DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton

Andrew R. J. Curson[1], Beth T. Williams[1], Benjamin J. Pinchbeck[1], Leanne P. Sims[1], Ana Bermejo Martínez[1], Peter Paolo L. Rivera[1], Deepak Kumaresan[2], Elena Mercadé[3], Lewis G. Spurgin[1], Ornella Carrión[1], Simon Moxon[1], Rose Ann Cattolico[4], Unnikrishnan Kuzhiumparambil[5], Paul Guagliardo[6], Peta L. Clode[6,7], Jean-Baptiste Raina[5], Jonathan D. Todd[1*]

[1]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK. [2]School of Biological Sciences and Institute for Global Food Security, Queen's University Belfast, Belfast BT9 7BL, UK. [3]Laboratori de Microbiologia, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain. [4]Department of Biology, University of Washington, Seattle, Washington, USA. [5]Climate Change Cluster (C3), Faculty of Science, University of Technology, Sydney, NSW 2007 Australia. [6]The Centre for Microscopy Characterisation and Analysis, University of Western Australia, Crawley, Australia. [7]Oceans Institute, University of Western Australia, Crawley, Australia.

*corresponding author

Dimethylsulfoniopropionate (DMSP) is a globally important organosulfur molecule, and the major precursor for dimethyl sulfide (DMS). These compounds are important info-chemicals, key nutrients for marine microorganisms, and are involved in global sulfur cycling, atmospheric chemistry and cloud formation[1–3]. DMSP production was thought to be confined to eukaryotes, but heterotrophic bacteria can also produce DMSP, via the pathway used by most phytoplankton[4], and the DsyB enzyme catalysing the key step

of this pathway in bacteria was recently identified[5]. However, eukaryotic phytoplankton likely produce most of Earth's DMSP, yet no DMSP biosynthesis genes have been identified in any such organisms. Here we identify functional *dsyB* homologues, termed *DSYB*, in many phytoplankton and corals. DSYB is a methylthiohydroxybutryate (MTHB) methyltransferase enzyme localised in the chloroplasts and mitochondria of the haptophyte *Prymnesium parvum*, and stable isotope tracking experiments support these organelles as sites of DMSP synthesis. *DSYB* transcription levels increased with DMSP concentrations in different phytoplankton and were indicative of intracellular DMSP. The identification of the eukaryotic *DSYB* sequences, along with bacterial *dsyB*, provide the first molecular tools to predict the relative contributions of eukaryotes and prokaryotes to global DMSP production. Furthermore, evolutionary analysis suggests that eukaryotic *DSYB* originated in bacteria and was passed to eukaryotes early in their evolution.

Not all phytoplankton produce DMSP, and in those that do, intracellular DMSP concentrations vary considerably across groups and within genera[6]. Previous studies identified candidate genes[7,8] involved in DMSP synthesis via the transamination pathway (Fig. 1a), which is common to DMSP-producing bacteria[5] and algae[4]. A proteomic study of the diatom *Fragilariopsis cylindrus* identified putative DMSP synthesis enzymes[7], including the MTHB methyltransferase reaction catalysed by DsyB in bacteria. Another study on corals identified homologues of two of the *F. cylindrus* enzymes in *Acropora millepora*, one being a candidate MTHB methyltransferase[8]. None of these enzymes have been functionally ratified, and the putative MTHB methyltransferases share no significant sequence similarity to DsyB. When we cloned and expressed the *F. cylindrus* and *A. millepora* putative MTHB methyltransferase genes they had no such enzyme activity (Supplementary Table 1), suggesting that the identity of an algal MTHB methyltransferase was still unknown.

We identified homologues to the bacterial MTHB methyltransferase gene *dsyB*[5] in available genomes and/or transcriptomes of all marine prymnesiophytes; most dinoflagellates, some corals, and ~20% of diatoms and Ochrophyta (Fig. 1b, Supplementary Table 2, Supplementary Table 3 and Supplementary Data 1). The only dinoflagellate transcriptomes lacking *dsyB* were from *Oxyrrhis marina*, a heterotroph which produces no detectable DMSP[6,9]. Furthermore, many dinoflagellates, and some haptophytes, diatoms and corals, have multiple *dsyB* homologs. The grouping of these multiple homologues across the phylogeny was consistent with multiple gene duplication and gene loss events over the evolutionary history of eukaryotes[10] (Fig. 1b, Supplementary Table 2, and Supplementary Table 3). These *dsyB*-like genes, termed *DSYB*, from representatives of the corals (*Acropora cervicornis*), diatoms (*F. cylindrus*), dinoflagellates (*Alexandrium tamarense*, *Lingulodinium polyedrum*, *Symbiodinium microadriaticum*) and prymnesiophytes (*Chrysochromulina tobin*, *Prymnesium parvum*) were cloned and shown to have MTHB methyltransferase activity, at similar levels to bacterial DsyB from *Labrenzia* (Supplementary Table 1). These algal DSYB enzymes fully complement bacterial *dsyB*$^-$ mutants, defective in DMSP production. Furthermore, enzyme assays with purified DSYB and MTHB substrate alone showed no activity, but *in vitro S*-adenosyl methionine (SAM)-dependent MTHB methyltransferase activity was observed when the same assays were incubated with heat-denatured *P. parvum* cell lysates (Supplementary Table 4). This suggests that a co-factor(s) present in *P. parvum* lysates might be required for activity. The K_M values of DSYB for MTHB and SAM were 88.2 μM and 60.1 μM respectively (Supplementary Table 4, Supplementary Fig. 1). DSYB showed no detectable methyltransferase activity with other potential substrates (including methionine (Met), 4-methylthio-2-oxobutyrate (MTOB) and methylmercaptopropionate (MMPA); Supplementary Table 4). Thus, *DSYB* encodes the first DMSP synthesis enzyme to be identified and functionally ratified from any eukaryotic algae.

DSYB is found across many, but by no means all, major groups of eukaryotes, and eukaryotes are monophyletic in the DsyB/DSYB phylogeny, suggesting either i) that *DSYB* was present in the last eukaryotic common ancestor (LECA) and has been lost across many eukaryotic groups, or ii) that *dsyB* has been transferred to eukaryotes multiple times. Homology and phylogenetic analyses place Alphaproteobacteria as the sister clade to the eukaryotes for this gene (Fig. 1b); we note that Alphaproteobacterial genes make up a significant proportion of eukaryotic genomes, due to endosymbiotic events with the ancestor of mitochondria[11]. We suggest that DMSP production originated in prokaryotes, and was transferred to the eukaryotes, either via endosymbiosis at the time of mitochondrial origin, or more recently via horizontal gene transfer (HGT). Interestingly, coral DSYB paralogs grouped with dinoflagellate sequences from coral symbionts of the genus *Symbiodinium* (Fig. 1b). This is consistent with HGT between corals and their symbionts, as documented for other genes[12], and suggests that DMSP production in corals may be a result of recent HGT of *DSYB* from dinoflagellates. However, we cannot discount the possibility that coral DSYB sequences might be contaminant sequences unintentionally extracted from their symbionts.

No *DSYB* homologs were identified in available transcriptomes from marine ascomycota, cercozoa, chlorophyta, ciliophoran, cryptophyta, euglenozoa, glaucophyta, labyrinthista, perkinsozoa, or rhodophyta (Supplementary Table 3), although some members of these taxa, such as chlorophyta and rhodophyta[9,13], are known to produce DMSP. DSYB homologs were also absent in the genomes of the DMSP-producing diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*[14]. Some marine eukaryotes lack *DSYB* simply because they do not produce DMSP[6,9]. Others may (i) have *DSYB* but not express it under the tested transcriptome conditions, (ii) contain a MTHB methyltransferase isoform, or (iii) produce DMSP via a different synthesis pathway[15].

Intracellular DMSP concentrations are generally high in dinoflagellates (reported up to 3.4 M, but unlikely to be this high given seawater osmolarity is ~1 Osm l^{-1}) and haptophytes (up to 413 mM), but significant intra-group variance exists, with some representatives not producing DMSP at detectable levels[6,9]. Since eukaryotic DSYB enzymes had MTHB methyltransferase rates similar to bacterial DsyB enzymes (Supplementary Table 1), it is unlikely that variation in DsyB and DSYB amino acid sequences is responsible for the differing intracellular DMSP concentrations in these organisms (Supplementary Table 1). To understand this variance, we studied model DMSP-producing phytoplankton, starting with *Chrysochromulina tobin* CCMP291 and *Chrysochromulina* sp. PCC307, two haptophytes adapted to different salinity levels (fresh-brackish and marine waters[16,17], respectively). Both *Chrysochromulina* strains produced very low intracellular DMSP concentrations (Supplementary Table 1, Supplementary Fig. 2), which were unaffected by variation in salinity and nitrogen availability, conditions that have been shown previously to affect DMSP production in bacteria[5] and phytoplankton[7,18]. Consistent with these findings, *C. tobin* CCMP291 *DSYB* was transcribed at very low levels (Supplementary Fig. 2), perhaps indicating a DMSP function in these haptophytes that only requires low concentrations. Many haptophytes produce high DMSP concentrations, consistent with an osmoregulatory function, but this contrasts the low *C. tobin* DMSP concentrations and highlights the variability in the process and requirement for a methodology to predict which phytoplankton are high and low DMSP producers. Perhaps other compatible solutes, possibly sugars or amino acids, are the major osmolytes in CCMP291 and PCC307. Consistent with this, the osmolyte glycine betaine (551 ± 6 nmol) was present in ~10-fold higher amounts than DMSP (52 ± 6 nmol) in CCMP291.

Next, we investigated DMSP production in six *Prymnesium* strains, from brackish/marine sources, and found they had similar intracellular DMSP concentrations, which were much

higher than those for *C. tobin* (Supplementary Fig 2). *P. parvum* CCAP946/6 *DSYB* transcription was also higher than that for *C. tobin DSYB* under standard conditions (Supplementary Fig 2). Interestingly, *DSYB* transcription, DSYB protein levels and DMSP concentration in *P. parvum* were all enhanced by increased salinity but unaffected by other environmental conditions, including nitrogen availability or temperature (Supplementary Fig. 2; Supplementary Fig. 3). Increased salinity enhances DMSP production in many phytoplankton, notably *P. parvum*, where DMSP is thought to be a significant osmolyte[19]. Our findings, and those of Dickson and Kirst[19], are consistent with DMSP playing an osmoregulatory role in this haptophyte. However, *dsyB* transcription and DMSP production is regulated by salinity in bacteria, yet no detrimental effect on growth was observed in a bacterial *dsyB*$^-$ mutant when grown in saline conditions[5]. Thus, increased DSYB expression and DMSP production with raised salinity does not necessarily indicate a major role for DMSP in osmoprotection.

P. parvum DSYB protein was concentrated to the chloroplasts and mitochondria (Fig. 2; Supplementary Fig. 4). We propose these organelles as sites of DMSP synthesis in *P. parvum* and perhaps other eukaryotic phytoplankton. Although DMSP production in mitochondria has not been reported, DMSP is produced in the chloroplasts of the higher plant *Wollastonia*, albeit using a different pathway[20]. Based on *in silico* sequence analysis (see Methods), DSYB from *P. parvum* and some other phytoplankton are predicted to be targeted to the mitochondria and/or chloroplasts (Supplementary Table 5). However, chromophyte algae, such as haptophytes and diatoms, have complex plastids[21], which may render such *in silico* predictions less reliable.

Nanoscale secondary-ion mass spectrometry (NanoSIMS), with a cryopreservation method previously shown to preserve cytosolic DMSP[22], was used to identify potential sub-cellular

147 sites of DMSP production and storage in *P. parvum* through tracking $^{34}SO_4$ uptake. After 48 h
148 incubation, ~23.2 % ± 0.2 of the *P. parvum* intracellular DMSP pool was labelled with ^{34}S
149 (^{34}S-DMSP; Supplementary Fig. 5). Within the cells, the ^{34}S appeared to be localized in sub-
150 cellular compartments, with increasing levels appearing over time in the chloroplasts ($^{34}S/^{32}S$:
151 3.4 ± 0.17 after 48 h) and in submicrometer hotspots ($^{34}S/^{32}S$: 3.1 ± 0.15 after 48 h) (Fig. 3).
152 Given the size and location of these hotspots, they are likely to be mitochondria or small lipid
153 vesicles (Fig. 3). Although many sulfur compounds are present in algal cells, DMSP
154 represents more than 50% of the total organosulfur compounds in marine phytoplankton[23]
155 and it is expected to account for a significant fraction of the ^{34}S signal detected by
156 NanoSIMS. However, it cannot be discounted that the increased ^{34}S content in the chloroplast
157 could be due to transport of sulfur and subsequent assimilation via plastid-located enzymes,
158 such as ATP sulfurylase, APS reductase and sulfite reductase. Nonetheless, the simultaneous
159 increase in ^{34}S in the chloroplasts and potentially mitochondria supports our hypothesis that
160 these organelles are indeed sites of DMSP synthesis and storage in *P. parvum* and likely other
161 phytoplankton. Given the role of these organelles in energy production, it is perhaps not
162 surprising that DMSP production, an energy-demanding process[24], may occur at these sites.
163 With DMSP being far less concentrated in the cytosol, it is less likely that its primary
164 function in *P. parvum* is as a typical cytosolic osmolyte, but it may be a key osmolyte in the
165 chloroplasts and/or mitochondria, as proposed in *Wollastonia* chloroplasts[20]. Also,
166 considering reactive oxygen species (ROS) are generated in the mitochondria and
167 chloroplasts, and that DMSP is an effective scavenger of ROS[25], the production of DMSP in
168 these organelles is in line with its putative role in oxidative stress protection[24,25].

169 Diatoms are thought to produce the lowest intracellular DMSP levels (typically < 50 mM)[9].
170 We studied DMSP production in the polar ice diatom *Fragilariopsis cylindrus*, one of the few
171 diatoms with a functional *DSYB* (Supplementary Table 2), finding that, under standard

conditions, intracellular DMSP levels and *DSYB* transcription were relatively low, when compared to (e.g.) *P. parvum* (Supplementary Fig. 2). However, consistent with work in other diatoms[18], both *F. cylindrus* DMSP production and *DSYB* transcription increased with nitrogen limitation and increased salinity (Supplementary Fig. 2). The latter supports a role for DMSP in osmoregulation and salinity-induced oxidative stress protection in *F. cylindrus*, as suggested by Lyon et al.[7]. DSYB was not detected as one of the salinity-induced proteins in Lyon et al.[7], despite using the same salinity conditions for our experiments, reflecting the nature of 2D gel electrophoresis studies, whereby not all proteins are identified.

Given the trend of intracellular DMSP concentration increasing with *DSYB* transcription, we studied *Symbiodinium microadriaticum* CCMP2467, a dinoflagellate from a genus producing high DMSP concentrations[6]. *S. microadriaticum* gave the highest intracellular DMSP (282 mM) and cumulative *DSYB* transcription of the tested phytoplankton (Supplementary Fig. 2). Similarly, available transcriptomic data showed that high DMSP-producing dinoflagellate and haptophyte phytoplankton (see above) had the highest average *DSYB* transcription, which was ~3 and 8-fold higher, respectively, than that in diatoms (Supplementary Table 2). Transcriptomic data was also congruent with high variability in intracellular DMSP levels within dinoflagellates and haptophytes[6,9]. While additional factors, such as DSYB protein levels, DMSP excretion, DMSP catabolism and cell volume, will affect an organism's intracellular DMSP concentration, the data presented here on a small number of phytoplankton supports the hypothesis that *DSYB* transcription is a reasonably good indicator of DMSP concentration. Some *DSYB*-containing phytoplankton may also contain MTHB methyltransferase isoform enzymes or utilise other DMSP synthesis pathways, in which case such predictions may be inaccurate. Further work is required to substantiate this hypothesis.

The prominence of environmental DMSP-producing bacteria and eukaryotes was examined in the ocean microbial reference gene catalogue (OM-RGC) metagenomic dataset, generated from samples fractionated to < 3 μm[26] (Supplementary Table 6 and Supplementary Fig. 6). The *dsyB* gene was predicted to be present in 0.35% of total bacteria in these samples. For comparison, DMSP lyase genes (*dddD, dddL, dddK, dddP, dddQ, dddW, dddY* and *Alma1*)[27], were also used. The *dsyB* gene was more abundant than *dddL*, *dddW*, *dddY*, and the algal DMSP lyase gene *Alma1*, but was less abundant than *dddD*, *dddK*, *dddP* and *dddQ* in the OM-RGC dataset. Despite only 3% of the OM-RGC microorganisms likely being eukaryotes[26], *DSYB* genes were detected and were ~25-fold less abundant than bacterial *dsyB*. Since no *DSYB* sequences have been identified in bacteria, we conclude that picoeukaryotes in these samples contain *DSYB* and thus, the genetic potential to make DMSP. The production of DMSP by *DSYB*-containing picoeukaryotes could contribute, along with DMSP-producing bacteria, to the DMSP measured from particles <2 μm in size in seawater samples[28].

We also investigated the occurrence of *dsyB* and *DSYB* in marine metatranscriptomes (Supplementary Table 7). *dsyB* transcripts were detected in all tested *Tara* oceans metatranscriptomic datasets apportioned to marine bacteria (Supplementary Table 8 and Supplementary Fig. 6). *dsyB* transcript abundance (normalised to total sequence numbers) was similar to *dddD* and greater than *dddL*, *dddW*, *dddY* and *Alma1*, but was far less than *dddK, dddP* and *dddQ*. Although these datasets do not consider phytoplankton >3 μm, *DSYB* transcripts, likely from picoeukaryotes, were detected at levels only 3-fold lower than the bacterial *dsyB* gene, again suggesting that these smaller eukaryotes, like bacteria, should be considered as potentially significant DMSP producers (Supplementary Table 8).

We also analysed the North Pacific Ocean metatranscriptomes (GeoMICS) which used appropriate fractionation methods for bacteria and larger phytoplankton[29]. As expected, eukaryotic *DSYB* transcript numbers were higher than those of bacterial *dsyB* in all of the 2-53 µm fractions, which should contain relatively more phytoplankton than bacteria, and the opposite was true in most of the 0.2-2 µm fractions, which should have relatively more bacteria but not contain the larger phytoplankton (Supplementary Table 9). Analysing data from both the large and small size fractions at different sites allowed us to gauge the relative total transcript numbers of *DSYB* and *dsyB* in these samples, as well as those of the DMSP lyase genes. Prokaryotic *dsyB* transcripts (normalised to the recovery of an internal standard) were more abundant than those for the bacterial DMSP lyase genes *dddK*, *dddL*, *dddQ*, *dddY* and *dddW,* 3-fold less than *dddP* and *Alma1* and 27-fold less than *dddD* (Supplementary Table 9). Eukaryotic *DSYB* transcripts were slightly less abundant than those for the eukaryotic DMSP lyase (*Alma1*), but, were ~2-fold more abundant than those for bacterial *dsyB*. With similar DsyB and DSYB enzyme rates (Supplementary Table 1), this metatranscriptomic data suggests that eukaryotic phytoplankton may be the major contributors to DMSP production via the DsyB/DSYB pathway in these samples. However, direct extrapolation from these data to predict eukaryotic versus bacterial DMSP production (via DsyB/DSYB) is not likely accurate since other factors, such as DsyB/DSYB protein stability or the differing expression and activities of other enzymes in the pathway, may also affect DMSP production. Nonetheless, *dsyB* and *DSYB* sequences provide invaluable tools for future, in-depth studies to investigate the relative contribution of bacterial and algal DMSP production in varied marine environments. Molecular studies are also required to identify DMSP synthesis genes in DMSP-producing organisms which lack *dsyB* or *DSYB*.

Methods

Media and general growth of algae and bacteria

Prymnesium parvum CCAP941/1A, *Prymnesium parvum* CCAP941/6, *Prymnesium parvum* CCAP946/1B, *Prymnesium parvum* CCAP946/1D, *Prymnesium parvum* CCAP946/6, *Prymnesium patelliferum* CCAP946/4, *Chrysochromulina* sp. PCC307 and *Symbiodinium microadriaticum* CCMP2467 were grown in F/2[30] medium made with ESAW artificial seawater[31] and without any added Na_2SiO_3. Axenic *Fragilariopsis cylindrus* CCMP1102 was supplied by Mock et al.[32] and grown in F/2 medium made with ESAW artificial seawater at 4 °C with a light intensity of 120 μE m^{-2} s^{-1} and constant illumination. *Chrysochromulina tobin* CCMP291 was grown in the proprietary medium RAC-5[33]. All algal cultures (except *F. cylindrus*) were grown at 22 °C with a light intensity of 120 μE m^{-2} s^{-1} and a light dark cycle of 16 h light/8 h dark, unless otherwise stated. Where necessary, media for algal growth were modified according to the requirements of the experimental conditions being tested. Where strains were not already known to be axenic, cultures were treated with multiple rounds of antibiotic treatment prior to experiments. Test cultures with and without antibiotic treatments showed no significant difference in total DMSP in samples. For *P. parvum* CCAP946/6, and *Chrysochromulina* sp. PCC307 cultures, streptomycin (400 μg ml^{-1}), chloramphenicol (50 μg ml^{-1}), gentamicin (20 μg ml^{-1}) and ampicillin (100 μg ml^{-1}) were added, and for *S. microadriaticum* cultures, streptomycin (100 μg ml^{-1}) and neomycin (100 μg ml^{-1}) were added. *E. coli* was grown in LB[34] complete medium at 37 °C. *R. leguminosarum* was grown in TY[35] complete medium or Y[35] minimal medium (with 10 mM succinate as carbon source and 10 mM NH_4Cl as nitrogen source) at 28 °C. *L. aggregata* J571 was grown in YTSS[36] complete medium or MBM[37] minimal medium (with 10 mM succinate as carbon source and 10 mM NH_4Cl as nitrogen source) at 30 °C. Where necessary, antibiotics were added to

bacterial cultures at the following concentrations: streptomycin (400 μg ml^{-1}), kanamycin (20 μg ml^{-1}), spectinomycin (200 μg ml^{-1}), gentamicin (20 μg ml^{-1}), ampicillin (100 μg ml^{-1}). Strains used in this study are listed in Supplementary Table 10.

Staining with 4',6-diamidino-2-phenylindole (DAPI)

The absence of bacterial contamination was confirmed by epifluorescence microscopy of culture samples stained with DAPI[38]. Briefly, 13 ml of culture was removed and fixed with 765 μl paraformaldehyde, then 130 μl of DAPI stain (1 mg ml^{-1} in H_2O) was added and samples were stored in the dark at 4 °C for 16 h. After staining, 3 ml of the stained cells were removed and filtered onto a Whatman Nuclepore track-etched membrane (25 mm, 0.2 μm, polycarbonate). To prepare slides, one drop of immersion oil was added onto the slide then the sample filter was placed on the oil and another drop of immersion oil was added onto the filter. A cover slip was then placed on top of the filter and pressed down with forceps to remove air bubbles. The slide was then tilted and left on absorbent paper towel to allow any excess oil to drain/wick away. Slides were examined using an Olympus BX40 microscope equipped with an Olympus Camedia C-7070 digital camera.

General *in vivo* and *in vitro* genetic manipulations

Plasmids (Supplementary Table 10) were transferred to *E. coli* by transformation, and *Rhizobium leguminosarum* J391 or *Labrenzia aggregata* J571 by conjugation in a triparental mating using the helper plasmid pRK2013[39]. Routine restriction digestions and ligations for cloning were performed essentially as in Downie et al.[40]. The oligonucleotide primers used

for molecular cloning were synthesised by Eurofins Genomics and are detailed in Supplementary Table 11. Sequencing of plasmids and PCR products was performed by Eurofins Genomics.

The *DSYB* gene from *P. parvum* CCAP946/6 was PCR-amplified from cDNA and cloned into the IPTG-inducible wide host range expression plasmid pRK415[41]. All other *DSYB* genes were synthesised by Eurofins Genomics, from sequences codon-optimised (using Invitrogen GeneArt) for expression in *E. coli*, in the vector pEX-K4 (Eurofins Genomics). The synthesised genes were then subcloned into pLMB509[42], a taurine-inducible plasmid for the expression of genes in *Rhizobium* and *Labrenzia*, using *Nde*I and *BamH*I or *EcoR*I restriction enzymes. All plasmid clones are described in Supplementary Table 10.

MTHB methyltransferase (MMT) assays

To measure MMT activity from pLMB509 clones expressing the *dsyB* or *DSYB* gene in *R. leguminosarum* J391, cultures were grown (in triplicate) overnight in TY complete medium, 1 ml of culture was centrifuged at 20,000*g* for 2 min, resuspended in the same volume of Y medium and then diluted 1:100 into 5 ml Y with 10 mM taurine (to induce expression, Sigma-Aldrich, T0625), 0.5 mM DL-MTHB (Sigma-Aldrich, 55875), 0.1 mM L-methionine and gentamycin, and incubated at 28 °C for 60 h before sampling for gas chromatography (GC) analysis (see 'Quantification of DMS and DMSP by gas chromatography') to determine the amount of DMSP product.

To measure MMT activity from pLMB509 clones expressing the *DSYB* gene in the *L. aggregata dsyB*$^-$ mutant strain J571, cultures were grown (in triplicate) overnight in YTSS complete medium. Following incubation, 1 ml of culture was then centrifuged at 20,000*g* for 2 min, resuspended in the same volume of MBM medium and then diluted 1:50 into 5 ml

MBM with 10 mM taurine (to induce expression, Sigma-Aldrich), rifampicin and gentamycin, and incubated at 30 °C for 24 h. Samples were taken for GC analysis and determining protein concentration (t = 0 h timepoint). DL-MTHB (0.5 mM) and L-methionine (0.1 mM) were then added as substrates to the remaining cultures and these were incubated for 4 h at 30 °C before sampling for GC and protein again (t = 4 h timepoint), with activity calculated based on the difference in measured DMSP product between t=0 h and t=4 h.

To measure DMSP in *Rhizobium* or *Labrenzia* assay mixtures, 200 µl of culture was added to a 2 ml glass serum vial then 100 µl 10 M NaOH was added and vials were crimped immediately, incubated at 22 °C for 24 h and monitored by GC assay (see 'Quantification of DMS and DMSP by gas chromatography'). DsyB/DSYB activity is expressed as pmol DMSP mg protein^{-1} min^{-1}, assuming that all the DMSP is derived from DMSHB through DDC activity. LC-MS analysis shows no detectable DMSHB in *Rhizobium* or *Labrenzia* expressing DsyB/DSYB, presumably due its conversion to DMSP by DDC activity, so DMSP production is used as a proxy for DsyB activity. Protein concentrations were determined using the Bradford method (BioRad). Control assays of *Rhizobium* or *Labrenzia* J571 containing pLMB509 were carried out, as above, and gave no detectable DsyB/DSYB activity.

Growth of algae under non-standard conditions

For all *P. parvum*, *F. cylindrus* and *C. tobin* cultures described here, all samples were taken in mid-exponential phase growth before growth rates started to decline (checked by continuing to monitor growth following sampling). To measure DMSP production or *DSYB*/DSYB expression in *P. parvum* CCAP946/6 under different conditions, the growth

conditions or F/2 medium were modified as follows. Standard growth conditions were a temperature of 22 °C, light intensity of 120 μE m^{-2} s^{-1}, salinity of 35 practical salinity units (PSU) and nitrogen concentration of 882 μM. For increased or decreased salinity, the amount of salts added to the artificial seawater were adjusted to give a salinity of 50 or 10 PSU respectively. For reduced nitrogen concentration cultures, the F/2 medium contained 88.2 μM (10% of standard F/2). For changes in temperature, cultures were grown at 15 °C or 28 °C. To measure the effect of increased salinity and nitrogen limitation in *F. cylindrus* CCMP1102, this strain was grown in F/2 medium with increased salts in the artificial seawater (to 70 PSU) or reduced nitrogen (88.2 μM, 10% of standard F/2). To measure the effect of increased salinity and nitrogen limitation in *C. tobin* CCMP291, this strain was grown in F/2 medium with sea salts added to the RAC-5 medium (to 5 PSU) or reduced nitrogen (85 μM, 10% of standard RAC-5).

Sampling methods

To measure growth of algal cultures, samples were removed, diluted (dependent on level of growth) in artificial seawater and cell counting was done using a Multisizer 3 Coulter counter (Beckman Coulter). The effect of stress on photosystem II was determined by measuring Fv/Fm values using a Phyto-Pam phytoplankton analyzer (Heinz Walz, Germany). To obtain samples for DMSP quantification by GC or liquid chromatography-mass spectrometry (LC-MS), 25 ml of culture was filtered onto 47 mm GF/F glass microfiber filters (Fisher Scientific, UK) using a Welch WOB-L 2534 vacuum pump, and filters were then blotted on paper towel to remove excess liquid and stored at -80 °C in 2 ml centrifuge tubes for particulate DMSP (DMSPp) measurement. To obtain samples for RNA, 50 ml of culture was filtered onto 47 mm 1.2 μm RTTP polycarbonate filters (Fisher Scientific, UK) and filters

were stored in 2 ml centrifuge tubes at -80 °C. To obtain samples for protein for Western blotting, 50 ml of culture was centrifuged at 600*g* for 10 min in a 50 ml centrifuge tube, the supernatant was decanted and cells were transferred in the residual liquid to a 2 ml centrifuge tube and centrifuged at 600*g* for 5 mins. All residual liquid was then aspirated and the pelleted cells were stored at -80 °C.

Quantification of DMS and DMSP by GC

All GC assays involved measurement of headspace DMS, either directly produced or via alkaline lysis of DMSP or DMSHB, using a flame photometric detector (Agilent 7890A GC fitted with a 7693 autosampler) and a HP-INNOWax 30 m x 0.320 mm capillary column (Agilent Technologies J&W Scientific). All GC measurements were performed using 2 ml glass serum vials containing 0.3 ml liquid samples and sealed with PTFE/rubber crimp caps. Quantification of DMSP from algal samples filtered on GF/F glass microfiber filters (see 'Sampling methods') was performed following methanol extraction. Filters were folded, placed in a 2 ml centrifuge tube and 1 ml 100% methanol was added. Samples were stored for 24 h at -20 °C to allow the extraction of cellular metabolites, then 200 µl of the methanol extract was added to a 2 ml vial, 100 µl 10 M NaOH was added, vials were crimped immediately, incubated at 22 °C for 24 h in the dark and monitored by GC. Control samples in which DMSP standards were added to algal sample filters prior to methanol extraction showed that all standard was recovered following our extraction and measurement procedure. Calibration curves were produced by alkaline lysis of DMSP standards in water (for *Rhizobium*/*Labrenzia* MMT assays) or 100% methanol (for algal methanol extracts), or DL-DMSHB (chemically synthesised as in Curson et al.[5]) standards in water with heating at 80 °C for 10 mins (to release DMS from DMSHB) (for assays with purified DSYB protein). The

384 detection limit for headspace DMS from DMSP was 0.015 nmol in water and 0.15 nmol in
385 methanol, and from DMSHB was 0.3 nmol in water.

386

387 **Quantification of DMSP by LC-MS**

388 LC-MS was used to confirm that phytoplankton were producing DMSP and at similar levels
389 to that shown by GC, ruling out the possibility that DMS detected by GC was due to some
390 other compound and not DMSP. Samples were extracted as follows: GF/F filters of
391 phytoplankton (see 'Sampling methods') were resuspended in 1 ml of 80% LC-MS grade
392 acetonitrile (extraction solvent), and mixed by pipetting and vortexing for 2 min. The
393 resulting mixture was transferred into a fresh 2 ml Eppendorf tube. For a second round of
394 extraction, another 1 ml of the extraction solvent was then added and mixed as previously
395 described. Then the filters were centrifuged at 18,000*g* for 10 min and the supernatant was
396 collected, giving a total volume of 2 ml of the collected supernatant. The collected
397 supernatant was then centrifuged at 18,000*g* for 10 min and 1.5 ml of the supernatant was
398 collected for LC-MS analysis. To extract the metabolites from *Chrysochromulina* sp.
399 CCMP291, 20 ml of sample was centrifuged at 600*g* for 10 min and the cell pellet was
400 resuspended in a total volume of 0.7 ml of the extraction solvent and mixed by pipetting and
401 vortexing for 2 min. Samples were then centrifuged at 18,000*g* for 10 min and 0.5 ml of the
402 supernatant was collected for LC-MS analysis.

403 LC-MS was carried out using a Shimadzu Ultra High Performance Liquid Chromatography
404 (UHPLC) system formed by a Nexera X2 LC-30AD Pump, a Nexera X2 SIL-30AC
405 Autosampler, a Prominence CTO-20AC Column oven, and a Prominence SPD-M20A Diode
406 array detector; and a Shimadzu LCMS-2020 Single Quadrupole Liquid Chromatograph Mass

Spectrometer. Samples were analysed in hydrophilic interaction chromatography (HILIC) mode using a Phenomenex Luna NH2 column (100 x 2 mm with a particle size of 3 µm) at pH 3.75. Mass spectrometry spray chamber conditions were capillary voltage 1.25 kV, oven temperature 30 °C, desolvation temperature 250 °C and nebulising gas flow 1.50 L min^{-1}. Solvent A is 5% acetonitrile + 95% 5 mM ammonium formate in water. Solvent B is 95% acetonitrile + 5% 100 mM ammonium formate in water. Flow rate was 0.6 ml min^{-1} and gradient (% solvent A/B) was t = 1 min, 100% B; t = 3.5 min, 70% B; t = 4.1 min, 58% B; t = 4.6 min, 50% B; t = 6.5 min, 100% B; t = 10 min, 100% B. The injection volume was 15 µl. All samples were analysed immediately after being extracted. The targeted mass transition corresponded to $[M+H]^+$ of DMSP (m/z 135) and of glycine betaine (m/z 118) in positive mode. A calibration curve was performed for quantification of DMSP and glycine betaine using a mixture of DMSP and glycine betaine standards in the extraction solvent.

Reverse transcription quantitative PCR (RT-qPCR)

For each culture, RNA was extracted as follows: 1 ml Trizol reagent (Sigma-Aldrich), prewarmed at 65 °C, was added directly to the frozen phytoplankton filter (see 'Sampling methods'), followed by 600 mg of < 106 µm glass beads (Sigma-Aldrich). Cells were disrupted using an MP FastPrep®-24 instrument set at maximum speed for 3 x 30 seconds. Following a 5 min recovery time at 22 °C, samples were centrifuged at 13,000*g*, 4 °C, for 2 min. The supernatant was transferred to a 2 ml screwcap tube containing 1 ml 95% ethanol and RNA was extracted using a Direct-zol™ RNA MiniPrep kit (Zymo Research, R2050), according to the manufacturer's specifications.

Genomic DNA was removed by treating samples with TURBO DNA-free™ DNAse (Ambion®) according to the manufacturer's protocol. The quantity and quality of the RNA

was determined by a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific) using 1 µl of sample.

Reverse transcription of 1 µg DNA-free RNA was achieved using the QuantiTect® Reverse Transcription Kit (Qiagen). Primers (Supplementary Table 11) were designed, using Primer3Plus[43], to amplify ~130 bp region, with an optimum melting temperature of 60 °C. Melting temperature difference between primers in a pair was 2 °C and GC content was kept between 40-60%.

Quantitative PCR was performed with a C1000 Thermal cycler equipped with a CFX96 Real-time PCR detection system (BioRad), using a SensiFAST^{TM} SYBR® Hi-ROX Kit (Bioline) as per the manufacturer's instructions for a 3-step cycling programme. Reactions (20 µl) contained 50 ng cDNA and a final concentration of 400 nM of each primer, with a 60 °C annealing temperature. Gene expression for each condition was performed upon three biological replicates, each with three technical replicates. Control DNA consisted of pGEMT-Easy (Promega) containing the fragment created by the RT-qPCR primer pair for each gene tested (made through PCR on synthesised cDNA, cloning in *E. coli* 803 and purifying via a Miniprep Kit [Qiagen]).

For each condition and gene, the cycle threshold (Ct) values of the technical and biological replicates were averaged and manually detected outliers were excluded from further analysis. Standard curves of control DNA were calculated from 3 points of 1:10 serial dilutions, starting with 0.01 ng, to absolutely quantify the *DSYB* transcripts for comparison between organisms[44]. For an individual organism, relative *DSYB* expression was normalised to the β-actin housekeeping gene, and calculated using the $2^{-\Delta\Delta CT}$ method[45] to observe changes in response to various conditions.

Analysis of DSYB expression by Western blotting

A polyclonal rabbit IgG was designed against *P. parvum* DSYB using the OptimumAntigen™ software (GenScript Ltd.). The purified IgG was used as a primary antibody in Western blotting and immunogold labelling (see 'DSYB immunogold labelling'). The specificity of this antibody was ensured by Western blot analysis of DSYB expressed in the heterologous host *R. leguminosarum* J391. J391 strains containing pBIO2275 (positive control) and pRK415 with no cloned insert (negative control) were grown overnight in TY medium with 0.5 mM IPTG. Proteins were extracted by harvesting 1 ml culture, resuspending cell pellet in 200 µl 20 mM HEPES, 150 mM NaCl, pH 7.5 and disrupting with an ultrasonic processor (Cole Palmer) for 2 x 10 s cycles on ice. Cell debris was separated by centrifugation at 18,000*g* for 10 mins, following which the supernatant was mixed with SDS sample buffer and incubated at 95 °C for 5 min, before resolution on a 15 % (v/v) acrylamide gel.

The specificity of the anti-DSYB antibody was additionally tested on *P. parvum* 946/6, where protein samples were prepared from cell pellets (see 'Sampling methods') as for *R. leguminosarum*, without the removal of cell debris. Cell lysate containing 5.5 µg protein was mixed with SDS sample buffer and heat-treated at 95 °C for 20 min, before resolution on a 15 % (v/v) acrylamide gel.

Following SDS-PAGE, proteins were transferred to a PVDF membrane (Amersham Hybond™-P, GE Healthcare) by semi-dry Western blot as outlined by Mahmood and Yang[46]. After 1 hour blocking with 5 % (w/v) skimmed milk powder in TBS (20 mM Tris, 150 mM NaCl, pH 7.5), the anti-DSYB antibody was added at a final concentration of 0.386 µg ml^{-1}. Specific interactions were left to form overnight at 4 °C, before the membrane was washed 4 x 10 min with TBST (TBS + 0.1 % (v/v) Tween 20). TBST (20 ml) was added with 3 µl anti-

rabbit IgG-alkaline phosphatase at 1 mg ml^{-1} (Sigma). Following 1 h incubation, the membrane was washed as before with two 10 min TBS washes. Colorimetric detection with NBT/BCIP (Thermo Fisher) was used to detect the target protein as per the manufacturer's instructions. All SDS-PAGE gels were run with Bio-Rad Precision Plus Dual Colour protein size standards and stained with Coomassie using InstantBlue Protein stain (Expedeon).

Purification of DSYB and *in vitro* catalytic assays

A 1.1 kb fragment of DNA containing the synthesised coding region of *Chrysochromulina tobin DSYB* was subcloned (from pBIO2272) into pET16b as an *Nde*I/*Eco*RI restriction fragment, downstream of a 10-histidine coding sequence, and transformed into *E. coli* BL21 DE3 (New England BioLabs), for protein purification. Batch cultures were grown aerobically in LB medium at 37 °C until reaching an OD_{600} value of ~0.6 and were then supplemented with 0.2 mM IPTG and incubated at 28 °C overnight to induce recombinant protein expression. Cells were harvested at 5,000*g* for 20 min and resuspended in buffer A (20 mM HEPES, 150 mM NaCl, 25 mM imidazole, pH 7.5). The mixture was supplemented with protease inhibitor (Roche cOmplete Tablets, Mini EDTA-free, *EASY*pack (cat. no. 04 693 159 001)), lysed via sonication and separated at 15,000*g*, 4 °C for 30 min.

DSYB was purified via an immobilized metal affinity chromatography (IMAC, HiTrap Chelating HP, GE Healthcare) column charged with $NiSO_4$ and equilibrated with buffer A. All steps were performed at 24 °C with a flow rate of 1 ml min^{-1}. Soluble cell lysate was loaded and washed through with 4 column volumes of buffer A. Bound protein was eluted into 1 ml fractions using a stepped gradient of 25 to 150 mM imidazole, applied for 2 column volumes each. Fractions were visualised via SDS-PAGE analysis (Supplementary Fig. 7) and

those containing DSYB were pooled and dialysed at 4 °C overnight against 20 mM HEPES, 150 mM NaCl, pH 7.5.

P. parvum lysate was prepared by centrifuging 100 ml of culture at late exponential phase for 10 min at 2,500*g*. The pellet was washed with 20 mM HEPES, 150 mM NaCl, pH 7.5 and resuspended in 2 ml buffer supplemented with EDTA-free protease inhibitor (Roche cOmplete Tablets, Mini EDTA-free, *EASY*pack (cat. no. 04 693 159 001)). Cells were sonicated 3 x 10 s to lyse, with a 50 s recovery time at 4 °C. Resulting lysate was heat-treated at 80 °C for 10 min to denature proteins (ensuring no activity from native DSYB protein) and centrifuged for 2 min 14,000*g*. Supernatant was removed to a fresh Eppendorf tube and used for downstream catalytic assays.

DSYB MTHB methyltransferase activity was monitored by performing *in vitro* enzyme assays in 400 µl reactions with 50 µl *P. parvum* lysate and 350 µl purified DSYB (~0.1 mg ml^{-1}) or buffer. All enzyme substrates were added to a final concentration of 1 mM and reactions were incubated at 28 °C for 30 mins. Following this, 800 µl of finely ground charcoal (38 mg ml^{-1} in 0.1 M acetic acid) was added to the samples and mixed to remove SAM. Samples were centrifuged for 10 mins, 14,000*g* and the supernatant was retained. For GC analysis, 200 µl of the supernatant was added to a 2 ml vial, 100 µl 10 M NaOH was added, vials were crimped immediately, then heated at 80 °C for 10 minutes (to release DMS from DMSHB) and finally incubated at 22 °C for 24 h in the dark. These samples were subsequently used for quantification of DMSHB by GC analysis as described earlier and activities are reported as nmol DMSHB mg $protein^{-1}$ min^{-1}. DMS produced from background DMSHB/DMSP present in the *P. parvum* lysate was subtracted from the reported activities.

DSYB immunogold labelling

Cells from *P. parvum* 946/6 were cryoimmobilized using a Leica EMPACT High-Pressure Freezer (Leica Microsystems), freeze-substituted in an EM AFS (Leica Microsystems) and embedded in Lowicryl HM20 resin (EMS, Hatfield, USA) as in Perez-Cruz et al.[47]. Gold grids containing Lowicryl HM20 ultrathin sections were immunolabeled with a specific primary antibody to *P. parvum* DSYB (polyclonal rabbit IgG, GenScript), whose stock concentration was 0.550 mg ml^{-1} and this was diluted 1:15,000. Secondary antibody was an IgM anti-rabbit coupled to 12 nm diameter colloidal gold particles (Jackson) diluted 1:30. As controls, pre-immune rabbit serum was used as primary antibody, or the gold-conjugated secondary antibody was used without the primary antibody. Sections were observed in a Tecnai Spirit microscope (FEI, Eindhoven, The Netherlands) at 120 kV.

Prymnesium **growth and experimental conditions for NanoSIMS**

P. parvum were grown as previously described in F/2 medium (35 PSU)[30]. Sodium sulfate (Na_2SO_4, 25 mM) was used as the sole sulfur source, with either ^{34}S (90% ^{34}S (Sigma-Aldrich, USA; hereafter called ^{34}S-F/2) or natural abundance of ^{32}S (95% ^{32}S, 0.7% ^{33}S, 4.2% ^{34}S; hereafter called ^{nat}S-F/2). Consequently, the composition of the both the trace metals and vitamin complement had to be slightly modified (with Riboflavin replacing the sulfur-containing Biotin and Thiamine)[22]. *P. parvum* cells in late exponential phase (grown in ^{nat}S-F/2) were centrifuged at low speed (1,000*g*) for 5 mins, rinsed with ^{34}S-F/2 (to remove potential leftover ^{nat}S) and transferred in ^{34}S-F/2, whereas a batch incubated only in ^{nat}S-F/2 acted as a control. Culture were sampled at four time-points: directly after the medium exchange, and after 6 hrs, 24 hrs and 48 hrs. At each timepoint, cultures were sampled for NanoSIMS, mass-spectrometry and cell counts (see below).

550 **Flow cytometry for NanoSIMS samples**

551 Cells were enumerated in triplicate *via* flow cytometry (BD Accuri C6, Becton Dickinson,
552 USA). For each sample, forward scatter (FSC), side scatter (SSC), and red (chlorophyll)
553 fluorescence were recorded. The samples were analysed at a flow rate of 35 µl min^{-1}.
554 *Prymnesium* populations were characterized according to SSC and chlorophyll fluorescence
555 and cell abundances were calculated by running a standardized volume of sample (50 µl).

556

557 **Sample collection for mass spectrometry (NanoSIMS)**

558 At each time point, 1 ml of culture was centrifuged at low speed (1,000*g*) for 5 mins, the
559 supernatant was discarded and the cell pellet was extracted with 80% methanol, sonicated on
560 ice for 30 mins and dried.

561 Dried extracts were reconstituted in methanol to perform LC-MRM-MS analysis. The LC-
562 MS system consisted of an Agilent 1290 series LC interfaced to an Agilent G6490A QQQ
563 mass spectrometer (Agilent, Santa Clara, CA, USA). The MS was equipped with an
564 electrospray ionization source and was controlled by Mass Hunter workstation (version B07)
565 software. A HILIC column (Luna Phenomenex, 150×3 mm, 5 µm, 300 Å) was used for the
566 on-line separations, at a flow rate of 1 ml min^{-1}. The gradient used consisted of a 95 %
567 solvent B (Acetonitrile, 0.1% formic acid), followed by a 2 min linear gradient to 40%
568 solvent A (Milli Q, 0.1 % formic acid), then a 10 min linear gradient to 90% A, and returning
569 to initial conditions at 12.25 min. The injection volume was 2 µl. The MS acquisition
570 parameters were: positive ion mode; capillary voltage, 3,000 V; gas flow 12 l min^{-1}; nebulizer
571 gas, 20 p.s.i.; sheath gas flow rate 7 l/min^{-1} at a temperature of 250 °C. Acquisition was done
572 in MRM mode with transitions m/z 135- > 63 and m/z 137- > 65 for quantifying ^{32}DMSP and

573 ^{34}DMSP respectively. The collision energy was optimised as 10 eV to detect the highest
574 possible intensity.

575

576 **Sample collection and preparation for NanoSIMS**

577 Samples for NanoSIMS were collected and processed following the method described by
578 Raina et al.[22]. Briefly, samples were snap-frozen, and embedded following by a water-free
579 embedding procedure to effectively prevent the loss of highly soluble compounds such as
580 DMSP from the samples. This method does retain elements in solution by effectively
581 replacing the 'solution' with resin, without displacing the ions and osmolytes. *Prymnesium*
582 cultures (20 µl) were dropped onto Thermanox strips (Thermo Fisher Scientific, Waltham,
583 USA, 4×18 mm) and placed in humidified chambers. After 20 min, the cells settled onto the
584 strips and the excess medium was carefully removed with filter paper. The strips were then
585 immediately snap-frozen by immersion into liquid nitrogen slush[22]. Samples were stored in
586 liquid nitrogen until required. Frozen samples for NanoSIMS were freeze-substituted in
587 anhydrous 10% acrolein in diethyl ether, and warmed progressively to room temperature over
588 three weeks in an EM AFS2 automatic freeze-substitution unit (Leica Microsystems, Wetzlar,
589 Germany) as described recently in step-by-step detail by Kilburn and Clode[48]. The samples
590 were subsequently infiltrated and embedded in anhydrous Araldite 502 resin, after which the
591 Thermanox strip was removed and the sample re-embedded and stored in a desiccator. No
592 sulfur was present in processing or resin components. Resin sections (1 mm thick) of
593 embedded *Prymnesium* cells were cut dry using a Diatome-Histo diamond knife on an EM
594 UC6 Ultramicrotome (Leica Microsystems, Wetzlar, Germany), mounted on a silicon wafer
595 and coated with 10 nm of gold.

596

NanoSIMS analysis

The NanoSIMS-50L (Cameca, Gennevilliers, France) at the Centre for Microscopy, Characterisation and Analysis (CMCA) at the University of Western Australia was used for all subsequent analyses. The NanoSIMS-50L allows simultaneous collection and counting of seven isotopic species, which enables the determination of $^{34}S/^{32}S$ ratio. Enrichments of the rare isotope ^{34}S was confirmed by an increase in the sulfur ($^{34}S/^{32}S$) ratio above natural abundance values recorded in controls (0.0438). NanoSIMS analysis was undertaken by rastering a 2.5 pA Cs^+ beam (~100 nm diameter) across defined 20 μm^2 sample areas (256×256 pixels), with a dwell time of 30 ms per pixel. The isotope ratio values are represented hereafter using a colour-coded transform (hue saturation intensity (HSI)) showing natural abundance levels in blue, and grading to high enrichment in pink. Images were processed and analysed using Fiji (http://fiji.sc/Fiji)[49] with the Open-MIMS plug-in (http://nrims.harvard.edu/software). All images were dead-time corrected[50]. Ratio data were tested for QSA (quasi-simultaneous arrivals) by applying different beta values from 0.5 to 162. No differences in the data were observed, indicating that the secondary ion count rates were too low to be affected by QSA. Quantitative data were extracted from the mass images through manually drawn regions of interest, at T0 (whole cells n = 7, hotspot n = 10, chloroplasts n = 3), at T6 (whole cells n = 14, hotspot n = 10, chloroplasts n = 6), at T24 (whole cells n = 12, hotspot n = 10, chloroplasts n = 9), and at T48 (whole cells n = 6, hotspot n = 10, chloroplasts n = 4).

Statistics

Statistical methods for RT-qPCR are described in the relevant section above. All measurements for DMSP production or DSYB/DsyB enzyme activity (in algal strains or

621 enzyme assays) are based on the mean of at least three biological replicates per
622 strain/condition tested, with all experiments performed at least twice. To identify statistically
623 significant differences between standard and experimental conditions in Supplementary Fig.
624 2, a two-tailed independent Student's *t*-test ($P<0.05$) was applied to the data, using R[51].

625

626 **Identification of DSYB proteins in eukaryotes**

627 BLASTP and TBLASTN searches[52] were used to identify homologues of the *Labrenzia* DsyB
628 protein in available eukaryotic genomes and/or transcriptome assemblies at NCBI or JGI.
629 Any eukaryotic DsyB-like proteins (E values $\leq 1e^{-30}$), were aligned to ratified bacterial DsyB
630 sequences and to non-functional DsyB-like proteins, e.g., in *Streptomyces varsoviensis*, see
631 below. Representative DsyB-like proteins, more similar to DsyB than to non-functional *S.*
632 *varsoviensis* DsyB-like proteins, were cloned and assayed for MMT activity (as above).

633 Ratified eukaryotic DSYB peptide sequences were used in BLASTP searches of 119
634 eukaryotic transcriptomes (with replicates) downloaded from the Marine Microbial
635 Eukaryote Transcriptome Sequencing Project (MMETSP)[53] via the sequencing repositories
636 iMicrobe (http://imicrobe.us/project/view/104) and ENA (European Nucleotide Archive)[54].
637 Of these, 45 contained at least one hit to DSYB (E values $\geq 1e^{-30}$) (Supplementary Table 3).
638 Each potential DsyB/DSYB sequence was manually curated by BLASTP analysis against the
639 RefSeq database and discounted as a true DSYB sequence if the top hits were not to ratified
640 DSYB sequences detailed in Fig. 1b. DSYB sequences identified from iMicrobe
641 transcriptomes were aligned to ratified DsyB and DSYB sequences and included in the
642 evolutionary analysis (Fig. 1b). All DsyB and DSYB protein sequences identified from
643 genomes or transcriptomes are listed in Supplementary Data 1. Kallisto[55] was used to

quantify transcript abundances. Firstly, Kallisto indexes were created for the combined nucleotide assemblies of each organism. Next, Kallisto quant was used to obtain Transcripts Per kilobase Million (TPM) expression values for all datasets using the relevant reference transcriptome index for that organism. Nucleotide sequences corresponding to the DSYB hits were obtained using TBLASTN, and the CAMNT ID number was used to identify the TPM values for each *DSYB* read, giving an estimate of gene expression for organisms grown in standard conditions.

Phylogenetic analysis of DSYB and DsyB proteins

All prokaryotic DsyB and eukaryotic DSYB amino acid sequences were aligned in MAFFT[56,57] version 7 using default settings, then visually checked. Prior to phylogeny construction, model selection was carried out and the best supported model of sequence evolution based on the Bayesian Information Criterion (BIC)[58] was selected for phylogeny construction (the LG+I+G4 model[59]). A maximum likelihood phylogeny was then constructed using IQ-TREE[60] version 1.5.3, implemented in the W-IQ-TREE web interface[61], with 1,000 ultrafast bootstrap replicates[62] used to assess node support. The resulting tree was rooted using a non-DsyB methyltransferase sequence from *Streptomyces varsoviensis*[5], and was formatted for publication using the ggtree package[63] in R[51].

Analysis of DSYB sequences for localisation signals

Searches for localisation signals in the DSYB protein sequences used the prediction software packages SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), TargetP 1.1

(http://www.cbs.dtu.dk/services/TargetP/) and ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/).

Analysis of marine metagenomes and metatranscriptomes

Hidden Markov Model (HMM)-based searches for *dsyB* and *DSYB* homologs in metagenome and metatranscriptome datasets were performed as described in[64] using HMMER tools (version 3.1, http://hmmer.janelia.org/). The DsyB/DSYB protein sequences, shown in Fig. 1b, and ratified DddD[65–68], DddK[69], DddL[70], DddP[71], DddQ[72], DddY[73] DddW[74] and Alma1[75] sequences were used as training sequences to create the HMM profiles. Profile HMM-based searches eliminate the bias associated with single sequence BLAST queries[76]. HMM profiles for the *recA* gene were downloaded from the functional gene pipeline and repository (FunGene[77]). The *Ruegeria pomeroyi* DddW[74] sequence was used to search metagenome and metatranscriptome datasets via BLASTP[52] since it is the only ratified DddW. HMM and BLASTP searches were performed against peptide sequences predicted from OM-RGC database assemblies (Supplementary Table 6) and all hits with an E value cut-off of $1e^{-30}$ were retrieved. In the case of metatranscriptome datasets (*Tara* Oceans and GeoMICS metatranscriptomes), homologs with an E value cutoff of $1e^{-5}$ were retrieved. Each potential DsyB/DSYB sequence retrieved from the analysis of metagenomes and metatranscriptomes was manually curated by BLASTP analysis against the RefSeq database and discounted as a true DsyB sequence if the top hits were not to DsyB or DSYB sequences detailed in Fig. 1b. If the top hits were to eukaryotic DSYB then the sequence was counted as a true DSYB sequence, and vice versa for bacterial DsyB. Each of the DddD, DddK, DddL, DddP, DddQ, DddW, DddY and Alma1 peptide sequences retrieved were aligned to curated reference sequences using hmmalign and an approximate maximum likelihood tree was constructed

using FastTree[78] v2.1. Putative Ddd or Alma1 peptide sequences not aligning most closely to functional Ddd or Alma1 enzymes were removed. To estimate the percentage of bacteria containing *dsyB*, the number of unique hits to DysB in metagenomes was normalised to the number of RecA sequences. Retrieved DsyB/DSYB homolog sequences were aligned to the training sequences using the *dsyB* HMM alignment and this was used to construct an approximately maximum likelihood phylogenetic tree inferred using FastTree[78] v2.1. The resulting tree (Supplementary Fig. 6) was visualised and annotated using the Interactive Tree Of Life (iTOL)[79] version 3.2.4.

The GeoMICS metatranscriptome database[29] generated from North Pacific Ocean samples offered an opportunity to compare prokaryotic and eukaryotic gene expression. Sequences from both the 0.2 μm – 2 μm and 2 μm – 53 μm filtrate fractions for sites P1 and P6 (those samples that had duplicates) were obtained from NCBI (Accession: PRJNA272345) (Supplementary Table 7). Sequences were trimmed using TrimGalore (default parameters, paired-end mode, https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and overlapping paired-end reads were joined using PandaSeq[80]. To create peptide databases, the joined reads were translated using the translate function in Sean Eddy's squid package (http://selab.janelia.org/software.html) to generate all ORFs above 20 amino acids in length. The resulting peptide sequences were used to retrieve *dsyB* and *DSYB* sequences using HMM searches and BLASTP (as above). Read numbers for *dsyB*/*DSYB* were normalised to the read numbers of internal standard[29] recovered in each sample by dividing the number of reads by the internal standard number and multiplying by 100. Normalised reads from the same site and fraction were averaged (Supplementary Table 9) and ratios of *dsyB*/*DSYB* calculated.

Data availability statement The datasets analysed during the current study are available in the iMicrobe (https://www.imicrobe.us/#/projects/104), European Nucleotide Archive (https://www.ebi.ac.uk/ena), NCBI (https://www.ncbi.nlm.nih.gov/) and Ocean Microbiome (http://ocean-microbiome.embl.de/companion.html) repositories or are available within the paper in Methods section 'Analysis of marine metagenomes and metatranscriptomes' and in Supplementary Tables 7, 8 and 9. All data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Nevitt, G. A. The neuroecology of dimethyl sulfide: a global-climate regulator turned marine infochemical. *Integr. Comp. Biol.* **51,** 819–825 (2011).

2. Sievert, S. M., Kiene, R. P. & Schulz-Vogt, H. N. The sulfur cycle. *Oceanography* **20,** 117–123 (2007).

3. Curson, A. R., Todd, J. D., Sullivan, M. J. & Johnston, A. W. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nat. Rev. Microbiol.* **9,** 849–859 (2011).

4. Summers, P. S. *et al.* Identification and stereospecificity of the first three enzymes of 3-dimethylsulfoniopropionate biosynthesis in a chlorophyte alga. *Plant Physiol.* **116,** 369–378 (1998).

5. Curson, A. R. *et al.* Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. *Nat. Microbiol.* **2,** 17009 (2017).

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

7. Lyon, B. R., Lee, P. A., Bennett, J. M., DiTullio, G. R. & Janech, M. G. Proteomic

736 analysis of a sea-ice diatom: salinity acclimation provides new insight into the
737 dimethylsulfoniopropionate production pathway. *Plant Physiol.* **157,** 1926–1941
738 (2011).

739 8. Raina, J. B. *et al.* DMSP biosynthesis by an animal and its role in coral thermal stress
740 response. *Nature* **502,** 677–680 (2013).

741 9. Keller, M. D., Bellows, W. K. & Guillard, R. R. L. *Dimethyl sulfide production in*
742 *marine phytoplankton*. *Biogenic sulfur in the environment* (American Chemical
743 Society, 1989).

744 10. Nei, M. & Rooney, A. P. Concerted and birth-and-death evolution of multigene
745 families. *Annu. Rev. Genet.* **39,** 121–152 (2005).

746 11. Ku, C. *et al.* Endosymbiotic origin and differential loss of eukaryotic genes. *Nature*
747 **524,** 427–432 (2015).

748 12. Baumgarten, S. *et al.* The genome of *Aiptasia*, a sea anemone model for coral
749 symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 11893–11898 (2015).

750 13. Van Alstyne, K. L. & Puglisi, M. P. DMSP in marine macroalgae and
751 macroinvertebrates: Distribution, function, and ecological impacts. *Aquat. Sci.* **69,**
752 394–402 (2007).

753 14. Spielmeyer, A. & Pohnert, G. Influence of temperature and elevated carbon dioxide on
754 the production of dimethylsulfoniopropionate and glycine betaine by marine
755 phytoplankton. *Mar. Environ. Res.* **73,** 62–69 (2012).

756 15. Dickschat, J. S., Rabe, P. & Citron, C. A. The chemical biology of
757 dimethylsulfoniopropionate. *Org. Biomol. Chem.* **13,** 1954–1968 (2015).

758 16. Hovde, B. T. *et al.* Genome sequence and transcriptome analyses of *Chrysochromulina*

tobin: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae). *PLoS Genet.* **11,** (2015).

17. Jones, H. L. J., Leadbeater, B. S. C. & Green, J. C. Mixotrophy in marine species of *Chrysochromulina* (Prymnesiophyceae) - ingestion and digestion of a small green flagellate. *J. Mar. Biol. Assoc. United Kingdom* **73,** 283–296 (1993).

18. Kettles, N. L., Kopriva, S. & Malin, G. Insights into the regulation of DMSP synthesis in the diatom *Thalassiosira pseudonana* through APR activity, proteomics and gene expression analyses on cells acclimating to changes in salinity, light and nitrogen. *PLoS One* **9,** (2014).

19. Dickson, D. M. J. & Kirst, G. O. Osmotic adjustment in marine eukaryotic algae - the role of inorganic-ions, quaternary ammonium, tertiary sulfonium and carbohydrate solutes .2. Prasinophytes and Haptophytes. *New Phytol.* **106,** 657–666 (1987).

20. Trossat, C. *et al.* Salinity promotes accumulation of 3-dimethylsulfoniopropionate and its precursor S-methylmethionine in chloroplasts. *Plant Physiol.* **116,** 165–171 (1998).

21. Gruber, A. *et al.* Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol. Biol.* **64,** 519–530 (2007).

22. Raina, J. B. *et al.* Subcellular tracking reveals the location of dimethylsulfoniopropionate in microalgae and visualises its uptake by marine bacteria. *Elife* **6,** (2017).

23. Matrai, P. A. & Keller, M. D. Total organic sulfur and dimethylsulfoniopropionate in marine phytoplankton: intracellular variations. *Mar. Biol.* **119,** 61–68 (1994).

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

25. 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).

26. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science (80-.).* **348,** (2015).

27. Johnston, A. W. B., Green, R. T. & Todd, J. D. Enzymatic breakage of dimethylsulfoniopropionate - a signature molecule for life at sea. *Curr. Opin. Chem. Biol.* **31,** 58–65 (2016).

28. Belviso, S. *et al.* Size distribution of dimethylsulfoniopropionate (DMSP) in areas of the tropical northeastern Atlantic Ocean and the Mediterranean Sea. *Mar. Chem.* **44,** 55–71 (1993).

29. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522,** 98–101 (2015).

30. Guillard, R. R. L. Culture of phytoplankton for feeding marine invertebrates. in *Culture of Marine Invertebrate Animals* (ed. Smith Chanley, M. H., W. L.) 26–60 (Plenum Press, 1975).

31. Berges, J. A., Franklin, D. J. & Harrison, P. J. Evolution of an artificial seawater medium: Improvements in enriched seawater, artificial water over the last two decades. *J. Phycol.* **37,** 1138–1145 (2001).

32. Mock, T. *et al.* Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541,** 536–540 (2017).

33. Fixen, K. R. *et al.* Genome sequences of eight bacterial species found in coculture with the haptophyte *Chrysochromulina tobin*. *Genome Announc.* **4,** e01162-16 (2016).

34. Sambrook, J., Fritsch, E. F., Maniatis, T. & Nolan, C. *Molecular cloning, a laboratory*

manual. **3,** (Cold Spring Harbor Laboratory Press, 1989).

35. Beringer, J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84,** 188–198 (1974).

36. Gonzalez, J. M., Whitman, W. B., Hodson, R. E. & Moran, M. A. Identifying numerically abundant culturable bacteria from complex communities: An example from a lignin enrichment culture. *Appl. Environ. Microbiol.* **62,** 4433–4440 (1996).

37. Baumann, P. & Baumann, L. *The marine Gram-negative eubacteria: genera* Photobacterium, Benecekea, Alteromonas, Pseudomonas *and* Alcaligenes. *In The Prokaryotes* (Springer-Verlag, 1981).

38. Porter, K. G. & Feig, Y. S. The use of DAPI for identifying and counting aquatic microflora1. *Limnol. Oceanogr.* **25,** 943–948 (1980).

39. Figurski, D. H. & Helinski, D. R. Replication of an origin-containing derivative of plasmid Rk2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 1648–1652 (1979).

40. Downie, J. A. *et al.* Cloned nodulation genes of *Rhizobium leguminosarum* determine host range specificity. *Mol. Gen. Genet.* **190,** 359–365 (1983).

41. Keen, N. T., Tamaki, S., Kobayashi, D. & Trollinger, D. Improved broad-host-range plasmids for DNA cloning in Gram-negative bacteria. *Gene* **70,** 191–197 (1988).

42. Tett, A. J., Rudder, S. J., Bourdes, A., Karunakaran, R. & Poole, P. S. Regulatable vectors for environmental gene expression in alphaproteobacteria. *Appl. Environ. Microbiol.* **78,** 7137–7140 (2012).

43. Untergasser, A. *et al.* Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **40,** (2012).

828 44. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR.
829 *Genome Res.* **6,** 986–994 (1996).

830 45. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-
831 time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25,** 402–408
832 (2001).

833 46. Mahmood, T. & Yang, P. C. Western blot: technique, theory, and trouble shooting. *N.*
834 *Am. J. Med. Sci.* **4,** 429–434 (2012).

835 47. Perez-Cruz, C. *et al.* New type of outer membrane vesicle produced by the Gram-
836 negative bacterium *Shewanella vesiculosa* M7T: implications for DNA content. *Appl.*
837 *Environ. Microbiol.* **79,** 1874–1881 (2013).

838 48. Kilburn, M. R. & Clode, P. L. Electron microscopy. in (Humana Press, 2014).

839 49. Schindelin, J. *et al.* Fiji: An open source platform for biological image analysis. *Nat.*
840 *Methods* **9,** 676–682 (2012).

841 50. Hillion, F., Kilburn, M. R., Hoppe, P., Messenger, S. & Webers, P. K. The effect of
842 QSA on S, C, O and Siisotopic ratio measurements. *Geochim. Cosmochim. Acta* **72,**
843 A377 (2008).

844 51. Team, R. D. C. *R: A Language and Environment for Statistical Computing*. (R
845 Foundation for Statistical Computing, 2008).

846 52. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein
847 database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

848 53. Keeling, P. J. *et al.* The marine microbial eukaryote transcriptome sequencing project
849 (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans
850 through transcriptome sequencing. *PLoS Biol.* **12,** (2014).

851 54. Toribio, A. L. *et al.* European nucleotide archive in 2016. *Nucleic Acids Res.* **45,** D32–
852 D36 (2017).

853 55. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-
854 seq quantification. *Nat. Biotechnol.* **34,** 525–527 (2016).

855 56. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid
856 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,**
857 3059–3066 (2002).

858 57. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
859 Improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).

860 58. Schwarz, G. Estimating dimension of a model. *Ann. Stat.* **6,** 461–464 (1978).

861 59. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol.*
862 *Biol. Evol.* **25,** 1307–1320 (2008).

863 60. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and
864 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*
865 *Biol. Evol.* **32,** 268–274 (2015).

866 61. Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast
867 online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44,**
868 W232–W235 (2016).

869 62. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast approximation for
870 phylogenetic bootstrap. *Mol. Biol. Evol.* **30,** 1188–1195 (2013).

871 63. Yu, G. C., Smith, D. K., Zhu, H. C., Guan, Y. & Lam, T. T. Y. GGTREE: an R
872 package for visualization and annotation of phylogenetic trees with their covariates
873 and other associated data. *Methods Ecol. Evol.* **8,** 28–36 (2017).

64. Kumaresan, D. *et al.* Aerobic proteobacterial methylotrophs in Movile Cave: genomic and metagenomic analyses. *Microbiome* **6,** 1 (2018).

65. Todd, J. D. *et al.* Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science (80-.).* **315,** 666–669 (2007).

66. Todd, J. D. *et al.* Molecular dissection of bacterial acrylate catabolism--unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. *Environ. Microbiol.* **12,** 327–343 (2010).

67. Curson, A. R. J., Sullivan, M. J., Todd, J. D. & Johnston, A. W. B. Identification of genes for dimethyl sulfide production in bacteria in the gut of Atlantic Herring (*Clupea harengus*). *ISME J.* **4,** 144–146 (2010).

68. Curson, A. R. J., Fowler, E. K., Dickens, S., Johnston, A. W. B. & Todd, J. D. Multiple DMSP lyases in the gamma-proteobacterium *Oceanimonas doudoroffii*. *Biogeochemistry* **110,** 109–119 (2012).

69. Sun, J. *et al.* The abundant marine bacterium *Pelagibacter* simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol. *Nat. Microbiol.* **1,** (2016).

70. Curson, A. R., Rogers, R., Todd, J. D., Brearley, C. A. & Johnston, A. W. Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine alpha-proteobacteria and *Rhodobacter sphaeroides*. *Environ. Microbiol.* **10,** 757–767 (2008).

71. Todd, J. D., Curson, A. R. J., Dupont, C. L., Nicholson, P. & Johnston, A. W. B. The *dddP* gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi (vol 11, pg 1376, 2009). *Environ. Microbiol.* **11,**

1624–1625 (2009).

72. Todd, J. D. *et al.* DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. *Environ. Microbiol.* **13,** 427–438 (2011).

73. Curson, A. R. J., Sullivan, M. J., Todd, J. D. & Johnston, A. W. B. DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of Proteobacteria. *ISME J.* **5,** 1191–1200 (2011).

74. Todd, J. D., Kirkwood, M., Newton-Payne, S. & Johnston, A. W. B. DddW, a third DMSP lyase in a model Roseobacter marine bacterium, *Ruegeria pomeroyi* DSS-3. *ISME J.* **6,** 223–226 (2012).

75. Alcolombri, U. *et al.* Identification of the algal dimethyl sulfide-releasing enzyme: A missing link in the marine sulfur cycle. *Science (80-.).* **348,** 1466–1469 (2015).

76. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7,** (2011).

77. Fish, J. A. *et al.* FunGene: the functional gene pipeline and repository. *Front. Microbiol.* **4,** (2013).

78. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).

79. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44,** (2016).

80. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: PAired-eND Assembler for Illumina sequences. *BMC Bioinformatics* **13,** (2012).

Correspondence and requests for materials should be addressed to Jonathan D. Todd (jonathan.todd@uea.ac.uk).

Acknowledgements Funding from the Natural Environment Research Council (NE/J01138X/1, NE/M004449/1, NE/N002385/1 and NE/P012671/1) supported work in J.D.T.'s laboratory. B.T.W. was supported by a NERC EnvEast grant (NE/L002582/1) and A.B.M was supported by a BBSRC Norwich Research Park Biosciences Doctoral Training Partnership grant (BB/M011216/1). The NanoSIMS work was supported by an Australian Research Council Grant (DE160100636) to J-B.R. We thank Pamela Wells and Marco Giardina for general technical support, Thomas Mock for supplying *Fragilariopsis cylindrus*, and Robert Green, Ji Liu and Colin Murrell for advice and discussion of results. We also acknowledge the *Tara* Oceans Consortium for providing metagenomic sequence data, and the facilities at the Australian Microscopy & Microanalysis Research Facility at the Centre for Microscopy, Characterisation & Analysis, University of Western Australia, a facility funded by the University, State and Commonwealth Governments.

Author contributions J.D.T. wrote the paper, designed experiments and performed experiments (gene cloning, enzyme assays, bioinformatics) and analysed data; A.R.J.C. wrote the paper, designed experiments, performed experiments (gene cloning, enzyme assays, gas chromatography to quantify DMSP/DMSHB, phytoplankton growth experiments), analysed data and prepared figures/tables; B.T.W. performed experiments (bioinformatics analysis of DsyB/DSYB in transcriptomes, metagenomes and metatranscriptomes, phylogenetic tree construction), analysed data and prepared figures/tables; B.J.P. performed experiments (gene

cloning, RNA isolation, qRT-PCR experiments, protein purification, *in vitro* enzyme assays and Western Blots) and analysed data; L.P.S. performed experiments (gene cloning) and analysed data; A.B.M. performed experiments (LC-MS detection of DMSP and glycine betaine) and analysed data; P.P.L.R. performed experiments (phytoplankton growth experiments); D.K. performed experiments (bioinformatic analysis and phylogenetic tree construction); E.M. performed experiments (immunogold labelling, microscopy) and prepared figures; L.G.S. wrote the paper, performed experiments (evolutionary analysis of DsyB and DSYB sequences and phylogenetic tree construction) and prepared figures/tables; J-B.R. wrote the paper, performed experiments (NanoSIMS, LC-MRM-MS) and prepared figures; U.K. performed experiments (LC-MRM-MS); P.L.C. and P.G. performed experiments (NanoSIMS); O.C. designed antibodies and prepared materials for microscopy; S.M. performed experiments (bioinformatic analysis); R.A.C. supplied *C. tobin* CCMP291 strain. All authors reviewed the manuscript before submission.

Competing interests

The authors declare no competing financial interests.

Additional Information

Supplementary Information is linked to the online version of the paper.

Reprints and permissions information is available at www.nature.com/reprints.

Figure legends

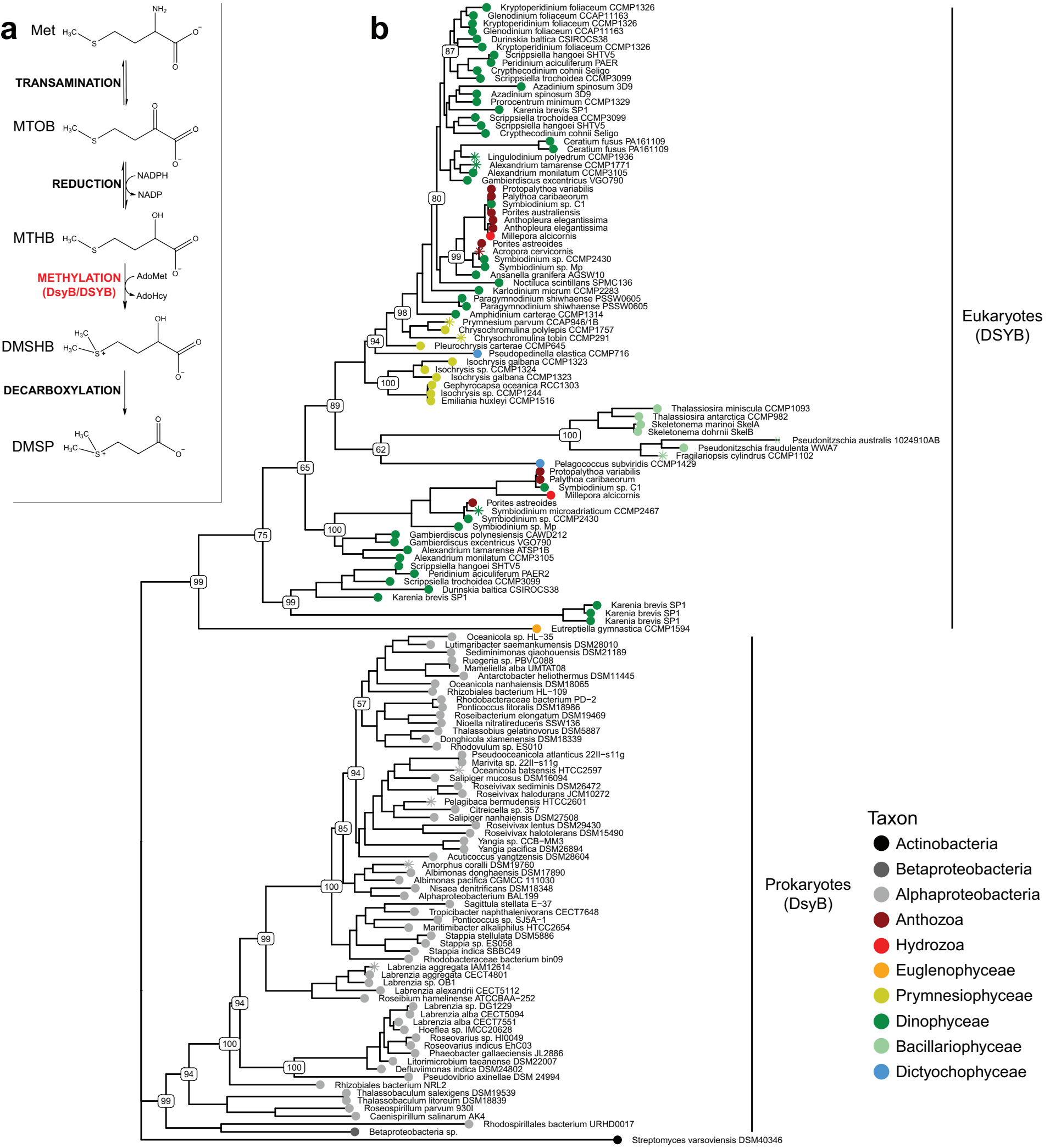
Figure 1. Transamination pathway for DMSP biosynthesis pathway in bacteria and marine algae, and phylogenetic tree of DsyB/DSYB proteins

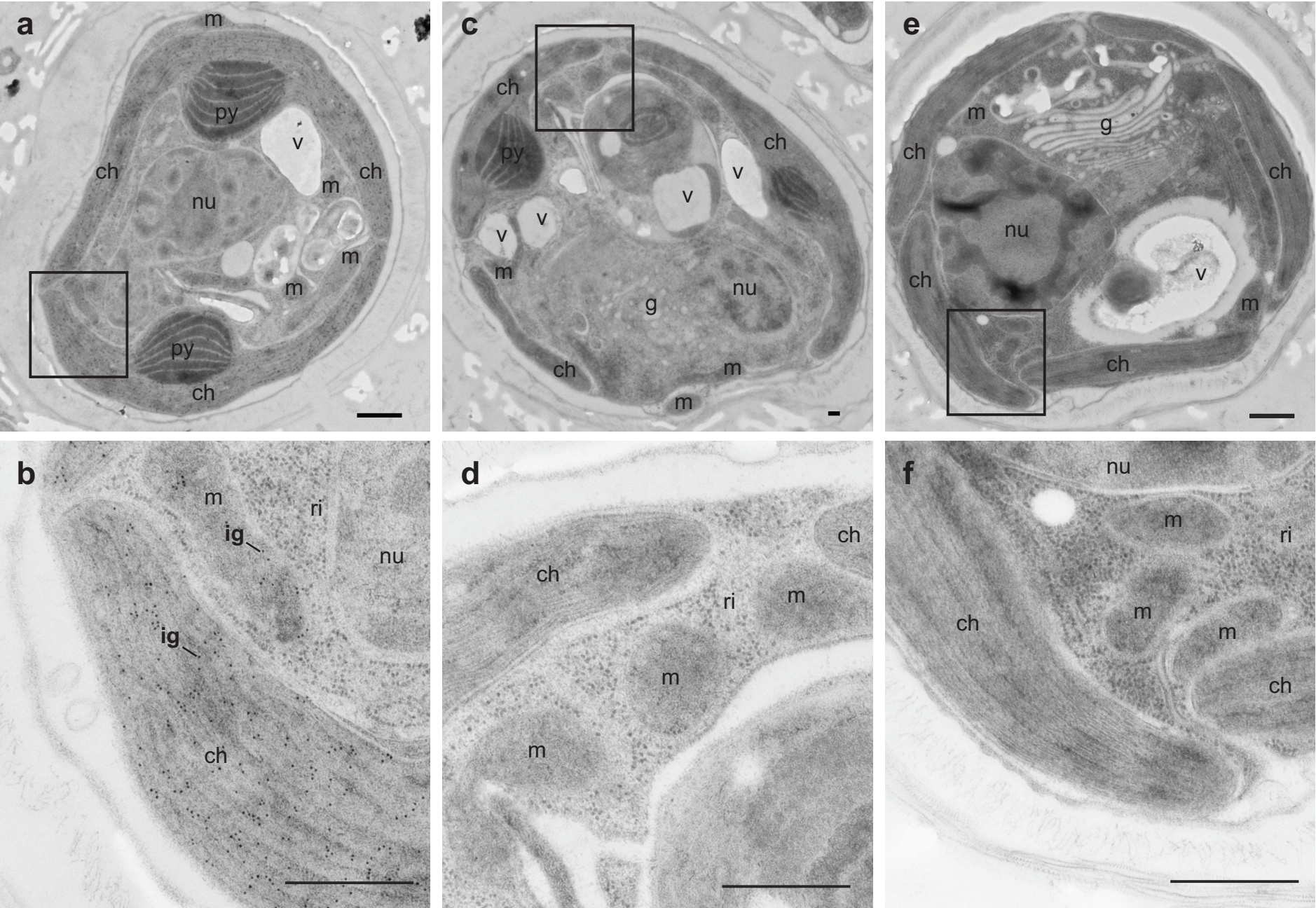
a, Predicted pathway for DMSP biosynthesis in bacteria (*Labrenzia*), macroalgae (*Ulva*, *Enteromorpha*), diatoms (*Thalassiosira*, *Melosira*), prymnesiophytes (*Emiliania*) and prasinophytes (*Tetraselmis*). Abbreviations: Met, methionine; MTOB, 4-methylthio-2-oxobutyrate; MTHB, 4-methylthio-2-hydroxybutyrate; DMSHB, 4-dimethylsulphonio-2-hydroxybutyrate. **b**, Maximum likelihood phylogenetic tree of DsyB/DSYB proteins. Species are colour-coded according to taxonomic class as shown in the key, with proteins shown to be functional marked with an asterisk. Bootstrap support for nodes is marked. Based on 145 protein sequences.

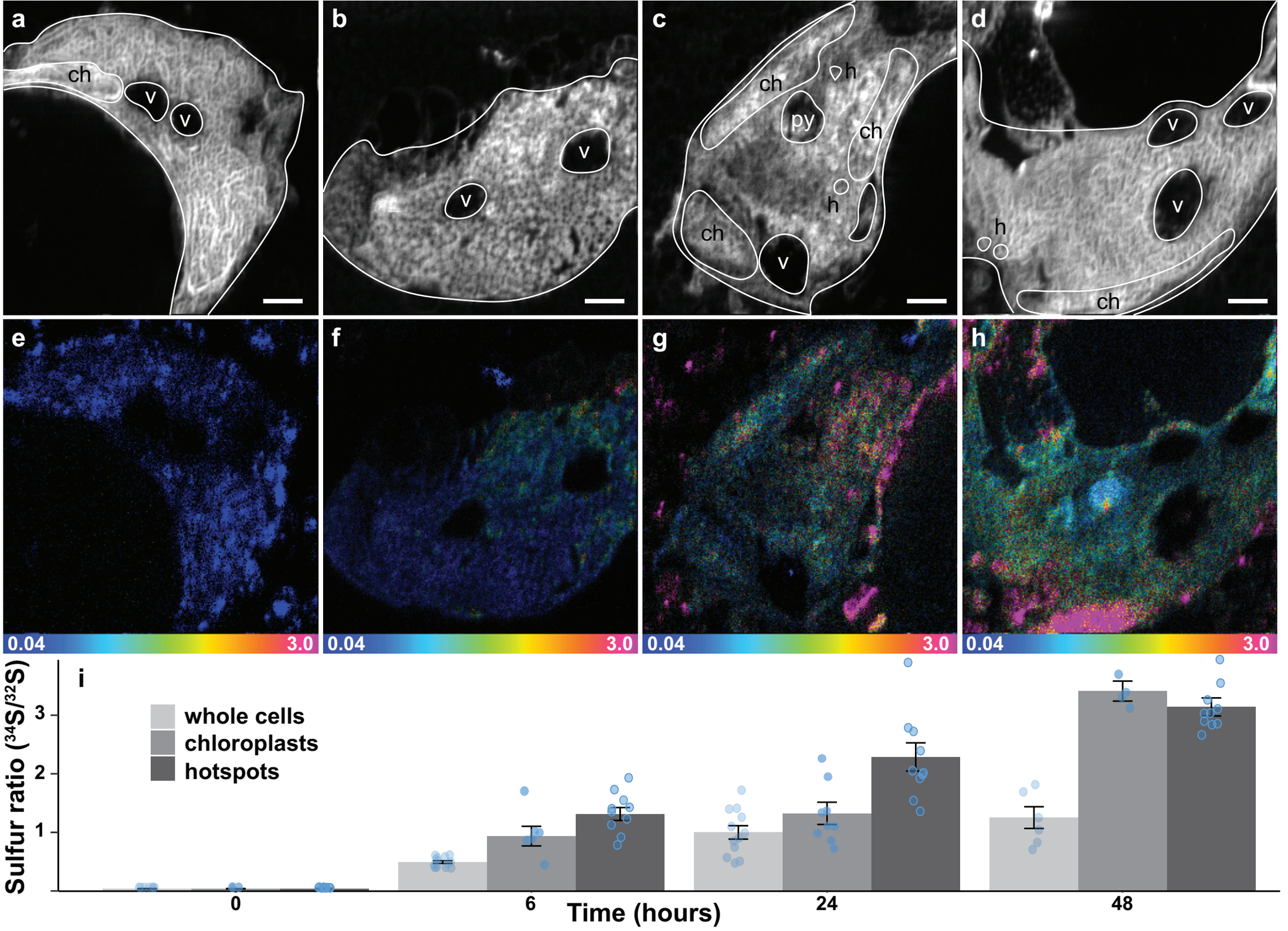
Figure 2. Immunogold localisation of DSYB in *Prymnesium parvum* CCAP946/6

Representative electron micrographs of *P. parvum* cells showing location of DSYB by immunogold labelling. **a**, **b**, Immunostaining of cell with DSYB antibody and secondary antibody with gold. **c**, **d**, Control immunostaining with pre-immune serum. **e**, **f**, Control immunostaining with only secondary antibody. Boxes in **a**, **c**, and **e**, correspond to area magnified in **b**, **d**, and **f** respectively. Scale bars are all 500 nm. Abbreviations: ch, chloroplast; g, golgi apparatus; ig, immunogold; m, mitochondrion; nu, nucleus; py, pyrenoid; ri, ribosome; v, vacuole. Experiments were repeated twice and two samples (n=2) were used for each experiment.

Figure 3. Sub-cellular distribution of ^{34}S in *Prymnesium parvum* CCAP946/6 following sulfur uptake for 48 h. a-d, Representative $^{12}C^{14}N/^{12}C_2$ mass images showing cellular structures of *P. parvum* cells. The cells were imaged straight after the start of the incubation (**a**), and after 6 h (**b**), 24 h (**c**) and 48 h (**d**). **e-h,** $^{34}S/^{32}S$ ratio of the same cells, shown as Hue Saturation Intensity (HSI) images where the colour scale indicates the value of the $^{34}S/^{32}S$ ratio, with natural abundance in blue, changing to pink with increasing ^{34}S levels. Each image was only acquired once. **i**, Isotope ratio of $^{34}S/^{32}S$ in different cellular regions (biological replicates, number of cells analysed: T0: whole cells n = 7, chloroplasts n = 3, hotspot n = 10; T6: whole cells n = 14, chloroplasts n = 6, hotspot n = 10; T24: whole cells n = 12, chloroplasts n = 9, hotspot n = 10; and T48: whole cells n = 6, chloroplasts n = 4, hotspot n = 10; error bars are shown for standard error). Abbreviations, ch: chloroplast; h: hotspot; py: pyrenoid; v: vacuole. Scale bars: 1 μm.


a
Met
TRANSAMINATION
MTOB
REDUCTION
NADPH
NADP
MTHB
METHYLATION
(DsyB/DSYB)
AdoMet
AdoHcy
DMSHB
DECARBOXYLATION
DMSP
b
Kryptoperidinium foliaceum CCMP1326
Glenodinium foliaceum CCAP11163
Kryptoperidinium foliaceum CCMP1326
Glenodinium foliaceum CCAP11163
Durinskia baltica CSIROCS38
Kryptoperidinium foliaceum CCMP1326
Scrippsiella hangoei SHTV5
Peridinium aciculiferum PAER
Crypthecodinium cohnii Seligo
Scrippsiella trochoidea CCMP3099
Azadinium spinosum 3D9
Azadinium spinosum 3D9
Prorocentrum minimum CCMP1329
Karenia brevis SP1
Scrippsiella trochoidea CCMP3099
Scrippsiella hangoei SHTV5
Crypthecodinium cohnii Seligo
Ceratium fusus PA161109
Ceratium fusus PA161109
Lingulodinium polyedrum CCMP1936
Alexandrium tamarense CCMP1771
Alexandrium monilatum CCMP3105
Gambierdiscus excentricus VGO790
Protopalythoa variabilis
Palythoa caribaeorum
Symbiodinium sp. C1
Porites australiensis
Anthopleura elegantissima
Anthopleura elegantissima
Millepora alcicornis
Porites astreoides
Acropora cervicornis
Symbiodinium sp. CCMP2430
Symbiodinium sp. Mp
Ansanella granifera AGSW10
Noctiluca scintillans SPMC136
Karlodinium micrum CCMP2283
Paragymnodinium shiwhaense PSSW0605
Paragymnodinium shiwhaense PSSW0605
Amphidinium carterae CCMP1314
Prymnesium parvum CCAP946/1B
Chrysochromulina polylepis CCMP1757
Chrysochromulina tobin CCMP291
Pleurochrysis carterae CCMP645
Pseudopedinella elastica CCMP716
Isochrysis galbana CCMP1323
Isochrysis sp. CCMP1324
Isochrysis galbana CCMP1323
Gephyrocapsa oceanica RCC1303
Isochrysis sp. CCMP1244
Emiliania huxleyi CCMP1516
Thalassiosira miniscula CCMP1093
Thalassiosira antarctica CCMP982
Skeletonema marinoi SkelA
Skeletonema dohrnii SkelB
Pseudonitzschia australis 1024910AB
Pseudonitzschia fraudulenta WWA7
Fragilariopsis cylindrus CCMP1102
Pelagococcus subviridis CCMP1429
Protopalythoa variabilis
Palythoa caribaeorum
Symbiodinium sp. C1
Millepora alcicornis
Porites astreoides
Symbiodinium microadriaticum CCMP2467
Symbiodinium sp. CCMP2430
Symbiodinium sp. Mp
Gambierdiscus polynesiensis CAWD212
Gambierdiscus excentricus VGO790
Alexandrium tamarense ATSP1B
Alexandrium monilatum CCMP3105
Scrippsiella hangoei SHTV5
Peridinium aciculiferum PAER2
Scrippsiella trochoidea CCMP3099
Durinskia baltica CSIROCS38
Karenia brevis SP1
Karenia brevis SP1
Karenia brevis SP1
Karenia brevis SP1
Eutreptiella gymnastica CCMP1594
Oceanicola sp. HL−35
Lutimaribacter saemankumensis DSM28010
Sediminimonas qiaohouensis DSM21189
Ruegeria sp. PBVC088
Mameliella alba UMTAT08
Antarctobacter heliothermus DSM11445
Oceanicola nanhaiensis DSM18065
Rhizobiales bacterium HL−109
Rhodobacteraceae bacterium PD−2
Ponticoccus litoralis DSM18986
Roseibacterium elongatum DSM19469
Nioella nitratireducens SSW136
Thalassobius gelatinovorus DSM5887
Donghicola xiamenensis DSM18339
Rhodovulum sp. ES010
Pseudooceanicola atlanticus 22II−s11g
Marivita sp. 22II−s11g
Oceanicola batsensis HTCC2597
Salipiger mucosus DSM16094
Roseivivax sediminis DSM26472
Roseivivax halodurans JCM10272
Pelagibaca bermudensis HTCC2601
Citreicella sp. 357
Salipiger nanhaiensis DSM27508
Roseivivax lentus DSM29430
Roseivivax halotolerans DSM15490
Yangia sp. CCB−MM3
Yangia pacifica DSM26894
Acuticoccus yangtzensis DSM28604
Amorphus coralli DSM19760
Albimonas donghaensis DSM17890
Albimonas pacifica CGMCC 111030
Nisaea denitrificans DSM18348
Alphaproteobacterium BAL199
Sagittula stellata E−37
Tropicibacter naphthalenivorans CECT7648
Ponticoccus sp. SJ5A−1
Maritimibacter alkaliphilus HTCC2654
Stappia stellulata DSM5886
Stappia sp. ES058
Stappia indica SBBC49
Rhodobacteraceae bacterium bin09
Labrenzia aggregata IAM12614
Labrenzia aggregata CECT4801
Labrenzia sp. OB1
Labrenzia alexandrii CECT5112
Roseibium hamelinense ATCCBAA−252
Labrenzia sp. DG1229
Labrenzia alba CECT5094
Labrenzia alba CECT7551
Hoeflea sp. IMCC20628
Roseovarius sp. HI0049
Roseovarius indicus EhC03
Phaeobacter gallaeciensis JL2886
Litorimicrobium taeanense DSM22007
Defluviimonas indica DSM24802
Pseudovibrio axinellae DSM 24994
Rhizobiales bacterium NRL2
Thalassobaculum salexigens DSM19539
Thalassobaculum litoreum DSM18839
Roseospirillum parvum 930I
Caenispirillum salinarum AK4
Rhodospirillales bacterium URHD0017
Betaproteobacteria sp.
Streptomyces varsoviensis DSM40346
Eukaryotes
(DSYB)
Prokaryotes
(DsyB)
Taxon
Actinobacteria
Betaproteobacteria
Alphaproteobacteria
Anthozoa
Hydrozoa
Euglenophyceae
Prymnesiophyceae
Dinophyceae
Bacillariophyceae
Dictyochophyceae


a
m
py
v
ch
ch
nu
m
m
m
py
ch
c
ch
py
ch
v
v
v
v
m
g
nu
ch
m
m
e
m
g
ch
ch
nu
ch
v
m
ch
b
m
ri
ig
nu
ig
ch
d
ch
ch
ri
m
m
m
f
nu
m
ri
ch
m
m
ch


a
ch
v
v
b
v
v
c
ch
h
py
ch
h
ch
v
d
v
v
h
v
ch
e
0.04
3.0
f
0.04
3.0
g
0.04
3.0
h
0.04
3.0
i
whole cells
chloroplasts
hotspots
Sulfur ratio ($^{34}S/^{32}S$)
1
2
3
0
6
24
48
Time (hours)