nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-020-00814-7

Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean

Natalie R. Cohen[1]✉, Matthew R. McIlvin[1], Dawn M. Moran[1], Noelle A. Held[1], Jaclyn K. Saunders[1], Nicholas J. Hawco[2], Michael Brosnahan[3], Giacomo R. DiTullio[4], Carl Lamborg[5], John P. McCrow[6], Chris L. Dupont[6], Andrew E. Allen[6,7]✉ and Mak A. Saito[1]✉

Marine microeukaryotes play a fundamental role in biogeochemical cycling through the transfer of energy to higher trophic levels and vertical carbon transport. Despite their global importance, microeukaryote physiology, nutrient metabolism and contributions to carbon cycling across offshore ecosystems are poorly characterized. Here, we observed the prevalence of dinoflagellates along a 4,600-km meridional transect extending across the central Pacific Ocean, where oligotrophic gyres meet equatorial upwelling waters rich in macronutrients yet low in dissolved iron. A combined multi-omics and geochemical analysis provided a window into dinoflagellate metabolism across the transect, indicating a continuous taxonomic dinoflagellate community that shifted its functional transcriptome and proteome as it extended from the euphotic to the mesopelagic zone. In euphotic waters, multi-omics data suggested that a combination of trophic modes were utilized, while mesopelagic metabolism was marked by cytoskeletal investments and nutrient recycling. Rearrangement in nutrient metabolism was evident in response to variable nitrogen and iron regimes across the gradient, with no associated change in community assemblage. Total dinoflagellate proteins scaled with particulate carbon export, with both elevated in equatorial waters, suggesting a link between dinoflagellate abundance and total carbon flux. Dinoflagellates employ numerous metabolic strategies that enable broad occupation of central Pacific ecosystems and play a dual role in carbon transformation through both photosynthetic fixation in the euphotic zone and remineralization in the mesopelagic zone.

A natural biogeochemical gradient extends across the central Pacific Ocean where equatorial upwelling of macronutrient-rich seawater fuels phytoplankton growth and creates a productive ecosystem compared to adjacent oligotrophic gyres. Low iron supplies limit phytoplankton growth at the equator[1–3], whereas nitrate depletion occurs in nearby tropical gyre waters[4]. The specific molecular mechanisms underlying the physiological response of phytoplankton across this nutrient resource gradient and their impact on carbon and nutrient cycling is not well understood, particularly for eukaryotic phytoplankton.

Eukaryotic phytoplankton are primarily responsible for the transfer of organic carbon to higher trophic levels in marine systems[5] with many capable of mixotrophy[6]. Mixotrophic species fix carbon via photosynthesis and consume organic carbon via heterotrophy, with one common form being endocytosis-mediated phagotrophy[7–9]. This trophic strategy can shape the biogeochemical landscape of an ecosystem by transferring carbon more efficiently up the food chain and vertically to the deep ocean[10]. Mixotrophy may be common in the central Pacific Ocean since recent modelling-based studies suggest it is an advantageous nutritional strategy relative to autotrophy in low-nutrient oligotrophic environments[11], especially in low latitudes experiencing simultaneous carbon and nutrient limitation[12,13].

Despite recent progress in mapping microeukaryotic (protistan) distributions and their biogeochemical importance across the ocean[14–20], the factors driving mixotrophy are unclear[21] and protistan ecological roles are not well understood in deep waters[22]. The mesopelagic zone (approximately 200–1,000 m), residing just below the euphotic layer, is characterized by minimal light penetration, cooler temperatures and high concentrations of inorganic nutrients due to active microbial degradation, recycling and repackaging of organic material[23]. Field surveys have begun to characterize the abundance, diversity and activity of mesopelagic protists using meta-omics analyses[22,24–26] to understand their functional roles and contributions to biogeochemistry.

In this study, we used a combined multi-omics and geochemical approach to explore how gradients in the chemical environment influence protistan distributions, trophic strategies, nutrient physiology and carbon cycling dynamics in the euphotic and mesopelagic ocean. Our findings illustrate that dinoflagellates are relatively abundant throughout the central Pacific and utilize numerous growth strategies to survive in diverse environments, including phototrophy and phagotrophy in euphotic waters, phagotrophy and intracellular recycling in the mesopelagic zone and by restructuring iron and nitrogen metabolism depending on environmental availability.

Results

Multi-omics and pigment community composition reveal prevalence of dinoflagellates. We conducted a combined multi-omics

[1]Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [2]Department of Earth Science, University of Southern California, Los Angeles, CA, USA. [3]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [4]Department of Biology, College of Charleston, Charleston, SC, USA. [5]Ocean Sciences Department, University of California, Santa Cruz, Santa Cruz, CA, USA. [6]Microbial & Environmental Genomics, J. Craig Venter Institute, La Jolla, CA, USA. [7]Scripps Institution of Oceanography, Integrative Oceanography Division, University of California, San Diego, La Jolla, CA, USA. ✉e-mail: ncohen@whoi.edu; aallen@jcvi.org; msaito@whoi.edu

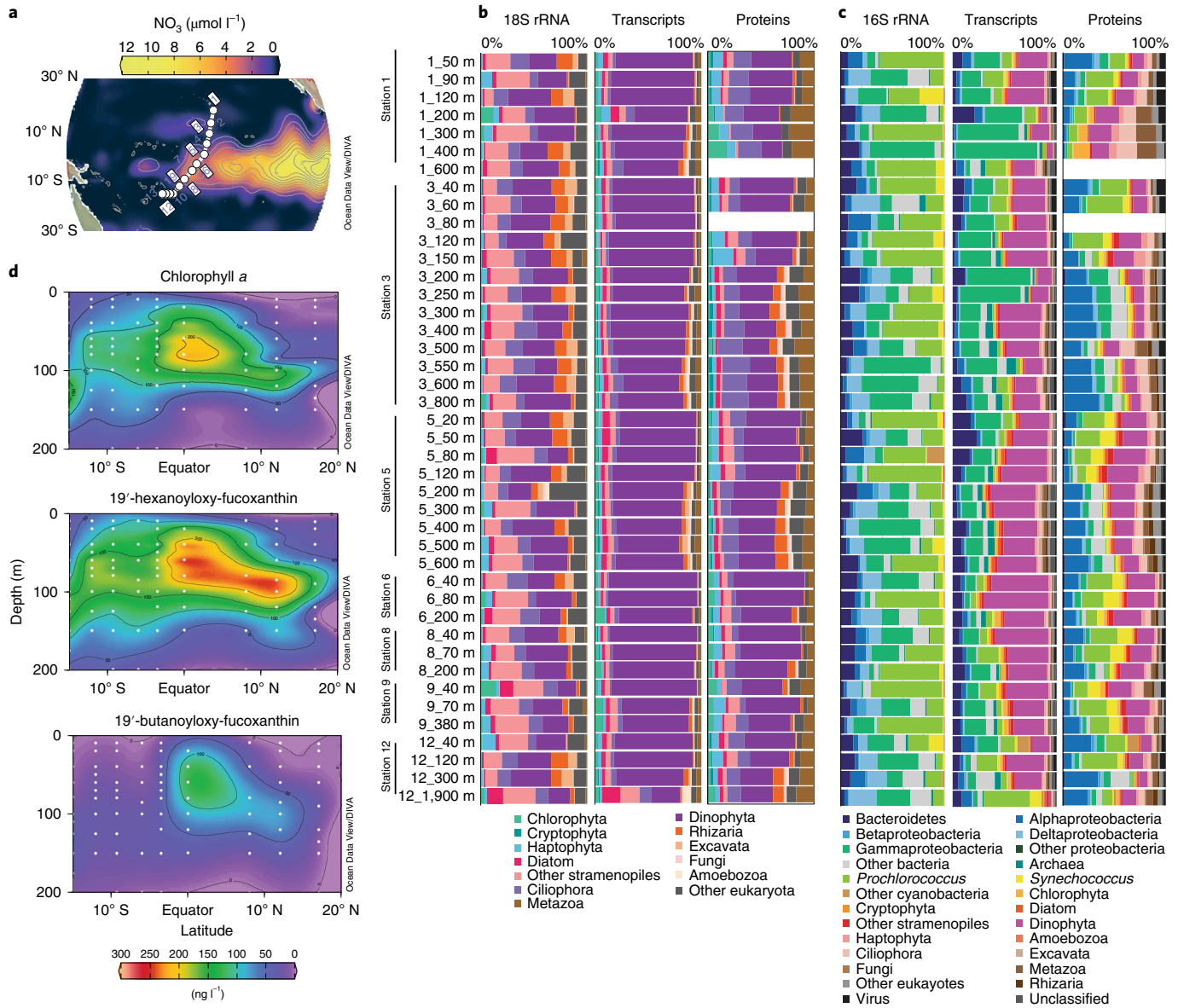


Fig. 1 | Protistan community composition across the METZYME transect. a, The meridional transect is shown with World Ocean Atlas-interpolated mean October surface NO_3 concentrations (μmol l^{-1})[147] plotted using Ocean Data View (odv.awi.de). Stations 1–12 are labelled with white squares denoting locations where combined multi-omics analyses were performed. **b**, Eukaryotic phylum-level relative community abundance determined through 18S rRNA sequencing, metatranscriptomics and metaproteomics. **c**, Whole-community phyla and supergroup-level relative community abundance determined through 16S rRNA sequencing, metatranscriptomics and metaproteomics on the 3–51-μm size fraction. For the metatranscriptome and metaproteome, only reads or spectral counts mapping to ORFs with a taxonomic annotation are shown. **d**, Vertical sections of chlorophyll *a*, 19′-hexanoyloxy-fucoxanthin and 19′-butanoyloxy-fucoxanthin pigments across the meridional transect, plotted using DIVA interpolation in Ocean Data View (Extended Data Fig. 3 for full pigment profiles).

and geochemical analysis of protistan community dynamics and functionality across an approximately 4,600-km section (Fig. 1). Forty-two biomass samples were collected primarily between 40 m and 600 m depth in the 3–51-μm size range by filter fractionation from 7 sites (Extended Data Fig. 1), providing even coverage across the transect. Of the microeukaryotes, dinoflagellates were in high relative abundance in transcript, protein and 18S ribosomal RNA pools throughout both the euphotic and mesopelagic zones (Fig. 1b). We also detected a considerable number of transcripts and proteins from prokaryotic lineages (Fig. 1c). These organisms are much smaller than 3 μm; hence, their presence suggests particle attachment, aggregation onto filters, consumption by heterotrophic organisms and/or endosymbiotic relationships with eukaryotic hosts. The dinoflagellates comprised an average 31 ± 13% of transcript read counts and 18 ± 6% of protein spectral counts, respectively, relative to the entire microbial community. In the eukaryotic subset of the community, dinoflagellates comprised 69 ± 11% of transcriptomic read counts and 41 ± 11% of protein spectral counts with the majority of these proteins associated with the Kareniaceae family and specifically, *Karlodinium* and *Karenia*-like genera (Fig. 1b and Extended Data Fig. 2). This family contains gymnodinoid dinoflagellates known to contain mixotrophic members[27–30]. While there has been concern that sequencing-based approaches may overestimate dinoflagellate populations due to their large genome size and tendency to post-transcriptionally regulate gene expression (Supplementary Information), the metaproteomic dataset provides

an independent assessment of composition based on biomass[31] and is consistent with high dinoflagellate relative abundance. Finally, although prymnesiophytes (haptophytes) have been observed to be abundant in the equatorial Pacific[32], they comprised only 2.7 ± 1% of transcripts and 3 ± 2% of protein spectral counts. Their relatively low transcripts and proteins in this study may be due to low cell abundance during this expedition, small cells <3 µm passing through filters, lysing during in situ filtration or low transcript/protein production relative to biomass.

We used a V9 18S ribosomal RNA sequencing approach to assign finer scale classification to the protistan community than was possible using metatranscriptomics. Approximately 31 ± 7% of the 18S rRNA reads were assigned to dinoflagellates (Fig. 1b). Of the 3,096 operational taxonomic units (OTUs) generated, a single OTU belonging to the dinoflagellate class Dinophyceae recruited the highest number of reads across all locations and depths (13 ± 3%) (Supplementary Fig. 2); however, subclass-level annotation was not possible using this approach (Supplementary Information). These observations are consistent with other 18S rRNA-based assessments of protistan composition in the South Pacific Ocean[17,18].

Pigments collected from the euphotic zone provided additional information about phytoplankton composition (Fig. 1d and Extended Data Fig. 3, using 0.7-µm GF/F filters). Divinyl chlorophyll *a*, zeaxanthin, 19′-hexanoyloxy-fucoxanthin and 19′-butanoyloxy-fucoxanthin were major pigments. The first is unique to *Prochlorococcus*[33], zeaxanthin is used by cyanobacteria and chlorophytes and the last two are indicative of dinoflagellates, pelagophytes and haptophytes[34]. Divinyl chlorophyll *a* reached similar concentrations as chlorophyll *a* and supports the well-characterized abundance of *Prochloroccocus* in the tropical Pacific[4,35]. Notably, pigments representative of certain dinoflagellates (peridinin) and diatoms (fucoxanthin) were an order of magnitude lower in concentration. Diatoms rapidly bloom after iron addition to equatorial upwelling waters[3,36], yet they comprised only a small fraction of the natural assemblage based on pigments and multi-omics along this transect (Fig. 1b). The high concentrations of 19′-hexanoyloxy-fucoxanthin and 19′-butanoyloxy-fucoxanthin, yet low levels of peridinin in this study support metatranscriptomic assignment of Kareniaceae dinoflagellates (Fig. 1b and Extended Data Fig. 3). Genera within this family have undergone tertiary endosymbiosis where their peridinin-containing plastid was replaced with that of a haptophyte resulting in possession of 19′-hexanoyloxy-fucoxanthin and 19′-butanoyloxy-fucoxanthin[37–40]. These interpretations are consistent with previous epifluorescence microscopy observations in the equatorial Pacific where small dinoflagellates accounted for up to 30% of the community biomass and reached 50,000–400,000 cells l^{-1} in surface waters[32,41], and gymnodinoid dinoflagellates 4 × 7 µm in size contributed to approximately 50% of the autotrophic dinoflagellate community[32]. Thus, dinoflagellates were the eukaryotic group in highest relative abundance across the transect, extending from the euphotic zone into the mesopelagic zone, based on four independent measurements and consistent with previous microscopic identifications[32,41].

Dinoflagellate growth strategies in the euphotic zone. Metabolic profiles suggested that the euphotic dinoflagellate community engaged in both phototrophy and phagotrophy, supporting the notion that both trophic strategies commonly co-occur[26,42]. Metaproteomes were compared using hierarchical clustering with Euclidean distance and resulted in separation primarily based on depth rather than latitude (Extended Data Fig. 4 and Supplementary Fig. 3). The individual proteins driving depth patterns were identified using a general independence permutation test followed by multiple test correction ($P < 0.1$). Using this conservative approach, six Pfam-annotated proteins were differentially abundant in euphotic waters compared to the mesopelagic zone (Fig. 2 and Supplementary Table 4), including proteins involved in light-driven processes, such as proteorhodopsin ($P = 0.03$) (Supplementary Fig. 4) and the carbon fixation protein ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which was undetected deeper than 120 m. Energy-associated processes were differentially abundant with inorganic pyrophosphatase ($P = 2 \times 10^{-3}$), an ATPase ($P = 0.01$), and the oxidative phosphorylation protein mitochondrial cytochrome c oxidase ($P = 0.01$) more abundant in the euphotic zone and indicative of enhanced growth and cellular respiration, by autotrophic and/or heterotrophic cells.

Dinoflagellate transcriptomes also shifted with depth, reflecting changes in the trophic strategies utilized (Fig. 3a and Supplementary Figs. 5 and 6). A weighted correlation network analysis (WGCNA) clustered 1,663 dinoflagellate Kyoto Encyclopaedia of Genes and Genomes (KEGG)-annotated genes into module eigengenes, which were further examined for association with physiochemical parameters[43]. One module contained genes highly expressed in the euphotic zone and positively correlated with surface conditions; another contained genes more highly expressed in the mesopelagic zone and correlated with deeper ocean physiochemical parameters (Extended Data Fig. 5). Similar to protein signatures, euphotic zone dinoflagellates expressed transcripts encoding photosynthetic carbon fixation components (carbonic anhydrases, RuBisCO, photosystem II) and light-driven proteorhodopsin, and also imported nitrogen and trace metals to fulfil cellular nutrient demands (ammonium, urea, metal transporters) (Supplementary Table 7). In addition to photosynthesizing, dinoflagellates were likely phagotrophic given the detection of lysosomal components including nine V-type proton-transporting ATPase subunits, phosphatidylinositol 3-kinase, lysophospholipase III and cathepsin C and H[26,44,45]. Such proteins may also play a role in internal nutrient cycling, similar to autotrophic processes[46]. These genes were detected in the same genera that were photosynthesizing, as indicated through chlorophyll *a*-binding proteins and RuBisCO (Extended Data Fig. 6 and Supplementary Fig. 7), and could represent mixotrophic organisms or distinct trophic strategies being used among species. Accordingly, a KEGG pathway enrichment analysis conducted on transcripts classified genes into photosynthetic carbon fixation, carbon metabolism, secondary metabolite synthesis and oxidative phosphorylation pathways that were significantly enriched in the surface module (hypergeometric test, Benjamin–Hochberg adjusted P values < 0.05; Extended Data Fig. 5). These transcriptional and protein patterns suggest an actively growing phototrophic and phagotrophic dinoflagellate community in the euphotic zone and are consistent with the metabolic profiles of mixotrophic protistan cultures grown in the light[47], natural dinoflagellate communities along the California coast[26] and coastal dinoflagellate blooms[48,49].

Altered metabolic functionality in the mesopelagic zone. Mesopelagic waters showed a distinct dinoflagellate metabolic profile dominated by cytoskeletal components with indications of nutrient recycling and phagotrophy (Fig. 2). Three dinoflagellate proteins were significantly more abundant in the mesopelagic zone, including the cytoskeletal components tubulin ($P = 6 \times 10^{-3}$) and actin ($P = 4 \times 10^{-2}$), which were among the most abundant proteins at depth in the metaproteomic dataset (Supplementary Table 3), and EF hand domains ($P = 0.06$). Actin family proteins are key components of phagocytosis and cytoskeleton remodelling in addition to flagellar motility, adhesion, cell division and the feeding apparatus[50–54]. Accordingly, heightened expression may reflect phagotrophy, prevalence of resting stages and/or altered swimming behaviour. The elevated actin and tubulin proteins in the mesopelagic zone could be driven by higher phagotrophic activity since cytoskeletal components are upregulated in mixotrophic and heterotrophic protists under grazing conditions[44,55–58]. However,

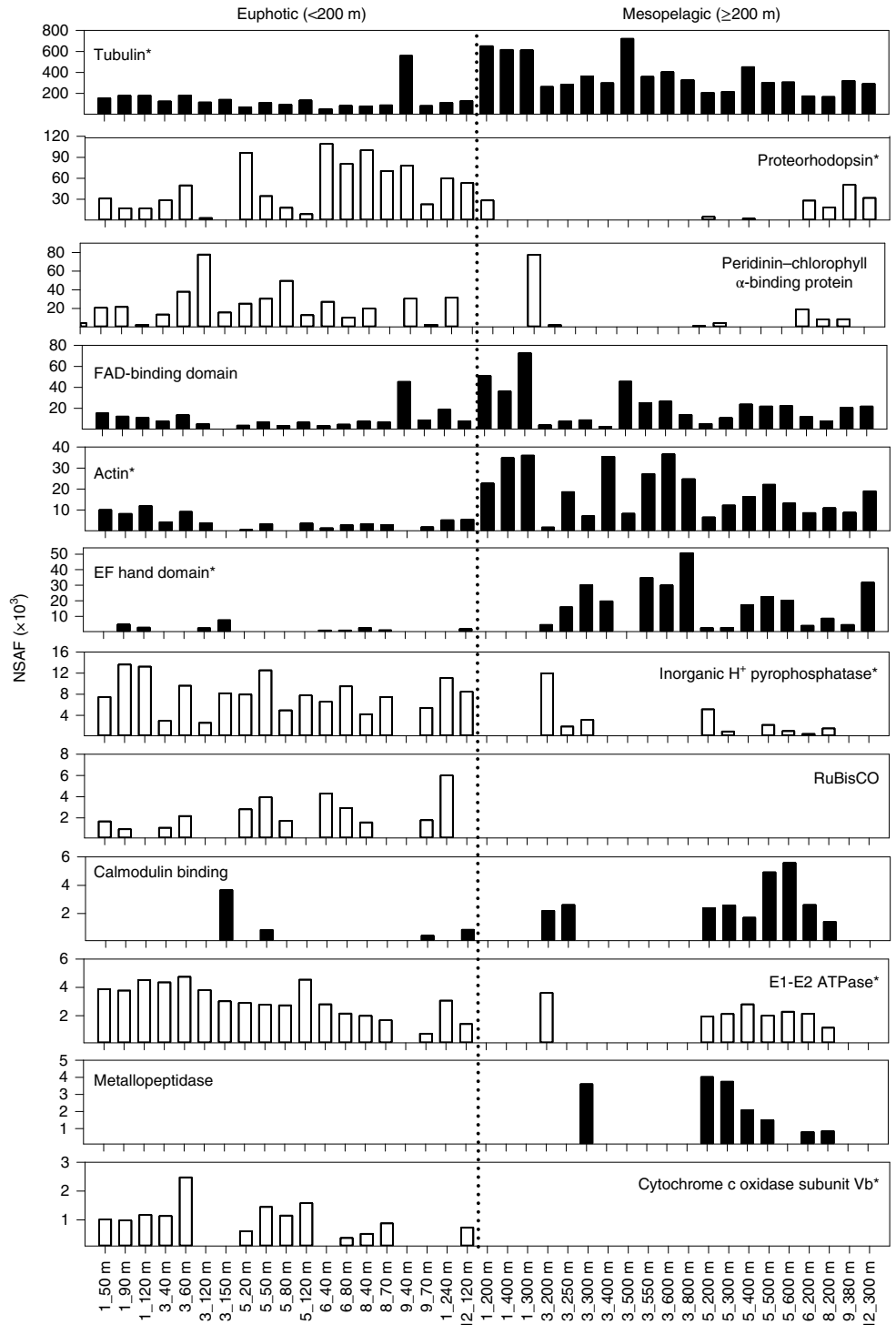


Fig. 2 | Differentially abundant dinoflagellate proteins between the euphotic and mesopelagic zones. The asterisks represent significantly differentially abundant proteins, determined using a two-tailed general independence permutation test followed by multiple test correction (<200 m, $n = 19$ biologically independent seawater samples; ≥200 m, $n = 20$ biologically independent seawater samples; $P < 0.1$; Supplementary Table 4). Pfam-annotated proteins more abundant in the surface are shown in white and those more abundant in the deep ocean are shown in black. Exclusive spectral counts were normalized using the NSAF approach.

the cytoskeletal machinery is not upregulated in all grazing mixotrophs[47,59] and the trait may therefore be taxon-specific and depend on feeding mechanism. A reprolysin-like zinc-binding metallopeptidase was the only peptidase detected to be elevated in mesopelagic waters (Fig. 2) and may be an important contributor to protein degradation. In addition, calcium-signalling proteins including calmodulin-binding proteins and EF hand domains were abundant and may facilitate cell signalling efforts in mesopelagic waters[60,61].

A collection of genes were more highly expressed in mesopelagic waters and positively correlated with nutrient concentrations elevated at depth, including nitrate, phosphate, iron and cobalt (Extended Data Fig. 5). These genes encode proteins involved in

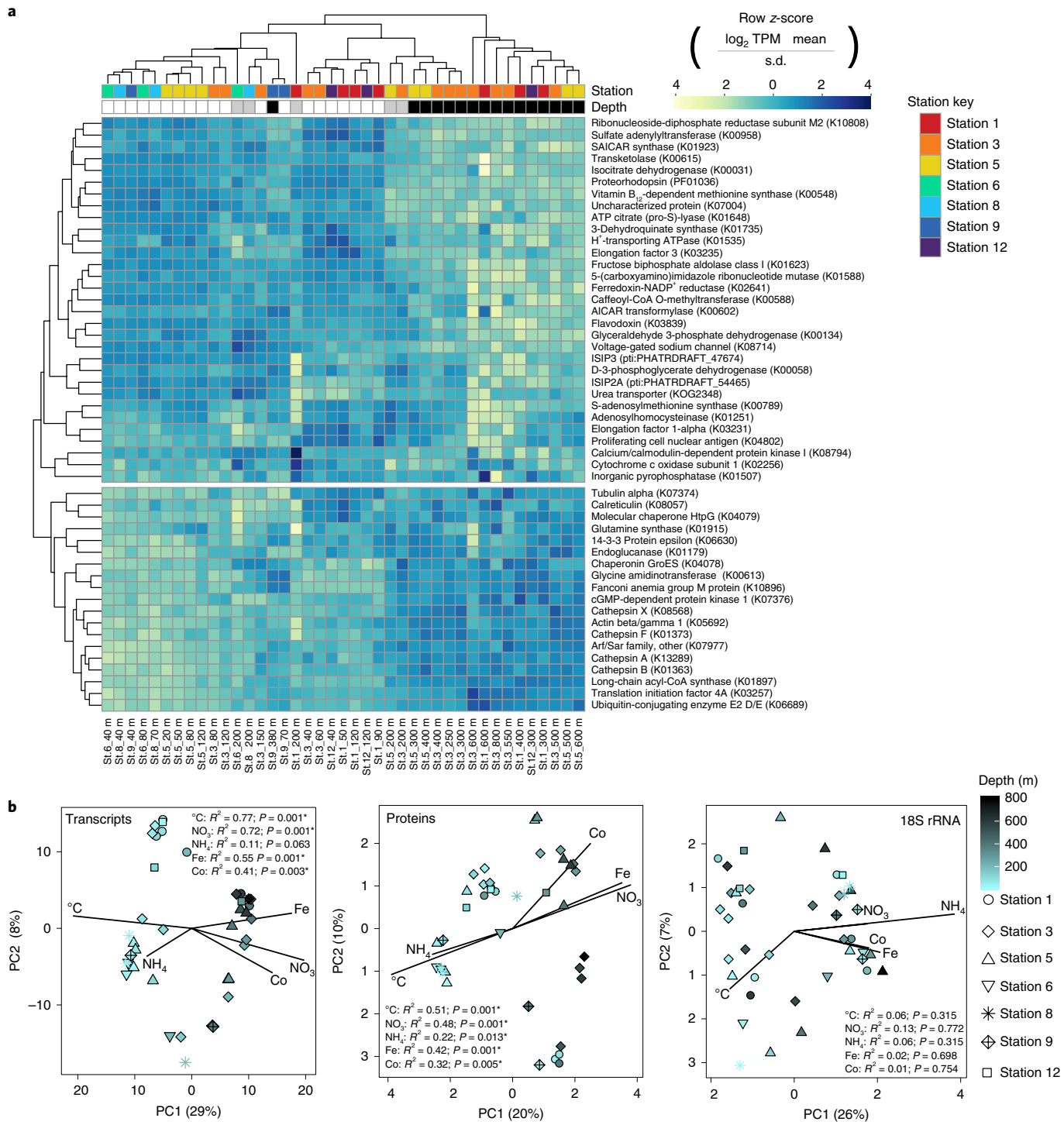


Fig. 3 | Distinct dinoflagellate functional metabolism between the euphotic and mesopelagic zones of the central Pacific. a, Heatmap displaying TPM-normalized gene expression of the top 50 KEGG-annotated genes with the highest transcript deviations from the mean (variances) across samples, shown as the *z*-score ($\log_2$ TPM − mean)/s.d.). The depth annotation bar highlights samples from the surface (<200 m, white), deep (>200 m, black) and exactly 200 m (grey). The dendrograms show similarity in transcript abundances as determined with Euclidean distance and hierarchical clustering. Each row represents a unique KEGG-annotated gene, with ISIPs and a KOG-annotated urea transporter manually added. The colour gradient represents low (yellow) to high (blue) gene expression. **b**, PCA of dinoflagellate metatranscriptomes, metaproteomes and 18S rRNA OTU data, with overlaid vectors indicating the relationship between environmental parameters and ordination axes as determined by a two-tailed permutation test (asterisks denote $P < 0.05$). Axes percentages represent the proportion of variance explained by each principal component.

intracellular recycling of polysaccharides and fatty acids (endoglucanase, long-chain acyl-CoA synthase), lysosomal degradation (acid phosphatase, lysosomal acid lipase, vesicle-associated clathrin, dynamin GTPase and seven cathepsin proteases) and cytoskeletal and motility components (actin, tubulin) (Fig. 3a and Supplementary Table 7).

Congruency between transcripts and proteins was evaluated by comparing the average fold change in expression between euphotic and mesopelagic zones (Extended Data Fig. 7). Most transcript and protein signatures were not tightly correlated, which was expected given these are mixed dinoflagellate communities and protein and transcript inventories are under different regulatory controls[62]. However, key processes were differentially abundant and correlated between fractions, including carbon fixation through RuBisCO and light-driven ATP production via proteorhodopsin in surface waters and cytoskeletal tubulin and cathepsin protease in the mesopelagic zone (Supplementary Table 6), indicating concerted cellular efforts to maintain both molecular pools.

Shifting functional and ecological roles within a continuous taxonomic community. The combined multi-omics survey revealed that euphotic zone dinoflagellates utilize light energy to generate ATP through photosynthesis and proteorhodopsin, while acquiring inorganic and organic nutrients and likely also performing phagotrophy. In particular, proteorhodopsin enabled a form of phototrophy complementary to photosynthesis[63,64] and is widely utilized among prokaryotes of the oligotrophic North Pacific[65]. Transcripts and proteins for proteorhodopsin were attributed to the phototrophic genera *Karenia, Prorocentrum* and *Alexandrium*, and the primarily heterotrophic genus *Oxyrrhis*[66,67]. The proteorhodopsin proton pump can be driven by V-type ATPases; hence, these proteins may have a dual role in acidifying food vacuoles as well as producing ATP via proteorhodopsin[63]. Little evidence of photoautotrophic metabolism remained deeper than 120 m, with RuBisCO proteins no longer detected. The equatorial Pacific euphotic depth ranges between 100 and 150 m based on historical measurements[68–70], which is consistent with our observed transitions in protein metabolism. Mesopelagic dinoflagellates increased expression of cytoskeletal components and nutrient recycling, and maintained cellular degradation and phagotrophic processes observed in surface waters. One potential explanation for this distinct metabolic profile is cyst formation. Dinoflagellates can form temporary cysts during unfavourable conditions, resulting in a reduced metabolic state with major cytoskeletal rearrangements[71,72]. *Lingulodinium polyedrum* cysts show enhanced cytoskeleton, ubiquitinylation, enzymatic degradation, calcium-signalling and RNA splicing phosphoproteins[72], similar to our observations in the mesopelagic zone. Cysts are not known to regulate buoyancy and would be expected to sink to the seafloor as observed among coastal species[73], with this evolved strategy beneficial in coastal regions where seawater resuspension occurs. Alternatively, elevated expression of phagocytosis-related genes, including cathepsin digestive proteases, calreticulin and clathrin have also been detected in actively feeding protists[44,55,74], and changes in trophic activity may also be driving the observed pattern. Despite the apparent disadvantages of cyst formation in offshore environments where encysted cells are lost to the deep ocean, we cannot rule out the simultaneous presence of cysts and phagotrophy in the present study; the possibility of maladaptive oceanic cyst formation in dinoflagellates could be the subject of future studies.

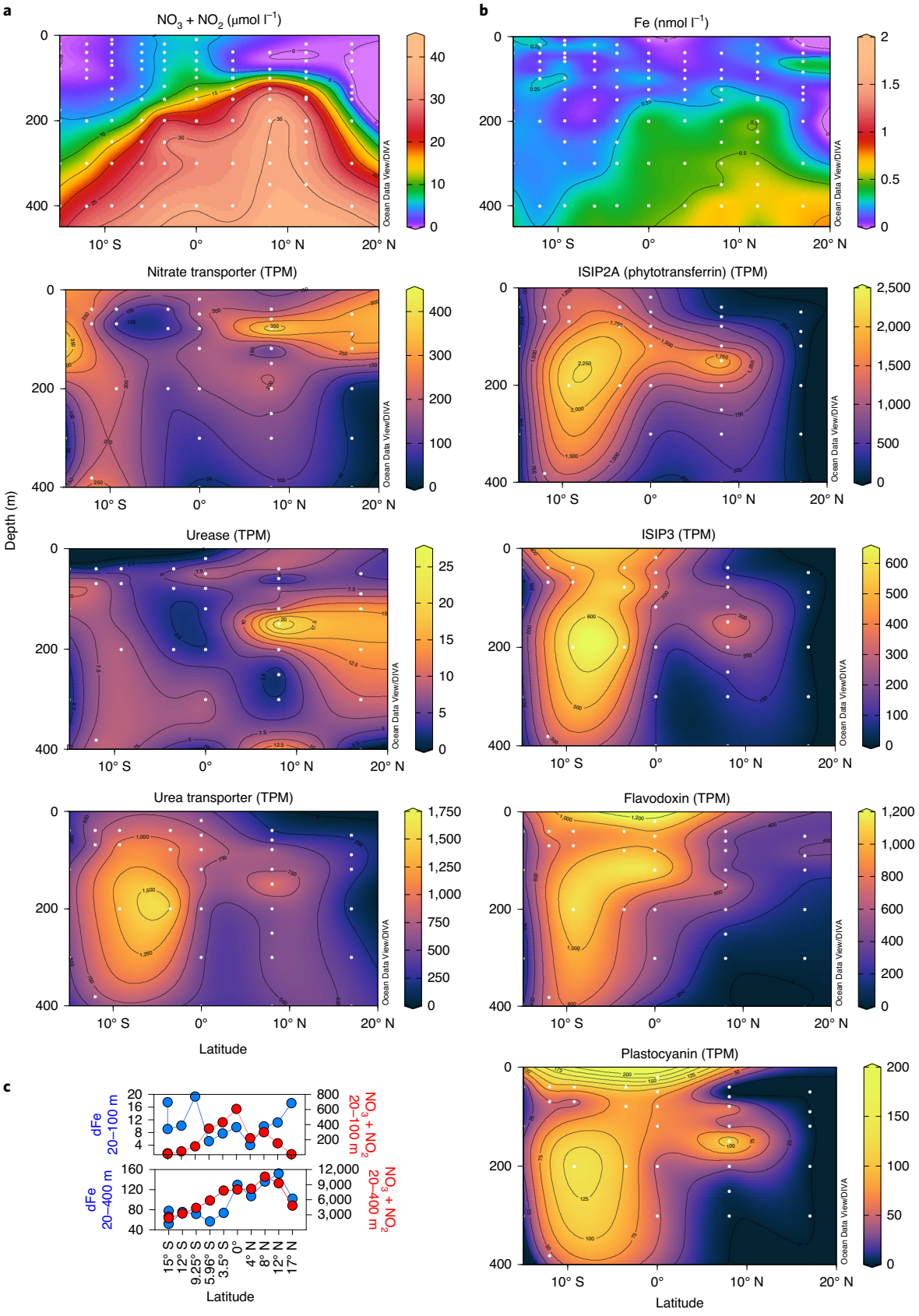
Taken together, these differences in functional profiles between depths reflects metabolic shifts within the same taxonomic groups rather than changes in dinoflagellate community composition. First, taxonomic and functional relationships across latitudes and depths were assessed using principal component analysis (PCA) with dinoflagellate transcripts, proteins and 18S rRNA OTUs. Transcripts and proteins separated by depth along the first component axis (PC1) while 18S rRNA OTUs did not (Fig. 3b). Physiochemical parameters characteristically changing with depth, including temperature, nitrate + nitrite, cobalt and iron, explained a portion of the variability in transcript and protein ordinations, yet did not in the 18S rRNA ordination since no separation by depth was observed. This indicates that the dinoflagellate community, at a coarse taxonomic level, did not undergo pronounced changes with depth. Second, since the V9 18S rRNA analysis did not resolve the fine taxonomy of the dinoflagellate community, we constructed a phylogenetic tree to determine whether separate taxonomic populations in the water column, or ecotypes, may explain the contrasting phenotypes between depth zones (Extended Data Fig. 8). The phylogenetic tree and subsequent cladogram was built using dinoflagellate-annotated tubulin contigs and reference dinoflagellate proteins derived from the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP)[75]. Diverse dinoflagellates contributed to tubulin gene expression in the mesopelagic zone, with no lineages distinctly responding in the mesopelagic zone in a manner consistent with deep-adapted ecotypes. Third, genera-based assessments can be made using metatranscriptomic and metaproteomic data (Extended Data Fig. 2), which show that dinoflagellate relative abundance was generally uniform with depth and latitude. Therefore, the dinoflagellate communities across the central Pacific Ocean appear to belong to a continuous taxonomic community demonstrating physiological plasticity between depth zones.

Nitrogen and iron stress across the biogeochemical gradient. Strong spatial biogeochemical gradients were observed along the transect. Consistent with historical measurements[76], equatorial surface waters between 0 and 6° S (stations 5–7) were elevated in nitrate + nitrite with an average of 5.6 μmol l^{-1} in surface waters above 100 m, while higher latitude regions were comparatively lower with an average of 0.95 μmol l^{-1} in northern (17–8° N; stations 1–3) and 0.24 μmol l^{-1} in southern tropical gyre surface waters (12–15° S; stations 9–12) (Fig. 4a). Total dissolved iron concentrations reflected a different latitudinal pattern with low concentrations (<0.25 nmol l^{-1}) south of the equator extending 300–400 m into the mesopelagic zone (Fig. 4b). Biomass inferred from total protein concentrations also indicated the highest levels south of the equatorial upwelling region, suggesting iron was depleted by growth (Supplementary Fig. 8).

Established molecular biomarkers of nitrogen[77,78] and iron stress[79–82] indicated physiological responses to shifts in nutrient regimes. Iron stress biomarkers characterized in other algal taxa were also utilized by dinoflagellates, including phytotransferrin (iron starvation-induced protein 2A (ISIP2A))[82–84], ISIP3 (refs. [81,83,84]) and the iron-independent photosynthetic electron acceptors flavodoxin[81,85,86] and plastocyanin[87]. Biomarkers demonstrated on average greater than twofold higher gene expression in equatorial upwelling and south Pacific surface waters (0–12° S, ≤200 m) containing low dissolved iron concentrations (<0.25 nM) compared to higher latitude waters (15° S, 8–17° N) (Mann–Whitney U-test = 1–20, $n_1 = 12$, $n_2 = 13$, $P \leq 0.002$) (Fig. 4b and Supplementary Table 8). Although a few previous studies discussed the presence of dinoflagellate ISIP2A[15,82,83], its relationship with iron scarcity has been weak[15]. In

Fig. 4 | Dinoflagellates restructure iron and nitrogen metabolism depending on external concentrations. a,**b**, Mid-depth vertical sections of dissolved nitrate + nitrite (**a**) and iron (**b**) plotted using Ocean Data View with DIVA interpolation (top). Characterized iron and nitrate stress biomarkers are depicted as TPM-normalized dinoflagellate gene expression across the transect (bottom). **c**, Depth-integrated nitrate + nitrite (mmol m^{-2}) and iron (μmol m^{-2}) between 20 and 100 m (top) and 20 and 400 m (bottom) across latitudes. Depth integrations were performed by computing the area below the curve following the trapezoidal rule in SigmaPlot v.14. For stations 3–5, 20-m samples were not collected and 40-m values were instead used as surface estimates. If 100-m measurements were not available, a linear extrapolation was performed using adjacent depths.

this study, dinoflagellate origins for low iron-induced ISIP2A transcripts were supported by a phylogenetic analysis (Supplementary Figs. 9 and 10) and their distribution was consistent with physiological acclimation to low iron. These patterns indicate that equatorial dinoflagellates possess mechanisms to cope with iron stress, similar to cyanobacteria in this ecosystem[4].



In addition to these iron stress signals, differences in nitrogen metabolism were also observed across the transect. Selected nitrogen stress biomarkers in dinoflagellates included transporters for nitrate (nitrate transporter)[88,89] and urea (urea transporter)[89] and the enzyme urease, which converts urea to ammonium[89,90]. The urea transporter displayed twofold higher gene expression in equatorial upwelling and southern tropical waters compared to adjacent oligotrophic regions, matching the distributions of the iron stress biomarkers (Fig. 4). This urea transporter could aid in nitrogen assimilation via reduced nitrogen compounds rather than relying on iron-intensive nitrate and nitrite reductases. Higher gene expression in equatorial waters may also be due to higher flux of organic nitrogen yet lower urea concentrations, which could stimulate transporter production[89]. In contrast, the nitrate transporter showed 1.5-fold higher gene expression in oligotrophic surface waters between 8 and 17° N and at 15° S compared to equatorial upwelling-influenced waters, demonstrating an inverse relationship with nitrate + nitrite concentrations (Fig. 4a). Urease also reached peak levels in oligotrophic waters at 10° N and is consistent with a low-nitrogen cellular response[89]. This suggests that the dinoflagellate community was using distinct nitrogen acquisition systems across the transect: nitrate in low-nitrogen northern tropical waters and external urea in iron-stressed southern tropical waters. These patterns are in contrast to *Prochlorococcus*, which primarily uses urea transport proteins in low-nitrogen northern tropical waters[4] and generally prefers urea and ammonia over nitrate[91,92]; therefore, it has minimal iron requirements for nitrogen assimilation. These results suggest that the iron costs associated with nitrogen assimilation could alter the physiology of dinoflagellates, yet not *Prochlorococcus*, with the two lineages preferentially relying on urea in different regions.

Contrasting metabolic profiles between oligotrophic and equatorial upwelling regimes. Beyond nitrogen and iron stress biomarkers, other metabolic processes responding to transitions in nitrogen or iron availability across the gradient were investigated (<100 m; Extended Data Fig. 9). Gene expression profiles generally separated into two divisions: equatorial upwelling and oligotrophic sites. In addition to the iron stress biomarkers, an α-carbonic anhydrase, a metalloenzyme responsible for concentrating CO_2 for carbon fixation[93], was more highly expressed in the low-iron equatorial upwelling waters. α-Carbonic anhydrase has been observed to scale with growth rate in marine diatoms to assist in carbon aquisition[94] and may therefore be more abundant in macronutrient-rich upwelling waters supporting elevated growth rates. Conversely, genes differentially expressed during protistan grazing conditions, including lysosomal cathepsin proteases, calreticulin involved in the phagosome and tubulin cytoskeletal proteins[44,55,74], as well as ubiquitin degradation, were elevated in expression in nitrogen-deficient oligotrophic regions (Extended Data Fig. 9). Cathepsin proteases, cytoskeletal and ubiquitin degradation proteins were likewise more highly expressed in dinoflagellates residing in the mesopelagic zone compared to the euphotic zone.

These metabolic profiles derived from meta-omics data reflect a composite of the dinoflagellate community likely consisting of autotrophs, heterotrophs and mixotrophs. It is therefore difficult to determine whether the transitions in metabolism from nitrogen-deficient oligotrophic to nitrate-rich upwelling region are due to changes in dinoflagellate species with differing trophic strategies, or the extent of phagotrophy performed within a uniform mixotrophic community. Laboratory evidence supports mixotrophy as a strategy for fulfilling nitrogen demands, with nitrogen-starved cultures of the mixotrophic *Karenia brevis* increasing growth by grazing on the cyanobacterium *Synechococcus*[30]. The metabolic and geographical patterns presented in this study suggest dinoflagellates may respond to changes in macronutrient and metal supplies across the gradient by altering trophic modes, variably expressing nutrient acquisition genes, and by substituting iron- or nitrate-dependent processes with alternative mechanisms (for example, flavodoxin and urea), akin to the strategies utilized by marine diatom populations acclimating to low iron[86,95,96] or nitrate[77,97].

Dinoflagellate co-occurrence with enhanced export flux and biogeochemical importance. Microbial biomass collected from the water column was compared with particulate carbon flux derived from sediment traps at three depths along the transect (Fig. 5). Non-normalized dinoflagellate spectral counts linearly scaled with particulate fluxes and increased equatorward by approximately twofold ($R^2 = 0.99$, $n = 3$; Extended Data Fig. 10 and Supplementary Table 9). The carbon fluxes were fitted to a power law function to calculate the coefficient of flux attenuation[98] (b value), which increased along the transition from oligotrophic (1.15) to equatorial (1.73) waters and was consistent with shallower remineralization[99] of the increased flux. Equatorial dinoflagellate spectral counts remained elevated into the mesopelagic zone (600 m). These protein distributions contrasted from that of *Prochlorococcus*, which decreased equatorward and towards the mesopelagic zone (Fig. 5). These patterns were most striking in the upper mesopelagic (200–400 m), where depth-integrated spectral counts decreased by 67-fold for *Prochlorococcus* and increased by sevenfold for dinoflagellates from oligotrophic to equatorial waters (Extended Data Fig. 10). These results suggest that dinoflagellates benefit from and/or contribute to the enhanced nutrient fluxes associated with equatorial upwelling compared to oligotrophic conditions. Supporting this, a recent global multi-omics study identified a positive correlation between dinoflagellate lineages and carbon export at 150 m, inferred from particle size and abundance[19]. The capability for phototrophic and heterotrophic nutritional modes likely allows dinoflagellates to contribute to export flux through their euphotic zone photosynthetic activities and remineralization processes within the mesopelagic zone in the same regional environment.

Conclusions and implications

This sectional analysis elucidated connections between biogeochemical cycles, resource availability and microbial physiology across gradients of the central Pacific Ocean. Dinoflagellates were relatively abundant and prevalent throughout the transect, appearing as a continuous community at the genus level from the euphotic to the mesopelagic zone. Distinct vertical shifts in metabolism were observed through a combination of photosynthetic carbon fixation and phagotrophy in euphotic waters and a heavy investment in cytoskeletal proteins and internal nutrient recycling by heterotrophic communities at depth. Vertical migration via flagellar motility is known to occur in dinoflagellates and could explain euphotic–mesopelagic transitioning[100,101], allowing for photosynthesis in the euphotic zone and access to inorganic resources and prey in the nutrient-rich mesopelagic zone[100]. However, swimming speeds of dinoflagellates span approximately 9–60 m day^{-1} (ref. [100]) and could not explain dinoflagellate presence down to 800 m. Instead, we hypothesize that mesopelagic communities are seeded from large surface mixotrophic populations. These organisms may settle in the water column and continue to survive by relying solely on heterotrophic metabolism. Alternatively, Kareniaceae dinoflagellates may be particularly well adapted to the tropical Pacific, with distinct autotrophic/mixotrophic species in euphotic environments and heterotrophic species separately thriving in the mesopelagic zone. Future research efforts will be required to confirm mixotrophy, examine taxonomic relationships between euphotic and mesopelagic populations and determine whether these dinoflagellates possess the functional capabilities to perform euphotic to mesopelagic transitioning.

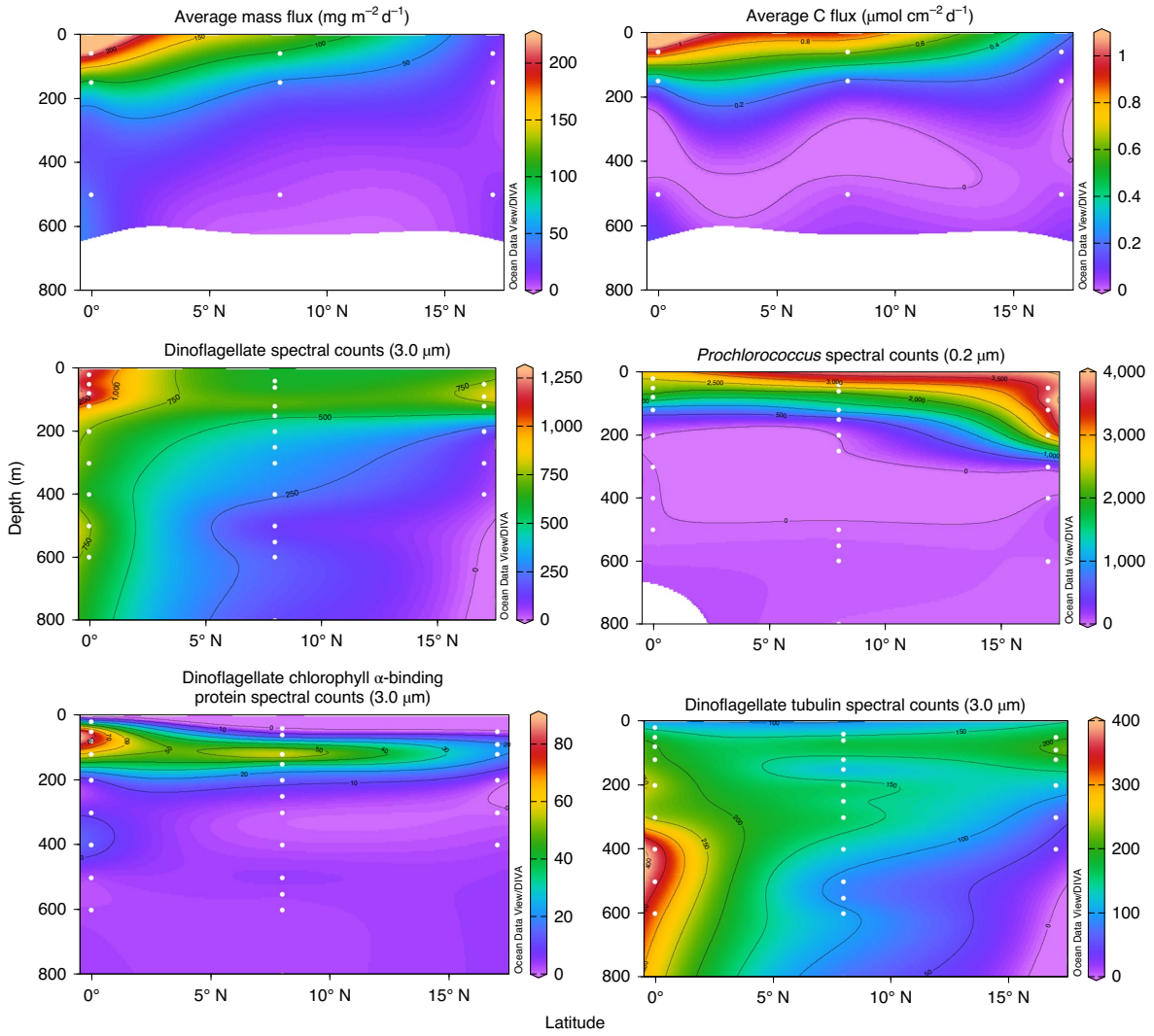


Fig. 5 | Dinoflagellate protein abundance is linked to carbon export in the central Pacific Ocean. Top: particulate mass flux and carbon flux along the oligotrophic to equatorial biogeochemical gradient. Dinoflagellate protein spectral counts derived from the 3–51-μm filter size fraction. Middle: *Prochlorococcus* spectral counts from the 0.2–3-μm filter size fraction. Bottom: changes in dinoflagellate protein inventory between the euphotic and mesopelagic zones are illustrated with chlorophyll *a*-binding protein (Pfam ID: PF00504), abundant in the euphotic zone, and tubulin (Pfam IDs: PF00091 and PF03953) primarily associated with mesopelagic water. The dinoflagellate- and *Prochlorococcus* spectral counts depicted in this figure were not normalized as performed in the functional analysis to allow for geographical comparisons.

Adaptive responses were also observed across the horizontal sectional gradient and illustrated key differences between dinoflagellates and the abundant cyanobacterium *Prochlorococcus*. Dinoflagellate metabolic profiles reflected the geochemical transition from oligotrophic gyres to an equatorial upwelling regime sustaining higher dinoflagellate biomass. Transcript profiles were consistent with mixotrophic populations relying on phagotrophy to a greater extent in oligotrophic waters as a means to confront increased nitrogen scarcity. In the equatorial upwelling zone, dinoflagellate populations experienced iron stress and utilized urea potentially to minimize the iron costs associated with nitrate assimilation, whereas *Prochlorococcus* primarily relied on urea in northern oligotrophic waters where it achieves higher abundance[4]. Dinoflagellates were positively correlated with carbon flux, either directly contributing to export or benefiting from elevated organic and inorganic resources. These observations may be consistent with a global plankton model linking mixotrophy to enhanced vertical carbon flux[10].

The ecological success of dinoflagellates across these diverse habitats is likely related to their large genome size, hypothesized to be the result of selective pressure to maintain genetic capabilities and exploit broad resources across variable environments[102–104]. This is in contrast to the reverse strategy of genome streamlining typical of *Prochlorococcus* and its limited abilities to utilize diverse resources[104,105]. Interestingly, both genome expansion and reduction strategies co-exist in these respective populations to facilitate resource partitioning, allowing dinoflagellates to thrive as an abundant microeukaryote possessing multiple functional modes across vertical and latitudinal gradients in the central Pacific ecosystem.

Methods

Microbial biomass collection. Oceanographic sampling during the METZYME expedition (KM1128) occurred between 1 and 25 October 2011 onboard the R/V Kilo Moana as also described in Saito et al.[4]. The meridional transect (17° N–15° S) began off the Hawaiian Islands (154.4° W) and terminated in the Tonga-Fiji region (173.1° W) (Fig. 1a). Forty-two samples were collected for biomass at 7 sites and between 3 and 13 depths per site, with the majority between 40 and 600 m (Extended Data Fig. 1). Biomass collection along the vertical profiles for metaproteomic, metatranscriptomic and 18S rRNA metabarcoding analyses were performed using battery-operated underwater McLane pumps (McLane Research

Laboratories) outfitted with custom filter head units secured onto a trace metal clean winch line. Sampling depths were predetermined based on oceanographic features (for example, surface, deep chlorophyll maximum, oxygen minimum zone) (Supplementary Fig. 1). Each McLane pump head was fitted with three 142-mm Supor filters (Pall Corporation) for targeting specific size classes of the microbial community: 0.2–3 µm (prokaryotes and picoeukaryotes[4]); 3–51 µm (filamentous prokaryotes, protists and zooplankton); and ≥51 µm (mesozooplankton and sinking particles). For this analysis focused on the eukaryotic phytoplankton community, the 3–51-um fraction was mainly considered. *Prochlorococcus* proteins used as a pelagic prokaryotic comparison belong to the 0.2–3-µm fraction[4,106]. Pumps filtered between 165 and 1,384 l of seawater with an initial filtering rate of 4 or 8 l per minute. After pump retrieval, filters were promptly sectioned for multi-omics analyses (rRNA, transcripts, proteins) and frozen at −80 °C until onshore processing. High resolution multi-omics sampling was achieved across the meridional transect using the McLane pumps, resulting in three mid-depth profiles (20–50 to 600–800 m) at stations 1, 3 and 5, and 4 upper-water column profiles (40 to 200–400 m) at stations 6, 8, 9 and 12 (Extended Data Fig. 1). Physiochemical data collection consisted of conductivity, temperature and depth casts, measurements of dissolved trace metals, macronutrients, pigments and estimates of particulate flux from sediment traps, resulting in mid- and full-depth vertical profiles.

Dissolved trace metals. Seawater samples intended for dissolved trace metal quantification were collected using a trace metal clean rosette consisting of 12 approximately 8-l X-Niskins (General Oceanics) on a non-conducting line outfitted with temperature, salinity, oxygen, chlorophyll fluorescence, transmissivity, conductivity and pressure sensors[4]. After seawater collection, the X-Niskins were brought into a fabricated class-100 clean room where they were pressurized with high purity nitrogen gas and seawater was filtered through 0.2-µm Supor membranes to remove particulate material. HEPA filters were used to minimize particle contamination on the ship and polyethylene gloves were used when handling samples. Trace metal clean guidelines were followed throughout the collection, filtering and sample handling processes as outlined by the GEOTRACES international community[107,108]. Polyethylene collection bottles were rigorously cleaned by soaking for 3 d in citranox detergent followed by ultrapure Milli-Q (Merck Millipore) water rinses, soaking for 2 weeks in 10% HCl (Baker instra-analysed grade) and finally by rinsing with HCl (pH 2). Bottles were stored empty and double-bagged. Dissolved Co was measured via cathodic stripping voltammetry following a 1-h ultraviolet oxidation[4]. Total dissolved Fe was preconcentrated using the magnesium hydroxide precipitation method and analysed on an Element-2 inductively coupled plasma mass spectrometer (Thermo Fisher Scientific)[109]. Concentrations were determined using isotope dilution calculations with indium (^{115}In) as an internal standard to account for matrix effects and recovery efficiency. GEOTRACES intercalibration consensus standards were used to verify accuracy, with North Atlantic GS and GD dissolved Fe (dFe) measurements of 0.58 ± 0.10 and 1.06 ± 0.16 nmol l^{-1}, respectively, falling within the GEOTRACES dFe consensus ranges of 0.56 ± 0.05 and 1.03 ± 0.10 nmol l^{-1} for GS and GD standards. Of the 110 upper water column seawater samples measured, three dFe points were excluded from the section plot visualizations because they were anomalously high, inconsistent with nearby profile measurements and may have been the result of contamination. A table of the dFe measurements spanning surface to 500 m along the METZYME transect can be found in Supplementary Table 8.

Dissolved nutrients and pigments. Dissolved nitrate + nitrite ($NO_3 + NO_2$), nitrite (NO_2) and silicate (SiO_4) were measured on an Alpkem Rapid Flow Analyzer while ammonium (NH_4) and phosphate (PO_4) were measured on a Technicon AutoAnalyzer II using 2× the s.d. to estimate the limit of detection. For pigment determination, 4 l of seawater was filtered through GF/F filters and frozen in liquid nitrogen until analysis on an Agilent 1000 HPLC[4].

Sediment traps. Sinking particles were collected in acid-cleaned 250-ml low-density polyethylene bottles attached to polycarbonate particle collection tubes. Replicate tubes were placed in metal-free racks attached to a non-metallic line at depths of 60, 150 and 500 m at stations 1 (17° N), 3 (8° N) and 5 (0° N) using a surface-tethered system[110]. The bottles were deployed full of borate-buffered (pH 8.2) seawater brine prepared from freezing filtered seawater. Above the brine, each tube contained filtered seawater. Three capped tubes were deployed in the trap array as process blanks. On recovery, the tubes were allowed to sit for 1 h to allow for particles to settle. The collection bottles were then removed from the trap tubes and the contents filtered on either pre-weighed polycarbonate membranes (1 µm; Nucleopore) or combusted quartz fibre filters (QMA). The membranes were used to determine mass fluxes. The QMA filters were used for C and N flux determinations. Particulate carbon and nitrogen determinations were made on both sinking and suspended matter through high-temperature combustion analysis of quartz fibre filter subsamples on a Flash EA1112 (Thermo Fisher Scientific). For these samples, no attempt was made to distinguish between organic and inorganic particulate carbon. The *b* value was determined via nonlinear flux curve fitting to the Martin power law.

Global metaproteomics. Frozen filters were thawed and extracted according to an approach similar to the protocol described for 0.2-µm filters by Saito et al.[4]. In this study, 3.0-µm filters were extracted using an SDS detergent to solubilize both membrane and soluble proteins and subjected to a 10-min heating incubation at 95 °C. Protein material was subsequently purified, alkylated, reduced and digested with trypsin (mass spectrometry grade; Promega Corporation) while embedded in a polyacrylamide gel electrophoresis tube gel[111]. Proteins were quantified using a colorimetric BSA assay (DC Protein Assay; Bio-Rad Laboratories). Digestion was performed using the protease trypsin at a trypsin:protein ratio of 1:20 and peptides were purified using C18 tips (Pierce C18 Tips, 100-µl bed; Thermo Fisher Scientific). Purified peptides were diluted to 0.1 µg µl^{-1} and 20 µl (2 µg) was injected onto a Dionex UltiMate 3000 RSLCnano LC system (Thermo Fisher Scientific) with an additional RSLCnano pump run in online two-dimensional active modulation mode coupled to a Thermo Fusion Orbitrap mass spectrometer. The first separation used a nonlinear 8-h pH 10 gradient (10 mmol l^{-1} ammonium formate and 10 mmol l^{-1} ammonium formate in 90% acetonitrile) on a PLRP-S column (200 µm × 150 mm, 3 µm bead size, 300 Å pore size; nanoLCMS Solutions), diluted inline (10 µl min^{-1}, 0.1% formic acid) and eluted every 30 min on alternating dual traps (300 µm × 5 mm, 5 µm bead size, 100 Å pore size, C18 PepMap100; Thermo Fisher Scientific) (Supplementary Table 10). Alternating traps were eluted at a rate of 500 nl min^{-1} onto a C18 column (100 µm × 150 mm, 3 µm particle size, 120 Å pore size, C18 ReproSil Gold (Dr. Maisch) packed in a New Objective PicoFrit column) with a 30-min nonlinear gradient (0.1% formic acid and 0.1% formic acid in 99.9% acetonitrile) on a Thermo Flex ion source. The mass spectrometer acquired MS1 scans from 380 to 1,580 *m/z* at 240,000 resolution with MS2 acquisition of charge states 2 to 10 and exclusion of undetermined charged states. MS2 scans had a 1.6 *m/z* isolation window, 50 ms maximum injection time and 5 s dynamic exclusion time at 5,000 resolution. The top speed data-dependent mode was used with a cycle time of 2-s, where as many dependent scans as possible were performed within the cycle time.

The 3–51-µm size fraction translated metatranscriptome (see below) was used as a reference protein database and peptide spectra matches were performed using the SEQUEST algorithm within Proteome Discoverer v.2.1 (Thermo Fisher Scientific) with a fragment tolerance of 0.6 Da and parent tolerance of 10 parts per million. Identification criteria consisted of a peptide threshold of 95% (false discovery rate (FDR) = 0.08%) and protein threshold of 99% (1 peptide minimum, FDR = 0.9%) in Scaffold v.4.8.7 (Proteome Software) resulting in 9,796 and 99,143 peptides identified. The mass spectrometry raw files, metatranscriptomic database and Proteome Discoverer search result files have been deposited with the ProteomeXchange Consortium through the Proteomics Identification Database (PRIDE)[112] repository under accession no. PXD014230. Approximately 75% of the proteins were associated with a taxonomic and/or functional annotation (Supplementary Table 3). To avoid double-counting mass spectra, exclusive spectral counts were used for the downstream proteomic analysis. Using this approach, only spectral counts assigned to a single protein were considered. Exclusive dinoflagellate spectral counts were normalized using the normalized spectral abundance factor (NSAF) calculation[113] to allow for a comparison of dinoflagellate protein abundance across samples and remain consistent with the metatranscriptomic procedure:

$$\mathrm{NSAF} = \frac{x_i}{l_i} \times \frac{1}{\sum_i^x} \times 10^3$$

where x represents the spectral counts of protein i and l represents the length of the open reading frame (ORF) in the amino acid residues. These normalizations were performed separately for each taxonomic group of interest (for example, dinoflagellates). Counts associated with redundant ORFs (sharing identical taxonomic and functional assignments) were summed together. We present data using both NSAF (Fig. 2, Extended Data Fig. 4 and Supplementary Fig. 3) and non-normalized (exclusive) spectral counts (Fig. 5 and Extended Data Fig. 10). The NSAF approach enabled a relative comparison between euphotic and mesopelagic samples and non-normalized counts provide information on protein inventory. Although the same amount of peptides (2 µg, as measured using predigested protein concentrations) were injected into the mass spectrometer across all samples, total spectral counts (peptide spectrum matches (PSMs)) decreased by 1.8-fold in the mesopelagic zone (Mann–Whitney U-test = 74, $n_1 = 19$, $n_2 = 21$, $P < 0.001$; Supplementary Fig. 11 and Supplementary Table 11). The MS2 counts and resulting PSM/MS2 ratios also decreased in mesopelagic samples by 1.4-fold (Mann–Whitney U-test = 42, $n_1 = 19$, $n_2 = 21$, $P < 0.001$; Mann–Whitney U-test = 99, $n_1 = 19$, $n_2 = 21$, $P = 0.007$, respectively). We hypothesized that the decrease in PSMs in the mesopelagic zone was due to a combination of factors, including poorer database coverage at a depth where diverse mesopelagic microorganisms are less represented and lower spectra quality due to organics interference and enhanced protein degradation/modification. Therefore, the relative analysis was used to observe differences in contribution to functional processes across latitudes and depths. To complement this, non-normalized (exclusive) spectral counts were also used to present an alternate visualization to further illustrate geographical and vertical differences in proteins. Absolute quantification using targeted proteomics could be employed in future analyses to determine concentrations of protein per volume of seawater[4,106].

MS2 were obtained from raw mass spectrometry (RAW) files and converted with ProteoWizard v.3 MSConvertGUI. The ProteoWizard msaccess function was used to filter mass spectra between 380 and 1,580 *m/z*. Means of PSM/MS2 ratios, MS2 PSMs (total spectral counts) between euphotic and mesopelagic depths were statistically compared in Sigmaplot v.14 using the Mann–Whitney *U*-test.

Metatranscriptomics. Frozen and sectioned 3–51-µm size fraction filters were thawed on ice and subjected to an RNA extraction and messenger RNA purification protocol closely following Bertrand et al.[96]. Briefly, RNA was extracted using the TRIzol reagent according to the manufacture's guidelines (Thermo Fisher Scientific). An RNeasy MinElute Cleanup Kit (QIAGEN) was used to purify RNA and ribosomal RNA was removed using the Ribo-Zero Magnetic Kit (Illumina). The mRNA fraction was amplified and converted to complementary DNA using the Ovation RNA-seq System V2 (NuGEN) and subsequently fragmented to 200 base pairs (bp). Libraries were prepared using Truseq RNA Sample Prep Kit V2 and paired-end sequences were generated on the Illumina HiSeq platform resulting in an average of 13 ± 2.5 million paired-end raw reads per sample. Metatranscriptomic reads are publicly available through the National Center for Biotechnology (NCBI) under Bioproject number PRJNA555787.

Raw reads were quality-trimmed and adaptors and rRNA sequences were removed using riboPicker v.0.4.3 (ref. [114]). CLC Assembly Cell (QIAGEN) was used for the de novo assembly with ORFs predicted using FragGeneScan v.1.16 (ref. [115]). Reads were mapped to ORFs using the Burrows–Wheeler Aligner-MEM[116]. ORFs were assigned taxonomy and function via BLASTp[117] with an *E*-value cut-off of 10^{-3} using the custom-built database PhyloDB, which consists of marine prokaryotic and eukaryotic genomes and transcriptomics (https://github.com/allenlab/PhyloDB), including protistan isolates from the MMETSP[75]. A lineage probability index was additionally used with the eukaryotic community to assign taxonomy more conservatively to closely related organisms[96,118]. Dinoflagellate ORFs with a conservative lineage probability index >80% were used for the functional analysis. To maximize functional assignments, sequences were searched against the KEGG[119] and EuKaryotic Orthologous Groups (KOG)[120] tools and conserved protein domain families were identified using HMMER v.3.1b2 (ref. [121]) with Pfam[122]. Of the assembled contigs, 62% were associated with a taxonomic annotation and approximately 54% were assigned a functional annotation (Supplementary Table 1). Dinoflagellate reads were normalized following the transcripts per million (TPM) approach[123,124] (Supplementary Table 2), similar to the proteomic NSAF normalization, allowing for a relative analysis across samples:

$$\mathrm{TPM} = \frac{x_i}{l_i} \times \frac{1}{\sum \frac{x}{l}} \times 10^6$$

where x represents the read counts of ORF i and l represents the length of the ORF I in the nucleotide base pairs. TPM normalizations were performed separately for each taxonomic group of interest (for example, dinoflagellates). Counts associated with ORFs sharing identical taxonomic and functional assignments were summed together at the supergroup level.

18S and 16S ribosomal RNA metabarcoding. Eukaryotic and prokaryotic taxa were characterized using V9 18S and V3–V5 16S ribosomal RNA metabarcoding, respectively, from the 3–51-µm filter size fraction. A 500-bp region of V3–V5 16S rRNA was amplified with the eubacterial primers 341F (5′-CCTACGGGNGGCWGCAG-3′)[125] and 926R (5′-CCGTCAATTCMTTTRAGT-3′)[126]. A 130-bp region of V9 18S rRNA was amplified with 1389F (5′-TTGTACACACCGCCC-3′) and 1510R (5′-CCTTCYGCAGGTTCACCTAC-3′)[17,127,128]. FLX Titanium adaptors were added to primers for Roche 454 sequencing (A adaptor sequence: 5′-CCATCTCATCCCTGCGTGTCTCCGACTCAG -3′; B adaptor sequence: 5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3′). The 18S rRNA raw reads are available on the NCBI under Biosample accession nos. SAMN12331629–SAMN12331670 and the 16S rRNA raw reads under Biosample accession nos. SAMN12332710–SAMN12332751.

The detailed cDNA prep and sequencing protocol is described in Bertrand et al.[96]. The 16S rRNA fraction was taxonomically classified using FASTA36 and the SILVA rRNA database release 111 (ref. [129]), while the 18S rRNA OTUs were searched against the eukaryotic 18S ribosomal taxonomic database Protist Ribosomal Reference v.4.11.1 (ref. [130]), containing curated dinoflagellate 18S ribosomal sequences from the dinoREF database[131]. Eukaryotic 16S rRNA plastid sequences were separated from the prokaryotic 16S rRNA fraction by searching against phytoREF release 1, a database containing aquatic and terrestrial plastid representatives from major eukaryotic lineages[132]. A modest number of sequences was generated to capture the dominant microbial community members since deep sequencing of the rare biosphere was not the intended goal. An average of 2,500 post-processed sequences per sample were generated for 18S rRNA, 6,000 for 16S rRNA (non-plastid) and 144 for 16S rRNA plastid fractions corresponding to approximately 3,100 18S rRNA, 4,800 16S rRNA (non-plastid) and 200 16S rRNA plastid OTUs (Supplementary Table 5). 18S rRNA OTU data were visualized using the R package phyloseq v.1.25.2. Nucleotide MUSCLE alignments[133] were constructed in Geneious v.11.1.4.

Metagenomic analysis. Fractions from the 0.2–3-µm filter were extracted for DNA and submitted to the Integrated Microbial Genomes and Microbiomes division of the Joint Genome Institute[134] for metagenomic sequencing. Quality-trimmed reads were assembled with SPAdes (using the -meta flag) and annotated as described in Dupont et al.[135]. The reference assembly was used as a protein database to identify prokaryotic proteins of interest. *Prochloroccocus* ORFs were identified using *Prochlorococcocus* NCBI taxon IDs. The 0.2–3-µm metagenomic assembly is publicly available through the NCBI under accession no. GCA_900411625.

Statistical analyses and data visualization. Beyond selecting individual genes or proteins of interest, a goal of the study was to determine whether broad metabolic differences exist between latitudes or depth regions. A WGCNA[43] approach was used to identify clusters of highly expressed genes across locations and depths and quantify relationships with physiochemical metadata. Eigengene modules were created using $\log_2$ TPM-normalized, KEGG-annotated dinoflagellate transcripts. At least 75 genes were required per module and modules displaying similar eigengene values across samples were merged (MEDissThres = 0.3). Eigengene values representing module expression were correlated with physiochemical data (Pearson's correlation test; $P < 0.05$). The default 'turquoise' and 'blue' colour modules were changed to 'white' and 'black', respectively, for clarity. Functional characterization of modules was performed using a KEGG enrichment analysis with the enrich function in clusterProfiler v.3.12.0, which performs a hypergeometric test to calculate overrepresentation of KEGG pathways compared to the total genes identified in the dataset (Benjamin–Hochberg adjusted $P < 0.05$)[136].

More Pfam annotations were available than KEGG annotations for the metaproteomic analysis; therefore, Pfam was selected to maximize the amount of information used to evaluate differences in proteins between depths. Differential abundance was determined using normalized Pfam-annotated dinoflagellate spectral counts between the surface (<200 m, $n = 19$) and deep (≥200 m, $n = 20$) layers with a two-sided asymptotic general independence (permutation) test using the R package coin v.0.6.6[137] with a maximum test statistic, default parameters and *P* values adjusted for multiple testing according to the max-*T* method (method = 'step-down').

To compare 18S rRNA community composition, relative transcript abundance and protein functional profiles across samples, a PCA was performed. 18S rRNA OTU data were log-transformed, subset to include only dinoflagellates and components calculated using phyloseq (ordinate function, method = 'RDA')[138]. TPM-normalized dinoflagellate transcript counts and NSAF-normalized, exclusive protein spectral counts were log-transformed. PCA was performed using the rda function in vegan v.2.3-0 (ref. [139]) and environmental vector fitting was performed using the envfit function. Environmental variables including temperature, ammonium, nitrate + nitrite, dissolved iron and cobalt were measured at the same depths and locations as multi-omics material. The PC1 axis of the protein ordination was flipped for visual consistency with the transcripts.

A phylogenetic analysis was performed using the protein tubulin to address whether distinct dinoflagellate taxonomic groups could be responsible for the contrasting metabolic features observed between the euphotic and mesopelagic zones and to investigate dinoflagellate ISIP2A phylogeny. Dinoflagellate reference sequences were obtained by searching contig 'contig_2077622_1_918', annotated as *K. brevis* CCMP2229 and tubulin C-terminal domain (Pfam PF03953), against dinoflagellates from the recently reassembled MMETSP database[75,140] using BLASTp v.2.7.1 (ref. [117]). The top five-hundred BLASTp output entries with the highest bit score ($E \leq 10^{-40}$) were manually curated to retain long and unique representative contigs from each isolate, with most isolates possessing multiple gene copies. Sequences were aligned using MUSCLE v.3.8 (ref. [141]) and trimmed in Jalview v.2.10.5 (ref. [142]) to allow for a maximum number of contigs (49) to be retained of the same the length, which resulted in 478 amino acid residues. Using the same approach, ISIP2A sequences were obtained by searching contig 'contig_1526186_1_756', annotated as *K. brevis* CCMP2229 and ISIP2A (pti:PHATRDRAFT_54465) against the full MMETSP database. The top 515 blast hits ($E \leq 10^{-40}$) were used to create a reference alignment of 110 contigs with 335 amino acids in length. These reference sequences were used to a build a phylogenetic tree onto which the metatranscriptomic dinoflagellate contigs were aligned. The maximum-likelihood tree was created using RAxML v.8.2.11 using the PROTGAMMALG model and 100 bootstrap replicates[143]. Contigs were placed on the reference tree using pplacer v.1.1alpha19 (ref. [144]). The cladogram was visualized with the R package ggtree v.1.16.0 (ref. [145]).

Section plots were created in Ocean Data View 4 using Data-Interpolating Variational Analysis (DIVA) interpolation[146]. Gene expression and protein abundance heatmaps were created with ggplot2 v.3.2.1. Bar plots and scatter plots were created in Sigmaplot v.14.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The mass spectrometry global proteomics data and metatranscriptome-derived FASTA file has been deposited with the ProteomeXchange Consortium through

the PRIDE[112] repository under accession number PXD014230. Metaproteomic annotations and total spectral counts from this analysis are also available on the Ocean Protein Portal (proteinportal.whoi.edu/). Nutrients, dissolved cobalt, pigments and conductivity, temperature and depth physiochemical information is available through the NSF's Biological and Chemical Oceanography Data Management Office (BCO-DMO) repository under project number 2236. Metatranscriptomic reads have been deposited with the NCBI under Bioproject no. PRJNA555787. The 16S rRNA raw reads are available on the NCBI under Biosample accession nos. SAMN12331629–SAMN12331670 and the 18S rRNA raw reads under Biosample accession nos. SAMN12332710–SAMN12332751. The 0.2–3-µm metagenomic assembly has been deposited with the NCBI under accession no. GCA_900411625.

Code availability

The R code used to create the heatmaps, ordinations and the 18S rRNA and WGCNA analyses are available on github (github.com/cnatalie/METZYME).

Received: 7 November 2019; Accepted: 13 October 2020;
Published online: 04 January 2021

References

1. Pennington, J. T. et al. Primary production in the eastern tropical Pacific: a review. *Prog. Oceanogr.* **69**, 285–317 (2006).
2. Moore, J. K., Doney, S. C., Glover, D. M. & Fung, I. Y. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep Sea Res. II* **49**, 463–507 (2002).
3. Landry, M. R. et al. Biological response to iron fertilization in the eastern equatorial Pacific (IronEx II). I. Microplankton community abundances and biomass. *Mar. Ecol. Prog. Ser.* **201**, 27–42 (2000).
4. Saito, M. A. et al. Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* **345**, 1173–1177 (2014).
5. Falkowski, P. G. et al. The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).
6. Stoecker, D. K., Hansen, P. J., Caron, D. A. & Mitra, A. Mixotrophy in the marine plankton. *Ann. Rev. Mar. Sci.* **9**, 311–335 (2017).
7. Flynn, K. J. et al. Misuse of the phytoplankton–zooplankton dichotomy: the need to assign organisms as mixotrophs within plankton functional types. *J. Plankton Res.* **35**, 3–11 (2013).
8. Beisner, B. E. et al. A guide to methods for estimating phago-mixotrophy in nanophytoplankton. *J. Plankton Res.* **41**, 77–89 (2019).
9. Caron, D. A., Countway, P. D., Jones, A. C., Kim, D. Y. & Schnetzer, A. Marine protistan diversity. *Ann. Rev. Mar. Sci.* **4**, 467–493 (2012).
10. Ward, B. A. & Follows, M. J. Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proc. Natl Acad. Sci. USA* **113**, 2958–2963 (2016).
11. Ward, B. A., Dutkiewicz, S., Barton, A. D. & Follows, M. J. Biophysical aspects of resource acquisition and competition in algal mixotrophs. *Am. Nat.* **178**, 98–112 (2011).
12. Edwards, K. F. Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proc. Natl Acad. Sci. USA* **116**, 6211–6220 (2019).
13. Ward, B. A. Mixotroph ecology: more than the sum of its parts. *Proc. Natl Acad. Sci. USA* **116**, 5846–5848 (2019).
14. Carradec, Q. et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **9**, 373 (2018).
15. Caputi, L. et al. Community-level responses to iron availability in open ocean planktonic ecosystems. *Global Biogeochem. Cycles* **33**, 391–419 (2019).
16. Malviya, S. et al. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl Acad. Sci. USA* **113**, E1516–E1525 (2016).
17. de Vargas, C. et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
18. Le Bescot, N. et al. Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environ. Microbiol.* **18**, 609–626 (2016).
19. Guidi, L. et al. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465–470 (2016).
20. Gorsky, G. et al. Expanding *Tara* oceans protocols for underway, ecosystemic sampling of the ocean-atmosphere interface during *Tara* Pacific expedition (2016–2018). *Front. Mar. Sci.* **6**, 750 (2019).
21. Wilken, S. et al. The need to account for cell biology in characterizing predatory mixotrophs in aquatic environments. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190090 (2019).
22. Edgcomb, V. P. Marine protist associations and environmental impacts across trophic levels in the twilight zone and below. *Curr. Opin. Microbiol.* **31**, 169–175 (2016).
23. Robinson, C. et al. Mesopelagic zone ecology and biogeochemistry: a synthesis. *Deep Sea Res. 2 Top. Stud. Oceanogr.* **57**, 1504–1518 (2010).
24. Pernice, M. C. et al. Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.* **10**, 945–958 (2016).
25. López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607 (2001).
26. Hu, S. K. et al. Shifting metabolic priorities among key protistan taxa within and below the euphotic zone. *Environ. Microbiol.* **20**, 2865–2879 (2018).
27. Jeong, H. J. et al. Mixotrophy in the phototrophic dinoflagellate *Takayama helix* (family Kareniaceae): predator of diverse toxic and harmful dinoflagellates. *Harmful Algae* **60**, 92–106 (2016).
28. Hansen, P. J. The role of photosynthesis and food uptake for the growth of marine mixotrophic dinoflagellates. *J. Eukaryot. Microbiol.* **58**, 203–214 (2011).
29. Adolf, J. E. et al. Species specificity and potential roles of *Karlodinium micrum* toxin. *Afr. J. Mar. Sci.* **28**, 415–419 (2006).
30. Glibert, P. M. et al. Grazing by *Karenia brevis* on *Synechococcus* enhances its growth rate and may help to sustain blooms. *Aquat. Microb. Ecol.* **55**, 17–30 (2009).
31. Kleiner, M. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* **8**, 1558 (2017).
32. Chavez, F. P., Buck, K. R. & Barber, R. T. Phytoplankton taxa in relation to primary production in the equatorial Pacific. *Deep Sea Res. A.* **37**, 1733–1752 (1990).
33. Goericke, R. & Repeta, D. The pigments of *Prochlorococcus marinus*: the presence of divinyl chlorophyll *a* and *b* in a marine procaryote. *Limnol. Oceanogr.* **37**, 425–433 (1992).
34. Irigoien, X., Meyer, B., Harris, R. & Harbour, D. Using HPLC pigment analysis to investigate phytoplankton taxonomy: the importance of knowing your species. *Helgol. Mar. Res.* **58**, 77–82 (2004).
35. Binder, B. J., Chisholm, S. W., Olson, R. J., Frankel, S. L. & Worden, A. Z. Dynamics of picophytoplankton, ultraphytoplankton and bacteria in the central equatorial Pacific. *Deep Sea Res. 2 Top. Stud. Oceanogr.* **43**, 907–931 (1996).
36. de Baar, H. J. W. et al. Synthesis of iron fertilization experiments: from the iron age in the age of enlightenment. *J. Geophys. Res. Oceans* **110**, C09S16 (2005).
37. Bodył, A. & Moszczyński, K. Did the peridinin plastid evolve through tertiary endosymbiosis? A hypothesis. *Eur. J. Phycol.* **41**, 435–448 (2006).
38. Ishida, K.-I. & Green, B. R. Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc. Natl Acad. Sci. USA* **99**, 9294–9299 (2002).
39. De salas, M. F. et al. *Takayama* gen. nov. (Gymnodiniales, Dinophyceae), a new genus of unarmored dinoflagellates with sigmoid apical grooves, including the description of two new species. *J. Phycol.* **39**, 1233–1246 (2003).
40. Yoon, H. S., Hackett, J. D. & Bhattacharya, D. A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc. Natl Acad. Sci. USA* **99**, 11724–11729 (2002).
41. Chavez, F. P., Buck, K. R., Service, S. K., Newton, J. & Barber, R. T. Phytoplankton variability in the central and eastern tropical Pacific. *Deep Sea Res. 2 Top. Stud. Oceanogr.* **43**, 835–870 (1996).
42. Jeong, H. J. et al. Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Sci. J.* **45**, 65–91 (2010).
43. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
44. Labarre, A., Obiol, A., Wilken, S., Forn, I. & Massana, R. Expression of genes involved in phagocytosis in uncultured heterotrophic flagellates. *Limnol. Oceanogr.* **65**, S149–S160 (2020).
45. Burns, J. A., Pittis, A. A. & Kim, E. Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. *Nat. Ecol. Evol.* **2**, 697–704 (2018).
46. Zhang, Y. Metatranscriptomic signatures associated with phytoplankton regime shift from diatom dominance to a dinoflagellate bloom. *Front. Microbiol.* **10**, 590 (2019).
47. Liu, Z., Campbell, V., Heidelberg, K. B. & Caron, D. A. Gene expression characterizes different nutritional strategies among three mixotrophic protists. *FEMS Microbiol. Ecol.* **92**, fiw106 (2016).
48. Yu, L. et al. Comparative metatranscriptomic profiling and microRNA sequencing to reveal active metabolic pathways associated with a dinoflagellate bloom. *Sci. Total Environ.* **699**, 134323 (2020).
49. Zhuang, Y., Zhang, H., Hannick, L. & Lin, S. Metatranscriptome profiling reveals versatile N-nutrient utilization, CO_2 limitation, oxidative stress, and active toxin production in an *Alexandrium fundyense* bloom. *Harmful Algae* **42**, 60–70 (2015).
50. Yutin, N., Wolf, M. Y., Wolf, Y. I. & Koonin, E. V. The origins of phagocytosis and eukaryogenesis. *Biol. Direct* **4**, 9 (2009).
51. Perret, E., Davoust, J., Albert, M., Besseau, L. & Soyer-Gobillard, M. O. Microtubule organization during the cell cycle of the primitive eukaryote dinoflagellate *Crypthecodinium cohnii*. *J. Cell Sci.* **104**, 639–651 (1993).

52. Brown, D. L., Cachon, J., Cachon, M. & Boillot, A. The cytoskeletal microtubular system of some naked dinoflagellates. *Cell Motil.* **9**, 361–374 (1988).
53. Gagnon, C. et al. The polyglutamylated lateral chain of alpha-tubulin plays a key role in flagellar motility. *J. Cell Sci.* **109**, 1545–1553 (1996).
54. Okamoto, N. & Keeling, P. A comparative overview of the flagellar apparatus of dinoflagellate, perkinsids and colpodellids. *Microorganisms* **2**, 73–91 (2014).
55. McKie-Krisberg, Z. M., Sanders, R. W. & Gast, R. J. Evaluation of mixotrophy-associated gene expression in two species of polar marine algae. *Front. Mar. Sci.* **5**, 273 (2018).
56. Rubin, E. T., Cheng, S., Montalbano, A. L., Menden-Deuer, S. & Rynearson, T. A. Transcriptomic response to feeding and starvation in a herbivorous dinoflagellate. *Front. Mar. Sci.* **6**, 246 (2019).
57. Lie, A. A. Y. et al. Effect of light and prey availability on gene expression of the mixotrophic chrysophyte, *Ochromonas* sp. *BMC Genomics* **18**, 163 (2017).
58. Massana, R. et al. Gene expression during bacterivorous growth of a widespread marine heterotrophic flagellate. *ISME J.* https://doi.org/10.1038/s41396-020-00770-4 (2020).
59. Santoferrara, L. F., Guida, S., Zhang, H. & McManus, G. B. De novo transcriptomes of a mixotrophic and a heterotrophic ciliate from marine plankton. *PLoS ONE* **9**, e101418 (2014).
60. Bouché, N., Yellin, A., Snedden, W. A. & Fromm, H. Plant-specific calmodulin-binding proteins. *Annu. Rev. Plant Biol.* **56**, 435–466 (2005).
61. Crivici, A. & Ikura, M. Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 85–116 (1995).
62. Becker, K. et al. Quantifying post-transcriptional regulation in the development of *Drosophila melanogaster*. *Nat. Commun.* **9**, 4970 (2018).
63. Slamovits, C., Okamoto, N., Burri, L. James, E. R. & Keeling, P. J. A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat. Commun.* **2**, 183 (2011).
64. Lin, S. Genomic understanding of dinoflagellates. *Res. Microbiol.* **162**, 551–569 (2011).
65. Olson, D. K., Yoshizawa, S., Boeuf, D., Iwasaki, W. & Delong, E. F. Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. *ISME J.* **12**, 1047–1060 (2018).
66. Guo, Z., Zhang, H., Liu, S. & Lin, S. Biology of the marine heterotrophic dinoflagellate *Oxyrrhis marina*: current status and future directions. *Microorganisms* **1**, 33–57 (2013).
67. Guo, Z., Zhang, H. & Lin, S. Light-promoted rhodopsin expression and starvation survival in the marine dinoflagellate *Oxyrrhis marina*. *PLoS ONE* **9**, e114941 (2014).
68. Taylor, A. G., Landry, M. R., Selph, K. E. & Yang, E. J. Biomass, size structure and depth distributions of the microbial community in the eastern equatorial Pacific. *Deep Sea Res. 2 Top. Stud. Oceanogr.* **58**, 342–357 (2011).
69. Takahashi, M., Satake, K.-I. & Nakamoto, N. Chlorophyll distribution and photosynthetic activity in the north and equatorial Pacific Ocean along 155°W. *J. Oceanogr. Soc. Japan* **28**, 27–36 (1972).
70. Ducklow, H. W. The bacterial component of the oceanic euphotic zone. *FEMS Microbiol. Ecol.* **30**, 1–10 (1999).
71. Behrmann, G. & Hardeland, R. Ultrastructural characterization of asexual cysts of *Gonyaulax polyedra* Stein (Dinoflagellata). *Protoplasma* **185**, 22–27 (1995).
72. Roy, S., Letourneau, L. & Morse, D. Cold-induced cysts of the photosynthetic dinoflagellate *Lingulodinium polyedrum* have an arrested circadian bioluminescence rhythm and lower levels of protein phosphorylation. *Plant Physiol.* **164**, 966–977 (2014).
73. Bravo, I. & Figueroa, R. I. Towards an ecological understanding of dinoflagellate cyst functions. *Microorganisms* **2**, 11–32 (2014).
74. Gotthardt, D. et al. High-resolution dissection of phagosome maturation reveals distinct membrane trafficking phases. *Mol. Biol. Cell* **13**, 3508–3520 (2002).
75. Keeling, P. J. et al. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).
76. Garcia, H. E. et al. *World Ocean Atlas 2013. Volume 4, Dissolved Inorganic Nutrients (Phosphate, Nitrate, Silicate)* NOAA Atlas NESDIS Series 76 (NOAA, 2013); https://www.nodc.noaa.gov/OC5/woa13/pubwoa13.html
77. Alexander, H., Jenkins, B. D., Rynearson, T. A. & Dyhrman, S. T. Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc. Natl Acad. Sci. USA* **112**, E2182–E2190 (2015).
78. Bender, S. J., Parker, M. S. & Armbrust, E. V. Coupled effects of light and nitrogen source on the urea cycle and nitrogen metabolism over a diel cycle in the marine diatom *Thalassiosira pseudonana*. *Protist* **163**, 232–251 (2012).
79. Groussman, R. D., Parker, M. S. & Armbrust, E. V. Diversity and evolutionary history of iron metabolism genes in diatoms. *PLoS ONE* **10**, e0129081 (2015).
80. Marchetti, A. et al. Development of a molecular-based index for assessing iron status in bloom-forming pennate diatoms. *J. Phycol.* **53**, 820–832 (2017).
81. Chappell, P. D. et al. Genetic indicators of iron limitation in wild populations of *Thalassiosira oceanica* from the northeast Pacific Ocean. *ISME J.* **9**, 592–602 (2015).
82. McQuaid, J. B. et al. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature* **555**, 534–537 (2018).
83. Morrissey, J. et al. A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Curr. Biol.* **25**, 364–371 (2015).
84. Allen, A. E. et al. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl Acad. Sci. USA* **105**, 10438–10443 (2008).
85. Erdner, D. L. & Anderson, D. M. Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. *Limnol. Oceanogr.* **44**, 1609–1615 (1999).
86. La Roche, J., Boyd, P. W., McKay, R. M. L. & Geider, R. J. Flavodoxin as an in situ marker for iron stress in phytoplankton. *Nature* **382**, 802–805 (1996).
87. Peers, G. & Price, N. M. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **441**, 341–344 (2006).
88. Morey, J. S. et al. Transcriptomic response of the red tide dinoflagellate, *Karenia brevis*, to nitrogen and phosphorus depletion and addition. *BMC Genomics* **12**, 346 (2011).
89. Jing, X., Lin, S., Zhang, H., Koerting, C. & Yu, Z. Utilization of urea and expression profiles of related genes in the dinoflagellate *Prorocentrum donghaiense*. *PLoS ONE* **12**, e0187837 (2017).
90. Fan, C., Glibert, P. M., Alexander, J. & Lomas, M. W. Characterization of urease activity in three marine phytoplankton species, *Aureococcus anophagefferens*, *Prorocentrum minimum*, and *Thalassiosira weissflogii*. *Mar. Biol.* **142**, 949–958 (2003).
91. Shilova, I. N. et al. Differential effects of nitrate, ammonium, and urea as N sources for microbial communities in the North Pacific Ocean. *Limnol. Oceanogr.* **62**, 2550–2574 (2017).
92. Casey, J. R., Lomas, M. W., Mandecki, J. & Walker, D. E. *Prochlorococcus* contributes to new production in the Sargasso Sea deep chlorophyll maximum. *Geophys. Res. Lett.* **34**, L10604 (2007).
93. Price, N. M. & Morel, F. M. M. Cadmium and cobalt substitution for zinc in a marine diatom. *Nature* **344**, 658–660 (1990).
94. McGinn, P. J. & Morel, F. M. M. Expression and regulation of carbonic anhydrases in the marine diatom *Thalassiosira pseudonana* and in natural phytoplankton assemblages from Great Bay, New Jersey. *Physiol. Plant.* **133**, 78–91 (2008).
95. Marchetti, A. et al. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl Acad. Sci. USA* **109**, E317–E325 (2012).
96. Bertrand, E. M. et al. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl Acad. Sci. USA* **112**, 9938–9943 (2015).
97. Bender, S. J., Durkin, C. A., Berthiaume, C. T., Morales, R. L. & Armbrust, E. V. Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front. Mar. Sci.* **1**, 3 (2014).
98. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. VERTEX: carbon cycling in the northeast Pacific. *Deep Sea Res. A* **34**, 267–285 (1987).
99. Gloege, L., McKinley, G. A., Mouw, C. B. & Ciochetto, A. B. Global evaluation of particulate organic carbon flux parameterizations and implications for atmospheric pCO_2. *Global Biogeochem. Cycles* **31**, 1192–1215 (2017).
100. Smayda, T. J. Adaptations and selection of harmful and other dinoflagellate species in upwelling systems. 2. Motility and migratory behaviour. *Prog. Oceanogr.* **85**, 71–91 (2010).
101. Raven, J. A. & Richardson, K. Dinophyte flagella: a cost–benefit analysis. *New. Phytol.* **98**, 259–276 (1984).
102. Hou, Y. & Lin, S. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. *PLoS ONE* **4**, e6978 (2009).
103. Lin, S. The smallest dinoflagellate genome is yet to be found: a comment on LaJeunesse et al. '*Symbiodinium* (Pyrrophyta) genome sizes (DNA content) are smallest among dinoflagellates'. *J. Phycol.* **42**, 746–748 (2006).
104. Fuhrman, J. Genome sequences from the sea. *Nature* **424**, 1001–1002 (2003).
105. Rocap, G. et al. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047 (2003).
106. Saito, M. A. et al. Needles in the blue sea: sub-species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome. *Proteomics* **15**, 3521–3531 (2015).
107. Cutter, G. et al. Sampling and sample-handling protocols for GEOTRACES cruises. *EPIC Eprint* https://epic.awi.de/id/eprint/34484/ (2010).

108. Anderson, R. F. & Henderson, G. M. GEOTRACES: a global study of the marine biogeochemical cycles of trace elements and their isotopes. *Oceanography* **18**, 76–79 (2005).
109. Saito, M. A. & Schneider, D. L. Examination of precipitation chemistry and improvements in precision using the $Mg(OH)_2$ preconcentration inductively coupled plasma mass spectrometry (ICP-MS) method for high-throughput analysis of open-ocean Fe and Mn in seawater. *Anal. Chim. Acta* **565**, 222–233 (2006).
110. Munson, K. M., Lamborg, C. H., Swarr, G. J. & Saito, M. A. Mercury species concentrations and fluxes in the Central Tropical Pacific Ocean. *Global Biogeochem. Cycles* **29**, 656–676 (2015).
111. Lu, X. & Zhu, H. Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Mol. Cell. Proteomics* **4**, 1948–1958 (2005).
112. Vizcaíno, J. A. et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013).
113. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82**, 2272–2281 (2010).
114. Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012).
115. Rho, M., Tang, H. & Ye, Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191 (2010).
116. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
117. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
118. Kolody, B. C. et al. Diel transcriptional response of a California Current plankton microbiome to light, low iron, and enduring viral infection. *ISME J.* **13**, 2817–2833 (2019).
119. Ogata, H. & et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
120. Nordberg, H. et al. The genome portal of the Department of Energy joint genome institute: 2014 updates. *Nucleic Acids Res.* **42**, D26–D31 (2014).
121. Hancock, J. M., Zvelebil, M. J., Hancock, J. M. & Bishop, M. J. in *Dictionary of Bioinformatics and Computational Biology* (eds Hancock, J. M. & Zvelebil, M. J.) (Wiley, 2004).
122. Finn, R. D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
123. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).
124. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2010).
125. Lane, D. J. et al. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl Acad. Sci. USA* **82**, 6955–6959 (1985).
126. Herlemann, D. P. et al. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**, 1571–1579 (2011).
127. Stoeck, T. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
128. Hugerth, L. W. et al. Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS ONE* **9**, e95567 (2014).
129. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
130. Guillou, L. et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604 (2013).
131. Mordret, S. et al. dinoref: a curated dinoflagellate (Dinophyceae) reference database for the 18S rRNA gene. *Mol. Ecol. Resour.* **18**, 974–987 (2018).
132. Decelle, J. et al. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Mol. Ecol. Resour.* **15**, 1435–1445 (2015).
133. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
134. Chen, I.-M. A. et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2018).
135. Dupont, C. L. et al. Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J.* **9**, 1076–1092 (2015).
136. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
137. Hothorn, T., Hornik, K., de Wiel, M. & Zeileis, A. coin: Conditional inference procedures in a permutation test framework. R package version 0.6.6 https://rdrr.io/cran/coin/ (2006).
138. McMurdie, P. J. & Holmes, S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
139. Oksanen, J. et al. vegan: Community Ecology Package. R package version 2.3-0 https://cran.r-project.org/web/packages/vegan/index.html (2015).
140. Johnson, L. K., Alexander, H. & Brown, C. T. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. *Gigascience* **8**, giy158 (2019).
141. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
142. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2: a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
143. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
144. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**, 538 (2010).
145. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
146. Brown, M. Ocean Data View 4.0. *Oceanography* **11**, 19–21 (1998).
147. Garcia, H. E. et al. *World Ocean Atlas 2009, Volume 4: Nutrients (Phosphate, Nitrate, and Silicate)* (ed. Levitus, S.) (US Government Printing Office, 2010).

Acknowledgements

We thank the captain, crew and scientific party of the October 2011 R/V Kilo Moana cruise. This research was funded through National Science Foundation (NSF) grant nos. OCE-1031271, 1924554, 1850719 and 1736599, Gordon and Betty Moore Foundation grant nos. 2724 and 3782, the Center for Microbial Oceanography Research and Education and the Woods Hole Oceanographic Institution Ocean Life Institute to M.A.S. N.R.C. was supported by grant no. 544236 from the Simons Foundation. A.E.A. acknowledges support from the NSF-OCE-1756884 grant and a Gordon and Betty Moore Foundation grant no. GBMF3828. G.R.D. acknowledges support from NSF grant no. OCE-1061876. N.A.H. was supported by NSF Graduate Research Fellowship no. 1122274, and J.K.S. was supported by a NASA Postdoctoral Program Fellowship. We acknowledge T. Geopfert (Arizona State University) and D. Wang (ExxonMobil) for sampling assistance, and thank J. Jennings (Oregon State University) for processing the macronutrient samples and S. Davies (Boston University) for providing WGCNA code. J. Bowman's (Scripps Institution of Oceanography) pplacer blog provided invaluable support during phylogenetic tree construction. We thank M. Johnson (Woods Hole Oceanographic Institution (WHOI)), S. Hu (WHOI), A. Frank (WHOI), J.P. Balmonte (Uppsala), Lisa Nigro (University of Connecticut), C. Moreno (University of North Carolina-Chapel Hill (UNC)) and S. H. Jang (UNC) for helpful feedback and suggestions on the manuscript.

Author contributions

N.R.C. analysed the data and wrote the first draft of the manuscript. D.M.M., M.R.M., N.A.H., M.A.S. and C.L. collected multi-omics samples. M.R.M. analysed the proteomic samples by mass spectrometry. D.M.M. processed, extracted and prepared the protein samples for the proteomic analysis. N.A.H. and J.K.S. contributed to the development of the metaproteomic pipeline. N.J.H. quantified the dissolved cobalt concentrations. M.B. contributed to dinoflagellate physiology interpretations. G.R.D. collected and generated the pigment data. C.L. deployed the sediment traps, calculated carbon flux measurements and organized the expedition as co-chief scientist. C.L.D. processed and annotated the metagenomics data. A.E.A. and J.P.M. contributed the 18S and 16S rRNA metabarcoding and metatranscriptomic data. M.A.S. designed the study, organized the expedition as co-chief scientist and guided the interpretations. All authors contributed to the final version of the manuscript.

Competing interests

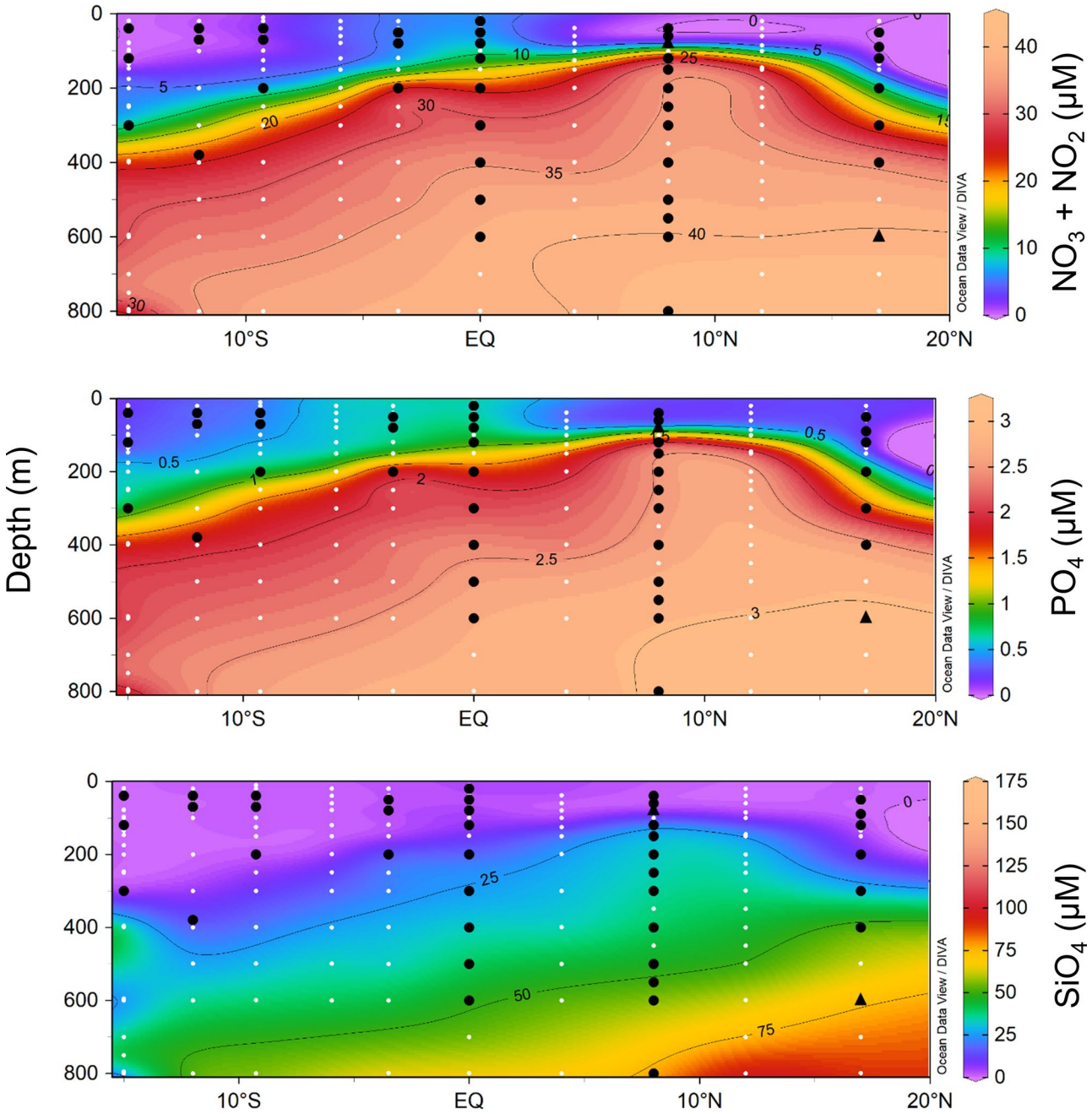
The authors declare no competing interests.

Additional information

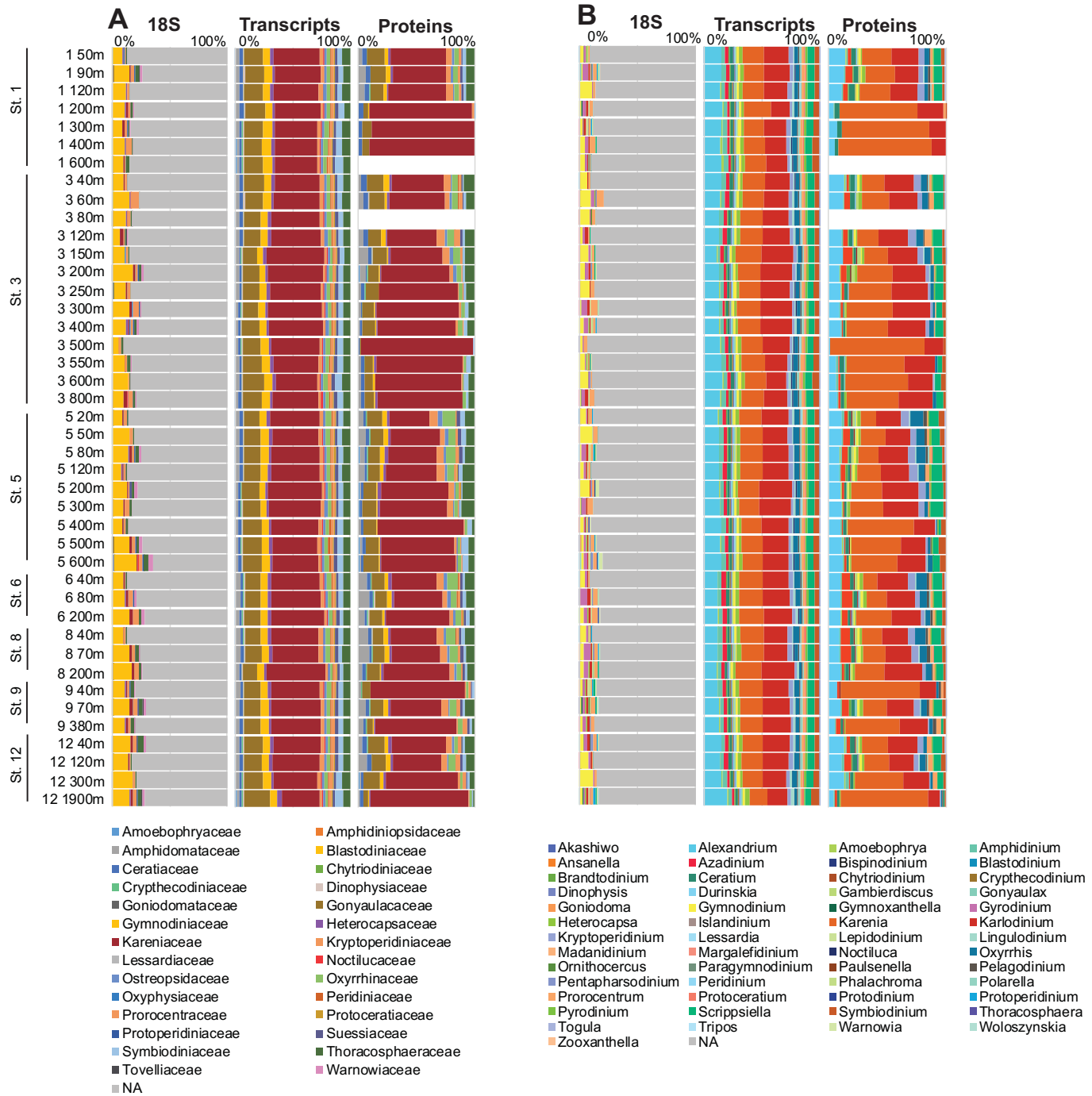
Extended data is available for this paper at https://doi.org/10.1038/s41564-020-00814-7.

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-00814-7.

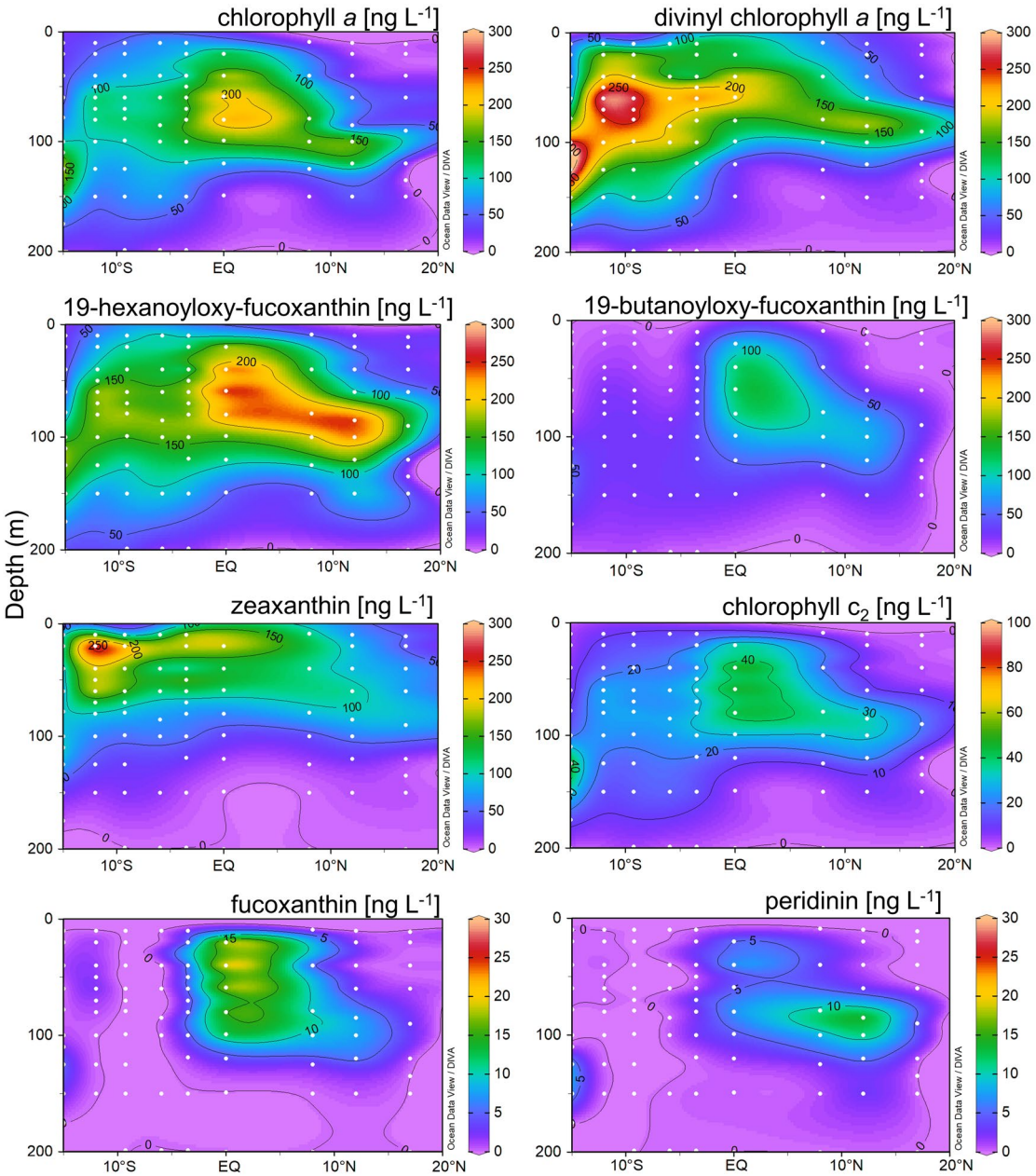
Correspondence and requests for materials should be addressed to N.R.C., A.E.A. or M.A.S.



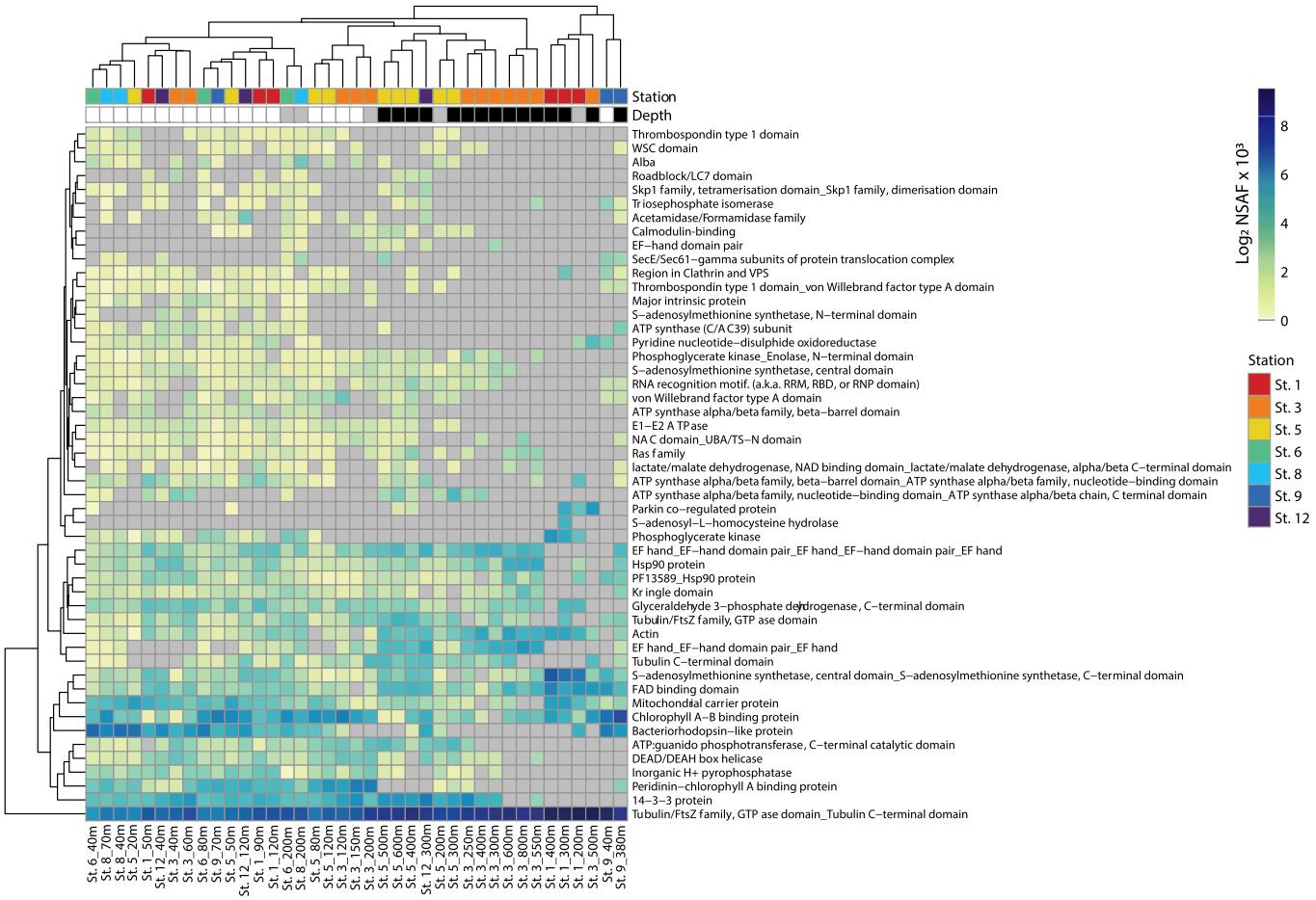
Extended Data Fig. 1 | Nitrate + nitrite, phosphate and silicate concentrations along the transect. White dots represent CTD sampling depths where physiochemical data, pigments, macronutrients and trace metals were collected. Black dots indicate locations where filters were processed for metaproteomic, metatranscriptomic and 18s rRNA analyses. Two black triangles represent depths at which only metatranscriptomic and 18S rRNA information is available (St. 1, 600 m; St. 3, 80 m).



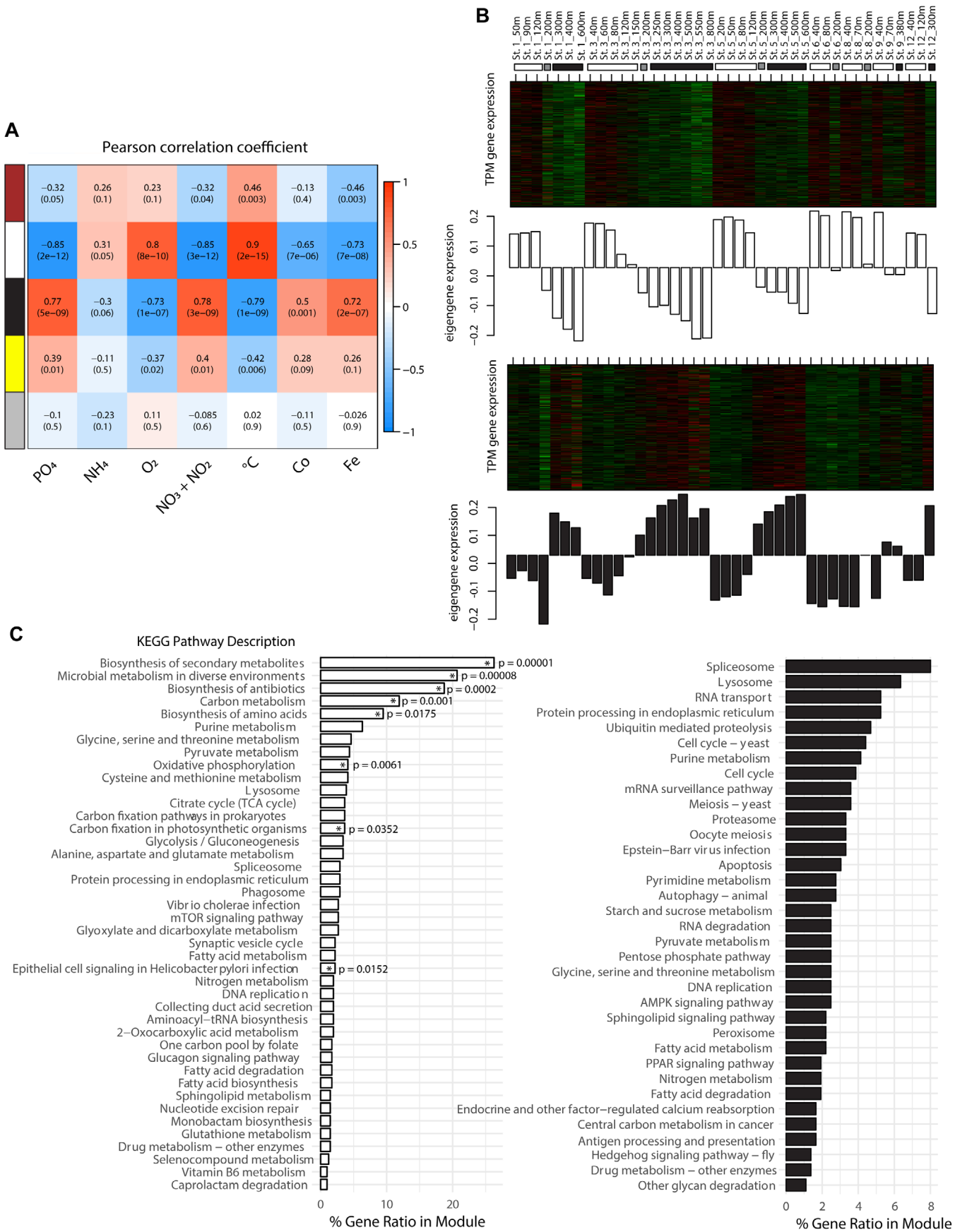
Extended Data Fig. 2 | Dinoflagellate taxonomic annotations across the transect. a, Dinoflagellate family-level and (**b**) genus-level relative community composition determined through 18S rRNA, metatranscriptomic and metaproteomic analyses from 3-51 µm filter fractions, highlighting the abundance of the *Kareniaceae*-like family and *Karlodinium* and *Karenia*-like genera in transcripts and proteins. Taxonomic annotations were assigned based on assembled metatranscriptome matches to the PhyloDB database, containing marine protistan[21] and bacterial transcriptomes and genomes. 18S rRNA annotations were assigned using the PR[2,22] database. 'NA' represents dinoflagellate OTUs without family or genera-level taxonomy available in the PR[2] database. In the transcript pool, *Karenia* and *Karlodinium* genera together comprised an average 42 ± 0.02% of the total dinoflagellate reads in the euphotic zone (< 200m) and 42 ± 0.04% in the mesopelagic (≥ 200m). In the protein pool, these two genera comprised an average 45 ± 10% of the spectral counts in the euphotic zone and 71 ± 12% in the mesopelagic.



Extended Data Fig. 3 | Pigment profiles along the METZYME transect determined by high performance liquid chromatography (HPLC). White dots represent sampling depths. Graphed in Ocean Data View (ODV) using DIVA interpolation.

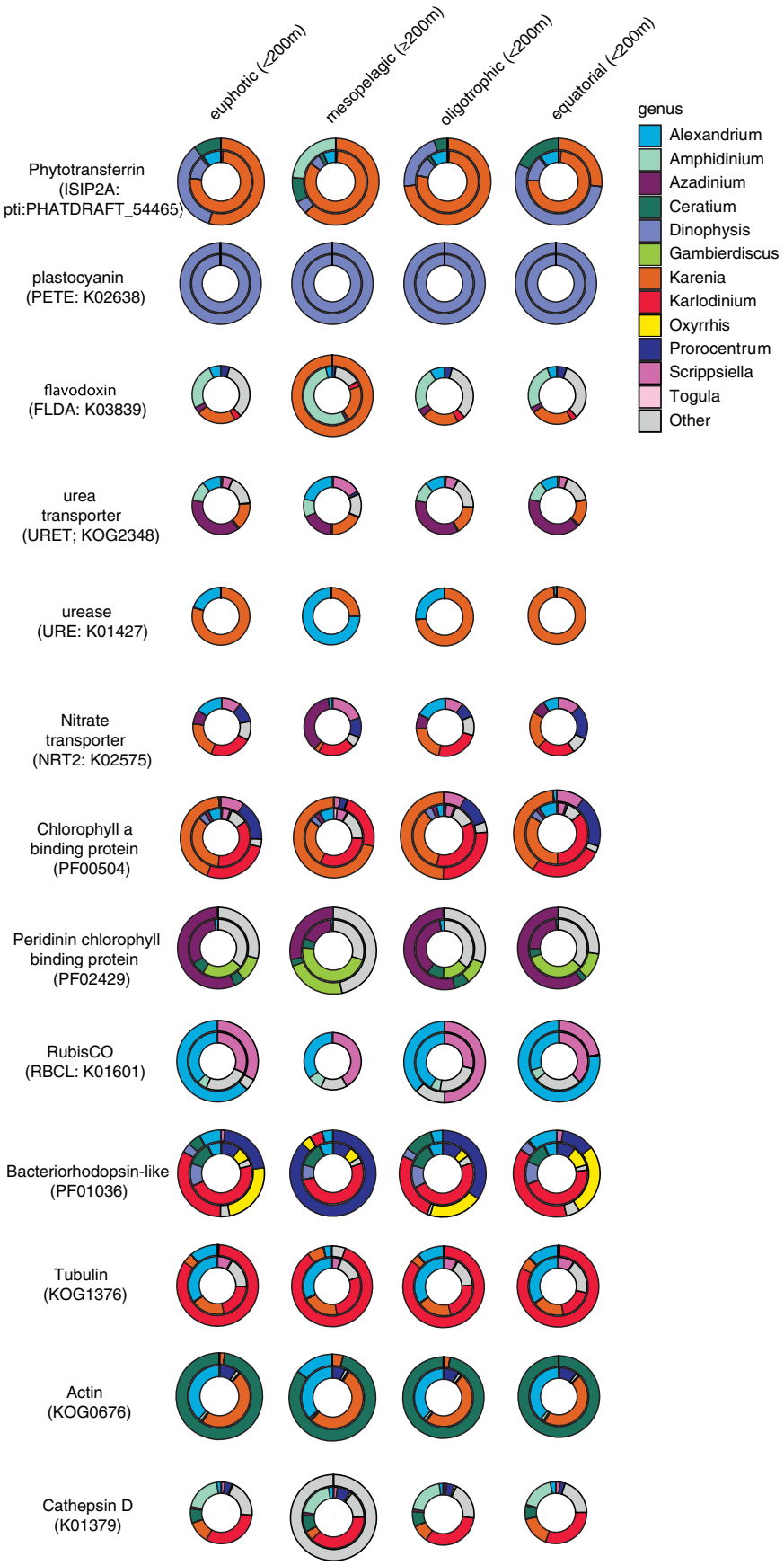


Extended Data Fig. 4 | Log_2 NSAF-normalized dinoflagellates spectral count heatmap displaying relative protein abundance at the PFam annotation level. The top 50 PFam-annotated genes with highest deviations from the mean (variances) across samples are shown. The depth annotation bar highlights samples from the surface (<200 m, white), deep (>200 m, black) and 200 m (gray). Dendrogram shows similarity in spectral abundance among samples based on Euclidean distance and hierarchical clustering. Each row represents a PFam annotation, with spectral counts associated with identical PFams summed together. Multiple PFam annotations of the same contig are separated by an underscore.



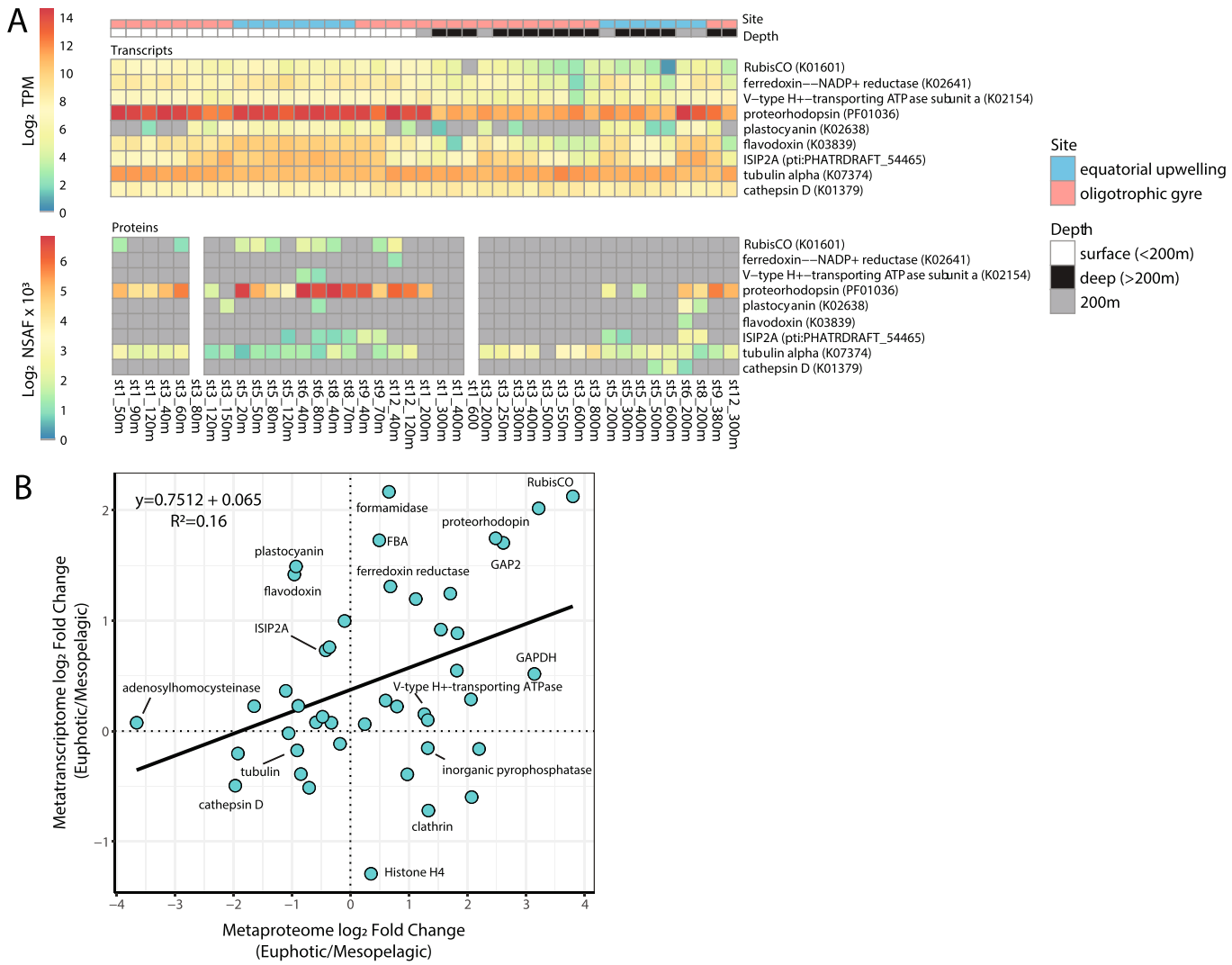
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Weighted correlation network analysis (WGCNA) eigengene modules using log_2 TPM-normalized KEGG-annotated dinoflagellate transcripts. The 'module eigengene' represents the first principal coordinate of the module and summarizes the module gene expression profile. Signed network analysis was performed using the WCGNA package in R[23] with at least 75 genes per eigengene module, and modules merged displaying similar eigengene values across samples (MEDissThres= 0.3). **a**, Color scale bar represents Pearson correlation coefficients between environmental metadata and eigengene modules; correlation coefficients and two-tailed Student test unadjusted (default) *p* values are shown in each box. **b**, Eigenene expression values plotted alongside log_2 TPM gene expression in the white (surface; top) and black (deep; bottom) modules. **c**, KEGG pathway identity of white and black WCGNA modules. A KEGG enrichment analysis was performed using clusterProfiler's *enrich* function, which calculates overrepresentation of KEGG pathways compared to the total genes identified in the data set using a two-tailed hypergeometric test (Benjamin-Hochberg adjusted p-value < 0.05)[24]. Significantly enriched pathways are denoted with an asterisk (*).

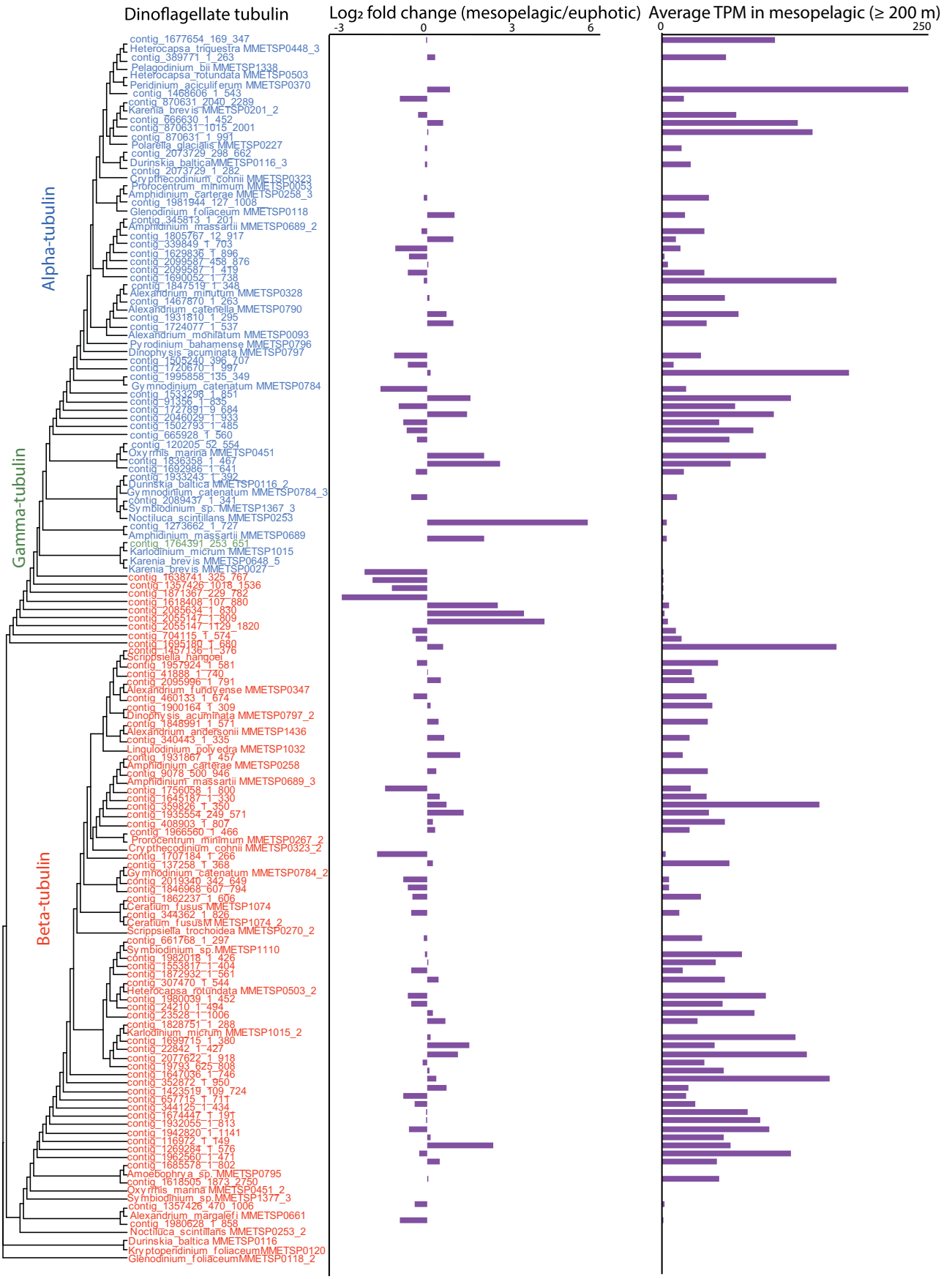


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Stacked pie charts depicting dinoflagellate genus-level relative community composition for genes of interest shown to be responsive to shallow (< 200m), deep (≥ 200m), oligotrophic (St. 1, 3, 9, 12) and equatorial upwelling (St. 5, 6, 8) environments. The inner rings show genera composition based on normalized transcript read counts, and the outer rings show normalized protein spectral counts. Transcript and protein were averaged across samples within each of the four environments. Only shallow depths (< 200m) were included for the oligotrophic and equatorial pie charts. Abbreviations used are shown for each gene of interest, along with their IDs from KEGG, KOG or PFam databases. 'Other' represents other dinoflagellate genera in minor relative abundance and not included here (see Extended Data Fig. 2 for dinoflagellate genus-level relative abundance across samples).

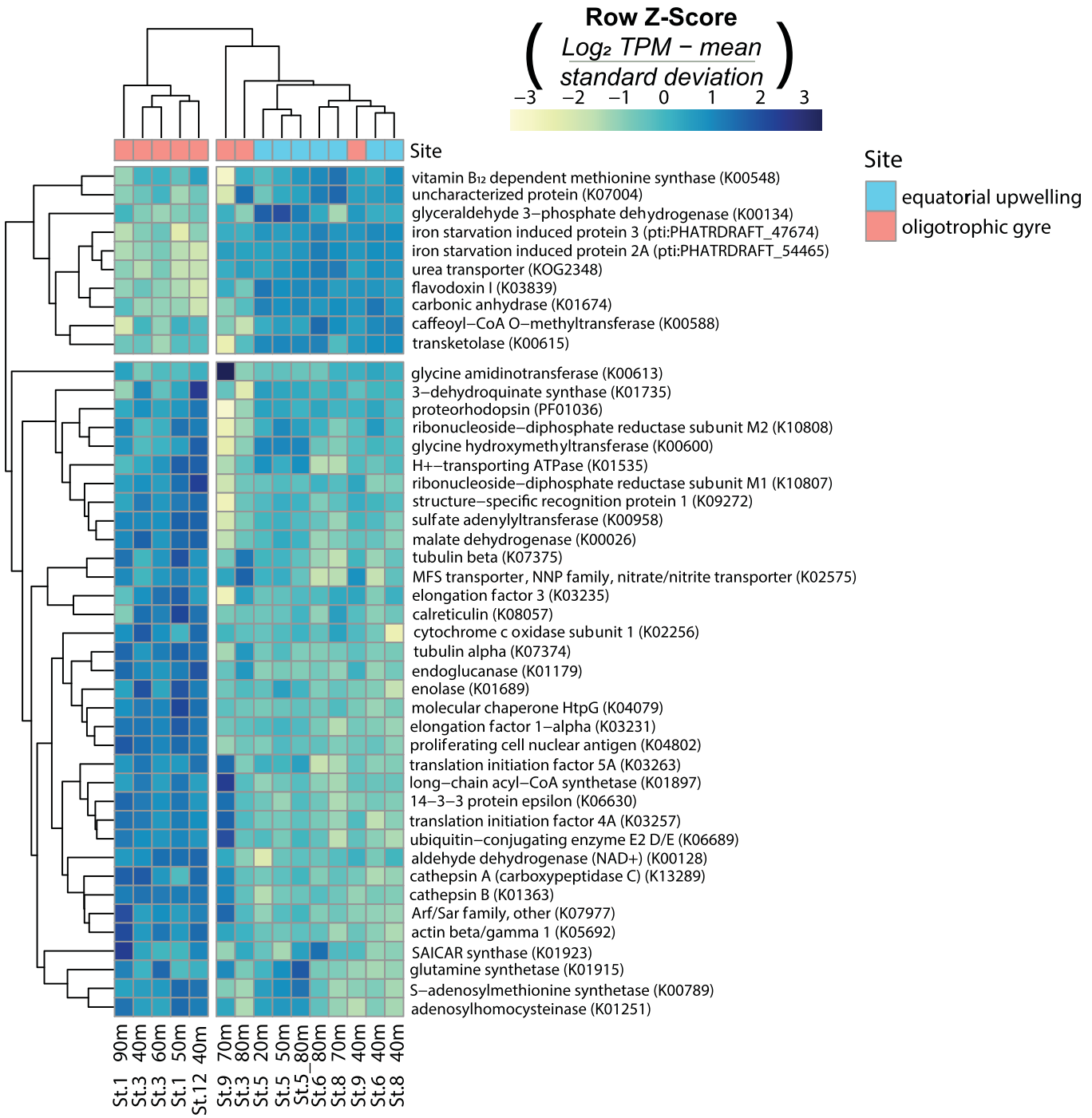


Extended Data Fig. 7 | Relative transcript and protein abundance for dinoflagellate genes of interest between depths and across latitudes. a, Relative gene expression is shown as $\log_2$ transcripts per million (TPM; top), relative protein abundance as $\log_2$ normalized spectral abundance factor (NSAF; bottom). Undetected transcripts and proteins are indicated in gray. The depth annotation bar indicates samples from the surface (<200 m, white), deep (>200 m, black) and 200 m (gray), and the site annotation bar shows whether samples were collected from the oligotrophic gyres (St. 1, 3, 9, 12) or the equatorial upwelling zone (St. 5, 6, 8). **b**, Comparison of average transcript and average protein abundance fold changes between euphotic and mesopelagic zones. Only shared KEGG genes (KOs) are shown that were detected in at least one metatranscriptome and one metaproteome, with proteorhodopin (PFam PF01036) and ISIP2a (KEGG gene pti:PHATDRAFT_54465) manually added. Values of zero were changed to a small value (0.1) to allow for fold changes estimates. The black line illustrates the linear relationship between protein and transcript fold changes. Genes in the top right (quadrant 1) represent transcripts/proteins abundant in euphotic waters, genes in the bottom left (quadrant 3) represent transcripts/proteins abundant in the mesopelagic (FBA = fructose biphosphate aldolase; GAP2 = glyceraldehyde-3-phosphate dehydrogenase (NAD(P)); ISIP2A = iron starvation induced protein 2 [phytotransferrin]; GAPDH = glyceraldehyde 3-phosphate dehydrogenase).

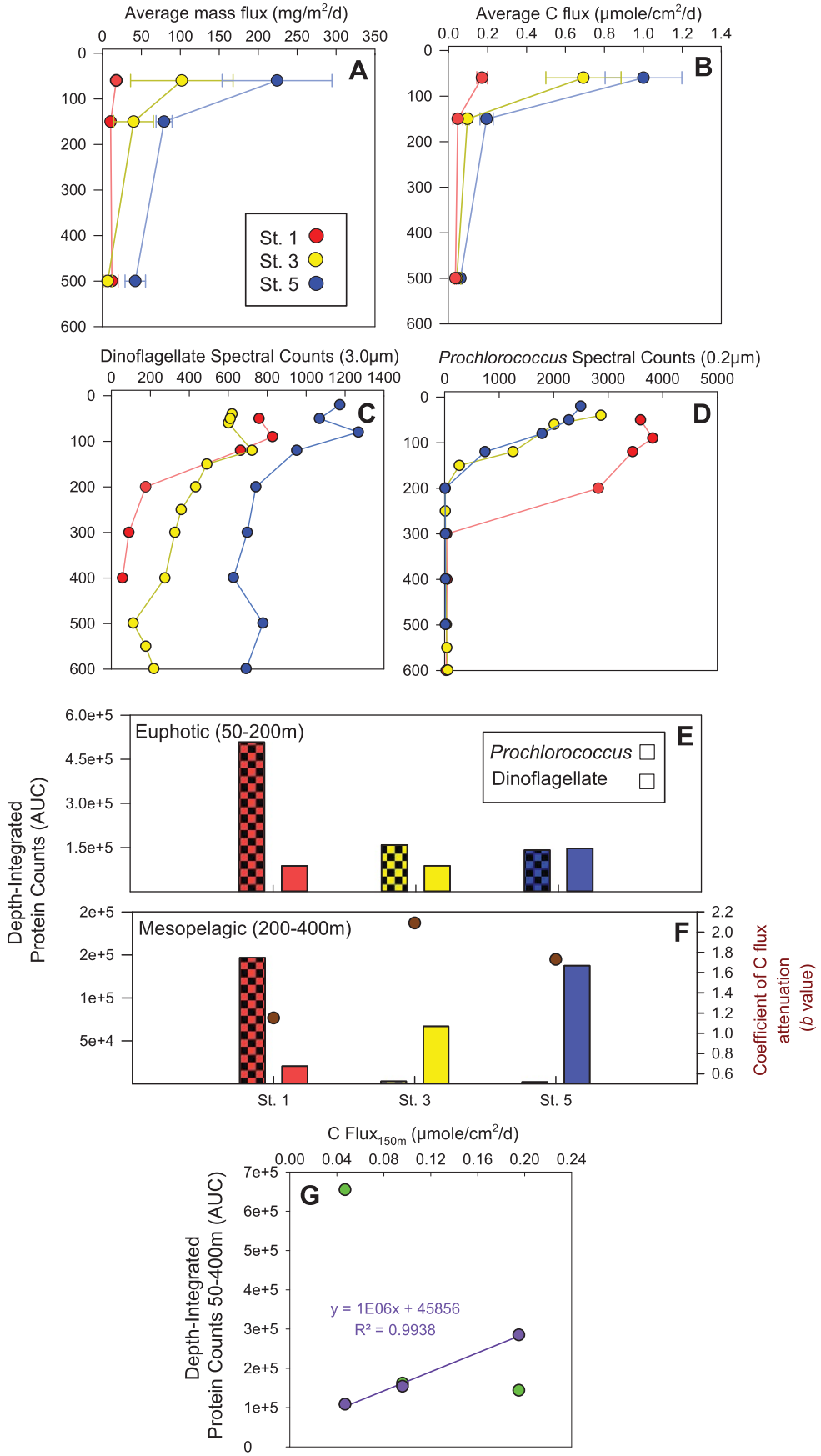


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Cladogram of translated dinoflagellate tubulin contigs with reference MMETSP dinoflagellate proteins (left) shown alongside tubulin gene expression in the mesopelagic compared to euphotic zone (log_2 average fold change) (middle) and average expression levels (TPM-normalized transcript abundance) in the mesopelagic (right). Alpha-tubulin is shown in blue, Gamma-tubulin in green, and beta-tubulin in orange. Reference sequences were aligned using MUSCLE v3.8 and the maximum-likelihood tree was created using RAxML v8.2.11 with the PROTGAMMALG model and 100 bootstrap replicates. Contigs were placed on the reference tree using pplacer v1.1alpha19. The cladogram is visualized with the R package ggtree v1.16.0.



Extended Data Fig. 9 | Heatmap displaying relative dinoflagellate TPM-normalized gene expression from each station < 100m, displayed as row Z-score [(log2 TPM - mean)/standard deviation]. The top 45 KEGG annotated genes with highest transcript deviations from the mean (variances) are displayed. Along with the KEGG entries, five annotations not included in KEGG were manually added. Dendrogram shows similarity in transcript abundances based on Euclidean distance and hierarchical clustering, created with pheatmap v1.0.12. Each row represents a unique KEGG-annotated gene. The site annotation bar indicates whether samples were collected from the oligotrophic gyres (St. 1, 3, 9, 12) or the equatorial upwelling zone (St. 5, 6, 8). Color gradients represents low (yellow) to high (blue) gene expression.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Positive relationship between dinoflagellate protein abundance and carbon flux. Particulate export estimates through average mass flux (**a**) and particulate carbon flux (**b**) from the oligotrophic (St. 1, red) to equatorial upwelling region (St. 5, blue), as visualized by latitude in Fig. 5 (n = 3 tubes per depth from the same sediment trap array; error bars represent standard deviation.). **c**, Vertical profiles of absolute dinoflagellate exclusive protein spectral counts derived from the 3-51 µm size fraction and (**d**) absolute *Prochlorococcus* exclusive spectral counts from the 0.2-3 µm fraction along the surface gradient. Absolute spectral counts were not NSAF-normalized as performed in the functional analysis. Depth-integrated spectral counts from (**e**) 50-200m and (**f**) 200-400m highlight changes to protein inventory across the biogeochemical gradient and between the euphotic and mesopelagic zones, with the coefficient of carbon flux attenuation (*b* value) shown in brown (three depths per station were used to calculate the slope (*b*) via non-linear flux curve fitting to the Martin power law). Depth-integrated values were obtained by calculating the area under the profile for spectral counts versus depth. Depth-integrated dinoflagellate spectral counts show a positive relationship with carbon flux to 150 m (**g**) in contrast to *Prochlorococcus* which demonstrates a negative relationship.

nature research

Corresponding author(s): Natalie Cohen, Mak Saito, Andrew Allen

Last updated by author(s): Sep 22, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Proteome Discoverer v2.1 was used to perform peptide-spectral matches, and Scaffold 4.8.7 was used to identify protein groups and generate spectral counts
Data analysis	Ribopicker v.0.4.3, FragGeneScan v1.16, BLAST+ 2.2.31, PhyloDB v1.076, HMMer 3.0, FASTA36 v36.3.6, SWARM v2.1.10, Protist Ribosomal Reference (PR2) v4.11.1 , WGCNA v.168, clusterProfiler v3.12.0, coin v.1.3-1, phyloseq v.1.28.0, vegan v.2.5-6, ggplot2 v3.2.1 v., MUSCLE v3.8, Jalview v2.10.5, RAxML v8.2.11, seqmagick v.0.6.2, taxtastic v0.8.11, hmmer v3.1b2, pplacer v1.1alpha19, guppy v1.1, ggtree v1.16.0, Ocean Data View 4 v.4.7.10, Sigmaplot 14, R v3.6.0, Geneious v11.1.4, ProteoWizard v3, BWA-MEM v0.7.12-r1039, CLC Assembly Cell (clc_novo_assemble) v3.22.55708, spades v3.8.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The mass spectrometry global proteomics data and metatranscriptome-derived FASTA file has been deposited to the ProteomeXchange Consortium through the PRIDE 99 repository under accession number PXD014230. Metaproteomic annotations and total spectral counts from this analysis have additionally been uploaded

onto the Ocean Protein Portal (proteinportal.whoi.edu/) for user-friendly visualization and exploration. Raw metatranscriptomic reads have been deposited to the National Center for Biotechnology (NCBI) under Bioproject number PRJNA555787. The 18S raw reads are available on NCBI under Biosample accession numbers SAMN12331629-SAMN12331670, and 16S raw reads under Biosample accession numbers SAMN12332710-SAMN12332751. The 0.2-3 µm assembly used to extract Prochlorococcus proteins has been deposited onto NCBI under accession GCA_900411625. Nutrients, dissolved cobalt, pigments and CTD physiochemical information is available at NSF's BCO-DMO repository under project number 2236. The PhyloDB database used to obtain taxonomic and functional annotations can be accessed from github.com/allenlab/PhyloDB. There are no restrictions on the data.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Multi-'omics and geochemical analyses were performed across a ~4,600 km section in the central Pacific Ocean to capture microbial community metabolism along a surface gradient and across ocean depths. Biomass was collected onto forty-two 3-51 um filters using underwater McLane pumps.
Research sample	The research samples comprised of natural plankton communities residing in the surface and mesopelagic zones of the tropical and equatorial Pacific Ocean. The exact makeup of these plankton communities was unknown at the time of sampling. The microeukaryote fraction (in the 3-51 um size range) was the focus of this analysis given their under-explored biogeographical distributions across this region of the ocean.
Sampling strategy	Pumps filtered between 165-1,384 L of seawater, and concentrated microbial biomass consisting of prokaryotes, eukaryotes, and viruses. Pumps were deployed for 4-6 hours to allow for biomass to accumulate, with the upper limit determined by the time allotted to hold on station. The amount of volume filtered depended on biomass present and its influence on filtration flow rates.
Data collection	McLane pump logs document the depth targeted and duration. At each station where McLane pumps were deployed, a trace metal rosette CTD collected accompanying physiochemical parameters. A cruise log was used to record each operation under of the supervision of the Chief Scientists.
Timing and spatial scale	Sampling occurred between October 1st and 25th 2011 onboard the R/V Kilo Moana. Stations were chosen before the research cruise to evenly capture biological and chemical transitions along the surface gradient, including the oligotrophic tropical northern and southern gyres and equatorial upwelling regime. Depths were selected to characterize differences in metabolic strategies between surface and mesopelagic communities. Seven sites were chosen for sample collection between 7N to 15S (154W-173W). Three to thirteen depths were sampled per station, with the majority between 40-600m. Sampling commenced when the ship arrived on station, with the gap between sampling depending on transit time and length spent on station. Exact sampling times for each station can be accessed from the cruise BCO-DMO data repository: https://www.bco-dmo.org/project/2236
Data exclusions	The meta-'omic analysis focused on the 3-51 um fraction collected from the McLane pumps to isolate eukaryotic contributions. The prokaryotic 0.2-3um and filamentous/higher trophic level (51-150um) size fractions were not the focus of this analysis. A single sample was additionally collected at 1,900 m to target hydrothermal vent activity. This was the only sample collected at a bathypelagic depth, with all others collected in euphotic and mesopelagic waters. It was pre-established that this sample would be the focus of a separate analysis linking hydrothermal geochemistry to microbial ecology.
Reproducibility	Oceanographic samples were collected once to generate a spatial section in a highly heterogeneous and dynamic environmental system
Randomization	Randomization does not apply to field collection as sample depths and sites were predetermined based on oceanographic features.
Blinding	Blinding was not possible as all samples analyzed were associated with a specific site and depth. Clustering approaches allowed for structural patterns in the data to be computationally identified regardless of assigned category.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Sampling occurred during October of 2011. Supplementary Fig. 1 and Extended Data Fig. 1 and 3 present environmental data at time of sample collection including, water temperature, chlorophyll, pigments, salinity, oxygen concentration, density, and macronutrients nitrate+nitrite, phosphate, and silicic acid. These datasets and other co-occurring additional expedition datasets are available on the expedition project page at BCO-DMO (https://www.bco-dmo.org/project/2236)
Location	The meridional transect (17°N-15°S) began off the Hawaiian Islands (154.4°W) and terminated in the Tonga-Fiji region (173.1°W). Most sampling depths ranged between 40-600m.
Access & import/export	Samples were collected during the cruise onboard the R/V Kilo Moana with conduct permitted by the National Science

Access & import/export	Foundation research vessel crew. Three permits were issued. Cook Islands High Commission Note No 304/11 granted access to waters near the Cook Islands from Oct. 9-22 2011; The Ministry of Foreign Affairs and Immigration of the Republic of Kiribati FA:48/12/395A provided clearance for sampling from Oct 1-30 2011; The Ministry of Foreign Affairs of the Kingdom of Tonga F.7/2/3 allowed access to waters under the jurisdiction of the Kingdom of Tonga from Oct 1-30 2011.
Disturbance	No disturbance was caused by oceanographic sample collection

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging