpMT2* are the only identified eukaryotic DMSP synthesis genes with confirmed methyltransferase activity to perform the third step of DMSP synthesis (Curson *et al*., 2018; Kageyama *et al*., 2018). These two gene families form distinct phylogenetic clades (Fig. 3), and exhibit amino-acid sequence similarities of <40% (Fig. S1). As well, *TpMT2* homologues were not found in prokaryotes, whereas *DSYB* is considered to be a homologue of *dsyB* that originated from prokaryotes (Curson *et al*., 2018). Altogether, these differences support our hypothesis that the presence of *DSYB* or *TpMT2* is integral to exhibition of the two different HiDP and LoDP phenotypes.

While DMSP has been directly measured in many isolates, changes in TpMT2 and DSYB protein concentrations have been measured in only a small number of isolates (Curson *et al*., 2017, 2018; Kageyama *et al*., 2018). Specifically, TpMT2 protein concentrations have only been measured in *Thalassiosira pseudonana*, and were found to be regulated in response to both salinity changes and nitrogen limitation (Kageyama *et al*., 2018). Protein concentrations of dsyB in Alphaproteobacteria and DSYB in the diatom *Fragilariopsis cylindrus* were

also found to be regulated in response to both salinity and nitrogen stress (Curson *et al*., 2017, 2018). A synthesis of LoDP response to environmental changes showed that intracellular DMSP concentrations in a majority of other members of the LoDP *DSYB* sub-clade (primarily Bacillariophyta) change in response to nutrient stress (McParland and Levine, 2019). In contrast, DSYB protein concentrations in two HiDP prymnesiophytes either only responded to salinity changes (*Prymnesium parvum*), or did not respond to either condition (*Chrysochromulina tobin*) (Curson *et al*., 2018). Intracellular DMSP in other members of the HiDP *DSYB* sub-clade has also been shown to change in response to salinity but not nutrients (McParland and Levine, 2019; McParland *et al*., 2020). Under the same environmental stressors, differences in DSYB and TpMT2 protein regulation mirror previously described differences in phenotypic responses of HiDP and LoDP intracellular DMSP concentrations. If the regulatory differences are assumed to be representative of their respective sub-clades, the phylogenetic divisions of *TpMT2* and the three *DSYB* sub-clades support an evolutionary history that resulted in the observed differences in HiDP and LoDP phenotypes. It is very likely *DSYB* originated in prokaryotes. However, without knowing more about the origins of *TpMT2*, and the relation of *TpMT2* to *DSYB*, it is unclear whether the genes were evolutionarily sourced from a common ancestral gene family, or were a product of convergent evolution (Fig. 3).

While most known LoDPs exclusively exhibited *TpMT2* genotypes, some exhibited a *TpMT2* + *DSYB* genotype (Fig. 2). Gene redundancy is common in protists (Lee *et al*., 2014; Gruber and Kroth, 2017), but why so many Bacillariophyta would retain both genes remains unresolved. Nonetheless, as described above, Bacillariophyta regulate both TpMT2 and DSYB protein abundance in response to the environment, and this LoDP-like regulation was not observed in members of the HiDP *DSYB* sub-clade. Thus, the presence of a *TpMT2* + *DSYB* genotype appears to correspond with a LoDP phenotype, and the *DSYB*-only genotype corresponds with the HiDP phenotype.

DMSP is observed throughout the global surface ocean, and both DMSP synthesis genes were also globally abundant in *in situ* eukaryotic communities (Fig. 4). As *in situ* DMSP concentrations are thought to be dependent on HiDP abundance (McParland and Levine, 2019), it is not surprising that *DSYB* unigenes, which correlate with the taxonomy of HiDP phenotypes, are also globally abundant. While the contribution of LoDPs to global *in situ* DMSP concentrations is predicted to be relatively minimal (McParland and Levine, 2019), *TpMT2* unigenes were also found globally (Fig. 4). However, *TpMT2* typically represented ≤ ∼10% of the combined *DSYB* and *TpMT2* unigene abundances, except for two stations in the Arabian Sea where high abundances of Chlorophyta/Pelagophyta taxa likely resulted in a significant contribution of *TpMT2* to combined abundances (Fig. 5C). Although generally less abundant, the consistent expression of *TpMT2* unigenes suggests that DMSP synthesis is a critical metabolism for LoDPs.

*DSYB* unigenes annotated as Dinophyta dominated the combined *DSYB* and *TpMT2* unigene abundance at almost all *Tara* stations (Fig. 5A). Dinophyta contain high intracellular DMSP concentrations and contribute significantly to *in situ* DMSP (Caruana and Malin, 2014; Kiene *et al*., 2019). However, it is known that other HiDP taxa often dominate *in situ* DMSP production, particularly at higher latitudes (Ditullio *et al*., 2000; Archer *et al*., 2001). Indeed, at south Atlantic and Southern Ocean *Tara* stations, *DSYB* unigenes annotated as Haptophyta dominated both the combined *DSYB* and *TpMT2* unigene abundance (Fig. 5B). Vorobev *et al*. (2020) similarly showed that at the same Southern Ocean stations, as much as 45% of total *DSYB* expression can be attributed to two metagenomics-based transcriptomes affiliated with the Haptophyta genus *Phaeocystis*. The global dominance of Dinophyta was also likely influenced by their large transcript pools and *DSYB* paralogs (Hackett *et al*., 2004; Curson *et al*., 2018). This is a common challenge for Dinophyta and suggests that additional work is needed to constrain the relationship between Dinophyta *DSYB* transcript abundance and the contribution of Dinophyta to *in situ* DMSP production.

The *DSYB* genotype and *TpMT2* (or *TpMT2* + *DSYB*) genotype are significantly correlated with the HiDP and LoDP phenotype, respectively, across a diverse array of DMSP producers. While the exact cellular roles of DMSP remain unknown, it is clear that HiDPs and LoDPs exhibit contrasting responses to environmental stressors and possible that the identified genotypes are involved in these differential intracellular DMSP changes. Characterizing DMSP producers by their DMSP genotype is also likely a more accurate reflection of DMSP phenotype as it eliminates the challenges associated with measurements of cell volume. Additionally, DMSP producer abundances have previously been estimated *in situ* with traditional methods such as HPLC pigments or satellite-based products, but accurate quantification is challenging as these groups are often not the dominant primary producers (McParland and Levine, 2019). Using *DSYB* and *TpMT2* as marker genes has the potential to not only provide a direct quantification of DMSP producers within a mixed community, but also differentiate the fraction of the DMSP phenotype present.

The ecological and environmental significance of DMSP production has been studied for more than three decades (Charlson *et al*., 1987). However, DMSP synthesis genes were only identified in the past three years

 

making it previously impossible to survey diverse *in situ* microbial communities for potential DMSP production. Here we show for the first time that *DSYB* and *TpMT2* genotypes are globally abundant in *in situ* eukaryotic microbial communities, and are correlated to the known taxonomy of the HiDP and LoDP phenotypes. Like other biogeochemical cycles previously quantified with omics techniques (Zehr *et al*., 1998; Van Mooy *et al*., 2006; Webb *et al*., 2007; Beman *et al*., 2012), direct quantification of the environmental abundance of DMSP synthesis genes is promising for much improved future predictions of *in situ* DMSP production.

### *Methodology*

*DMSP measured in isolates.* The reported DMSP phenotypes were derived from a previous analysis of known prokaryotic and eukaryotic DMSP producers (supp table 1 in McParland and Levine, 2019). The phenotypes were assayed in both axenic and non-axenic isolates.

*DMSP synthesis genes survey.* We performed TBLASTN searches with previously published amino-acid sequences of *dsyB*, *DSYB*, and *TpMT2*. A single search was performed for *TpMT2* (XP_002291473) from Kageyama *et al*. (2018) and *dsyB* from Curson *et al*., 2017 (AOR83342.1), and a recursive search was performed with all *DSYB* sequences reported by Curson *et al*. (2018).

TBLASTN searches (Gerts *et al*., 2006) for prokaryotic sequences were performed on genomes from the NCBI RefSeq database (O'Leary *et al*., 2016) that were ranked as 'reference' by NCBI. If a previously reported prokaryotic DMSP producer was not represented in this subset of the RefSeq database, the species' sequences were searched in the NCBI nr database (these included: *Anabaena cylindrica*, *Anabaena sp*., *Coleofasciculus chthonoplastes*, *Crocosphaera watsonii*, *Lyngbya aestuarii*, *Marichromatium gracile*, *Nostoc minutum*, *Nostoc sp*., *Rhodopseudomonas palustris*, *Rhodovulum adriaticum*, *Rhodovulum euryhalinum*, *Rhodovulum robiginosum*, *Rhodovulum sp*., *Rhodovulum sulfidophilum*, *Synechococcus elongatus*, *Synechococcus sp*., *Synechocystis sp*., *Thiocapsa marina*, *Thiocapsa roseopersicina*, *Thiocystis violascens*, *Trichodesmium erythraeum*). Prokaryotic TBLASTN results were filtered by requiring an e-value ≤ 1e-30 and > 70% coverage of the query sequence.

TBLASTN searches for eukaryotic sequences were performed on a custom BLAST database built with the recently re-assembled MMETSP transcriptomes (Johnson *et al*., 2018). We also searched a small number of genomes from NCBI or JGI for species that are not present in MMETSP: *Fragilariopsis cylindrus*, *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, *Chrysochromulina tobin*, *Ostreococcus tauri*, *Nannochloropsis oceanica*, and *Euglena gracilis*. Eukaryotic TBLASTN results were less stringently filtered (e-value ≤ 2e-15 and > 75% coverage of the query sequence) in order to build hidden Markov model (HMM) profiles for subsequent HMM searches of the MMETSP proteins. Sequences identified with TBLASTN were aligned with MUSCLE (v.3.8.2) with default settings (Edgar, 2004), manually curated in JalView (v.2.10.5) (Waterhouse *et al*., 2009), and trimmed to regions of high coverage (∼190 amino-acids long). The *DSYB* and *TpMT2* alignments were used to create HMM profiles with hmmbuild (v.3.2.1) (hmmer.org) with default settings. The HMM searches were performed against the protein sequences predicted from the MMETSP assemblies. Alignments with an e-value ≤ 1e-30 were retained. If multiple alignments for a DMSP synthesis gene were recovered for a single genome or transcriptome, only the alignment with the lowest e-value was retained. Some isolates in the MMETSP database had multiple transcriptomes from different culturing conditions (e.g. phosphorus limitation and replete conditions). We searched all transcriptomes for a given isolate and if any returned an alignment, the organism was assigned the respective DMSP genotype.

As some MMETSP transcriptomes are not axenic (Keeling *et al*., 2014), recursive BLASTP searches of every identified eukaryotic sequence against the NCBI nr database were performed to control for prokaryotic contamination. The top 100 alignments were retained and filtered by requiring an e-value ≤ 1e-30. Sequences were removed if ≥ 97% of the filtered alignments were prokaryotic. The topology of these sequences with prokaryotic groups supports their removal from further analyses (Fig. S3).

*Phylogenetic analysis.* Alignments of all *dsyB*, *DSYB*, and *TpMT2* amino-acid sequences were built with MUSCLE, and manually curated with JalView. Two alignments were built: one with only sequences containing a corresponding phenotype measurement ($n = 51$) and one with all sequences ($n = 302$). The alignments were trimmed to a region of high coverage (∼ 340 amino-acids long) (Fig. 1B). If sequences were missing more than half of the alignment, they were removed. After trimming, the alignments were refined with the refine setting with MUSCLE. Twenty maximum-likelihood searches were performed with RAxML (v.8.0.0) using the PROTGAMMLG protein model (Stamatakis, 2014). A bootstrap analysis was performed on the best-scoring maximum-likelihood tree using a frequency-based convergence criterion. iTol was used to midpoint-root the trees and annotate (Letunic and Bork, 2019).

Tara *eukaryotic query.* Searches of the *DSYB* and *TpMT2* HMM profiles described above were performed against the Marine Atlas of *Tara* Ocean Unigenes database (MATOU; Carradec *et al*., 2018) using the Ocean Gene Atlas portal (Villar *et al*., 2018). The resulting alignments were filtered by requiring an e-value ≤ 1e-30. Reads from all 441 *Tara* metatranscriptome samples were previously mapped to MATOU. The *DSYB* and *TpMT2* abundances were estimated by normalizing to total mapped reads per sample (Villar *et al*., 2018). Combined normalized abundances were summed across all size fractions ≤ 180 μm for each station. Taxonomy was previously assigned to the MATOU with reference marine protist genomes and a least common ancestor method (Carradec *et al*., 2018). MATOU taxonomy was used to classify the *DSYB* and *TpMT2* unigenes by four major eukaryotic groups (Dinophyta, Haptophyta, Chlorophyta/Pelagophyta, and Bacillariophyta).

*Code and data.* The code used for the previously described workflow and figures can be found on GitHub: https://github.com/emcparland/DMSPgenes. All raw datasets that are required to replicate results are publicly available on Figshare: https://doi.org/10.6084/m9.figshare.13244750.v2.

**Acknowledgements**

We thank Sarah Hu, Elaina Graham, and Ben Tully for invaluable discussions. We thank the NCBI, MMETSP, and *Tara* Oceans communities for providing databases of raw and curated sequences to the public. Finally, we thank two anonymous reviewers for their time and constructive feedback. This work was supported by the Rose Hills Foundation, the University of Southern California, and a National Defense Science and Engineering Graduate Research Fellowship.

**Conflict Of Interest**

The authors declare no conflicts of interest.

**References**

Archer, S.D., Ragni, M., Webster, R., Airs, R.L., and Geider, R.J. (2010) Dimethylsulfoniopropionate and dimethylsulfide production in response to photoinhibition in *Emiliania huxleyi*. *Limnol Oceanogr* **55**: 1579–1589.

Archer, S.D., Widdicombe, C.E., Tarran, G.A., Rees, A.P., and Burkill, P.H. (2001) Production and turnover of particulate dimethylsulphoniopropionate during a coccolithophore bloom in the northern North Sea. *Aquat Microb Ecol* **24**: 225–241.

Beman, J.M., Bertics, V.J., Braunschweiler, T., and Wilson, J.M. (2012) Quantification of ammonia oxidation rates and the distribution of ammonia-oxidizing archaea and bacteria in marine sediment depth profiles from Catalina Island, California. *Front Microbiol* **3**: 263.

Bucciarelli, E., Ridame, C., Sunda, W.G., Dimier-Hugueney, C., Cheize, M., and Belviso, S. (2013) Increased intracellular concentrations of DMSP and DMSO in iron-limited oceanic phytoplankton *Thalassiosira oceanica* and *Trichodesmium erythraeum*. *Limnol Oceanogr* **58**: 1667–1679.

Bullock, H.A., Luo, H., and Whitman, W.B. (2017) Evolution of dimethylsulfoniopropionate metabolism in marine phytoplankton and bacteria. *Front Microbiol* **8**: 1–17.

Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2019) The new tree of eukaryotes. *Trends Ecol Evol* **0**: 1–13.

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., *et al*. (2018) A global ocean atlas of eukaryotic genes. *Nat Commun* **9**: 373.

Caruana, A.M.N., and Malin, G. (2014) The variability in DMSP content and DMSP lyase activity in marine dinoflagellates. *Prog Oceanogr* **120**: 410–424.

Charlson, R., Lovelock, J., Andrae, M., and Warren, S. (1987) Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. *Nature* **326**: 655–661.

Curson, A., Williams, B., Pinchbeck, B., Sims, L., Bermejo Martínez, A., Rivera, P., *et al*. (2018) DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. *Nat Microbiol* **3**: 430–439.

Curson, A.R.J., Liu, J., Bermejo Martínez, A., Green, R.T., Chan, Y., Carrión, O., *et al*. (2017) Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. *Nat Microbiol* **2**: 1–9.

Derelle, R., López-García, P., Timpano, H., and Moreira, D. (2016) A Phylogenomic framework to study the diversity and evolution of Stramenopiles (=Heterokonts). *Mol Biol Evol* **33**: 2890–2898.

Ditullio, G.R., Grebmeier, J.M., Arrigo, K.R., and Lizotte, M. P. (2000) Rapid and early export of *Phaeocystis Antarctica* blooms in the Ross Sea, Antarctica. *Nature* **404**: 595–598.

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

Edvardsen, B., Eikrem, W., Throndsen, J., Sáez, A.G., Probert, I., and Medlin, L.K. (2011) Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (haptophyta). *Eur J Phycol* **46**: 202–228.

Gage, D., Rhodes, D., Nolte, K., Hicks, W.A., Leustek, T., Cooper, A.J.L., and Hanson, A. (1997) A new route for synthesis of in marine algae. *Nat Lett* **387**: 891–894.

Gerts, E.M., Yu, Y.K., Agarwala, R., Schäffer, A.A., and Altschul, S.F. (2006) Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol* **4**: 1–14.

Gruber, A., and Kroth, P.G. (2017) Intracellular metabolic pathway distribution in diatoms and tools for genome-enabled experimental diatom research. *Philos Trans R Soc B Biol Sci* **372**: 20160402.

Hackett, J.D., Anderson, D.M., Erdner, D.L., and Bhattacharya, D. (2004) Dinoflagellates: a remarkable evolutionary experiment. *Am J Bot* **91**: 1523–1534.

Harrison P.J., Zingone A., Mickelson M.J., Lehtinen S., Ramaiah N., and Kraberg A.C. (2015). Cell volumes of

marine phytoplankton from globally distributed coastal data sets. *Estuarine, Coastal and Shelf Science*, **162**: 130–142.

Ito, T., Asano, Y., Tanaka, Y., and Takabe, T. (2011) Regulation of biosynthesis of dimethylsulfoniopropionate and its uptake in sterile mutant of *Ulva pertusa* (Chlorophyta). *J Phycol* **47**: 517–523.

Janouškovec, J., Gavelis, G.S., Burki, F., Dinh, D., Bachvaroff, T.R., Gornik, S.G., *et al*. (2017) Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc Natl Acad Sci U S A* **114**: E171–E180.

Johnson, L.K., Alexander, H., and Brown, C.T. (2018) Reassembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. *Gigascience*, **8**: 1–12.

Johnson, W.M., Soule, M.C.K., and Kujawinski, E.B. (2016) Evidence for quorum sensing and differential metabolite production by the marine heterotroph, *Ruegeria pomeroyi*, in response to DMSP. *ISME J* **10**: 2304–2316.

Kageyama, H., Tanaka, Y., Shibata, A., Waditee-Sirisattha, R., and Takabe, T. (2018) Dimethylsulfoniopropionate biosynthesis in a diatom *Thalassiosira pseudonana*: identification of a gene encoding MTHB-methyltransferase. *Arch Biochem Biophys* **645**: 100–106.

Karsten, U., Kück, K., Vogt, C., and Kirst, G.O. (1996) Dimethylsulfoniopropionate production in phototrophic organisms and its physiological functions as a cryoprotectant. In *Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds*, Keller, M.D., Kiene, R.P., Kirst, G.O., and Visscher, P.T. (eds). Boston, MA: Springer, pp. 143–153.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., *et al*. (2014) The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* **12**: e1001889.

Keller, M., Bellows, W., and Guillard, R. (1989) Dimethyl sulfide production in marine phytoplankton. In *Biogenic Sulfur in the Environment*, Saltzman, E., and Cooper, W. (eds). Washington D.C.: American Chemical Society, pp. 167–182.

Kiene, R.P., Linn, L.J., and Bruton, J.A. (2000) New and important roles for DMSP in marine microbial communities. *J Sea Res* **43**: 209–224.

Kiene, R.P., Nowinski, B., Esson, K., Preston, C., Marin, R., Birch, J., *et al*. (2019) Unprecedented DMSP concentrations in a massive dinoflagellate bloom in Monterey Bay, CA. *Geophys Res Lett* **46**: 12279–12288.

Lana, A., Bell, T.G., Simó, R., Vallina, S.M., Ballabrera-Poy, J., Kettle, A.J., *et al*. (2011) An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean. *Global Biogeochemical Cycles* **25**.

Lavoie, M., Levasseur, M., and Babin, M. (2015) Testing the potential ballast role for dimethylsulfoniopropionate in marine phytoplankton: a modeling study. *J Plankton Res* **37**: 699–711.

Lee, R., Lai, H., Malik, S.B., Saldarriaga, J.F., Keeling, P.J., and Slamovits, C.H. (2014) Analysis of EST data of the marine protist *Oxyrrhis marina*, an emerging model for alveolate biology and evolution. *BMC Genomics* **15**: 1471–2164.

Letunic, I., and Bork, P. (2019) Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**: W256–W259.

Lyon, B.R., Lee, P.a., Bennett, J.M., DiTullio, G.R., and Janech, M.G. (2011) Proteomic analysis of a sea-ice diatom: salinity acclimation provides new insight into the dimethylsulfoniopropionate production pathway. *Plant Physiol* **157**: 1926–1941.

McParland, E.L., and Levine, N.M. (2019) The role of differential DMSP production and community composition in predicting variability of global surface DMSP concentrations. *Limnol Oceanogr* **64**: 757–773.

McParland, E.L., Wright, A., Art, K., He, M., and Levine, N. M. (2020) Evidence for contrasting roles of dimethylsulfoniopropionate production in Emiliania huxleyi and Thalassiosira oceanica. *New Phytol* **226**: 396–409.

Menden-Deuer, S., and Lessard, E.J. (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnology and Oceanography*, **45**: 569–579.

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 U S A* **103**: 8607–8612.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., *et al*. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**: D733–D745.

Reisch, C.R., Moran, M.A., and Whitman, W.B. (2011) Bacterial catabolism of dimethylsulfoniopropionate (DMSP). *Front Microbiol* **2**: 1–12.

Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., *et al*. (2018) Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nat Commun* **9**: 310.

Seymour, J.R., Simó, R., Ahmed, T., and Stocker, R. (2010) Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. *Science* **329**: 342–345.

Spielmeyer, A., Gebser, B., and Pohnert, G. (2011) Dimethylsulfide sources from microalgae: improvement and application of a derivatization-based method for the determination of dimethylsulfoniopropionate and other zwitterionic osmolytes in phytoplankton. *Mar Chem* **124**: 48–56.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stefels, J. (2000) Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *J Sea Res* **43**: 183–197.

Stefels, J., and Van Leeuwe, M.A. (1998) Effects of iron and light stres on the biochemical composition of antarctic *Phaeocystis sp*. I. Intracellular DMSP concentrations. *J Phycol* **34**: 486–495.

Stefels, J., Steinke, M., Turner, S., Malin, G., and Belviso, S. (2007) Environmental constraints on the production and

removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. *Biogeochemistry* **83**: 245–275.

Sunda, W., Kieber, D.J., Kiene, R.P., and Huntsman, S. (2002) An antioxidant function for DMSP and DMS in marine algae. *Nature* **418**: 317–320.

Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., Wu, C.H., and Consortium, the U. (2015) UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**: 926–932.

Tripp, H.J., Kitner, J.B., Schwalbach, M.S., Dacey, J.W.H., Wilhelm, L.J., and Giovannoni, S.J. (2008) SAR11 marine bacteria require exogenous reduced Sulphur for growth. *Nature* **452**: 741–744.

Villar, E., Vannier, T., Vernette, C., Lescot, M., Cuenca, M., Alexandre, A., *et al*. (2018) The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res* **46**: W289–W295.

Vorobev, A., Dupouy, M., Carradec, Q., Delmont, T.O., Annamalé, A., Wincker, P., and Pelletier, E. (2020) Transcriptome reconstruction and functional analysis of eukaryotic marine plankton communities via high-throughput metagenomics and metatranscriptomics. *Genome Res* **30**: 647–659.

Waterhouse, A.M., Martin, D.M.A., Barton, G.J., Procter, J. B., and Clamp, M. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Webb, E.a., Jakuba, R.W., Moffett, J.W., and Dyhrman, S.T. (2007) Molecular assessment of phosphorus and iron physiology in Trichodesmium populations from the western central and western South Atlantic. *Limnol Oceanogr* **52**: 2221–2232.

Zehr, J.P., Mellon, M.T., and Zani, S. (1998) New nitrogen-fixing microorganisms detected in oligotrophic oceans by the amplification of nitrogenase (nifH) genes. *Appl Environ Microbiol* **64**: 3444–3450.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix S1**: Supporting Information.