Phosphate Limitation Responses in Marine Green Algae Are Linked to Reprogramming of the tRNA Epitranscriptome and Codon Usage Bias

Elisabeth Hehenberger [ID],[1,2] Jian Guo,[3,8] Susanne Wilken,[3,9] Kenneth Hoadley,[1,10] Lisa Sudek,[3] Camille Poirier,[1,11] Richard Dannebaum,[4] Edward Susko,[5] and Alexandra Z. Worden [ID] [1,3,6,7,*]

[1]Ocean EcoSystems Biology Unit, RD3, GEOMAR Helmholtz Centre for Ocean Research, 24148 Kiel, DE

[2]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 370 05 České Budějovice, CZ

[3]Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA 95064, USA

[4]Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[5]Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia B3H 4R2, CA

[6]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, USA

[7]Max Planck Institute for Evolutionary Biology, 24306 Plön, DE

[8]Present address: Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, CA

[9]Present address: Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, NE

[10]Present address: Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, USA

[11]Present address: Department of Biology, University of Oxford, Oxford, UK

*Corresponding author: E-mail: azworden@mbl.edu.

Associate editor: Deepa Agashe

Abstract

Marine algae are central to global carbon fixation, and their productivity is dictated largely by resource availability. Reduced nutrient availability is predicted for vast oceanic regions as an outcome of climate change; however, there is much to learn regarding response mechanisms of the tiny picoplankton that thrive in these environments, especially eukaryotic phytoplankton. Here, we investigate responses of the picoeukaryote *Micromonas commoda*, a green alga found throughout subtropical and tropical oceans. Under shifting phosphate availability scenarios, transcriptomic analyses revealed altered expression of transfer RNA modification enzymes and biased codon usage of transcripts more abundant during phosphate-limiting versus phosphate-replete conditions, consistent with the role of transfer RNA modifications in regulating codon recognition. To associate the observed shift in the expression of the transfer RNA modification enzyme complement with the transfer RNAs encoded by *M. commoda*, we also determined the transfer RNA repertoire of this alga revealing potential targets of the modification enzymes. Codon usage bias was particularly pronounced in transcripts encoding proteins with direct roles in managing phosphate limitation and photosystem-associated proteins that have ill-characterized putative functions in "light stress." The observed codon usage bias corresponds to a proposed stress response mechanism in which the interplay between stress-induced changes in transfer RNA modifications and skewed codon usage in certain essential response genes drives preferential translation of the encoded proteins. Collectively, we expose a potential underlying mechanism for achieving growth under enhanced nutrient limitation that extends beyond the catalog of up- or downregulated protein-encoding genes to the cell biological controls that underpin acclimation to changing environmental conditions.

Key words: green algae, codon usage, nutrient limitation, tRNA modification, marine primary production.

Introduction

To accomplish the primary production that supports ocean food webs, algae must acclimate to seasonally and spatially varying nutrient conditions (Wu et al. 2000; Björkman et al. 2018). Nutrient limitation is predicted to increase in vast areas of the ocean with warming, likely posing additional challenges to acclimation and growth of the phytoplankton responsible for marine primary production (Bopp et al. 2001; Behrenfeld et al. 2006). Some aspects of algal responses to elemental growth limitation appear to follow a generalized acclimation program, combining initial resource acquisition and storage mobilization steps with more drastic measures that implement

Open Access

elemental-sparing and elemental-recycling mechanisms if the deficiency persists (Merchant and Helmann 2012; Dyhrman 2016). In eukaryotic marine algae, these responses have been observed on the level of gene transcripts or protein expression or both based on comparisons of expression data from cells exposed to replete, limiting, or starvation conditions (Dyhrman et al. 2012; Liu et al. 2015; Haley et al. 2017; Guo et al. 2018; Zhang et al. 2019). The identified commonalities appear to transcend major algal groups, likely in part because they hinge on identification of protein function—which is biased toward well-studied processes in model organisms, such as protein synthesis, nutrient transport, and photosynthesis, functions that have a high degree of conservation across multiple lineages (Grossman and Aksoy 2015; Harke et al. 2017).

The alga *Micromonas* is known for its importance in the oceans (Worden et al. 2009; van Baren et al. 2016), and for its role in elucidating aspects of evolution in Viridiplantae, to which both green algae and land plants belong (Faktorová et al. 2020; Bachy et al. 2022). *Micromonas* is an abundant marine picoeukaryote (≤2 μm cell diameter) and comprises diverse lineages (Monier et al. 2016; Tragin and Vaulot 2019) that together have been proposed to function as sentinels for ocean warming (Worden et al. 2009; Demory et al. 2019). Clade A, represented by *Micromonas commoda*, is among the most important species in subtropical regions where warming will dramatically alter nutrient availability due to increased stratification (Demory et al. 2019; Bolaños et al. 2020). In addition to being predicted to become even more scarce, phosphate (P) is currently a major limiting nutrient in the subtropical North Atlantic gyre where this alga has been observed (Carlson et al. 2009; Cuvelier et al. 2010). With respect to nutrient limitation, this species exhibits a conserved proteomic response to P-limiting conditions, including upregulation of P metabolism and transport proteins, as well as a role for some proteins of narrow phylogenetic distributions that have unknown functions (Guo et al. 2018). This raises questions about what other factors shape the individual trajectories of algal species and influence their cellular modulation in response to changing environmental conditions.

Transfer RNAs (tRNAs) and their modifications (epitranscriptome) have also recently been studied within the context of stress response mechanisms in bacterial and eukaryotic model organisms (Chan et al. 2010, 2012; Endres et al. 2015; Chionh et al. 2016; Chan et al. 2018). The tRNA epitranscriptome captures the entirety of enzyme-catalyzed chemical modifications of ribonucleosides present in tRNAs —such that various chemical modifications of the tRNA ribonucleosides can be added at several stages of tRNA biosynthesis and affect different aspects of tRNA function, depending on their location on the tRNA molecule (Berg and Brandl 2021). Modifications that occur away from the anticodon region tend to be essential for tRNA structure and stability, while those within affect translation rate and fidelity by preventing frameshifts and influencing codon anticodon pairing (El Yacoubi et al. 2012; Berg and Brandl 2021).

Modified ribonucleosides are numerous (>100) and diverse in tRNAs relative to other RNAs, with some modifications being present across the domains of life (Boccaletto et al. 2018). Based on studies performed in model organisms (i.e. mycobacterium, yeast, and animal; Chan et al. 2010, 2012; Endres et al. 2015; Chionh et al. 2016; Chan et al. 2018), it has been proposed that an interplay between alterations in tRNA modifications and specific transcripts with biased patterns of codon usage influences the translation of proteins responding to an underlying stressor—resulting in new concepts surrounding stress response mechanisms in non-photosynthetic models (Chan et al. 2018). Aspects of this stress response mechanism have been studied in vascular plant models where several tRNA modifications correlate with cold temperature, drought, or increased soil salinity and also developmental stages (Wang et al. 2017). Thus far, tRNA modifications are unexplored in free-living unicellular algae that dominate pelagic marine environments where key stressors, such as variable nutrient availability, differ from those previously examined in plants.

Here, we exposed *M. commoda* to transitions in P availability using continuous culture photobioreactors. We discovered a large number of transcripts encoding enzymes catalyzing tRNA modifications were upregulated in *M. commoda* when grown under P-limiting conditions. We therefore reinvestigated this protein type—resulting in identification and annotation of a total of 94 tRNA modification enzymes in *M. commoda*. The results point to an unprecedented response to limiting nutrient "stress" conditions for marine and freshwater algae—operating at the level of tRNA modifications. Concordantly, we identified biases in codon usage in transcripts more abundant in P-limiting conditions in comparison to transcripts with higher abundance in P-replete conditions, and these had known functional roles in P limitation responses. These results highlight a regulatory mechanism that underpins green algal acclimation to shifting conditions in the oceans and likely extends to diverse marine algae.

Results and Discussion

Physiological Responses to Varying P Availability

To understand algal responses to environmentally relevant conditions, we performed controlled experiments using a specialized photobioreactor (Guo et al. 2018) with *M. commoda*. During the experiments, inorganic P was manipulated in a limiting but not starvation-inducing manner, while other resources were held constant. A stable P-limited phase was achieved wherein *M. commoda* had significantly slower exponential growth ($0.14 \pm 0.11 \ d^{-1}$) than in P-replete "precultures" ($0.61 \pm 0.10 \ d^{-1}$; Fig. 1a, supplementary dataset S1A and fig. S1, Supplementary Material online). These growth rates fit with the few field measurements that are available for picoeukaryotic algae in samples where *M. commoda* was notable (Worden et al. 2004; Cuvelier et al. 2010). Coincident with the growth phases, we observed P drawdown (due to algal growth) in the P-limited phase, and cell size increased
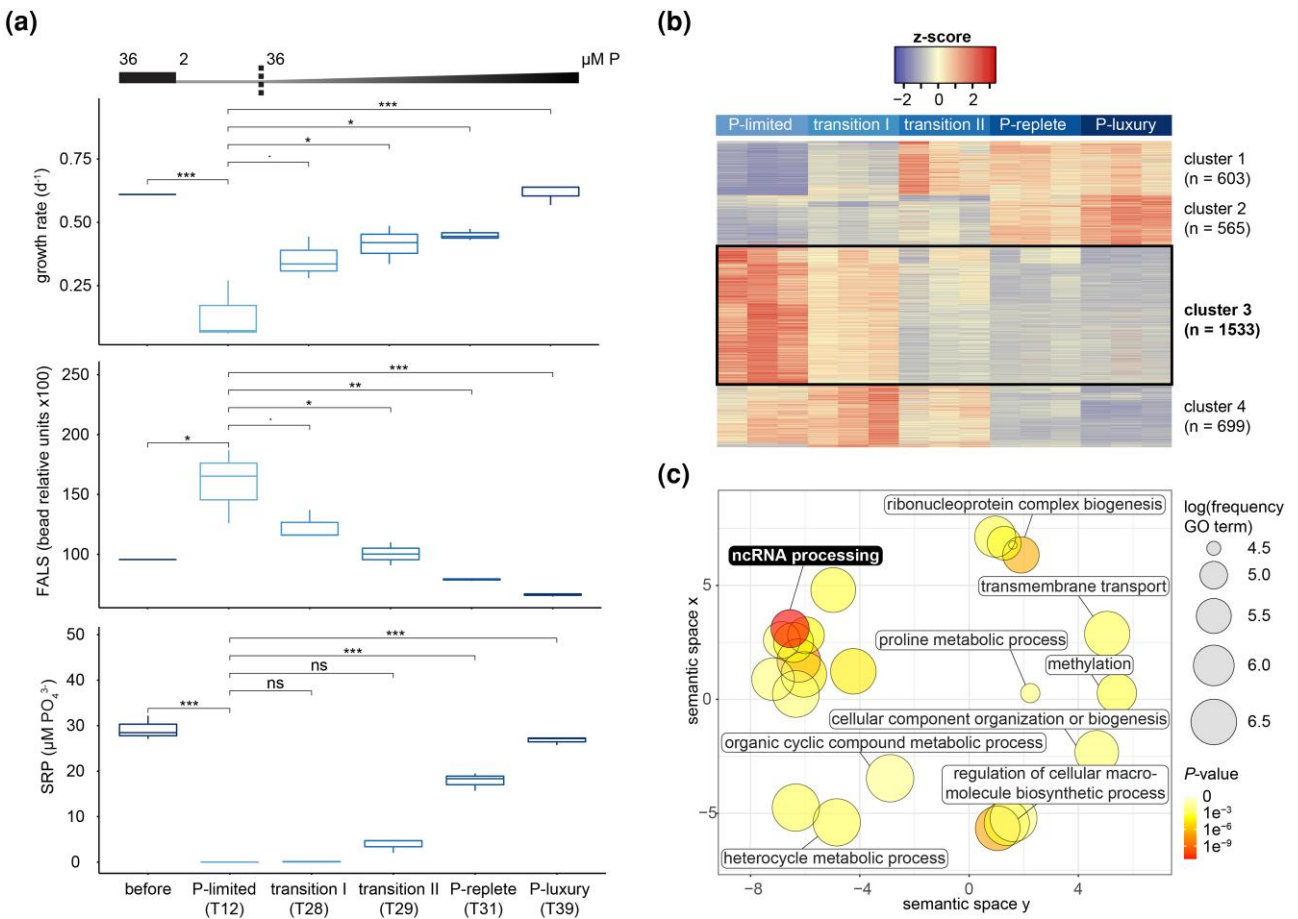
Fig. 1. Changes in P availability correlate with trends in physiology and distinct clusters of gene expression. a) Box plots for growth rate (top), FALS as a proxy for cell size (center), and residual SRP (bottom, after phytoplankton growth) for each condition across the 3 photobioreactors: 3 h before inoculation into photobioreactors (before; 1 h after lights on) and on samplings Days T12 (P limited), T28 (transition I), T29 (transition II), T31 (P replete), and T39 (P luxury) 1 h after lights on. After inoculation from semicontinuous replete cultures (36 μM PO_4^{3-} in medium), the photobioreactors were acclimated to limiting P conditions (1.5 to 2 μM PO_4^{3-}, see Materials and Methods) in which the standing stock was drawn down to 0.02 ± 0.04 μM PO_4^{3-} (as SRP) due to algal growth. The horizontal bar along top and numbers above it indicate the P concentration in the medium supplied and the gradient triangle indicates increasing P concentration from 2 to 36 μM in the photobioreactor columns after the supply was switched on Day 27 (indicated by a vertical dashed line; supplementary fig. S1 and dataset S1A, Supplementary Material online). Significant differences are indicated as: •$P < 0.05$; *$P < 0.01$; **$P < 0.001$; and ***$P < 0.0001$. ns, not significantly different. b) Heatmap of z-scores of all differentially expressed genes ($n = 3{,}400$) in the experiment (FC ≥ 2, $P \geq 0.01$ in at least 1 all-vs.-all pairwise comparison), clustered by k-means ($k = 4$, based on a within-cluster sum of squared error plot). The recovered clusters are numbered from top to bottom of the heatmap, and a black box indicates the genes ($n = 1{,}533$) upregulated during P limitation. c) Semantic similarity clustering of enriched GO terms (BP) in the cluster upregulated during P limitation (black box in b). Bubble size corresponds to the frequency of the GO term in the EBI GOA database as implemented in REVIGO.

relative to the preculture (based on the proxy forward angle light scatter [FALS]).

After the P-limited phase, we transitioned *M. commoda* back to a P-replete state, leading to a rapid increase in growth rate and sharp decrease in cell size (Fig. 1a; supplementary fig. S1A to C, Supplementary Material online). Residual P in the spent medium began to increase between 24 and 48 h after P refeeding was initiated. Here, mean cell size increased as growth rate decreased, due to a greater percentage of the population being stalled before division, or simply not reaching division, whereas it decreased as a great proportion of cells divided each day at higher growth rates. These results were similar to rates and shifts in cellular characteristics observed under a similar photobioreactor experiment exploring P limitation in

M. commoda (Guo et al. 2018). F_v/F_m, representing the maximum potential quantum efficiency of photosystem II, was significantly lower at T12 (0.34 ± 0.01) than P replete (0.45 ± 0.00) as well as other phases ($P < 0.05$), while no significant differences were observed between transition I, transition II, and P replete to P luxury (supplementary dataset S1A, Supplementary Material online).

Nutrient Limitation Response Delineated by Transcriptomes

Just over a third of *M. commoda* genes ($n_{TOTAL} = 9{,}777$ genes) exhibited differential expression (DE; $n_{DE} = 3{,}400$, fold change [FC] ≥ 2, $P \leq 0.01$) in the experimental time course (Fig. 1b). The percentage of genes responding to limitation was in line

3

with studies of other marine algae (Dyhrman et al. 2012; Liu et al. 2015; Haley et al. 2017; Harke et al. 2017; Guo et al. 2018; Zhang et al. 2019), although the bulk of those experiments explored P starvation, wherein growth was completely terminated. Here, 4 gene clusters were distinguished using k-means clustering of expression data, and these corresponded to different response phases (Fig. 1b). The largest transcriptional cluster, Cluster 3, represented genes predominantly upregulated during P limited (time point T12), with a rapid shift to much lower transcript abundances as the P supply increased, and comprised nearly half of all differentially expressed genes ($n = 1,533$). Functional enrichment analysis based on Gene Ontology (GO) terms indicated these were most enriched in the biological process (BP) "noncoding RNA processing" (Fig. 1c; supplementary dataset S1B, Supplementary Material online). These observations were unprecedented in that noncoding RNA processing has not been reported as a response to nutrient shifts in marine algae and are explored further in the chapters below.

The other 3 clusters of DE genes connected to transcripts upregulated after the P-limited phase (Fig. 1b). Cluster 4 was primarily upregulated during transition I (time point T28, $n = 699$), Cluster 1 during transition II and following phases (time points T29 to T39, $n = 603$), and Cluster 2 during the P-luxury phase (time point T39, $n = 565$; Fig. 1b). Functional enrichment analysis associated the transition I and the transition II Clusters 4 and 1 most significantly with the BP term "cellular protein modification process" (supplementary dataset S1C and D, Supplementary Material online, respectively), which mainly encode undescribed kinases or phosphatases, as well as numerous proteins involved in protein folding, repair, and degradation. The latter group indicated a "repair-and-recovery phase" after nutrient limitation. Several of the proteins encoded by these genes were analyzed phylogenetically, including the state transition kinase 7 (STN7) and state transition kinase 8 (STN8, discussed below; supplementary dataset S1C and D, Supplementary Material online). For Cluster 2 (P luxury), the most significantly enriched term is described as "microtubule-based process," including flagellar-associated proteins and numerous undescribed motor domain–proteins involved in intracellular transport and cell division (supplementary dataset S1E, Supplementary Material online). These results suggested increased motility and growth, corresponding to the higher availability of P in the P-luxury stage. Our findings point to several unexplored avenues in response to P limitation. We therefore designed and implemented approaches to examine findings from the gene enrichment analysis more deeply, particularly for the cluster associated with the P-limited phase.

The tRNA Epitranscriptome and tRNA Repertoire of *M. commoda*

To confirm the enrichment results, an independent clustering approach was used to identify genes associated with growth under P limitation. This weighted correlation network analysis (WGCNA, supplementary fig. S2, Supplementary Material online) returned the same top GO term ("noncoding RNA processing") as the above approaches after functional enrichment analysis (see supplementary dataset S1F, Supplementary Material online for associated genes). We then used phylogenetic inference to identify all genes labeled with this GO term, which revealed genes involved in ribosomal RNA (rRNA) processing, ribosome assembly, and modification of various RNA types, particularly tRNAs (supplementary dataset S1G, Supplementary Material online). Thus, 2 different analytical approaches recovered numerous transcripts involved in modification of tRNA ribonucleosides that were upregulated under P limitation. These results are similar to proposed stress response mechanism described in bacterial, yeast, and animal models (Chan et al. 2018), involving stress-induced changes in tRNA modifications and selective translation of codon-biased messenger RNAs (mRNAs).

To date, little is known about tRNA modification enzymes in *Micromonas*—or other marine algae. Therefore, we aimed to (i) identify all tRNA modification enzymes in *M. commoda* and their program of expression across the experimental manipulations and (ii) identify codon usage of genes differentially expressed during P-limited conditions and determine possible biases in codon usage between nutrient-limited and replete states. Using phylogenetic approaches, we identified 94 *M. commoda* proteins that were homologous to described tRNA-modifying proteins in other organisms (supplementary dataset S2A, Supplementary Material online; see also "tRNA modifying enzyme phylogenies" in the figshare repository at https://doi.org/10.6084/m9.figshare.21164086). While the overall number in *M. commoda* is similar to that in *Arabidopsis thaliana* (~90; Chen et al. 2010), the composition of the modification enzyme complement partially differs. Several gene families in plant appear expanded compared to *Micromonas* (e.g. the Pus family), while multiple others in *Micromonas* appear to be lacking in *Arabidopsis*, particularly those of putative bacterial origins. The latter may have come from marine bacteria, although taxonomic origins could not be pinpointed due to inherent limitations of single-gene trees.

The modifications catalyzed by the *Micromonas* proteins spread across the ~80-ribonucleoside-long tRNA molecule and include modifications within and remote from the anticodon loop (Fig. 2a). Modifications in the anticodon loop are known to affect translation rate and fidelity. For example, the modification at Position 34 (or the "wobble" position) modulates recognition of specific codons in other organisms (Agris et al. 2018), ultimately allowing decoding of the degenerate genetic code. As another example, those at Position 37 (as also seen herein) stabilize codon–anticodon pairing, thus helping minimize frameshifts (El Yacoubi et al. 2012).

Given the marked expression of annotated modification enzymes, we next examined and annotated the suite of tRNAs present in *M. commoda* using a variety of search
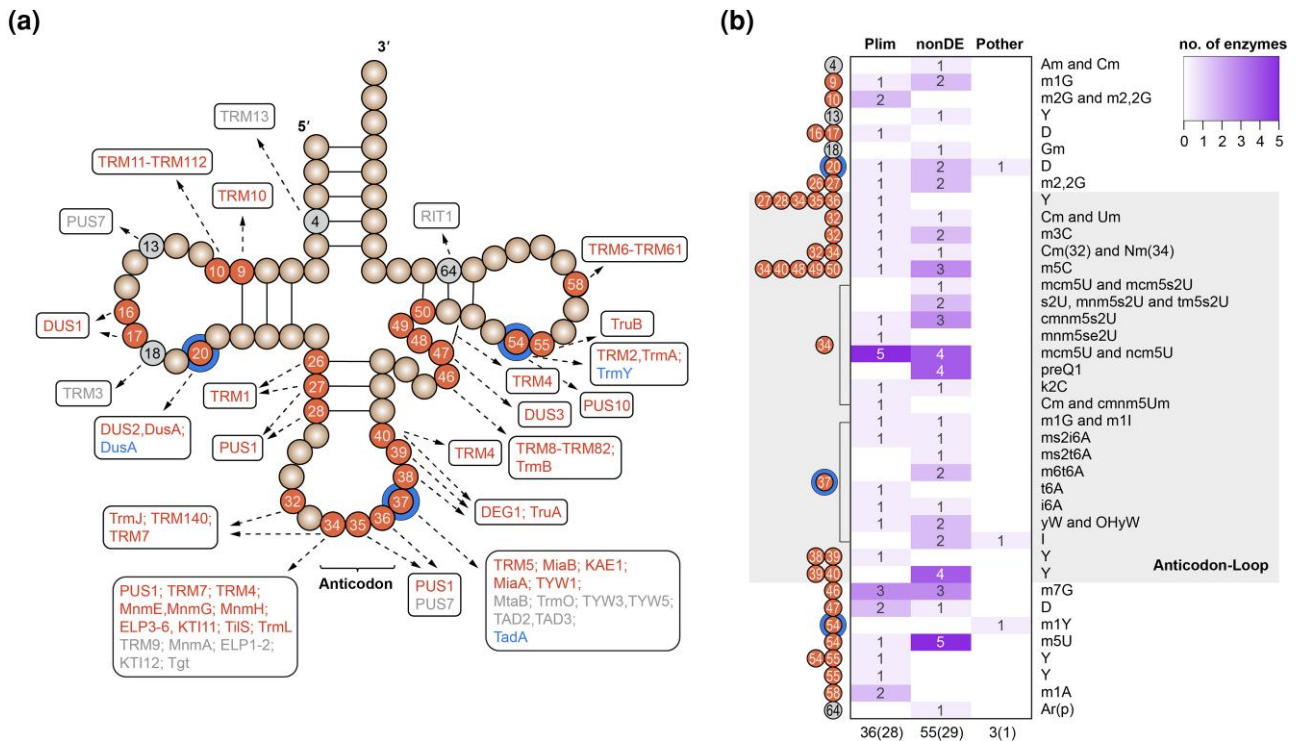
(a) / **(b)**

Plim	nonDE	Pother	Modification
	1		Am and Cm
1	2		m1G
2			m2G and m2,2G
	1		Y
1			D
	1		Gm
1	2	1	D
1	2		m2,2G
1			Y
1	2		Cm and Um
1	1		m3C
1	1		Cm(32) and Nm(34)
1	3		m5C
	1		mcm5U and mcm5s2U
	2		s2U, mnm5s2U and tm5s2U
	3		cmnm5s2U
1			mnm5se2U
5	4		mcm5U and ncm5U
	4		preQ1
1	1		k2C
1			Cm and cmnm5Um
1	1		m1G and m1I
	1		ms2i6A
	1		ms2t6A
	2		m6t6A
1	1		t6A
1	1		i6A
2	2	1	yW and OHyW
1			I
			Y
	4		Y
3	3		m7G
2	1		D
		1	m1Y
1	5		m5U
1			Y
1			Y
2			m1A
	1		Ar(p)
36(28)	**55(29)**	**3(1)**	

Anticodon-Loop

no. of enzymes: 0 1 2 3 4 5

Fig. 2. Most tRNA-modifying enzymes in *M. commoda* are upregulated under P-limiting conditions, including numerous enzymes involved in anticodon loop modification. a) Schematic tRNA molecule with ribonucleotides that can potentially be modified in *M. commoda* indicated by numbered spheres. Ribonucleotides are colored red if at least 1 enzyme modifying this residue is upregulated during P-limiting conditions, and additional blue rings around spheres indicate that 1 of the 3 enzymes upregulated after P resupply acts on this residue. Gray spheres indicate positions only modified by enzymes that are not differentially expressed herein. Names of enzymes responsible for modifying the corresponding residue are listed for each numbered ribonucleotide position, with red text if ≥1 representative catalyzing the corresponding modification is upregulated during P-limiting conditions or blue if at later stages of the experiment. Gray text indicates enzymes that were identified in the *M. commoda* genome but were not differentially expressed. Commas separate enzymes involved in the same modification, and dashes indicate enzymes acting together in complexes. b) Heatmap indicating the number of enzymes for each modification either upregulated during P limitation (Plim), upregulated at another stage (Pother), or not differentially expressed (nonDE). Some enzymes act on >1 ribonucleotide, and they are listed according to the lowest-numbered residue within the tRNA molecule on which they act. Each row corresponds to a specific modification, the spheres on the left (colors corresponding to a) indicate the ribonucleotides where this modification can occur, while the text on the right lists the abbreviation for the given modification (see supplementary dataset S2, Supplementary Material online for complete names). A gray box indicates modifications that are found in the anticodon loop of the tRNA molecule that have been shown to be involved in the regulation of codon–anticodon recognition in model organisms (Berg and Brandl 2021).

approaches. The 94 modification enzymes identified can putatively act on 49 nuclear, 34 mitochondrial, and 26 plastid tRNAs predicted to be present in *M. commoda* (supplementary dataset S2C, Supplementary Material online). For each of the 20 amino acids, at least 1 nuclear and mitochondrial tRNA is present, as is a nuclear tRNA-recognizing selenocysteine. Plastid tRNAs were predicted for all amino acids except valine (V) and histidine (H). These tRNA numbers are similar to the tRNA repertoire reported for the related green alga *Ostreococcus tauri*, which encodes 45 nuclear, 28 mitochondrial, and 27 plastidial tRNAs (Cognat et al. 2022).

Reprogramming of the tRNA Epitranscriptome Underpins Responses to P Limitation

Thirty-nine of 94 transcripts encoding tRNA-modifying enzymes in *M. commoda* were differentially expressed in our experiment. Remarkably, 36 were upregulated during P-limiting conditions (T12) relative to P-replete/P-luxury conditions (T31/T39) (Fig. 2b; supplementary fig. S3, Supplementary Material online), while just 3 were specifically upregulated at later stages of the experiment (transitional and/or replete/luxury stages) but not P limitation. Although less than half of the tRNA modification enzyme complement changes expression levels in our experiment, the majority of the different tRNA modifications were nevertheless affected by P limitation because many modifying functions are represented by more than 1 protein in *M. commoda* (Fig. 2b). This redundancy may partially be explained by different subcellular localizations; for example, we found orthologs to a cytoplasmic and a plastidial form of the modification enzyme TRM5 described for *Arabidopsis* (supplementary dataset S2A, Supplementary Material online).

Among the 36 tRNA-modifying enzymes upregulated during P limitation, several appear to catalyze modifications involved in responses to various stressors in model organisms. Paired transcriptomic and proteomic data that could help confirm the "consequences' of our observations are

5

not available for *M. commoda* or related marine prasinophytes, especially at similar detection levels. Nevertheless, we examined 2 case studies with considerable knowledge from model organisms (Zhang et al. 2022) and having significant changes in *M. commoda*, using the transcriptional data generated herein and prior proteomic data from a study of similar shifts in P availability (Guo et al. 2018). For the first case study, we examined TRM4, which is responsible for the methylation of the anticodon wobble Position C34 in tRNA^{Leu}(CAA) in yeast, and increased levels of this modification instigated by oxidative stress promoted the selective translation of mRNA enriched for the target codon UUG (e.g. a ribosomal protein; Chan et al. 2012). Furthermore, loss of TRM4 or the ribosomal protein resulted in hypersensitivity to this type of stress. Here, the TRM4 transcript is upregulated under P limitation, and examination of proteomic data (Guo et al. 2018) shows the protein itself increased. Of note, TRM4 is 1 of only 3 modification enzymes detected in the prior proteomics study. The other 2 were KAE1, part of a complex that catalyzes the formation of the modification t6A (Beenstock and Sicheri 2021), and EXPP1, a paralog to KTI11 which together with the elongator complex is involved in complex modifications of wobble uridines (Glatt et al. 2015). Neither KAE1 nor EXPP1 showed statistically significant changes in the proteomic study under P limitation (Guo et al. 2018). With respect to TRM4, the tRNA targeted in yeast, tRNA^{Leu}(CAA), is also present in *M. commoda*, implicating TRM4 in nutrient limitation responses of this alga.

For the second case study, we examined METTL2A and METTL6, which catalyze the formation of m^{3}C32 in cytosolic tRNAs (in tRNA^{Ser}/tRNA^{Thr}/tRNA^{Arg}) in human cell lines and mice (Xu et al. 2017; Ignatova et al. 2020). In yeast, stress response studies show that increased m^{3}C32 in tRNA^{Thr}(IGU) leads to preferential recognition of 2 out of 4 Thr codons (ACC and ACU), leading to selective expression of membrane protein genes enriched for these codons (Chan et al. 2015). We identified both METTL2A and METTL6 in *M. commoda*, with the latter displaying increased expression under P limitation. Neither enzyme was detected at the proteome level, potentially due to the low sensitivity or digests used for proteomics (Guo et al. 2018). Finally, inosine, which can pair with A, C, or U in mRNA, in the anticodon of tRNA^{Thr}(IGU) (a target of METTL 2 and 6) requires modification of adenosine at the anticodon wobble position. This modification is catalyzed by a tRNA-specific heterodimeric enzyme composed of ATAD2 and ATAD3 at tRNA precursor level (Schaub and Keller 2002). Orthologs of these enzymes are present in *M. commoda*, and *M. commoda* encodes for a tRNA gene with the anticodon AGU that can be modified to IGU, but neither ATAD2 nor ATAD3 was significantly upregulated during P limitation. Unfortunately, in general, there was restricted overlap between the transcriptomic data generated herein and the prior proteomic data set (Guo et al. 2018), largely due to the detection limits of the proteomic methodology wherein peptides were recovered from only 2,118 of the 10,306 protein-encoding genes

predicted in the *M. commoda* genome, versus 9,777 transcripts recovered herein. Collectively, these results indicated that cascades through to the protein level can be observed, but observation is likely impaired due to the lack of paired mRNA and protein stress response data sets. Overall, these findings call for additional studies to fully examine downstream consequences of our findings.

Our observations point to tRNA modifications as an important cellular response mechanism during growth under P limitation, an insight that extends beyond the classically investigated shifts in proteins involved in nutrient transport and other functions highlighted by most studies of transcript and protein expression levels (Dyhrman et al. 2012; Liu et al. 2015; Haley et al. 2017; Harke et al. 2017; Guo et al. 2018; Zhang et al. 2019). To contextualize our findings, we employed the phylogenetic analyses used to identify modification enzymes, to examine their distributions. This analysis indicated that 28 out of 36 upregulated enzymes during P limitation are present across eukaryotes (supplementary dataset S2A and B, Supplementary Material online; see also "Phylogenetic reconstructions of tRNA modifying enzymes" in the figshare repository at https://doi.org/10.6084/m9.figshare.21164086), including the above discussed examples (TRM4 and METTL6). Of the 55 enzymes showing no significant change, approximately half appear to be present in just a subset of eukaryotes, representing putative bacterial gene transfers (excluding endosymbiotic gene transfers from the plastid or mitochondrion) or paralogs to conserved eukaryotic proteins that are not present across eukaryotic groups (supplementary dataset S2A and B, Supplementary Material online). Presumably, these enzymes have roles outside responses to P limitation or potentially are involved in different life cycle stages. Thus, the phylogenetic distributions of the modification enzymes responding herein, in general, appear to be somewhat biased toward ancient eukaryotic proteins, largely conserved among eukaryotes, relative to those that did not respond. We note that taxonomic sampling for many eukaryotic groups is lower than for green algae; thus, a more complete set of these enzymes may yet be recovered in other algal lineages.

Codon Usage of Individual Codons Is Biased during P Limitation

The observed DE of enzymes catalyzing modifications on the tRNA molecule included numerous enzymes acting on the anticodon loop, influencing codon–anticodon pairing (Fig. 2b). These data raised the question of whether genes preferentially expressed in 1 condition display different codon usage than those in the other condition. We first examined the distribution of tRNAs recognizing codons in G or C and A or T across the genome, disregarding expressional patterns. This analysis showed that they are evenly represented in *M. commoda*, with 25 out of 49 tRNAs recognizing codons ending in G or C (supplementary dataset S2C, Supplementary Material online), precluding a genome-inherent bias.

Codon usage analysis was then performed on the subset of differentially expressed genes ($n = 1,811$) that showed the most significant logFC (based on adjusted P-value) for the comparison P-limited versus P-luxury (T12 vs. T39) relative to all other pairwise comparisons. These were divided into 2 groups based on having positive (1,185 upregulated genes) or negative (626 downregulated genes) logFC at T12. To compare codon usage over all sites for these 2 groups, we investigated the use of individual codons using z-statistics, controlling for selection by conducting the comparisons separately for each amino acid. Codons with positive z-statistics were used more frequently in P luxury (downregulated at T12) and all ended in G or C. In contrast, codons used more frequently in P limited (upregulated at T12) all ended in A or T (supplementary fig. S4A, Supplementary Material online). The observed AT bias in the T12-upregulated group is confirmed by χ^2 tests of homogeneity of AT usage. For most amino acids, the test was highly significant ($P < 0.0001$), with the exceptions lysine (K; $P = 0.84$), aspartic acid (D; $P = 0.32$), H ($P = 0.02$), and glutamic acid (E; $P = 0.01$), but regardless of whether AT content differed, it was greater for all amino acids in the T12-upregulated group.

We next identified genes contributing the most to codon usage bias. To this end, we identified genes in the T12-upregulated group with average AT usage that differed significantly from the downregulated group. Twenty-five genes were recovered with average AT usage > 0.50, and these were predominantly encoded in the large, low-GC region of Chromosome 1, that is, 50% GC content rather than the average (64%; supplementary dataset S3A, Supplementary Material online). When further comparing the genomic locations of the 1,185 upregulated and 625 downregulated genes at T12, 2.5% and 2.1%, respectively, were located in the Chromosome 1 low-GC region. These observations suggest that the bias in AT content may be partially explained by genomic location on the putative sex chromosome (i.e. Chromosome 1; Worden et al. 2009). Functional examination based on homology and domain search efforts indicated that about half the AT-enriched genes lack known domains and are specific to certain green algal groups or even *M. commoda* itself (supplementary dataset S3A, Supplementary Material online). Those that could be annotated appeared to be functionally diverse, and none belonged to known P-response mechanisms (Dyhrman 2016). These results suggest that the correlation between bias on the individual codon level and P limitation has no direct cause, and the observed AT bias may be indirectly induced by other factors. Alternatively, the genes recovered may have hitherto unknown functions in managing responses to P limitation.

Third Codon Position Biased Genes Are Highly Expressed Responders to P Limitation

We also investigated the more subtle bias at the third codon position that pairs with the ribonucleoside 34 at the wobble position of the tRNA anticodon. Specifically, we conducted a test of homogeneity of C usage across the T12-upregulated and T12-downregulated groups, to analyze usage of G versus C (and A vs. T) for each amino acid but only where either a C or a G (or an A or a T) was present at the third codon position. Six out of the 8 amino acids having codons that ended either in G or in C gave significant differences in proportions (Fig. 3a; supplementary dataset S3B, Supplementary Material online). All amino acids tested showed increased usage of C in the T12-upregulated group. In contrast, considering the usage of A versus T for the same 6 amino acids, only V gave a significant difference ($P = 0.01$), with A being used more frequently in the T12-upregulated group than the T12-downregulated group (supplementary dataset S3C, Supplementary Material online). The *M. commoda* genome encodes between 3 and 6 nuclear tRNAs for each of these 8 amino acids. For each of these 8 amino acids, except glycine (G) and serine (S), the majority of tRNAs recognize codons ending in A or T, suggesting preferential codon recognition possibly steered by tRNA modification (supplementary dataset S2C, Supplementary Material online). An alternative explanation is that tRNAs decoding codons ending in G or C are in general (or specifically under P limitation) more highly expressed in *M. commoda*, overcoming this imbalance. Together, these observations suggest a correlation between P limitation and codon usage at the third codon position specific to amino acids ending in G or C.

Again, we investigated which genes in terms of function or evolutionary relationships contributed most to the observed bias in the third codon position. We focused here on genes with an average C|GC usage > 0.80. In contrast to those biased in overall GC content (supplementary dataset S3A, Supplementary Material online), most genes enriched in C at the third codon position and upregulated at P limited (T12) were conserved among eukaryotes and functionally annotated (via orthologs; Fig. 3b; supplementary dataset S3D, Supplementary Material online). Further, the 30 recovered genes contained 3 of the top 10 differentially expressed transcripts in the entire experiment. These 3 encoded proteins clearly involved in acquisition of P: 2 members of the P:Na+ symporter (PNaS) family and an alkaline phosphatase. The latter was orthologous to the P-repressible alkaline phosphatase PHOX in the nonmarine green alga *Chlamydomonas reinhardtii* (Quisel et al. 1996; Moseley et al. 2006). The PHOX transcript increased $>14,000$-fold in our study, and the corresponding protein showed the second-highest FC in expression in the proteomic study of *M. commoda* (Guo et al. 2018). Phylogenetic reconstruction suggests PHOX originated through horizontal gene transfer (HGT) from bacteria to eukaryotes (supplementary fig. S4B, Supplementary Material online; see also below). Moreover, putative eukaryotic homologs of PHOX all came from organisms that reside in aquatic environments, indicating an optimal function of PHOX in these environments as already suggested by Grossman and Aksoy (2015). In contrast, the phylogeny of the PNaS3 and PNaS2 (upregulated $> 6,800$-fold and $>2,500$-fold, respectively) suggests PNaS is an ancient eukaryotic protein that

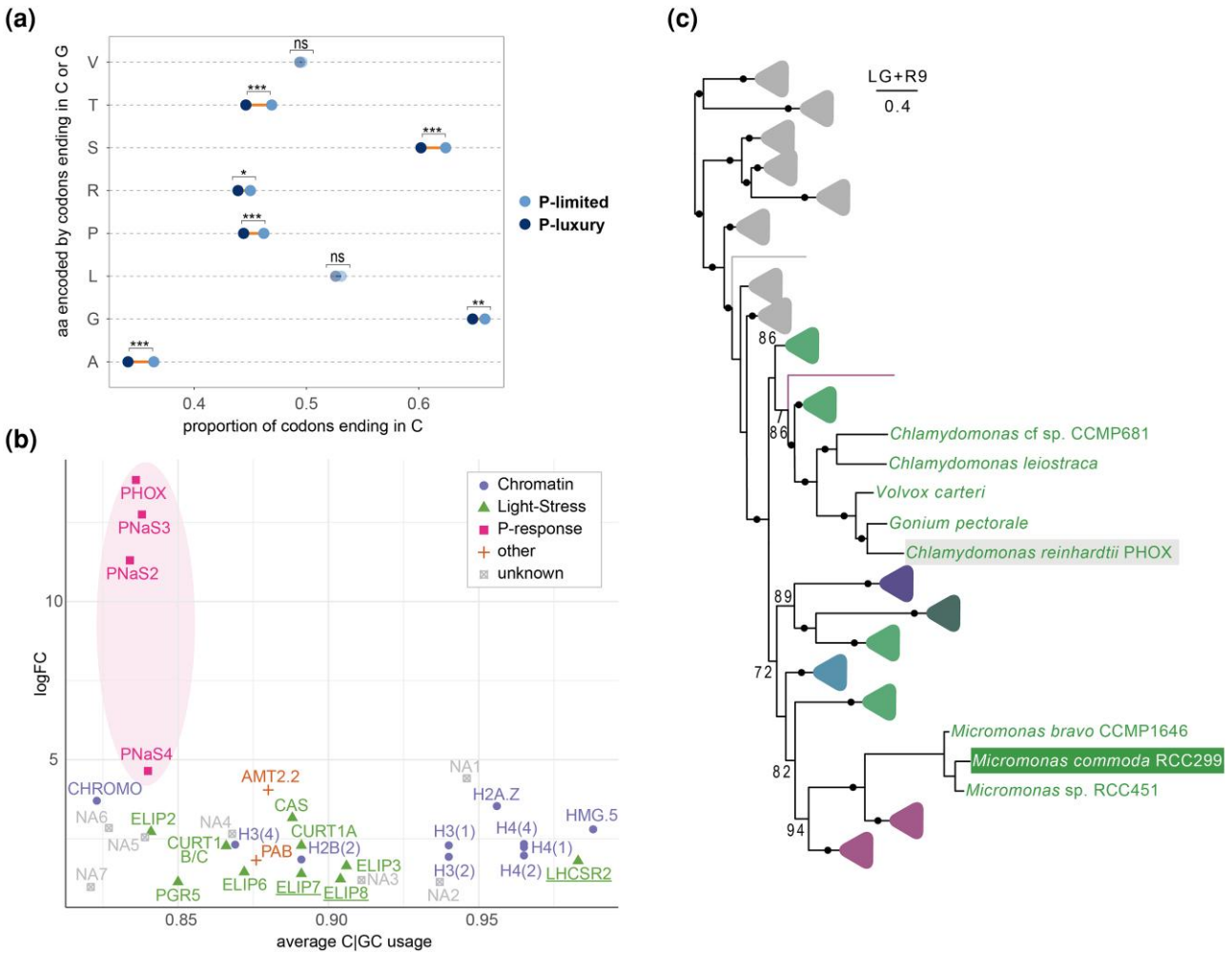**Fig. 3.** Under differing P conditions, codon usage varies at the third position, and highly differentially expressed genes with codon usage bias are involved in P scavenging. a) Dot plot showing, for each amino acid encoded by codons ending in C or G, the proportion of C at the third position for genes upregulated at T12 (P limited) and T39 (P luxury) for the comparison T12 versus T39. Statistical significances between the conditions were determined by testing for homogeneity; * indicates significant differences with $P < 0.05$, ** with $P < 0.005$ and *** with $P < 0.0005$; ns, not significantly different (supplementary dataset S3B, Supplementary Material online). b) Genes upregulated during T12 with average C|GC usage >0.80 and giving $P < 1e-5$ for a test of C|GC usage compared to the genes upregulated during T39 plotted against their logFC in the pairwise comparison T12 versus T39. Colors indicate functional groups (see supplementary dataset S3D, Supplementary Material online for full names). Underlined gene names indicate candidates with the most significant logFC, according to the associated *P*-value, in a pairwise comparison other than that between T12 and T39. NA, no functional annotation available. c) Schematic tree of the alkaline phosphatase PHOX enriched for C at the third codon position and highly upregulated at T12 (see b). The tree was inferred using the LG + R9 model, and node support was calculated using 1,000 ultrafast bootstrap (UFBoot) replicates. Only bootstrap support >70% is shown, with support ≥95% indicated by black dots. The scale bar and number beneath it indicate the estimated number of substitutions per site. Eukaryotic groups are the following: green, green algae/charophytes; dark green, red algae; turquoise, cryptophytes; dark purple, stramenopiles; purple, dinoflagellates/chromerids. Gray branches/clades are prokaryotic. The gray and green boxes highlight *C. reinhardtii* and *M. commoda* PHOX orthologs, respectively.

underwent expansion in some taxa while becoming lost from other lineages since (supplementary fig. S4C, Supplementary Material online). Although exhibiting lower FC between T12 and T39, another protein with C-bias at the third position was placed in this reconstruction, PNaS4 (25-fold upregulated; supplementary fig. S4C and dataset S3D, Supplementary Material online). Notably, 7 transcripts, including PHOX and PNaS4, with C|GC > 0.80 also exhibited high expression of the corresponding protein in a lower resolution proteomic study of P-limitation responses by *M. commoda* (Guo et al. 2018) (22 out of the 30 C|GC > 0.80 genes were recovered on the protein level; see supplementary dataset S3D, Supplementary Material online). Collectively,

these results point to a response to P limitation that hinges at least in part on the differential usage of nucleotides in the third codon position, underpinning a coordinated interaction between the shift in tRNA modifications, codon usage bias, and expression of proteins involved in acclimation to P limitation.

Third Codon Position Bias Associated with P Limitation Connects to Light Acquisition Management

In addition to high expression of genes encoding proteins that are directly involved in accomplishing growth under

P limitation, a group of "light stress–related" genes emerged as connected to the codon usage–based acclimation mechanism (Fig. 3b). Most of these show similar expression patterns across time points and photobioreactors, particularly during P limitation (Fig. 4a). A number of these were ELIPs, proteins affiliated with the light-harvesting complex that are specific to the green lineage (green algae and land plants; Bag 2021) and still have ill-characterized functional roles. Some accumulate under light and other stressors in plants (Rochaix and Bassi 2019). In *M. commoda*, a study of changes in light demonstrated that UV plus high light stress resulted in increased transcript levels for FAS-ELIP and CBR/ELIP7, while other ELIPs could not be distinguished methodologically (Cuvelier et al. 2017). Under P limitation, FAS-ELIP was also upregulated at the protein level in *M. commoda* (Guo et al. 2018). Also responding was LHCSR2 as well as the thylakoid Ca^{2+} sensor protein CAS, with the latter appearing to be exclusive to green algae and land plants (supplementary fig. S5A, Supplementary Material online; see also "Phylogenetic reconstructions of light stress response proteins" in the figshare repository at https://doi.org/10.6084/m9.figshare.21174397), indicating that, like ELIPs, this protein evolved within the green lineage. In *C. reinhardtii*, CAS is essential for the high light–induced expression of light-harvesting complex stress–related protein 3 (LHCSR3; Petroutsos et al. 2011). LHCSR3 is associated with both photosystems and induces nonphotochemical quenching (NPQ) via energy quenching qE (which dissipates excess energy as heat) under shifting light conditions (Peers et al. 2009) as well as under nitrogen, P, and sulfur deprivation conditions (Schmollinger et al. 2014). Likewise, in *M. commoda*, NPQ is accompanied by higher protein expression of LHCSR2 (orthologous to LHCSR3 in *C. reinhardtii*) during P limitation and has been proposed to play a key role in allowing the light-harvesting antennae to remain intact, poising the alga for increased growth when released from nutrient limitation (Guo et al. 2018). Herein, LHCSR2 transcripts were upregulated during the P-limited stage when F_v/F_m was significantly lower and NPQ was higher (Fig. 4; supplementary dataset S1A, Supplementary Material online). These results are similar to prior observations of LHCSR protein expression and coordinated NPQ (Guo et al. 2018); thus, upregulation of LHCSR2 appears to be a central mechanism for dissipating light energy as heat to minimize stress during reductions in use of photosynthate and overall growth under P-limiting conditions. Concordant expression of LHCSR2 and CAS, alongside information from model organisms, further suggests the latter regulates LHCSR2 in *M. commoda*.

A directed search for other candidates with putative roles in light and other stress responses recovered significant changes for PsbS and 3 additional ELIPs (CBR/ELIP1, 7, and 8), all of which were upregulated under P limitation (Fig. 4a and b). Similar to LHCSR2, PsbS can elicit an NPQ response in *Arabidopsis* and *Chlamydomonas* after exposure to high light (Li et al. 2002; Tibiletti et al. 2016). LHCSR2 and the 3 additional ELIPS were not included in our initial codon usage analysis, as their most significant logFC was between T12 and T29 or T31. When we investigated the C|GC usage for these 4 candidates, we found the codons for LHCSR2, CBR/ELIP7, and CBR/ELIP8 were strongly enriched in C at the third position (Fig. 3b; supplementary dataset S3E, Supplementary Material online). Finally, the codon-biased transcripts of LHCSR2, CBR/ELIP2, CBR/ELIP6, and CBR/ELIP8 also show increased expression of the corresponding protein under P-limiting conditions (supplementary dataset S3D and E, Supplementary Material online) in the prior proteomic study of *M. commoda* (Guo et al. 2018). These findings support the proposed interplay between tRNA modifications, codon usage bias, and regulation of stress-associated protein levels.

Cyclic electron flow (CEF) is a prerequisite for efficient NPQ via qE and is crucial for an effective response to changes in environmental conditions and stress in plants (Yamori and Shikanai 2016). CAS has also been reported to regulate the activity of a CEF "supercomplex" in *Chlamydomonas* (Terashima et al. 2012), connecting it to yet another light stress–related protein identified herein by its codon usage bias, the CEF-mediating protein PGR5. PGR5 mediates the more major of 2 known pathways for CEF in plants and green algae in association with PGRL1 (Fig. 4b) and is present in all photosynthetic lineages (supplementary fig. S5B, Supplementary Material online; see also "Phylogenetic reconstructions of light stress response proteins" in the figshare repository at https://doi.org/10.6084/m9.figshare.21174397). Although the specific role of PGR5 in CEF is still debated (Buchert et al. 2020), *C. reinhardtii* mutants lacking PGR5 display decreased NPQ during light stress, a consequence of the absence of CEF (Johnson et al. 2014). In conclusion, the identification of CAS and PGR5 transcripts and their specific upregulation during P limitation not only corroborates prior observations (Guo et al. 2018) that NPQ plays an important role in nutrient stress response mechanisms in *M. commoda* but also introduces a previously unrecognized role of CEF in P-limitation responses.

Shifts to P-Replete Conditions Trigger a Different Type of Light "Stress" Response than Limiting Conditions

The last group of putative light stress–related proteins recovered via their codon usage bias was 2 subunits of the CURT1 complex, that is, the curvature thylakoid 1 complex. The function of the CURT1 complex has only been described in plants where it localizes to the grana margins (Fig. 4b) and is involved in maintaining thylakoid plasticity and adjusting grana stack diameter under variable light conditions (Armbruster et al. 2013). These mechanisms, in turn, are a prerequisite for the efficient operation of processes depending on protein diffusion, such as PSII repair or state transitions (qT), another component of NPQ that involves partial relocation of light-harvesting complex II (LHCII) between PSI and PSII (Pribil et al. 2018). Intriguingly, we found the 2 protein kinases regulating
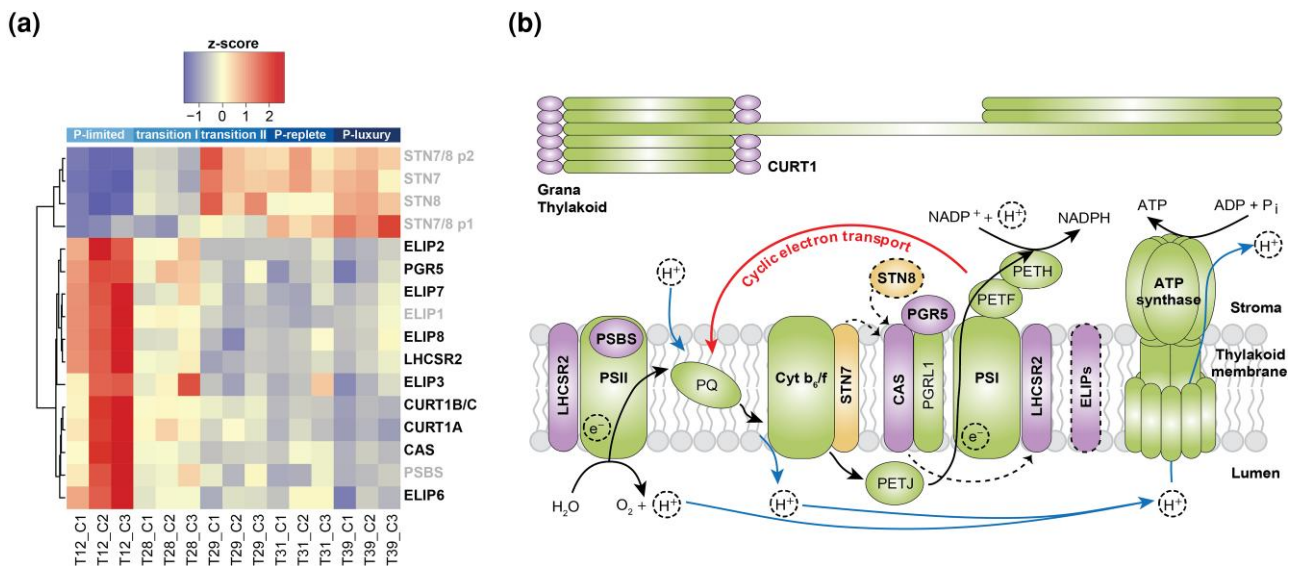
Fig. 4. Identification of strongly codon-biased genes reveals a concerted light management response during P limitation. a) Heatmap of hierarchically clustered z-scores for transcripts encoding proteins known to respond to light stress (and differentially expressed herein). Row labels indicate gene names. Column labels below the heatmap indicate sampling days (T12 to T39) and photobioreactor biological triplicates, Columns 1 to 3 (C1 to C3), while column labels correspond to Fig. 1. Gene names in black and gray text correspond to transcripts with average C|GC usage >0.80 and C|GC usage <0.80, respectively. p1, Paralog 1; p2, Paralog 2. b) Schematic representation of the electron transport chain in the thylakoid membrane including select *M. commoda* proteins involved in managing incoming light energy. Arrows indicate electron flow (black), CEF (red), and proton movement (blue). Fills indicate proteins upregulated during P limitation (purple) or afterwards (during transition II to P luxury; orange). Likely CAS mediation of LHCSR is indicated by a dashed arrow. Targeting of CAS by STN8 and STN7 is also indicated by dashed arrows, and STN8 and ELIPs are surrounded by dashed lines to indicate that the location of these proteins relative to the other shown thylakoid membrane proteins is not known.

these processes, STN8 and STN7, were also upregulated in our experiment, including 2 paralogs, but only under increasing P supply (T29 to T39; Fig. 4a). STN7, closely associated with the cytochrome b_6f complex (Shapiguzov et al. 2016; Fig. 4b), induces migration of LHCII from PSII to PSI (Depège et al. 2003), and STN8 is involved in PSII repair (Theis and Schroda 2016). Both also mediate phosphorylation of CAS, linking calcium signaling and protein phosphorylation (Fig. 4b; Stael et al. 2012; Cutolo et al. 2019). These 4 transcripts were not detected in the protein data of Guo et al. (2018) and were not enriched in C at the third codon position (supplementary dataset S3F, Supplementary Material online).

The increased abundance of the STN7 and STN8 transcripts and their close paralogs after P resupply suggests the possibility that the transition from limited to replete P triggers a mechanism in *M. commoda* that involves NPQ via state transition and photosystem repair mechanisms. Together, these observations suggest that the response involving light stress–associated genes does not end with the limiting P conditions. Rather, in response to increased P availability, the cell switches to a different mechanism. Collectively, these results establish that *M. commoda* reacts to limiting as well as replete P conditions with mechanisms often simplistically classified as "light stress response."

In phylogenies, STN7 and STN8 were found present across the green lineage and in lineages with secondary green plastids (supplementary fig. 5D, Supplementary Material online), suggesting another green invention.

However, cryptophytes are also embedded within the STN8 branch of the tree, indicating either emergence of this gene prior to the green lineage or a lateral gene transfer from the green lineage into the cryptophyte lineage. Additionally, for unicellular algae, there are only a very few green algal orthologs of the CURT1 complex subunits documented in GenBank. Our phylogenetic reconstructions revealed a cyanobacterial origin of CURT1, likely present in the ancestor of the Archaeplastida (comprising green algae, land plants, red algae, and glaucophytes), followed by putative duplication events and subsequent spread into lineages with secondarily acquired green algal plastids (supplementary fig. S5C, Supplementary Material online; see also "Phylogenetic reconstructions of light stress response proteins" in the figshare repository at https://doi.org/10.6084/m9.figshare.21174397). The restriction of the "light stress response" proteins discussed here to predominantly the green lineage (with the exception of the CEF protein PGR5) suggests that these types of light management responses to changing P conditions are largely specific to green algae and other marine algal lineages may use different sets of proteins to regulate photosynthesis under varying nutrient conditions.

Conclusion

Collectively, our studies indicate that the tRNA epitranscriptome acts as an acclimatization mechanism enabling continued growth under fluctuating conditions frequently encountered in open ocean environments. The observed

shifts in expression of numerous enzymes modifying tRNA nucleosides in *M. commoda* suggest reprogramming of the tRNA epitranscriptome in response to P limitation, in addition to known responses of proteins involved in P acquisition and other cellular processes. Our observations parallel the recently recognized stress response mechanism in nonphotosynthetic model organisms, in which stress-induced changes in the tRNA epitranscriptome are linked to codon-biased transcripts, ultimately leading to the preferred translation of putative "stress"-associated genes (Endres et al. 2015; Chionh et al. 2016; Chan et al. 2018).

For *M. commoda*, shifts in tRNA modifications under P-limiting conditions coincide with a bias in third position codon usage in upregulated transcripts. It should be noted that the nutrient limitation conditions used herein still support exponential growth, albeit slow growth, of *M. commoda*. This is in contrast to the extreme stress conditions tested thus far in other organisms where similar tRNA-related responses have been reported, e.g. hypoxia in bacteria (Chionh et al. 2016) and exposure to cytotoxins in yeast (Chan et al. 2010). For *Micromonas*, particularly biased genes were directly involved in stress alleviation and included others that contribute to light management, including some previously associated with P-limiting conditions in a proteomic study (Guo et al. 2018). Herein, as P limitation was alleviated, a different set of proteins involved in light management was expressed that do not display codon bias. Additionally, at least 1 gene identified in *Micromonas* that appears to result from HGT (i.e. PHOX) was ameliorated in a manner that accommodates control by the tRNA-based modulation. Our comparative phylogenetic analyses using data available from outside the green algae indicate other algal lineages possess a comprehensive tRNA modification enzyme complement, although more complete genomic and transcriptomic reference data is needed to recover complete information. Thus, we propose that the framework for a dynamic response to shifting nutrient conditions on the level of the tRNA epitranscriptome is important in distant marine algae as well.

We postulate that the identified response mechanism provisions a swift, flexible, and niche-specialized reaction to changes in nutrient availability. "Swiftness" would arise because modifications on the tRNA molecules can be set during several stages of tRNA processing (Berg and Brandl 2021), sidestepping the necessity of de novo biosynthesis of a new, different tRNA molecule necessary to shift translation toward required stress-involved proteins. Because the enzymes responsible for tRNA modifications are mostly conserved across the tree of eukaryotes, this may reflect a common nutrient stress response mechanism in eukaryotic algae. Future research must scrutinize a greater number of stressors, as well as pairing transcriptomic and proteomic data to examine this hypothesis more fully. Moreover, it remains unclear which modifications, if any, on the tRNA molecule are specifically affected by P or whether limitation by other resources results in

distinct modification patterns, as observed between desiccation, temperature, and salinity stress tested in rice and *Arabidopsis* (Wang et al. 2017). While the stress response mechanism we have identified appears to be conserved with those in animals and other nonmarine models, it identifies a new approach to adjusting growth under nonlethal nutrient fluctuations. Our studies highlight aspects of a possible conserved but hitherto unknown cell biological mechanism for acclimating to dynamic resource shifts in the ocean.

Materials and Methods

Micromonas Strain and Culturing

Axenic *M. commoda* RCC299 was grown in sterile L1 medium (-Si; Guillard and Hargraves 1993) in artificial seawater, with modified P concentration for experiments (see below). Cells were grown under a 14:10 h light:dark cycle at 21 °C. Cultures were monitored daily by flow cytometry using an Accuri C6 (BD Biosciences) before and during experiments. Prior to the bioreactor inoculation, semicontinuous batch cultures were acclimated and maintained in midexponential growth for 10 generations in sterile polystyrene flasks with 150 µmol photons m^{-2} s^{-1} photosynthetically active radiation (PAR). Experiments were performed in custom-built photobioreactor columns (Wilson et al. 2010; Guo et al. 2018) and run in biological triplicate, with 150 µmol photons m^{-2} s^{-1} PAR. The photobioreactors were operated in a way that incorporated aspects of turbidostats and chemostats, allowing us to capture algal responses to changing environmental factors (P in this case). First, they were inoculated with midexponential growth cells that had been switched from replete P (36 µM supplied as NaH$_2$PO$_4$.H$_2$O, as in standard L1 medium) to limited P (1.1 µM) after centrifugation. Briefly, precultures were spun (8,000 × *g*, 12 min) and, after discarding most of the supernatant, resuspended in the remaining volume and enumerated by flow cytometry. The cells were then diluted in L1 lacking P resulting in a final concentration of 1.1 µM PO$_4$ and 5.5 × 10^6 cells mL^{-1} and then added to the photobioreactors (final volume, 1.8 L per column). The initial concentration of amended P in the medium (in the sterile supply reservoir) was 1.5 µM and was adjusted on Days 8 and 27 (see below). Additionally, the pump rate was adjusted, based on daily flow cytometry counts, for example, to prevent washout of cells associated with reduced growth rates in the first days of the experiment.

Sample Collection, Growth Rates, and Cell Size

Cultures were sampled 1 h after lights on (9 AM) across the photobioreactor time course, flow cytometry samples were collected daily, and RNA as indicated in supplementary fig. S1, Supplementary Material online. For RNA, 90 mL was filtered onto 47-mm Supor 800 filters with 0.8-µm pore size. For residual soluble reactive phosphate (SRP) analysis, 45 mL was filtered through 47-mm

Durapore DVPP filters with 0.65-μm pore size (see supplementary fig. S1, Supplementary Material online for sampling days). RNA samples were flash frozen in liquid nitrogen, and all samples were stored at −80 °C (or −20 °C for SRP) until further processing. SRP samples were measured spectrophotometrically (Murphy and Riley 1962).

While the photobioreactor dilution rate sets the nominal growth rate during steady-state conditions, this approach is not valid for transitional periods. Hence, the reported growth rates (μ) were calculated using cell abundances and the following formula:

$$\mu = D_1 + \ \ln\left(\frac{N_1}{N_2}\right)/(T_2 - T_1)$$

where D_1 is the dilution rate between time point 1 (T_1) and time point (T_2), and N_1 and N_2 are cell numbers at T_1 and T_2, respectively. Growth rates began to decline between T0 and T2 and continuously dropped until Day 8, when the medium was amended to reach a 2 μM supply concentration. Growth mostly stabilized after Day 8, and the period from T9 to T27 represented acclimated P-limited growth (P-limited; T12 for the transcriptome analysis). Refeeding was accomplished by switching the supply from concentrations of 2 to 36 μM P on Day 27. The pump dilution rate D_1 was turned up on T30 (and again on T31, T33, and T40) from 0.16 to ultimately 0.65 d^{-1} (0.53 d^{-1} in Bioreactor 3) due to increasing cell abundances. Complete replacement of P-limited medium occurred on T31. T28 and T29 were defined as transitional stages (transition I and transition II), T31 represents the replete stage (P replete), and T39 the luxury stage (P luxury). It should be noted that because all sampling was performed within the context of the light:dark cycle, observed changes in cell size over the experimental phases could be separated from those occurring diurnally due to light-synchronized growth and division state (Cuvelier et al. 2017). Statistical significance between conditions was determined using one-way Analysis of varience (ANOVA) followed by Tukey's honest significance difference (HSD) tests to perform multiple pairwise comparisons.

RNA Sequencing and Analysis

RNA was extracted using the ToTALLY RNA kit (Ambion) adding a mechanical step involving glass beads (Duanmu et al. 2014). PolyA-selected RNA was isolated from 5 μg total RNA by 2 rounds of isolation, followed by purification and reverse transcription of fragmented RNA (200 to 300 nt) and further purification and amplification of cDNA fragments. Strand-specific, paired-end sequencing (150 bp each read) was performed on the Illumina HiSeq platform. On average, 36.45 million reads were sequenced per sample, with a SD of 3.13 million across the whole plate. Reads were aligned to the *M. commoda* RCC299 genome build v3 (Worden et al. 2009) using TopHat (Trapnell et al. 2009) v. 1.4.067, and a custom R script was used to convert and filter the TopHat results into an annotated count file used as

input for differential gene expression analysis with the edgeR package (Robinson et al. 2010). Genes (version "MicpuN3v2_GeneCatalog_genes_20160404.gff3" from van Baren et al. 2016 [available at https://mycocosm.jgi.doe.gov/MicpuN3v2/MicpuN3v2.home.html]) with count per million (CPM) $\geq$ 1 in at least 2 samples were further analyzed, and library size normalization was performed before and after filtering, resulting in 9,777 of 10,306 predicted protein-encoding genes being used for DE analysis. Analysis was performed with all time points sampled for RNA sequencing, and T12, T28, T29, T31, and T39 were selected as representing specific phases of P availability (see Fig. 1). Dispersion estimates and all versus all time point pairwise comparisons were performed with the general linearized model (glm) functionality of edgeR. Significantly differentially expressed genes were defined as genes with FC $\geq$ 2 and $P < 0.01$.

Clustering of z-scores of significantly expressed genes was based on k-means clustering, using the k-means function in R and estimating the optimal numbers of clusters using the wss (within-cluster sum of square) approach. The results were visualized using the heatmap3 function in R. Additionally, all genes present in the annotated count file were subjected to hierarchical clustering using the WGCNA package in R (Langfelder and Horvath 2008) with the automatic, 1-step network construction and module detection approach (chosen soft-thresholding power = 9). Identified modules were related to the following sample "traits": the 5 sampling time points (T12, T28, T29, T31, and T39), cell abundance, growth rate, FALS, and SRP as well as combined time points T12 and T28 ("limited combined" in supplementary fig. S2, Supplementary Material online) and T28 to T31 ("transition combined" in supplementary fig. S2, Supplementary Material online). The associations between modules and sample traits were calculated as Pearson's correlation coefficients.

Five gene sets were subjected to GO enrichment analysis: all 4 gene clusters resulting from k-means clustering (Fig. 1b) and all genes in the WGCNA module with the strongest positive correlation with time point T12 ($n = 2,030$; supplementary fig. S2, Supplementary Material online). Functional enrichment analysis was performed using the topGO package in R, testing with Fisher's exact test and taking the GO hierarchy into account with the parent–child algorithm (see supplementary dataset S1B to F, Supplementary Material online for complete results). GO terms used for the enrichment analysis were created by a local InterProScan (Blum et al. 2021) v. 5.31-70.0 analysis (using the -goterms option and running all analyses) of proteins from "MicpuN3v2_GeneCatalog_proteins_20160404" (https://mycocosm.jgi.doe.gov/MicpuN3v2/MicpuN3v2.home.html). Genes belonging to the most enriched GO term resulting from k-means clustering that were predominantly upregulated at T12 (Cluster 3, Fig. 1b) and in the WGCNA module with the strongest positive correlation with T12 (GO:0034470 "ncRNA processing" in both gene sets) were investigated using preliminary phylogenetic reconstructions with FastTree (Price et al. 2010; see below; supplementary

dataset S1G, Supplementary Material online). GO enrichment analysis results of those in *k*-means Cluster 3 was summarized and visualized using the R script provided by REVIGO (Supek et al. 2011). Enriched GO terms (BP) were clustered by semantic similarity using a similarity cutoff value of 0.7.

tRNA Modification Enzyme Search and Phylogenetic Reconstructions

Searches for tRNA modification enzymes other than those recovered by functional enrichment analysis initially used keyword searches for all pfam domains associated with tRNA modification enzymes (http://pfam.xfam.org) and used to query our *M. commoda* InterProScan analysis that also included a search against the Pfam database. All *M. commoda* candidate genes recovered were then subjected to BLASTP (Altschul et al. 1990) searches against GenBank to help distinguish between tRNA-modifying enzymes and enzymes acting on other noncoding RNAs. Additionally, we searched the literature for tRNA modification enzymes, leading to use of enzymes from *A. thaliana*, a yeast, or bacterial representative from Table 2 in Chen et al. (2010) and Table 1 in Hori (2014) to query the *M. commoda* proteins. Candidates from both approaches were used as queries to search for homologs.

Micromonas commoda candidate proteins were used as BLASTP queries against a custom database containing representatives from most major eukaryotic groups (390 data sets), with a focus on green algae (83 data sets), and RefSeq data from all bacterial phyla at NCBI. The database was subjected to CD-HIT (Fu et al. 2012) with a similarity threshold of 85% to reduce redundant sequences and paralogs, except for *M. commoda* proteins. BLASTP search results were generally parsed for hits with an *e*-value threshold ≤1e−25 and query coverage ≥50% to reduce the possibility of paralogs and short sequences. In cases where the query was distinctly longer than potential homologs (due to the often unusually long *M. commoda* homologs), or the query proteins were generally short or very divergent, the parsing parameters were adjusted in individual cases to recover sufficient homologous sequences for tree reconstruction. Bacterial hits were restrained to 20 per phylum (for FCB group, most classes of Proteobacteria, PVC group, Spirochaetes, Actinobacteria, Cyanobacteria [unranked], and Firmicutes) or 10 per phylum (remaining bacterial phyla) as defined by NCBI taxonomy. Parsed hits were aligned with MAFFT (Katoh and Standley 2013), using the --auto option, poorly aligned regions were eliminated using trimAl (Capella-Gutiérrez et al. 2009) v. 1.2 with a gap threshold of 80%, and maximum likelihood tree reconstructions were performed with FastTree v. 2.1.7 using the default options in a preliminary analysis. The resulting phylogenies and underlying alignments were inspected manually to remove contaminating, divergent, and/or low-quality sequences. When the recovered phylogenies contained a small number of taxa, parsing parameters were adjusted as described above, and additional queries (primarily described proteins of *A. thaliana* and *Homo sapiens* or in addition orthologs

from other genome-sequenced algae, such as the diatoms *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, or the cryptophyte *Guillardia theta*) were used in blast searches against our database. Similarly, when the recovered phylogenies contained additional *M. commoda* candidates (usually paralogs to more conserved proteins), those were also used as queries in searches against our database. BLAST results for all additional queries were combined with those for the initial *M. commoda* query and, after a de-duplication step, sequences were aligned, trimmed, and tree reconstruction, and inspection was performed. Cleanup and alignment/tree reconstruction steps were iterated as necessary. The final cleaned, unaligned sequences were subjected to filtering with PREQUAL using default options (Whelan et al. 2018) to remove nonhomologous residues introduced by poor-quality sequences, followed by alignment with MAFFT G-INS-i using the VSM option (−unalignlevel 0.6) to control overalignment (Katoh and Standley 2016). The alignments were subjected to Divvier (Ali et al. 2019) using the -divvygap and the -mincol 4 option to improve homology inference before removing ambiguously aligned sites with trimAl (-gt 0.01). Final trees were calculated with IQ-TREE v. 1.6.5 (Nguyen et al. 2015), using the -mset option to restrict model selection to LG for ModelFinder (Kalyaanamoorthy et al. 2017), while branch support used 1,000 ultrafast bootstrap replicates (Hoang et al. 2018). All other phylogenies represented by figures in the manuscript were reconstructed using IQ-TREE as described above, and the proteins in supplementary datasets S1 and S3, Supplementary Material online that were annotated via phylogenetic reconstruction were subjected to preliminary analysis, using the MAFFT (−auto), trimAl (-gt 0.8), and FastTree approach described above.

For heatmaps (Fig. 4a; supplementary fig. S3, Supplementary Material online), z-scores of selected subsets of transcripts were hierarchically clustered using the hclust function (agglomeration method: ward.D2) in the Pvclust package in R. The clustering results were visualized using the heatmap3 function in R.

For presence/absence comparisons with a *M. commoda* proteomics study (Guo et al. 2018), we used the 2,118 proteins that were matched by at least 2 peptides and analyzed further in that study. From those, 441 proteins met the coverage, significance, and FC criteria to be classified as differentially expressed, while here 9,777 transcripts were analyzed with 3,400 exhibiting DE. These differences preclude examination of many candidates of interest in the proteomic study.

Identification of the *M. commoda* tRNA Repertoire

To identify the tRNAs encoded in the *M. commoda* genome, we subjected the reference genome of *M. commoda* (ASM9098v2, downloaded from NCBI) to tRNAscan-SE 2.0 (Chan et al. 2021). We used the default option (-E) to search for eukaryotic tRNAs but also performed searches with the -B (search for bacterial tRNAs) and -O (other organellar tRNAs) and compared the search outputs. If

annotations differed, we chose the annotation with the highest Infernal Score as assigned by the sequence search engine Infernal 1.1 implemented in tRNAscan-SE 2.0.

Gene and Protein Naming Conventions

If available, user annotations at https://mycocosm.jgi.doe.gov/MicpuN3v2/MicpuN3v2.home.html were used. For unannotated *M. commoda* genes, we based naming on annotated orthologs—as listed in the UniProt database—that we recovered in our phylogenetic reconstructions. For the unannotated tRNA modification genes, we used conventions for yeast or bacterial orthologs, if available (see Fig. 2a; supplementary dataset S2, Supplementary Material online). For unannotated genes recovered via their codon usage bias (supplementary dataset S3, Supplementary Material online), we used the names of described orthologs (*A. thaliana* CAS; CURT1A, B, and C; PAB proteins; PGR5; STN7 and STN8; and *C. reinhardtii* PHOX) or the domain name (CHROMO). Where multiple recent *M. commoda* paralogs are present in a tree (the PNaS family), the genes were numbered according to their transcript identifier (see supplementary dataset S3, Supplementary Material online). For consistency with the nonitalicized domain names, and with all phylogenies being amino acid based, we use the nonitalicized versions of gene names throughout.

Codon Usage Analysis

To compare codon usage in genes upregulated versus downregulated at T12 (in relation to T39), we used specialized z-tests. For a given amino acid a and a given codon c corresponding to the amino acid a, the z-statistic is given by the following:

$$z_c = \frac{p_{cu} - p_{cd}}{\sqrt{p_c(1 - p_c) \cdot (1/n_{au} + 1/n_{ad})}}.$$

Here, n_{au} and n_{ad} give the number of times the amino acid a arose for the upregulated and downregulated groups. The proportions p_{cu} and p_{cd} are the proportions of times codon c arose for up- and downregulated groups but only among those sites where amino acid a was present. Using the proportions within amino acids controls for selection as it is amino acid usage that may differ between groups. The proportion p_c is the overall proportion of times codon c arose among all sites where amino acid a was present. Under the null hypothesis that codon usage of c is the same for both groups, z_c should be approximately normal. To adjust for multiple comparisons, we used a Bonferroni correction that replaces an α-level threshold by an α/m-level threshold where m is the number of tests; $m = 59$ in the case that z-tests are run for all codons except those, ATG (methionine) and TGG (tryptophan), that correspond to amino acids represented by a single codon.

For a given amino acid a, we used χ^2 tests to test whether codon usage distributions are the same over

both the upregulated and downregulated groups. The χ^2 statistic is as follows:

$$\chi^2 = \sum_j \sum_c (O_{cj} - E_{cj})^2 / E_{cj} \qquad (1)$$

Here, j indexes group (up/down) and the sum over codons c is over all codons corresponding to amino acid a. The observed number of times the codon c came up for group j is denoted O_{cj}. Under the null hypothesis that codon usage is the same for the 2 groups, the number of times that the codon c was expected to come up for group j is $E_{cj} = n_{aj}p_c$. Under the null hypothesis, χ^2 should have an approximate χ^2 distribution with degrees of freedom equal to the number of distinct codons (among 61) corresponding to amino acid a.

Tests of GC content effectively used the same χ^2 tests of homogeneity. To adjust for selection, tests were conducted separately for separate amino acids. The χ^2 statistic is as in Equation (1) but c is binary, indexing whether the codon ended in C or G ($c = 0$) or ended in A or T ($c = 1$). Thus, for instance, restricting attention to sites with amino acid a in the upregulated group, O_{0u} is the number of codons that ended in C or G. Similarly, O_{1d} is the number of codons in the downregulated group that ended in A or T, restricting attention to sites for that group with amino acid a.

Finally, considering sites for both groups with amino acid a, p_0 is the proportion of sites ending in C or G and p_1 the proportion ending in A or T, among sites where amino acid a was present.

A χ^2 test was also used to test for biased use of G within GC usage sites. To adjust for selection, tests were conducted separately for separate amino acids. For amino acid a, since interest was in whether G was used differentially given GC usage, attention was restricted to sites with amino acid a and G or C in the last codon position in calculating the O_{cj} and E_{cj} in Equation (1). Here, c indicates whether the codon ended in C or G. So, for group j (up/down), O_{Cj} is the number of sites ending in $c = C$ and O_{Gj} is the number of sites ending in $c = G$, restricting attention to sites with amino acid a. The expected is $E_{cj} = n_{aj}^{GC} p_c$, where n_{aj}^{GC} is the number of times amino acid a occurred with last nucleotide G or C in group j. The proportion p_c is the proportion of codons ending in c over both groups but restricting attention to codons corresponding to amino acid a that end in either a G or C.

To check genes responsible for the GC usage bias, we calculated the average proportion of GC. The null hypothesis is that GC usage for a gene in the upregulated group is the same as overall GC usage for the downregulated group. To adjust for selection, rather than use the proportion of codons that ended in G or C, we calculated that proportion separately for each amino acid and then averaged over amino acids. The reason for considering averages rather than separately considering GC usage bias for separate amino acids is to deal with sparseness concerns. It is often

the case that there are few instances of a given amino acid for a gene. The z-statistic is calculated as follows:

$$z = \{p_g^{GC} - p_d^{GC}\}/SE[p_g^{GC} - p_d^{GC}],$$

where p_g^{GC} gives the average GC proportion for gene g in the upregulated group and p_d^{GC} gives the average GC proportion for all genes present in the downregulated group. The SE is calculated from the following equation:

$$SE[p_g^{GC} - p_d^{GC}] = K_g^{-1}\sqrt{\sum_a p_a^{GC}[1 - p_a^{GC}]/n_{ag}}.$$

The sum is over the K_g distinct amino acids that appeared for gene g. The proportion p_a^{GC} is the overall proportion of times a codon ended in G or C among those corresponding to amino acid a and over both the gene and downregulated group. The count n_{ag} gives the number of times the amino acid a occurred for gene g in the upregulated group. The same equations were used to calculate z-statistics for usage of G given GC but with p_g^{GC} and p_d^{GC} replaced by the average proportions of codons ending in G among those ending in G or C and averaged over amino acids. Similarly, p_a^{GC} was replaced by the proportion ending in G among those ending in G or C for amino acid a. Finally, n_{ag} was replaced by the numbers of sites with amino acid a and ending in G or C.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank D. Klimov for support with photobioreactors and S. Sudek for general lab support as well as MBARI.

Author Contributions

J.G., S.W., and A.Z.W. designed the experiments; J.G., S.W., L.S., C.P., and A.Z.W. performed the experiments and sequencing; E.H., J.G., S.W., K.H., L.S., C.P., and R.D. analyzed the data; E.S. developed the statistical methods and analyzed the data; E.H. performed phylogenetic analyses; and E.H. and A.Z.W. wrote the paper.

Funding

Data Availability

RNA-seq data are available from the NCBI Short Read Archive (SRA) under the accession numbers SRX712380 to SRX712411. Phylogenetic reconstructions for all trees discussed in the manuscript have been deposited to figshare repositories under the DOIs 10.6084/m9.figshare.21164086 and 10.6084/m9.figshare.21174397.

References

Agris PF, Eruysal ER, Narendran A, Väre VYP, Vangaveti S, Ranganathan SV. Celebrating wobble decoding: half a century and still much is new. *RNA Biol.* 2018:**15**(4-5):537–553. https://doi.org/10.1080/15476286.2017.1356562.

Ali RH, Bogusz M, Whelan S, Tamura K. Identifying clusters of high confidence homologies in multiple sequence alignments. *Mol Biol Evol.* 2019:**36**(10):2340–2351. https://doi.org/10.1093/molbev/msz142.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990:**215**(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Armbruster U, Labs M, Pribil M, Viola S, Xu W, Scharfenberg M, Hertle AP, Rojahn U, Jensen PE, Rappaport F, et al. Arabidopsis CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. *Plant Cell* 2013:**25**(7):2661–2678. https://doi.org/10.1105/tpc.113.113118.

Bachy C, Wittmers F, Muschiol J, Hamilton M, Henrissat B, Worden AZ. The land–sea connection: insights into the plant lineage from a green algal perspective. *Annu. Rev. Plant Biol.* 2022:**73**(1):585–616. https://doi.org/10.1146/annurev-arplant-071921-100530.

Bag P. Light harvesting in fluctuating environments: evolution and function of antenna proteins across photosynthetic lineage. *Plants* 2021:**10**(6):1184. https://doi.org/10.3390/plants10061184.

Beenstock J, Sicheri F. The structural and functional workings of KEOPS. *Nucleic Acids Res.* 2021:**49**(19):10818–10834. https://doi.org/10.1093/nar/gkab865.

Behrenfeld MJ, O'Malley RT, Siegel DA, McClain CR, Sarmiento JL, Feldman GC, Milligan AJ, Falkowski PG, Letelier RM, Boss ES. Climate-driven trends in contemporary ocean productivity. *Nature* 2006:**444**(7120):752–755. https://doi.org/10.1038/nature05317.

Berg MD, Brandl CJ. Transfer RNAs: diversity in form and function. *RNA Biol.* 2021:**18**(3):316–339. https://doi.org/10.1080/15476286.2020.1809197.

Björkman KM, Duhamel S, Church MJ, Karl DM. Spatial and temporal dynamics of inorganic phosphate and adenosine-5'-triphosphate in the North Pacific Ocean. *Front. Mar. Sci.* 2018:**5**:1–14. https://doi.org/10.3389/fmars.2018.00235.

Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 2021:**49**(D1):D344–D354. https://doi.org/10.1093/nar/gkaa977.

Boccaletto P, MacHnicka MA, Purta E, Pitkowski P, Baginski B, Wirecki TK, De Crécy-Lagard V, Ross R, Limbach PA, Kotter A, et al. MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res.* 2018:**46**(D1):D303–D307. https://doi.org/10.1093/nar/gkx1030.

Bolaños LM, Karp-Boss L, Choi CJ, Worden AZ, Graff JR, Haëntjens N, Chase AP, Della Penna A, Gaube P, Morison F, et al. Small

phytoplankton dominate western North Atlantic biomass. *ISME J.* 2020:**14**(7):1663–1674. https://doi.org/10.1038/s41396-020-0636-0.

Bopp L, Monfray P, Aumont O, Dufresne JL, Le Treut H, Madec G, Terray L, Orr JC. Potential impact of climate change on marine export production. *Global Biogeochem Cycles.* 2001:**15**(1):81–99. https://doi.org/10.1029/1999GB001256.

Buchert F, Mosebach L, Gäbelein P, Hippler M. PGR5 is required for efficient Q cycle in the cytochrome b6f complex during cyclic electron flow. *Biochem J.* 2020:**477**(9):1631–1650. https://doi.org/10.1042/BCJ20190914.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009:**25**(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K. Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J.* 2009:**3**(3):283–295. https://doi.org/10.1038/ismej.2008.117.

Chan CTY, Deng W, Li F, Demott MS, Babu IR, Begley TJ, Dedon PC. Highly predictive reprogramming of tRNA modifications is linked to selective expression of codon-biased genes. *Chem Res Toxicol.* 2015:**28**(5):978–988. https://doi.org/10.1021/acs.chemrestox.5b00004.

Chan CTY, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* 2010:**6**(12):e1001247. https://doi.org/10.1371/journal.pgen.1001247.

Chan PP, Lin BY, Mak AJ, Lowe TM. TRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 2021:**49**(16):9077–9096. https://doi.org/10.1093/nar/gkab688.

Chan CTY, Pang YLJ, Deng W, Babu IR, Dyavaiah M, Begley TJ, Dedon PC. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat Commun.* 2012:**3**(1):937. https://doi.org/10.1038/ncomms1938.

Chan C, Pham P, Dedon PC, Begley TJ. Lifestyle modifications: coordinating the tRNA epitranscriptome with codon bias to adapt translation during stress responses. *Genome Biol.* 2018:**19**(1):1–11. https://doi.org/10.1186/s13059-018-1611-1.

Chen P, Jäger G, Zheng B. Transfer RNA modifications and genes for modifying enzymes in *Arabidopsis thaliana*. *BMC Plant Biol.* 2010:**10**(1):1–19. https://doi.org/10.1186/1471-2229-10-201.

Chionh YH, McBee M, Babu IR, Hia F, Lin W, Zhao W, Cao J, Dziergowska A, Malkiewicz A, Begley TJ, *et al.* TRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nat Commun.* 2016:**7**(1):1–12. https://doi.org/10.1038/ncomms13302.

Cognat V, Pawlak G, Pflieger D, Drouard L. PlantRNA 2.0: an updated database dedicated to tRNAs of photosynthetic eukaryotes. *Plant J.* 2022:**112**(4):1112–1119. https://doi.org/10.1111/tpj.15997.

Cutolo E, Parvin N, Ruge H, Pirayesh N, Roustan V, Weckwerth W, Teige M, Grieco M, Larosa V, Vothknecht UC. The high light response in *Arabidopsis* requires the calcium sensor protein CAS, a target of STN7- and STN8-mediated phosphorylation. *Front. Plant Sci.* 2019:**10**:1–15. https://doi.org/10.3389/fpls.2019.00974.

Cuvelier ML, Allen AE, Monier A, McCrow JP, Messié M, Tringe SG, Woyke T, Welsh RM, Ishoey T, Lee JH, *et al.* Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci U S A.* 2010:**107**(33):14679–14684. https://doi.org/10.1073/pnas.1001665107.

Cuvelier ML, Guo J, Ortiz AC, Van Baren MJ, Tariq MA, Partensky F, Worden AZ. Responses of the picoprasinophyte *Micromonas commoda* to light and ultraviolet stress. *PLoS One* 2017:**12**(3):e0172135. https://doi.org/10.1371/journal.pone.0172135.

Demory D, Baudoux AC, Monier A, Simon N, Six C, Ge P, Rigaut-Jalabert F, Marie D, Sciandra A, Bernard O, *et al.* Picoeukaryotes of the *Micromonas* genus: sentinels of a warming ocean. *ISME J.* 2019:**13**(1):132–146. https://doi.org/10.1038/s41396-018-0248-0.

Depège N, Bellafiore S, Rochaix JD. Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in *Chlamydomonas*. *Science* 2003:**299**(5612):1572–1575. https://doi.org/10.1126/science.1081397.

Duanmu D, Bachy C, Sudek S, Wong CH, Jiménez V, Rockwell NC, Martin SS, Ngan CY, Reistetter EN, Van Baren MJ, *et al.* Marine algae and land plants share conserved phytochrome signaling systems. *Proc Natl Acad Sci U S A.* 2014:**111**(44):15827–15832. https://doi.org/10.1073/pnas.1416751111.

Dyhrman ST. Nutrients and their acquisition: phosphorus physiology in microalgae. In: Borowitzka M Beardall J Raven J, editors. *The physiology of microalgae. Developments in applied phycology.* Vol. 6. Switzerland: Springer International Publishing; 2016. p. 155–183.

Dyhrman ST, Jenkins BD, Rynearson TA, Saito MA, Mercier ML, Alexander H, Whitney LAP, Drzewianowski A, Bulygin V V, Bertrand EM, *et al.* The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS One* 2012:**7**(3):e33768. https://doi.org/10.1371/journal.pone.0033768.

El Yacoubi B, Bailly M, De Crécy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet.* 2012:**46**(1):69–95. https://doi.org/10.1146/annurev-genet-110711-155641.

Endres L, Begley U, Clark R, Gu C, Dziergowska A, Małkiewicz A, Melendez JA, Dedon PC, Begley TJ. Alkbh8 regulates selenocysteine-protein expression to protect against reactive oxygen species damage. *PLoS One* 2015:**10**(7):e0131335. https://doi.org/10.1371/journal.pone.0131335.

Faktorová D, Nisbet RER, Fernández Robledo JA, Casacuberta E, Sudek L, Allen AE, Ares M, Aresté C, Balestreri C, Barbrook AC, *et al.* Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nat Methods.* 2020:**17**(5):481–494. https://doi.org/10.1038/s41592-020-0796-x.

Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012:**28**(23):3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Glatt S, Zabel R, Vonkova I, Kumar A, Netz DJ, Pierik AJ, Rybin V, Lill R, Gavin AC, Balbach J, *et al.* Structure of the Kti11/Kti13 heterodimer and its double role in modifications of tRNA and eukaryotic elongation factor 2. *Structure* 2015:**23**(1):149–160. https://doi.org/10.1016/j.str.2014.11.008.

Grossman AR, Aksoy M. Algae in a phosphorus-limited landscape. In: Plaxton WC, Lambers H, editors. *Annual plant reviews.* Volume 48. John Wiley & Sons, Ltd; 2015. p. 337–374.

Guillard RRL, Hargraves PE. *Stichochrysis immobilis* is a diatom, not a chrysophyte. *Phycologia* 1993:**32**(3):234–236. https://doi.org/10.2216/i0031-8884-32-3-234.1.

Guo J, Wilken S, Jimenez V, Choi CJ, Ansong C, Dannebaum R, Sudek L, Milner DS, Bachy C, Reistetter EN, *et al.* Specialized proteomic responses and an ancient photoprotection mechanism sustain marine green algal growth during phosphate limitation. *Nat. Microbiol.* 2018:**3**(7):781–790. https://doi.org/10.1038/s41564-018-0178-7.

Haley ST, Alexander H, Juhl AR, Dyhrman ST. Transcriptional response of the harmful raphidophyte *Heterosigma akashiwo* to nitrate and phosphate stress. *Harmful Algae* 2017:**68**:258–270. https://doi.org/10.1016/j.hal.2017.07.001.

Harke MJ, Juhl AR, Haley ST, Alexander H, Dyhrman ST. Conserved transcriptional responses to nutrient stress in bloom-forming algae. *Front Microbiol.* 2017:**8**:1279. https://doi.org/10.3389/fmicb.2017.01279.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018:**35**(2):518–522. https://doi.org/10.1093/molbev/msx281.

Hori H. Methylated nucleosides in tRNA and tRNA methyltransferases. *Front Genet.* 2014:**5**:1–26. https://doi.org/10.3389/fgene.2014.00144.

Ignatova V V, Kaiser S, Ho JSY, Bing X, Stolz P, Tan YX, Lee CL, Gay FPH, Lastres PR, Gerlini R, et al. METTL6 is a tRNA m3C methyltransferase that regulates pluripotency and tumor cell growth. Sci Adv. 2020:6(35):eaaz4551. https://doi.org/10.1126/sciadv.aaz4551.

Johnson X, Steinbeck J, Dent RM, Takahashi H, Richaud P, Ozawa SI, Houille-Vernes L, Petroutsos D, Rappaport F, Grossman AR, et al. Proton gradient regulation 5-mediated cyclic electron flow under ATP- or redox-limited conditions: a study of ΔATPase pgr5 and ΔrbcL pgr5 mutants in the green alga Chlamydomonas reinhardtii. Plant Physiol. 2014:165(1):438–452. https://doi.org/10.1104/pp.113.233593.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder:Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Katoh K, Standley DM. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. Bioinformatics 2016:32(13):1933–1942. https://doi.org/10.1093/bioinformatics/btw108.

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008:9(1):559. https://doi.org/10.1186/1471-2105-9-559.

Li XP, Müller-Moulé P, Gilmore AM, Niyogi KK. PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. Proc Natl Acad Sci U S A. 2002:99(23):15222–15227. https://doi.org/10.1073/pnas.232447699.

Liu Z, Koid AE, Terrado R, Campbell V, Caron DA, Heidelberg KB. Changes in gene expression of Prymnesium parvum induced by nitrogen and phosphorus limitation. Front Microbiol. 2015:6:1–13. https://doi.org/10.3389/fmicb.2015.00631.

Merchant SS, Helmann JD. Elemental economy: microbial strategies for optimizing growth in the face of nutrient limitation. In: Poole RK, editors. Advances in microbial physiology. Academic Press; Vol. 60. 2012. p. 91–210.

Monier A, Worden AZ, Richards TA. Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. Environ Microbiol Rep. 2016:8(4):461–469. https://doi.org/10.1111/1758-2229.12390.

Moseley JL, Chang CW, Grossman AR. Genome-based approaches to understanding phosphorus deprivation responses and PSR1 control in Chlamydomonas reinhardtii. Eukaryot Cell. 2006:5(1):26–44. https://doi.org/10.1128/EC.5.1.26-44.2006.

Murphy J, Riley JP. A modified single solution method for the determination of phosphate in natural waters. Anal Chim Acta.1962:27:21–36. https://doi.org/10.1016/S0003-2670(00)88444-5.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300.

Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, Hippler M, Niyogi KK. An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature 2009:462(7272):518–521. https://doi.org/10.1038/nature08587.

Petroutsos D, Busch A, Janßen I, Trompelt K, Bergner SV, Weinl S, Holtkamp M, Karst U, Kudla J, Hippler M. The chloroplast calcium sensor CAS is required for photoacclimation in Chlamydomonas reinhardtii. Plant Cell 2011:23(8):2950–2963. https://doi.org/10.1105/tpc.111.087973.

Pribil M, Sandoval-Ibáñez O, Xu W, Sharma A, Labs M, Liu Q, Galgenmüller C, Schneider T, Wessels M, Matsubara S, et al. Fine-tuning of photosynthesis requires CURVATURE THYLAKOID1-mediated thylakoid plasticity. Plant Physiol. 2018:176(3):2351–2364. https://doi.org/10.1104/pp.17.00863.

Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 2010:5(3):e9490. https://doi.org/10.1371/journal.pone.0009490.

Quisel JD, Wykoff DD, Grossman AR. Biochemical characterization of the extracellular phosphatases produced by phosphorus-deprived Chlamydomonas reinhardtii. Plant Physiol. 1996:111(3):839–848. https://doi.org/10.1104/pp.111.3.839.

Robinson MD, Mccarthy DJ, Smyth GK. Edger: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010:26(1):139–140. https://doi.org/10.1093/bioinformatics/btp616.

Rochaix JD, Bassi R. LHC-like proteins involved in stress responses and biogenesis/repair of the photosynthetic apparatus. Biochem J. 2019:476(3):581–593. https://doi.org/10.1042/BCJ20180718.

Schaub M, Keller W. RNA editing by adenosine deaminases generates RNA and protein diversity. Biochimie 2002:84(8):791–803. https://doi.org/10.1016/S0300-9084(02)01446-3.

Schmollinger S, Mühlhaus T, Boyle NR, Blaby IK, Casero D, Mettler T, Moseley JL, Kropat J, Sommer F, Strenkert D, et al. Nitrogen-sparing mechanisms in Chlamydomonas affect the transcriptome, the proteome, and photosynthetic metabolism. Plant Cell 2014:26(4):1410–1435. https://doi.org/10.1105/tpc.113.122523.

Shapiguzov A, Chai X, Fucile G, Longoni P, Zhang L, Rochaix JD. Activation of the Stt7/STN7 kinase through dynamic interactions with the cytochrome b6 f complex. Plant Physiol. 2016:171(1):82–92. https://doi.org/10.1104/pp.15.01893.

Stael S, Rocha AG, Wimberger T, Anrather D, Vothknecht UC, Teige M. Cross-talk between calcium signalling and protein phosphorylation at the thylakoid. J Exp Bot. 2012:63(4):1725–1733. https://doi.org/10.1093/jxb/err403.

Supek F, Bošnjak M, Škunca N, Šmuc T. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 2011:6(7):e21800. https://doi.org/10.1371/journal.pone.0021800.

Terashima M, Petroutsos D, Hüdig M, Tolstygina I, Trompelt K, Gäbelein P, Fufezan C, Kudla J, Weinl S, Finazzi G, et al. Calcium-dependent regulation of cyclic photosynthetic electron transfer by a CAS, ANR1, and PGRL1 complex. Proc Natl Acad Sci U S A. 2012:109(43):17717–17722. https://doi.org/10.1073/pnas.1207118109.

Theis J, Schroda M. Revisiting the photosystem II repair cycle. Plant Signal. Behav. 2016:11(9):1–8. https://doi.org/10.1080/15592324.2016.1218587.

Tibiletti T, Auroy P, Peltier G, Caffarri S. Chlamydomonas reinhardtii PsbS protein is functional and accumulates rapidly and transiently under high light. Plant Physiol. 2016:171(4):2717–2730. https://doi.org/10.1104/pp.16.00572.

Tragin M, Vaulot D. Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding. Sci Rep. 2019:9(1):1–14. https://doi.org/10.1038/s41598-019-41680-6.

Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009:25(9):1105–1111. https://doi.org/10.1093/bioinformatics/btp120.

van Baren MJ, Bachy C, Reistetter EN, Purvine SO, Grimwood J, Sudek S, Yu H, Poirier C, Deerinck TJ, Kuo A, et al. Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. BMC Genomics 2016:17(1):267. https://doi.org/10.1186/s12864-016-2585-6.

Wang Y, Pang C, Li X, Hu Z, Lv Z, Zheng B, Chen P. Identification of tRNA nucleoside modification genes critical for stress response and development in rice and Arabidopsis. BMC Plant Biol. 2017:17(1):1–15. https://doi.org/10.1186/s12870-017-1206-0.

Whelan S, Irisarri I, Burki F. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. Bioinformatics 2018:34(22):3929–3930. https://doi.org/10.1093/bioinformatics/bty448.

Wilson ST, Tozzi S, Foster RA, Ilikchyan I, Kolber ZS, Zehr JP, Karl DM. Hydrogen cycling by the unicellular marine diazotroph Crocosphaera watsonii strain WH8501. Appl Environ Microbiol. 2010:76(20):6797–6803. https://doi.org/10.1128/AEM.01202-10.

Worden AZ, Lee J, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett M V, et al. Green evolution and dynamic adaptations revealed by genomes of the marine

MBE

picoeukaryotes *Micromonas*. *Science* 2009:**324**(5924):268–272. https://doi.org/10.1126/science.1167222.

Worden AZ, Nolan JK, Palenik B. Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. *Limnol Oceanogr.* 2004:**49**(1):168–179. https://doi.org/10.4319/lo.2004.49.1.0168.

Wu J, Sunda W, Boyle EA, Karl DM. Phosphate depletion in the western North Atlantic Ocean. *Science* 2000:**289**(5480):759–762. https://doi.org/10.1126/science.289.5480.759.

Xu L, Liu X, Sheng N, Oo KS, Liang J, Chionh YH, Xu J, Ye F, Gao YG, Dedon PC, *et al.* Three distinct 3-methylcytidine (m3C) methyltransferases modify tRNA and mRNA in mice and humans. *J Biol Chem.* 2017:**292**(35):14695–14703. https://doi.org/10.1074/jbc.M117.798298.

Yamori W, Shikanai T. Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. *Annu. Rev. Plant Biol.* 2016:**67**(1):81–106. https://doi.org/10.1146/annurev-arplant-043015-112002.

Zhang SF, Chen Y, Xie ZX, Zhang H, Lin L, Wang DZ. Unraveling the molecular mechanism of the response to changing ambient phosphorus in the dinoflagellate *Alexandrium catenella* with quantitative proteomics. *J Proteomics.* 2019:**196**:141–149. https://doi.org/10.1016/j.jprot.2018.11.004.

Zhang W, Foo M, Eren AM, Pan T. tRNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes. *Mol. Cell.* 2022:**82**(5):891–906. https://doi.org/10.1016/j.molcel.2021.12.007.