

## ORIGINAL PAPER

# Metabolic Consequences of Cobalamin Scarcity in the Diatom *Thalassiosira pseudonana* as Revealed Through Metabolomics



Katherine R. Heal, Natalie A. Kellogg, Laura T. Carlson, Regina M. Lionheart, and Anitra E. Ingalls<sup>1</sup>

School of Oceanography, University of Washington, Seattle, WA 98195, USA

Submitted November 6, 2018; Accepted May 19, 2019  
Monitoring Editor: Chris Bowler

Diatoms perform an estimated 20% of global photosynthesis, form the base of the marine food web, and sequester carbon into the deep ocean through the biological pump. In some areas of the ocean, diatom growth is limited by the micronutrient cobalamin (vitamin B<sub>12</sub>), yet the biochemical ramifications of cobalamin limitation are not well understood. In a laboratory setting, we grew the diatom *Thalassiosira pseudonana* under replete and low cobalamin conditions to elucidate changes in metabolite pools. Using metabolomics, we show that the diatom experienced a metabolic cascade under cobalamin limitation that affected the central methionine cycle, transsulfuration pathway, and composition of osmolyte pools. In *T. pseudonana*, 5'-methylthioadenosine decreased under low cobalamin conditions, suggesting a disruption in the diatom's polyamine biosynthesis. Furthermore, two acylcarnitines accumulated under low cobalamin, suggesting the limited use of an adenosylcobalamin-dependent enzyme, methylmalonyl CoA mutase. Overall, these changes in metabolite pools yield insight into the metabolic consequences of cobalamin limitation in diatoms and suggest that cobalamin availability may have consequences for microbial interactions that are based on metabolite production by phytoplankton.

© 2019 Elsevier GmbH. All rights reserved.

**Key words:** Cobalamin; diatoms; metabolomics.

## Introduction

Cobalamin (vitamin B<sub>12</sub>) is an organic micronutrient that can shape and control primary productivity in marine ecosystems. Over half of surveyed species of eukaryotic algae require an exogenous source of the compound (Croft et al. 2006), which is produced through a biosynthetic pathway of over two

dozen steps by a select cohort of bacteria and archaea (Warren et al. 2002). In the ocean, Thaumarchaeota and certain lineages of heterotrophic bacteria are thought to be the main producers of cobalamin (Doxey et al. 2014; Heal et al. 2017), while the majority of cyanobacteria produce pseudocobalamin, a closely related compound that is less bioavailable to eukaryotic phytoplankton (Heal et al. 2017; Helliwell et al. 2016). Though few direct measurements of oceanic cobalamin exist, perturbation experiments of natural populations suggest

<sup>1</sup>Corresponding author.  
e-mail [aingalls@uw.edu](mailto:aingalls@uw.edu) (A.E. Ingalls).

that the scarcity of cobalamin in the marine environment can control which phytoplankton species thrive (Bertrand et al. 2007, 2015; Browning et al. 2017; Sañudo-Wilhelmy et al. 2006).

Cobalamin is used in a wide variety of enzymes that can shape dependencies in marine microbes (Bertrand et al. 2015; Romine et al. 2017). In eukaryotic algae, the dependency on exogenous cobalamin, or cobalamin auxotrophy, appears to arise from the compound's role as a cofactor in the methylcobalamin (Me-cobalamin) dependent methionine synthase (MetH) (Bertrand and Allen 2012; Helliwell et al. 2011), a key enzyme in the synthesis and regeneration of methionine via a methyl transfer from methyltetrahydrofolate (Banerjee et al. 2003). Some algae also have a cobalamin-independent isoform of the enzyme, MetE (Bertrand and Allen 2012; Helliwell et al. 2011); MetE performs the same function as MetH, but much less efficiently and without the use of cobalamin as its catalytic center (Bertrand et al. 2013; Gonzalez et al. 1992). The current understanding of cobalamin dependence in algae hinges on the presence of MetE — that is, if algae have only MetH, they are dependent on exogenous cobalamin, but if algae have both MetE and MetH they can switch between enzymes and can grow without cobalamin, even in cases where algae harbor other cobalamin-dependent enzymes (Bertrand and Allen 2012; Helliwell et al. 2011). Both algae that require exogenous cobalamin and those that can grow without cobalamin can grow over a wide range of cobalamin concentrations (Croft et al. 2005; Droop 1968; Provasoli and Carlucci 1974; Tang et al. 2010), and previous studies have elucidated the transcriptional and translational response that model diatoms have to cobalamin scarcity (Bertrand and Allen 2012; Bertrand et al. 2012, 2013; Helliwell et al. 2011). These analyses have revealed strategies diatoms employ to reduce cellular demand for cobalamin, increase cobalamin acquisition from the environment, and manage reduced functionality of cobalamin-dependent enzymes (Bertrand and Allen 2012; Bertrand et al. 2012, 2013).

One study revealed that the important methyl donor, S-adenosyl methionine (SAM), is depleted in diatom cells grown under cobalamin limitation (Heal et al. 2017), but no study has attempted to look at the whole-cell metabolic consequence of cobalamin depletion in algae to date. Specific hypotheses deduced from transcriptional changes include a decrease in the osmoregulators dimethylsulfoniopropionate (DMSP) and glycine betaine (GBT), an imbalance in the

methionine cycle, and changes in polyamine biosynthesis (Bertrand and Allen 2012), which all have yet to be validated on the metabolite level. Beyond these hypotheses, all sequenced diatoms code for an adenosylcobalamin-dependent enzyme (methylmalonyl-CoA mutase, MCM) (Bertrand and Allen 2012; Helliwell et al. 2011), but it is unclear if cobalamin availability affects diatom metabolism through decreased efficiency of this enzyme. Furthermore, it is not clear if or how algae cope with the accumulation of toxic compounds like S-adenosyl homocysteine (SAH) and homocysteine as seen in a cobalamin-stressed green alga (Croft et al. 2005). Transcriptomic and proteomic analyses have provided valuable insight into cobalamin limitation in diatoms, though there are likely additional metabolic consequences that are regulated post-translationally which can be unmasked by direct metabolite measurements.

In this study, we used targeted and untargeted metabolomics to explore the metabolic consequences that cobalamin scarcity has on the centric diatom *Thalassiosira pseudonana*. This diatom has an absolute requirement for cobalamin (Croft et al. 2005; Provasoli and Carlucci 1974), does not code for the cobalamin-independent MetE (Bertrand and Allen 2012; Helliwell et al. 2011), and has been well studied on the transcript and protein level (Bertrand and Allen 2012). In our targeted approach, we obtain relative abundances (among samples) of known primary metabolites like amino acids, osmolytes, and methionine cycle intermediates (Boysen et al. 2018). In untargeted metabolomics, we hope to gain a holistic understanding of both known and unknown metabolite pools and how they change under different conditions. Using a batch culture approach, we grew *T. pseudonana* under replete cobalamin and low cobalamin conditions and in both low and saturating light to differentiate metabolite pools that changed due to growth rate changes and those that changed due to cobalamin limitation. These experiments provide us the foundation we need to understand how these important primary producers experience and cope with cobalamin limitation in laboratory and natural settings.

## Results

### Growth Rates

We grew *T. pseudonana* under both saturating ( $120 \mu\text{mol photons m}^{-2} \text{sec}^{-1}$ ) and low light ( $50 \mu\text{mol photons m}^{-2} \text{sec}^{-1}$ ) conditions with low

**Table 1.** Experimental conditions and growth rates ( $\mu$ ) for *T. pseudonana* under the four experimental conditions.

Treatment	Light level ( $\mu\text{mol photons m}^{-2} \text{sec}^{-1}$ )	Cobalamin (pM)	$\mu \pm \text{sd}$ ( $\text{day}^{-1}$ )	<i>n</i>
Saturating light, replete cobalamin	120	200	$0.66 \pm 0.01$	9
Saturating light, low cobalamin	120	1	$0.41 \pm 0.09$	9
Low light, replete cobalamin	50	200	$0.58 \pm 0.04$	9
Low light, low cobalamin	50	1	$0.30 \pm 0.07$	9

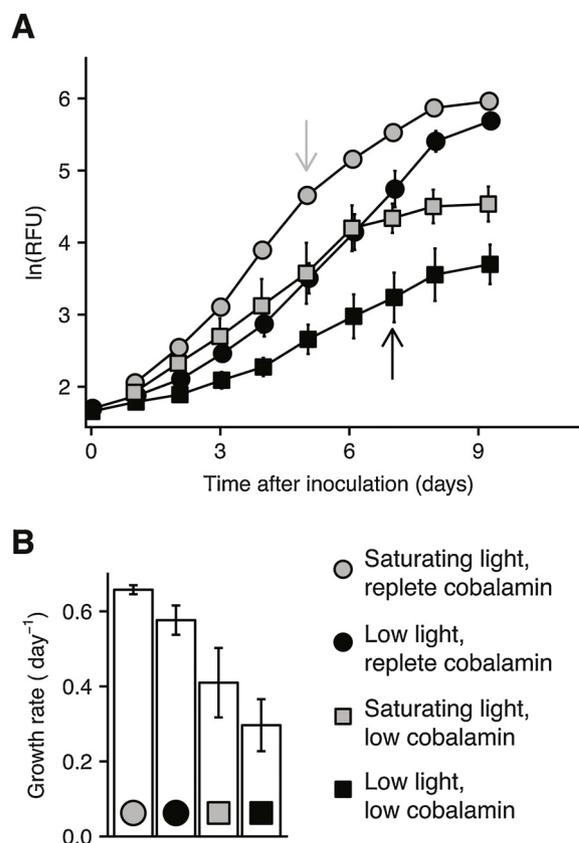
(1 pM) or replete (200 pM) cobalamin (control). Concentrations for low and replete cobalamin conditions were based on previous work (Bertrand et al. 2012; Heal et al. 2017). Cobalamin was provided as hydroxocobalamin (OH-cobalamin) in all experiments. Compared to the cobalamin replete control, the growth rate of *T. pseudonana* was lowered by 38% and 48% under low cobalamin when grown in saturating and low light, respectively ( $p < 0.001$ , Table 1 and Fig. 1). The low light conditions also lowered the growth rate 12% and 28% ( $p = 0.01$ ) under replete and low cobalamin, respectively (Table 1 and Fig. 1).

### Targeted Metabolomics

We obtained relative concentrations for compounds where the amount of a compound can be compared between samples, but the absolute concentration is not known. We detected 72 metabolites in *T. pseudonana* with our targeted approach (Supplementary Material Table S1). To assess if our metabolomics approach could discern between the tested experimental conditions, we performed a two-dimensional non-metric multidimensional scaling (NMDS) analysis. Using the relative concentrations that resulted from our targeted analysis, an NMDS ordination resulted in a low stress value ( $< 0.1$ , Fig. 2). The non-metric approach was successfully able to discern low light from saturating light treatments as well as low cobalamin from replete cobalamin treatments (Table 2). Compounds that significantly contributed to the NMDS plot are reported in Supplementary Material Table S2. Of the targeted metabolites that we observed, 36% showed a significant difference in abundance between cobalamin treatments under both light regimes in *T. pseudonana* ( $p < 0.05$ , Table 3 and Fig. 3).

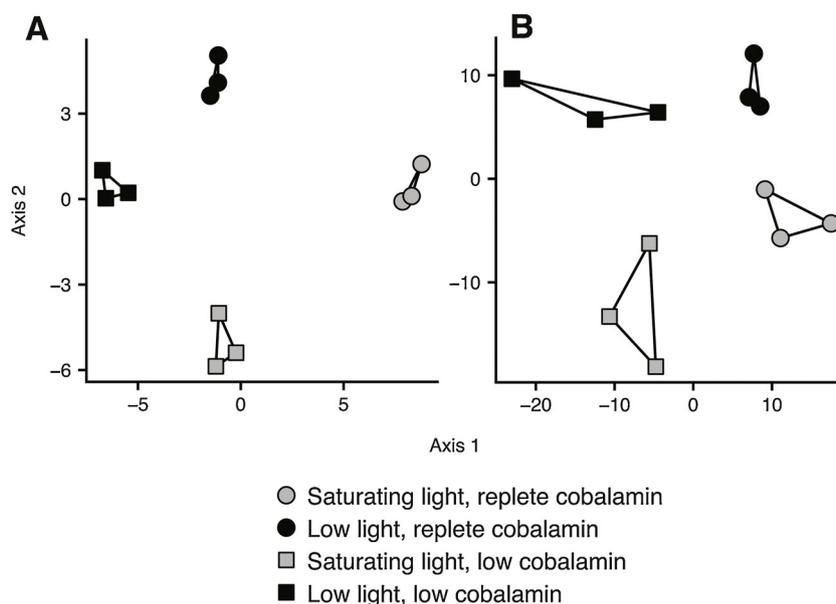
### Untargeted Metabolomics

A targeted approach to metabolomics is inherently biased as the only data acquired are on a set of prescribed compounds chosen by the ana-



**Figure 1.** Growth curves of *T. pseudonana* (A) under the four experimental conditions. Cells grown under saturating light were harvested at grey arrow, cells grown under low light were harvested at black arrow. RFU = relative fluorescence units, error bars are standard deviation and are often smaller than the markers ( $n = 9$  for all pre-harvest time points,  $n = 3$  for post-harvest time points). (B) Average growth rate with standard deviation for *T. pseudonana*,  $n = 9$ . Growth rates were significantly different between all treatments ( $t$ -test,  $p < 0.05$ ).

lyst. On the contrary, an untargeted approach is only biased towards compounds that are detectable using the analytical method chosen. Untargeted metabolomics yields areas of peaks that corre-



**Figure 2.** Results from NMDS analysis in targeted and untargeted data show clear differentiation between treatments. **(A)** Locations of the samples in NMDS space for 75 targeted compounds (stress = 0.026). **(B)** The same analysis on the 500 largest mass features from the untargeted analysis (stress = 0.083),  $p$  values <0.05 for both. Markers are biological replicates of each of the treatments.

**Table 2.** Results from ANOSIM analysis. ANOSIM statistic and  $p$  values are given for differentiating metabolomes depending on light or cobalamin in the targeted and untargeted analyses, with  $p$  values.

Variable	ANOSIM Statistic	$p$	Analysis
Light	0.602	0.003	Targeted
Cobalamin	0.507	0.006	Targeted
Light	0.380	0.008	Untargeted
Cobalamin	0.685	0.004	Untargeted

**Table 3.** Summarized results of targeted and untargeted metabolomics analyses. Percent of the mass features or compounds that showed a difference ( $p < 0.05$  between the cobalamin treatments).

Analysis	Quality Mass Features	% of Mass Features with Cobalamin Signal
Untargeted	1674	2.81
Targeted	71	36.51

spond to single mass and retention time pairs, these peaks are referred to as mass features. While each compound may correspond to several mass features (due to natural isotopes or adduct formation), each mass feature generally corresponds to

only one compound. Similar to our targeted analyses, several mass features displayed a differential response under cobalamin limitation and our multivariate approach yielded similar results to the targeted analyses (Fig. 2; Table 2).

A major bottleneck in metabolomics research is the identification of detected mass features. Challenges remain because there are outstanding gaps in metabolic pathways, existing databases have insufficient coverage in fragmentation spectra ( $MS^2$ ), and absolute identification relies on authentic standards which can be prohibitively expensive or commercially unavailable (Sumner et al. 2007). We focused our identification efforts on quality mass features that showed a univariate response to cobalamin limitation, and we used the ranking system outlined in Sumner et al. (2007). This ranking system matches to a compound's exact mass to charge ( $m/z$ ), fragmentation, and retention time (and combinations thereof, when possible). As previously noted (Johnson et al. 2016), these stringent guidelines yield a small subset of identifiable metabolites from an untargeted metabolomics analysis, but we can expect to see improvements in this yield in the future as metabolomics analyses become more common and the databases improve.

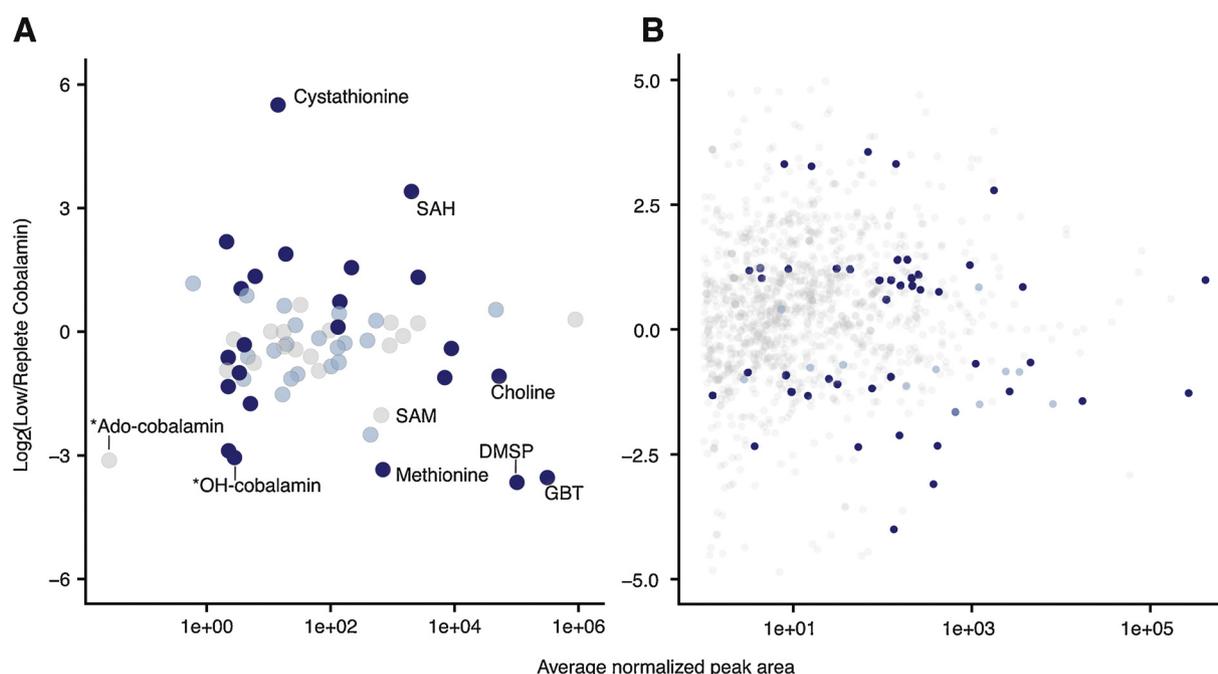
Using automated approaches, we were able to identify or putatively identify 1–2% of the quality mass features in *T. pseudonana* (Supplementary

**Table 4.** Mass features with significant differences between cobalamin treatments that were detected in our untargeted analysis ( $p < 0.05$ ). Identification annotations with compound names and ions (in brackets) observed. Identification confidence based on the rating from Sumner et al. (2007), where confidence level 1 is *unequivocal* and verified with standards, level 2 is *putative* and supported by a match based on MS<sup>2</sup> and  $m/z$ , level 3 should be considered a *possible* match based on  $m/z$  in the KEGG database, and level 4 yielded no helpful matches. Log<sub>2</sub> (fold change) between cobalamin treatments under saturating light (SL), or low light (LL). Many adducts of mass features in this list also showed significant differences between treatments which are not repeated here, see Supplemental Table 4 for full results including those adducts with annotations. RT = retention time. \*\*The isomers isobutylcarnitine and butylcarnitine are indistinguishable in our analysis, so this mass feature may be either compound or a combined signal from both.

$m/z$	RT (min)	Column, polarity	Log <sub>2</sub> (fold change)		Best ID [ion]	Confidence	Evidence
			SL	LL			
385.1288	10.55	HILIC, +	3.364	3.248	S-adenosyl homocysteine [M+H]	1	$m/z$ , RT, standard
218.1386	6.97	HILIC, +	2.021	3.808	Propionyl-L-carnitine [M+H]	1	$m/z$ , MS <sup>2</sup> , RT, standard
175.1189	18.42	HILIC, +	1.796	0.779	Arginine [M+H]	1	$m/z$ , RT, standard
205.0973	1.95	RP, +	1.377	1.411	Tryptophan [M+H]	1	$m/z$ , RT, standard
584.3941	6.23	HILIC, +	1.318	0.671		4	
144.102	7.3	HILIC, +	1.215	1.506	Proline betaine [M+H]	1	$m/z$ , RT, standard
147.0996	2.38	RP, +	1.104	1.320		4	
149.1093	2.37	RP, +	1.081	1.317		4	
166.1387	2.37	RP, +	1.078	1.280		4	
166.1359	2.37	RP, +	1.075	1.334		4	
149.1121	2.37	RP, +	1.030	1.416		4	
182.0813	0.99	RP, +	0.990	1.228	Tyrosine [M+H]	1	$m/z$ , RT, standard
133.0608	11.37	HILIC, +	0.904	0.695	Asparagine [M+H]	1	$m/z$ , RT, standard
278.1611	6.24	HILIC, +	0.897	0.786		4	
562.4104	9.46	HILIC, +	0.885	0.849		4	
138.0548	6.22	HILIC, +	0.871	1.092	Homarine [M+H]	1	$m/z$ , RT, standard
188.0706	8.28	HILIC, +	0.819	0.972	Tryptophan [M-OH+H]	1	$m/z$ , RT, standard
370.1628	1.07	RP, +	0.811	1.283		4	
366.0511	9.39	HILIC, +	0.809	1.225		4	

Table 4 (Continued)

<i>m/z</i>	RT (min)	Column, polarity	Log <sub>2</sub> (fold change)		Best ID [ion]	Confidence	Evidence
			SL	LL			
116.3942	9.46	HILIC, +	0.742	0.419		4	
232.1544	1.36	RP, +	0.643	1.381	(iso)Butyryl carnitine [M+H]**	1	<i>m/z</i> , MS <sup>2</sup> , RT, standard
247.0577	10.24	HILIC, +	0.455	1.189		4	
148.0603	11.94	HILIC, +	-0.897	-0.367	Glutamic acid [M+H]	1	<i>m/z</i> , RT, standard
490.7438	2.26	RP, +	-1.005	-0.743		4	
339.2279	0.92	RP, +	-1.102	-0.705		4	
613.1596	0.91	RP, +	-1.188	-1.331	Glutathione disulfide [M+H]	1	<i>m/z</i> , MS <sup>2</sup> , RT, standard
364.0651	13.83	HILIC, +	-1.436	-0.680	GMP [M+H]	1	<i>m/z</i> , RT, standard
120.0657	11.3	HILIC, +	-1.452	-0.851	Homoserine [M+H]	1	<i>m/z</i> , RT, standard
187.1229	6.26	HILIC, +	-1.469	-2.231	Calligonine [M+H]	3	<i>m/z</i>
155.0012	8.14	HILIC, -	-1.601	-1.251		4	
258.11	11.47	HILIC, +	-1.679	-0.763	Glycerophosphocholine [M+H]	1	<i>m/z</i> , MS <sup>2</sup> , RT, standard
298.0969	2.13	RP, +	-1.817	-0.802	MTA [M+H]	1	<i>m/z</i> , MS <sup>2</sup> , RT, standard
309.0438	9.85	HILIC, -	-2.073	-2.178		4	
275.1026	0.84	RP, +	-2.470	-2.183		4	
178.0715	7.91	HILIC, -	-2.623	-1.922		4	
278.5725	0.9	RP, +	-2.697	-1.884		4	
116.0707	7.9	HILIC, -	-4.543	-2.004	Glycine betaine [M-H]	1	<i>m/z</i> , RT, standard



**Figure 3.** Results from the analyses for targeted (A) and untargeted (B) metabolomics analysis. In (A), every compound detected with the targeted analysis is depicted as a dot; light blue compounds are significantly different between cobalamin treatments in either saturating or low light conditions, blue compounds are significantly different under both light conditions or when light treatments are pooled ( $p < 0.05$ ). Y-axis is  $\log_2$ (average peak size of cobalamin limited treatments / average peak size of replete cobalamin treatments). X-axis is average peak area of compound after adjustment and normalization; note that x-axis is log scaled. Specific compounds discussed in text are labeled. In (B), colors and location of dots are same as in panel (A), but each dot represents a mass feature. \*These compounds were only detected in replete cobalamin conditions, therefore the fold change and univariate statistics were calculated against an estimate of the analytical blank and represent a lower limit fold change.

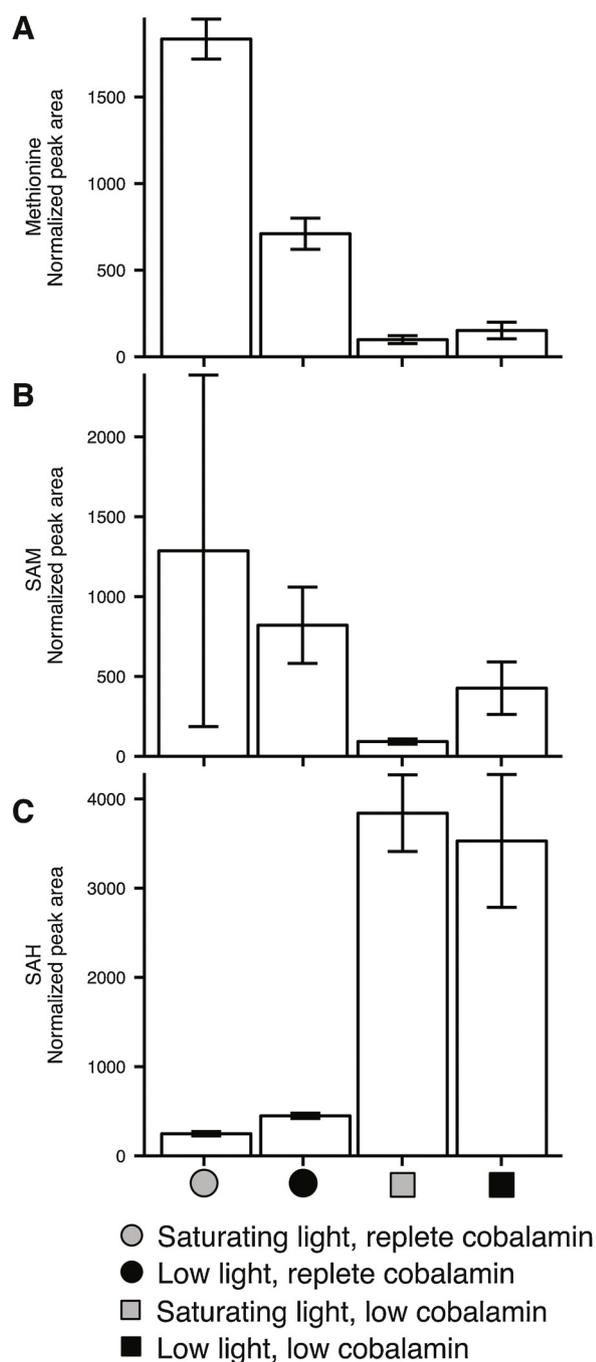
Material Table S3). With a more manual approach (described in methods section), we identified a larger percentage (47%) of the mass features that showed a response to cobalamin limitation (Table 4). Many of the mass features we identified matched those in our targeted analysis, and the change in pool size of these compounds under cobalamin limitation matched the signal from the targeted results. Compounds explicitly discussed in this manuscript were putatively identified with  $m/z$  and  $\text{MS}^2$  and validated with a newly purchased standard or matched to a standard in our existing targeted analysis (Boysen et al. 2018), and are therefore unequivocally identified.

### Changes in Specific Metabolite Pools

**Cobalamins.** We provided cobalamin to the diatom cultures in the form of OH-cobalamin as it is the major form of cobalamin we have observed dis-

solved in seawater (Heal et al. 2014). This oxidized form of cobalamin must be enzymatically converted to either adenosylcobalamin (Ado-cobalamin) or methylcobalamin (Me-cobalamin) before it can be used as a cofactor. We observed that under replete cobalamin, intracellular OH-cobalamin and Ado-cobalamin concentrations were higher for *T. pseudonana* when provided with more cobalamin (Fig. 3; Supplementary Material Table S1).

**Methionine Cycle.** Many of the compounds with significant differences between cobalamin treatments in *T. pseudonana* can be tied directly to changes in the methionine cycle. This phenomenon has been previously hypothesized because cobalamin is used as the catalytic center in the methionine synthase (MetH) enzyme in these algae (Bertrand and Allen 2012; Bertrand et al. 2012; Croft et al. 2006). In the low cobalamin treatments, we see that *T. pseudonana* experienced a major change in metabolite pools in the methion-



**Figure 4.** Normalized peak areas of methionine (A), SAM (B), and SAH (C) in *T. pseudonana*, summarized by each of the growing conditions. Error bars are standard deviation with  $n = 3$  of biological replicates. All these data were obtained in our targeted analysis.

ine cycle compared to cobalamin replete controls (Fig. 4). Methionine and SAM decreased while SAH increased in cobalamin-depleted *T. pseudonana* cells (Figs 3, 4, and 5). In the low cobalamin

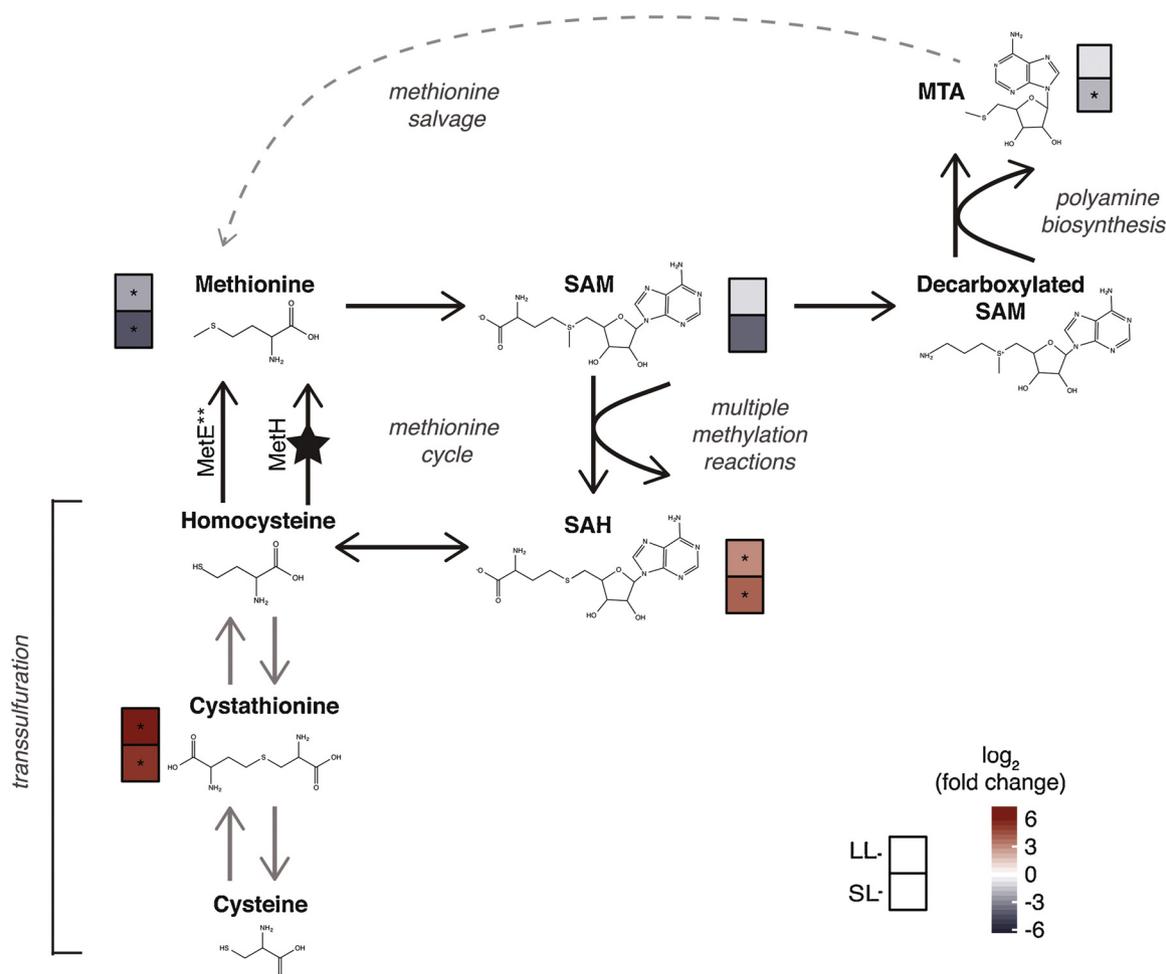
treatments, cystathionine was on average 42 times more abundant compared to cobalamin-replete cells ( $p=0.02$ , Figs 3, 5, and 6). Although previous studies have suggested that the folate cycle in algae may also be affected by cobalamin availability due to its link to the methionine cycle at MetH (Bertrand and Allen 2012; Bertrand et al. 2012; Croft et al. 2006), we were not able to detect any folates in this study to support or refute this hypothesis.

**Compatible solutes.** Compatible solutes, or osmolytes, are compounds in a cell's cytosol that maintain osmotic pressure (Brown 1976) and are known to play other roles in cells (Yancey 2005). Many compounds that are commonly described as compatible solutes changed under low cobalamin conditions in *T. pseudonana*. We observed on average 12 times less dimethylsulfoniopropionate (DMSP) under low cobalamin ( $p=0.02$  under saturating light,  $p=0.04$  under low light, Figs 3 and 7). Glycine betaine (GBT) also decreased in low cobalamin treatments compared to cobalamin replete controls (on average 11 times less GBT under cobalamin limitation,  $p=0.02$  under saturating light,  $p=0.007$  under low light, Figs 3 and 8). Compounds related to GBT also decreased under cobalamin limitation (Fig. 8), while compounds that may act as replacements for compatible solutes increased under cobalamin limitation in *T. pseudonana* (Fig. 9).

**5'-methylthioadenosine.** In *T. pseudonana*, 5'-methylthioadenosine (MTA) was less abundant in cells grown under low cobalamin conditions (Fig. 10,  $p=0.02$ ). We detected MTA in our untargeted analysis and confirmed its identity with commercial standards.

**Acylcarnitines.** Our untargeted analysis revealed that two acylcarnitines, propionylcarnitine and butyrylcarnitine, accumulated in cells that were grown with low cobalamin (Fig. 11). This increase was statistically significant for both molecules ( $p < 0.05$ ).

**Broad patterns in metabolite pools.** Several other compounds showed a univariate response to cobalamin in *T. pseudonana* (Fig. 3; Table 4). The concentrations of many amino acids were affected by low cobalamin conditions: glutamine and glutamic acid concentrations were roughly halved while tyrosine, asparagine, and tryptophan increased ( $p < 0.05$  for all, Fig. 3; Table 4, and Supplementary Material Table S1). Interestingly, both purine nucleosides showed changes under cobalamin limitation, though in opposite directions (guanosine accumulated while adenosine decreased).



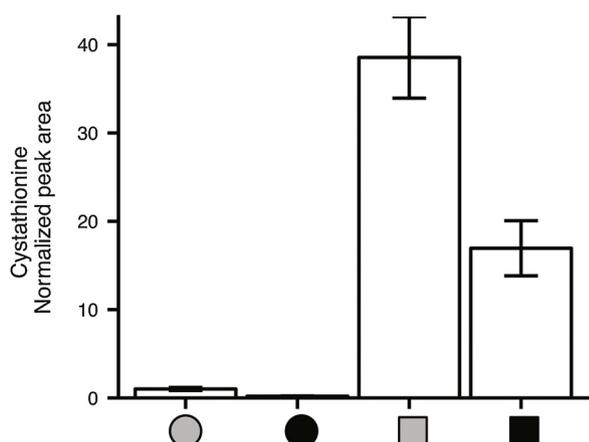
**Figure 5.** Changes in the methionine cycle and related pathways as observed in metabolite concentrations. Log<sub>2</sub>(fold change) in metabolite concentrations are shown between cobalamin treatments in low light (LL) and saturating light (SL) conditions and *T. pseudonana* as noted in the key. Significant differences in concentrations between cobalamin treatments in individual light treatments are designated with an asterisk ( $n = 3$ ,  $p < 0.05$ ). Reactions with corresponding enzymes that are well described in diatoms are shown in black, with less certain pathways in grey; note that the methionine salvage pathway consists of multiple enzymes (dashed arrow).

## Discussion

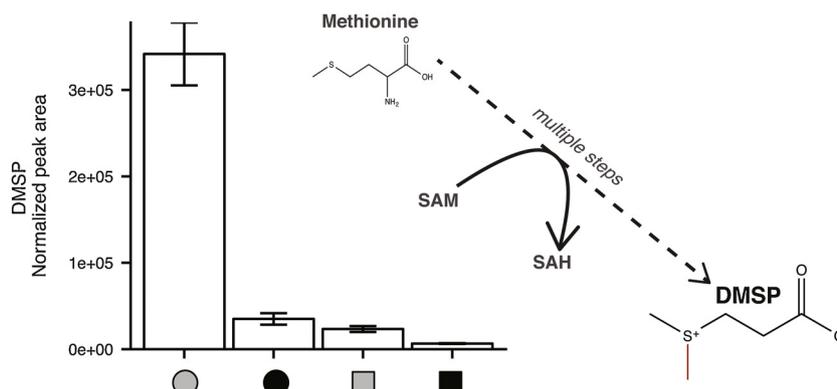
### Intracellular Cobalamins

The Ado-cobalamin form of cobalamin has been previously observed in *T. pseudonana* cultures (Heal et al. 2017), but has not been well explained. Here we observed Ado-cobalamin in *T. pseudonana* that were supplied only OH-cobalamin. This implies that the diatoms enzymatically added the adenosyl  $\alpha$  ligand; the gene encoding the enzyme for the adenylation of cobalamin (cob(I)yrinic acid a,c-diamide adenosyltransferase) has been identified in all sequenced diatom genomes (Helliwell et al. 2011). We only observed OH- and

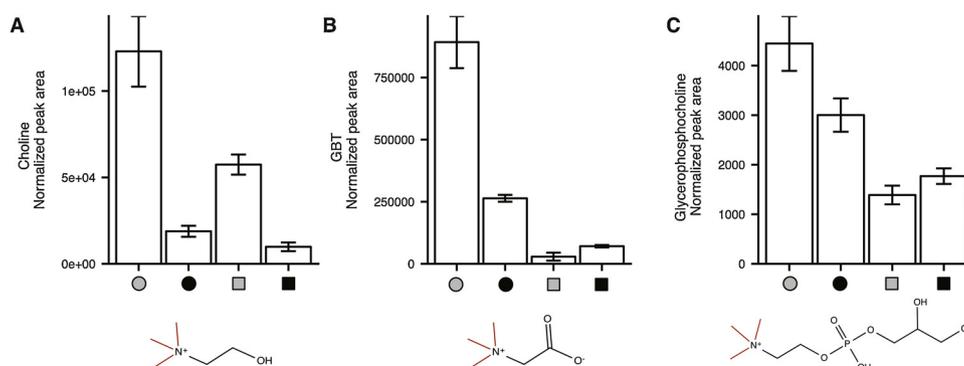
Ado-cobalamin under replete cobalamin (Fig. 3; Supplementary Material Table S1). We did not observe Me-cobalamin in any of our samples. This is likely due to limits in our analytical detection with the method used here; in previous work from our group we were able to detect Me-cobalamin in *T. pseudonana* using a more sensitive analysis (Heal et al. 2017). Since Ado-cobalamin is not thought to be an intermediate to Me-cobalamin use and regeneration (Banerjee et al. 2003), our data suggest that *T. pseudonana* use Ado-cobalamin for another function. The most likely candidate for this is as a cofactor for MCM, which has been identified in all sequenced diatom genomes (Bertrand and Allen 2012; Helliwell et al. 2011).



**Figure 6.** Normalized peak areas of cystathionine in *T. pseudonana* summarized by each of the growing conditions (labeled the same as Fig. 4). Error bars are standard deviation with  $n=3$  of biological replicates. We detected cystathionine in our targeted analysis.



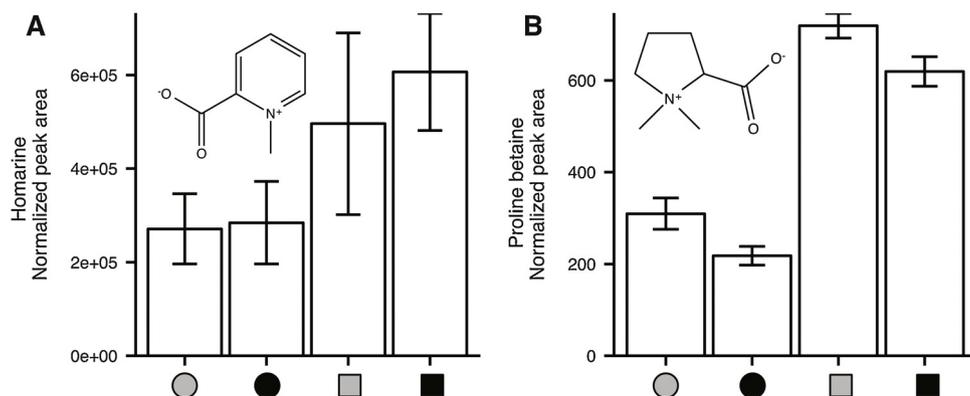
**Figure 7.** Normalized peak areas of DMSP summarized by each of the growing conditions (labeled the same as Fig. 4). Inset shows the structure of DMSP and its connection to the methionine cycle, where it is directly derived from methionine and is methylated via SAM (red methyl group). Error bars are standard deviation with  $n=3$  of biological replicates. We detected DMSP in our targeted analysis.



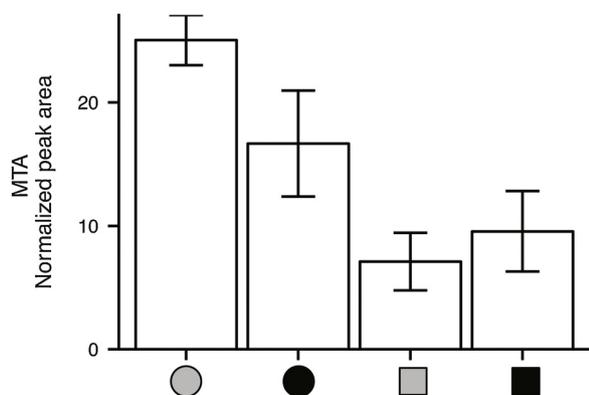
**Figure 8.** Normalized peak areas of choline (A), GBT (B), and glycerophosphocholine (C) summarized by each of the growing conditions (labeled the same as Fig. 4), with structures. In each structure, the red methyl groups are likely directly from SAM. Error bars are standard deviation with  $n=3$  of biological replicates. Data shown for GBT and DMSP are from our targeted analysis; glycerophosphocholine was detected in our untargeted analysis.

## Disruption of the Methionine Cycle

The methionine cycle showed obvious changes under low cobalamin conditions in *T. pseudonana* (Figs 3, 4, and 5). SAM likely only follows ATP in the variety and number of cellular reactions in which it serves as a cofactor (Lu 2000). SAM is a methylating agent for a wide variety of reactions in both primary and secondary metabolism, including DNA methylation (affecting gene expression) and osmoprotectant biosynthesis (Roje 2006). Once SAM has donated its methyl group during a methylation reaction, the resulting compound (SAH) must be re-methylated via homocysteine and then methionine in the central methionine cycle (Fig. 5). SAM depletion has been hypothesized to be a major consequence of cobalamin limitation in diatoms by transcriptomic and proteomic inference (Bertrand and Allen 2012; Bertrand et al. 2012), and a



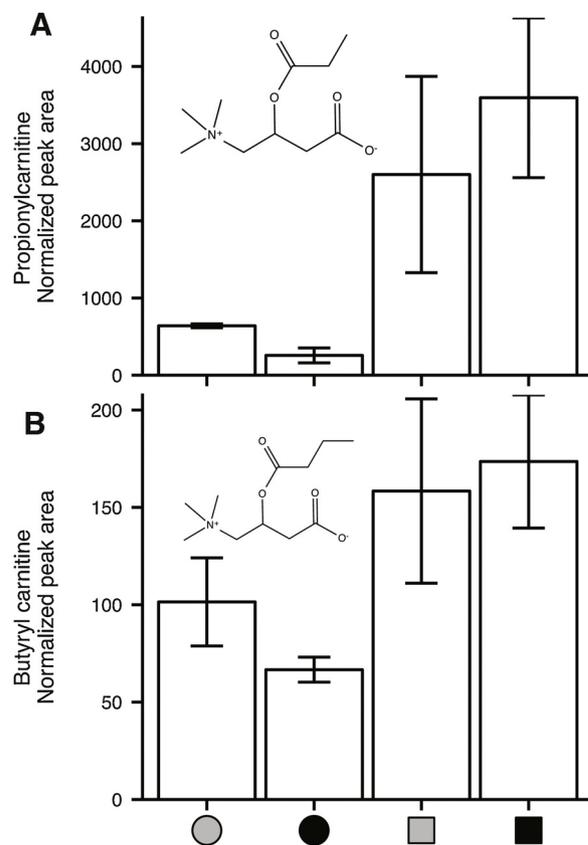
**Figure 9.** Normalized peak areas of homarine (**A**) and proline betaine (**B**) summarized by each of the growing conditions (labeled the same as Fig. 4), with structures in insets. Error bars are standard deviation with  $n=3$  of biological replicates. These compounds were detected in our untargeted analysis.



**Figure 10.** Normalized peak areas of MTA, summarized by each of the growing conditions (labeled the same as Fig. 4). Error bars are standard deviation with  $n=3$  of biological replicates. MTA was detected in our untargeted analysis.

recent study showed that this metabolite is less abundant in cobalamin-limited *T. pseudonana* cells (Heal et al. 2017). We also found that SAM was generally less abundant under low cobalamin conditions in *T. pseudonana* (on average, 4 times less SAM under cobalamin limited conditions, Figs 3, 4, and 5), though the relationship was not statistically robust in a strict univariate comparison between treatments due to high variability ( $p=0.1$ ). SAM significantly contributed to the NMDS results of *T. pseudonana*'s metabolome ( $p=0.006$ , Supplementary Material Table S3), indicating that the compound plays a significant role in driving differences between the experimental treatments we tested.

We detected changes in the size of the pools of two other metabolites in the central methionine cycle. Methionine and SAH both showed signifi-



**Figure 11.** Normalized peak areas of propionylcarnitine (**A**) and butyrylcarnitine (**B**), summarized by each of the growing conditions (labeled the same as Fig. 1, with structures). Error bars are standard deviation with  $n=3$  of biological replicates. We detected these compounds in our untargeted analysis and confirmed the structures with authentic standards.

cant differences between cobalamin treatments in *T. pseudonana* ( $p=0.04$ ,  $p<0.001$ , respectively), where SAH was 10 times more abundant and methionine 10 times less abundant under cobalamin limitation (Figs 3, 4, and 5).

### Changes in the Transsulfuration Pathway

Two intermediates in the methionine cycle, SAH and homocysteine, are potentially cytotoxic compounds due to their affinity for methyl groups, high reducing power, and structural similarity to SAM or proteinogenic amino acids (Hoffman et al. 1980; Jakubowski 2004; Roe et al. 2002; Selhub 1999). To our knowledge, no mechanisms have been proposed to cope with or prevent the buildup of these metabolites during cobalamin limitation in diatoms. We were unable to detect homocysteine in any samples; this is likely due to metabolite instability or considerable matrix effects confounding the instrumental signal. However, using different methodologies, others have shown that homocysteine builds up under cobalamin limitation in a green algae (Croft et al. 2005), humans (Allen et al. 1993), and rats (Stabler et al. 1997) indicating this is a common consequence of cobalamin limitation in cobalamin dependent organisms.

In humans, a major consequence of cobalamin limitation is an increase in the traffic of metabolites through the transsulfuration pathway (the metabolic pathway that converts homocysteine to cysteine and vice versa) that results in an accumulation of cystathionine (Hannibal et al. 2016; Stabler et al. 1993) (Fig. 5). Of the targeted metabolites we observed in this experiment, the sulfur-containing cystathionine was the most drastically affected by cobalamin availability, increasing greatly under low cobalamin (Figs 3, 5, and 6). Others have attempted to uncover the exact enzymology of the transsulfuration pathway through a phylogenetic analysis in *T. pseudonana*, but these attempts were unsuccessful due to the high sequence similarity and common cofactor binding sites in the four enzymes involved in cystathionine synthesis and lysis (cystathionine- $\beta$ -synthase, cystathionine- $\beta$ -lyase, cystathionine- $\gamma$ -synthase, and cystathionine- $\gamma$ -lyase) and high divergence from characterized enzymes (Bromke and Hesse 2015). Several sequences in *T. pseudonana* and other diatom genomes have high sequence similarity to characterized enzymes in the transsulfuration pathway (Bromke and Hesse 2015), so it is plausible that diatoms have the ability to move molecules both directions through the transsulfuration pathway, but without experimental verification, we are unable to reliably distinguish or

identify these genes. Regardless of the particular direction of metabolite traffic, our data suggest that either more cystathionine is produced or less cystathionine is converted to cysteine or homocysteine under cobalamin limitation. We propose that this change in the transsulfuration traffic is a strategy to prevent or cope with the accumulation of homocysteine and SAH under cobalamin limitation.

In other eukaryotes, the activity of at least one of the enzymes in the transsulfuration pathway is regulated allosterically. Specifically, the production of cystathionine from homocysteine via cystathionine- $\beta$ -synthase depends on the concentration of small metabolites in the methionine cycle (Banerjee et al. 2003; Finkelstein et al. 1975). The regulation of the other enzymes in the transsulfuration pathway is less understood, but if they are also regulated allosterically, it may explain why previous studies have not observed a change in the transcription or translation of genes in the transsulfuration pathway in diatoms as a result of cobalamin limitation. Detecting the differential abundance of this compound in *T. pseudonana* under low cobalamin demonstrates the power of measuring metabolites in concert with transcriptomic and proteomic studies when investigating the physiological responses and adaptations of these important organisms.

### Depletion of Major Osmolyte Pools

Two of *T. pseudonana*'s abundant pools of osmolytes, DMSP and GBT, are greatly reduced under cobalamin scarcity (Figs 3, 7, and 8). Osmolytes present as large pools of polar compounds that algae use to balance osmotic pressure in a saline environment; in *T. pseudonana*, DMSP and GBT are two of the three largest peaks we observed in the targeted analyses, indicating they are very abundant molecules and supporting their role as osmolytes. This observation is consistent with other published metabolomes of *T. pseudonana* (Kujawinski et al. 2017). Though the biosynthetic pathway for DMSP is not fully elucidated in diatoms (Curson et al. 2018), it has been shown that DMSP biosynthesis uses methionine as a starting material and a methyl group from SAM (Fig. 7). Thus, DMSP production essentially removes methionine from the methionine cycle (Gage et al. 1997). Our work suggests that cobalamin availability affects DMSP production in *T. pseudonana*, as this compound dramatically decreases under cobalamin limitation (Figs 3 and 7). Like previous work (Kettles et al. 2014), we see a change in intracellular DMSP under the different light regimes, where cultures

grown under saturating light had more DMSP when compared to low light cultures. This supports the osmolyte's alternate roles as an antioxidant or in the storage of excess carbon and sulfur (Kettles et al. 2014; Stefels 2000). Due to our experimental design, we cannot rule out that the change we observed in DMSP between cobalamin treatments is due to changes in growth rate as the pattern followed growth rate changes. DMSP is the precursor for the climatically important gas dimethylsulfide (Andreae and Raemdonck 1983), and our results suggest that cobalamin status may be a key factor in controlling DMSP production in algae.

We also observed a dramatic decrease of GBT under low cobalamin conditions (Figs 3 and 8), supporting former conjectures that GBT production in diatoms depends on the presence of copious amounts of SAM (Bertrand and Allen 2012), which is true for other organisms (Kurepin et al. 2015). In higher plants, GBT is synthesized from choline, which receives three methyl groups from three SAM molecules (Summers and Weretilnyk 1993; Stekol et al. 1958). Choline is also less abundant in *T. pseudonana* under cobalamin limitation, though not to such a dramatic degree as GBT (on average two times less abundant,  $p=0.049$  under saturating light,  $p=0.041$  under low light, Fig. 8, Supplementary Material Table S1). We detected another choline derivative, glycero-phosphocholine, in our untargeted analysis; this compound also decreased under cobalamin limitation (Fig. 8, Table 4). Interestingly, choline and GBT were also affected by the different light regimes, which may point to other biochemical roles for these compounds in diatoms. In our targeted analysis, we did not see an obvious replacement for GBT and DMSP in the osmolyte pools of *T. pseudonana* under low cobalamin conditions, but our untargeted analysis revealed that two other possible osmolytes (homarine and proline betaine) increased under low cobalamin conditions (Fig. 9). Both DMSP and GBT have been proposed as important energy and carbon sources for heterotrophic bacteria living among algae in the surface ocean due to the high concentrations of these molecules in algal biomass (Kiene et al. 2000; Sun et al. 2011). Furthermore, DMSP availability has been shown to change the metabolite pools within marine heterotrophs (Johnson et al. 2016). Cobalamin availability clearly influences which osmolyte pools diatoms produce and therefore provide to a wider community. In this way, cobalamin availability may change which members of the heterotrophic community can thrive, even if those organisms are not directly reliant on cobalamin.

The diversity and distribution of osmolytes among organisms is an outstanding area of research in evolutionary and molecular biology, and this work reveals new insight regarding the role of cobalamin availability on the osmolyte suite produced by algae. The apparent dependencies of GBT and DMSP on SAM and methionine put extra strain on the methionine cycle (and therefore demand more cobalamin) compared to using SAM-independent osmolytes like small carbohydrates or amino acids like proline. The pervasive use of GBT and DMSP in diatoms despite this SAM dependency suggest that these compounds act as superior chemical osmolytes or have functions beyond osmoprotection. For instance, DMSP has been linked with antioxidant capacity and predation defense (Yancey 2005). Another possibility is that there has been little evolutionary pressure for cobalamin-dependent organisms like *T. pseudonana* to use SAM-independent osmolytes because there has been a consistent source of cobalamin to these organisms (i.e. from a persistent cobalamin producing bacterial partner (Croft et al. 2005)); this is supported by the likely loss of the *metE* gene and subsequent cobalamin dependence proposed by others (Helliwell et al. 2015).

## Depletion of MTA

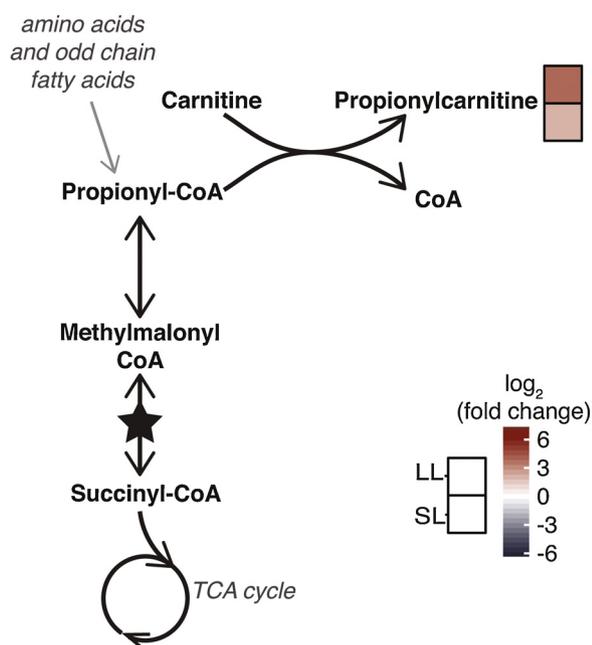
In *T. pseudonana*, 5'-methylthioadenosine (MTA) was about half as abundant in cells in low cobalamin conditions (Fig. 10). This compound is a sulfur-containing nucleoside that is produced from SAM (via decarboxylated-SAM) during polyamine biosynthesis in a two-step process (Fig. 5). We found that diatoms have proteins that are closely related to characterized adenosylmethionine decarboxylase in *Homo sapiens* (Supplementary Material Table S4, Fig. S1), and the second MTA-producing step has been experimentally verified in *T. pseudonana* (Knott et al. 2007). Polyamines are likely essential for growth of all organisms and are the building blocks of long chain polyamines that diatoms use to build their silica frustules (Michael 2011; Kröger et al. 2000). Due to the high demand that polyamine biosynthesis has for SAM, others have hypothesized that polyamine biosynthesis could be disrupted by cobalamin stress (Bertrand et al. 2012), and our observed decrease in MTA corroborates this hypothesis. Unfortunately, our LC-MS analysis does not yield quality data on polyamines, so we cannot measure these molecules using our metabolomics approach. In other organisms, MTA is recycled back to methionine via the methionine

salvage pathway (Albers 2009) (Fig. 5); unfortunately the salvage pathway is poorly characterized in diatoms and our attempts to identify genes associated with this pathway were not fruitful due to high divergence from characterized enzymes. Thus, it is also possible that the observed change of MTA is a result of an increased efficiency of the methionine salvage pathway.

### Accumulation of Acylcarnitines

In *T. pseudonana*, propionylcarnitine showed a similar pattern as SAH, with the cultures grown under low cobalamin having nearly seven times more propionylcarnitine than cobalamin-replete cultures (Fig. 11). Accumulation of propionylcarnitine has been observed in cobalamin deficient mammals (Brass and Stabler 1988; Sarafoglou et al. 2011), and is used as a diagnostic for cobalamin deficiency in newborn humans (Hannibal et al. 2016; Sarafoglou et al. 2011). We also observed an increase in butyrylcarnitine under cobalamin limitation; an increase in this compound has also been observed in cobalamin-deficient animals to a

lesser extent than a concurrent increase in propionylcarnitine (Kelmer et al. 2007), similar to what we observed (Fig. 11). In mammals, the production of acylcarnitines (like propionylcarnitine and butyrylcarnitine) is a method for liberating Coenzyme A (CoA) from acyl-CoAs when cells cannot use these particular acyl-CoAs. Previous work has hypothesized that increased propionylcarnitine observed under cobalamin limitation is a result of the inability to process propionyl- or methylmalonyl-CoA through the cobalamin-dependent enzyme MCM (Hannibal et al. 2016) (Fig. 12). There is strong genomic evidence for this pathway in *T. pseudonana* and other sequenced diatoms. In particular, we found that diatoms encode genes for two enzymes beyond MCM, propionyl CoA carboxylase and carnitine palmitoyl transferase that could explain the increased propionylcarnitine under low cobalamin (Fig. 12, Supplementary Material Figs S2 and S3). Some eukaryotic phytoplankton encode MCM in their genomes, including *T. pseudonana* (Bertrand and Allen 2012; Helliwell et al. 2011), though the presence of MCM in diatom genomes does not result in an absolute cobalamin requirement — for instance, the model diatom *Phaeodactylum tricornutum* codes for and transcribes the enzyme, but can thrive in the absence of cobalamin (Bertrand et al. 2013; Maheswari et al. 2010). Overall, the role and importance of MCM in phytoplankton remains unclear, but our data suggest that *T. pseudonana* uses MCM and experience a decreased efficiency of the pathway under cobalamin limitation.



**Figure 12.** Accumulation of propionylcarnitine and our proposed mechanism in relation to MCM decreased efficiency based on Hannibal et al. (2016).  $\log_2$ (fold change) in metabolite concentrations are shown between cobalamin treatments in saturating light (SL) and low light (LL) conditions as noted in the key. The enzymatic conversion that is cobalamin-dependent methylmalonyl-CoA mutase (MCM) is starred.

### Conclusions

Our work demonstrates that metabolite pools in diatoms are affected by the availability of cobalamin. Many of the changes we observed could be directly linked to the role that cobalamin plays as a coenzyme in Me-cobalamin dependent methionine synthase (MetH), including an imbalanced methionine cycle, altered transsulfuration activity, and rearrangement of osmolyte pools. Our untargeted approach revealed that two acylcarnitines increased under cobalamin limitation, which we propose to be directly linked to a decreased function of the Ado-cobalamin dependent methylmalonyl-CoA mutase (MCM). In areas of the ocean, diatoms experience ephemeral or sustained cobalamin scarcity, and this work elucidates how the availability of this organic micronutrient has far reaching metabolic consequences for these important marine primary producers.

## Methods

**Culture maintenance:** *Thalassiosira pseudonana* CCMP1335 was acquired from National Center for Marine Algae and Microbiota (NCMA, [ncma.bigelow.org](http://ncma.bigelow.org)) and maintained at 13 °C with a 12 hr light:dark cycle (light levels reported in Table 1) in 50 mL combusted borosilicate tubes with 35 mL filter sterilized artificial seawater media (41 g instant ocean salts in 1 L MilliQ water with f/2 nutrients added, except cobalamin). *T. pseudonana* was maintained in 1 pM OH-cobalamin (a concentration known to limit its growth (Bertrand et al. 2013; Heal et al. 2017)) for at least 4 transfers on maintenance media before starting the experiment and could not be maintained without additional cobalamin. The starter culture and media was tested for bacterial contamination via marine purity test broth (Saito et al. 2002) at the start of the experiment. Though test broths are not wholly reliable indicators of axenicity, we did not detect a bacterial signal using a particle counter (see below) for our cell counts, supporting a lack of contamination by bacteria. Since our metabolomics analysis is a direct measurement, it is unlikely that the bulk metabolite signals could be coming from any other source than from the diatoms themselves.

**Experimental conditions:** Using a single inoculum, we inoculated nine 35 mL cultures under four conditions: replete cobalamin, saturating light; replete cobalamin, low light; low cobalamin, saturating light; low cobalamin, low light. For all media, we also prepared a media blank (un-inoculated media treated the same as the samples). Light was manipulated using a neutral density photographic lighting film (Lee Filters, Burbank, CA, USA), and photosynthetically available radiation (PAR) was measured with a LI-250A Light Meter (LI-COR Biosciences, Lincoln, NE, USA). Light levels were defined as low and saturating based on previous work (Stramski et al. 2002). Light levels and initial cobalamin concentrations for each treatment and organism are given in Table 1.

We monitored growth of these four conditions by relative fluorescence over the course of 12 days, taking chlorophyll *a* fluorescence measurements at approximately 11 a.m. each day with a Turner Designs model 10-AU fluorometer (in vivo chlorophyll optical kit). To calculate growth rate, we used days 1–5 for saturating light treatments of *T. pseudonana*, and days 3–7 of low light treatments of *T. pseudonana* using in vivo fluorescence ( $n=9$  for each treatment). Using a combusted glass apparatus and gentle vacuum filtration, we harvested two 35 mL cultures onto one 47 mm 0.2  $\mu\text{m}$  PTFE filters when the cultures were in exponential phase (on days noted on Fig. 1), in triplicate. In order to have enough biomass for the metabolite analysis, we harvested on different days for the low and saturating light conditions. The remaining three cultures were allowed to grow out to observe final culture density. Cultures were harvested at approximately 11 a.m. Throughout exponential growth, we took 1 mL samples (fixed with 1% formaldehyde) for cell counts. We used a Beckman Coulter Z2 Particle Count and Size Analyzer (Beckman Coulter) to measure *T. pseudonana* densities and diameters. At each light level, we determined the in vivo fluorescence versus cell concentration ( $R^2 > 0.9$  for all) and converted our fluorescence at harvest to cell density on filters. Between treatments, we did not observe any significant differences in the size of *T. pseudonana*, so we used cell counts as a proxy for biomass for each species for normalization (see data processing section).

**Metabolite extraction:** Filters of samples and media blanks were extracted using a modified Bligh Dyer extraction (Bligh and Dyer 1959) as described in detail in Boysen et al. (2018), which resulted in a non-polar organic fraction and a polar aqueous fraction. In this study, we were interested in the primary

metabolites (not lipids that are retained in the organic fraction) and therefore only analyzed the aqueous fraction. To aid in normalization, we added a cocktail of internal standards (listed in Supplementary Material Table S5) before and after extraction.

**Data acquisition:** To obtain metabolomic data, we separated compounds using reversed phase (RP) and hydrophilic interaction liquid chromatography (HILIC), using the exact specifications as previously described (Boysen et al. 2018). After extraction, samples were run within 24 hours for HILIC and within 96 hours for RP; between extraction and analysis, samples were stored at  $-80^\circ\text{C}$ . We obtained targeted and untargeted metabolomics data on triple quadrupole (TQS) and Q-Exactive (QE) mass spectrometers (MSs), respectively. Full scan (non-fragmented mass to charge ( $m/z$ )) data acquisition from the QE and all data acquisition from the TQS are described fully in a previous publication from our group (Boysen et al. 2018). Throughout the run, we ran a pooled sample several times in order to monitor signal stability and train normalization.

For the untargeted metabolomics analysis, we collected fragmentation spectra ( $\text{MS}^2$ ) on our pooled samples using data dependent acquisition (DDA). For samples separated via HILIC, separate injections were analyzed in positive and negative ion modes.  $\text{MS}^2$  spectra were collected from 50–500  $m/z$  at resolution of 30,000 at 200  $m/z$ . For each MS scan, DDA was set to perform on the top five most abundant ions with dynamic exclusion of 20 seconds using a normalized collision induced dissociation at 35 V. For RP, samples were only run in positive mode; DDA was performed in the same manner but with a dynamic exclusion time of 10 seconds due to narrower peaks than in HILIC.

**Targeted data processing:** Targeted data (from the TQS) were subject to an in-house quality control, blank subtracted, and normalized via best-matched internal standard (B-MIS) normalization (Boysen et al. 2018), using a 20% improvement to the relative standard deviation (RSD) of each mass feature in a pooled sample as a criteria to apply normalization. We arrived at the 20% improvement criteria using tools in the B-MIS normalization package (Boysen et al. 2018). Like our previous work (Boysen et al. 2018), we did not normalize any mass features that had a  $\text{RSD}_{\text{raw}}$  of  $<10\%$  in the raw pooled area, instead defaulting to the raw area. This normalization process resulted in adjusted peak areas that minimize obscuring variation—the non-biological variation inherent to LC–MS analyses. We then normalized the adjusted peak areas to cell counts of each biological replicate to account for different amounts of cellular material on each filter. For subsequent analyses, the normalized adjusted peak areas were used. For univariate and multivariate statistical analyses (see sections below), we only included compounds that were detected in at least two treatments. For our targeted analysis, when targeted compounds were detected in only a subset of the treatments or replicates (or small peaks were removed during the quality control step), we assigned a value for the remaining treatments or replicates that represents an upper estimate of how large a peak could be present and still remain below our detection limit (3 *peak area in blank* + 100). This value underwent the same normalization (B-MIS and to cell counts).

**Untargeted data processing:** Data collected for MS and  $\text{MS}^2$  from the QE were first converted to .mzXML using MSConvert (Chambers et al. 2012) with positive and negative scans processed separately. These raw data were submitted to the Metabolights data repository (<http://www.ebi.ac.uk/metabolights>) (Haug et al. 2013) and can be accessed under study ID MTBLS703.

For MS data, each fraction (RP, positive scans on HILIC, and negative scans on HILIC) were processed separately through XMCS (Benton et al. 2010; Smith et al. 2006; Tautenhahn et al. 2008) with parameters optimized via Isotopologue Parameter Optimization (Libiseller et al. 2015) (reported in Supplementary Material Table S6). For each sample, this resulted in a list of peak areas of mass features (peaks with unique retention time (RT) and  $m/z$ ). We filtered out any mass features with an average peak area less than three times larger than the methodological blank or with an average peak size less than 100. Next, we normalized for obscuring variation using B-MIS normalization (Boysen et al. 2018), using the same criteria as in our targeted analyses. We calculated the coefficient of variation (CV) of each mass feature in injections of the pooled sample and removed peaks that did not demonstrate acceptable replicability (CV >30%). Next, we identified mass features that were likely  $^{13}\text{C}$ ,  $^{15}\text{N}$ , or  $^{34}\text{S}$  isotopologues of other mass features because we did not want to give these compounds extra weight in our statistical analyses. To do this, we searched for expected differences in  $m/z$  and intensity between mass features for each of the aforementioned isotopes within each three second (for RP) or six second (for HILIC) corrected retention time window. We excluded these  $^{13}\text{C}$ ,  $^{15}\text{N}$ , or  $^{34}\text{S}$  mass features from downstream analyses. Like in the targeted analysis, we normalized the adjusted peak areas of quality peaks to cell counts of each replicate to account for different amount of cellular material on each filter, and for subsequent analyses, the normalized adjusted peak areas of the filtered mass features were used.

**MS<sup>2</sup> data processing:** For quality mass features, we searched the .mzXML files collected for DDA analysis for MS<sup>2</sup> scans that match the parent  $m/z$  (at 0.5 Da tolerance) and retention time (at 10 and 20 second tolerance for RP and HILIC chromatography, respectively). The isolation chamber of the QE is low resolution, so it is possible that a parent scan will match at 0.5 Da but the fragmentation spectra present is from a different parent  $m/z$  with the same nominal mass — this is especially problematic for low abundance peaks. Therefore, for each matched MS<sup>2</sup> scan (at 0.5 Da), we filtered out any scans where the parent  $m/z$  was not present in the high-resolution fragmentation spectra at 0.02 Da tolerance. If multiple scans were found, we summed the scans' intensities for each fragment across the scans as fragments' intensities can vary between scans. We filtered out peaks that contributed less than 0.5% of the intensity of the most intense fragment.

**Compound identification:** Using the ranking system outlined in Sumner et al. (2007), we attempted to identify the quality mass features present in our sample sets. First we removed any mass features that were plausible contaminants by searching for common contaminants (Keller et al. 2008). For the remaining mass features, we searched an internal database of standards run in the exact manner on our instruments for matches to exact  $m/z$  and retention time, yielding an *unequivocal identification* (confidence level 1); for current list of the standards run in this exact manner see [https://github.com/kheal/Example\\_Untargeted\\_Metabolomics\\_Workflow/blob/master/Ingalls\\_Lab\\_Standards.csv](https://github.com/kheal/Example_Untargeted_Metabolomics_Workflow/blob/master/Ingalls_Lab_Standards.csv). For mass features without a standard match but with MS<sup>2</sup>, we searched against the publicly available LC/MS MS<sup>2</sup> spectral databases, including MassBank (Horai et al. 2010), Global Natural Products Social Molecular Networking (Wang et al. 2016), MetiTree (Vaniya and Fiehn 2015), RIKEN tandem mass spectral database (Sawada et al. 2012), and the Human Metabolome Database (Wishart et al. 2007) (all downloaded from <http://mona.fiehnlab.ucdavis.edu/downloads> on November 17, 2017). We searched for spectra with matching parent

$m/z$  and a cosine similarity >0.8 from an ESI spectrum, using the same algorithms as MassBank (Horai et al. 2010). When a compound in the MassBank database matched  $m/z$  and MS<sup>2</sup>, we assigned a *putative identification* (confidence level 2). For some of these compounds (including all of those explicitly discussed in this work), we later obtained standards and upgraded their identification to confidence level 1 by matching MS<sup>2</sup> and retention time. For mass features without confidence levels 1 or 2 identification, we searched for possible matches to compounds in the KEGG database (Kanehisa et al. 2016, 2017) based only on  $m/z$ , these identifications are considered *possible identification* (confidence level 3). For many mass features, the aforementioned identification attempts were not fruitful, so we attempted to identify quality mass features that showed a univariate response to cobalamin limitation, using a manual approach with the larger, but not batch-searchable or publicly downloadable, Metlin database (Smith et al. 2005). We searched mass features by  $m/z$  (assuming M+H, M+NH<sub>4</sub>, or M+Na for positive ionization, M-H, M+Cl for negative ionization, at a 5 ppm tolerance) and compared MS<sup>2</sup> when obtained by DDA and available in the database. We assigned putative identifications using the same confidence levels as discussed above.

**Multivariate statistics:** For all multivariate statistics, we used data that were standardized to  $z$ -scores for each compound or mass feature, where  $z = (X - \mu) / \sigma$ , where  $z$  is the  $z$ -score,  $X$  is the adjusted normalized peak area in each replicate,  $\mu$  is the mean peak area, and  $\sigma$  is the standard deviation of the peak areas across each sample set (for each metabolite, separated by organism). We employed two multivariate analyses to analyze changes in the metabolome of *T. pseudonana*. We used a non-metric dimensional scaling (NMDS) approach (Kruskal and Wish 1978) based on a euclidean distance to analyze differences in the treatments. We selected this non-metric approach due to our low sample numbers, high variable numbers, and the non-normal nature of our dataset to avoid overfitting common in metric approaches of ordination in metabolomics (Saccenti et al. 2013). We assessed dimensionality of the NMDS by examining a scree plot and calculated the probability with a monte carlo permutation. We paired this NMDS with an analysis of similarities (ANOSIM) to determine whether our metabolite data could distinguish differences between light and cobalamin conditions, using 999 permutations. All multivariate statistics were performed in R using the vegan package (V2.4-2).

**Univariate statistics:** We used an unpaired  $t$ -test to compare growth rates between experimental treatments. For targeted and untargeted analyses, compounds (from our targeted analysis) and mass features (from our untargeted analysis) were investigated for significant differences between low and replete cobalamin treatments and low and saturating light treatments in each organism. For all univariate statistics on metabolomes, we corrected  $p$  values for false discovery rate (Benjamini and Hochberg 1995) and report those corrected values throughout the text. We calculated fold changes and  $p$ -values (via an unpaired  $t$ -test) between low and replete cobalamin treatments (regardless of light status) and low and saturating light treatments (regardless of cobalamin status,  $n = 6$ ). Finally, we calculated fold changes and  $p$  values between cobalamin treatments within each light status and between light treatments for each cobalamin status ( $n = 3$  for each treatment). For mass features in the untargeted data, we visually examined peaks that had resulted in significant  $p$  values between cobalamin treatments and removed any mass features that did not have quality peak shapes or integrations from downstream analyses.

**Gene searches:** To corroborate our metabolomics analysis, we performed a BLASTP (Altschul et al. 1990) search on the *T. pseudonana* genome models vX (Armbrust et al. 2004) for proteins related to propionylcarnitine metabolism and polyamine biosynthesis using functionally verified sequences from either *Arabidopsis thaliana* or *Homo sapiens* as queries. To gain an understanding if these pathways were well distributed among diatoms, we also searched publicly available genome models from *Phaeodactylum tricornutum* (Bowler et al. 2008), *Fragilariopsis cylindrus* (Mock et al. 2017), *Thalassiosira oceanica* (Lommer et al. 2012), and *Pseudo-nitzschia multiseriata* CLN-47.

For each query protein BLASTP, we searched for potential orthologs in an in-house, curated protein database of publicly available genomes and assembled transcriptomes from marine eukaryotes available through JGI and the Marine Microbial Eukaryote Transcriptome Sequence Project (MMETSP) (Keeling et al. 2014), which has been expanded to include marine bacterial and archaeal genomes available through NCBI (Groussman et al. 2015). We used an E-value cut off of  $<1e-15$  for our BLASTP search. Sequences were clustered at  $>80\%$  sequence identity (Edgar 2010) and a single representative from each cluster was aligned using mafft E-INS-i v7.407 (Katoh et al. 2002). The resulting alignment was trimmed using trimAL v1.2 (Capella-Gutiérrez et al. 2009). The most appropriate substitution models were determined based on the Akaike information criterion using ProtTest v3.4.2 (Abascal et al. 2005). Phylogeny was inferred with maximum-likelihood amino acid phylogenetic trees generated using RAxML v8.1.20 (Stamatakis 2014). Bootstrap support values were computed using Booster (Lemoine et al. 2018). Archaeopteryx software (Han and Zmasek 2009) was used to generate graphics. We identified conserved domains between putative diatom proteins and experimentally verified proteins using Interpro-scan sequence search (Jones et al. 2014).

## Acknowledgements

The authors would like to acknowledge E. Armbrust, B. Durham, R. Lundeen, G. Rocap, R. Groussman, A. Boysen, and J. Young for input on the experimental design, data analysis, and manuscript editing. This work was supported by grants from the Simons Foundation (LS Award ID: 385428, A.E.I.; SCOPE Award ID 329108, A.E.I.; Award ID 598819, KRH), NSF OCE-1228770 and OCE-1205232 to A.E.I., NSF GRFP to K.R.H.

## Appendix A. Supplementary Data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.protis.2019.05.004>.

## References

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

Albers E (2009) Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5'-methylthioadenosine. *IUBMB Life* **61**:1132–1142

Allen RH, Stabler P, Savage G, Lindenbaum J (1993) Metabolic abnormalities in cobalamin (vitamin B<sub>12</sub>) and folate deficiency. *FASEB J* **7**:1344–1353

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**:403–410

Andreae MO, Raemdonck H (1983) Dimethyl sulfide in the surface ocean and the marine atmosphere: a global view. *Science* **221**:744–747

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**:79–86

Banerjee R, Evande R, Kabil Ö, Ojha S, Taoka S (2003) Reaction mechanism and regulation of cystathionine  $\beta$ -synthase. *Biochim Biophys Acta Proteins Proteom* **1647**:30–35

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B (Methodol)* **57**:289–300

Benton HP, Want EJ, Ebbels TMD (2010) Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics* **26**:2488–2489

Bertrand EM, Allen AE (2012) Influence of vitamin B auxotrophy on nitrogen metabolism in eukaryotic phytoplankton. *Front Microbiol* **3**:1–16

Bertrand EM, Moran DM, McIlvin MR, Hoffman JM, Allen AE, Saito MA (2013) Methionine synthase interreplacement in diatom cultures and communities: implications for the persistence of B<sub>12</sub> use by eukaryotic phytoplankton. *Limnol Oceanogr* **58**:1431–1450

Bertrand EM, Allen AE, Dupont CL, Norden-Krichmar TM, Bai J, Valas RE, Saito MA (2012) Influence of cobalamin scarcity on diatom molecular physiology and identification of a cobalamin acquisition protein. *Proc Natl Acad Sci USA* **109**:E1762–E1771

Bertrand EM, Saito MA, Rose JM, Riesselman CR, Lohan MC, Noble AE, Lee PA, Di Tullio GR (2007) Vitamin B<sub>12</sub> and iron colimitation of phytoplankton growth in the Ross Sea. *Limnol Oceanogr* **52**:1079–1093

Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, Post AF, Sipler RE, Spackeen JL, Xu K (2015) Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc Natl Acad Sci USA* **112**:9938–9943

Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* **37**:911–917

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, O'tillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falcione A, Fournet J, Haruta M, Huysman MJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kröger N, Kroth PG,

- La Roche J, Lindquist E, Lommer M, Martin-Jézéquel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, MPOL Secq, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, Von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van De Peer Y, Grigoriev IV (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**:239–244
- Boysen A, Heal K, Carlson L, Ingalls A (2018) Best-matched internal standard normalization in liquid chromatography–mass spectrometry metabolomics applied to environmental samples. *Anal Chem* **9**:1363–1369
- Brass EP, Stabler SP (1988) Carnitine metabolism in the vitamin B<sub>12</sub> deficient rat. *Biochem J* **255**:153–159
- Bromke MA, Hesse H (2015) Phylogenetic analysis of methionine synthesis genes from *Thalassiosira pseudonana*. *SpringerPlus* **4**:1–14
- Brown AD (1976) Microbial water stress. *Bacteriol Rev* **40**:803–846
- Browning TJ, Achterberg EP, Rapp I, Engel A, Bertrand EM, Tagliabue A, Moore CM (2017) Nutrient co-limitation at the boundary of an oceanic gyre. *Nature* **551**:242–246
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973
- Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**:918–920
- Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG, Croft MT, Raux-Deery E, Warren MJ, Smith AG (2005) Algae acquire vitamin B<sub>12</sub> through a symbiotic relationship with bacteria. *Nature* **438**:90–93
- Croft MT, Warren MJ, Smith AG (2006) Algae need their vitamins. *Eukaryot Cell* **5**:1175–1183
- Curson AR, Williams BT, Pinchbeck BJ, Sims LP, Martínez AB, Rivera PPL, Kumaresan D, Mercadé E, Spurgin LG, Carrión O, Moxon S, Cattolico RA, Kuzhiumparambil U, Guagliardo P, Clode PL, Raina JB, Todd JD (2018) DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. *Nat Microbiol* **3**:430–439
- Doxey AC, Kurtz DA, Lynch MDJ, Sauder LA, Neufeld JD (2014) Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J* **9**:461–471
- Droop MR (1968) Vitamin B<sub>12</sub> and marine ecology. IV. The kinetics of uptake, growth and inhibition in *Monochrysis lutheri*. *J Mar Biol Assoc UK* **48**:689–733
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:2460–2461
- Finkelstein JD, Kyle WE, Martin JJ, Pick AM (1975) Activation of cystathionine  $\beta$ -synthase by adenosylmethionine and adenosylethionine. *Biochem Biophys Res Commun* **66**:81–81
- Gage DA, Rhodes D, Nolte KD, Hicks WA, Leustek T, Cooper AJL, Hanson AD (1997) A new route for synthesis of dimethylsulphoniopropionate in marine algae. *Nature* **387**:891–894
- Gonzalez JC, Banerjee RV, Huang S, Sumner JS, Matthews RG (1992) Comparison of cobalamin-independent and cobalamin-dependent methionine synthases from *Escherichia coli*: two solutions to the same chemical problem. *Biochemistry* **31**:6045–6056
- Groussman RD, Parker MS, Armbrust EV (2015) Diversity and evolutionary history of iron metabolism genes in diatoms. *PLoS ONE* **10**(6):e0129081
- Han MV, Zmasek CM (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**:1–6
- Hannibal L, Lysne V, Bjørke-Monsen AL, Behringer S, Grunert SC, Spiekerkoetter U, Jacobsen DW, Blom HJ (2016) Biomarkers and algorithms for the diagnosis of vitamin B<sub>12</sub> deficiency. *Front Mol Biosci* **3**:27
- Haug K, Salek RM, Conesa P, Hastings J, De Matos P, Rijnbeek M, Mahendraker T, Williams M, Neumann S, Rocca-Serra P, Maguire E, González-Beltrán A, Sansone SA, Griffin JL, Steinbeck C (2013) MetaboLights — an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res* **41**:D781–D786
- Heal KR, Carlson LT, Devol AH, Armbrust E, Moffett JW, Stahl DA, Ingalls AE (2014) Determination of four forms of vitamin B<sub>12</sub> and other B vitamins in seawater by liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectr* **28**:2398–2404
- Heal KR, Qin W, Ribalet F, Bertagnolli AD, Coyote-Maestas W, Hmelo LR, Moffett JW, Devol AH, Armbrust EV, Stahl DA, Ingalls AE (2017) Two distinct pools of B<sub>12</sub> analogs reveal community interdependencies in the ocean. *Proc Natl Acad Sci USA* **114**:364–369
- Helliwell KE, Wheeler GL, Leptos KC, Goldstein RE, Smith AG (2011) Insights into the evolution of vitamin B<sub>12</sub> auxotrophy from sequenced algal genomes. *Mol Biol Evol* **28**:2921–2933
- Helliwell KE, Collins S, Kazamia E, Purton S, Wheeler GL, Smith AG (2015) Fundamental shift in vitamin B<sub>12</sub> eco-physiology of a model alga demonstrated by experimental evolution. *ISME J* **9**:1446–1455
- Helliwell KE, Lawrence AD, Holze A, Kudahl UJ, Sasso S, Kräutler B, Scanlan DJ, Warren MJ, Smith AG (2016) Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B<sub>12</sub>. *Curr Biol* **26**:999–1008
- Hoffman DR, Marion DW, Cornatzer WE, Duerre JA (1980) S-Adenosylmethionine and S-adenosylhomocysteine metabolism in isolated rat liver. *J Biol Chem* **255**:10822–10827
- Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T (2010)

- MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectr* **45**:703–714
- Jakubowski H** (2004) Molecular basis of homocysteine toxicity in humans. *Cell Mol Life Sci* **61**:470–487
- Johnson WM, Kido Soule MC, Kujawinski EB** (2016) Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP. *ISME J* **10**:2304–2316
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S** (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240, [arXiv:1501.3606](https://arxiv.org/abs/1501.3606)
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K** (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**:D353–D361, [arXiv:1611.06654](https://arxiv.org/abs/1611.06654)
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M** (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**:D457–D462
- Katoh K, Misawa K, Ki Kuma, Miyata T** (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059–3066
- Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, Petroni G, Piganeau G, Posewitz MC, Rengefors K, Romano G, Rumpho ME, Rynearson T, Schilling KB, Schroeder DC, Simpson AG, Slamovits CH, Smith DR, Smith GJ, Smith SR, Sosik HM, Stief P, Theriot E, Twary SN, Umale PE, Vaulot D, Wawrik B, Wheeler GL, Wilson WH, Xu Y, Zingone A, Worden AZ** (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**(6):e1001889
- Keller BO, Sui J, Young AB, Whittall RM** (2008) Interferences and contaminants encountered in modern mass spectrometry. *Anal Chim Acta* **627**:71–81
- Kelmer E, Shelton GD, Williams DA, Ruaux CG, Kerl ME, O'Brien DP** (2007) Organic acidemia in a young cat associated with cobalamin deficiency. *J Vet Emerg Crit Care* **17**:299–304
- Kettles NL, Kopriva S, Malin G** (2014) 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**(4):e94795
- Kiene RP, Linn LJ, Bruton JA** (2000) New and important roles for DMSP in marine microbial communities. *J Sea Res* **43**:209–224
- Knott JM, Römer P, Sumper M** (2007) Putative spermine synthases from *Thalassiosira pseudonana* and *Arabidopsis thaliana* synthesize thermospermine rather than spermine. *FEBS Lett* **581**:3081–3086
- Kröger N, Deutzmann R, Bergsdorf C, Sumper M** (2000) Species-specific polyamines from diatoms control silica morphology. *Proc Natl Acad Sci USA* **97**:14133–14138
- Kruskal JB, Wish M** (1978) Multidimensional Scaling. In Uslaner EM (ed) *Sage University Paper Series on Quantitative Applications in the Social Sciences*. Sage Publications, Beverly Hills and London, pp 7–28
- Kujawinski EB, Longnecker K, Alexander H, Dyhrman ST, Fiore CL, Haley ST, Johnson WM** (2017) Phosphorus availability regulates intracellular nucleotides in marine eukaryotic phytoplankton. *Limnol Oceanogr Lett* **2**:119–129
- Kurepin LV, Ivanov AG, Zaman M, Pharis RP, Allakhverdiev SI, Hurry V, Hüner NP** (2015) Stress-related hormones and glycine betaine interplay in protection of photosynthesis under abiotic stress conditions. *Photosynth Res* **126**:221–235
- Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O** (2018) Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**:452–456
- Libiseller G, Dvorzak M, Kleb U, Gander E, Eisenberg T, Madeo F, Neumann S, Trausinger G, Sinner F, Pieber T, Magnes C** (2015) IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics* **16**:118
- Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, LaRoche J** (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol* **13**:R66
- Lu SC** (2000) S-Adenosylmethionine. *Int J Biochem Cell Biol* **32**:391–395
- Maheswari U, Jabbari K, Petit JL, Porcel BM, Allen AE, Cadoret JP, De Martino A, Heijde M, Kaas R, La Roche J, Lopez PJ, Martin-Jézéquel V, Meichenin A, Mock T, Schnitzler Parker M, Vardi A, Armbrust EV, Weissenbach J, Katinka M, Bowler C** (2010) Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol* **11**:R85
- Michael AJ** (2011) Molecular machines encoded by bacterially-derived multi-domain gene fusions that potentially synthesize, N-methylate and transfer long chain polyamines in diatoms. *FEBS Lett* **585**:2627–2634
- Mock T, Otilar RP, Strauss J, McMullan M, Paaanen P, Schmutz J, Salamov A, Sanges R, Toseland A, Ward BJ, Allen AE, Dupont CL, Frickenhaus S, Maumus F, Veluchamy A, Wu T, Barry KW, Falcatore A, Ferrante MI, Fortunato AE, Glöckner G, Gruber A, Hipkin R, Janech MG, Kroth PG, Leese F, Lindquist EA, Lyon BR, Martin J, Mayer C, Parker M, Quesneville H, Raymond JA, Uhlig C, Valas RE, Valentin KU, Worden AZ, Armbrust EV, Clark MD, Bowler C, Green BR, Moulton V, Van Oosterhout C, Grigoriev IV** (2017) Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**:536–540
- Provasoli L, Carlucci AF** (1974) Vitamins and Growth Regulators. In Stewart WDP, Baker HG, Beevers H, Whatley FR (eds)

Algal Physiology and Biochemistry. Botanical Monographs, University of California Press, Berkeley and Los Angeles, pp 741–787

Roe AJ, O'Byrne C, McLaggan D, Booth IR (2002) Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. *Microbiology* **148**:2215–2222

Roje S (2006) S-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry* **67**:1686–1698

Romine MF, Rodionov DA, Maezato Y, Osterman AL, Nelson WC (2017) Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities. *ISME J* **11**:1434–1446

Saccenti E, Hoefsloot HC, Smilde AK, Westerhuis JA, Hendriks MM (2013) Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **10**:361–374

Saito MA, Moffett JW, Chisholm SW, Waterbury JB (2002) Cobalt limitation and uptake in *Prochlorococcus*. *Limnol Oceanogr* **47**:1629–1636

Sañudo-Wilhelmy SA, Gobler CJ, Okbamichael M, Taylor GT (2006) Regulation of phytoplankton dynamics by vitamin B<sub>12</sub>. *Geophys Res Lett* **33**:10–13

Sarafoglou K, Rodgers J, Hietala A, Matern D, Bentler K (2011) Expanded newborn screening for detection of vitamin B<sub>12</sub> deficiency. *JAMA* **305**:1198–1200

Sawada Y, Nakabayashi R, Yamada Y, Suzuki M, Sato M, Sakata A, Akiyama K, Sakurai T, Matsuda F, Aoki T, Hirai MY, Saito K (2012) RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry* **82**:38–45

Selhub J (1999) Homocysteine metabolism. *Annu Rev Nutr* **19**:217–246

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**:779–787

Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G (2005) Metlin A metabolite mass spectral database. *Ther Drug Monit* **27**:747–751

Stabler SP, Lindenbaum J, Savage DG, Allen RH (1993) Elevation of serum cystathionine levels in patients with cobalamin and folate deficiency. *Blood* **81**:3404–3413

Stabler SP, Sampson DA, Wang LP, Allen RH (1997) Elevations of serum cystathionine and total homocysteine in pyridoxine-, folate-, and cobalamin-deficient rats. *J Nutr Biochem* **8**:279–289

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

Stekol JA, Anderson EI, Weiss S (1958) S-Adenosyl-L-methionine in the synthesis of choline, creatine, and cysteine *in vivo* and *in vitro*. *J Biol Chem* **233**:425–429

Stramski D, Sciandra A, Claustre H (2002) Effects of temperature, nitrogen, and light limitation on the optical properties of the marine diatom *Thalassiosira pseudonana*. *Limnol Oceanogr* **47**:392–403

Summers PS, Weretilnyk EA (1993) Choline synthesis in spinach in relation to salt stress. *Plant Physiol* **103**:1269–1276

Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J (2007) Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**:211–221

Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, Landry ZC, Giovannoni SJ (2011) One carbon metabolism in SAR11 pelagic marine bacteria. *PLoS ONE* **6**(8):e239736

Tang YZ, Koch F, Gobler CJ (2010) Most harmful algal bloom species are vitamin B<sub>1</sub> and B<sub>12</sub> auxotrophs. *Proc Natl Acad Sci USA* **107**:20756–20761

Tautenhahn R, Böttcher C, Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatic* **9**:504

Vaniya A, Fiehn O (2015) Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. *Trends Anal Chem* **69**:52–61

Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kaponov CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedió BE, Klitgaard A, Larson CB, Boya CA, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJ, Granatosky EA, Glukhov E, Ryyfel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AM, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson B, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**:828–837

Warren MJ, Raux E, Schubert HL, Escalante-Semerena JC (2002) The biosynthesis of adenosylcobalamin (vitamin B<sub>12</sub>). *Nat Product Rep* **19**:390–412

Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivas-

tava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, MacInnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L (2007) HMDB:

the human metabolome database. *Nucleic Acids Res* **35**: 521–526

Yancey PH (2005) Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J Exp Biol* **208**:2819–2830

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

**ScienceDirect**