


Shifting metabolic priorities among key protistan taxa within and below the euphotic zone

Sarah K. Hu ^{1*}, Zhenfeng Liu,¹
Harriet Alexander,² Victoria Campbell,³
Paige E. Connell,¹ Sonya T. Dyhrman,⁴
Karla B. Heidelberg¹ and David A. Caron¹

¹Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA.

²Department of Population Health and Reproduction, University of California Davis, Davis, CA, USA.

³Division Allergy and Infectious Diseases, UW Medicine, Seattle, WA, USA.

⁴Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA.

Summary

A metatranscriptome study targeting the protistan community was conducted off the coast of Southern California, at the San Pedro Ocean Time-series station at the surface, 150 m (oxycline), and 890 m to link putative metabolic patterns to distinct protistan lineages. Comparison of relative transcript abundances revealed depth-related shifts in the nutritional modes of key taxonomic groups. Eukaryotic gene expression in the sunlit surface environment was dominated by phototrophs, such as diatoms and chlorophytes, and high abundances of transcripts associated with synthesis pathways (e.g., photosynthesis, carbon fixation, fatty acid synthesis). Sub-euphotic depths (150 and 890 m) exhibited strong contributions from dinoflagellates and ciliates, and were characterized by transcripts relating to digestion or intracellular nutrient recycling (e.g., breakdown of fatty acids and V-type ATPases). These transcriptional patterns underlie the distinct nutritional modes of ecologically important protistan lineages that drive marine food webs, and provide a framework to investigate trophic dynamics across diverse protistan communities.

Introduction

Natural assemblages of microbial eukaryotes are dominated by a huge diversity of protistan species. Protists fulfill multiple roles in marine food webs due to their diverse morphologies, behaviours and metabolic capabilities (De Vargas *et al.*, 2015; Worden *et al.*, 2015). Understanding their nutritional strategies, including phototrophy, heterotrophy and mixotrophy, is essential for characterizing their ecological interactions, responses to environmental conditions and for modelling the emergent properties of these communities. Recent global surveys of protistan diversity based on molecular approaches have uncovered higher species richness at all depths of the water column (De Vargas *et al.*, 2015; Pernice *et al.*, 2016). Understanding the *in situ* functional roles of these species is necessary to assess their roles in marine biogeochemical cycles.

Transcriptomic analyses have augmented traditional physiological studies of protists by identifying the core genes and pathways that serve as a way to determine if an organism is reliant on a primarily autotrophic or heterotrophic mode of nutrition (Koid *et al.*, 2014; Liu *et al.*, 2016; Beisser *et al.*, 2017). Primarily photoautotrophic modes of nutrition in phytoplankton are generally characterized by higher expression of genes related to photosynthetic machinery and the downstream fixation and intracellular partitioning of carbon in the cell (Smith *et al.*, 2012).

Further, transcript-based studies have revealed information on the metabolic adaptations of phytoplankton to specific environmental conditions, such as light or nutrient availability (e.g., Moustafa *et al.*, 2010; Beszteri *et al.*, 2012; Dyhrman *et al.*, 2012; Frischkorn *et al.*, 2014; Liu *et al.*, 2015a; Harke *et al.*, 2017).

Compared with phototrophy, the metabolic pathways for phagotrophic ingestion of prey are not as well documented, but recent work on mixed nutrition in protists (combined phototrophy and heterotrophy; mixotrophy) has identified genes enriched in organisms dependent on prey ingestion. These studies have included investigations of organelle sequestration (e.g., Johnson *et al.*, 2007; Johnson, 2015; Lasek-Nesselquist *et al.*, 2015), endosymbioses (Balzano *et al.*, 2015) and phytoflagellate mixotrophy (Liu *et al.*, 2015b, 2016; Lie *et al.*, 2017; 2018). For instance, increased expression of genes associated with the catabolic

Received 11 November, 2017; revised 24 April, 2018; accepted 26 April, 2018. *For correspondence. E-mail sarah.hu@usc.edu; Tel. +(213) 821-1800, Fax: (213) 740-8123

breakdown of compounds or vacuolar activity in mixotrophic species provided putative target genes for a heterotrophic physiology (Liu *et al.*, 2016). Beisser *et al.* (2017) found heterotrophic strains of chrysophyte species to have reduced expression of genes integral for carbon fixation and photosynthetic machinery and increased expression of genes involved in the breakdown and adsorption of ingested organic material.

Metatranscriptomic surveys of natural communities of microbial eukaryotes have recently become feasible with advances in sequencing capability and improved reference databases, and are rapidly becoming a tool for probing the 'black box' of environmental microbial ecology. This approach has enabled the identification of shifts in metabolic potential of multiple taxonomic groups simultaneously with respect to environmental forcing factors (e.g., coastal vs. oligotrophic conditions, light or salinity gradients; Alexander *et al.*, 2015a; Aylward *et al.*, 2015; Dupont *et al.*, 2015; Pearson *et al.*, 2015; Grossmann *et al.*, 2016; Zielinski *et al.*, 2016).

To date, *in situ* protistan-specific metatranscriptomic surveys have largely focused on communities within the euphotic zone (Marchetti *et al.*, 2012; Alexander *et al.*, 2015a,b; Dupont *et al.*, 2015; Carradec *et al.*, 2018). Efforts to characterize subsurface protistan diversity have demonstrated the important contribution that these species make in sub-euphotic, low-oxygen microbial food webs (Stoeck *et al.*, 2014; Edgcomb, 2016; Pernice *et al.*, 2016). Previous work has found evidence for transcriptionally active unicellular eukaryotes (mainly fungi) in deep hypersaline anoxic basins based on the presence of actin and tubulin transcripts (Edgcomb *et al.*, 2016) and other subsurface environments (Orsi *et al.*, 2013). However, details of the metabolic potential of protistan species at water

column depths below the euphotic zone still remain poorly understood, warranting further exploration.

Gene expression profiles in and below the euphotic zone at the San Pedro Ocean Time-series (SPOT) station were used to examine the distribution of ecologically significant nutritional modes in five taxonomic groups: dinoflagellates, ciliates, haptophytes, diatoms and chlorophytes. The surface environment was characterized by synthesis-related pathways such as photosynthesis and downstream fixation of carbon, while below the euphotic zone, at 150 and 890 m, there were higher abundances of transcripts associated with lysosomes, and vacuolar type ATPases, and the breakdown of fatty acids. Depth-related differences in transcript abundance (based on several replicates) allowed us to form molecular-level descriptions of *in situ* nutritional strategies among dominant taxonomic groups, which included evidence for anaerobic metabolism among ciliates at 150 and 890 m, and adaptation to prolonged darkness in diatoms and chlorophytes. Exploring the physiological traits of dominant protists throughout the water column allowed us to infer depth-specific roles of protistan-mediated production and turnover of carbon.

Results

Environmental parameters and cell counts

Water temperatures at the SPOT station (Fig. 1A) in May were 17.1°C at the surface, 9.6°C at 150 m and 5.2°C at 890 m (Fig. 1B). Chlorophyll *a* fluorescence was 0.03 µg/L at the surface, decreasing to un-detectable (< 0.01 µg/L) at ≈ 80 m (Fig. 1B). Oxygen levels at the surface were 4.95 ml/L, relatively hypoxic at 150 m (1.98 ml/L) and nearly anoxic at 890 m (0.15 ml/L; Fig. 1B, Supporting Information Table S4). All nutrient concentrations (µM) were low at the

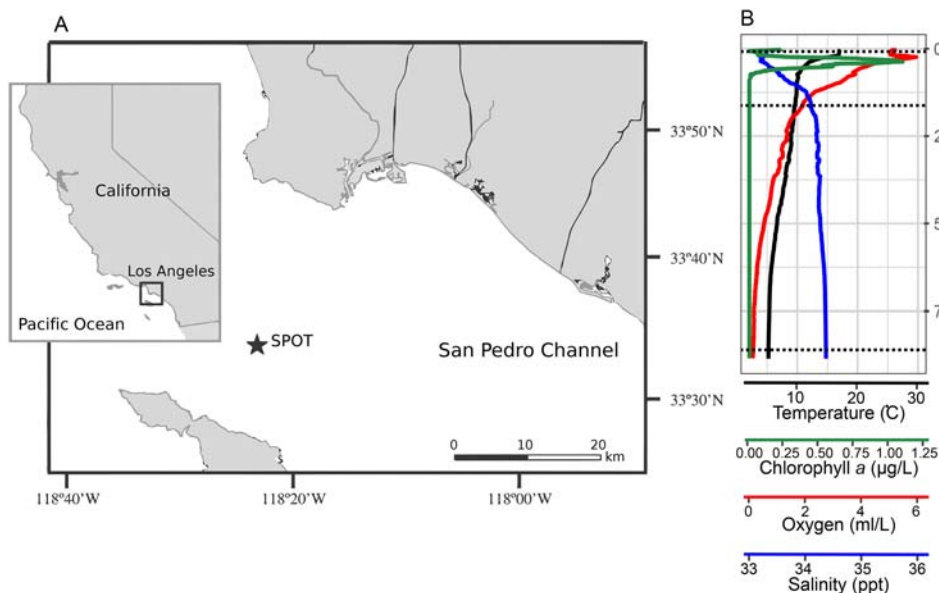


Fig. 1. **A.** Map of San Pedro Ocean Time-series (SPOT) station, located ≈ 20 km from Los Angeles in the San Pedro Channel (33° 33' N, 118° 24' W). Samples were collected from the surface (5 m), 150 m, and 890 m in late May. **B.** Vertical profile based on CTD cast data, including temperature (°C), chlorophyll *a* fluorescence (µg/L), oxygen (ml/L) and salinity (ppt). Dotted horizontal lines indicate sampling depths.

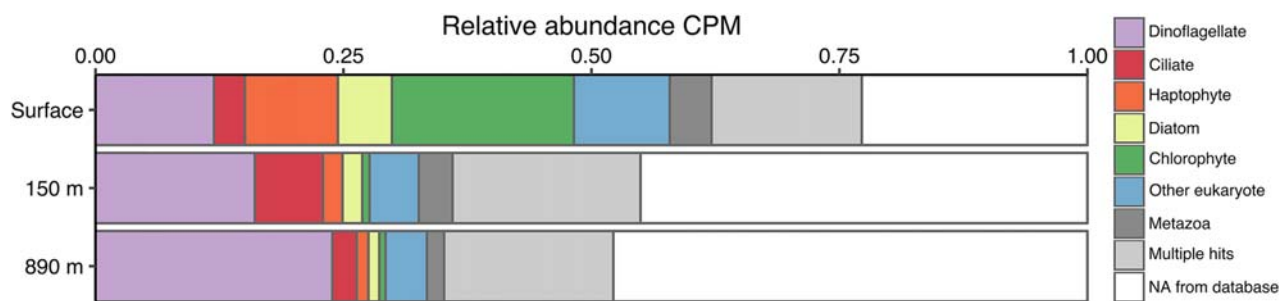


Fig. 2. Average (across replicates) relative abundance of transcript counts per million (CPM). Contigs were assigned a taxonomic identity based on a custom database (see section on ‘Experimental procedures’). Contigs that hit more than one reference with no consensus among the references were labelled ‘Multiple hits’. When no reference was hit, the contig was labelled ‘NA from database’. There were more total hits (excluding ‘NA from database’ and ‘Multiple hits’) to the reference database in the surface sample (62%), relative to 150 m (36%) and 890 m (35%). The ‘Other eukaryotes’ category was comprised of taxonomic groups that made up less than 2% of the total composition of contigs; a full summary of ‘Other eukaryote’ can be found in Supporting Information Table S6.

surface, consistent with the oligotrophic nature of the site, relative to values at 150 and 890 m (Supporting Information Table S4).

Microscopic cell counts of all microbial groups decreased from the surface to 150 and 890 m (Supporting Information Table S4). Microplankton (20–200 μm), pigmented (phototrophic and mixotrophic) and heterotrophic nanoplankton (2.0–20 μm) and picophytoplankton (0.2–2.0 μm), including picoeukaryotes, *Synechococcus*, *Prochlorococcus*, were highest at the surface (Supporting Information Table S4), and representative of spring-time abundances for these assemblages (Caron *et al.*, 2017).

Annotation

Assemblies from each depth yielded a total of 7.8 million contigs, generating 5.4 million putative protein sequences (Supporting Information Table S5). Contigs were clustered into \approx 3.9 million unique ortholog groups (4.2 million total; Supporting Information Table S5). 40.5% of the contigs were assigned taxonomic identities (excluding ‘NA from database’ and ‘Multiple hits’) and 25% of the contigs were assigned KEGG identities.

Taxonomic composition

Taxonomic assignments of transcripts revealed the top groups that contributed to metabolic potential (Fig. 2, Supporting Information Table S6). The total number of transcripts successfully assigned taxonomy was highest from the surface community (62%), nearly double the percentage at 150 or 890 m (36% and 35% respectively; Fig. 2, Supporting Information Table S6). Major taxonomic groups that made up $>$ 2% of transcript counts per million (CPM) across all depths were: dinoflagellates (17%), ciliates (4.5%), haptophytes (3.7%), diatoms (2.6%) and chlorophytes (5.3%) (Fig. 2, Supporting Information Table

S6). The surface was mainly comprised of dinoflagellate, haptophyte and chlorophyte transcripts ($>$ 9% each; Fig. 2, Supporting Information Table S6). Below the euphotic zone (150 and 890 m), the majority of protistan transcripts were dinoflagellates (16%–24%; Fig. 2, Supporting Information Table S6). The relative abundance of ciliate transcripts at 150 m (6.9%) was more than twice the percentage found at the surface or 890 m (3.1% and 2.5% respectively; Fig. 2, Supporting Information Table S6). Taxonomic groups that made up less than 2% of the community were collapsed into the ‘Other eukaryote’ category but are summarized in Supporting Information Table S6.

The majority of rRNA reads (5.8% total sequence reads; Supporting Information Table S5) were found to be protistan (28%–50%; Supporting Information Fig. S1A). Metagenomic Illumina tag (miTag, as defined by Logares *et al.*, 2014) results at the class-level and with more than 500 count were reported in Supporting Information Fig. S2 (also see Supporting Information Table S7), and the distribution of miTags at each depth was shown in Supporting Information Fig. S3. However, we refrain from over interpreting the results, as sequence lengths shorter than 400 bps provide less accurate assessments of protistan diversity (Hu *et al.*, 2015).

Ortholog groups

More than 1 million ortholog groups were unique to each sampling depth (Fig. 3). Of the contigs making up these ortholog groups, 13%–19% had KEGG identities (Fig. 3B). Clustering revealed a higher number of shared ortholog groups between 150 and 890 m relative to other comparisons ($>$ 120 000 ortholog groups; Fig. 3B). More than half of the ortholog groups shared among all depths were given functional assignments (right-most bar: 62.9%; Fig. 3B), which mainly belonged to the KEGG Class 2 category,

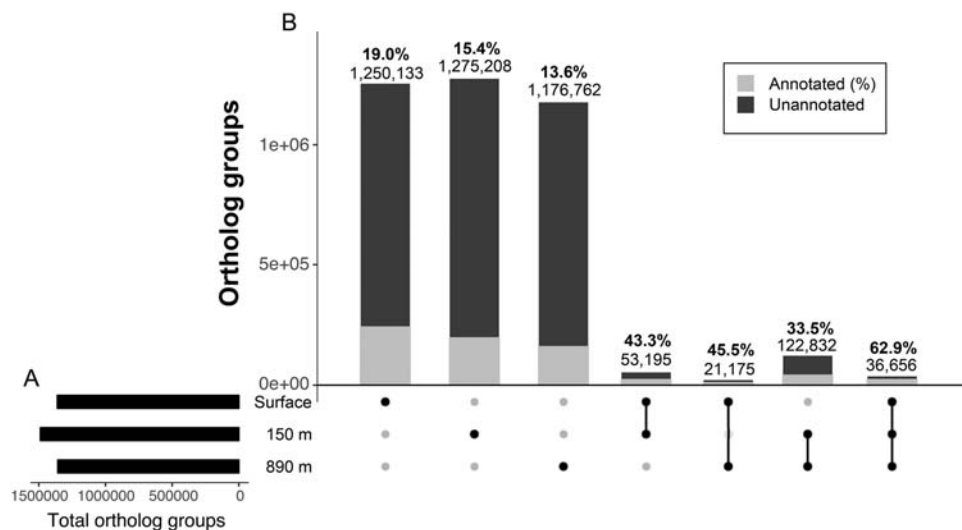


Fig. 3. Distribution of ortholog groups among depths sampled. **A.** There were a total of 3.9 million ortholog groups found at the surface, 150 and 890 m (top to bottom). **B.** Bar plots show the total number of shared or unique ortholog groups across the three depths. Depths included in a set are represented by filled dots in the matrix below each bar. Most ortholog clusters were unique to a single depth (three left-most bars), although 36 656 orthologous groups were common across depths (right most bar). The number of ortholog groups that contain at least one contig with an annotated KEGG identity are represented by light grey (also given as percentages). Plot generated using 'UpSetR' R package Conway *et al.* (2017).

'Genetic information processing' (Supporting Information Fig. S4).

Metabolic potential of the whole community with respect to depth

Community-wide gene expression of targeted pathways at the surface was distinct from that observed from the two depths below the euphotic zone (Fig. 4). Photosynthesis, carbon fixation (Calvin cycle, glycolysis and gluconeogenesis), pyruvate dehydrogenase (PDH) and fatty acid biosynthesis related transcripts were significantly more abundant at the surface relative to deeper depths (Fig. 4, upper portion of ternary plot; Supporting Information Table S8; $p < 0.05$). Transcripts associated with fatty acid breakdown, TCA and glyoxylate cycles and vacuolar-type (V-type) ATPases had similar abundances at all three depths (centre of plot; Fig. 4). Pyruvate-ferredoxin oxidoreductase (por), lysosome-associated genes and chitinase genes were significantly more abundant at 150 and/or 890 m relative to the surface (Fig. 4, lower portion of ternary plot; Supporting Information Table S8; $p < 0.05$).

Taxon-specific trends in metabolic potential

Ordination analysis of transcripts with KEGG annotations shared among taxonomic groups revealed that samples clustered by taxonomic group and depth. Dinoflagellates and ciliates clustered regardless of depth, while haptophytes, diatoms and chlorophytes separated by depth (Fig.

5). Haptophytes, diatoms and chlorophytes at the surface clustered away from all 150 and 890 m samples (Fig. 5). The 150 and 890 m samples from haptophytes, diatoms and chlorophytes more closely clustered with dinoflagellates (Fig. 5).

In order to examine depth-specific shifts in expressed genes by taxon, transcript counts were separated by taxonomic group and re-normalized in edgeR (see section on 'Experimental procedures'). Functions with similar transcript abundances at each depth, indicative of unchanging metabolism, were positioned centrally in each ternary plot (Fig. 6). The majority of dinoflagellate transcripts were similar in abundance at all depths (Fig. 6A, centre of plot), with exceptions including nitrate/nitrite uptake (NRT), which was significantly higher among dinoflagellates at the surface and 890 m relative to 150 m, and photosynthesis-related genes expressed at the surface (Fig. 6A, Supporting Information Table S9; $p < 0.05$). Ciliate relative transcript abundance was highest at 150 m (Figs. 2A and 6B). Below the euphotic zone, ciliate transcripts that were significantly more abundant relative to the surface included fatty acid metabolism, chitinase, TCA and glyoxylate cycles, por, lysosome and V-type ATPases (Fig. 6B, Supporting Information Table S9; $p < 0.05$).

Distinct patterns of gene expression were observed between the surface and sub-euphotic zone samples for haptophytes. Haptophyte species at the surface had significantly higher abundances of photosynthesis, NRT, Calvin cycle, glycolysis, gluconeogenesis and PDH transcripts (Fig. 6C, Supporting Information Table S9; $p < 0.05$).

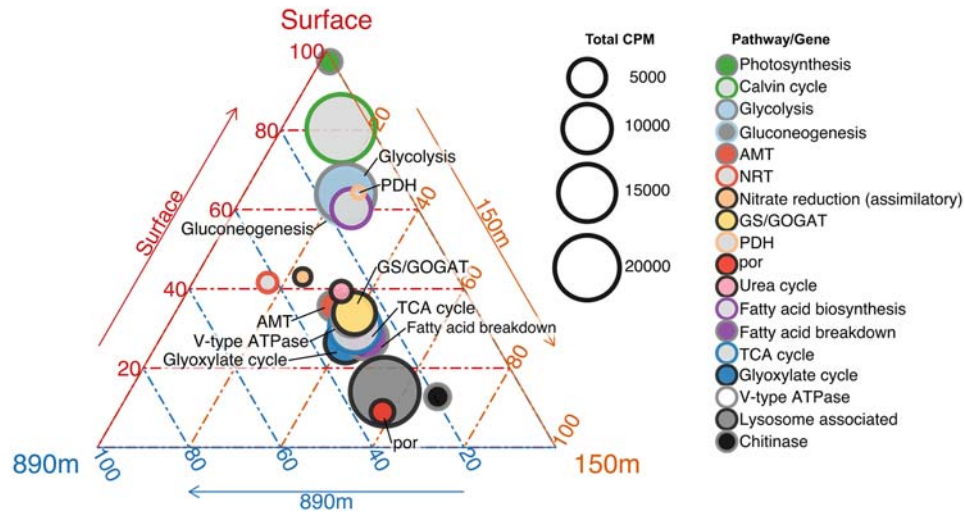


Fig. 4. Ternary plot of relative transcript abundance in counts per million (CPM) with respect to depth for key metabolic pathways (colours). Circle placement is representative of relative transcript abundance among the three depths (one depth per axis, clockwise with increasing abundance) and circle size is proportional to the total CPM for the three depths. A full list of genes and pathways is in Supporting Information Table S3. Related statistical analyses presented in Supporting Information Table S7. Plot generated using ggtern R package, Hamilton (2016). Abbreviations: AMT, ammonium transporter (K03320); NRT, nitrate/nitrite transporter (K02575); GS/GOGAT, glutamate synthase/glutamine oxoglutarate aminotransferase pathway; PDH, pyruvate dehydrogenase subunits alpha (K00161) and beta (K00162); por, pyruvate-ferredoxin/flavodoxin oxidoreductase (K03737).

Below the euphotic zone, dominant haptophyte transcripts consisted of higher abundances of urea cycle ($p < 0.05$), lysosome ($p < 0.05$), chitinase ($p < 0.05$), glyoxylate cycle and glutamate synthase/glutamine oxoglutarate aminotransferase pathway (GS/GOGAT) associated transcripts (Fig. 6C, Supporting Information Table S9).

Diatom and chlorophyte species at the surface had significantly more abundant transcripts associated with photosynthesis, Calvin cycle, glycolysis and gluconeogenesis (Fig. 6D,E, Supporting Information Table S9; $p < 0.05$). Numbers of diatom and chlorophyte reads were low at 150 and 890 m (Fig. 2, pie charts in Fig. 6D,E), where metabolisms were characterized by V-type ATPases, chitinase, lysosome, TCA cycle and glyoxylate cycle related genes (Fig. 6D,E).

Discussion

A metatranscriptome study was conducted from a vertical depth profile at the San Pedro Ocean Time-series (SPOT) station in order to gain a better understanding of the ecological activities of marine protists throughout the water column. This community-wide view of gene expression compared the sunlit surface versus sub-euphotic zone depths to highlight the metabolic flexibility or adaptations of protistan lineages to alternate nutritional strategies. High relative abundances of transcripts associated with core carbon fixation pathways and fatty acid synthesis indicated an overwhelming contribution of phytoplankton at the surface, with strong contributions by dinoflagellates,

haptophytes, diatoms and chlorophytes (Figs. 2 and 4). Transcripts associated with the breakdown of organic carbon compounds and intracellular nutrient cycling dominated at 150 and 890 m, particularly among

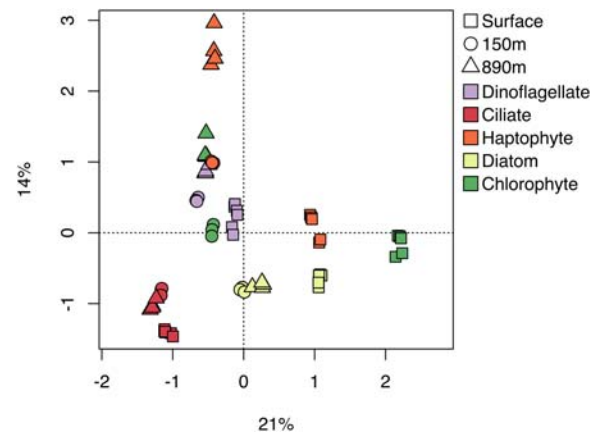


Fig. 5. Ordination analysis (constrained correspondence analysis [CCA]) plot shows the placement of major taxonomic groups at each depth based on normalized transcript counts per million (CPM). Contigs included in this analysis had common KEGG identities among all five taxonomic groups ($n = 1603$). Transcript CPM was normalized by sample to the total number of transcripts belonging to each taxonomic group (colours). Shapes depict depth at the SPOT station: squares – surface, circles – 150 m and triangles – 890 m. Replicate samples are shown as the same symbol and colour; there were 6 replicates from the surface, 3 from 150 m, and 4 from 890 m (Supporting Information Table S1). The x- and y-axes demonstrate the percent variance explained by the data.

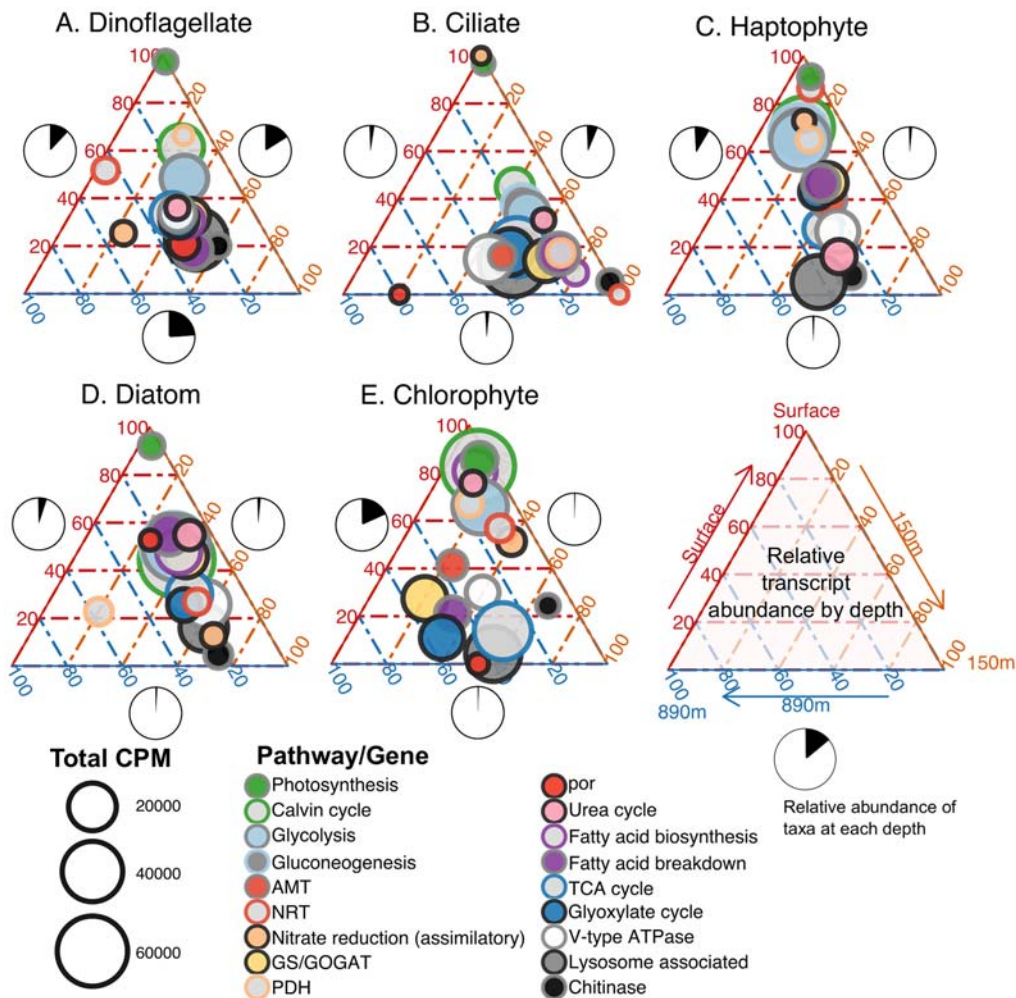


Fig. 6. Ternary plots of the relative transcript abundance in counts per million (CPM) with respect to depth for key metabolic pathways (colours, listed in Supporting Information Table S3) found in each taxonomic group: **(A)** dinoflagellates, **(B)** ciliates, **(C)** haptophytes, **(D)** diatoms and **(E)** chlorophytes. Relative CPM was normalized to taxonomic group and depth, white and black pie charts represent the relative abundance of taxa at each depth (Fig. 2). Circle placement is representative of transcript abundance relative to the three depths (one depth per axis, clockwise with increasing abundance) and circle size is proportional to the total CPM. A full list of genes and pathways is in Supporting Information Table S3. Related statistical analyses presented in Supporting Information Table S8. Abbreviations: AMT, ammonium transporter (K03320); NRT, nitrate/nitrite transporter (K02575); GS/GOGAT, glutamate synthase/glutamine oxoglutarate aminotransferase pathway; PDH, pyruvate dehydrogenase subunits alpha (K00161) and beta (K00162); por, pyruvate-ferredoxin/ferredoxin oxidoreductase (K03737).

dinoflagellates and ciliates. Lysosome-associated enzymes are involved in many cellular processes, including the internal breakdown and recycling of biomolecules (Eskelinen and Saftig, 2009; Settembre *et al.*, 2013), while V-type ATPases play a role in digestive processes by lowering the pH in phagosomes (Finbow and Harrison, 1997). Distinct transcriptional profiles with respect to depth demonstrated the variable roles protists contribute to elemental cycling throughout the water column at the SPOT station.

Generally, taxonomic composition and transcriptional profiles from 150 to 890 m for the whole community were more similar to one another relative to the euphotic zone; the only significant difference between the two sub-

euphotic zone depths was an abundance of transcripts associated with chitinase and lysosome activity (Figs. 3 and 4, Supporting Information Table S8; $p < 0.05$). These observations are in concordance with previous work at the SPOT station, where the composition of euphotic and sub-euphotic zone communities were distinct and there was a relative increase in metabolic activity at the oxycline (Hu *et al.*, 2016).

Depth-specific differences among taxonomic groups

The ordination analysis (Fig. 5) and ternary plot (Fig. 6) illustrated taxon-specific transcriptional patterns likely

indicative of differences in species composition within major taxonomic groups, physiological acclimatization among the same species, or a result of evolutionary adaptation within a lineage at each depth. In addition to verifying high reproducibility among replicates, the ordination analysis appeared to divide some groups by different nutritional strategies exhibited at different depths in the water column. That is, primarily phototrophic taxa (chlorophytes, haptophytes and diatoms) at the surface clustered away from samples collected at sub-euphotic depths, and from phagotrophic-capable taxonomic groups (Fig. 5, positions on x-axis). This discrimination was similar to that observed in culture-based transcriptome studies, where gene expression profiles of species clustered by nutritional mode (MMETSP transcriptomes; Koid *et al.*, 2014; chryso-phytes; Beisser *et al.*, 2017). The distribution of miTags with respect to depth among most taxonomic groups provided some evidence that closely related species were found throughout the water column (Supporting Information Fig. S2), as there were many miTags shared among all depths.

In contrast to well-known pathways associated with photosynthesis, there is little known about the pathways that govern phagotrophy in protists. Yet, heterotrophic protists are important consumers of microbial prey, and contribute significantly to food web function and nutrient remineralization (Strom *et al.*, 1997; Sherr and Sherr, 2002). Additionally, phagotrophy among phytoplankton (mixotrophy) is a widespread and ecologically important strategy for those species, and may increase the transport of carbon to higher trophic levels of plankton communities (Worden *et al.*, 2015; Ward and Follows, 2016).

Dinoflagellate genes made up a significant portion of the transcripts observed in our dataset (Fig. 2), and much of their expression (Fig. 6A) and class-level taxonomic composition (miTag results, Supporting Information Fig. S2A) was similar at each depth. In part this may reflect the large transcript pools or post-transcriptional gene regulation found among dinoflagellates (Hackett *et al.*, 2004; Mustafa *et al.*, 2010; Morey *et al.*, 2011). However, those pathways that were significantly differentially expressed indicated a primarily phototrophic assemblage of dinoflagellates in the photic zone (photosynthesis-related transcripts in Fig. 6A), while pathways that may be associated with heterotrophic nutrition were observed at 150 and 890 m (fatty acid breakdown associated transcripts; Fig. 6A). These results are consistent with a wide range of nutritional modes exhibited by dinoflagellates (Hackett *et al.*, 2004; Jeong *et al.*, 2010).

Interestingly, dinoflagellate assemblages at different depths appeared to employ different metabolic strategies with respect to nitrogen transport or transformation (NRT, AMT and nitrate reduction). Nitrogen-related transcripts and pathways were significantly increased at the surface

and 890 m at the SPOT station, where the availability of inorganic nitrogen was lowest and highest respectively (Fig. 6A, Supporting Information Table S4). Phototrophic dinoflagellate species at the surface were likely scavenging inorganic nitrogen (upregulation of NRT in Fig. 6A), as was observed in *Karenia brevis* under N-starved conditions (Morey *et al.*, 2011), whereas, heterotrophic dinoflagellate species found at 890 m were unlikely to be actively taking up inorganic nitrogen. We speculate that high abundances of nitrogen metabolism transcripts at 890 m reflected repartitioning of intracellular nitrogen (Dagenais-Bellefeuille and Morse, 2013) or the release of nitrogen acquired from ingested prey. This is notable because it implies flexibility in the nitrogen metabolism strategies in sub-euphotic zone dinoflagellates. However, additional work to characterize these taxa are required as dinoflagellate transcripts detected at sub-euphotic zone depths may have included species (or life stages) for which we do not have appropriate reference databases (Supporting Information Fig. S2A, 'Uncertain' and 'uncultured'), including parasitic species (e.g., syndiniales; Guillou *et al.*, 2008) or encysted life stages (McMinn and Martin, 2013).

Sources of labile organic matter that support higher protistan phagotrophy at oxyclines (and other transition zones), might partially explain high ciliate diversity and activity observed at subsurface depths in this and previous studies (Stock *et al.*, 2009; Pachiadaki *et al.*, 2016). Ciliate gene expression provided evidence of elevated activity at the oxycline and also illustrated that core ciliate metabolisms did not vary greatly with depth (clustering of red symbols in Fig. 5). Ciliate reads were highest at 150 m (Fig. 2), in agreement with previous work at the SPOT station that indicated higher relative activity at the oxycline (Hu *et al.*, 2016). Transcripts associated with fatty acid breakdown were significantly increased at 150 m relative to other depths (Fig. 6B, Supporting Information Table S9; $p < 0.05$). Class-level taxonomic distribution based on miTag results showed little variation by depth; ciliates below the euphotic zone were mainly dominated by *Choreotrichia* and *Oligohymenophora* (Supporting Information Fig. S2B), which include species previously found at oxic/anoxic boundaries and anoxic environments (Fenchel and Finlay, 1995; Edgcomb *et al.*, 2011a). Further, compared with other taxonomic groups, there was a higher number of unique ciliate miTags detected at 150 and 890 m (Supporting Information Fig. S3C); the presence of surprisingly diverse and active populations of ciliates below the euphotic zone warrants further study.

Ciliates in low oxygen environments may adopt anaerobic metabolisms using a modified mitochondrial organelle (hydrogenosome), in which pyruvate (from glycolysis) is fermented into acetyl coenzyme-A (acetyl-CoA), acetate, and hydrogen to generate energy (Müller, 1988; Fenchel and Finlay, 1995). Pyruvate:ferredoxin oxidoreductase

(por) transcripts, responsible for the initial breakdown of pyruvate into acetyl-CoA and CO₂ (Fenchel and Finlay, 1995), were higher at 150 m relative to the surface, and significantly higher at 890 m, where oxygen was lowest, compared with the surface (0.15 ml/L; Fig. 6B and Supporting Information Tables S4 and S9; $p < 0.05$). The presence of this gene indicated that some species may have adopted an anaerobic metabolism at those depths. In agreement with the frequency of microaerophilic and anaerobic ciliates in the nearby oxygen-deplete Santa Barbara Basin (Bernhard *et al.*, 2000; Beinart *et al.*, 2018) and other low-oxygen environments (Orsi *et al.*, 2012), some ciliates at 150 and 890 m may possess ectosymbiotic (Fenchel *et al.*, 1977) or endosymbiotic bacteria and archaea to support an anaerobic metabolism (Edgcomb *et al.*, 2011b; Nowack and Grossman, 2012). Specifically, methanogens or sulfate reducing or oxidizing bacteria associated with ciliates may serve to deplete the end products from the fermentation of pyruvate; however, molecular characterizations of the interactions between protistan hosts and symbionts at anoxic or low-oxygen depths in the water column remain largely uncharacterized (Stoeck *et al.*, 2014; Edgcomb, 2016).

Haptophytes at the surface had higher transcript abundances associated with nitrogen uptake, consistent with a nitrogen-limited environment (Fig. 6C, Supporting Information Tables S4 and S9; $p < 0.05$). *Prymnesium parvum*, *Isochrysis galbana* and *Emiliania huxleyi*, have been shown to significantly upregulate nitrate transporters under N-deplete conditions (Dyhrman *et al.*, 2006; Liu *et al.*, 2015a). Conversely, decreased expression of nitrogen uptake genes by *P. parvum* was observed when prey were abundant (Liu *et al.*, 2015b), and a natural population of haptophytes exhibited decreased expression of nitrogen transporters when nitrogen deficiency was alleviated by supplementation with nutrient-rich water (Alexander *et al.*, 2015b). Combined, these results are consistent with the use of NRT as a putative biomarker for N-stressed haptophyte populations, where decreased abundances of NRT may reflect an alternate N acquisition strategy among haptophytes, such as phagotrophy.

Haptophyte transcriptional patterns were similar to other phototrophic groups at the surface, but clustered with dinoflagellates in the ordination analysis for samples collected below the euphotic zone (Fig. 5), perhaps reflecting the mixotrophic capabilities of many of these species (Liu *et al.*, 2009; Cuvelier *et al.*, 2010). Haptophytes from deep samples in this study had high relative transcript abundances in pathways related to fatty acid breakdown and the glyoxylate cycle relative to the surface, similar to the response of *P. parvum* in the presence of prey (Fig. 6C; Liu *et al.*, 2015b). Taken together, these data provide evidence of the cellular breakdown of fatty acids (likely from consumed prey) to produce acetyl-CoA, which enters the

glyoxylate cycle. Transcripts related to the urea cycle also increased in abundance among haptophytes at 150 and 890 m, suggesting that the urea cycle may play an important role in how haptophytes process nitrogen below the euphotic zone, where phagotrophy is potentially occurring. While there is molecular evidence for urea metabolism in haptophytes (culture-based studies), integrating *in situ* biogeochemical measurements of urea uptake and utilization are required to fully understand intracellular regulation (or re-distribution) of different nitrogen pools among haptophytes, and their role in the urea cycle (Solomon *et al.*, 2010). Our results imply that phylogenetically related haptophytes (mainly *Prymnesiales*; Supporting Information Figs. S2C and S3D) occupy different depths in the water column. Additionally, while the data suggest a predominantly phagotrophic metabolism, sinking haptophytes may reallocate intracellular carbon compounds as a stress response to sub-euphotic zone conditions.

Diatoms and chlorophytes

High abundances of diatoms and chlorophytes in the euphotic zone (Fig. 2) and high expression of phototrophy and carbon fixation-related genes at the surface (Figs. 4 and 6D,E, Supporting Information Table S9; $p < 0.05$) was anticipated, and our study confirmed a primarily phototrophic lifestyle for these taxa in the euphotic zone. Diatoms and picoeukaryotic prasinophytes such as *Ostreococcus* are common inhabitants of coastal waters of the eastern Pacific Ocean where they contribute significantly to primary production (Countway and Caron, 2006; Worden, 2006; Aylward *et al.*, 2015). MiTag results confirmed the presence of important coastal diatoms such as *Pseudonitzschia*, and the chlorophytes *Micromonas* and *Ostreococcus* (Supporting Information Fig. S2D,E). Their presence in dark ocean environments is typically overlooked, and generally attributed to deep vertical mixing, active sinking of cells or association with sinking particles (Jochem, 1999; Agusti *et al.*, 2015). Although found at very low abundances, the presence of transcripts and unique miTags (Supporting Information Fig. S3E,F) from these species below the euphotic zone indicated that some diatoms and chlorophytes may have the capability to alter their physiology to survive at 150 and 890 m, at least for some limited time.

Phytoplankton metabolism below the euphotic zone is not well understood, but several studies have examined the ability of diatoms to survive periods of prolonged darkness (e.g., up to 10 months; Smayda and Mitchell-Innes, 1974; Peters and Thomas, 1996; Nymark *et al.*, 2013). Subsurface diatoms and chlorophytes transcribed photosynthesis-related genes, but at lower levels than at the surface (Fig. 6D, photosynthesis transcripts; Fig. 6E, photosynthesis, Calvin cycle, glycolysis-associated

transcripts); similar to *Phaeodactylum tricornutum* exposed to darkness for 48 h (Nyman *et al.*, 2013). Continued transcription of photosynthesis-related genes may be a mechanism for these phytoplankton to retain the ability to perform photosynthesis immediately when light becomes available (Peters and Thomas, 1996; Nyman *et al.*, 2013). This conjecture is supported by evidence that even though transcription of chloroplast-encoded genes may continue in phytoplankton, the transcripts may not be translated into protein without sunlight, thus maintaining responsiveness while minimizing energy expenditure (Lee and Herrin, 2002).

Diatoms and chlorophytes below the euphotic zone demonstrated a possible reliance on intracellular lipids for energy acquisition (high abundances of fatty acid, glyoxylate cycle, V-type ATPases and lysosome transcripts; Fig. 6D,E). Previous studies have reported increased lipid accumulation in phytoplankton under sub-optimal conditions, including polar darkness (McKnight *et al.*, 2000; McMinn and Martin, 2013), silicon-deficiency (Roessler, 1988) or low temperature and N-limitation (Mock and Kroon, 2002), consistent with the transcriptional patterns reported here.

The abundance of transcripts associated with chitinase production was higher below the euphotic zone in both diatoms and chlorophytes (Fig. 6D,E). Increased chitinase in diatoms has been linked to altering cellular structure in order to reduce sinking (by increasing drag), decreasing susceptibility to fungal infection or entering a resting state (Round *et al.*, 1990; Mulisch, 1993). Additionally, transcriptional signatures of diatoms and chlorophytes at 150 and 890 m may be indicative of resting cysts, as many taxa are known to enter a dormant cell cycle (cyst) in response to unfavourable environmental conditions (McMinn and Martin, 2013). However, the transcriptional signatures associated with encysted cells or resting spores among diatom and chlorophyte species in dark ocean environments are not well characterized.

Improving future metatranscriptome analyses

Community composition at each depth in this study, as determined from the metatranscriptomes, was in agreement with previous tag-sequencing surveys of protistan diversity at the SPOT station, confirming and expanding our current knowledge of *in situ* protistan ecological roles at this coastal site. However, a high percentage of transcripts and miTags still could not be annotated or sufficiently assigned taxonomy, especially for samples collected at sub-euphotic depths (Fig. 2). While there was some evidence of taxa such as excavates, rhizaria and the stramenopile group known as MARine STRamenopiles (MAST) in the results, transcripts from these taxa were not well-represented in our dataset (compare Fig. 2 and

Supporting Information Fig. S1, Supporting Information Tables S6 and S7), presumably due to the lack of reference data for those species. One explanation for this is the use of a prefilter (see section on 'Experimental procedures'), which would have excluded rhizaria larger than 80 µm; however, using the same collection procedures, an earlier study found rhizaria to be a potentially significant grazer below the euphotic zone at the SPOT station (Hu *et al.*, 2016). Further, rhizaria are thought to be an important group throughout the world's oceans (De Vargas *et al.*, 2015; Biard *et al.*, 2016; Guidi *et al.*, 2016). We therefore anticipated a significant contribution of these species to our metatranscriptome survey, but <2% of the community metatranscriptome was identified as cercozoan, radiolarian or foraminiferan (Supporting Information Table S6). These findings indicate that available rhizarian reference transcriptomes (28/844 in our custom database, Supporting Information Table S2) were phylogenetically distant from those found in the water column at the SPOT station (Del Campo *et al.*, 2014). Increasing the number of available reference transcriptomes and genomes for poorly represented groups (e.g., rhizaria or MAST) and increasing sequence read lengths should dramatically improve the amount of information that can be derived from metatranscriptomic studies of microbial eukaryotes. Efforts to enhance genetic databases for uncultured eukaryotes are well underway (Mangot *et al.*, 2017; Carradec *et al.*, 2018).

Further, we chose to keep taxonomic assignments uniform for all protistan lineages (e.g., same taxonomic level for all major groups); therefore we refrained from describing the protistan community to the class or genus level based on transcript or miTag results. Some protistan lineages are currently underrepresented in transcriptome databases (see previous section), therefore the alignment of short reads (125 bps in this study) to assembled contigs does not offer the phylogenetic resolution required to accurately assign all protistan lineages to the species level. This attests to the need for improved protistan taxonomic assignment criteria and strategies for metatranscriptome data analyses. We also acknowledge that sample handling procedures may have an impact on gene expression (Pachiadaki *et al.*, 2016). To combat these two factors, we included a high number of replicates, carried out quick and careful sampling procedures, and focused our study on abundant taxa (>2%, see section on 'Experiment procedures').

Summary

This study contributes a rich amount of gene expression data underpinning the ecological activities of marine protists throughout the water column at a coastal study site. Metatranscriptomic approaches are a way forward to uncover the metabolic activities of ecologically significant

microorganisms which we otherwise have no other way of accessing. Recent global surveys of marine diversity have revealed the presence and ecological impact of protists to be more significant than previously thought (De Vargas *et al.*, 2015), highlighting a pressing need to develop tools and techniques to investigate *in situ* trophic strategies in diverse microbial communities. Identifying the mechanistic drivers of these trophic modes will help to better parameterize food web models in the coastal zone and other planktonic ecosystems.

Experimental procedures

Sample collection and handling

Seawater was collected from the San Pedro Ocean Time-series (SPOT) station off the coast of Southern California near the surface (5 m), 150 and 890 m, in late May 2015 (Fig. 1A; Supporting Information Table S1). Briefly, seawater was pre-filtered (80 μ m) into 20 L carboys to minimize the presence of multicellular eukaryotes. Replicate samples (ranging in volume from 1.5–3.5 L, Supporting Information Table S1) from each depth were filtered onto sterile GF/F filters (nominal pore size 0.7 μ m, Whatman, International Ltd. Florham Park, NJ). While we cannot avoid some impact that sample handling (i.e., bringing samples to the surface) may have had on our results, filters were immediately placed in 1.5 mL of lysis buffer and flash frozen in liquid nitrogen in < 40 min and away from light to minimize RNA degradation (see Supporting Information Table S1). Detailed protocols for field collection and preservation of samples can be found at protocols.io (dx.doi.org/10.17504/protocols.io.hisb4ee) and in Supporting Information. Temperature, chlorophyll *a* fluorescence, oxygen and salinity at the SPOT station were obtained during each CTD cast (Sea-bird Electronics, Inc., Bellevue, WA). Samples for inorganic nutrients and cell counts were also taken (Supporting Information).

Molecular sample processing and quality control

Total RNA was extracted from each filter using a DNA/RNA AllPrep kit (Qiagen, Valencia, CA, #80204) with an in-line genomic DNA removal step (RNase-free DNase reagents, Qiagen #79254) (dx.doi.org/10.17504/protocols.io.hk3b4yn). Extracted RNA was quality checked and low biomass samples were pooled. Six replicates were processed and sequenced from the surface, while pairs of filters were pooled for either 150 or 890 m, yielding 3 and 4 replicates respectively (Supporting Information Table S1). RNA concentrations were normalized before library preparation (Supporting Information). ERCC spike-in was added before sequence library preparation with Kapa's Stranded mRNA library preparation kit using poly-A tail selection beads to select for eukaryotic mRNA (Kapa Biosystems, Inc., Wilmington, MA, #KK8420). HiSeq High Output 125 bp PE sequencing was performed at UPC Genome Core at University of Southern California, Los Angeles, CA (BioProject: PRJNA391503).

Sequence adapters, low quality (phred score < 10, from 5' and 3' ends, and within a 25 bp sliding window) or short sequences (< 50 bps), and sequences containing more than

50 consecutive As or Ts were removed using Trimmomatic v. 0.32 (Bolger *et al.*, 2014). All quality trimmed sequences were aligned to ERCC sequences using 'align_and_estimate_abundance.pl' in the Trinity v. 2.1.1 (Grabherr *et al.*, 2011) package. Reads that were aligned to ERCC sequences were removed using a custom PERL script (available: https://github.com/shu251/SPOT_metatranscriptome).

miTag analysis

The rRNA and mRNA reads were separated using SortMeRNA v. 2.0 (Kopylova *et al.*, 2012). Forward rRNA reads were further quality checked (Q > 16) in QIIME (Caporaso *et al.*, 2010) and taxonomic identities were assigned to each read using uclust (Edgar, 2010); each read had to have at least 97% similarity to the SILVA database v.128 (Quast *et al.*, 2012). Metagenomic Illumina tags, or miTags, are defined as rRNA reads that serve as an alternative to amplicon sequences for diversity analyses (Logares *et al.*, 2014).

Assembly and protein prediction

mRNA sequences were assembled using MEGAHIT v. 1.0.3 (Li *et al.*, 2015) with default parameters. Redundancy among assembled contigs was reduced using CD-HIT at 98% similarity (Li *et al.*, 2001). Estimated transcript abundances based on paired end reads were determined using Salmon v. 0.8.2 with quasi mapping strategy (kmer size = 31; Patro *et al.*, 2017). Coding sequences for each contig were predicted using GeneMarkS-T v. 5.1, which specifically predicts coding regions in eukaryotic transcripts (Tang *et al.*, 2015). GeneMarkS-T should predict one coding sequence per contig, but in rare cases (< 0.001%) more than one coding sequence was predicted. In those cases, the longest coding sequence was used. Predicted protein sequences (> 300 bps) were clustered to produce orthologous groups with 75% identity cut-off using uclust v. 1.2.22 (Edgar, 2010). Contigs without predicted coding sequences were not analyzed.

Annotation

Contigs were assigned taxonomic identities using BLASTX (e-value cutoff: 1e-5) and then MegaBLAST (e-value cutoff: 1e-5) to customized protein and cDNA reference databases, respectively, which included data from the Marine Microbial Eukaryote Transcriptome Sequencing Project and other publicly available protistan genomes and transcriptomes (Supporting Information Table S2 and Supporting Information). First, contigs were aligned to the custom protein database using a faster BLASTX search provided by DIAMOND (Buchfink *et al.*, 2015) with the sensitive alignment mode, an e-value cutoff of 1e-5, and an identity cutoff of 40%. The best hit and all hits with bit scores within 90% of the best score were considered. A taxonomic identity was assigned to a contig if all extracted hits were from the same taxonomic group (Supporting Information Table S2). If no consensus could be reached among all top hits from the BLASTX results, the contig was assigned to 'Multiple hits'. Second, to increase the number of contigs with taxonomic assignments, any contigs that fell into the 'Multiple hits' category or did not hit any

reference in the protein database ('NA from database') were searched against the custom cDNA database using MegaBLAST (Camacho *et al.*, 2009) with the same cutoffs. We elected to use nucleotide sequences for taxonomic assignments in order to evade missing taxonomy information from shorter or missing predicted protein sequences. We found that taxonomic assignment using this approach was reasonably accurate (> 90%) at the phylum or class level, but unreliable at genus or species level for every protistan lineage. While assignment to the class or genus level may be accurate among well-represented taxa in the database, we chose to assign taxonomy uniformly across all lineages, thus taxonomic assignment was restricted to phylum/class levels (see Supporting Information Table S2 for taxonomic grouping).

KEGG annotation of putative protein sequences (GeneMark S-T) was performed using GhostKOALA (Kanehisa *et al.*, 2016). Protein annotations for each contig included KEGG module information when available. Pathways and genes were chosen to represent the core biogeochemical functions relevant to known metabolisms (Supporting Information Table S3). Genes that appeared in multiple KEGG modules, and subsequent pathways, were counted multiple times after normalization steps (Supporting Information Table S3).

Data synthesis and statistical analyses

Transcript counts were normalized across libraries (replicates) using the trimmed mean normalization method in 'edgeR' (Lund *et al.*, 2012) to generate transcript counts per million (CPM) for all downstream analyses. To evaluate differences at the pathway level, analysis of variance statistical tests (ANOVA) were performed on normalized data using replicates from each depth. ANOVAs were followed by Tukey's Honestly Significantly Different test to obtain *p*-values for pairwise comparisons among depths. Library normalization with 'edgeR' and ANOVA statistical tests were performed twice, once with the whole community (all transcripts) and a second time for each individual taxonomic group (i.e., dinoflagellates, ciliates, haptophytes, diatoms and chlorophytes). The mean CPM of the replicates for a given depth was used to visualize depth-specific differences in gene expression, unless specified in the figure legend. Transcripts with KEGG identities in all five taxonomic groups were used for ordination analysis. All figures were generated in R using the vegan, UpSetR, ggplot and ggtern packages (Wickham, 2009; R Core Team, 2014; Hamilton, 2016; Conway *et al.*, 2017).

Access to data

All sequences are publicly available with accession numbers SAMN07269826-SAMN07269838, in the Short Read Archive (Supporting Information Table S1). Processed data is available at zenodo (DOI: 10.5281/zenodo.1202041) and all source code is available https://github.com/shu251/SPOT_metatranscriptome.

Acknowledgements

Authors would like to thank the Captain and Crew of the R/V Yellow Fin, Troy Gunderson, members of Doug Capone's lab,

Erin Fichot, Diane Kim and the Wrigley Institute for Environmental Studies for fieldwork support. Support was in part provided by the World Surf League PURE in partnership with Columbia University's Centre for Climate and Life to SD. Funding for this project was provided by the Gordon and Betty Moore Foundation #3299 (DAC and KBH) and the National Science Foundation #1136818 (DAC).

Conflict of Interest

The submitted manuscript is original research and has not been previously published. There are neither conflicts with editorial policies nor competing interests with the authors.

References

- Agusti, S., González-Gordillo, J.I., Vaqué, D., Estrada, M., Cerezo, M.I., and Salazar, G. (2015) Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nat Commun* **6**: 7608.
- Alexander, H., Jenkins, B.D., Rynearson, T.A., and Dyhrman, S.T. (2015a) Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc Natl Acad Sci USA* **112**: E2182–E2190.
- Alexander, H., Rouco, M., Haley, S.T., Wilson, S.T., Karl, D.M., and Dyhrman, S.T. (2015b) Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc Natl Acad Sci USA* **112**: E5972–E5979.
- Aylward, F.O., Eppley, J.M., Smith, J.M., Chavez, F.P., Scholin, C.A., and DeLong, E.F. (2015) Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc Natl Acad Sci USA* **112**: 5443–5448.
- Balzano, S., Corre, E., Decelle, J., Sierra, R., Wincker, P., Da Silva, C., *et al.* (2015) Transcriptome analyses to investigate symbiotic relationships between marine protists. *Front Microbiol* **6**: 1–14.
- Beinart, R.A., Beaudoin, D.J., Bernhard, J.M., and Edgcomb, V.P. (2018) Insights into the metabolic functioning of a multi-partner ciliate symbiosis from oxygen-depleted sediments. *Mol Ecol* **27**: 1794–1807.
- Beisser, D., Graupner, N., Bock, C., Wodniok, S., Grossmann, L., Vos, M., *et al.* (2017) Comprehensive transcriptome analysis provides new insights into nutritional strategies and phylogenetic relationships of chrysophytes. *PeerJ* **5**: e2832.
- Bernhard, J.M., Buck, K.R., Farmer, M.A., and Bowser, S.S. (2000) The Santa Barbara Basin is a symbiosis oasis. *Nature* **403**: 77–80.
- Beszteri, S., Yang, I., Jaeckisch, N., Tillmann, U., Frickenhaus, S., Glöckner, G., *et al.* (2012) Transcriptomic response of the toxic prymnesiophyte *Prymnesium parvum* (N. Carter) to phosphorus and nitrogen starvation. *Harmful Algae* **18**: 1–15.
- Biard, T., Stemann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., *et al.* (2016) *In situ* imaging reveals the biomass of giant protists in the global ocean. *Nature* **532**: 504–516.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

- Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinform* **10**: 421–429.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Caron, D.A., Connell, P.E., Schaffner, R.A., Schnetzer, A., Fuhrman, J.A., Countway, P.D., and Kim, D.Y. (2017) Planktonic food web structure at a coastal time-series site: I. Partitioning of microbial abundances and carbon biomass. *Deep-Sea Res Part I* **121**: 14–29.
- Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018) A global ocean atlas of eukaryotic genes. *Nat Commun* **9**: 373.
- Conway, J.R., Lex, A., and Gehlenborg, N. (2017) UpSetR: an R Package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**: 2938–2940.
- Countway, P.D., and Caron, D.A. (2006) Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Appl Environ Microb* **72**: 2496–2506.
- Cuvelier, M., Allen, A., Monier, A., McCrow, J., Messié, M., Tringe, S., et al. (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci USA* **107**: 14679–14684.
- Dagenais-Bellefeuille, S., and Morse, D. (2013) Putting the N in dinoflagellates. *Front Microbiol* **4**: 1–14.
- De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605–1261612.
- Del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. (2014) The others: our biased perspective of eukaryotic genomes. *Trends Ecol Evol* **29**: 252–259.
- Dupont, C.L., McCrow, J.P., Valas, R., Moustafa, A., Walworth, N., Goodenough, U., et al. (2015) Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *Isme J* **9**: 1076–1092.
- Dyhrman, S.T., Haley, S.T., Birkeland, S.R., Wurch, L.L., Cipriano, M.J., and McArthur, A.G. (2006) Long serial analysis of gene expression for gene discovery and transcriptome profiling in the widespread marine coccolithophore *Emiliana huxleyi*. *Appl Environ Microb* **72**: 252–260.
- Dyhrman, S.T., Jenkins, B.D., Rynearson, T.A., Saito, M.A., Mercier, M.L., Alexander, H., et al. (2012) The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS One* **7**: e33768.
- Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Edgcomb, V.P. (2016) Marine protist associations and environmental impacts across trophic levels in the twilight zone and below. *Curr Opin Microbiol* **31**: 169–175.
- Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., et al. (2011a) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *Isme J* **5**: 1344–1356.
- Edgcomb, V.P., Leadbetter, E.R., Bourland, W., Beaudoin, D., and Bernhard, J.M. (2011b) Structured multiple endosymbiosis of bacteria and archaea in a ciliate from marine sulfidic sediments: a survival mechanism in low oxygen, sulfidic sediments?. *Front Microbiol* **2**: 1–16.
- Edgcomb, V.P., Pachiadaki, M.G., Mara, P., Kormas, K.A., Leadbetter, E.R., and Bernhard, J.M. (2016) Gene expression profiling of microbial activities and interactions in sediments under haloclines of E. Mediterranean deep hypersaline anoxic basins. *Isme J* **10**: 2643–2657.
- Eskelinen, E.-L., and Saftig, P. (2009) Autophagy: a lysosomal degradation pathway with a central role in health and disease. *BBA-Mol Cell Res* **1793**: 664–673.
- Fenchel, T., and Finlay, B.J. (1995) *Ecology and Evolution in Anoxic Worlds*. New York, NY: Oxford University Press Inc.
- Fenchel, T., Perry, T., and Thane, A. (1977) Anaerobiosis and symbiosis with bacteria in free-living ciliates. *J Protozool* **24**: 154–163.
- Finbow, M.E., and Harrison, M.A. (1997) The vacuolar H⁺-ATPase: a universal proton pump of eukaryotes. *Biochem J* **324**: 697–712.
- Frischkorn, K.R., Harke, M.J., Gobler, C.J., and Dyhrman, S.T. (2014) *De novo* assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. *Front Microbiol* **5**: 1–16.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644–652.
- Grossmann, L., Beisser, D., Bock, C., Chatzinotas, A., Jensen, M., Preisfeld, A., et al. (2016) Trade-off between taxon diversity and functional diversity in European lake ecosystems. *Mol Ecol* **25**: 5876–5888.
- Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**: 465–470.
- Guillou, L., Viprey, M., Chambouvet, A., Welsh, R.M., Kirkham, A.R., Massana, R., et al. (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (*Alveolata*). *Environ Microbiol* **10**: 3349–3365.
- Hackett, J.D., Anderson, D.M., Erdner, D.L., and Bhattacharya, D. (2004) Dinoflagellates: a remarkable evolutionary experiment. *Am J Bot* **91**: 1523–1534.
- Hamilton, N. (2016) ggtern: an Extension to 'ggplot2', for the Creation of Ternary Diagrams. R package version 2.1.5. <https://CRAN.R-project.org/package=ggtern>.
- Harke, M.J., Juhl, A.R., Haley, S.T., Alexander, H., and Dyhrman, S.T. (2017) Conserved transcriptional responses to nutrient stress in bloom-forming algae. *Front Microbiol* **8**: 1279.
- Hu, S.K., Liu, Z., Lie, A.A.Y., Countway, P.D., Kim, D.Y., Jones, A.C., et al. (2015) Estimating protistan diversity using high-throughput sequencing. *J Eukaryot Microbiol* **62**: 688–693.
- Hu, S.K., Campbell, V., Connell, P., Gellene, A.G., Liu, Z., Terrado, R., and Caron, D.A. (2016) Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. *FEMS Microbiol Ecol* **92**: 1–13.

- Jeong, H.-J., Yoo, Y.-D., Kim, J.S., Seong, K.A., Kang, N.S., and Kim, T.H. (2010) Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Sci J* **45**: 65–91.
- Jochem, F.J. (1999) Dark survival strategies in marine phytoplankton assessed by cytometric measurement of metabolic activity with fluorescein diacetate. *Mar Biol* **135**: 721–728.
- Johnson, M.D. (2015) Inducible Mixotrophy in the Dinoflagellate *Prorocentrum minimum*. *J Eukaryot Microbiol* **62**: 431–443.
- Johnson, M.D., Oldach, D., Delwiche, C.F., and Stoecker, D.K. (2007) Retention of transcriptionally active cryptophyte nuclei by the ciliate *Myrionecta rubra*. *Nature* **445**: 426–428.
- Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726–731.
- Koid, A.E., Liu, Z., Terrado, R., Jones, A.C., Caron, D.A., and Heidelberg, K.B. (2014) Comparative transcriptome analysis of four prymnesiophyte algae. *PLoS One* **9**: e97801.
- Kopylova, E., Noe, L., and Touzet, H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.
- Lasek-Nesselquist, E., Wisecaver, J.H., Hackett, J.D., and Johnson, M.D. (2015) Insights into transcriptional changes that accompany organelle sequestration from the stolen nucleus of *Mesodinium rubrum*. *BMC Genom* **16**: 1–14.
- Lee, J., and Herrin, D.L. (2002) Assessing the relative importance of light and the circadian clock in controlling chloroplast translation in *Chlamydomonas reinhardtii*. *Photosynth Res* **72**: 295–306.
- Li, W., Jaroszewski, L., and Godzik, A. (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**: 282–283.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676.
- Lie, A.A.Y., Liu, Z., Terrado, R., Tatters, A.O., Heidelberg, K.B., and Caron, D.A. (2017) Effect of light and prey availability on gene expression of the mixotrophic chrysophyte, *Ochromonas* sp. *BMC Genom* **18**: 1–16.
- Lie, A.A.Y., Liu, Z., Terrado, R., Tatters, A.O., Heidelberg, K.B., and Caron, D.A. (2018) A tale of two mixotrophic chrysophytes: insights into the metabolisms of two *Ochromonas* species (Chrysophyceae) through a comparison of gene expression. *PLoS One* **13**: e0192439.
- Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., *et al.* (2009) Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci USA* **106**: 12803–12808.
- Liu, Z., Koid, A.E., Terrado, R., Campbell, V., Caron, D.A., and Heidelberg, K.B. (2015a) Changes in gene expression of *Prymnesium parvum* induced by nitrogen and phosphorus limitation. *Front Microbiol* **6**: 1–13.
- Liu, Z., Jones, A.C., Campbell, V., Hambricht, K.D., Heidelberg, K.B., and Caron, D.A. (2015b) Gene expression in the mixotrophic prymnesiophyte, *Prymnesium parvum*, responds to prey availability. *Front Microbiol* **6**: 1–12.
- Liu, Z., Campbell, V., Heidelberg, K.B., and Caron, D.A. (2016) Gene expression characterizes different nutritional strategies among three mixotrophic protists. *FEMS Microbiol Ecol* **92**: fiw106–fiw111.
- Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmiento, H., *et al.* (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* **16**: 2659–2671.
- Lund, S.P., Nettleton, D., McCarthy, D.J., and Smyth, G.K. (2012) Detecting differential expression in rna-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol* **11**: 1–44.
- Mangot, J.-F., Logares, R., Sánchez, P., Latorre, F., Seeleuthner, Y., Mondy, S., *et al.* (2017) Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. *Sci Rep* **7**: 41498–41412.
- Marchetti, A., Schrueth, D.M., Durkin, C.A., Parker, M.S., Kodner, R.B., Berthiaume, C.T., *et al.* (2012) Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc Natl Acad Sci USA* **109**: E317–E325.
- McKnight, D.M., Howes, B.L., Taylor, C.D., and Goehring, D.D. (2000) Phytoplankton dynamics in a stably stratified Antarctic lake during winter darkness. *J Phycol* **36**: 852–861.
- McMinn, A., and Martin, A. (2013) Dark survival in a warming world. *Proc R Soc B-Biol Sci* **280**: 20122909.
- Mock, T., and Kroon, B. (2002) Photosynthetic energy conversion under extreme conditions - I: important role of lipids as structural modulators and energy sink under N-limited growth in Antarctic sea ice diatoms. *Phytochemistry* **61**: 41–51.
- Morey, J.S., Monroe, E.A., Kinney, A.L., Beal, M., Johnson, J.G., Hitchcock, G.L., and Dolah, F.M.V. (2011) Transcriptomic response of the red tide dinoflagellate, *Karenia brevis*, to nitrogen and phosphorus depletion and addition. *BMC Genom* **12**: 1–18.
- Moustafa, A., Evans, A.N., Kulis, D.M., Hackett, J.D., Erdner, D.L., Anderson, D.M., and Bhattacharya, D. (2010) Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. *PLoS One* **5**: e9688.
- Mulisch, M. (1993) Chitin in protistan organisms: distribution, synthesis and deposition. *Eur J Protistol* **29**: 1–18.
- Müller, M. (1988) Energy metabolism of protozoa without mitochondria. *Annu Rev Microbiol* **42**: 465–488.
- Nowack, E., and Grossman, A.R. (2012) Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc Natl Acad Sci USA* **109**: 5340–5345.
- Nymark, M., Valle, K.C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., *et al.* (2013) Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricoratum*. *PLoS One* **8**: e58722.
- Orsi, W., Charvet, S., Vďačný, P., Bernhard, J.M., and Edgcomb, V.P. (2012) Prevalence of partnerships between bacteria and ciliates in oxygen-depleted marine water columns. *Front Microbiol* **3**: 1–8.

- Orsi, W.D., Edgcomb, V.P., Christman, G.D., and Biddle, J.F. (2013) Gene expression in the deep biosphere. *Nature* **499**: 205–208.
- Pachiadaki, M.G., Taylor, C., Oikonomou, A., Yakimov, M.M., Stoeck, T., and Edgcomb, V. (2016) *In situ* grazing experiments apply new technology to gain insights into deep-sea microbial food webs. *Deep-Sea Res Part II* **129**: 223–231.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**: 417–410.
- Pearson, G.A., Lago-Leston, A., Cánovas, F., Cox, C.J., Verret, F., Lasternas, S., et al. (2015) Metatranscriptomes reveal functional variation in diatom communities from the Antarctic Peninsula. *Isme J* **9**: 2275–2289.
- Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J.L., Acinas, S.G., Duarte, C.M., et al. (2016) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *Isme J* **10**: 945–958.
- Peters, E., and Thomas, D.N. (1996) Prolonged darkness and diatom mortality I: marine Antarctic species. *J Exp Mar Biol Ecol* **207**: 25–41.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.
- R Core Team (2014) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Roessler, P.G. (1988) Changes in the activities of various lipid and carbohydrate biosynthetic enzymes in the diatom *Cyclotella cryptica* in response to silicon deficiency. *Arch Biochem Biophys* **267**: 521–528.
- Round, F.E., Crawford, R.M., and Mann, D.G. (1990) *The Diatoms: Biology and Morphology of the Genera*. Cambridge: Cambridge University Press.
- Settembre, C., Fraldi, A., Medina, D.L., and Ballabio, A. (2013) Signals from the lysosome: a control centre for cellular clearance and energy metabolism. *Nat Rev Mol Cell Biol* **14**: 283–296.
- Sherr, E.B., and Sherr, B.F. (2002) Significance of predation by protists in aquatic microbial food webs. *Antonie Van Leeuwenhoek* **81**: 293–308.
- Smayda, T.J., and Mitchell-Innes, B. (1974) Dark survival of autotrophic, planktonic marine diatoms. *Mar Biol* **25**: 195–202.
- Smith, S.R., Abbriano, R.M., and Hildebrand, M. (2012) Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Res* **1**: 2–16.
- Solomon, C.M., Collier, J.L., Berg, G.M., and Glibert, P.M. (2010) Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. *Aquat Microb Ecol* **59**: 67–88.
- Stock, A., Jürgens, K., Bunge, J., and Stoeck, T. (2009) Protistan diversity in suboxic and anoxic waters of the Gotland Deep (Baltic Sea) as revealed by 18S rRNA clone libraries. *Aquat Microb Ecol* **55**: 267–284.
- Stoeck, T., Filker, S., Edgcomb, V.P., Orsi, W., Yakimov, M.M., Pachiadaki, M., et al. (2014) Living at the limits: evidence for microbial eukaryotes thriving under pressure in deep anoxic, hypersaline habitats. *Adv Ecol Res* **2014**: 1.
- Strom, S.L., Benner, R., Ziegler, S., and Dagg, M.J. (1997) Planktonic grazers are a potentially important source of marine dissolved organic carbon. *Limnol Oceanogr* **42**: 1364–1374.
- Tang, S., Lomsadze, A., and Borodovsky, M. (2015) Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res* **43**: e78.
- Ward, B.A., and Follows, M.J. (2016) Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proc Natl Acad Sci USA* **113**: 2958–2963.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer Publishing Company, Incorporated.
- Worden, A.Z. (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43**: 165–175.
- Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.
- Zielinski, B.L., Allen, A.E., Carpenter, E.J., Coles, V.J., Crump, B.C., Doherty, M., et al. (2016) Patterns of transcript abundance of eukaryotic biogeochemically-relevant genes in the amazon river plume. *PLoS One* **11**: e0160929.

Supporting information

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

Fig. S1. Bar plots representing metagenomic Illumina tags (miTags) relative abundance at each depth at the (A) domain level and (B) major protistan taxonomic group. Following taxonomy assignment (uclust at 97% identity against the PR2 database, see section on 'Experimental procedures' for more details), miTags were subsampled to select the five main taxonomic groups that were the focus of this study. 'Other eukaryotes' are comprised of taxonomic groups that made up less than 10% of the total composition of miTags. Relative miTag abundances for dinoflagellates, ciliates, haptophytes, diatoms and chlorophytes were generally similar to mRNA-derived diversity (Fig. 2, Supporting Information Tables S6 and S7), and to that previously observed at the SPOT station (Hu et al., 2016). Marine Stramenopiles (MAST), syndiniales, excavates and rhizaria were represented in the miTag results (> 10% miTags, Supporting Information Table S7), but these same groups made up fewer than 2% of the mRNA-derived results (compare to Fig. 2, Supporting Information Tables S6 and S7).

Fig. S2. Bar plots representing metagenomic Illumina tags (miTags) relative abundance at the class level for (A) dinoflagellates, (B) ciliates, (C) haptophytes, (D) diatoms and (E) chlorophytes. Following taxonomy assignment (uclust at 97% identity against the PR2 database, see section on 'Experimental procedures' for more details), miTags were subsampled to select the five main taxonomic groups that were the focus of this study. miTags with more than 500 hits were selected and summed based on the approximately class level. Class level designations labelled as 'Uncertain' represent miTags assigned to only the phylum level (e.g., Haptophyte, diatom, etc.), meaning no additional taxonomic information was available. 'Uncultured eukaryote'

and 'Ambiguous_taxa' designations also denote assignments where no further details were available from the PR2 database. In this study, miTag results are mainly represented at the approximately phylum level, as short read alignments (~ 125 bps) for taxonomy assignment are considered less accurate than, for instance, longer hypervariable regions of the 18S rRNA gene (e.g., V4, ~ 400 bps). Full list of miTag results are reported in Supporting Information Table S7.

Fig. S3. Distribution of miTags among (A) all hits, (B) dinoflagellates, (C) ciliates, (D) haptophytes, (E) diatoms and (F) chlorophytes. Bar plots show the total number of shared or unique miTags among depths, depths included are represented by filled dots in the matrix below each barplot. Numbers represent the total number of miTags in each category. Most miTags were shared among all depths (left-most bars), except for chlorophytes where the highest abundance of miTags was unique to the surface. Plot generated using 'UpSetR' R package Conway *et al.* (2017).

Fig. S4. KEGG module assignments for ortholog groups found to be present at all depths (surface, 150 and 890 m, $n = 36\ 656$). See main text Fig. 3.

Fig. S5. Distribution of ortholog groups among the five targeted taxonomic groups. Bar plots show the total number of shared or unique ortholog groups among depths, depths included are represented by filled dots in the matrix below each barplot. Numbers represent the total number of ortholog clusters in each category. Most ortholog clusters were unique to a single taxonomic group (five left-most barplots). Ortholog groups associated with dinoflagellates overwhelmed sampling efforts. Very few ortholog groups were shared among all taxonomic group ($n = 196$, right-most barplot). Plot generated using 'UpSetR' R package Conway *et al.* (2017).

Table S1. Collection information for all samples, including: time of collection, location, depth, number of replicates, volume filtered, sampling time and sample pooling information for library preparation.

Table S2. List of contributing transcriptomes and genomes for custom database. Level 1, 2 and 3 correspond to manual taxonomic designation used in this analysis. Level 2

used in main analysis. See Supporting Information for more information on how the custom database was generated.

Table S3. List of gene names and KEGG IDs (K0_number) for labelled targeted metabolic pathways and genes in main text. Associated labels are based on KEGG modules.

Table S4. Summary of environmental variables and results from microscopy counts for each depth at the SPOT station. Detailed explanation of cell types for microplankton counted can be found in Supporting Information and methods.

Table S5. Summary of reads throughout sample processing, including total number of reads post trimming and quality checking, assembled contigs, predicted protein sequences and clustered ortholog groups.

Table S6. Transcript counts per million (CPM, based on the mean across replicates, n) and relative abundance (%) of taxonomic groups found at each depth. Percentage of the total is in the right-most columns. The first column details the taxonomic grouping referred to in the main text. Taxonomic groups that made up more than 2% of the community are highlighted in yellow and discussed further in main text.

Table S7. Total counts and percentages of miTag results at each depth (see section on 'Experimental procedures'). Top table summarizes the top taxonomic groups also represented in Supporting Information Fig. S1. The bottom table lists the taxonomic groups from Levels 1 to 3, which originate from the SILVA database. MiTag results needed to make up a large percentage of the population ($> 10\%$) to be represented in Supporting Information Fig. S1.

Table S8. Results from Tukey HSD post hoc test (following ANOVA) to look for significant differences among whole pathways based on transcript abundances (CPM) among replicates (Fig. 4 in main text). Significant values are bolded, $p < 0.05$.

Table S9. Results from Tukey HSD post hoc test (following ANOVA) to look for significant differences among whole pathways for individual taxonomic groups based on transcript abundances (CPM) among replicates (Fig. 6 in main text). Raw transcript counts were subset for each taxonomic group and normalized before this analysis. Significant values are bolded, $p < 0.05$.