A phosphate starvation response gene (*psr*1-like) is present and expressed in *Micromonas pusilla* and other marine algae

Cara L. Fiore[1,*], Harriet Alexander[2], Melissa C. Kido Soule, Elizabeth B. Kujawinski

Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

[1]Current address: Biology Department, Appalachian State University, Boone, NC, USA

[2]Current address: Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

*Corresponding author: Cara Fiore
572 Rivers St.
Boone, NC 28607, USA
fiorec@appstate.edu

Running title: Psr1 transcription factor in marine phytoplankton

Competing interest statement: The authors declare that they have no conflict of interest

Abstract

Phosphorus (P) limits primary production in regions of the surface ocean, and many plankton species exhibit specific physiological responses to P-deficiency. The metabolic response of *Micromonas pusilla*, an ecologically relevant marine photoautotroph, to P-deficiency was investigated using environmental metabolomics and comparative genomics. The concentrations of some intracellular metabolites were elevated in the P-deficient cells (e.g., xanthine) and genes involved in the associated metabolic pathways shared a significant motif in the non-coding regions of the gene. The presence of the conserved motif suggests that these genes may be co-regulated, and the motif may constitute a regulatory element for binding a transcription factor (i.e., Psr1). A putative phosphate starvation response gene (*psr*1-like) was identified in *M. pusilla* with homology to well characterized *psr*1/*phr*1 genes in freshwater algae and plants. This gene appears to be present and expressed in other marine algal taxa, such as the abundant haptophyte *Emiliania huxleyi*, in chronically phosphorus-limited field sites. Results from the present study have implications for our understanding of taxon-specific roles in mediating P cycling in the ocean.

Key words: *Micromonas pusilla*, phosphate stress response, marine algae, metabolomics, dissolved organic matter

Introduction

Phosphorus (P) is found in lipid membranes, genetic material, and energy storage compounds, making it a critical element for life. Because most marine microorganisms preferentially take up P as inorganic phosphate (PO_4^{3-}), concentrations of dissolved PO_4^{3-} are low in most of the surface ocean (<1 μM; ref. 1) thereby limiting phytoplankton productivity (2).

Marine microbes respond to chronic P-deficiency by investing resources in organic and inorganic P uptake (3-7), remodeling cellular metabolism and structures (8-13), and storing P (14,15). Different mechanisms for enacting these responses have evolved in widely distributed and often sympatric microbial taxa. For example, within the diatom, *Phaeodactylum tricornutum*, proteomics indicated broadly depressed metabolic activity as a response to P-starvation, with down-regulated energy metabolism, amino acid metabolism, nucleic acid metabolism, and photosynthesis, and up-regulated protein degradation, lipid accumulation, and photorespiration (16). In contrast, the prymnesiophytes *Prymnesium parvum* and *Emiliania huxleyi*, and the dinoflagellates *Prorocentrum donghaiense* and *Amphidinium carterae*, maintained energy-generating processes (i.e., photosynthesis) and carbon fixation in response to P-deficiency (17-20). Interestingly, *P. donghaiense* increased nitrate assimilation under P-deficiency (20), whereas *E. huxleyi* reduced nitrate uptake with a tight coupling between P and nitrogen (N) pools (19). These varied strategies all minimize non-critical P utilization and maximize P uptake, and thus play important roles in structuring phytoplankton assemblages and P biogeochemical cycling in the oceans (11, 19, 21-23).

Picoeukaryotes, including the abundant and cosmopolitan groups *Micromonas* and *Ostreococcus* spp., are members of the "green" lineage (Chlorophyta) and are important marine primary producers (24-26). Recent culture experiments with *Micromonas pusilla* indicated that this group is likely to be successful under conditions of increased acidification and low nutrient (P) availability, yet the mechanisms behind their success are elusive (27). Similar to other phytoplankton (4), *M. pusilla* may remodel lipids to increase non-P containing lipids (28) and up-regulate genes for P-transporters and polyphosphate accumulation (29). However, the broader physiological response of *M. pusilla* and other prasinophytes to P-deficiency has not been fully explored.

The green algae, Chlorophyta, share a more recent common ancestor with land plants than other marine algal groups such as diatoms and haptophytes (30,31). Consequently, their physiological response to P-deficiency may share more traits with land plants than with other algal taxa. Indeed, P-deficiency is common in lakes and terrestrial systems (32). The physiological response of model chlorophytes such as the mustard plant, *Arabidopsis thaliana*, and the freshwater green alga, *Chlamydomonas reinhardtii*, to P-deficiency led to the identification of a phosphate starvation response gene described as *psr*1 in *C. reinhardtii* (33) and as *phr*1 in plants (34). This gene encodes a transcription factor (TF), a protein that binds to specific DNA regions to activate or repress transcription of one or more genes. The specific region of DNA to which TFs bind is typically characterized by a short repetitive nucleotide sequence, or motif, found up- or downstream of a given gene or within introns (35,36). In *A. thaliana*, regulatory motifs for the Phr1 protein were more abundant in known P-responsive genes than in the rest of the genome (37), linking Phr1 to the regulation of P-responsive genes. Phr1/Psr1 TFs have not been described, to our knowledge, in marine algae.

Here, we investigate the impact of P-deficiency on the physiology of the ecologically relevant picoeukaryote, *Micromonas pusilla* CCMP1545. Many metabolites are produced by metabolic pathways under genetic regulatory control; thus, our combined metabolomics and genomics approach provides mechanistic insight into the physiological response of the organism to P-deficiency (e.g., 38, 39) and its underlying genetic regulation. We used a targeted metabolomics approach to analyze the suite of molecules produced by *M. pusilla* within the cells and in the external medium. In combination, we employed a comparative genetics approach to identify 1) a *psr*1-like gene in *M. pusilla* and other abundant marine phytoplankton species and 2) a potential regulatory element that may be the DNA-binding site of the Psr1-like protein in putatively P-responsive genes. We found evidence for the expression of *psr*1-like genes in *M. pusilla* and other marine algae under P-deficient conditions in cultures and in the field (i.e., *Tara* Oceans metatranscriptomes). These results have implications for better understanding and predicting the metabolic response of diverse phytoplankton groups to P-deficiency.

Materials and Methods

Culture of Micromonas pusilla *CCMP 1545*

All glassware was acid-cleaned and combusted at 450°C for at least 4 h. We grew an axenic culture of *Micromonas pusilla* CCMP1545 from the National Center for Marine Algae and Microbiota (Boothbay, ME, USA) in L1-Si media (https://ncma.bigelow.org/algal-recipes) with 0.2-μm filtered (Omnipore filters; Millipore, MA, USA) autoclaved seawater from Vineyard Sound (MA, USA). We maintained the cultures at 22ºC under a 12:12 light:dark regime (84 μmol m^{-2} s^{-1}). After at least three generations, we split the culture into two parallel cultures of L1-Si media amended with

(a) 36 μM phosphate (P-replete) and (b) 0.4 μM phosphate (P-deficient). We completed three transfers of these parallel cultures prior to this experiment (Fig S1 a,b). For each media type, we inoculated 9 flasks of media with exponential-phase cells to achieve 320 ml and ~300,000 cells in each flask (33 ml P-replete; 37 ml P-deficient), with three cell-free control flasks. We grew cultures for two weeks (Fig S2a, b, c) and removed samples approximately 1 hr into the light cycle at the time of experimental setup (T_0; day 0), in rapid growth phase (T_1; day 4 P-deficient; day 6 P-replete), and in stationary phase (T_2; day 5 P-deficient; day 13 P-replete).

At each sampling point, we removed 1 ml for cell counts and chlorophyll fluorescence, 30 ml for total organic carbon (TOC; Methods S1), and 20 ml of filtrate (see below) for nutrients. We monitored cell abundance daily by flow cytometry (Guava, Millipore) and assessed photosystem II efficiency by measuring the variable and maximum chlorophyll fluorescence (F_v/F_m) using fluorescence induction and relaxation (FIRe; Satlantic LP, Halifax, Cananda). We used chlorophyll *a* (692 nm) and side scatter of *M. pusilla* cultures to optimize settings for flow cytometry analysis. We assessed potential bacterial contamination by viewing DAPI-stained cells of each time point (Methods S1).

Metabolite extraction and instrument methods

We processed cultures for intracellular and extracellular metabolite extraction as described previously (ref. 40; Methods S1). For intracellular metabolites, we collected cells via filtration on 0.2 μm PTFE filters (Omnipore, Millipore) and stored them at -80°C. Metabolites were extracted in 500 μl of cold 40:40:20 acetonitrile:methanol:water with 0.1 M formic acid. Extracts were reduced to near dryness under vacuum centrifugation

and reconstituted in 500 μl of 95:5 MilliQ water: acetonitrile with deuterated biotin (final concentration 0.05 μg ml^{-1}) added as a high-performance liquid chromatography (HPLC) injection standard; 100 μl of the extract was used for targeted metabolomics analysis. We extracted extracellular metabolites using solid phase extraction (SPE) with 1 g/6 cc PPL cartridges (BondElut, Agilent, Santa Clara, CA, USA) as described previously (41) and modified by Longnecker (42) (Methods S1). Targeted metabolomics analysis was performed as described in Kido Soule *et al*. (43). All metabolomics data are available in the MetaboLights database (http://www.ebi.ac.uk/metabolights/) with accession number MTBLS295.

R statistical software (v3.0.2; R Core Team) was used for statistical analyses. We used a Welch's two sample *t*-test to compare the cell-normalized concentrations of specific metabolites (log transformed) between P-deficient and P-replete treatments.

*Characterization of the phosphate starvation response (*psr*1) gene and DNA-binding motifs*

We used the *psr*1 gene sequence from *C. reinhardtii* (ref. 33, NCBI XM_001700501.1) to query the genome of *M. pusilla* using the BLAST option through Integrated Microbial Genomes (IMG, blastx, e-value: $1e^{-5}$). Conserved domains within putative sequences were characterized using NCBI conserved domain search (CD-search). We then used the putative *M. pusilla psr*1-derived amino acid sequence to query several databases using BLAST (tblastn, e-value: $1e^{-5}$), including: a re-assembly of the Marine Microbial Eukaryote Transcriptional Sequencing Project (MMETSP) (44), NCBI non-redundant protein sequences (nr), the NCBI metagenome proteins (env nr), the *Tara* Oceans eukaryotic unigenes (45), and several eukaryotic algal genomes from IMG

(Methods S1). Visualization of the *Tara* Oceans queries were performed using the basemap library in Python and custom Python scripts (Methods S1). Inspection of sequences were performed with multiple sequences alignment with *C. reinhardtii psr*1-derived amino acid sequence. Sequence alignment and phylogenetic analysis was performed in MEGA 5.2.1 (46) using MUSCLE (47) with default settings in MEGA (Methods S1).

We searched for a regulatory binding element, or motif, for the Psr1-like protein within a now-archived gene model of *M. pusilla* CCMP1545 (FrozenGeneCatalog_20080206 (ver 1), Table S1), then again within the updated gene model from Bachy *et al*. (29) (Table 1). Specifically, we analyzed gene sequences of interest for a regulatory motif using the motif-based sequence analysis tools Multiple Em for Motif Elicitation (MEME Suite) 4.10.1 (48). The genes of interest (Table 1) were selected based on metabolites that were elevated or depressed in concentration in the P-deficient cultures although not necessarily statistically different between treatments (i.e., malate, several amino acids, vitamins, Fig 2) and gene expression information from Bachy *et al*. (29). These included genes related to central carbon metabolism, lipid metabolism, vitamin metabolism (pantothenic acid, folic acid), and nucleotide metabolism (aspartic acid). We also included genes within the carbohydrate and lipid metabolic pathways based on published work (49, 50) as well as the gene for proline oxidase (POX), a key enzyme up-regulated in *E. huxleyi* in response to N- and P-deficiency (19, 51). After initial motif discovery, we optimized our putative motif by varying discovery parameters such as motif width and number within genes of interest and within genes that were not expected to be regulated in response to P-deficiency, including asparagine synthetase (JGI gene ID: 9681548) and acetyltransferase like/FAD

linked oxidase (9687568)). We attempted to identify motifs using the motif comparison tool (Tomtom; ref. 52) against the *A. thaliana* database (Protein Binding Microarray (PBM) motifs, (36)) and significant matches were used to guide the optimization of motif discovery settings and inclusive sequences.

Using the selected sequences and optimized parameters, we searched for motifs that are overrepresented in the query sequences ("positive") relative to a set of background sequences ("negative"), the latter we defined using the genome scaffolds. These positive and negative sequences were used to create a position specific prior (PSP) distribution file (48). Discriminative motif discovery used the PSP file and the set of *M. pusilla* gene sequences of interest (parameters: minw = 5, maxw = 10, nmotifs = 3, mod = oops (one occurrence per sequence)). Each gene sequence consisted of concatenated 500 nucleotides upstream and downstream of the gene as well as untranslated regions (UTRs) and intronic regions, thereby excluding the coding regions.

Two recent publications provided transcriptomes for testing our hypotheses regarding expected DE genes in P-deficient and P-replete conditions. Whitney & Lomas (53) published transcriptomes of *M. commoda* (formerly *M. pusilla*) RCC299, while Bachy *et al*. (29) published transcriptomes of *M. pusilla* CCMP1545. We utilized these data to test our hypotheses based on results of our metabolomics analyses with *M. pusilla* CCMP1545 and *in silico* genomic analysis. We performed the same protocol for motif discovery and identification on gene sequences from *M. commoda* RCC299 (gene model: *Micromonas pusilla* NOUM17/FrozenGeneCatalog_20090404) and CCMP1545 (gene model: *Micromonas pusilla* CCMP1545/MBARI_models(ver_1)).

Results

203 *Metabolic response of* Micromonas pusilla *CCMP1545 to phosphorus deficiency*

204 *M. pusilla* grew in P-replete (0.45 day^{-1}) and P-deficient (0.03 day^{-1}) media (Fig
205 S2). Total organic carbon concentrations mirrored growth curves, increasing
206 substantially in the P-replete cultures, and increasing slightly in the P-deficient cultures
207 (Fig S2). Inorganic nutrient concentrations and photochemical efficiency generally
208 decreased over time in both treatments (Results S2).

209 Metabolites were generally less abundant under P-deficiency in the intracellular
210 (Fig 1) and extracellular (Fig S3) fractions, although several intracellular metabolites
211 were most abundant during rapid growth (T_1) under P-deficiency, including nucleobases,
212 amino acids, carbohydrates, and vitamins (Fig 1). Only a few intracellular metabolites
213 showed significantly different concentrations between treatments: (a) the nucleobase
214 xanthine, which was significantly elevated under P-deficiency (*t*-test, $p = 0.01$), and (b)
215 the amino acid derivatives, *N*-acetylglutamate and tryptamine, which were significantly
216 higher in P-replete samples (*t*-test, $p = 0.01$, $p = 0.02$, respectively) (Fig 1). We observed
217 significant decreases in the intracellular ratio of purine nucleosides to their nucleobases
218 under P-deficiency at T_1 (Fig 2). Intracellular malate, succinate, and citrate exhibited
219 divergent responses, although their abundances were not significantly different between
220 treatments (Fig 1). Malate was more abundant in the P-deficient cells, while succinate
221 concentrations were similar between treatments and replicates and citrate was detected in
222 only one of three replicates in each treatment (Fig 1). We noted similarly varied
223 responses to P-deficiency in the purine nucleosides xanthosine and inosine, with invariant
224 xanthosine abundances but higher average concentrations and variability of inosine (Fig
225 1).

Metabolite concentrations are, in part, determined by genetic regulation of the enzymes producing or degrading them and we hypothesized that the varied metabolite responses could provide insight into the regulation of these genes. The description of a phosphate starvation response (*psr*1) gene containing a myb DNA-binding domain, led us to investigate the presence of *psr*1, and a myb-like motif in genes that could be regulated by the *psr*1 gene product in *M. pusilla* and other marine algae.

*Characterization of the phosphate starvation response (*psr*1) gene in marine algae*

We found four statistically-significant sequences (Table 2) in response to querying the now-archived *M. pusilla* CCMP1545 genome with the *psr*1 gene from *C. reinhardtii* (strain cc125, NCBI accession AF174532). Only one of these sequences (JGI ID 61323) had the two characteristic myb domains of the *psr*1 gene in *C. reinhardtii* (myb-like DNA-binding domain, SHAQKYF class (TIGR01557) and myb predicted coiled-coil type transfactor, LHEQLE motif (pfam PF14379)) (Fig 3a). The glutamine-rich regions, characteristic of TF activation domains, and putative metal binding domain originally described for *psr*1 in *C. reinhardtii* (33) were not detected in the putative *M. pusilla psr*1 gene, thus we refer to this as a *psr*1-like gene. Conserved domain analysis (54) revealed two other domains in the putative *psr*1-like-derived amino acid sequence from *M. pusilla*: 1) PLN03162 super family (NCBI cl26028), a provisional golden-2 like transcription factor, and 2) a DUF390 super family domain (NCBI cl25642), comprising proteins of unknown function.

We identified putative *psr*1-like genes or transcripts in field datasets (Fig 3b, 4, 5, Table S2), representing diverse taxa such as Dinophyta (e.g., the symbiotic clade C *Symbiodinium* sp.), and Haptophyta (e.g., the coccolithophore *E. huxleyi).* These

representative sequences did not necessarily cluster by taxonomic origin (e.g., cluster III, Fig 4). Transcripts from MMETSP and *Tara* Oceans were generally short sequences and often contained one of the two characteristic domains for *psr*1-derived amino acid sequences (Results S2). Field-derived transcripts of the *psr*1-like gene were geographically dispersed in the *Tara* Oceans dataset with the highest relative expression in surface samples and in the North Atlantic Ocean and Mediterranean Sea (Fig 5a). The majority of transcripts from the *Tara* Oceans dataset occurred in samples with low phosphate concentrations (≤ 0.5 µM) but there was no significant correlation between *psr*1 expression and phosphate concentration (Fig 5b). We observed a significant positive relationship between the presence of *Micromonas psr*1-like genes and the relative expression of one of the TCA cycle genes in *Micromonas,* fumarase, which were up-regulated in both culture-based studies of *M. pusilla* (29) and *M. commoda* (53) under P-deficiency (Fig 5c, Discussion S3). We never found genes with homology to the *M. pusilla* putative *psr*1-like gene, in genomes or transcriptomes within diatoms, the Heterokonta superphylum or the cryptophytes.

Putative regulatory element in Micromonas pusilla *genes of interest*

MEME analysis yielded a significant motif in *M. pusilla* genes involved in central carbon metabolism, lipid metabolism, and nucleotide metabolism (Fig S4). The conserved motif is similar in nucleotide sequence to several TF-binding sites in *A. thaliana* (Table S3, Fig S5; 36), including myb-like TF-binding motifs. Several iterations of this analysis using genes not tied to the metabolites of interest yielded either no significant motif or motif sequences lacking similarity to myb-like domains. We observed a similar conserved motif in DE transcripts in *M. pusilla* CCMP1545 under P-deficiency

(29) (Table 1, Fig S4). While the actual nucleotide sequence recognized by a TF must be experimentally determined (36), the presence of a significant motif in each model and only in the expected genes, indicates that this motif is likely biologically significant in *M. pusilla*. Specifically, this motif, with similarity to myb-like domains in *A. thaliana*, may represent a regulatory element where the Psr*1*-like protein binds.

Comparison of metabolomics-based predictions of P-responsive genes in M. pusilla *CCMP1545 to transcriptomics analysis*

We queried the DE transcripts in *M. pusilla* CCMP1545 (29) and compared these to our gene predictions based on our metabolomics data. As Bachy and colleagues (29) used a more recent gene model than our original analysis, we confirmed the identity of the *psr*1-like gene in the updated gene model (Fig S5). We analyzed a similar set of genes in search of a regulatory element where the Psr1-like TF might bind (Fig S4), including DE genes in the P-deficient transcriptome relative to the replete (Table S1 from 29). Under P-deficiency, the *psr*1-like gene (JGI gene ID 360) was highly up-regulated (log-2 fold change = 4.3) and nearly all of the predicted P-responsive genes were DE (Table 1). In contrast to the *M. commoda* P-deficient transcriptome, *M. pusilla* exhibited many DE genes in the tricarboxylic acid (TCA) cycle (Results S2) but not pyruvate carboxylase (gene ID 6984). Additionally, between *M. pusilla* and *M. commoda*, different genes involved in nucleotide and lipid metabolism were DE and contained a conserved motif with similarity to myb-like regulatory element in *A. thaliana* (Fig 6). Lastly, the POX gene observed in *M. commoda* and in the archived gene model for *M. pusilla*, is not present, or at least not similarly annotated in the updated gene model. Instead, a copper

amine oxidase (gene ID 5863) contained a significant conserved motif (Table 1, Figure S4) and is DE under P-deficiency.

Comparison to Micromonas commoda *RCC299 gene expression under P-deficiency*

We queried the DE genes in *M. commoda* RCC299 (26) exposed to P-deficient conditions (53) for the *psr*1-like gene as well as for genes that we expected to be regulated by the Psr1-like TF based on our *M. pusilla* observations and the literature (e.g., 19, 55, 56). In *M. commoda*, the *psr*1-like gene (JGI gene ID: 60184) was significantly up-regulated in the P-deficient treatment (Table S4), and we found a significant conserved motif in 18 DE genes in *M. commoda*. The significant motif identified in *M. commoda* differed in sequence from that discovered in *M. pusilla* and occurred in a different set of genes. However, putative binding motifs for Psr1 are present in the same pathways in the two *Micromonas* species and are similar to other DNA-binding motifs in *A. thaliana* (Fig S6 Results S2; ref. 36). The *psr*1-like genes in RCC299 and CCMP1545 exhibited relatively low homology to each other but were equally similar to the *psr*1-derived amino acid sequence in *C. reinhardtii* (Table S5).

Discussion

Metabolomics and comparative genomic analysis of the response of *M. pusilla* CCMP1545 to P-deficiency has led to three conclusions: 1) there is a shift in intracellular metabolite composition, 2) a *psr*1-like gene is expressed in *M. pusilla* and other marine phytoplankton, and 3) the genes regulated by the Psr1-like protein appear to differ amongst algal species.

317 Micromonas pusilla *exhibits a metabolic shift in response to P-deficiency*

318 Most intracellular metabolites did not differ significantly in concentration
319 between treatments. Because these metabolites represent many central metabolic
320 pathways, it appears that *M. pusilla* does not respond to P-deficiency through a global
321 decrease in metabolic activity, as was observed for a diatom (16). Interestingly, we found
322 examples of functionally-related metabolites that behaved differently. For example,
323 glutamate, proline, and methionine were variable but higher on average under P-
324 deficiency, while phenylalanine and tryptophan were not. In the TCA cycle, malate was
325 variable but higher in concentration on average under P-deficiency, while citrate was
326 lower and succinate was relatively unchanged. The opposing responses of these TCA
327 intermediates was surprising as others (e.g., 49, 57) have shown that transcripts or
328 proteins involved in the TCA cycle tend to exhibit a coordinated response to nutrient
329 limitation. Additional non-uniform metabolite responses occurred, with vitamins and
330 glucosamine-6-phosphate higher on average under P-deficiency, but other P-containing
331 compounds (e.g. glycerol-3-phosphate) and sugars were depleted. Previous work has
332 suggested that organisms placed under nutrient stress are capable of diverting carbon flux
333 between pathways (reviewed in 39). Changes in carbon flux through the cell could result
334 in coordinated abundance shifts in seemingly unrelated metabolites, as observed in this
335 study. If the shunting of carbon is a coordinated response within *M. pusilla*, then we
336 would expect genes involved in the synthesis and/or catabolism of these compounds to be
337 co-regulated, and potentially mediated by a specific transcription factor.

338 *Presence and expression of* psr*1 in marine algae*

339 We detected a putative *psr*1-like gene in the genome of *M. pusilla* and in the
340 genomes and transcriptomes of major algal lineages (e.g., prasinophytes, dinoflagellates).

A gene annotated as *psr*1 is present in *Ostreococcus tauri* (58), although to our knowledge the potential role of this gene in the metabolism of *O. tauri* or other marine algae has not been discussed. The *M. pusilla psr*1-like gene contains two myb domains similar to those in *psr*1 from *C. reinhardtii* (33), and comparable to those in *phr*1 in *A. thaliana* (34). No *psr*1-like gene was identified in queried diatom transcriptomes or genomes, suggesting that a phosphate starvation response gene akin to *psr*1 is present in diverse but phylogenetically constrained phytoplankton groups.

We found significant similarity between the *M. pusilla psr*1-like gene and transcript sequences from the MMETSP and *Tara* Oceans datasets, indicating that this gene is expressed by marine algae *in situ*. High-identity hits to sequences in the GOS and *Tara* Oceans datasets underscore this gene's prevalence in the oceans, particularly in regions characterized by chronically low phosphorus concentrations. Although some MMETSP and *Tara* Oceans sequences contained only one (of two) characteristic Psr1 domains, these short sequences did not include enough of the C-terminal end to capture the myb coiled-coil domain. Thus, further investigation is required to confirm whether the identified transcripts are from *psr*1-like genes. Several GOS sequences, previously described as protein of unknown function, had homology to both characteristic Psr1 domains, suggesting that they should be re-annotated as *psr*1-like genes. The GOS sequences are likely derived from prasinophytes (i.e., *Micromonas* spp., *Ostreococcus* spp.) and haptophytes (e.g., *E. huxleyi*), highlighting the need to elucidate the role of *psr*1 in these organisms.

Potential role of Psr1 in the marine algal response to P-deficiency

We observed a conserved and enriched motif that may function as a regulatory element across genes involved in nucleotide biosynthesis, the TCA cycle and glycolysis, carbon fixation, fatty acid metabolism, and phosphate transport or salvage. The conserved motifs discovered in *M. pusilla* and *M. commoda* had unique sequences and were present in a slightly different set of genes (Fig 6). The motif detected in *M. pusilla* is similar in nucleotide sequence to a myb DNA-binding domain in *A. thaliana*. This similarity is potentially significant, because the *psr*1/*phr*1 genes in *C. reinhardtii* and *A. thaliana* contain a myb DNA-binding domain. Thus, we hypothesize that the conserved motif detected in the *Micromonas* species represents binding sites for the Psr1-like transcription factors (TF).

We observed the conserved motif only in DE genes under P-deficiency by each species (29, 53), suggesting that these genes could be co-regulated. In bacteria, genes within the well-characterized Pho-regulon, for example, all contain specific sequences (the PHO box) where the transcription factor binds to activate or repress the gene (59). Different bacterial species and strains contain distinct PHO box sequences. In *M. commoda*, the discovered motif was not significantly similar to myb-like DNA-binding motifs in *A. thaliana* and was not similar to the motif discovered in *M. pusilla*. This was surprising at first, as we expected that the myb domain of the Psr1-like proteins in the two species would interact with similar binding regions in the genome. However, *psr*1-like derived amino acid sequences from *M. pusilla* and *M. commoda* were only 47% similar. TF protein sequences with up to 79% similarity have been shown to have distinct DNA-binding motif profiles (36), so these two proteins could reasonably have distinct DNA-binding motifs. Moreover, as the conserved motif was present in different sets of genes in each *Micromonas* species, it is likely that the Psr1-like protein regulates

388 different genes in different taxa and may serve as a niche defining feature between these
389 taxa. The observed variability in the *psr*1-like gene and the associated binding motifs
390 may be an adaptation to distinct environments as *M. pusilla* was collected in the
391 temperate English Channel, while *M. commoda* was isolated from the tropical Atlantic
392 (26).

393 The transcriptomes of *M. commoda* (53) and *M. pusilla* (29) provided a means to
394 test our hypotheses regarding the up- or down-regulation of certain genes in response to
395 P-deficiency. The *psr*1-like gene in *M. commoda* and *M. pusilla* was one of the highest
396 DE genes in each experiment (29, 53), suggesting that this potential TF plays a critical
397 role in the response of *Micromonas* to P-deficiency. Based on previous studies, we
398 expected genes involved in triacylglycerol (TAG) production to increase under P-
399 deficiency, but found divergent responses between the two species. Two TAG production
400 genes were up-regulated in *M. commoda*, including the penultimate step in TAG
401 production (putative phosphatidic acid phosphatase), which contained a putative
402 regulatory element where the Psr1-like TF could bind. By contrast, neither of the two
403 TAG metabolism genes that were DE in *M. commoda* were DE in *M. pusilla*; instead two
404 different genes involved in fatty acid biosynthesis and metabolism were DE in *M. pusilla*.
405 Additionally, a starch-binding protein was highly up-regulated in the *M. pusilla*
406 transcriptome under P-deficiency (gene ID 9633, ref. 29). These observations suggest that
407 *M. pusilla* may invest resources in starch storage while *M. commoda* may invest in lipid
408 storage, a testable hypothesis for future work.

409 Variable and non-uniform dynamics of several TCA cycle metabolites between
410 treatments suggested that this pathway may be involved in the metabolic response to P-
411 deficiency. Indeed, nearly all of the genes involved in the end of glycolysis and in the

412 TCA cycle were up-regulated in *M. pusilla* and contained the conserved motif that may
413 function as a regulatory element where the Psr1-like TF could bind. We propose that
414 carbon flow through the TCA cycle may increase as a means to fuel TAG or starch
415 production and to pull potentially damaging energy away from the photosystems (60, 61).
416 Further work is needed to test this hypothesis and to explore the physiological
417 implications for *Micromonas* spp. *in situ*. In contrast, only a few genes involved in the
418 TCA cycle were up-regulated in *M. commoda*, including genes involved in malate and
419 oxaloacetate (OAA) metabolism and a gene for pyruvate carboxylase (PC) which
420 converts pyruvate to oxaloacetate. The expression of these TCA-cycle genes may reflect
421 an increase in the malate-aspartate shuttle, a redox reaction that drives the production of
422 NAD^+ from NADH in the mitochondrial membrane. In *M. commoda*, it is possible that
423 PC-based conversion of pyruvate to OAA combined with an increase in the malate-
424 aspartate shuttle function as a stress response that allows continued glycolysis and ATP
425 production *via* oxidative phosphorylation (62, 63). In this scenario, carbon flow into the
426 TCA cycle would be limited and the genes controlling malate, aspartate, oxaloacetate and
427 pyruvate would be co-regulated. This hypothesis is supported by the observation of a
428 conserved motif that may function as a regulatory element in the relevant genes and by
429 preliminary gene expression analysis of one of the TCA cycle genes that was up-
430 regulated in both *M. pusilla* and *M. commoda* transcriptomes under P-deficiency, the
431 fumarase gene. If fumarase expression is regulated by the Psr1-like protein, expression of
432 these two genes (fumarase and *psr*1) should be correlated in field populations. We
433 observed support for this hypothesis within the *Tara* Oceans dataset, although we observe
434 weaker relationships between Psr1-like gene expression and other genes that were up-
435 regulated in either *M. pusilla* or *M. commoda* only (Discussion S3).

436 We observed shifts for several purine nucleosides between treatments for *M.*
437 *pusilla* and detected the conserved motif in the genes for a nucleoside phosphatase and
438 aspartate transcarbamylase. These results suggest increased nucleotide salvage through
439 purine nucleotides and nucleosides, a mechanism described in *M. commoda* (53) and in
440 other phytoplankton (13, 64). Interestingly, while a 5'-nucleotidase was observed to be
441 DE in *M. commoda* (53) under P-deficiency, it was not DE in *M. pusilla* (29). In the *M.*
442 *pusilla* transcriptome (29), a nucleotide phosphatase was up-regulated and may function
443 in a similar capacity in nucleotide salvage.
444 Another point of contrast between *M. pusilla* and *M. commoda* is the presence and
445 expression of a proline oxidase (POX) gene. The POX gene was highlighted by Rokitta *et*
446 *al*. (19,51) in nutrient limitation experiments (including P-deficiency) with the
447 haptophyte *E. huxleyi* where this enzyme may have a role in stabilizing the mitochondrial
448 membrane and in detecting cellular nitrogen levels. The POX gene was up-regulated in *M.*
449 *commoda* but not present in *M. pusilla*. By contrast, a copper amine oxidase was up-
450 regulated in *M. pusilla*. The POX gene in *M. commoda* and the copper amine oxidase
451 gene in *M. pusilla* contained the conserved motif that may function as a regulatory
452 element for the Psr1-like TF. The role of these genes in nutrient limitation should be
453 targets for further investigations (Discussion S3).
454 The potential impact of Psr1-regulated genes on the ecological roles of
455 *Micromonas* spp. is unknown, but our results suggest that the Psr1-like protein
456 coordinates a metabolic shift in these organisms under P-deficiency, altering the
457 intracellular flow of carbon and other elements. More comprehensive examination of
458 these metabolic responses, which likely vary to some extent among these organisms, will
459 be paramount to improving models of trophic carbon flow. More experiments are needed

460 to characterize the structure and role of the *psr*1-like gene in *Micromonas* spp. and other
461 phytoplankton, including: (1) confirming the presence of the *psr*1-like gene with genetic
462 experiments, (2) determining whether the Psr1-like protein is more abundant under P-
463 deficiency, (3) identifying the taxon-specific genes affected by the Psr1-like protein, (4)
464 verifying the interaction of the Psr1-like protein with the hypothesized binding sites, and
465 (5) comparing the genetic and metabolic responses to P-deficiency between organisms
466 containing the *psr*1-like and those that do not. Exploring the underlying biology of the
467 *psr*1 gene will facilitate mechanistic understanding of the complex metabolic response of
468 these organisms to P-limitation and will enhance our ecological and biogeochemical
469 predictions.

470

471

472 **Acknowledgements**

473 The authors would like to thank Krista Longnecker for performing the TOC analysis and
474 Matthew Johnson and Elizabeth Harvey for assistance with flow cytometry and FIRe
475 analysis. This research was funded by the Gordon and Betty Moore Foundation through
476 Grant GBMF3304 to E. Kujawinski.

477

478 **References**

479 1. Karl DM. Microbially mediated transformations of phosphorus in the sea: new views
480 of an old cycle. *Ann Rev Mar Sci.* 2014;6:279-337.

481 2. Tyrrell T. The relative influences of nitrogen and phosphorus on oceanic primary
482 production. *Nature* 1999;400:525–531.

483 3. Björkman KM, Karl DM. Bioavailability of dissolved organic phosphorus in the
484 euphotic zone at Station ALOHA, North Pacific Subtropical Gyre. *Limnol Oceanogr.*
485 2003;48:1049–1057.

486 4. Chung C-C, Hwang S-PL, Chang J. Identification of a high-affinity phosphate
487 transporter gene in a prasinophyte alga, *Tetraselmis chui*, and its expression under

488 nutrient limitation. *Appl Environ Microbiol.* 2003;69:754-759.

489 5. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy
490 M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J,
491 Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gunnar G,
492 Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R. Substrate-controlled
493 succession of marine bacterioplankton populations induced by a phytoplankton bloom.
494 *Science* 2012;336:608-611.

495 6. Dyhrman ST. Nutrients and their acquisition: phosphorus physiology in microalgae. In:
496 Borowitzka MA, Beardall J, and Raven JA (eds). *The Physiology of Microalgae*.
497 Heidelberg, Germany: Springer Publishing, 2016, pp155-183.

498 7. Lin S, Litaker RW, Sundra WG. 2016. Phosphorus physiological ecology and
499 molecular mechanisms in marine phytoplankton. *J Phycol.* 2016;52:10-36.

500 8. Berdalet E, Latasa M, Estrada M. Effects of nitrogen and phosphorus starvation on
501 nucleic acid and protein content of Heterocapsa sp. *J Plank Res.* 1994;16: 303–316.

502 9. Van Mooy BAS, Rocap G, Fredricks HF, Evans CT, Devol AH. Sulfolipids
503 dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine
504 environments. *Proc Nat Acad Sci.* 2006;103:8607-8612.

505 10. Van Mooy BAS, Fredricks HF, Pedler BE, Dyhrman ST, Karl DM, Koblížek M, *et al.*
506 Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity.
507 *Nature* 2009;457:69–72.

508 11. Martin P, Van Mooy BA, Heithoff A, Dyhrman ST. Phosphorus supply drives rapid
509 turnover of membrane phospholipids in the diatom *Thalassiosira pseudonana*. *ISME J.*
510 2011;5:1057–1060.

511 12. Shemi A, Schatz D, Fredricks HF, Van Mooy BAS, Porat Z, Vardi A. Phosphorus
512 starvation induces membrane remodeling and recycling in *Emiliania huxleyi*. *New Phytol.*
513 2016;211:886-898.

514 13. Kujawinski EB, Longnecker K, Alexander H, Dyhrman ST, Fiore CL, Haley ST, *et al.*
515 Phosphorus availability regulates intracellular nucleotides in marine eukaryotic
516 phytoplankton. *Limnol Oceanogr Lett.* 2017;2:119-129. doi:10.1002/lol2.10043.

517 14. Orchard ED, Ammerman JW, Lomas MW, Dyhrman ST. Dissolved and organic
518 phosphorus uptake in Trichodesmium and the microbial community: The importance of
519 phosphorus ester in the Sargasso Sea. *Limnol Oceanogr.* 2010;55:1390-1399.

520 15. Martin P, Dyhrman ST, Lomas MW, Poulton NJ, Van Mooy BAS. Accumulation
521 and enhanced cycling of polyphosphate by Sargasso Sea phytoplankton in response to
522 low phosphorus. *Proc Nat Acad Sci.* 2014;111:8089-8094.

523 16. Feng T-Y, Yan Z-K, Zheng J-W, Xie Y, Li D-W, Bala Murugan S, Yang W-D, Liu J-
524 S, Li H-Y. Examination of metabolic responses to phosphorus limitation via proteomic

525 analyses in the marine diatom *Phaeodactylum tricornutum*. *Sci Rep*. 2015;5:10373.

526 17. Lai J, Yu Z, Song X, Cao X, Han X. Responses of the growth and biochemical
527 composition of *Prorocentrum donghaiense* to different nitrogen and phosphorus
528 concentrations. *J Exp Mar Biol Ecol*. 2015;405:6-17.

529 18. Li M, Shi X, Guo C, Lin S. Phosphorus deficiency inhibits cell division but not
530 growth in the dinoflagellate *Amphidinium carterae*. *Front Microbiol*. 2016;7:826.

531 19. Rokitta SD, von Dassow P, Rost B, John U. P- and N-depletion trigger similar
532 cellular responses to promote senescence in eukaryotic phytoplankton. *Front Mar Sci*
533 2016;3:109.

534 20. Shi X, Lin X, Li L, Palenik B, Lin S. Transcriptomic and microRNAomic profiling
535 reveals multi-faceted mechanisms to cope with phosphate stress in a dinoflagellate. *ISME*
536 *J*. 2017;11:2209-2218.

537 21. Cavender-Bares KK, Karl DM, Chisholm SW. Nutrient gradients in the western
538 North Atlantic Ocean: Relationship to microbial community structure and comparison to
539 patterns in the Pacific Ocean. *Deep Sea Res Part I: Oceanogr Res Pap*. 2001;48:2373–
540 2395.

541 22. Dyhrman ST, Ruttenberg KC. Presence and regulation of alkaline phosphatase
542 activity in eukaryotic phytoplankton from the coastal ocean: Implications for dissolved
543 organic phosphorus remineralization. *Limnol Oceanogr*. 2006;51:1381–1390.

544 23. Dyhrman ST, Benitez-Nelson CR, Orchard ED, Haley ST, Pellechia PJ. A microbial
545 source of phosphonates in oligotrophic marine systems. *Nat Geosci*. 2009;2:696–699.

546 24. Throndsen J, Kristiansen S. Micromonas pusilla (Prasinophyceae) as part of pico-and
547 nanoplankton communities of the Barents Sea. *Polar res*. 1991;10:201–208.

548 25. Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N. A single species,
549 *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the
550 Western English Channel. *Appl Environ Microbiol*. 2004;70: 4064–4072.

551 26. Worden AZ. Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat*
552 *Microb Ecol*. 2006;43:165–175.

553 27. Maat DS, Crawfurd KJ, Timmermans KR, Brussaard CPD. Elevated CO_2 and
554 phosphate limitation favor *Micromonas pusilla* through stimulated growth and reduced
555 viral impact. *Appl Environ Microbiol*. 2014;80:3119–3127.

556 28. Maat DS, van Bleijswijk JDL, Witte HJ, Brussaard PD. Virus production in
557 phosphorus-limited *Micromonas pusilla* stimulated by a supply of naturally low
558 concentrations of different phosphorus sources, far into the lytic cycle. *FEMS Microbiol*
559 *Ecol*. 2016;92:fiw136

560 29. Bachy C, Charlesworth CJ, Chan AM, Finke JF, Wong C-H, Wei C-L, *et al*.

Transcriptional responses of the marine green alga *Micromonas pusilla* and an infecting prasinovirus under different phosphate conditions. *Environ Microbiol.* 2018;20:2898-2912.

30. Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, *et al.* The evolution of modern eukaryotic phytoplankton. *Science* 2004;16:354-360.

31. Lewis LA, McCourt RM. Green algae and the origin of land plants. *Amer J Bot.* 2004;91:1535–1556.

32. Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, *et al.* Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecoystems. *Ecol Lett.* 2007;20:1135-1142.

33. Wykoff DD, Grossman AR, Weeks DP, Usuda H, Shimogawara K. Psr1, a nuclear localized protein that regulates phosphorus metabolism in *Chlamydomonas. Proc Nat Acad Sci.* 1999;96:15336–15341.

34. Rubio V, Linhares F, Solano R, Martín AC, Iglesias J, Leyva A, *et al.* A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes Devel.* 2001;15: 2122–2133.

35. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Molec Life Sci.* 2012;69:3613-3634.

36. Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Nat Acad Sci.* 2014;111:2367–2372.

37. Müller R, Morant M, Jarmer H, Nilsson L, Nielsen TH. Genome-wide analysis of the *Arabidopsis* leaf transcriptome reveals interaction of phosphate and sugar metabolism. *Plant Physiol.* 2006;143:156–171.

38. de Carvalho CCCR, Fernandes P. Production of metabolites as bacterial responses to the marine environment. *Marine Drugs* 2010;3:705-727.

39. Markou G, Nerantzis E. Microalgae for high-value compounds and biofuels production: a review with focus on cultivation under stress conditions. *Biotechn Advan.* 2013;31:1532-1542.

40. Fiore CL, Longnecker K, Kido Soule MC, Kujawinski EB. Release of ecologically relevant metabolites by the cyanobacterium *Synechococcus elongatus* CCMP 1631. *Environ Microbiol.* 2015;17:3949–3963.

41. Dittmar T, Koch B, Hertkorn N, Kattner G. A simple and efficient method for solid-phase extracion of dissolved organic matter (SPE-DOM) from seawater. *Limn. Oceangr. Meth.* 2008;6: 230–235.

42. Longnecker K. Dissolved organic matter in newly formed sea ice and surface seawater. *Geochim Cosmochim Acta* 2015;171:39–49.

43. Kido Soule MCK, Longnecker K, Johnson WM, Kujawinski EB. Environmental Metabolomics: Analytical Strategies. *Mar Chem.* 2015;177:1–62.

44. Johnson LK, Alexander H, Brown CT. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. *Giga Sci*. bioRxiv p 323576 doi:10.1101/323576 (In review).

45. Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, *et al*. A global ocean atlas of eukaryotic genes. *Nature Comm.* 2017;9:373.

46. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molec. Biol. Evol.* 2011;28: 2731–2739.

47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res.* 2004;32:1792–1797.

48. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology,* AAAI Press, Menlo Park, California, 1994, pp. 28-36,

49. Hockin NL, Mock T, Mulholland F, Kopriva S, Malin G. The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* 2012;158:299-312.

50. Goncalves EC, Johnson JV, Rathinasabapathi B. Conversion of membrane lipid acyl groups to triacylglycerol and formation of lipid bodies upon nitrogen starvation in biofuel green algae *Chlorella* UTEX29. *Planta* 2013;238:895-906.

51. Rokitta SD, von Dassow P, Rost B, John U. *Emiliana huxleyi* endures N-limitation with an efficient metabolic budgeting and effective ATP synthesis. *BMC Genomics* 2014;15:1051.

52. Gupta S, Stamatoyannopolous JA, Bailey T, Noble WS. Quantifying similarity between motifs. *Gen Biol*. 2007;8:R24.

53. Whitney LP, Lomas MW. Growth on ATP elicits a P-stress response in the picoeukaryote *Micrmonas pusilla*. *PLoS ONE* 2016;11:e0155158.

54. Marchler-Bauer A, Bo Y, Han L, Lanczycki CJ, Lu S, Chitsaz F, *et al*. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucl Acids Res.* 2017;4:D200-D203.

55. Moseley JL, Chang C-W, Grossman AR. Genome-based approaches to understanding phosphorus deprivation responses and PSR1 control in *Chlamydomonas reinhardtii*. *Euk Cell* 2006;5:26–44.

634 56. Sharma KK, Schuhmann H, Schenk PM. High lipid induction in microalgae for
635 biodiesel production. *Energies* 2012;5:1532–1553.

636 57. Mock T, Samanta MP, Iverson V, Berthiaume C, Robison M, Holtermann K, *et al*.
637 Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana*
638 identifies genes involved in silicon bioprocesses. *Proc Nat Acad Sci*. 2008;105:1579-
639 1584.

640 58. Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, *et al.* Genome
641 analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique
642 features. *Proc Nat Acad Sci.* 2006;103:11647–11652.

643 59. Santos-Beneit F. The Pho regulon: a huge regulatory network in bacteria. *Front*
644 *Microbiol.* 2015;6:402.

645 60. Norici A, Bazonni AM, Pugnetti A, Raven JA, Giordano M. Impact of irradiance on
646 the C allocation in the coastal marine diatom *Skeletonema marinoi* Sarno and Zingone.
647 *Plant Cell Environ.* 2011;34:1666–1677.

648 61. Klok AJ, Martens DE, Wijffels RH, Lamers PP. Simultaneous growth and neutral
649 lipid accumulation in microalgae. *Biores technol.* 2013;134:233–243.

650 62. Yoshida K, Terashima I, Noguchi K. Up-regulation of mitochondrial alternative
651 oxidase concomitant with chloroplast over-reduction by excess light. *Plant Cell Physiol.*
652 2007;48:606-614.

653 63. Easlon E, Tsang F, Skinner C, Wang C, Lin S-J. The malate-aspartate NADH shuttle
654 components are novel metabolic longevity regulators required for calorie restriction-
655 mediated life span extension in yeast. *Genes Devel.* 2008;22:931-944.

656 64. Dyhrman ST, Palenik B. Characterization of ectoenzyme activity and phosphate-
657 regulated proteins in the coccolithophorid *Emiliania huxleyi*. *J Plank Res.* 2003;25:1215-
658 1225.

659

660

661

662

663

664

665

666 **Figure Legends**

Figure 1. Average ratio of the P-deficient to P-replete intracellular metabolite concentrations during rapid growth (T_1) in *Micromonas pusilla* CCMP1545. Concentrations are normalized to cell number in each treatment and means with one standard deviation are shown (N=3 unless otherwise noted). Metabolites are listed in order of descending concentration within three groups; those that were detected in enough replicates for standard deviation to be calculated, those where only one ratio could be calculated (black line), and those that were not detected (ND) in any of the three replicates for one treatment (R = P-replete, D = P-deficient). Metabolite names marked with an asterisk indicates a significant difference in concentration between treatments (*t*-test, $p < 0.05$). Only one replicate in the P-deficient treatment contained a non-zero concentration for *N*-acetylmuramic acid, thus while there was a significant difference in concentration for this metabolite, only one ratio could be calculated. (DMSP = dimethylsulfoniopropionate, NAD = nicotinamide adenine dinucleotide)

Figure 2. The intracellular ratios of two purine nucleosides to their nucleobases for *Micromonas pusilla* CCMP1545 in P-deficient (orange) and P-replete (blue) treatments during rapid growth (T_1). Xanthosine and xanthine, and guanosine and guanine were quantified in the targeted metabolomics method and normalized to cell abundance. The average ratio and high and low values are shown based on three replicate cultures. The asterisk marks a significant difference between the treatments (*t*-test, $p = 0.03$). The dashed line represents a ratio of 1. Two samples in the P-replete treatment had guanine levels below the limit of detection and the one ratio is shown as a line.

Figure 3. Predicted amino acid sequences of the *psr*1-like gene in *Micromonas pusilla* CCMP1545 and other organisms. The amino acid sequence (with IMG gene ID) is shown as described in the Joint Genome Institute Integrated Microbial Genomes *M. pusilla* CCMP1545 database. The myb-like DNA-binding and myb coiled-coiled domains are highlighted in blue (a). Predicted amino acid sequence alignment of *psr*1 and *psr*1-like genes in the region of the myb-like DNA-binding domain (SHAQKYF class) and LHEQL coiled-coil domain (b). Asterisks at the top of each column indicate 100% conserved residues across species surveyed. Numbers at the end of each row indicate position of the last shown residue in the amino acid sequence. Accession (MMETSP, NCBI, *Tara*) or gene identification (IMG) numbers are listed for each sequence.

Figure 4. Occurrence and phylogenetic relationship of *psr*1-like genes in marine phytoplankton. Maximum likelihood tree based on the derived amino acid sequences for *psr*1 and *psr*1-like genes from eukaryotic phytoplankton. The alignment for the phylogenetic tree is based on 44 amino acid positions. Bootstrap values above 50% from 500 replications are shown for each node. Accession or IMG gene identification numbers are shown with the taxon name and the star indicates *Micromonas pusilla* CCMP1545 cultured in this study. Four clusters are highlighted containing the Chlamydomonadales (I), prasinophytes (II), dinoflagellates and other taxa (III), and the haptophyte, *Emiliania huxleyi* and other taxa (IV). The outgroup used to root the tree was *Arabidopsis thaliana* (NC_194590). The scale indicates the number of amino acid substitutions per site.

Figure 5. Geographic distribution of *Micromonas psr*1-like genes and relationship with phosphate concentration in the *Tara* Oceans eukaryotic gene atlas (ref 45). The geographic distribution and relative abundance of *Micromonas psr*1-like genes in surface samples are depicted. Black dots represent stations where *Micromonas* was detected; the

size of purple circles represents the relative abundance of *psr*1-like gene, where the FPKM of the psr1-like gene was normalized to the total FPKM of all *Micromonas*-associated transcripts (a). Relative abundance of *Micromonas* is plotted against phosphate concentrations for surface stations where *Micromonas* was detected. Grey circles indicate samples where *psr*1-like genes were not detected, while samples where psr1-like genes were detected are colored based on the relative expression of *Micromonas psr*1-like gene, again normalized to total *Micromonas* transcript abundance (b). A split violin plot depicting the distribution of expression of *Micromonas* fumarase synthase genes as a function of presence or absence of *psr*1-like gene. Significant difference in the distribution of expression between the presence and absence is assessed with Kolmogorov-Smirnov test ($p<0.005$) (c).

Figure 6. Conceptual model of Psr1-like TF regulation in *Micromonas pusilla* CCMP154 (top half) and *M. commoda* RCC299 (bottom half) in three major metabolic pathways under phosphorus deficiency. The top half of the figure illustrates the relevant metabolites quantified in the present study and indicates if they are elevated (orange, ratio > 1 Figure 2) or depressed (blue, ratio < 1 Figure 2) in concentration in the P-deficient cells relative to the P-replete cells. Genes that contain the conserved motif with similarity to the myb-like regulatory element in *Arabidopsis thaliana* are marked with a red star. The orange and blue text refer to elevated or depressed transcript levels for each gene (ref 29). The same pattern applies to the bottom half of the figure with gene expression data from Whitney & Lomas (53); metabolite data is not available for this organism. Descriptions of putative metabolic responses in each pathway are shown and are based on data in the present study and data from the literature ([1]ref 63; [2]ref 56). Genes encode for: A) pyruvate kinase, B) pyruvate dehydrogenase, C) citrate synthase, D) aconitase, E) α-ketoglutarate dehydrogenase, F) succinate dehydrogenase, G) fumarase, H) malate dehydrogenase, I) pyruvate carboxylase, J) ribulose-1,5-bisphosphate carboxylase, K) triose sugar transporter, L) nucleotide phosphatase, M) 5'-nucleotidase, N) glycerol kinase, O) 1-acyl-sn-glycerol-3-phosphate acyltransferase, P) phosphatidate phosphatase, Q) triacylglycerol lipase, LPAT = 1-acylglycerol-3-phosphate O-acyltransferase

Tables

751 Table 1. Enzymes that contain a significant motif in the gene sequences of *M. pusilla* CCMP1545 (MBARI_models (ver 1)). Column 2: the
752 Enzyme Commission (E.C.) number or pfam identifier; Column 3: main pathway(s) in which the enzyme is involved; Columns 4-6: fold
753 change and *q*-value of the transcript in P-deficient treatment relative to P-replete conditions from Bachy *et al*. (29) and corresponding gene ID.

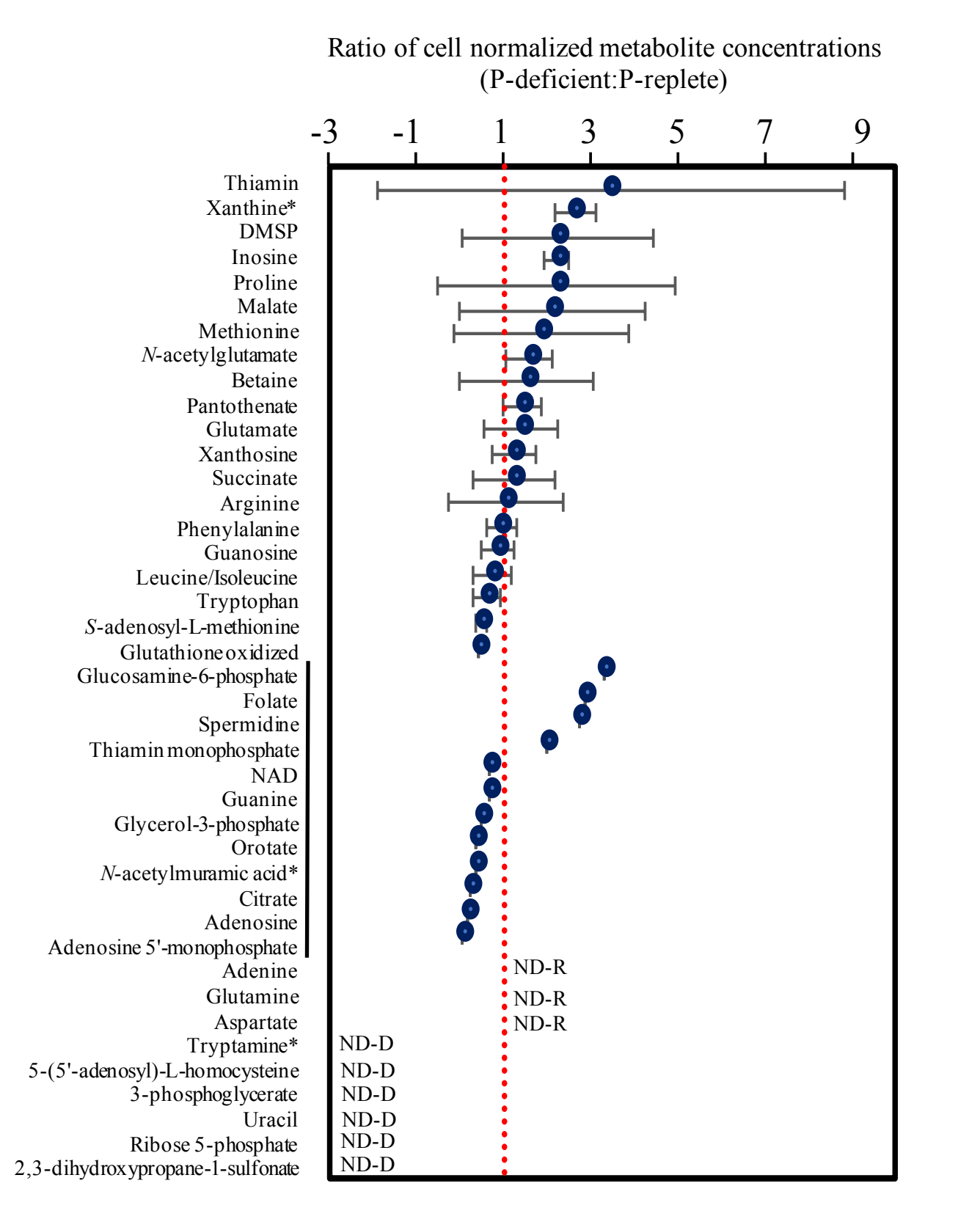
Enzyme	E.C. Number	Pathway	Log 2 fold change	*q*-value	JGI gene ID
Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	Calvin cycle/glycolysis	2.73	0.000037	4267
Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	Calvin cycle/glycolysis	-0.644	0.000037	700
Phosphoglycerate kinase	2.7.2.3	Glycolysis	-0.699	0.000037	7661
Pyruvate kinase	2.7.1.40	Glycolysis	-1.32	0.000037	9370
Pyruvate dehydrogenase	1.2.4.1	TCA[a] cycle/Glycolysis	0.756	0.000037	3535
Citrate synthase	2.3.3.1	TCA cycle	1.68	0.000037	1278
Citrate synthase	2.3.3.1	TCA cycle	1.10	0.000037	3714
Aconitase	4.2.1.3	TCA cycle	1.00	0.000037	2901
Succinate dehydrogenase	1.3.5.1	TCA cycle and ETC[b]	1.16	0.001515	2182
Fumarase	4.2.1.2	TCA cycle	1.68	0.000037	600
Fumarase	4.2.1.2	TCA cycle	1.73	0.000068	6662
Copper amine oxidase	1.4.3.21-22	Arginine and proline metabolism	2.87	0.000037	5863
Aspartate transcarbamylase	2.1.3.2	Pyrimidine biosynthesis	-0.609	0.000037	3615
Nucleoside phosphatase	3.6.1.3	Pyrimidine metabolism	0.711	0.000037	3322
Pantoate-beta-alanine ligase	6.3.2.1	Pantothenate and CoA biosynthesis	-0.616	0.000037	5304
Lysophospholipase	3.1.1.5	Glycerophospholipid metabolism	1.65	0.000037	3652
Acyl-CoA synthetase	6.2.1.3	Fatty acid metabolism	3.20	0.000037	1751
Long-chain acyl-CoA synthetases	6.2.1.3	Fatty acid metabolism	0.736	0.000037	416
$Na+/PO_4$ transporter	PF02690	inorganic nutrient transport	4.54	0.000037	5963

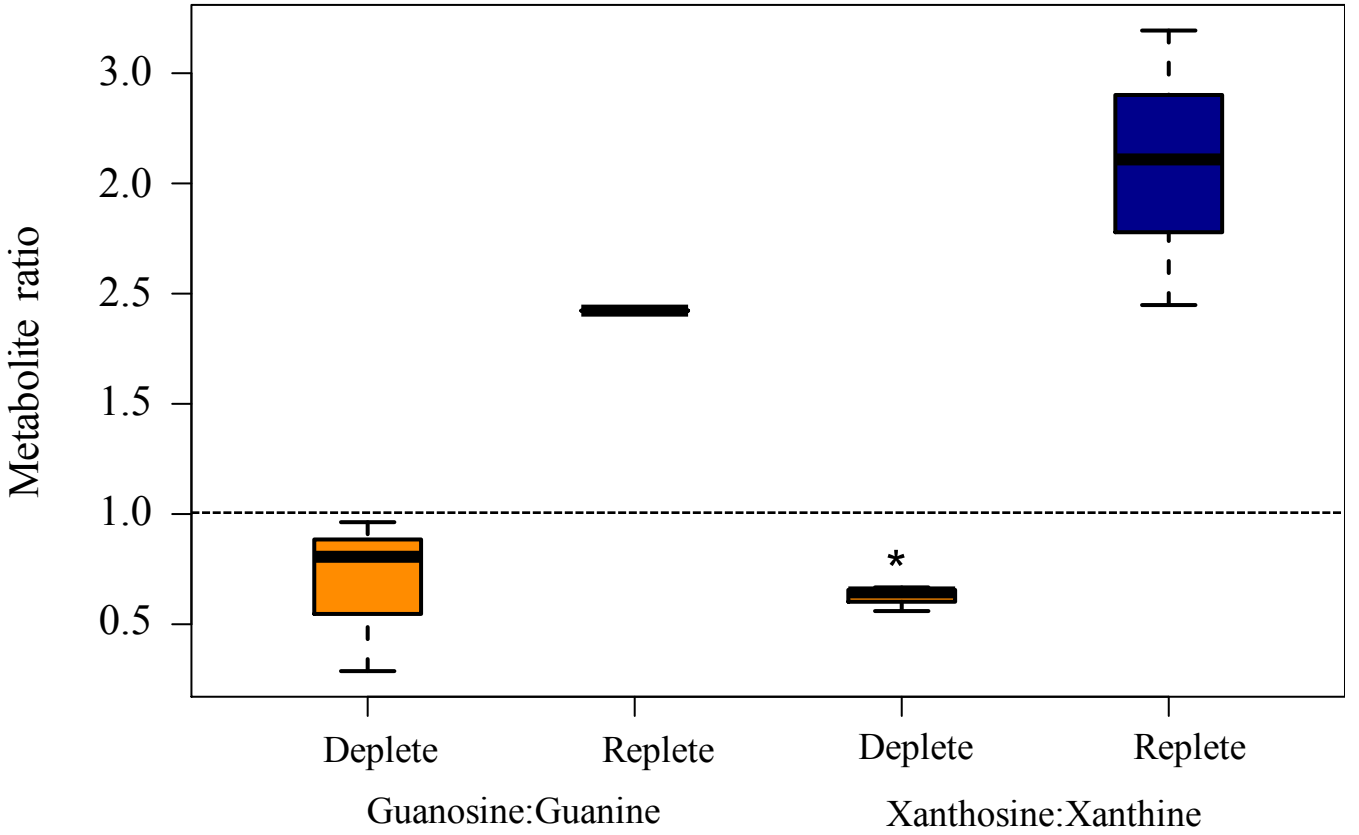
754 [a]Tricarboxylic acid cycle

[b]Electron transport chain
[c]Triacylglycerol

Table 2. Gene matches to *psr*1 in *Micromonas pusilla* CCMP1545 as originally discovered in the Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) database and the corresponding gene in the updated gene model for CCMP1545 (MBARI_models (ver 1)). The *M. pusilla* CCMP 1545 database in IMG contained 10,660 sequences and 4,795,637 letters. The top-scoring gene was considered to be a *psr*1-like gene and the JGI gene ID is shown in parentheses. The corresponding *psr*1-like gene in the updated gene model is shown for reference, and includes the JGI gene ID, the gene description, and the E-value from the BLAST search (see Methods) using the top scoring gene (613233) from CCMP1545.

IMG Gene ID	Locus tag	Bit score	E-value
2615011133 (JGI 61323)	MicpuC2.est_orfs.1_306_4269596:1	90.5	2×10^{-19}
2615008475	MicpuC2.EuGene.0000130210	74.3	4×10^{-14}
2615010887	MicpuC2.est_orfs.10_1861_4270447:1	70.5	3×10^{-13}
2615007841	MicpuC2.EuGene.0000090108	67.8	4×10^{-12}
JGI ID 360	Myb domain-containing protein		8×10^{-156}


Ratio of cell normalized metabolite concentrations
(P-deficient:P-replete)
-3
-1
1
3
5
7
9
Thiamin
Xanthine*
DMSP
Inosine
Proline
Malate
Methionine
N-acetylglutamate
Betaine
Pantothenate
Glutamate
Xanthosine
Succinate
Arginine
Phenylalanine
Guanosine
Leucine/Isoleucine
Tryptophan
S-adenosyl-L-methionine
Glutathione oxidized
Glucosamine-6-phosphate
Folate
Spermidine
Thiamin monophosphate
NAD
Guanine
Glycerol-3-phosphate
Orotate
N-acetylmuramic acid*
Citrate
Adenosine
Adenosine 5'-monophosphate
Adenine
Glutamine
Aspartate
Tryptamine*
5-(5'-adenosyl)-L-homocysteine
3-phosphoglycerate
Uracil
Ribose 5-phosphate
2,3-dihydroxypropane-1-sulfonate
ND-R
ND-R
ND-R
ND-D
ND-D
ND-D
ND-D
ND-D
ND-D

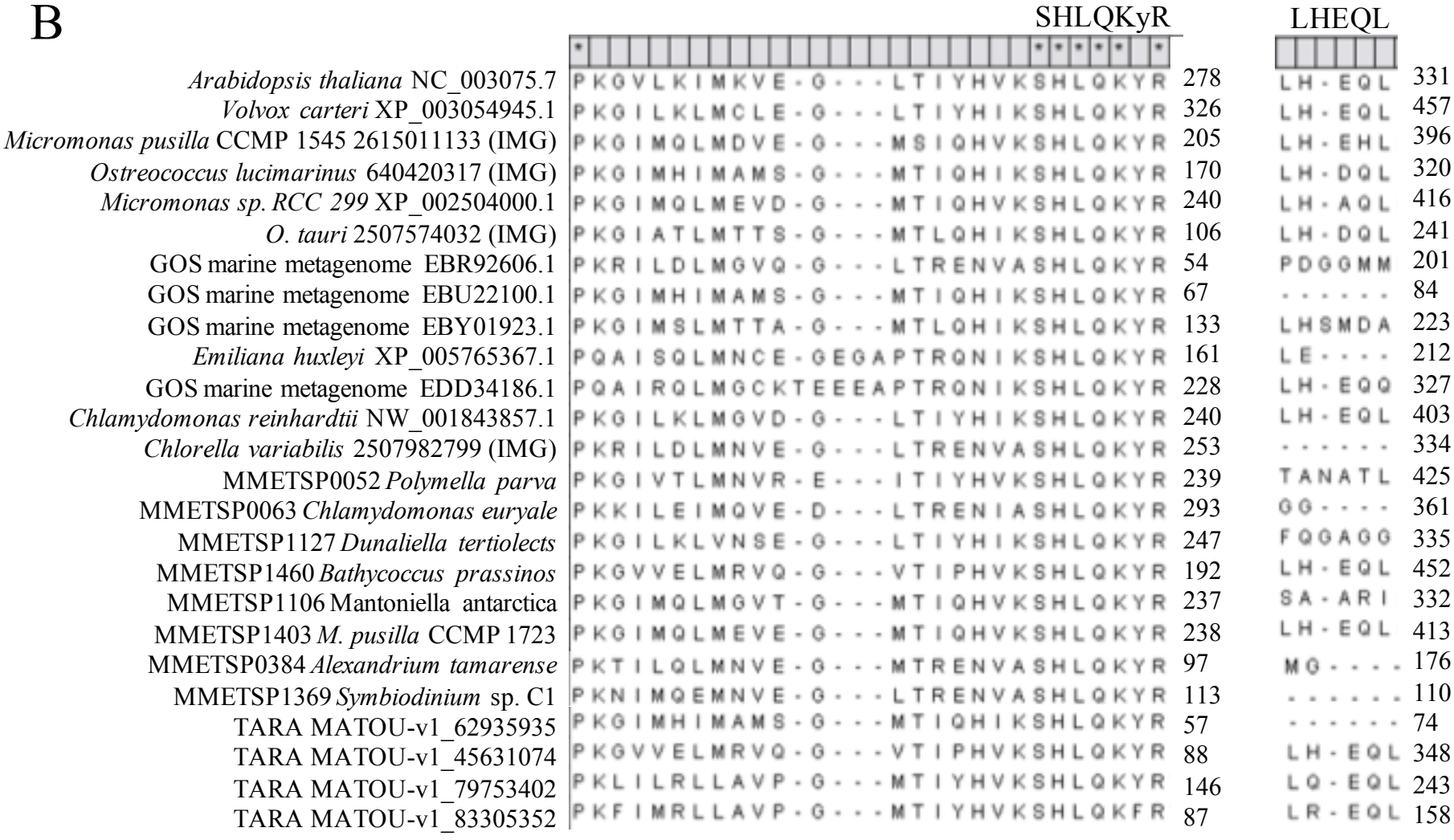

Metabolite ratio
3.0
2.0
2.5
1.5
1.0
0.5
*
Deplete
Replete
Deplete
Replete
Guanosine:Guanine
Xanthosine:Xanthine

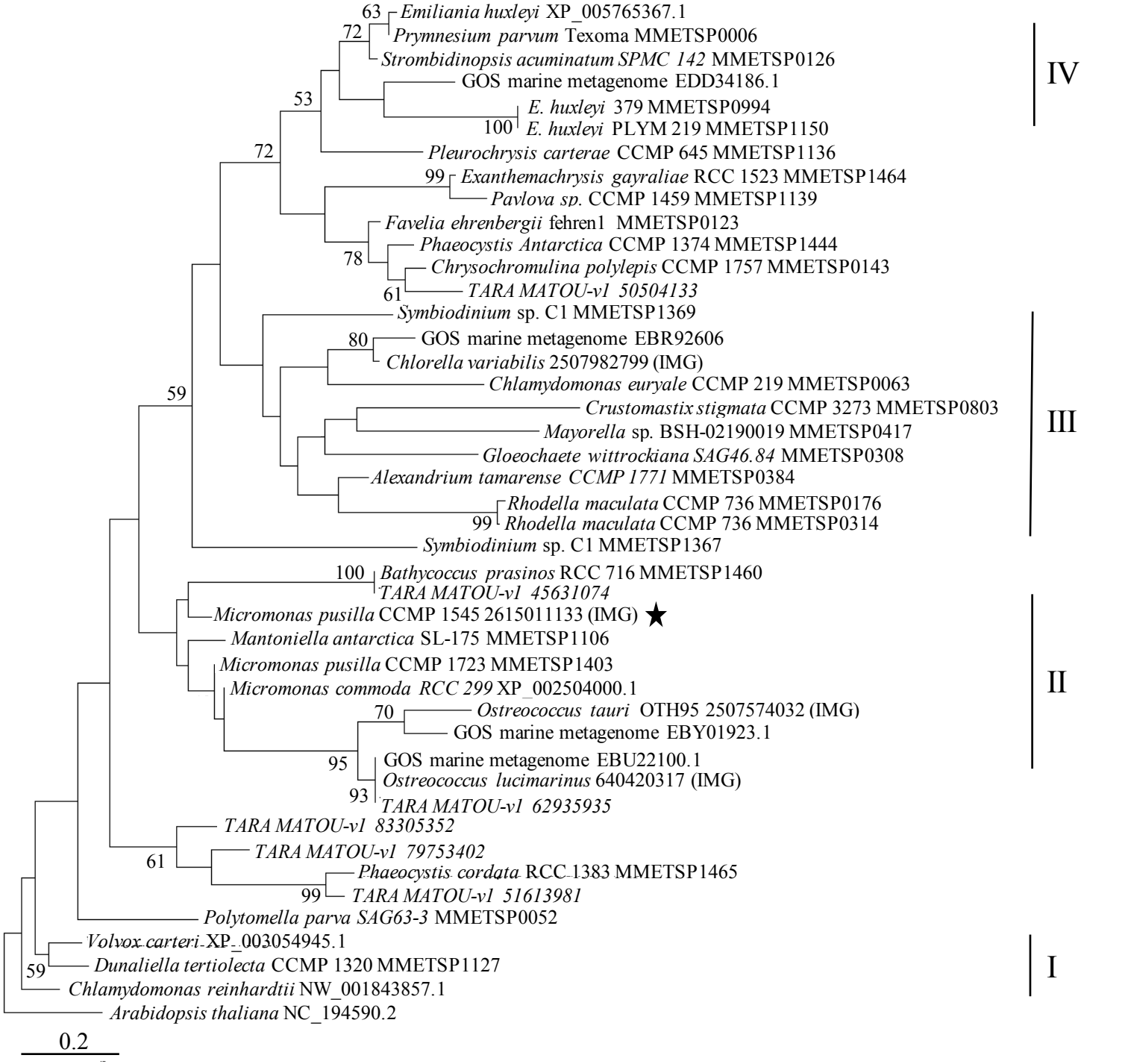
A

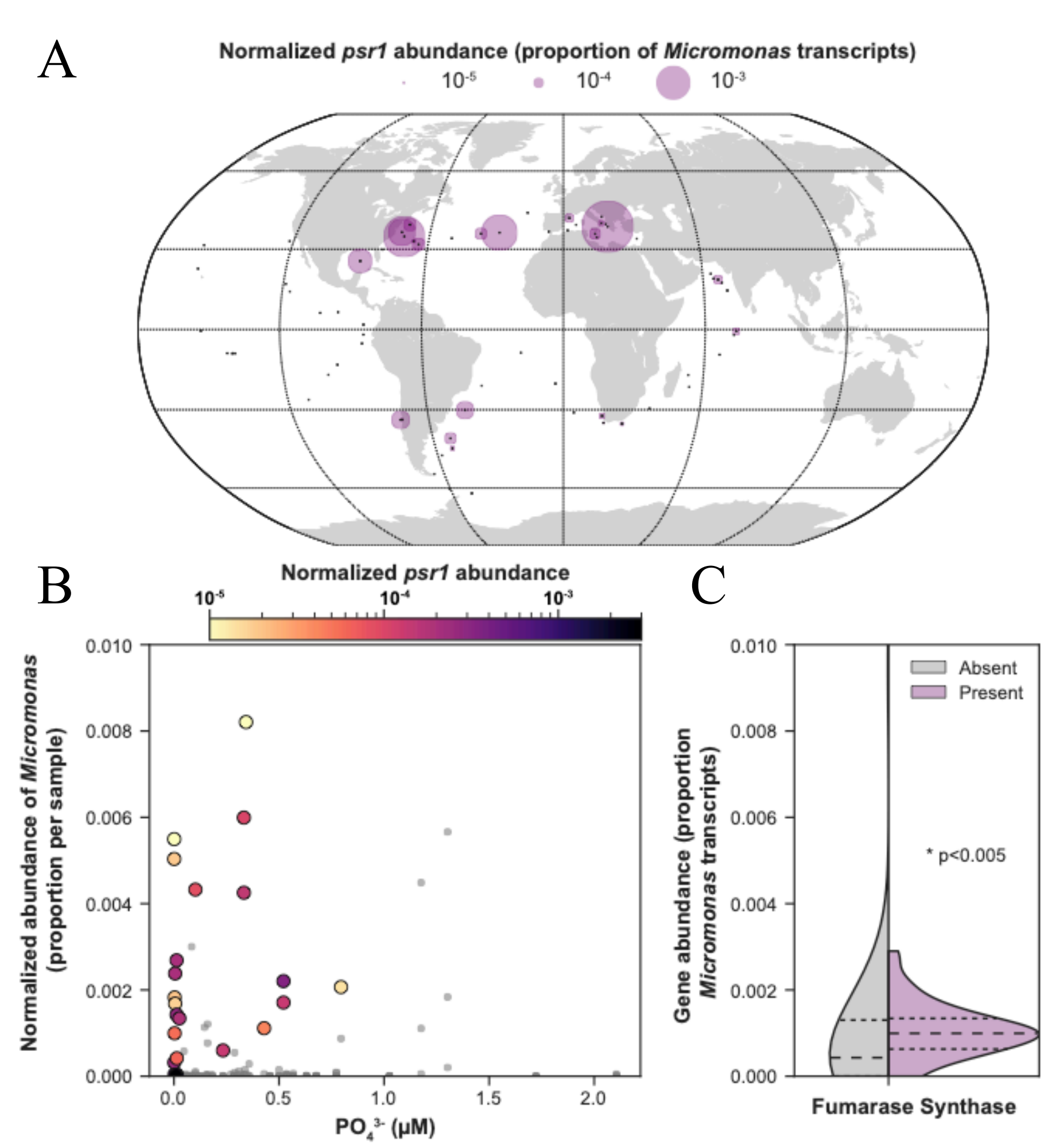
Micromonas pusilla CCMP1545: 2615011133

MTRMSGIEDDSFSYSFLLISGQLRDPTRLTPRPQTSSAPLGAGGGAAAASGMNTSGTFWPSDLDLDLESSDFL
DSLLLGGEMMTTQQQHHHAGDERLGHGPVPLGARAAGEDHDPRVVLAERDAGDGLLPRALADERREPVAR
AVVASNNNKQRLRWTPELHKMFVDAVKRLGGLDLATPKGIMQLMDVEGMSIQHVKSHLQKYRLQDSGGG
ASEFRVSPDASASGKRPRSEEDDAGGNGKDGNNSAGKTRRRPSAAERSAARLRAAEEKAREERDAARSMAA
AAAAAAAAAVERQNVELALLTGDAAGARDALEMSSHHAASHVGGAYDDVLDLVAGPGESGDGDGAWGD
VLHDHAIVGAAGDDVGLVDDVGSDPEAAAAMLKQLELQKKLHEHLMSQRRLQQQVEAHGVYLETILDQQK
RRRGVE

B

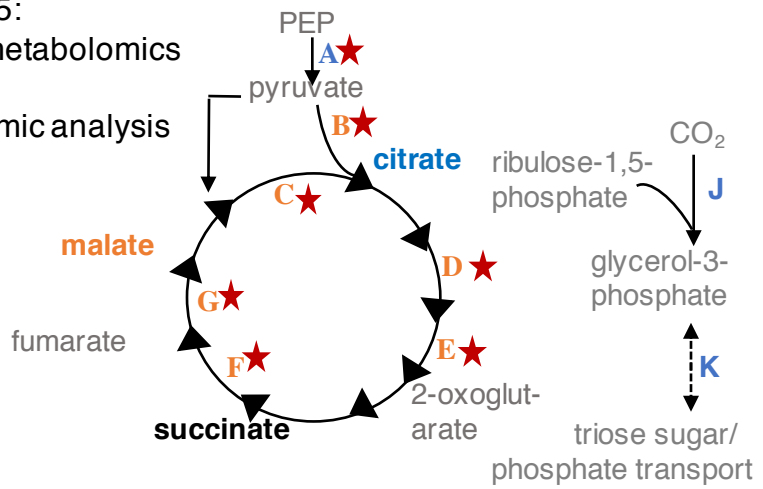



63
72
Emiliania huxleyi XP_005765367.1
Prymnesium parvum Texoma MMETSP0006
Strombidinopsis acuminatum SPMC 142 MMETSP0126
GOS marine metagenome EDD34186.1
53
E. huxleyi 379 MMETSP0994
100
E. huxleyi PLYM 219 MMETSP1150
72
Pleurochrysis carterae CCMP 645 MMETSP1136
99
Exanthemachrysis gayraliae RCC 1523 MMETSP1464
Pavlova sp. CCMP 1459 MMETSP1139
Favelia ehrenbergii fehren1 MMETSP0123
Phaeocystis Antarctica CCMP 1374 MMETSP1444
78
Chrysochromulina polylepis CCMP 1757 MMETSP0143
61
TARA MATOU-v1 50504133
Symbiodinium sp. C1 MMETSP1369
80
GOS marine metagenome EBR92606
Chlorella variabilis 2507982799 (IMG)
Chlamydomonas euryale CCMP 219 MMETSP0063
59
Crustomastix stigmata CCMP 3273 MMETSP0803
Mayorella sp. BSH-02190019 MMETSP0417
Gloeochaete wittrockiana SAG46.84 MMETSP0308
Alexandrium tamarense CCMP 1771 MMETSP0384
Rhodella maculata CCMP 736 MMETSP0176
99
Rhodella maculata CCMP 736 MMETSP0314
Symbiodinium sp. C1 MMETSP1367
100
Bathycoccus prasinos RCC 716 MMETSP1460
TARA MATOU-v1 45631074
Micromonas pusilla CCMP 1545 2615011133 (IMG)
Mantoniella antarctica SL-175 MMETSP1106
Micromonas pusilla CCMP 1723 MMETSP1403
Micromonas commoda RCC 299 XP_002504000.1
70
Ostreococcus tauri OTH95 2507574032 (IMG)
GOS marine metagenome EBY01923.1
95
GOS marine metagenome EBU22100.1
Ostreococcus lucimarinus 640420317 (IMG)
93
TARA MATOU-v1 62935935
TARA MATOU-v1 83305352
61
TARA MATOU-v1 79753402
Phaeocystis cordata RCC 1383 MMETSP1465
99
TARA MATOU-v1 51613981
Polytomella parva SAG63-3 MMETSP0052
Volvox carteri XP_003054945.1
59
Dunaliella tertiolecta CCMP 1320 MMETSP1127
Chlamydomonas reinhardtii NW_001843857.1
Arabidopsis thaliana NC_194590.2
0.2
IV
III
II
I


A
Normalized psr1 abundance (proportion of Micromonas transcripts)
10-5
10-4
10-3
B
Normalized psr1 abundance
10-5
10-4
10-3
Normalized abundance of Micromonas (proportion per sample)
0.010
0.008
0.006
0.004
0.002
0.000
0.0
0.5
1.0
1.5
2.0
PO43- (µM)
C
Absent
Present
* p<0.005
Gene abundance (proportion Micromonas transcripts)
0.010
0.008
0.006
0.004
0.002
0.000
Fumarase Synthase

TCA cycle & Carbon fixation

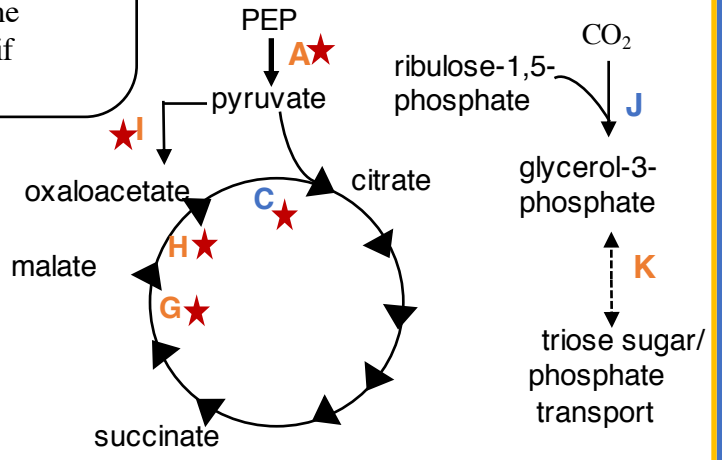
M. pusilla CCMP1545: Targeted metabolomics and transcriptomic analysis



Elevated in P-deficient
No significant difference
Depressed in P-deficient
Not quantified
★ Gene contains the conserved motif

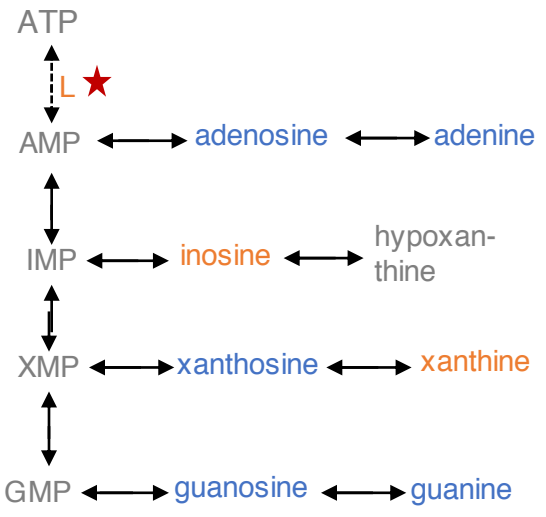
Potential P-deficiency strategy:
Decreased carbon fixation. Increased carbon flow through the TCA cycle to build metabolites for C, N, and P storage and salvage

M. commoda RCC299: Transcriptomic analysis

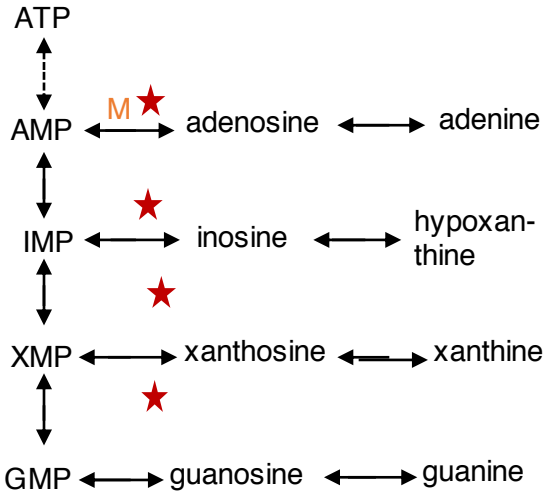


Potential P-deficiency strategy:
Decreased carbon fixation. Increased carbon flow through the malate-aspartate shuttle with minimal increase in flow through the entire TCA cycle[1]

Purine nucleotide metabolism



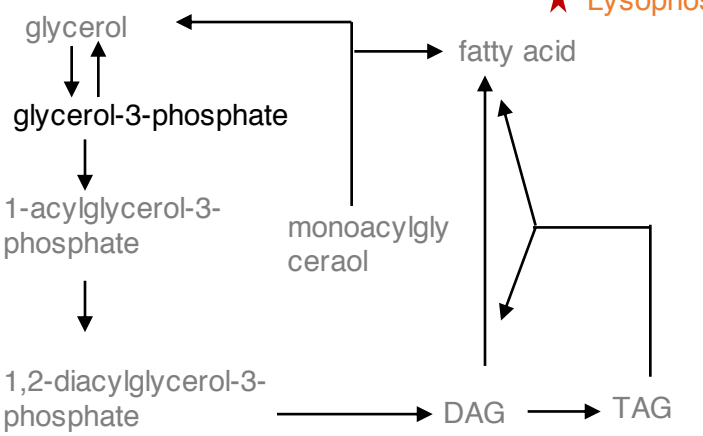
Shift nucleosides to nucleobases



Shift nucleosides to nucleobases

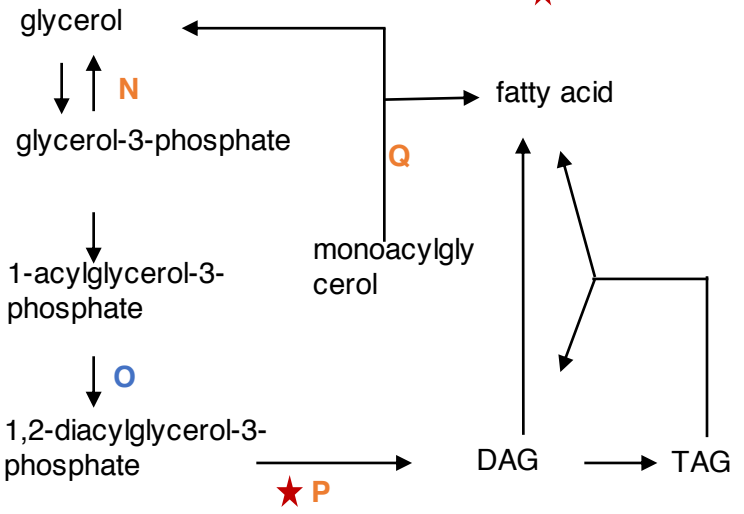
Lipid metabolism

★ Acyl-CoA synthetase
★ Lysophospholipase



Maintain TAG metabolism and potentially increase some fatty acid and other lipid biosynthesis and salvage

★ LPAT



Increase TAG biosynthesis to serve as energy storage molecules[2]