nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-021-01606-w

Check for updates

Complex marine microbial communities partition metabolism of scarce resources over the diel cycle

Daniel Muratore[1,2,17], Angela K. Boysen[3,4,17], Matthew J. Harke[5,14,17], Kevin W. Becker[6,7], John R. Casey[8,9], Sacha N. Coesel[3], Daniel R. Mende[8,10], Samuel T. Wilson[8], Frank O. Aylward[11], John M. Eppley[8], Alice Vislova[8], Shengyun Peng[12], Rogelio A. Rodriguez-Gonzalez[1,2], Stephen J. Beckett[1], E. Virginia Armbrust[3], Edward F. DeLong[8], David M. Karl[8], Angelicque E. White[8], Jonathan P. Zehr[13], Benjamin A. S. Van Mooy[7], Sonya T. Dyhrman[14,15], Anitra E. Ingalls[3] and Joshua S. Weitz[1,16] ✉

Complex assemblages of microbes in the surface ocean are responsible for approximately half of global carbon fixation. The persistence of high taxonomic diversity despite competition for a small suite of relatively homogeneously distributed nutrients, that is, 'the paradox of the plankton', represents a long-standing challenge for ecological theory. Here we find evidence consistent with temporal niche partitioning of nitrogen assimilation processes over a diel cycle in the North Pacific Subtropical Gyre. We jointly analysed transcript abundances, lipids and metabolites and discovered that a small number of diel archetypes can explain pervasive periodic dynamics. Metabolic pathway analysis of identified diel signals revealed asynchronous timing in the transcription of nitrogen uptake and assimilation genes among different microbial groups—cyanobacteria, heterotrophic bacteria and eukaryotes. This temporal niche partitioning of nitrogen uptake emerged despite synchronous transcription of photosynthesis and central carbon metabolism genes and associated macromolecular abundances. Temporal niche partitioning may be a mechanism by which microorganisms in the open ocean mitigate competition for scarce resources, supporting community coexistence.

Marine microbial dynamics in the surface oceans are driven, in part, by periodic light forcing on a diurnal cycle. In response to periodic light input, marine microbes (including complex assemblages of cyanobacteria, eukaryotes and heterotrophic bacteria) exhibit diel patterns of transcriptional regulation of key metabolic processes such as photosynthesis, nutrient assimilation and diazotrophy, and energy storage[1–12]. In turn, daily oscillations in aggregate measures of community activity (for example, particulate organic carbon in the North Pacific Subtropical Gyre (NPSG)[13]) suggest that the integrated rhythms in transcription and metabolic regulation scale up to influence biogeochemical processes such as light capture and export of matter and energy from the euphotic zone. However, it is unclear how complex assemblages of microbes coexist while competing for a limited set of scarce nutrients and how the aggregate effect of diverse metabolic processes, not necessarily shared by each taxon, leads to emergent patterns in community-level and biogeochemical dynamics.

Establishing a link between microbial diversity, gene content and ecosystem function can be particularly relevant when comparing the ecological state of different sites or seasons with significant underlying variation in diversity[14–16]. However, there are challenges in inferring ecosystem function from sequence data alone, for example, variable lags between transcription and translation[17] and potentially unknown transcript-specific degradation rates and protein activity and turnover rates[18,19]. In the ocean, challenges in inferring ecosystem function are exacerbated by incomplete characterization of metabolic pathways and transcriptional regulatory mechanisms for many microorganisms[20]. Direct observations of biomolecule concentrations across diel cycles have revealed diurnal variation in cellular N- and C-content[6,12,21–24], diel rhythmicity in lipid-body formation in coral symbionts[25], changes in the concentrations of low-molecular-weight dissolved organic matter[26] and periodicity in intracellular metabolic products[6,12,27] in surface marine ecosystems. Altogether, these studies suggest that temporal changes in the microbial metabolism linked to the diurnal light cycle have the potential to reveal principles underlying the uptake, use and assimilation of scarce nutrients.

Recent fine-scale temporal resolution studies revealed robust diel patterns in metatranscriptome oscillations within complex assemblages of both photoautotrophs and heterotrophic bacteria[4,8,28–30],

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. [2]Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, GA, USA. [3]School of Oceanography, University of Washington, Seattle, WA, USA. [4]Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA. [5]Gloucester Marine Genomics Institute, Gloucester, MA, USA. [6]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany. [7]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [8]Department of Oceanography, Daniel K. Inouye Center for Microbial Oceanography: Research and Education, University of Hawai'i at Mānoa, Honolulu, HI, USA. [9]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [10]Department of Medical Microbiology, Amsterdam University Medical Center, Amsterdam, the Netherlands. [11]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. [12]Adobe, San Jose, CA, USA. [13]Ocean Sciences Department, University of California, Santa Cruz, Santa Cruz, CA, USA. [14]Lamont-Doherty Earth Observatory, Biology and Paleo Environment, Columbia University, Palisades, NY, USA. [15]Department of Earth and Environmental Science, Columbia University, New York, NY, USA. [16]School of Physics, Georgia Institute of Technology, Atlanta, GA, USA. [17]These authors contributed equally: Daniel Muratore, Angela K. Boysen, Matthew J. Harke. ✉e-mail: jsweitz@gatech.edu

complementary to efforts to survey for microbial diversity at the broadest ocean scales[15,31–33] and to time series studies meant to infer associations within complex microbial communities at monthly or seasonal scales[28,34,35]. In this study, we analysed metatranscriptomes, lipidomes, macromolecules and metabolomes collected every 4 h over approximately 3 d in the summer of 2015 in the NPSG (http://scope.soest.hawaii.edu/data/hoelegacy/) (Fig. 1a,b) to examine the patterns underlying microbial community dynamics in this critical ecosystem. We used a combination of time series analytics and machine learning methods to examine regularity, clustering and (a)synchronization in transcripts among diverse taxa and in aggregate indicators of community metabolism. Our analysis demonstrates how quantifying and incorporating diel variation in resource utilization in response to daily light forcing can help advance efforts to understand the coexistence of diverse marine microbes and, in turn, help develop realistic multiscale models of the mechanistic links between diversity, nutrient availability and ecosystem functioning.

Results

Pervasive diel periodicity in cross-domain transcription, metabolites, lipids and macromolecules. We first set out to determine the extent of diel periodicity within time series of transcriptomes, metabolites and lipids. We leveraged gene reference databases tailored to marine microorganisms[36,37] to taxonomically classify transcripts from both >0.2 µm and >5 µm size fraction metatranscriptomes (Methods). To facilitate comparisons across taxa, a subset of transcripts with Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology annotations[38] were used as the basis for subsequent analyses. We used targeted metabolomics to measure the concentrations of particulate metabolites, including free amino acids, saccharides and vitamins. We also used lipidomic methods to measure six different classes of molecules: cell membrane-related lipids (phospholipids, betaine lipids); triacylglycerols (TAGs); chloroplast membrane-related lipids (mono- (MGDGs) and digalactosyldiacylglycerols (DGDGs), sulfoquinovosyldiacylglycerols); pigments; carotenoids; and quinones (Methods). The joint measurement of transcripts, lipids and metabolites allowed us to identify associations between suites of genes involved in the synthesis and degradation of molecules in particulate pools. In total, we tested for diel periodicity among 997 unique lipids, 77 metabolites, total hydrolysable amino acids (THAAs), total hydrolysable nucleobases (THNB) and 64,011 transcripts. The transcripts mapped to 5,540 KEGG orthologues from 27 major prokaryotic clades in a metatranscriptome sequenced from a >0.2 µm size fraction and 58,471 KEGG orthologues across a selected list of 14 major eukaryotic phyla (Methods) sequenced from the >5 µm size fraction. We assessed periodicity using a non-parametric method[39], accounting for non-stationarity and multiple testing (Methods).

We identified 6,273 time series with statistically significant diel rhythms out of more than 65,000 time series examined. These significant diel time series encompassed 501 out of 997 lipids (50.2%), 50 out of 77 metabolites (64.9%), 1,739 out of 5,540 of the > 0.2 µm fraction transcripts (31.4%) and 3,983 out of 58,471 of the >5 µm fraction transcripts (6.8%) (see Fig. 1 for a representative subset of diel time series from each analyte type). Diel rhythms were identified in all tested taxa in both size fractions, all classes of lipids measured and both primary and secondary metabolites, as well as in macromolecular measurements of THAAs and THNBs (Supplementary Table 1 and details in Supplementary Data 2). In addition to metabolites and macromolecules shared across all microbial groups, we identified diel transcriptional patterns in more than 3,000 unique KEGG orthologues, spanning functions that include photosystems, photosynthetic carbon fixation and central carbon metabolism as well as macro- and micronutrient and metal uptake. (A comprehensive list of all studied transcripts and associated pathway annotations is listed in Supplementary Data 2.)

Unsupervised learning approach to partition and interpret diel signal patterns. The cumulative set of 6,273 diel periodic signals differed in amplitude, shape of oscillation and peak timing. To identify recurring patterns among these signals, we implemented an unsupervised, self-organizing map (SOM) approach to determine the extent to which the data could be represented by a far smaller set of archetypes, each with its own characteristic temporal signature (Methods). SOM analysis revealed that diel signals robustly cluster into 4 archetypal time series with peaks at dusk (18:00 h), night (2:00 h), morning (6:00 h) and afternoon (14:00 h) (Fig. 2a,b; analysis of clustering resolution in Extended Data Fig. 1 and Supplementary Data 4). Notably, this clustering did not rely on a priori assumptions of sinusoidal patterns nor of preferred phase. Of the total signals, the 'dusk', 'night', 'afternoon' and 'morning' clusters comprised 36, 22, 18 and 24% of the signals, respectively (Supplementary Table 1). Archetypal diel clusters were heterogeneous in analyte type and taxonomic identity (that is, each including bacterial and eukaryotic transcripts, lipids and metabolites; Fig. 2c and Supplementary Table 1).

We implemented a pathway enrichment analysis to summarize the distributions of transcripts for metabolic pathways among the four SOM clusters (Methods). Briefly, diel transcripts were split among three broad categories: cyanobacteria; heterotrophic bacteria; and eukaryotes. Within each category, diel transcripts were grouped by their assigned KEGG pathway. Fisher's exact tests were used to identify KEGG pathways that were significantly overabundant in any of the four SOM clusters (Supplementary Data 1; see Supplementary Discussion for extended results and discussion). We found significant enrichments for cyanobacterial and eukaryotic photosynthesis transcripts in the morning cluster, as found previously across oceanic ecosystems[1,8,29,30]. We also found evidence for synthesis of nucleoside and amino acid precursors, carbohydrates and carbon fixation in the afternoon, carbohydrate catabolism at dusk and de novo protein synthesis at night (Supplementary Discussion). Notably, we found indications of synchronized diel responses by heterotrophic bacteria, for example, enrichment of the tricarboxylic acid (TCA) cycle pathway in the afternoon consistent with an increase in organic carbon catabolism corresponding to the accumulation of fixed carbon by photosynthetic organisms. We found evidence for protein synthesis occurring overnight for many eukaryotes and cyanobacteria, supporting a long-standing hypothesis of staggered diel patterns in resource acquisition, allocation and cell division[40]. Transcript abundance patterns corroborate with macromolecular analyses of THAAs and THNBs, which we used as a proxy for protein-associated and nucleic acid-associated nitrogen, respectively (Methods and Extended Data Fig. 2). Both THAAs and THNBs were found to have diel periodicity and were assigned to the night cluster along with ribosomal component transcripts and nitrogen assimilation gene transcripts.

Altogether, we observed a synchronized cascade across diverse bacterial and eukaryotic photoautotrophs from photosynthesis in the morning and afternoon to the accumulation of organic storage molecules at dusk. Specifically, Fig. 3 shows the average levels of TAGs peak at dusk and decline overnight, as do organic metabolites with putative functions such as carbon storage molecules, fatty acid precursors or compatible solutes[12]. The oscillation of these carbon storage components matches optically derived particulate organic carbon (POC) estimates. The accumulated carbon provides energy for respiration and a transition to protein synthesis overnight, culminating in the synthesis of photosynthetic machinery and pigments for the upcoming dawn. We also identified a synchronized feedback by heterotrophic bacteria to the primary productivity cascade (Fig. 3). Specifically, we found that heterotroph expression of sugar uptake transporters coincides with the decline of particulate stocks of sugars and lipids. The diel timing of heterotroph sugar transporter-associated transcription, further supported by diel

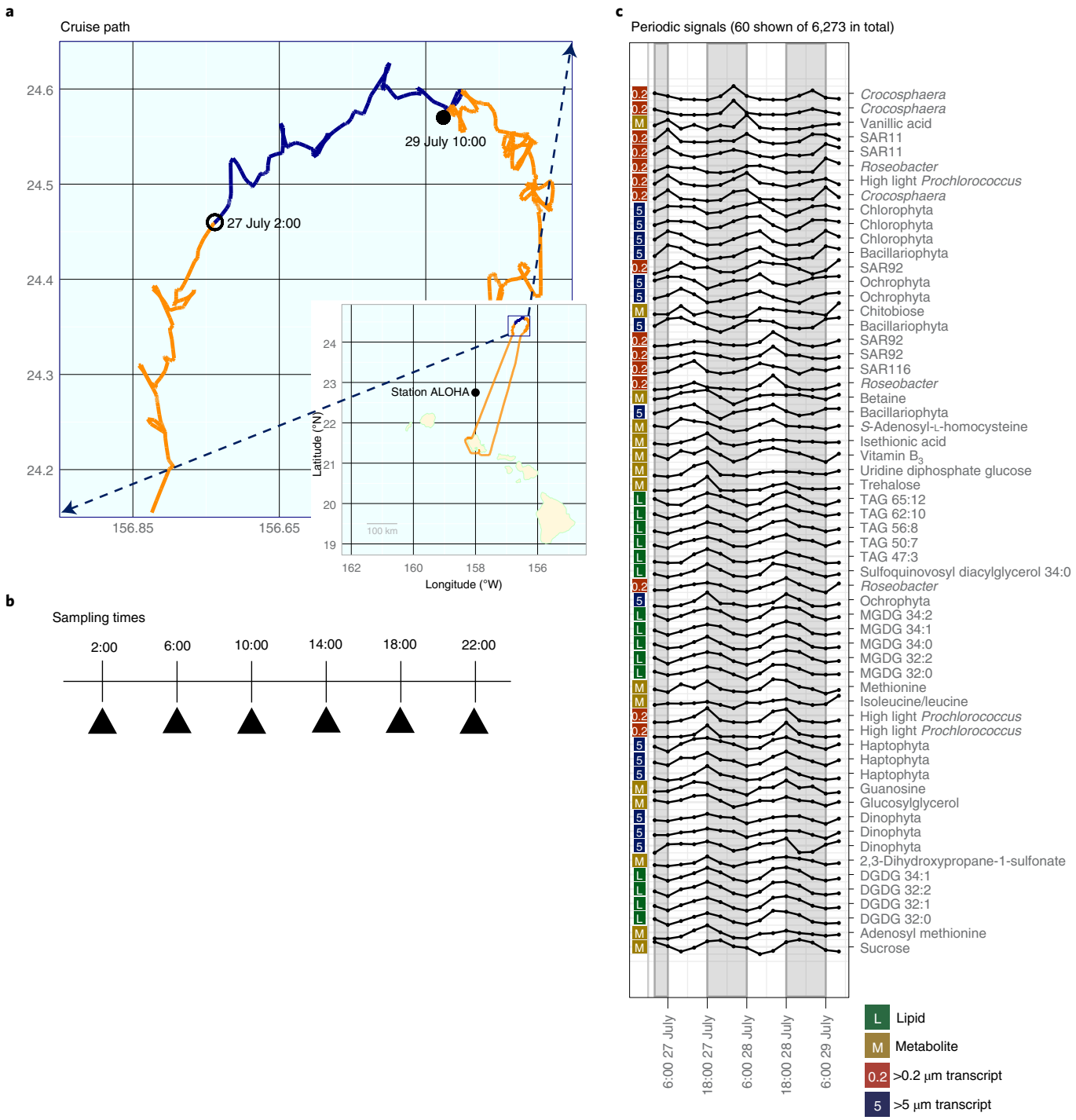


Fig. 1 | Diel patterns in diurnally resolved multi-omics at Station ALOHA. a, Map showing the Lagrangian cruise track of the HOE Legacy IIA cruise (orange line) with the samples used for this study taken over a 3 day period (blue line). **b**, Sampling times; 4 h between samples. **c**, Selected periodic signals determined by non-parametric analysis (Methods) ordered by peak time. The subset denotes the 15 signals from each dataset with the lowest *P* values (60 shown of 6,273 total), representing a combination of particulate metabolites, lipids and transcripts from both the >0.2 μm (predominantly prokaryotic organisms) and >5 μm size fractions. All measurements were scaled to have mean 0 and variance 1. The grey boxes indicate night-time hours (18:00–6:00).

oscillations in the TCA cycle pathway among heterotrophs (Fig. 3), suggests that cascades through the marine food web across broad domains follow the diel light cycle, as hypothesized previously[29,30]. A recent study using data collected in the same sampling period and location found a diversity of (diel-expressed) genes coding for proteins with light sensitive domains among eukaryotic taxa[41], including mixotrophic and heterotrophic taxa, further suggesting ecosystem-wide responses to the diel cycle.

Fine-grained diel transcriptional dynamics among taxa. Examining patterns of transcriptional timing across eukaryotes, cyanobacteria and bacterial heterotrophs can provide insights

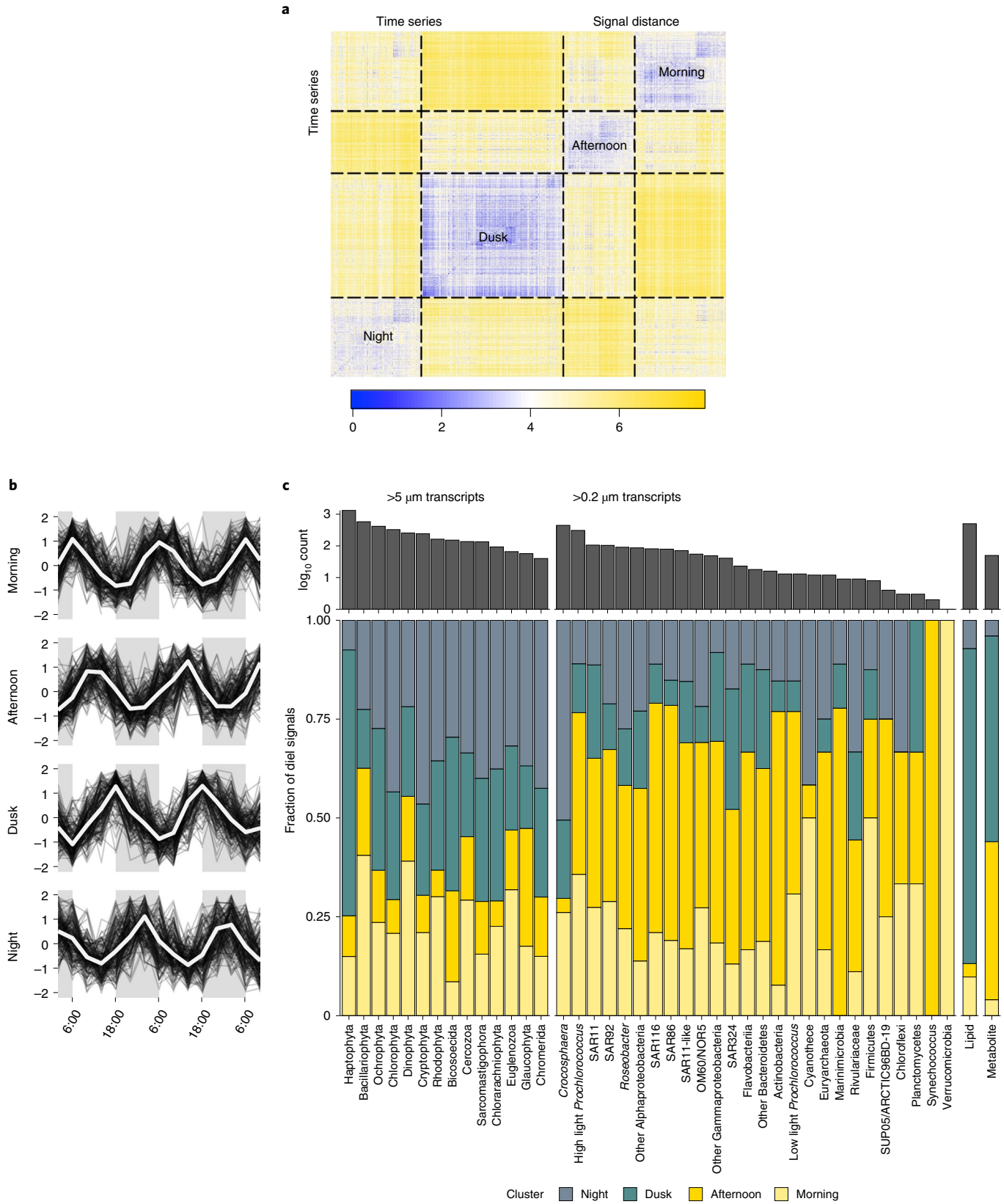


Fig. 2 | Unsupervised clustering analysis categorized diel patterns into four potential archetypes. a, Organized pairwise distance matrix for all diel measurements after clustering based on SOMs. Each pixel represents the Euclidean distance between the time series of two diel measurements, with blue indicating a small distance (similar time series) and yellow indicating a larger distance (less similar time series). Boxes are drawn around cluster boundaries. **b**, Archetypal time series for four clusters (beige lines); archetypes are a combination (determined via the SOM algorithm) of all time series in their cluster. A random sample of 200 time series belonging to each cluster are plotted as dark lines. **c**, Archetype affiliation by measurement class. Distribution of fractions of diel signals across clusters for transcripts assigned to taxa from (left to right) the >5 μm transcriptome, >0.2 μm transcriptome, lipidome and metabolome. The corresponding bar chart above indicates the quantity of signals found to be diel belonging to each group (note the log scale).

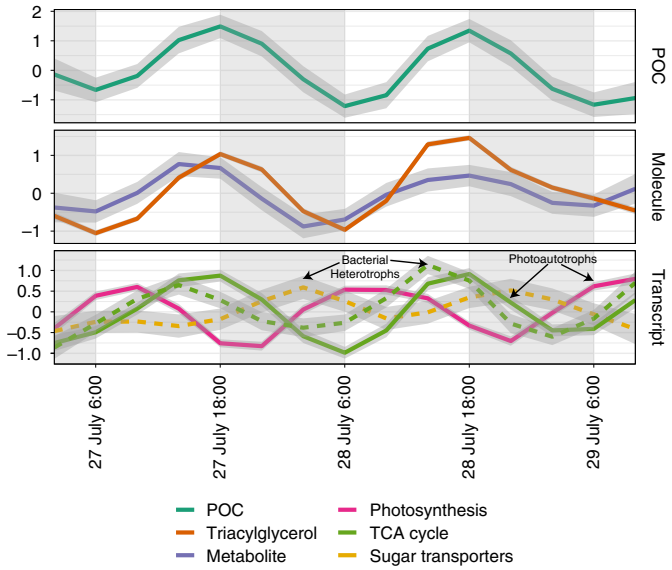


Fig. 3 | Carbon-related transcriptional and biogeochemical activity at the community scale. Averaged *z*-score-scaled time series of optically derived POC concentration, concentration of carbon fixation-associated lipids, metabolites associated with carbon storage and compatible solutes (specifically uridine diphosphate-glucosamine, trehalose, choline, glucosylglycerol, ribose-5 phosphate, 2,3-dihydroxypropane-1-sulphonate, gluconic acid, sucrose, vanillic acid, chitobiose, trans-retinal, dimethylsufoniopropionate and homarine) and transcript levels of transcripts involved in the photosynthesis and TCA cycle KEGG pathways, as well as transcripts with sugar transporter function assigned to heterotrophic bacteria. The dashed lines indicate transcripts assigned to heterotrophic bacteria. Data in these panels were smoothed using a generalized additive model with cubic spline smoothing with shrinkage penalties on all observations (95% confidence interval shown in the shaded area).

into specific metabolic functions that may be overlooked in pathway-level analysis. To compare diel transcriptional patterns between taxa, we projected the aggregate time series of diel signals into a lower-dimensional space using non-metric dimensional scaling (NMDS; Methods). The NMDS projection highlights the differences in overall transcriptional patterns between taxa (Fig. 4a) and metabolite/lipid abundances (Fig. 4b), and naturally projects the data into a 24 h 'clock-like' space (Fig. 4). For example, Haptophytes (a phylum of algae) transcribe a higher proportion of genes with diel expression at dusk, the diazotrophic cyanobacteria *Crocosphaera* transcribes many diel genes at night and the genes of the cyanobacteria *Prochlorococcus* with diel transcription tend to peak in the morning. Taxon-specific diel transcriptional peaks occur throughout the day. The differential distribution of diel transcripts across clusters are consistent with characteristic 'profiles' when comparing among different taxa (chi-squared test of homogeneity, d.f. = 135, $P < 1 \times 10^{-5}$). Identified taxa had genes with diel transcriptional peaks at all times of day, belonging to all SOM archetypes. Broad distributions of transcriptional peaks suggest organisms have changing transcriptional states over the daily cycle as opposed to a uniform biomass-specific transcript abundance changing as population abundances oscillate. Abundances were measured for a few key community members (for example, *Prochlorococcus* and *Crocosphaera*); in those cases, the timing of transcript peaks did not recapitulate the abundance peaks (see Supplementary Discussion for an extended analysis).

Next, we explored the functional basis underpinning differences in diel peak timing for different taxa. To do so, KEGG orthologues were classified into three categories: (1) 'taxonomically narrow' diel transcripts, corresponding to KEGG orthologues that only show diel transcriptional patterns in three or fewer taxa; (2) 'synchronous' diel transcripts, corresponding to KEGG orthologues that show diel expression in more than three taxa and have a uniform diel transcriptional pattern (that is, their expression peaks at similar times); and (3) 'asynchronous' diel transcripts, corresponding to KEGG orthologues that are expressed in more than three taxa but have discordant diel transcriptional patterns (that is, their expression peaks at different times). Of the 3,193 KEGG orthologues that showed a diel transcriptional pattern in any taxon, the vast majority (2,898) were taxonomically narrow, that is, a diel signature was only observed in three or fewer taxa (Supplementary Table 2). Hence, characteristic taxonomic profiles (Fig. 4) were driven primarily by transcripts that had diel expression in a small set of taxa in our dataset. This may point to taxon-specific differences in regulation and functional capacity. The remaining 295 widely shared KEGG orthologues were classified as synchronous or asynchronous based on the average difference in peak time across all associated taxa for that KEGG orthologue relative to a null expectation (Methods and Extended Data Fig. 3).

Of the 295 widely shared KEGG orthologues, we identified 80 as having synchronous diel transcription (Benjamini–Hochberg-adjusted $P < 0.1$) (Fig. 5 and Supplementary Table 2). The synchronous orthologues with the most associated diel signals included photosystem II components and cytochrome C oxidases. Cytochrome C oxidases are found in the photosynthetic electron chain as well as in oxidative phosphorylation in both photosynthetic and non-photosynthetic organisms (for example, heterotrophs). The widespread diel synchronicity of these orthologues is consistent with the hypothesis that biophysical and/or regulatory constraints transcend taxa in dictating peak time for functions related to photosynthesis and oxidative phosphorylation (Fig. 3), as seen in other community-level studies[29]. We interpret this evidence to imply that the emergent primary productivity cascade and resultant patterns in POC (Fig. 3) are driven, in part, by a few highly conserved genes with widespread synchronous diel expression.

Temporal niche partitioning and nitrogen metabolism. In evaluating transcripts that are 'widespread' (that is, found in more than 3 taxa), we identified 25 different KEGG pathways not related to photosynthesis or carbon fixation, some of which included asynchronous diel expression patterns. An initial curation of asynchronous orthologues with large average peak-time differences included transcripts annotated as *gltB* (glutamate synthase NADPH large chain subunit) and other components of the glutamate synthase-glutamine oxoglutarate aminotransferase (GS-GOGAT) system (Fig. 5a). Glutamate synthase (part of the GS-GOGAT system) is the first intracellular step in a widespread pathway for ammonia assimilation[42]. The expression of *gltB* spanned both heterotrophs and autotrophs, bacteria and eukaryotes and all four SOM clusters (Extended Data Fig. 4). Similarly, transcripts associated with other GS-GOGAT subunits exhibited asynchrony, including *gltD*, *gltS*, *glnA* and the nitrogen-regulatory PII response protein *glnK* (Fig. 5a). GS-GOGAT components associated with *Prochlorococcus* peaked in the afternoon and dusk, as reported previously[3]. Furthermore, transcripts annotated as the ammonium transmembrane transporter *amt* had diel expression across ten different taxa and showed strong asynchrony (Fig. 5a,b). We found that photoautotrophs had peak *amt* gene expression at dusk (concurrent with the average signal of GS-GOGAT component transcripts), while heterotrophs had peak expression in the morning (again, concurrent with GS-GOGAT component transcripts). As such, it appears that the expression of key steps required for nitrogen uptake and assimilation are periodic and asynchronous, that is, both GS-GOGAT components and ammonium transport transcript levels repeatedly peak at different

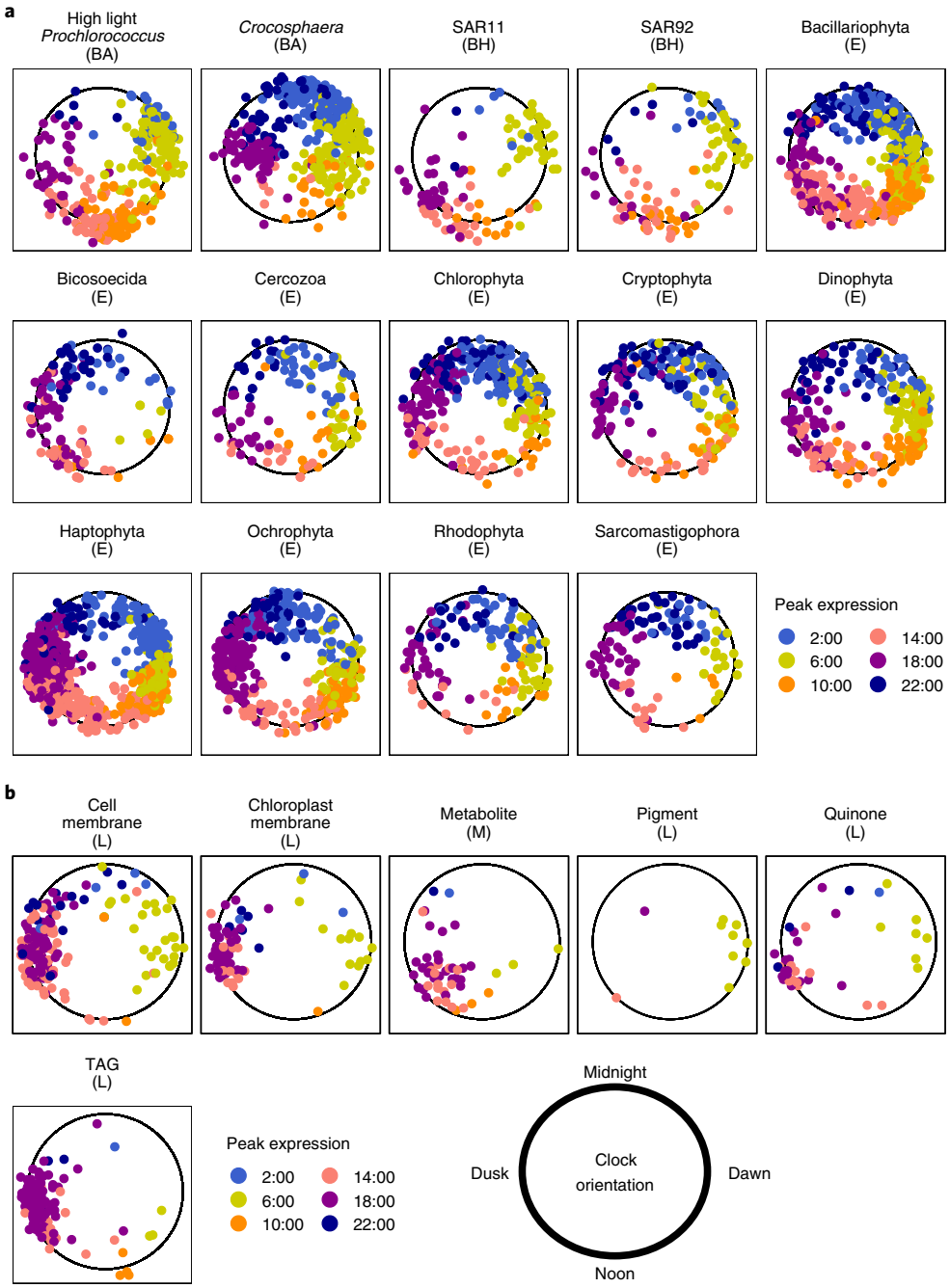


Fig. 4 | NMDS projection of the time series for each diel measurement in the transcriptomes, lipidomes and metabolomes. a, Diel taxonomic transcriptional profiles. Each point represents one gene with diel transcription. Diel transcriptional peaks are distributed around an emergent 24-hour 'clock' and among community members and metabolic functions. Points are coloured by calculated peak rank measurement time (Methods). For reference, the projection was rotated so that time proceeded clockwise from midnight (top of the circle) to noon (bottom of the circle). The parentheses indicate taxonomic affiliation: E, eukaryote; BH, bacterial heterotroph; BA, bacterial autotroph. Only taxa with at least 100 diel transcripts are shown. See Methods for details about taxonomic assignments. **b**, Diel lipid (L) and metabolite (M) profiles. Each point represents one metabolite or lipid. Lipids are separated by functional categories.

times in the day across taxonomic groups (Fig. 5a,b). We compared this result to other transcripts identified in ammonia assimilation via glutamate/glutamine as described by KEGG pathways. We found 11 KEGG orthologues that perform similar reactions (such as glutamate dehydrogenase *gdhA*; Extended Data Figs. 4 and 5). We found that these orthologues were generally not detected in any of the bacterial taxa we analysed; while they appeared in eukaryotic taxa, the vast majority did not have diel transcript abundances (Extended Data Fig. 4). Consequently, we determined that while other mechanisms of ammonia assimilation may be occurring in the ecosystem, it is noteworthy that GS-GOGAT-based assimilation is typically diel and asynchronous across taxa. Asynchronous timing of the expression of nitrogen assimilation-related transcripts suggests that marine microbes differ in their strategies of utilizing a key nutrient, even though nitrogen uptake and assimilation are not known to be directly linked to light forcing (Fig. 5c). While the

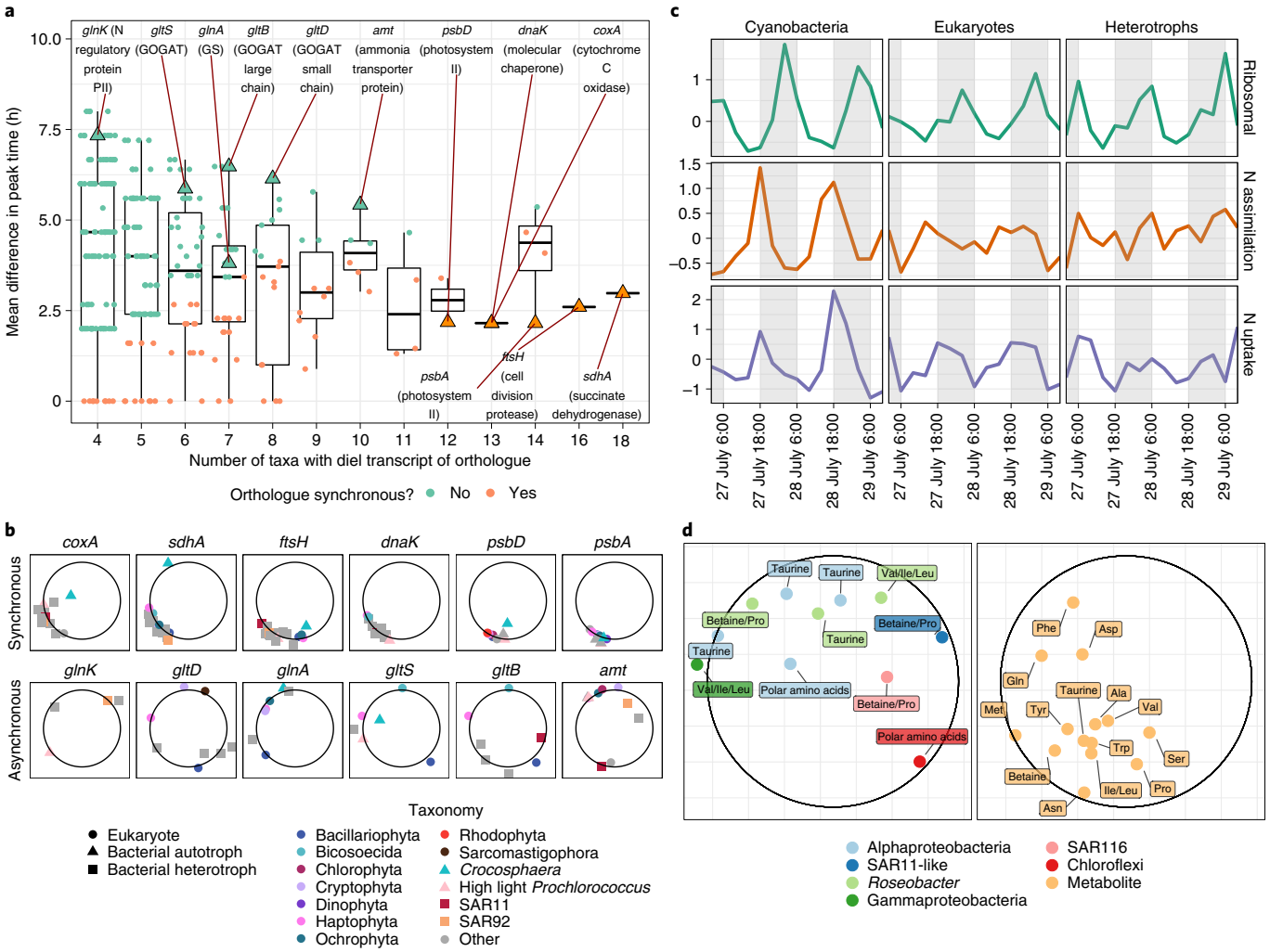


Fig. 5 | Niche partitioning and nitrogen metabolism at Station ALOHA. a, For each KEGG orthologue with diel expression in at least four different taxa, all pairwise differences in peak rank time were tabulated and averaged (Methods). Low average difference in peak time indicated that most taxa with diel transcription of that orthologue peaked at the same time of day. Orthologues with smaller average differences in peak times than would be expected from the population of all diel transcripts with a Benjamini–Hochberg-adjusted $P < 0.1$ are indicated in orange. Orthologues related to the GS-GOGAT system are labelled and indicated as green triangles. Orthologues with a low average peak-time difference expressed widely, related to primary production and central carbon metabolism, are labelled as orange triangles. **b**, NMDS projection of the subset of transcripts identified as examples in **a**. The top row shows all orthologues with synchronous timing; the bottom row shows asymmetric timing. The top of the projection roughly indicates midnight peak time and the bottom indicates noon. **c**, Transcription dynamics of nitrogen uptake and assimilation among different taxonomic groups. The lines indicate mean *z*-score transcription of all diel ribosomal genes (top row), GS-GOGAT pathway components (middle row) and *amt* ammonia transporters (bottom row). The shaded bars indicate dark hours. **d**, NMDS projection showing peaks in expression for heterotrophic bacteria amino acid uptake transporters (left) and diel amino acids from the metabolome (right). Points are labelled with which amino acid they are or which amino acid the transporter takes up. Colour indicates taxonomic affiliation.

timing of transcription does not always indicate simultaneous metabolic activity[17], the alignment between the bulk particulate THAA/THNB diel cycle (Extended Data Fig. 2) and nitrogen assimilation gene transcription of photoautotrophs—abundant community members consisting of a substantial portion of total detected biomass[12]—suggests that gene transcription timing is reflective of realized nitrogen assimilation. We hypothesized that diverse microbial community members distributing nitrogen demand over the course of the diel cycle reduced community-wide demand at any given time, thereby lessening overall competition. This apparent temporal partitioning of nitrogen uptake and metabolism suggests that nitrogen stress mitigation at the community scale may underpin the coexistence of diverse microbes in the NPSG. Nitrogen assimilation is of particular interest in the NPSG due to the exceedingly low background concentration of inorganic nitrogen species (nitrate, nitrite, ammonia), resulting in persistent and chronic nitrogen limitation and a strong reliance on remineralization to supply useable nitrogen[43].

Inorganic forms of nitrogen are extremely dilute in the epipelagic NPSG and comprise a small portion of total dissolved nitrogen[44]. However, uptake and assimilation of organic forms of nitrogen may also comprise key portions of diel nitrogen budgets[45–47]. To investigate the potential for diel uptake and assimilation of organic forms of nitrogen, we searched for uptake transporter genes with diel expression associated with heterotrophic taxa. We found that 11 of the 71 total diel transporter genes assigned to heterotrophic bacteria take up nitrogen-containing metabolites, including leucine/isoleucine, valine, glycine betaine, proline and taurine (Fig. 5d).

The particulate concentrations of these molecules are all identified as diel, indicating that they are synthesized and accumulate in cells on a diel cycle. The THAA pool is also diel and belongs to the night cluster, suggesting that some of these components may also contribute to the diel oscillation of nitrogen-rich macromolecules (protein) (Supplementary Table 3). In comparison with heterotrophic taxa, *Prochlorococcus* shows diel expression of the urea uptake transporter *urtA* in the night cluster, which is coincident with expression of ureases in picoeukaryotes (Supplementary Table 3). Notably, this timing follows after the peak transcription of *amt* at dusk, indicating potential sequential transport of ammonia and then urea overnight for these photoautotrophs. Urea is known to be among the important sources of organic nitrogen for photoautotrophs in the NPSG[48,49]. Night-time expression of urea uptake and catabolism genes may reflect an increase in supply due to exudation from active phagotrophic protists[50] or nocturnally feeding macroscopic zooplankton[51]. Thus, we found potential niche-specific preferential uptake of diverse organic nitrogen species throughout the day, which is analogous to our finding of temporal partitioning of uptake and assimilation of inorganic nitrogen species across the microbial community in the NPSG.

In addition to macronutrients, our data indicate diel regulation of uptake and synthesis of micronutrients (particularly cobalamins and iron) across photoautotrophs and heterotrophic bacteria (Supplementary Data 6; see Supplementary Discussion for details). Briefly, we found evidence for night-time production and excretion of pseudocobalamin by cyanobacteria and daytime cobalamin use by eukaryotes, and iron transport distributed between the morning (siderophore uptake by heterotrophs), afternoon (iron(III) uptake by *Prochlorococcus*) and dusk (iron remodelling by *Crocosphaera*). This diel regulation suggests that temporal partitioning may extend to diverse metabolic processes across microbial communities in the NPSG. Again, this temporal partitioning of resources could underpin coexistence of diverse microorganisms (see extended treatment in the Supplementary Discussion and related analysis of oscillations in the carotenoid biosynthesis pathways[52]) by reducing competition for scarce resources through time. Given the observation of periodicity in the uptake of organic nitrogen species, we suggest that a comprehensive accounting of community metabolism will also need to account for mortality/loss processes (for example, as caused by grazing and viral lysis), which also exhibit periodic regularity[53].

Discussion

The synthesis of in situ multi-omics data revealed pervasive and previously unrecognized synchronicity in pathways and signals at Station ALOHA in the NPSG, including both functions directly mediated by light (for example, photosynthesis) and other metabolic processes that may be indirectly affected by other ecological and environmental drivers (for example, nutrient and organic substrate uptake, vitamin synthesis and exchange). Machine learning methods revealed both coherence in carbon-associated metabolism and asynchronous timing of nitrogen uptake and assimilation across the community. We hypothesize that asynchronous timing is consistent with diel niche partitioning, such that taxa may minimize competition for scarce resources such as macro- and micronutrients by restricting their demand for these resources to narrow time intervals via mechanisms analogous to temporal shading and/or a temporal storage effect[54–56]. The temporal storage effect is an ecological mechanism where species specialize in growing at separate times instead of simultaneously competing for resources, thereby reducing competition for limiting resources in a focal niche[56,57]. The temporal storage effect has been empirically observed in terrestrial plant communities[57,58], where species exhibit differential levels of growth, survival and colonization in wet versus dry periods. A similar phenomenon has been observed in competing species of metazoan zooplankton in lakes[59] and is the theoretical basis for a model demonstrating coexistence of phytoplankton in a periodically fluctuating environment[56].

Characterizing a difference in the responses of competitor species to temporal environmental fluctuation provides a route to identify the temporal storage effect[57]. As demonstrated in Fig. 5c, we found asynchronous transcription of nitrogen assimilation genes between cyanobacteria, eukaryotes and heterotrophs and corresponding changes in particulate chemical composition. These temporal shifts in nitrogen assimilation are consistent with a temporal storage effect mechanism among marine microorganisms in the oligotrophic ocean surface on rapid, ecologically relevant diel timescales[60]. These temporal niches may underlie coexistence among microbes competing for persistently scarce, periodically supplied resources (for example, nitrogen in oligotrophic ocean waters). Asynchronous timing of nitrogen-associated metabolism points towards the ecological relevance of feedback mechanisms that modulate the availability of scarce nutrients over the diel period, further reinforcing temporal niches. Indeed, taking up a limiting resource at different times has been imputed in theoretical studies to facilitate coexistence in phytoplankton communities[61]. In this study, we found additional empirical evidence for mitigating competition for nitrogen. As shown in Fig. 5d, different community members exhibit asynchronous diel cycling of uptake receptor expression for diverse organic substrates, suggesting that multiple physiological mechanisms may be at work in overcoming community-wide nitrogen limitation. Expanding available niche space by taking up diverse substrates for the same resource, that is, resource partitioning, has been empirically shown to facilitate coexistence in multiple ecological contexts[57,62]. While our study focused on a link between transcription and the variation in biomolecules inside cells, we note that future work should aim to assess the specific impact of downstream mechanisms impacting enzyme activity or post-transcriptional modification on microbial function to directly link transcriptional shifts with mechanisms of coexistence and emergent, ecosystem-level behaviours.

In summary, our joint analysis of bacterial and eukaryotic transcription supplemented by direct analysis of biomolecular concentrations (metabolites, protein, nucleic acids and lipids) reveals how microorganisms respond directly to light input, to each other and to their environment via the use, assimilation and regeneration of limiting nutrients in the surface ocean. The emergence of community-scale synchronization and temporal niches may enable coexistence in complex microbial communities and reinforces the need to integrate mechanistic studies of diel cycles with global-scale models to understand the maintenance, diversity and resilience of carbon and nitrogen cycles in future ocean scenarios.

Methods

Fieldwork design and sampling. Fieldwork was conducted from 25 July to 5 August 2015 in the NPSG. A Lagrangian sampling strategy was implemented whereby World Ocean Circulation Experiment Surface Velocity Profile drifters from the Pacific Gyre were deployed within the centre of a mesoscale anticyclonic eddy. The mesoscale eddy fields were identified using Archiving, Validation, and Interpretation of Satellite Oceanographic data; when field sampling occurred, the target anticyclonic eddy was located north of the Hawaiian Islands at 24.4 °N and 156.5 °W, with a diameter of approximately 100 km. Over the 12 d sampling period, shipboard measurements were conducted alongside the drifters as they performed an almost complete circular pattern with a diameter of approximately 44 km (Fig. 1a). Water column seawater sampling for diel measurements took place every 4 h for a period of 4 d (26–30 July) and 3 d (31 July–3 August) at a depth of 15 m corresponding to the depth of the drogue. Water column sampling was achieved using a 24 × 12 l Niskin bottle rosette attached to a conductivity, temperature, depth package (SBE 911Plus; Sea-Bird Scientific) with additional fluorescence, oxygen and transmissometer sensors. The sampling and analytical protocols for the vertical profiles of nutrients, particulates and flow cytometry-enumerated phytoplankton populations and heterotrophic bacteria were identical to those employed by the Hawaii Ocean Time-series programme (http://hahana.soest.hawaii.edu/index.html). A more detailed explanation of the sampling strategy and resulting datasets can be found in Wilson et al.[9].

Metabolite sample collection, extraction and analysis. Metabolite data were collected as described previously[12,27]. Briefly, 3.5 l of seawater was filtered onto a

47 mm, 0.2 µm Omnipore filter using a peristaltic pump and flash-frozen in liquid nitrogen. Samples were collected in triplicate at every time point. Filters were stored in a −80 °C freezer until the time of metabolite extraction. Metabolites were extracted as reported previously[12] with a modified Bligh and Dyer extraction[63] using 1:1 methanol:water (aqueous phase) and dichloromethane (organic phase) to extract aqueous and organic metabolites. Select isotope-labelled internal standards were added before or after extraction to aid in normalization[64]. Metabolites were measured with a Waters Xevo TQ-S triple quadrupole mass spectrometer and a Thermo Fisher Scientific Q-Exactive HF hybrid quadrupole-orbitrap mass spectrometer with both reversed-phase and hydrophilic interaction liquid chromatography. Metabolite peaks were integrated with Skyline v.3.5.0.9319 for small molecules[65], followed by quality control and normalization. Details of data acquisition and processing have been reported elsewhere[12,64]. Blank filters were extracted alongside the samples. Metabolites that did not pass quality control in more than 10% of samples were discarded; this is discussed further in Boysen et al.[12]. For metabolites that passed the quality control in 90% of samples but not all samples, the remaining samples were filled in with values to reflect the limit of detection for that metabolite.

Macromolecular measurements. Macromolecules were hydrolysed as in Fountoulakis and Lahm[66] with some modifications as follows: samples were heated at 120 °C for 20 h instead of 110 °C for 20–24 h since initial recovery tests with BSA resulted in better recovery of the amino acids at 120 °C compared to 110 °C or a shorter hydrolysis with BSA at 150 °C. Punches of 142 mm 0.2 µm Durapore filters were transferred into acid (10% hydrochloric) and solvent (water, methanol, dichloromethane) cleaned 40 ml polytetrafluoroethylene centrifuge tubes. Enough 6N hydrochloric acid was added to cover the filter along with spikes of isotope-labelled amino acid and nucleobase standards. Each sample was purged under nitrogen gas for 30 s before immediately being sealed with a solvent-rinsed cap. Samples were heated at 120 °C for 20 h. The acid was then transferred to a clean, combusted glass vial. The original polytetrafluoroethylene vial and filter were rinsed with approximately 500 µl of Optima grade water and transferred to the new glass vial. A rinsing step was repeated with an equal volume of Optima grade methanol. The acid mixture was concentrated to dryness under nitrogen gas and on a heat block set to medium heat. Once dried, approximately 500 µl of water was used to rinse each vial and samples were returned to dry completely under nitrogen gas. Dried samples were redissolved in 1 ml of Optima grade water and syringe filtered into liquid chromatography–mass spectrometry (LC–MS) vials.

LC–MS analysis of nucleobases and amino acids used a SeQuant ZIC-pHILIC column (5 µm particle size, 2.1 × 150 mm; Merck Millipore) with 10 mM ammonium carbonate in 85:15 water to acetonitrile (solvent A) and 10 mM ammonium carbonate in 85:15 acetonitrile to water (solvent B) at a flow rate of 0.15 ml min^{-1}. The column was held at 100% B for 2 min, ramped to 64% A over 18 min, ramped up to 100% A over 1 min, held at 100% A for 7 min and equilibrated at 100% B for 22 min (total time was 50 min). The column was maintained at 30 °C. Compounds were detected on a Q-Exactive HF with a full scan method employing positive and negative switching, a scan range of 60–900 *m/z* and a resolution of 60,000. Capillary temperature was 320 °C, H-electrospray ionisation spray voltage was 3.5 kV and the auxiliary gas heater temperature was 90 °C. The S-lens radio frequency level was 65. Sheath, auxiliary and sweep gas flow rates were maintained at 16, 3 and 1, respectively.

Lipidome. Detailed methods for lipidomics are described in Becker et al.[6]. Briefly: lipids were extracted from triplicate samples using a modified Bligh and Dyer protocol[67]. The total lipid extract was analysed on an Agilent 1200 high-performance liquid chromatography (HPLC) system coupled to a Thermo Fisher Scientific Q-Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer equipped with an electrospray ion source. Analyte separation was achieved using reversed-phase HPLC on a C8 Xbridge column (particle size 5 µm, length 150 mm, width 2.1 mm; Waters Corporation). The HPLC and MS conditions are described in Collins et al.[68] (modified after Hummel et al.[69]). To identify and quantify lipids, we used LOBSTAHS v.1.0.0, an open-source lipidomics software workflow based on adduct ion abundances and several other orthogonal criteria[68]. Lipids identified using the LOBSTAHS software were quantified from MS data after preprocessing with XCMS[70] and CAMERA[71] and corrected for response factors of commercially available standards as described in Becker et al.[6] and Biller et al.[72]. As a means of validating the accuracy and reliability of LOBSTAHS identification and quantification, quality control samples of known composition were interspersed with the environmental samples.

The >0.2 µm metatranscriptomes. The >0.2 µm transcriptomes were collected, sequenced, quality-controlled, quantified and normalized, as described previously[9,73]. Briefly, 2 l of seawater was filtered onto 25 mm, 0.2 µm Supor PES Membrane Disc Filters (Pall Corporation) using a peristaltic pump. The filtration time was between 15 and 20 min and filters were placed immediately in RNALater (Thermo Fisher Scientific) and preserved at −80 °C until processing. RNA extractions were performed by removing RNALater followed by adding 300 µl of Ambion denaturing solution directly to the filter and then vortexing for 1 min. Next, 750 µl of nuclease-free water was added and the samples were robotically purified and DNase-treated using the Chemagen chemagic MSM I instrument with the tissue RNA CMG-1212A Kit (PerkinElmer). RNA quality was assessed using Fragment Analyzer high-sensitivity reagents (Advanced Analytical Technologies) and quantified using RiboGreen (Invitrogen).

Internal standard RNA mixtures used for quantitative transcriptomics were prepared as described elsewhere[9,73]. Internal standards were added at known concentrations to samples before sequencing by a NextSeq 500 system. Generated reads were trimmed of adaptor sequences with Trimmomatic v.0.27 (ref. [74]), end-joined with PANDAseq v.2.4 (ref. [75]) and filtered for quality using Sickle v.1.33 (ref. [76]). Reads containing ribosomal RNA sequences were removed in silico using SortMeRNA v.2.1 (ref. [77]). Spiked-in RNA internal standard sequences were identified using lastal v.756 (ref. [78]), quantified and then removed. The remaining transcript reads were mapped to the merged HOE Legacy II-ALOHA metagenomic gene catalogue using lastal[9,73]. For each time point, the average normalization coefficient (derived from five different internal standards) was multiplied by the reads mapped to each transcript to estimate transcripts per litre for each gene in the sample. This normalized transcript count table, and HOE Legacy II-ALOHA metagenomic gene catalogue annotations[9,36], were used in subsequent bioinformatics analyses of the >0.2 µm sample transcripts.

The >5 µm metatranscriptomes. The >5 µm metatranscriptomes were collected, sequenced and quality-controlled as described elsewhere[4]. Briefly, 20 l of water was filtered onto two 5 µm 47 mm polycarbonate filters with a peristaltic pump, passing approximately 10 l through each filter. Filtering time did not exceed 40 min, after which filters were placed into liquid nitrogen until processing. RNA extractions were performed using a QIAGEN RNeasy Mini Kit, modifying the lysis step by adding BioSpec 0.5 zirconia/silica beads. For each filter set ($n = 2$, representing 20 l of sample volume), lysis buffer and beads were added, vortexed for 1 min, placed on ice for 30 s and vortexed again for 1 min. Lysate was removed with a pipette and pooled into a single 5 ml microcentrifuge tube. The rest of the RNeasy Mini Kit protocol was then followed according to the manufacturer's instructions, adjusting volumes accordingly and incorporating the on-column DNase digestion step, using a QIAGEN RNase-free DNase kit. The resulting total RNA was eluted with RNase-free water and then purified and concentrated with a RNeasy MinElute Kit according to the manufacturer's instructions. The quantity and quality of the extracted total RNA were assessed on an Agilent 2100 Bioanalyzer. Illumina TruSeq libraries were prepared at the JP Sulzberger Columbia Genome Center according to centre protocols and sequenced on an Illumina HiSeq 2000 to produce 90 million 100 base pairs, paired-end poly(A) selected reads. Raw sequence quality was visualized and then cleaned and trimmed as described previously[4].

Mapping of the >5 µm metatranscriptome was conducted using the Burrows–Wheeler Aligner MEM (parameters --k 10 --aM; ref. [79]) against a reference database constructed from Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) transcriptomes after Alexander et al.[80]. The resulting alignments were counted using the HTSeq v.0.6.1 package (options --a 0, --m intersection-strict, --s no; ref. [81]). Read counts were then filtered for contigs with average read counts ≥10 across the time series; then, DESeq2 v.1.34.0's variance stabilizing normalization was implemented on the remaining data[82]. KEGG orthologues were assigned with UProC v.1.2.0 (ref. [83]) and putative taxonomic assignments at the phylum level were assigned from the MMETSP taxon designations[37]. KEGG orthology annotations link genes across diverse organisms, sets of genes comprising metabolic pathways, and in some instances, connections between transcripts for genes involved in the synthesis/transformation of a particular molecule and the concentration of that molecule, with the caveat that KEGG orthology annotations are incomplete and some reactions affecting molecular pools may be missed. Read counts for each KEGG orthologue were then summed over the genes assigned to each taxon, resulting in phylum-level signals.

Taxonomic resolution selection. For the >5 µm transcripts, phyla were selected as the level of taxonomic resolution to compromise between clarity in the overall features of the data and inclusion of the greatest number of transcriptional signals. We analysed 14 different eukaryotic phyla, as defined in the taxonomy provided by the MMETSP database. While some phyla are dominated by photosynthetic genera (for example, the Bacillariophyta), some phyla potentially include signals from mixotrophic or heterotrophic genera (for example, the Dinophyta) and so we conservatively use the term eukaryote throughout. In the >0.2 µm fraction because prokaryotes have fewer genes, the compromise between taxonomic resolution and data interpretability was less severe; thus, specific taxa of known importance[84,85] and interest were highlighted, such as *Prochlorococcus* (high light and low light ecotypes), *Synechococcus*, SAR11 and *Crocosphaera*, while other taxonomic groups were left at the phylum level (such as Actinobacteria and Planctomycetes). A complete list of taxa examined and the distribution of their diel transcriptional signals across clusters is available in Supplementary Data 5.

Determination of diel periodicity. For all datasets, diel periodicity was determined using the rank-based Jonckheere–Terpstra umbrella test implemented in the R RAIN package v.1.26.0 (ref. [39]). Data were first detrended (the linear regression with respect to time was subtracted from the time series) to increase power of rhythmicity detection using the detrend function in the R pracma

package v.2.3.3 (ref. [86]); after RAIN implementation, the Benjamini–Hochberg false discovery rate (FDR) control procedure was implemented to assess significance at the $P=0.05$ level for each data type, considering datasets separately because time series between data types were not fully overlapping. For the metabolite data, which was measured in triplicate, observations across all replicates were used to determine periodicity, while the averages across replicates for each time point were used for the clustering analysis. Only significantly diel signals were retained for further analysis.

Clustering analysis. Detrended diel time series were scaled to make data dimensionless and reduce the impact of magnitude on the construction of Euclidean distance matrices, emphasizing the shape of the periodic element of the time series over other features. The Hopkins statistic was calculated for this distance matrix to assess the meaningfulness of clustering; a value of $h=0.79$ was found, which indicates structure in the data that cannot be explained by a random distribution of distances. To determine a well-fitting clustering method, hierarchical clustering (implemented by the hclust function in the R stats package v.4.1.1), partitioning about medoid clustering (a version of *k*-nearest neighbour clustering more robust to outliers, calculated using the clara function from the R cluster package v.2.1.2; ref. [87]) and training of SOMs (using the R Kohonen package v.3.0.10; ref. [88]) were implemented. To evaluate the fit of each of these results, the Calinski–Harabasz metric (using the calinhara function from the R library fpc package v.2.2-9; ref. [89]) and average silhouette distance (using the silhouette function from the R cluster package[87]) were calculated and compared (Extended Data Fig. 1). On the basis of higher average silhouette width and Calinski–Harabasz score for all potential clusterings, SOM was selected as the clustering method for the data. We used the heuristic of identifying the 'elbow' in decreasing average silhouette width to initially select four as the operational number of clusters. To inspect the fits for more detail, we calculated the per-cluster average number of time series with negative silhouette widths (interpreted as misclassifications) for three, four and five clusters. We found that four clusters had the fewest average misclassifications per cluster. To further compare these potential clusterings, we generated ordered dissimilarity images and silhouette profiles for the four and five cluster cases. Silhouette width profiles were then constructed for each cluster for more detailed inspection of cluster coherence and four clusters were selected as the optimal number for these data.

Calculation of peak rank time. To estimate mean peak time for transcriptional profiles, a rank-based heuristic was calculated. For a given transcript or biomolecule, the expression levels at each measurement time point were ranked (1–20). The ranks from all the 2:00 measurements, 6:00 measurements and so on were averaged and the peak mean rank time was defined as the time with the highest average, where ties were summarized as the centre between tied times (for example, if a transcript had the same mean rank for the 2:00 and 6:00 measurements, the mean peak rank would be defined as 4:00). All peak rank time estimates along with original and rotated NMDS coordinates are provided in Supplementary Data 7. For figure plotting, peak rank times that were ties were rounded down (for example, a 4:00 mean peak rank time would be rounded down to a 2:00 mean peak rank time).

Assessing average peak rank time difference. For all KEGG orthologues with diel expression in at least four different taxonomic groups, we found all circular pairwise differences in peak times based on the peak times. For example, if an orthologue had a 6:00 peak time for one taxon and a 14:00 peak time for another taxon, the pairwise difference would be 8 h. The average of all of these pairwise differences was taken for each KEGG orthologue individually, across all taxa that had diel expression of that KEGG orthologue. For example, K02902 (rpmB ribosomal subunit L28) had diel transcript abundance for high light *Prochloroccous*, Bacillariophyta, *Crocosphaera* and Haptophyta. As such, we averaged the differences in peak time in K02902 transcript abundance between all potential pairwise combinations: high light *Prochloroccous* and Bacillariophyta; high light *Prochloroccous* and *Crocosphaera*; high light *Prochloroccous* and Haptophyta; Bacillariophyta and *Crocosphaera*; and *Crocosphaera* and Haptophyta. To assess whether or not an orthologue had significantly small average peak-time difference (that is, taxa tended to peak expression of this orthologue at the same time of day), we repeated a Monte Carlo simulation 10,000 times, using multinomial draws of size equal to the number of taxa with diel expression of the orthologue from the population of all observed diel transcripts. In the above example, we picked four random diel transcripts, calculated the average of all pairwise peak-time differences and repeated that process 10,000 times with new random draws. We used a bootstrap approach and selected transcripts from the data because the distribution of peak times was not uniform. The empirical distribution of average peak-time difference was then used as a null to calculate the simulated *P* value of each orthologue. Significance was thresholded using the Benjamini–Hochberg procedure at FDR 10% ($P<0.1$). See Extended Data Fig. 3 for examples showing the null distributions for the average peak-time difference between increasing numbers of taxa (histogram) and the test statistic values of all transcripts observed to have diel transcript abundance in that number of taxa (box plot with significantly synchronous KEGG orthologues coloured).

Pathway enrichment analysis. The 5,722 transcripts identified as significantly diel were assigned to 4,193 unique KEGG orthologues. Pathways were manually curated by inspection to eliminate redundant, ambiguous or otherwise inappropriate assignments. This resulted in 258 unique pathway assignments among 3,097 unique KEGG orthologues; the remaining 1,096 KEGG orthologues could not be unambiguously unassigned to a pathway (Supplementary Data 8). Diel transcripts with an assigned KEGG orthologue were also mapped to taxonomic classifiers—eukaryotes, cyanobacteria (photoautotrophic bacteria) and non-eukaryotic heterotrophs (nearly exclusively bacterial; Supplementary Data 8). Enrichment analysis was performed for each KEGG pathway using Fisher's exact test. The Benjamini–Hochberg adaptive FDR control procedure was implemented using a significance threshold of FDR = 10% ($P<0.1$).

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

Data availability

Sequence data for the >0.2 µm metatranscriptome have been deposited in the Sequence Read Archive through the National Center for Biotechnology Information under BioProject ID PRJNA358725. The Station ALOHA gene catalogue data are available under BioProject ID PRJNA352737 and iMicrobe (https://www.imicrobe.us/#/search/PRJNA352737). Sequence data for the >5 µm metatranscriptomes are available at the Sequence Read Archive under accession no. SRP136571 and BioProject no. PRJNA437978. Raw files for the metabolomics data are available at Metabolomics Workbench under Project ID PR000926. The lipidomics mass spectral raw data are available from the authors upon request.

Code availability

All code and feature/abundance tables used to complete this analysis are available at GitHub (https://github.com/WeitzGroup/community_scale_metabolism_NPSG) and are archived under https://zenodo.org/badge/latestdoi/262179139 (ref. [90]).

Received: 31 July 2020; Accepted: 1 November 2021;
Published online: 20 January 2022

References

1. Ottesen, E. A. et al. Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl Acad. Sci. USA* **110**, E488–E497 (2013).
2. Muñoz-Marín, M. D. C. et al. The transcriptional cycle is suited to daytime N_2 fixation in the unicellular cyanobacterium "*Candidatus* Atelocyanobacterium thalassa" (UCYN-A). *mBio* **10**, e02495-18 (2019).
3. Vislova, A., Sosa, O. A., Eppley, J. M., Romano, A. E. & DeLong, E. F. Diel oscillation of microbial gene transcripts declines with depth in oligotrophic ocean waters. *Front. Microbiol.* **10**, 2191 (2019).
4. Harke, M. J. et al. Periodic and coordinated gene expression between a diazotroph and its diatom host. *ISME J.* **13**, 118–131 (2019).
5. Hernández Limón, M. D. et al. Transcriptional patterns of *Emiliania huxleyi* in the North Pacific Subtropical Gyre reveal the daily rhythms of its metabolic potential.*Environ. Microbiol.* **22**, 381–396 (2020).
6. Becker, K. W. et al. Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean. *Nat. Commun.* **9**, 5179 (2018).
7. Frischkorn, K. R., Haley, S. T. & Dyhrman, S. T. Coordinated gene expression between *Trichodesmium* and its microbiome over day–night cycles in the North Pacific Subtropical Gyre. *ISME J.* **12**, 997–1007 (2018).
8. Ottesen, E. A. et al. Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345**, 207–212 (2014).
9. Wilson, S. T. et al. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocosphaera*. *Nat. Microbiol.* **2**, 17118 (2017).
10. Saito, M. A. et al. Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocosphaera watsonii*. *Proc. Natl Acad. Sci. USA* **108**, 2184–2189 (2011).
11. Strenkert, D. et al. Multiomics resolution of molecular events during a day in the life of *Chlamydomonas*. *Proc. Natl Acad. Sci. USA* **116**, 2374–2383 (2019).
12. Boysen, A. K. et al. Particulate metabolites and transcripts reflect diel oscillations of microbial activity in the surface ocean. *mSystems* **6**, e00896-20 (2021).
13. White, A. E., Barone, B., Letelier, R. M. & Karl, D. M. Productivity diagnosed from the diel cycle of particulate carbon in the North Pacific Subtropical Gyre: optically derived productivity. *Geophys. Res. Lett.* **44**, 3752–3760 (2017).
14. DeLong, E. F. et al. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496–503 (2006).
15. Sunagawa, S. et al. Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).

16. Coles, V. J. et al. Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* **358**, 1149–1154 (2017).
17. Walbauer, J. R., Rodrigue, S., Coleman, M. L. & Chisholm, S. W. Transcriptome and proteome dynamics of a light–dark synchronized bacterial cell cycle.*PLoS ONE* **7**, e43432 (2012).
18. Steiner, P. A. et al. Highly variable mRNA half-life time within marine bacterial taxa and functional genes. *Environ. Microbiol.* **21**, 3873–3884 (2019).
19. Moran, M. A. et al. Sizing up metatranscriptomics. *ISME J.* **7**, 237–243 (2013).
20. Tamames, J., Cobo-Simón, M. & Puente-Sánchez, F. Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC Genomics* **20**, 960 (2019).
21. DiTullio, G. R. & Laws, E. A. Diel periodicity of nitrogen and carbon assimilation in five species of marine phytoplankton: accuracy of methodology for predicting N-assimilation rates and N/C composition ratios. *Mar. Ecol. Prog. Ser.* **32**, 123–132 (1986).
22. Granum, E., Kirkvold, S. & Myklestad, S. M. Cellular and extracellular production of carbohydrates and amino acids by the marine diatom *Skeletonema costatum*: diel variations and effects of N depletion. *Mar. Ecol. Prog. Ser.* **242**, 83–94 (2002).
23. Lacour, T., Sciandra, A., Talec, A., Mayzaud, P. & Bernard, O. Diel variations of carbohydrates and neutral lipids in nitrogen-sufficient and nitrogen-starved cyclostat cultures of *Isochrysis* sp. *J. Phycol.* **48**, 966–975 (2012).
24. Follett, C. L., Dutkiewicz, S., Karl, D. M., Inomura, K. & Follows, M. J. Seasonal resource conditions favor a summertime increase in North Pacific diatom–diazotroph associations. *ISME J.* **12**, 1543–1557 (2018).
25. Chen, W.-N. U. et al. Diel rhythmicity of lipid-body formation in a coral-*Symbiodinium* endosymbiosis. *Coral Reefs* **31**, 521–534 (2012).
26. Zhou, X. & Mopper, K. Photochemical production of low-molecular-weight carbonyl compounds in seawater and surface microlayer and their air-sea exchange. *Mar. Chem.* **56**, 201–213 (1997).
27. Durham, B. P. et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean.*Nat. Microbiol.* **4**, 1706–1715 (2019).
28. Lambert, S. et al. Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. *ISME J.* **13**, 388–401 (2019).
29. 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).
30. Aylward, F. O. et al. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc. Natl Acad. Sci. USA* **112**, 5443–5448 (2015).
31. Rusch, D. B. et al. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
32. Bork, P. et al. Tara Oceans studies plankton at planetary scale. *Science* **348**, 873 (2015).
33. Delmont, T. O. et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813 (2018).
34. Fuhrman, J. A. et al. Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* **103**, 13104–13109 (2006).
35. Morris, R. M. et al. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. *Limnol. Oceanogr.* **50**, 1687–1696 (2005).
36. Mende, D. R. et al. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* **2**, 1367–1373 (2017).
37. 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).
38. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
39. Thaben, P. F. & Westermark, P. O. Detecting rhythms in time series with RAIN. *J. Biol. Rhythms* **29**, 391–400 (2014).
40. Cuhel, R. L., Ortner, P. B. & Lean, D. R. S. Night synthesis of protein by algae. *Limnol. Oceanogr.* **29**, 731–744 (1984).
41. Coesel, S. N. et al. Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities. *Proc. Natl Acad. Sci. USA* **118**, e2011038118 (2021).
42. Bolay, P., Muro-Pastor, M. I., Florencio, F. J. & Klähn, S. The distinctive regulation of cyanobacterial glutamine synthetase. *Life (Basel)* **8**, 52 (2018).
43. Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M. & Mahaffey, C. Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proc. Natl Acad. Sci. USA* **109**, 1842–1849 (2012).
44. Berman, T. & Bronk, D. A. Dissolved organic nitrogen: a dynamic participant in aquatic ecosystems. *Aquat. Microb. Ecol.* **31**, 279–305 (2003).
45. Lee, C. & Bada, J. L. Amino acids in equatorial Pacific Ocean water. *Earth Planet. Sci. Lett.* **26**, 61–68 (1975).
46. Bada, J. L. & Lee, C. Decomposition and alteration of organic compounds dissolved in seawater. *Mar. Chem.* **5**, 523–534 (1977).
47. Poretsky, R. S., Sun, S., Mou, X. & Moran, M. A. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ. Microbiol.* **12**, 616–627 (2010).
48. Berthelot, H. et al. NanoSIMS single cell analyses reveal the contrasting nitrogen sources for small phytoplankton. *ISME J.* **13**, 651–662 (2019).
49. Moore, L. R., Post, A. F., Rocap, G. & Chisholm, S. W. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **47**, 989–996 (2002).
50. Hu, S. K., Connell, P. E., Mesrop, L. Y. & Caron, D. A. A hard day's night: diel shifts in microbial eukaryotic activity in the North Pacific Subtropical Gyre. *Front. Mar. Sci.* https://doi.org/10.3389/fmars.2018.00351 (2018).
51. Hannides, C. C. S., Popp, B. N., Choy, C. A. & Drazen, J. C. Midwater zooplankton and suspended particle dynamics in the North Pacific Subtropical Gyre: a stable isotope perspective. *Limnol. Oceanogr.* **58**, 1931–1946 (2013).
52. Becker, K. W. et al. Combined pigment and metatranscriptomic analysis reveals highly synchronized diel patterns of phenotypic light response across domains in the open oligotrophic ocean.*ISME J.* **15**, 520–533 (2021).
53. Mruwat, N. et al. A single-cell polony method reveals low levels of infected *Prochlorococcus* in oligotrophic waters despite high cyanophage abundances. *ISME J.* **15**, 41–54 (2021).
54. Chesson, P. L. & Warner, R. R. Environmental variability promotes coexistence in lottery competitive systems. *Am. Nat.* **117**, 923–943 (1981).
55. Shmida, A. & Ellner, S. Coexistence of plant species with similar niches. *Vegetatio* **58**, 29–55 (1984).
56. Ellner, S. P., Snyder, R. E. & Adler, P. B. How to quantify the temporal storage effect using simulations instead of math. *Ecol. Lett.* **19**, 1333–1342 (2016).
57. Adler, P. B., Fajardo, A., Kleinhesselink, A. R. & Kraft, N. J. B. Trait-based tests of coexistence mechanisms. *Ecol. Lett.* **16**, 1294–1306 (2013).
58. Adler, P. B., HilleRisLambers, J., Kyriakidis, P. C., Guan, Q. & Levine, J. M. Climate variability has a stabilizing effect on the coexistence of prairie grasses. *Proc. Natl Acad. Sci. USA* **103**, 12793–12798 (2006).
59. Cáceres, C. E. Temporal variation, dormancy, and coexistence: a field test of the storage effect. *Proc. Natl Acad. Sci. USA* **94**, 9171–9175 (1997).
60. Padisák, J. Identification of relevant time-scales in non-equilibrium community dynamics: conclusions from phytoplankton surveys. *N. Z. J. Ecol.* **18**, 169–176 (1994).
61. Anderies, J. M. & Beisner, B. E. Fluctuating environments and phytoplankton community structure: a stochastic model. *Am. Nat.***155**, 556–569 (2000).
62. Wagg, C. et al. Functional trait dissimilarity drives both species complementarity and competitive disparity. *Funct. Ecol.* **31**, 2320–2329 (2017).
63. Bligh, E.G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
64. Boysen, A. K., Heal, K. R., Carlson, L. T. & Ingalls, A. E. Best-matched internal standard normalization in liquid chromatography–mass spectrometry metabolomics applied to environmental samples. *Anal. Chem.* **90**, 1363–1369 (2018).
65. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
66. Fountoulakis, M. & Lahm, H. W. Hydrolysis and amino acid composition analysis of proteins. *J. Chromatogr. A* **826**, 109–134 (1998).
67. Popendorf, K. J., Fredricks, H. F. & Van Mooy, B. A. S. Molecular ion-independent quantification of polar glycerolipid classes in marine plankton using triple quadrupole MS. *Lipids* **48**, 185–195 (2013).
68. Collins, J. R., Edwards, B. R., Fredricks, H. F. & Van Mooy, B. A. S. LOBSTAHS: an adduct-based lipidomics strategy for discovery and identification of oxidative stress biomarkers. *Anal. Chem.* **88**, 7154–7162 (2016).
69. Hummel, J. et al. Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. *Front. Plant Sci.* **2**, 54 (2011).
70. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
71. Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R. & Neumann, S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* **84**, 283–289 (2012).
72. Biller, S. J. et al. *Prochlorococcus* extracellular vesicles: molecular composition and adsorption to diverse microbes.*Environ. Microbiol.* https://doi.org/10.1111/1462-2920.15834 (2021).
73. Aylward, F. O. et al. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl Acad. Sci. USA* **114**, 11446–11451 (2017).
74. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
75. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).

76. Joshi, N. & Fass, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. Version 1.33. *GitHub* https://github.com/najoshi/sickle (2015).
77. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
78. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
79. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
80. Alexander, H. et al. Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl Acad. Sci. USA* **112**, E5972–E5979 (2015).
81. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
82. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
83. Meinicke, P. UProC: tools for ultra-fast protein domain classification. *Bioinformatics* **31**, 1382–1388 (2015).
84. Mende, D. R., Boeuf, D. & DeLong, E. F. Persistent core populations shape the microbiome throughout the water column in the North Pacific Subtropical Gyre. *Front. Microbiol.* **10**, 2273 (2019).
85. White, A. E. et al. Phenology of particle size distributions and primary productivity in the North Pacific subtropical gyre (Station ALOHA). *J. Geophys. Res. Oceans* **120**, 7381–7399 (2015).
86. Borchers, H. W. pracma: Practical numerical math functions. R package version 2 https://cran.r-project.org/web/packages/pracma/index.html (2019).
87. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. cluster: Cluster analysis basics and extensions. R package version 1.56 (2012).
88. Wehrens, R. & Buydens, L. M. C. Self- and super-organizing maps in R: the Kohonen package. *J. Stat. Softw.* **21**, 1–19 (2007).
89. Hennig, C. fpc: Flexible procedures for clustering. R package version 2.2-9 (2010).
90. Muratore, D. Code for complex marine microbial communities partition metabolism of scarce resources over the diel cycle. *Zenodo* https://doi.org/10.5281/zenodo.3817416 (2020).

Acknowledgements

We thank T. Clemente and E. Wood-Charlson for facilitating data collection. We also thank the captains and crew of the R/V Kilo Moana and research staff at the School of Ocean and Earth Science and Technology and S. Haley and K. Frischkorn for sample collection of eukaryotic transcriptomes. This work was supported by grants from the Simons Foundation as part of the SCOPE collaboration: no. 329108 to E.V.A., E.F.D., D.M.K., A.E.W., J.P.Z., A.E.I., B.A.S.V.M., S.T.D. and J.S.W.; no. 721244 to E.V.A.; no. 721223 to E.F.D.; no. 721252 to D.M.K.; no. 721256 to A.E.W.; no. 724220 to J.P.Z.; no. 723787 to A.E.I.; no. 721229 to B.A.S.V.M.; no. 721225 to S.T.D.; and no. 721231 to J.S.W. A.K.B. was supported by a National Science Foundation Graduate Research Fellowship. K.W.B. was further supported by the Postdoctoral Scholarship Program at the Woods Hole Oceanographic Institution & US Geological Survey and J.R.C. was supported by the Simons Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems (Simons Foundation grant no. 549894).

Author contributions

A.K.B., F.O.A., A.V., A.E.W. and S.T.D. contributed to data collection. A.K.B., M.J.H., K.W.B., F.O.A., J.M.E., D.R.M., A.E.I., B.A.S.V.M. and S.T.D. contributed to sample processing and data preparation. S.T.W. served as chief scientist for the research expedition. D.M., A.K.B., M.J.H., K.W.B., J.R.C., S.N.C., D.R.M., S.T.D. and J.S.W. developed the data analysis methods. D.M., A.K.B. and S.N.C. wrote the code. D.M., S.N.C., S.J.B., S.P., R.A.R.-G., A.E.I. and J.S.W. contributed to analysis design. D.M., A.K.B., M.J.H., J.R.C., B.A.S.V.M., A.E.I., S.T.D. and J.S.W. analysed the data. E.F.D., A.E.I., B.A.S.V.M., S.T.D. and J.S.W. designed the research, with contributions from all authors. D.M. and J.S.W. led the writing of the manuscript. A.K.B., M.J.H., K.W.B., D.R.M. and S.J.B. contributed to writing the manuscript. D.M., A.K.B., M.J.H., K.W.B., J.R.C., S.N.C., F.O.A., A.V., D.R.M., S.T.W., S.J.B., E.V.A., E.F.D., D.M.K., A.E.W., J.P.Z., A.E.I., B.A.S.V.M., S.T.D. and J.S.W. edited 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/s41559-021-01606-w.

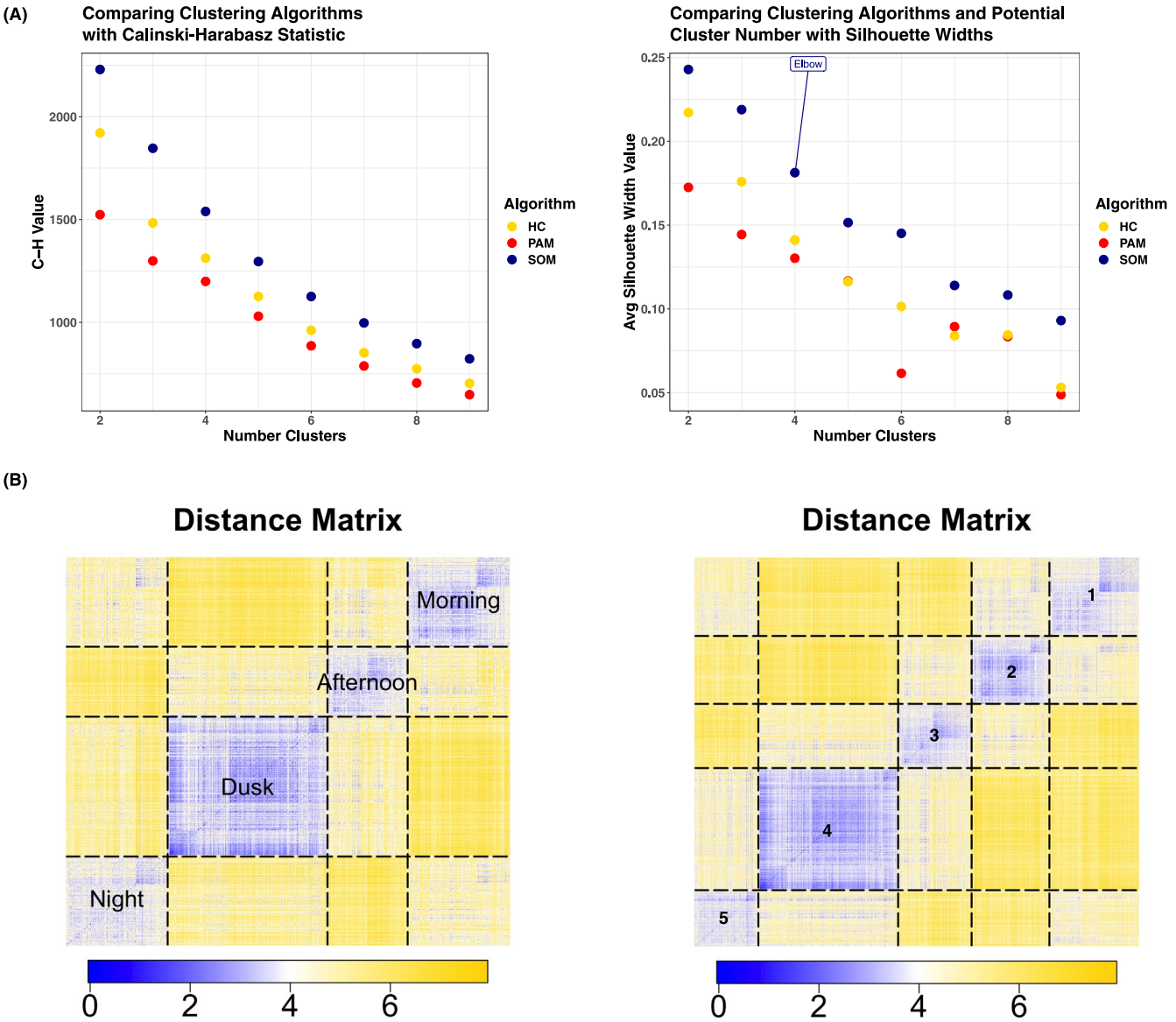
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-021-01606-w.

Correspondence and requests for materials should be addressed to Joshua S. Weitz.

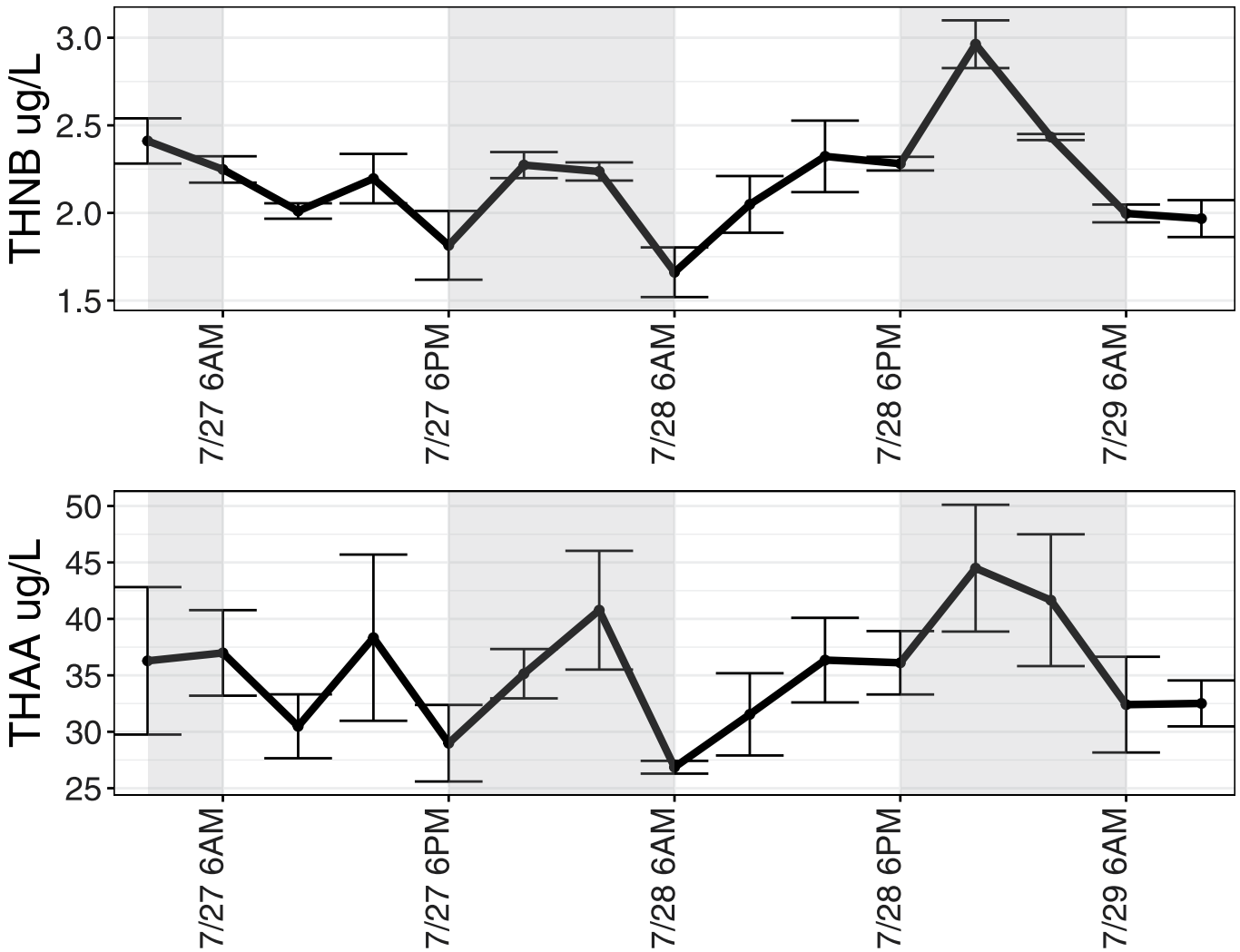
Peer review information *Nature Ecology & Evolution* thanks Daniel Garza and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

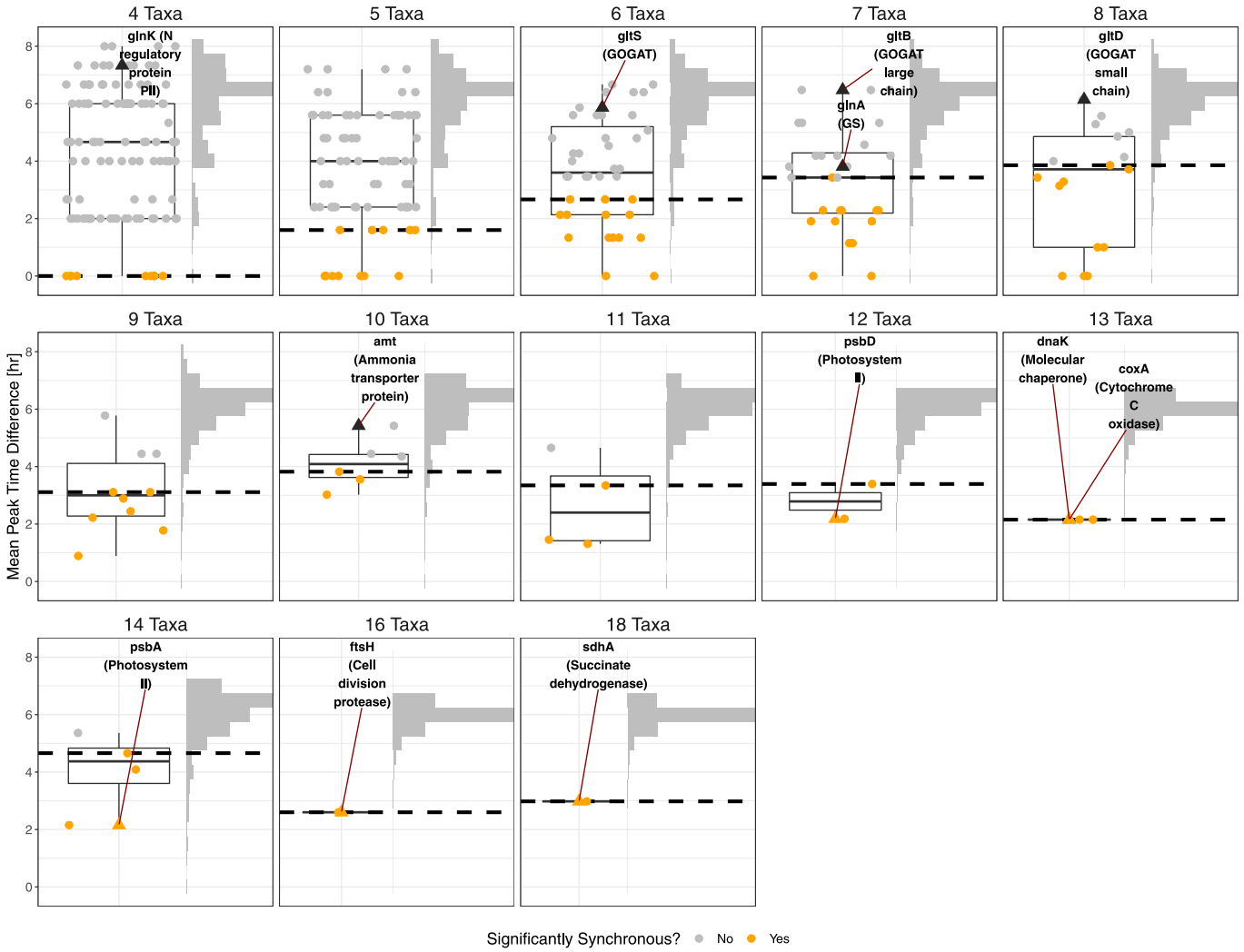
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



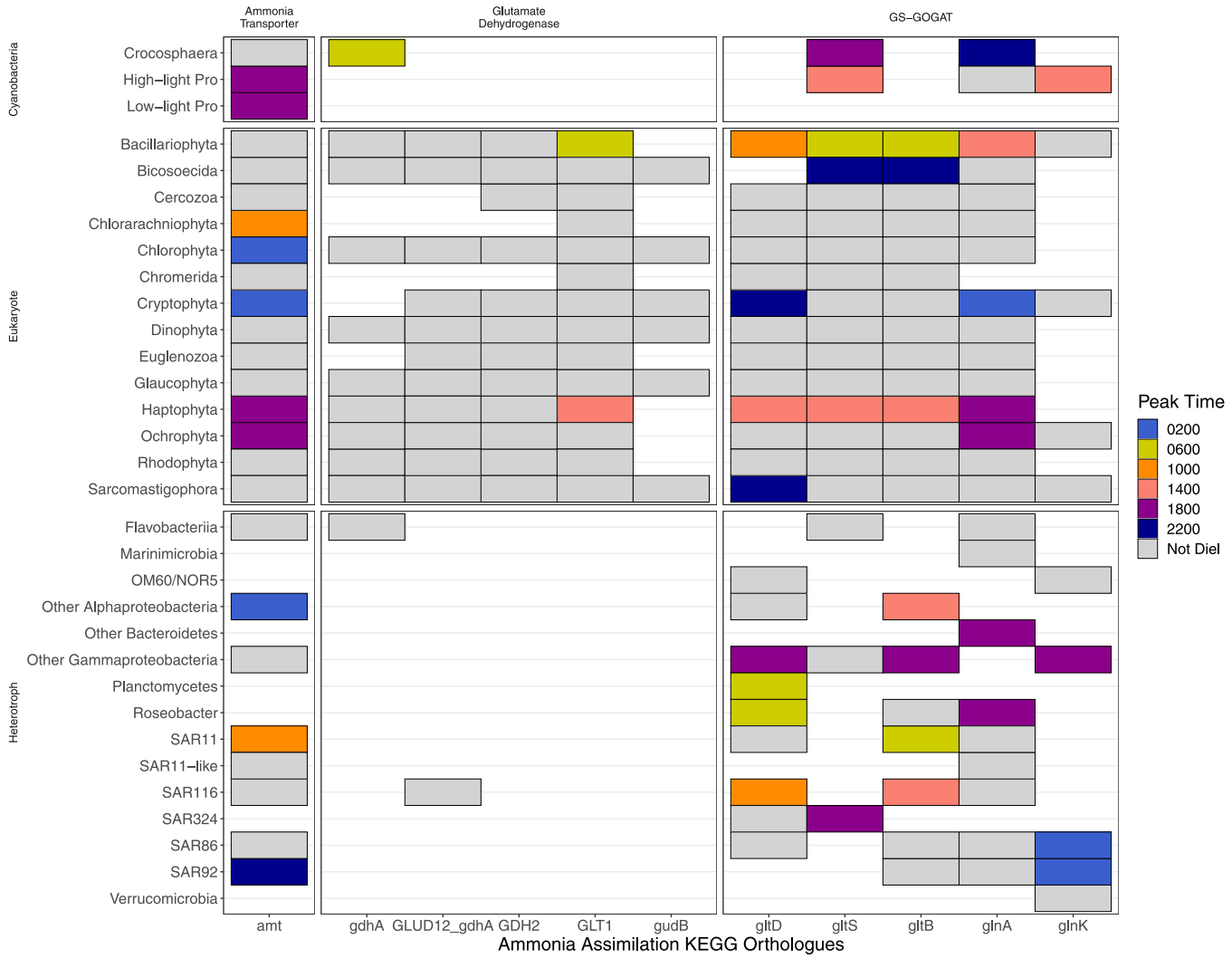
Extended Data Fig. 1 | Comparison of cluster metrics for archetype clustering. Comparison of cluster metrics for archetype clustering. Top panels show dynamics in clustering Calinski-Harabasz index and average silhouette width for increasing number of clusters comparing 3 clustering algorithms – self-organizing maps (SOM), hierarchical clustering (HC), and clustering about perimedoids (PAM). SOM was chosen for further clustering based on the advantage in C-H index and average silhouette width (A). The number of clusters was then selected based on the plateau in average silhouette width between four and five clusters. We then used ordered dissimilarity images (ODIs) to compare the four-cluster and five-cluster results (B). For additional information, silhouette profiles were constructed for all clusters in both clusterings. The 4 cluster clustering was chosen on the heuristic basis of higher maximum silhouette widths for all clusters in the 4 cluster SOM and fewer negative silhouette widths in all clusters (using negative silhouette width as a proxy for misclassification). Summary statistics for silhouette profiles are provided in Supplementary Data Item 6. Briefly, the SOM using 4 clusters had 235/6273 (3.7%) of silhouette widths less than 0 and maximum per-cluster silhouette widths between 0.322–0.456, while the SOM using 5 clusters had 353/6273 (5.6%) of silhouette widths less than 0 and maximum per-cluster silhouette widths between 0.265–0.419, indicating fewer misclassifications in the 4 cluster SOM and greater within-cluster similarity.



Extended Data Fig. 2 | Time series of hydrolysable nitrogenous bases (top) and amino acids (bottom). Details on quantification are available in Methods. Samples were measured in technical triplicates, with amino acids being separately measured in a biological duplicate. Time series shows mean total concentration with one standard deviation. Both THNB and THAA were detrended and RAIN analysis was conducted for 24-hour diel periodicity using replicates ($p < 1e\text{-}6, p = 0.00034$, respectively). These data were measured and incorporated into our analysis after the initial SOM clustering was conducted, so the detrended and z-score transformed time series were fit to the existing SOM model, and both THNB and THAA were assigned to the night cluster.



Extended Data Fig. 3 | Sampling distribution for taxon synchrony hypothesis test. As described in Methods section "Assessing Average Peak Time Rank Difference," for each KEGG orthologue with diel transcript abundance from at least 4 taxa, the average of all pairwise differences in peak time was calculated (points and boxplot to the left on each panel). To assess the significance of the calculated difference in peak time, empirical peak time distributions were calculated by randomly selecting groups of diel transcripts using a fixed number of taxa. We calculated the average of pairwise differences for these randomly selected transcripts and repeated the process 10,000 times for each hypothesis test conducted. The aggregate of all testing distributions is shown as a histogram on the right-hand side of each figure. A one-sided test of difference was conducted to identify those KEGG orthologues with a *lower* average peak time difference than could be expected by randomly selecting groups of diel orthologues of that size from the data (synchronous). Multiple testing corrections were conducted using the adaptive Benjamini-Hochberg method for a false discovery rate (FDR) of 0.1. Orthologues deemed significantly synchronous after this procedure are marked in orange. The dashed line shows the corresponding portion of the tail in the distribution where the highest average peak time difference was still deemed significant. The transcripts marked in main text Fig. 5a are also labeled.



Extended Data Fig. 4 | Diel asynchrony amongst ammonia-assimilation genes. All taxa were examined for the presence/absence of transcripts for the *amt* ammonia transporter, and all KEGG orthologues from the nitrogen metabolism/arginine biosynthesis pathways associated with reactions involving ammonia, glutamine, and glutamate (as well as the *glnK* PII regulatory gene). Transcripts that were present in the data but not identified to have diel changes in abundance are shown in gray. Colored blocks indicate taxa that did have diel transcript abundance for that transcript, and color indicates the associated peak time. All peak times were represented at least once in this set of diel transcripts associated with ammonia-assimilation genes. Note that taxa in Extended Data Figs. 4 and 5 are not equivalent; for this figure taxa are selected based on presence of ammonia-assimilation genes, irrespective of diel status.

Extended Data Fig. 5 | Diel synchrony amongst photosynthesis-related genes. All taxa were examined for the presence/absence of transcripts for photosystem subunits as defined by the KEGG photosynthesis pathway (photosystem I and photosystem II D1, D2, cp43, cp47, and cytochrome b559 shown). Transcripts that were present in the data but not identified to have diel changes in abundance are shown in gray. Colored blocks indicate taxa that did have diel transcript abundance for that transcript, and color indicates the associated peak time. Peak times for diel transcripts amongst photosynthesis-related genes were only found in the daytime (0600, 1000, and 1400). Note that taxa in Extended Data Figs. 4 and 5 are not equivalent; for this figure taxa are selected based on presence of photosynthesis-related genes, irrespective of diel status.

nature portfolio

Corresponding author(s): Joshua S. Weitz

Last updated by author(s): Oct 7, 2021

Reporting Summary

Nature Portfolio 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 Portfolio 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	The water-column sampling was achieved using a 24 x 12 L Niskin bottle rosette attached to a conductivity-temperature-depth (CTD) package (SBE 911Plus, SeaBird) with additional fluorescence, oxygen, and transmissometer sensors. Metabolites were measured with a Waters Xevo TQ-S triple quadrupole and a Thermo Scientific Q-Exactive Orbitrap HF with both reversed-phase and hydrophilic interaction liquid chromatography (HILIC). Metabolite peaks were integrated with Skyline for small molecules 69, followed by quality control and normalization. Details of the data acquisition and processing have been previously reported in Boyesen et al 2020. The total lipid extract was analyzed on an Agilent 1200 high performance liquid chromatography (HPLC) system coupled to a ThermoFisher Exactive Plus Orbitrap high resolution mass spectrometer (HRMS; ThermoFisher, Waltham, MA, USA) equipped with an electrospray ion source. For the identification and quantification of lipids, we used LOBSTAHS, an open-source lipidomics software workflow based on adduct ion abundances and several other orthogonal criteria. Lipids identified using the LOBSTAHS software were quantified from MS data after pre-processing with XCMS and CAMERA. For the >0.2 micron transcriptome, generated reads were trimmed of adapter sequences with Trimmomatic v 0.27, end-joined with PandaSeq v2.4, and filtered for quality using Sickle v1.33. Reads containing ribosomal RNA sequences were removed in silico using sortmerna v2.1. Spiked-in RNA internal standard sequences were identified using lastal v756, quantified, and then removed. For the >5 micron transcriptome, mapping was conducted using the Burrows–Wheeler Aligner (BWA-MEM, parameters –k 10 –aM;) against a reference database constructed from MMETSP transcriptomes. Resulting alignments were counted using the HTSeq 0.6.1 package (options –a 0, --m intersection-strict, -s no). Read counts were then filtered for contigs with average read counts ≥ 10 across the time series and then DESeq2's variance stabilizing normalization was implemented on remaining data. KEGG Orthologs were assigned with UProC and putative taxonomic assignments at the phylum level were assigned from MMETSP taxon designations.
Data analysis	All code and feature/abundance tables used to complete this analysis are available at https://github.com/WeitzGroup/community_scale_metabolism_NPSG archived under https://zenodo.org/badge/latestdoi/262179139. This analysis was conducted using R version 3.6, for the versions of specific packages implemented please see Methods.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequence data for the >0.2 µm metatranscriptome are deposited in the Sequence Read Archive through the National Center for Biotechnology Information under BioProject ID PRJNA358725. The Station ALOHA gene catalogue data are available under Bioproject no. PRJNA352737, and iMicrobe (http://datacommons.cyverse.org/browse/iplant/home/shared/imicrobe/projects/263/ALOHAgene cat_v1_nonredundant.annot). Sequence data for the >5 µm metatranscriptomes are available in the Sequence Read Archive through the National Center for Biotechnology Information under accession no. SRP136571, BioProject no. PRJNA437978. Raw files for metabolomics data are available in Metabolomics Workbench under project ID PR000797. Processed data will be available soon in Metabolomics Workbench under Project ID PR000926. Lipidomics mass spectral raw data are available from authors upon request. All code and feature/abundance tables used to complete this analysis are available at https://github.com/WeitzGroup/community_scale_metabolism_NPSG archived under https://zenodo.org/badge/latestdoi/262179139.

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	This study aimed to monitor microbial ecological activity over the diel cycle via repeated sampling of surface waters (15m) on a Lagrangian cruise track over the course of several days in the North Pacific Subtropical Gyre. Sampling was conducted via Niskin bottle rosette deployment every 4 hours and seawater samples were accordingly handled for >5 micron transcriptome sequencing, >0.2 micron transcriptome sequencing, or metabolite/lipid/macromolecular extraction and mass spectrometry.
Research sample	The population intended to be studied by this sample is that of the microorganisms living in an eddy in the North Pacific Subtropical Gyre in the sunlit open ocean surface. The samples we observed were extracted via niskin bottle sampling from an anticyclonic eddy in the NPSG at a depth of 15m taken at 6 times of day - 0200 hrs, 0600 hrs (corresponding to sunrise), 1000 hrs, 1400 hrs, 1800 hrs (corresponding to sunset), and 2200 hrs. Using serial filtration for transcriptomics, we intended to separate transcripts belonging to larger eukaryotic phytoplankton from smaller cyanobacteria and heterotrophic bacteria/archaea, while chemical measurements were filtered only once to attempt to catch the bulk community's intracellular concentrations of molecules of interest.
Sampling strategy	For specifics on sampling strategy, please refer to Wilson et al 2017, Becker et al 2018, Harke et al 2018, and Boysen et al 2020 in our references for sampling procedures corresponding to the >0.2 micron transcriptome, lipidome, >5 micron transcriptome, and metabolome respectively.
Data collection	For specifics on data collection, please refer to Wilson et al 2017, Becker et al 2018, Harke et al 2018, and Boysen et al 2020 in our references for sampling procedures corresponding to the >0.2 micron transcriptome, lipidome, >5 micron transcriptome, and metabolome respectively.
Timing and spatial scale	Water-column seawater sampling for diel measurements took place every 4 h for a period of 4 days (26-30 July) and 3 days (31 July-3 August) at a depth of 15 m corresponding to the depth of a drogue used to identify the Lagrangian track. The water-column sampling was achieved using a 24 x 12 L Niskin bottle rosette attached to a conductivity-temperature-depth (CTD) package (SBE 911Plus, SeaBird) with additional fluorescence, oxygen, and transmissometer sensors. The sampling and analytical protocols for vertical profiles of nutrients, particulates, and flow-cytometry enumerated phytoplankton populations and heterotrophic bacteria were identical to those employed by the Hawaii Ocean Time-series program (http://hahana.soest.hawaii.edu/index.html).
Data exclusions	Data collected from the second 3-day long diel sampling circuit were excluded from this analysis for several reasons. Partly, the gap between the two sampling circuits for the purpose of comparative analysis constitutes data missing not at random and our analysis lacked the inferential capability to account for those missing timepoints. Furthermore, a change in wind speed between the sampling periods resulted in mixed layer shoaling causing observable ecological changes such as in the abundances of Crocosphaera observed in Wilson et al 2017. This led us to believe that the two diel sampling periods represented different ecological regimes and were not intercomparable for the purposes of a systematic analysis.
Reproducibility	All data collected for this study were observational and no experiment was conducted.
Randomization	Randomization was not relevant to our study because data collected were observational and no treatment was applied.

Blinding	Blinding was not relevant to our study because data collected were observational and no treatment was applied.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	See description in 'Location'.
Location	Fieldwork was conducted during 25 July to 5 August, 2015 in the oligotrophic North Pacific Subtropical Gyre. To maximize the signal to noise ratio in an open ocean environment, a Lagrangian sampling strategy was implemented whereby World Ocean Circulation Experiment Surface Velocity Profile (WOCE SVP) drifters from Pacific Gyre, Inc. were deployed within the center of a mesoscale anticyclonic eddy. The mesoscale eddy fields were identified using Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO) and when the field sampling occurred, the target anticyclonic eddy was located north of the Hawaiian Islands at 24.4 N and 156.5 W, with a diameter of ~100 km. Over the 12-day sampling period, the shipboard measurements were conducted alongside the drifters as they performed an almost complete circular pattern with a diameter of ~44 km (Figure 1a). Water-column seawater sampling for diel measurements took place every 4 h for a period of 4 days (26-30 July) and 3 days (31 July-3 August) at a depth of 15 m corresponding to the depth of the drogue. The water-column sampling was achieved using a 24 x 12 L Niskin bottle rosette attached to a conductivity-temperature-depth (CTD) package (SBE 911Plus, SeaBird) with additional fluorescence, oxygen, and transmissometer sensors. The sampling and analytical protocols for vertical profiles of nutrients, particulates, and flow-cytometry enumerated phytoplankton populations and heterotrophic bacteria were identical to those employed by the Hawaii Ocean Time-series program (http://hahana.soest.hawaii.edu/index.html).
Access & import/export	The research expedition departed and returned from the port of Honolulu, Hawaii. All sampling was conducted in international oceanic waters using routine oceanographic operations that predominantly involved seawater filtration and subsequent freezing of the filters. Samples consisted of nucleic acids and other biological macromolecules such as proteins and lipids. No live biological samples were collected as part of the research activities. No permits were needed for this work and the samples were not subject to export controls.
Disturbance	Seawater was sampled using routine oceanographic methods in the open ocean with minimum disturbance to the surrounding waters.

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