Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities

Sacha N. Coesel[a,1], Bryndan P. Durham[b], Ryan D. Groussman[a], Sarah K. Hu[c], David A. Caron[d], Rhonda L. Morales[a], François Ribalet[a], and E. Virginia Armbrust[a]

[a]School of Oceanography, University of Washington, Seattle, WA 98195; [b]Department of Biology, Genetics Institute, University of Florida, Gainesville, FL 32610; [c]Marine Chemistry & Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543; and [d]Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089-0371

The 24-h cycle of light and darkness governs daily rhythms of complex behaviors across all domains of life. Intracellular photoreceptors sense specific wavelengths of light that can reset the internal circadian clock and/or elicit distinct phenotypic responses. In the surface ocean, microbial communities additionally modulate nonrhythmic changes in light quality and quantity as they are mixed to different depths. Here, we show that eukaryotic plankton in the North Pacific Subtropical Gyre transcribe genes encoding light-sensitive proteins that may serve as light-activated transcription factors, elicit light-driven electrical/chemical cascades, or initiate secondary messenger-signaling cascades. Overall, the protistan community relies on blue light-sensitive photoreceptors of the cryptochrome/photolyase family, and proteins containing the Light-Oxygen-Voltage (LOV) domain. The greatest diversification occurred within Haptophyta and photosynthetic stramenopiles where the LOV domain was combined with different DNA-binding domains and secondary signal-transduction motifs. Flagellated protists utilize green-light sensory rhodopsins and blue-light helmchromes, potentially underlying phototactic/photophobic and other behaviors toward specific wavelengths of light. Photoreceptors such as phytochromes appear to play minor roles in the North Pacific Subtropical Gyre. Transcript abundance of environmental light-sensitive protein-encoding genes that display diel patterns are found to primarily peak at dawn. The exceptions are the LOV-domain transcription factors with peaks in transcript abundances at different times and putative phototaxis photoreceptors transcribed throughout the day. Together, these data illustrate the diversity of light-sensitive proteins that may allow disparate groups of protists to respond to light and potentially synchronize patterns of growth, division, and mortality within the dynamic ocean environment.

photoreceptors | microbial eukaryotes | oligotrophic gyre | diel cycles | metatranscriptomics

Most life on Earth evolved under an unending 24-h cycle of light and dark. Eukaryotic organisms commonly rely on an internal circadian clock to generate an estimate of time (1–3) to coordinate the sequence of key biological events, to minimize cellular damage from ultraviolet (UV) radiation and light-induced reactive oxygen species and to optimize the timing of specific activities such as photosynthesis or foraging for prey. The molecular mechanism of endogenous circadian clocks typically pivots around a network of regulatory proteins that form negative and positive transcriptional/translational feedback loops that together generate a biological cycle of ∼24 h (3). The circadian clock is tuned to environmental light conditions by light-sensitive photoreceptors (4, 5) that are often under circadian control themselves (6–8). Photoreceptor proteins additionally allow organisms to dynamically respond to fluctuating light conditions (9). In well-studied terrestrial organisms, the light-sensitive domains and associated chromophores of photoreceptors are excited by specific wavelengths of light and through resulting conformational changes,

trigger secondary messenger pathways that lead to differential gene expression and phenotypic output. Some photoreceptors interact directly with key clock elements (10, 11) and include the blue light-sensitive F-box protein ZEITLUPE from plants and the animal-type cryptochromes. Diversity in light sensitivity and phenotypic responses across taxonomic lineages is exemplified by different light-sensitive protein domains that are combined with different effector domains within a protein (12). For example, phytochromes and phototropins both include protein kinase effector domains that regulate plant morphogenesis and phototropism, respectively. However, phytochromes in plants are excited by red/far-red wavelengths (13) due to the association with a bilin chromophore, and phototropins are activated by blue light (14) via a flavin mononucleotide (FMN) at the Light-Oxygen-Voltage (LOV) domain. A member of the Per-ARNT-Sim (PAS) superfamily, the versatile LOV domain can be found in all kingdoms of life (e.g., Archaea, Bacteria, Eukarya, as well as in Viruses) in combination with different effector domains, many of which have not yet been functionally characterized (12). Members of the ancient cryptochrome/photolyase family (CPF) can also be found in all kingdoms of life and show a variety of different functions (15).

Significance

Most organisms coordinate key biological events to coincide with the day/night cycle. These diel oscillations are entrained through the activity of light-sensitive photoreceptors that allow organisms to respond rapidly to changes in light exposure. In the ocean, the plankton community must additionally contend with dramatic changes in the quantity and quality of light over depth. Here, we show that the predominantly blue-light field in the open-ocean environment may have driven expansion of blue light-sensitive regulatory elements in open-ocean eukaryotic plankton derived from secondary and tertiary endosymbiosis. The diel transcription of genes encoding light-sensitive elements indicate that photosynthetic and heterotrophic marine protists respond to and anticipate fluctuating light conditions in the dynamic marine environment.

In the ocean, planktonic organisms experience dramatic changes in light quality and quantity as ocean currents and turbulent mixing transport them to different depths throughout the day. Available light decreases exponentially with depth, and longer wavelengths of light disappear within the first few meters (16, 17). However, the marine microbial community is often highly synchronized to the day/night cycle (18–22), and studies on the different components of the circadian clock of model marine eukaryotic algal species are emerging (23–26). Eukaryotic marine planktonic communities are evolutionarily diverse with representative species from all major lineages across the eukaryotic tree of life (27) that adopted different lifestyles including photoautotrophy, mixotrophy, and heterotrophy. Laboratory studies using model organisms suggest that marine protists possess diversified photoreceptors, some with experimentally verified functions and some with structures not found in terrestrial organisms. The pervasiveness and use of these proteins in natural communities remains unknown. Examples of distinct algal photoreceptors include the aureochrome photoreceptors that couple a LOV domain to a basic leucine zipper (bZIP) transcription factor domain that binds target DNA upon illumination (28–30) and regulates photomorphogenesis in the brackish-water macroalgal species *Vaucheria frigida* (28), as well as the light-dependent cell cycle in marine diatoms (29); the dual-functioning cryptochrome/photolyase (dual-function CPF) proteins that couple DNA-repair and regulatory properties (31, 32); the type I channel and sensory rhodopsins that control positive and negative phototaxis in flagellated organisms by converting the light signal into an electrical or chemical signal that directly affects flagellar rotation (33, 34); and phytochromes and proteorhodopsins with regulatory and phototrophic functions, respectively, that are spectrally tuned in some marine plankton toward shorter wavelengths of the light spectrum than their terrestrial counterparts (35–37). Here, we combine genetic surveys and metatranscriptomes of high temporal resolution to show that natural communities of open-ocean marine protists transcribe genes encoding putative photoreceptors and related light-sensitive proteins that potentially sense the UV/blue/green region of the light spectrum. Together, these regulatory elements may help coordinate the behavior of diverse taxonomic lineages in the dynamic aquatic light field.

Results and Discussion

Environmental Sampling in a Diel Context. Recent advances in photoreceptor studies of model marine algal species have highlighted the diversity in structural and/or functional regulators, many of which show light dependence or circadian control at the transcriptional level (15, 23–26, 29, 31, 38). However, most open-ocean plankton are not represented by cultured model organisms. To determine the diversity and potential diel transcriptional patterns of genes encoding light-signaling proteins in natural communities of surface open-ocean plankton, we collected samples for eukaryotic metatranscriptomes every 4 h for 4 d from the surface ocean (15 m) within the North Pacific Subtropical Gyre. Lagrangian tracking of free-floating drogues centered at 15 m allowed repeated sampling of a plankton community ~100 km northeast of Station ALOHA (A Long-Term Oligotrophic Habitat Assessment; 22.75°N, 158°W) (39). Organisms in the size range of 0.2 to 100 μm were collected; we focused on the single-celled eukaryotes (protists) and multicellular organisms such as crustaceans, cnidarians, and annelids with life-cycle stages that include small cells. Dominant and metabolically active eukaryotic photosynthetic plankton groups in this area include Dinophyceae (dinoflagellates), Haptophyta (haptophytes), and Bacillariophyceae (diatoms) (21, 40, 41) (*SI Appendix*, Table S1). Throughout the sampling period, the sun rose at ~0600 and set at ~1800 Hawaii–Aleutian Standard Time, with surface light intensities at noon reaching over 2,000 μmol m^{-2} s^{-1} (Fig. 1*A*). The oscillating increase in mean picoeukaryotic cell size during the day and decrease in mean cell size during the night, as estimated from

continuous flow cytometry measurements (42), reflected a light-driven synchronization of cell growth and cell division in the protist community (Fig. 1*B*). Similar oscillations in the cell diameter of larger plankton (>5 μm) over the day/night cycle (43) indicates that this light-driven synchrony extends across the plankton community. Cell concentrations of the most abundant eukaryotic picophytoplankton in this region (<5 μm in diameter) (44) remained relatively constant at ~1.5 × 10^6 cells per liter over the 4-d sampling period (Fig. 1*C*), reflecting a tight coupling between cell division and cell mortality. Differential attenuation of the light spectra with depth at our study site is illustrated by the loss of far-red light by 6 m, red light by 13 m, green light by 105 m, and blue light by 140 m (Fig. 1*D*). The depth of the surface mixed layer was estimated at 21 ± 5 m, as defined by a 0.03 kg/m^3 density offset from 10 decibar (db) (45), with an estimated mixing time scale of 2 to 4 h within this layer at this time of year (17). Over the course of the day, the plankton community of the surface mixed layer experienced light levels that varied at least threefold at any given time as cells were mixed within the upper ocean, with more dramatic variations in intensity for the longer red wavelengths of light.

Photoreceptors and Other Light-Sensitive Elements Transcribed by an Open-Ocean Protist Community. Transcripts within the metatranscriptomes were annotated based on a database of photoreceptors and other related light-sensitive protein sequences created by screening publicly available genomes and transcriptomes of over 500 marine protists, bacteria, archaea, and viruses (Dataset S1), using custom-made hidden Markov model (hmm)-profiles (Fig. 2*A* and Dataset S2; e < 0.001; hmmsearch) (46). Phylogenetic trees were generated for the thousands of distinct homologs to the microbial rhodopsins (*SI Appendix*, Fig. S1 and Dataset S3*A*), the cryptochrome/photolyase proteins (*SI Appendix*, Fig. S2 and Dataset S3*B*) and LOV domain-containing proteins (*SI Appendix*, Fig. S3 and Dataset S3*C*), and the few hundred distinct phytochrome sequences (*SI Appendix*, Fig. S4 and Dataset S3*D*). Phylogenetic placement analysis (pplacer; maximum likelihood mode) (47) was used to identify the putative taxonomy of the environmental homologs by mapping the short (~240 base pairs [bp]) amino acid-translated environmental sequences to the reference phylogenetic trees. The phylogenetic placement of the environmental transcripts was estimated to be ~90% accurate at the taxonomic-order level (*SI Appendix*, Fig. S5 and Dataset S4).

The most abundant eukaryotic light-sensitive protein-encoding transcripts detected near Station ALOHA encode rhodopsin homologs with an average of ~2 × 10^8 rhodopsin transcripts per liter detected over the 4-d sampling period. The greatest abundance of transcripts is associated with the Dinophyceae and the Bacillariophyceae (Table 1). Two orders of magnitude fewer transcripts were detected for the genes encoding the CPF (Cry/Phot) (~1 × 10^6 transcripts per liter) and LOV-containing (~5 × 10^5 transcripts per liter) proteins. Phytochrome transcripts were near the limit of detection at ~1 × 10^3 transcripts per liter. Diel patterns of the photoreceptor and related light-sensitive protein-encoding transcripts grouped at the order level varied across taxa (Table 1 and Dataset S5). At the two extremes, about half the transcripts from Haptophyta displayed diel patterns of transcript abundance (based on Rhythmicity Analysis Incorporating Non-Parametric Methods [RAIN] analysis; *P* < 0.001) (48), whereas the Dinophyceae displayed the fewest diel oscillating transcripts, possibly reflecting a preferential use of posttranscriptional regulation to tune their physiology to environmental conditions (49–52). A small subset of transcripts mapped most closely to those found in viruses (Table 1). Active viral-induced DNA repair in infected protists is suggested by detection of ~2 × 10^4 transcripts per liter for homologs of a class I to III cyclobutene–pyrimidine dimer (CPD) photolyase found in giant viruses of Amoebae (53–55). A few rhodopsin ion-pump transcripts (~3 × 10^2
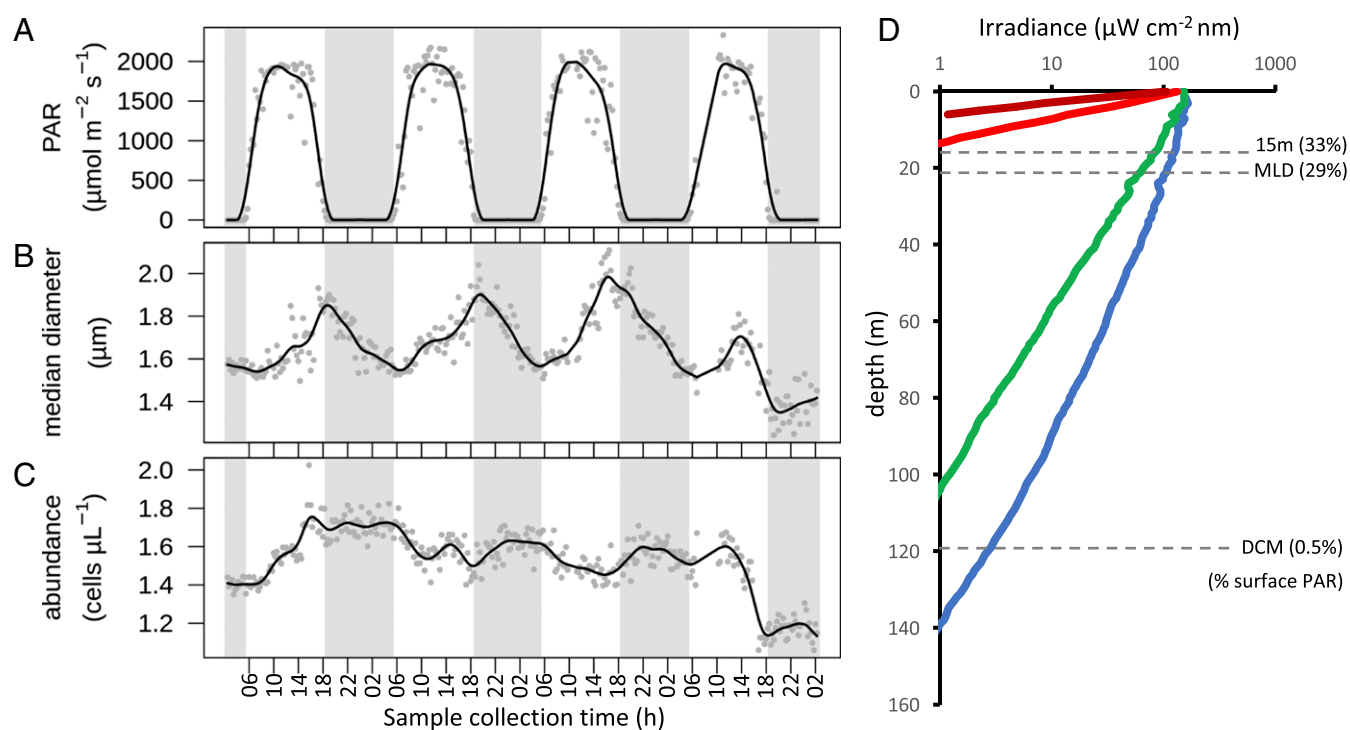
Fig. 1. Characteristics of the sampling site near Station ALOHA (July 26 to 30, 2015). (*A–C*) Photosynthetically active light (PAR) intensity at the ocean surface (*A*), median cell diameter of eukaryotic phytoplankton less than 5 μm in diameter (*B*), and abundance of eukaryotic phytoplankton less than 5 μm in diameter (*C*). Points indicate measurements, and solid lines represent smoothed data (spline of order 3). Collection times of metatranscriptome samples are indicated. (*D*) Depth profile of available irradiance at wavelengths of 430 to 480 nm (blue line), 500 to 560 nm (green line), 650 to 680 nm (light red line), and 700 to 740 nm (dark red line) were measured at noon on July 30, 2015. Dashed lines indicate the 15-m sampling depth, the 21-m mixed-layer depth (MLD), and the 119-m DCM; the percentage of PAR as compared to the surface PAR is indicated for these depths.

transcripts per liter) were detected that are most similar to those from giant double-stranded DNA viruses (56). Overall, the greatest number of environmental light-sensitive protein-encoding transcripts are derived from plastidic protists that must optimize photosynthesis and minimize photobleaching relative to the light/dark cycle. Environmental transcripts encoding potential photoreceptors and other light-sensitive proteins of heterotrophic (nonphotosynthetic) organisms within the Opisthokonta, Ciliphora, and Bigyra were proportionally underrepresented relative to their 18S ribosomal DNA (rDNA) abundances (Table 1 and *SI Appendix*, Table S1), suggesting a potential evolutionary divide in photoreceptor expansion and utilization driven by trophic mode.

Photoreceptor and Other Light-Sensitive Protein Classes Are Differentially Distributed over the Major Taxonomies. Taxonomic groups preferentially transcribed genes encoding different classes of photoreceptors and other light-sensitive proteins, suggesting that the observed synchrony of the protist community to the light/dark cycle (Fig. 1*B*) was regulated through distinct mechanisms. Environmental transcripts for phytochromes, typically (far-)red light-sensitive in plants (13), were restricted to a subset of Chlorophytes (Fig. 2*B* and *SI Appendix*, Fig. S4), despite their presence in the genomes and transcriptomes of cultured isolates of Cryptophytes, Bacillariophyceae, and other photosynthetic stramenopiles (Fig. 2*B*) and their demonstrated regulatory role in the model diatom *Phaeodactylum tricornutum* (57). Phytochromes from model protists can perceive shorter wavelengths of light up to the blue region of the light spectrum (35), perhaps of importance to marine phytoplankton that are routinely mixed to depths greater than that of red-light penetration. The lack of phytochrome transcripts associated with natural communities in the subtropical gyre suggests that perception of light via phytochrome-based signaling plays a relatively minor role in the open ocean. In contrast, transcripts associated with the UV/blue-light cryptochrome/

photolyase proteins (5, 11) were the most taxonomically widespread. These proteins are thought to have evolved from Precambrian-time cyanobacterial photolyases (58), and their taxonomic spread in modern organisms may indicate selective retention of these proteins. Within the CPF, relatively few transcripts were detected for the canonical animal (6-4) photolyase and animal type I cryptochrome proteins or the canonical plant cryptochrome and type I to III CPD photolyases that dominate terrestrial systems. Instead, detected environmental transcripts suggest that natural communities of marine protists predominantly rely on plant-like cryptochromes, dual-function cryptochrome/photolyase proteins (dual-function CPF) that are closely related to (6-4) photolyases, CryDASH proteins, and type II CPD photolyases (Fig. 2*B* and *SI Appendix*, Fig. S2). Plant cryptochromes help entrain the *Chlamydomonas reinhardtii* circadian clock and other light-dependent processes (59), and plant-like cryptochromes are thought to modulate the transcription levels of both phytochrome and dual-function CPF in *P. tricornutum* (60). Both dual-function CPF (also called CPF1) and CryDASH proteins remain understudied, and their exact role in DNA repair and light regulation is not yet fully understood. In the few examined marine organisms, these proteins appear to have cryptochrome-regulatory activity; dual-function CPF additionally performs (6-4) photolyase repair (31), and CryDASH additionally performs CPD photolyase activity (38, 61). Moreover, the dual-function CPF from *C. reinhardtii*, commonly referred to as "animal-like cryptochrome" (62), has an extended action spectrum with sensitivity of up to 680 nm (63). Our phylogenetic analysis cannot distinguish non–animal-derived (6-4) photolyase and dual-function CPF (*SI Appendix*, Fig. S2; clade A), and further studies are needed to resolve the molecular functions and action spectra of these proteins in the different classes of marine protists. Together, these data illustrate divergent evolutionary paths
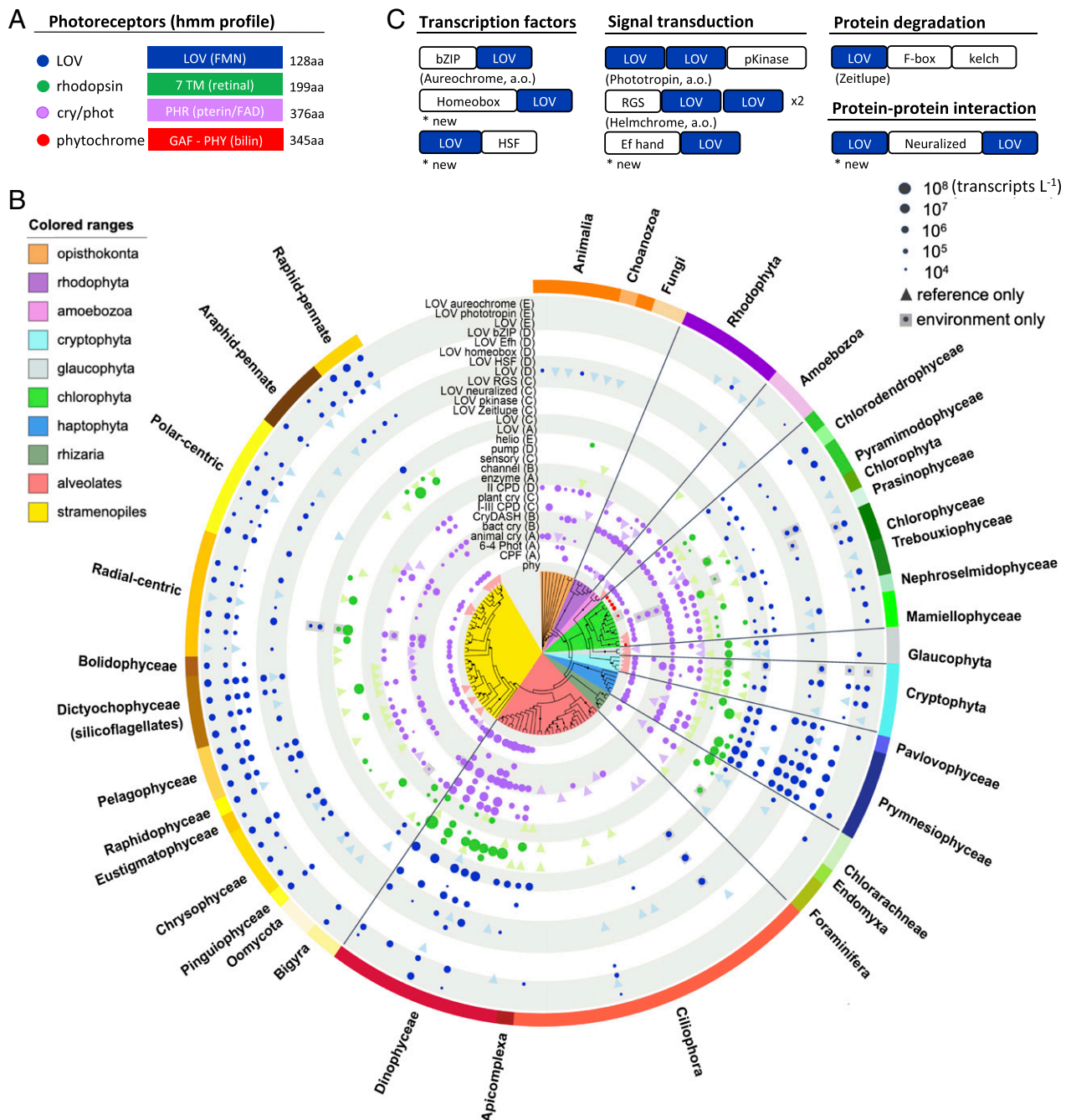
ENVIRONMENTAL SCIENCES

Fig. 2. Abundance and taxonomic distribution of environmental photoreceptor and other light-sensitive protein-encoding transcripts. (*A*) Schematic representation of the hmm-profiles used to identify environmental photoreceptor and other light-sensitive protein-encoding transcripts. The respective chromophores are indicated with parentheses: flavin mononucleotide (FMN) for the LOV domain, retinal for the seven transmembrane (7 TM) helices of rhodopsin, pterin and flavin adenine dinucleotide (FAD) for the photolyase-homologous region (PHR) of the CPF, and bilin for the GAF and PHY domain constituting the photosensory part of phytochrome. The length of the hmm-profile is indicated in amino acids (aa). (*B*) Environmental transcript abundance of phytochrome (red), cryptochromes/photolyase (violet), rhodopsin (green), and those with LOV domains (blue) visualized on an 18S ribosomal RNA maximum-likelihood phylogenetic tree, representing 117 different eukaryotic orders relevant for the marine environment. The taxonomic phylum and class-level classifications are indicated by the colored ranges further annotated in *SI Appendix*, Fig. S6. Protein subtypes (clades A ~ E) are derived from the respective reference trees (*SI Appendix*, Figs. S1–S4). Colored circles indicate transcripts detected in both the reference sequences and the environmental samples at the respective order level. The size of the circle corresponds to the mean environmental transcript concentrations (transcripts per liter) over the 4-d sampling period. Triangles indicate detection of transcripts in the reference sequences that are not detected in the environmental samples. Gray boxed circles indicate transcripts detected in the environmental samples only. (*C*) Schematic presentation of the domain structures of the LOV domain-containing sequences retrieved from our light-sensitive protein database of reference sequences. 6-4 Phot, (6-4) photolyase; animal cry, animal type I cryptochrome; bact cry, bacterial cryptochrome; channel, channel rhodopsin; CPF, (6-4) photolyase/cryptochrome dual-function proteins; CryDASH, cryptochrome-DASH; enzyme, enzyme rhodopsin; helio, heliorhodopsin; I-III CPD, type I to III CPD photolyase; II CPD, type II CPD photolyase; LOV (A ~ E), clade-aggregated counts of LOV transcripts (*SI Appendix*, Fig. S3) that are not included in *B*; plant cry, plant cryptochrome (found only in Chlorophyta and Rhodophyta) and plant-like cryptochrome (all other taxonomies; *SI Appendix*, Fig. S3); phy, phytochrome; pump, proteorhodopsin and other ion-pump rhodopsins; sensory, sensory rhodopsin.

Table 1. Mean environmental transcript concentrations (10^4 transcripts per liter) over the 4-d sampling period

Taxonomic group	Rhodopsin	Phytochrome	Cry/Phot	LOV	Total
Amoebozoa					
Amoebozoa	0.2 (0)	—	4 (0)	1 (0)	5
Opisthokonta					
Animalia, Fungi, Choanozoa	1 (0)	—	23 (13)	0.3 (0)	25
Rhizaria					
Chlorarachneae, Foraminifera	351 (99)	—	8 (12)	6 (0)	365
Alveolata					
Ciliophora	0.01 (0)	—	7 (15)	16 (22)	23
Apicomplexa	41 (0)	—	3 (0)	0.5 (0)	45
Dinophyceae	19,774 (2)	—	740 (20)	255 (1)	20,769
Cryptophyta					
Cryptophyta	9 (0)	—	11 (0)	6 (0)	26
Haptophyta					
Haptophyta	671 (57)	—	25 (62)	168 (66)	863
Stramenopiles					
Bigyra	2 (0)	—	5 (37)	1 (0)	8
Bacillariophyceae	14,28 (30)	—	21 (0)	26 (26)	1,475
Chrysophyceae	0.4 (0)	—	3 (0)	11 (55)	15
Dictyochophyceae	207 (0)	—	2 (0)	20 (48)	228
Pelagophyceae	4 (0)	—	5 (0)	17 (45)	26
Pinguiophyceae	20 (0)	—	0.3 (0)	2 (0)	23
Synurophyceae	0.1 (0)	—	0.01 (0)	5 (58)	5
Archaeplastida					
Chlorophyta	22 (0)	0.1 (0)	25 (14)	10 (41)	57
Rhodophyta	—	—	42 (0)	0.2 (0)	42
Glaucophyta	15 (0)	—	1 (0)	0.4 (0)	16
Viruses					
Viruses	0.03 (0)	—	2 (0)	—	2
Total	22,545	0.1	926	545	24,016

Parenthesis indicate the percentage of transcripts that displayed diel periodicity in abundance analyzed at the phylogenetic "order" level (RAIN; $P < 0.001$). Terms in bold indicate the different phyla. Dashes indicate transcripts not detected.

for the cryptochrome/photolyase protein family and the phytochromes, possibly shaped by the blue-light dominated "light-scape" of open-ocean waters.

Despite their overall abundance, microbial rhodopsin transcripts were restricted to specific subsets of organisms. The most abundant rhodopsin transcripts mapped to the light-activated ion-pump rhodopsins (Fig. 2B) that include proteorhodopsin, which is involved in adenosine triphosphate (ATP) synthesis rather than regulation (33). The green or blue light-sensitive (36, 37) ion-pump transcripts were scattered across the Dinophyceae, a subset of Bacillariophyceae, Prymesiophyceae, and a putative Chlorarachneae (Rhizaria; *SI Appendix*, Fig. S1; clade D), a seemingly haphazard pattern that reiterates the potential role of horizontal gene transfer, from either bacteria or viruses, in the spread of proteorhodopsin throughout eukaryotic lineages (64, 65). Blue-light tuning of proteorhodopsin (based on a defining amino acid sequence) has previously been shown in open-ocean bacteria, particularly at depth (36, 37). We detected a blue light-tuned proteorhodopsin from the parasitic Dinophyceae *Amoebophrya* in the reference database (Dataset S3A) as well as in a subset of poly(A)-selected environmental assembled contigs (Dataset S7D), indicating that marine protists are also able to tune rhodopsin to optimize their light-absorption spectrum for ATP generation.

Environmental transcripts for sensory and channel rhodopsins were identified in protists within representative taxa from the Chlorophyta, Cryptophyta, stramenopiles, Alveolata, and others (Fig. 2B), representing a wider range of taxonomic origins than proteorhodopsin and other ion-pump rhodopsins. However, transcript abundances of the sensory and channel rhodopsins were orders of magnitude lower than abundances of the enzymatic rhodopsin pumps (Fig. 2B). Sensory and channel rhodopsins are commonly found localized within the eyespot of motile algae to sense the direction of incoming light and allow light-dependent movement (33, 66–68), but they may also be implicated in other light-dependent processes, such as cellular differentiation in the multicellular green alga *Volvox carteri* (69). Sensory rhodopsin transcripts were widespread in the Cryptophyta, a flagellated group of algae known to phototax toward specific wavelengths of light (70). These observations suggest that marine protists in this open-ocean community respond to variations in spectral quality and photon flux to orient themselves within the water column or within their microenvironment. These proteins may also exert unexplored light-dependent regulatory functions in, for example, developmental processes (e.g., in sexual reproduction) or shifts in trophic mode. In addition, the recently described heliorhodopsin, with a putative sensory function (71), was transcribed at low levels by some orders within all phyla except Rhodophyta and Amoebozoa (Fig. 2B). Rhodopsin-histidine kinase transcripts, thought to contribute to the *Ostreococcus* circadian clock (26), were not detected in our dataset, and environmental enzyme rhodopsin (33) transcripts mapped primarily to a C2-domain membrane-targeted rhodopsin of unknown function (Dataset S5B).

The greatest diversity was detected for environmental transcripts that encode members of the LOV-containing protein family. These include well-studied blue light-sensitive photoreceptor sequences such as phototropin, ZEITLUPE, and aureochrome (72), as well as LOV domain-containing sequences retrieved from our light-sensitive protein database (Fig. 2C).

The reference genes segregate into five distinct clades (*SI Appendix*, Fig. S3), with clade D representing the largest and most divergent group of LOV domain sequences and including animal sequences that encode voltage-regulated potassium channels that are likely not excited by light as they lack the light-sensitive motif (Dataset S3D). We identified a total of 129 unique conserved domain (CD) annotations (73), located adjacent to the LOV domain in the full-length LOV references (Dataset S6A). The genetic mobility and plasticity of the LOV-domain proteins were also apparent within the environmental metatranscriptomes. Protists that arose from red-algal secondary or tertiary endosymbiosis, e.g., Dinophyceae, Haptophyta, and stramenopiles, transcribed multiple types of LOV-domain sequences in the environment (Fig. 2B). Interestingly, Rhodophyta themselves do not appear to transcribe these genes, suggesting that LOV-domain genes may have been lost over evolutionary time in Rhodophyta, in accordance with the reduced gene diversity reported for the genome of the red seaweed *Chondrus crispus* (74), or, alternatively, that LOV-domain gene-duplication/recombination events in Chromista and Dinophyceae occurred after the red-algal secondary and tertiary endosymbiosis events (<800 million years ago). Chlorophyta were the only taxonomic group with multiple parallels to known light-sensory pathways of higher plants. In addition to the plant-type cryptochrome and phytochrome, several orders of Chlorophyta transcribed phototropin and the F-box protein ZEITLUPE, indicating that the higher plant light-sensory pathways were already in place before colonization of land. Low levels of LOV-histidine kinase transcripts (24) were detected in Chlorophyta (Fig. 2B; clade A, Chlorophyta), and phototropin-like transcripts were detected for Cryptophyta in our environmental dataset (Fig. 2B).

Aureochromes are specific to photosynthetic stramenopiles and couple the LOV domain with the bZIP DNA-binding domain (Fig. 2C) (28, 75). As expected, environmental aureochrome transcripts were detected in Bacillariophyceae and other photosynthetic stramenopiles (Fig. 2B). Unexpectedly, they were also detected for Peridiniales, an order within the Dinophyceae known for kleptoplasty of diatom plastids (76, 77), suggesting that these aureochrome transcripts were derived from the engulfed diatom. We also detected additional putative transcription factors transcribed in the environment by Haptophyta and photosynthetic stramenopiles that combine a LOV domain with a DNA-binding domain (HSF, homeobox, and bZIP; Fig. 2 and *SI Appendix*, Fig. S3; clade D). The coupling of a potential light-sensitive domain with DNA-binding domains appears to be an innovation restricted to organisms derived from a secondary endosymbiosis, a critical group of phytoplankton in modern oceans (78, 79).

The flagellated protists possess members of the LOV protein families that are thought to be involved in phototaxis. Helmchrome proteins couple two LOV domains with two tandem repeats of a Regulator of G protein Signaling (RGS) domain (Fig. 2C) (80, 81) and are located at the base of the flagella in brown algae (80). Environmental helmchrome transcripts were associated with motile photosynthetic stramenopiles, such as Pelagophyceae and Dictyochophyceae (predominantly silicoflagellates; Fig. 2B and *SI Appendix*, Figs. S6 and S3; clade C). Detection of transcripts encoding both helmchrome and sensory rhodopsins in the Pelagophyceae may allow these organisms to sense low levels of light (~0.5% of surface levels) and thrive near the deep chlorophyll maximum, where these organisms are often observed (82, 83). Helmchrome-like sequences that couple LOV to a single RGS domain were also detected for Pavlovophyceae and Prymnesiophyceae, two members of the Haptophyta (Fig. 2B and *SI Appendix*, Fig. S3). These sequences have not been previously described in the literature, and their potential role in phototaxis remains unknown. A recently described RGS-LOV-DUF protein was found to rapidly associate with the plasma membrane upon blue-light excitation in the fungus *Botrytis cinerea* (84), and

similar mechanisms may persist in marine protists. The combination of a calcium-binding EF-hand motif with a LOV motif in several orders of Dictyochophyceae (Fig. 2 B and C and *SI Appendix*, Fig. S3; clade D) has also not been described previously. The EF-hand motifs serve as the Ca^{2+}-binding domain in the signaling proteins calmodulins and other regulatory proteins (85), suggesting a coupling between light and calcium-based signaling in these flagellated protists.

Dinophyceae display multiple divergent regulatory mechanisms, including a predominance of posttranscriptional and/or translational rather than transcriptional regulation (49–52). Detection of Dinophyceae-derived environmental transcripts with a LOV domain located on either side of a Neuralized Homology Repeat (NHR) domain, also found in E3 ubiquitin ligase (Fig. 2 B and C), suggests a potential for light-regulated protein–protein interaction (86). Although the LOV-neuralized transcript levels did not oscillate over the diel cycle (*SI Appendix*, Fig. S7), the encoded proteins may play a role in regulation by light-dependent protein binding. This may be especially important for Dinophyceae that may depend less on typical transcriptional regulatory mechanisms (49–52).

Protein Domain Structure and Light-Sensitive Motifs of Marine LOV-Domain Sequences. The diversity of the LOV domain-encoding transcripts implied modes of regulation in open-ocean communities. We focused on those subsets of transcripts that displayed diel patterns of abundance (RAIN; $P < 0.001$), as we hypothesized that these transcripts encoded proteins that may play a role in regulating the phasing of organisms to the light/dark cycle or other light-dependent regulatory roles. The metatranscriptomes were assembled de novo and environmental LOV domain-containing contigs were identified by hmmsearch ($e < 0.001$). A randomized axelerated maximum-likelihood (RAxML) phylogenetic tree was generated for the diel-transcribed environmental contigs (clustered at 90% identity) and their closest homologs retrieved from our light-sensitive protein database of reference sequences (Fig. 3A and Dataset S7B). Included are also the LOV-domain homologs used to generate the LOV hmm-profile in this work (Dataset S2A) and PAS-domain sequences that do not possess a LOV domain (PF00989) derived from the Pfam protein families database (87). The Pfam-derived PAS-domain sequences form a distinct outlier group (Fig. 3A, gray edges), confirming that the environmental sequences identified in this work correspond to LOV domain-containing proteins. The diel environmental sequences fell into three distinct clades, two of which contained proteins with associated effector domains (Fig. 3 B and C).

One clade contains a variety of well-characterized photoreceptor proteins with different effector domains: Chlorophyte homologs to phototropin include a pKinase domain, photosynthetic stramenopile homologs to aureochrome include a bZIP domain, and Haptophyte and photosynthetic stramenopile homologs to helmchrome include an RGS domain (Fig. 3B). Environmental contigs with homology to either aureochrome or helmchrome were found scattered across this clade. No environmental contigs with both a LOV and pKinase domain were found, perhaps reflecting the low abundance of Chlorophyte transcripts at the study site. As previously described by Krauss et al. (72), the potassium voltage-gated channel (Kv channel) LOV-domain proteins also grouped in this clade, as did the redox-sensitive bacterial NIFL proteins, albeit with low bootstrap support (Fig. 3 A and B); neither the Kv channel proteins nor the NIFL proteins are light-sensitive. The canonical light-sensitive motif of the LOV domain consists of GXNCRFLQG, with the cysteine residue commonly required for covalent linkage with the flavin-nucleotide chromophore upon blue-light activation. As expected, the Kv channel and NIFL sequences did not possess the canonical light-sensitive motif. The light-sensitive motif was largely conserved across all other members of the clade, including within the newly identified environmental
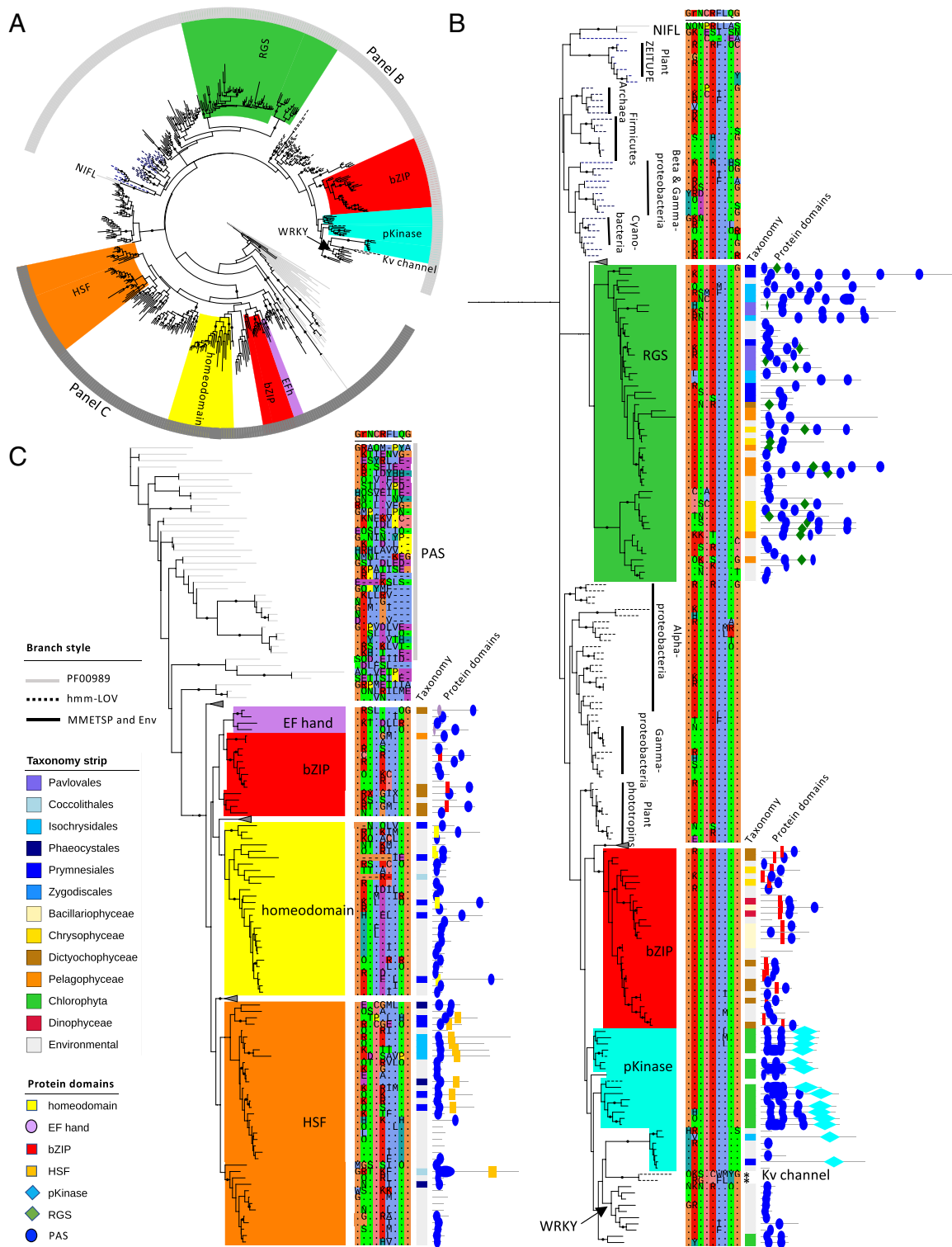
Coesel et al.
Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities

Fig. 3. Phylogenetic and domain analysis of LOV-containing environmental contigs. (*A*) Maximum-likelihood tree (RAxML) of protist environmental LOV-domain contigs that are transcribed with a diel rhythmicity and their closest reference database-derived homologs (black edges). Also included are PAS-domain (PF00989; gray edges) and LOV-domain (hmm-LOV; dashed edges) reference sequences. Effector domains are indicated by the colored ranges. Bootstrap values of 80% and higher (100 iterations) are indicated with black circles. The gray bars indicate the upper and lower sections of the tree that are expanded in *B* and *C*, respectively. The arrow indicates the placement site of the WRKY-LOV sequences (*SI Appendix*, Fig. S9) by pplacer. (*B* and *C*) Expansion of the upper and lower tree sections, respectively. Shown are only the clades in which the LOV domain was found to be associated with an effector domain. Indicated from left to right: alignment of the light-sensitive motif GXNCRFLQG within the LOV domain (ClustalX colorscheme; dotted repeats), color strip indicating the taxonomies of the sequences retrieved from our light-sensitive protein database with the environmental-derived contigs in gray, and schematic representation of the protein sequences with the locations of the protein domain motifs indicated by the various shapes color-coded as the tree ranges and with the PAS/LOV domain in blue. The non–light-sensitive NIFL and Kv channel sequences are indicated. The asterisk indicates that the light-sensitive motifs of the Kv channel proteins contain two additional amino acid residues (see also *SI Appendix*, Fig. S8).

Coesel et al.
Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities

contigs (Fig. 3*B* and *SI Appendix*, Fig. S8), suggesting that these proteins may serve as light-sensitive photoreceptors in the environment. Interestingly, phylogenetic analysis identified a set of seven closely related environmental contigs that group with this clade that were without homologs in our reference sequence database. These sequences combine a conserved LOV domain with a W-box DNA-binding WRKY domain (*SI Appendix*, Fig. S9*A*) that were subsequently also identified in datasets of sunlit oceans across the globe (*SI Appendix*, Fig. S9*B*), as deposited in the Oceans Gene Atlas (88). This domain combination has not previously been described in the literature and suggests a class of light-activated transcription factors, possibly belonging to the Chlorophyta (Fig. 3*B* and *SI Appendix*, Fig. S9*C*). A W-box *cis*-acting element presumably recognized by WRKY transcription factors has been previously identified in the promoter sequences of genes involved in biosynthesis of carotenoid pigments in the green alga *Dunaliella bardawil*. Transcription of these genes is induced by both light and salt, but the responsible transcription factor has not yet been identified (89). The environmental LOV-WRKY sequences identified here represent an avenue for identifying the potential type of transcription factor involved in regulating carotenoid biosynthesis in this biotechnologically valuable species.

A second clade (Fig. 3*C*) consists of LOV-domain sequences from Haptophyta and photosynthetic stramenopiles that possess either a homeobox, bZIP, or EF-hand motif on the amino-terminal side of the LOV domain or an HSF domain on the carboxyl-terminal side. A majority of environmental sequences within this clade were distinct from the reference sequences, and the light-sensitive motif within these sequences was more variable (Fig. 3*C* and *SI Appendix*, Fig. S8). In particular, the conserved cysteine residue was present within the light-sensitive motif of the HSF-containing variants. The cysteine was replaced with a leucine in the homeobox-, bZIP-, or EF-hand–containing variants. Recent mutation-based studies with *Neurospora* and bacteria indicate that photoexcitation and signal transduction can occur in the absence of the canonical cysteine residue within the light-sensitive motif, with LOV reactivity achieved through photoreduction (90). Subsequent detection of natural cysteine-lacking variants of LOV photoreceptors in archaeal halobacteria lent support for this proposal. Whether the environmental LOV-containing variants with the divergent light-sensitive motif nonetheless undergo light-induced conformational changes needs to be experimentally verified. However, the diel pattern of environmental transcript abundance suggests a potential role for these proteins in diel transcriptional or calcium-based regulation.

Phasing and Depth Prevalence of Environmental Photoreceptor and Other Light-Sensitive Protein-Encoding Transcript Levels. The light-sensitive proteins utilized by open-ocean protist communities could function in different ways, for example, as photoreceptors that initiate secondary messenger pathways to instigate light-regulated processes, as proteins involved in phototaxis, as light-activated transcription factors, or as a combination of the above. To further evaluate potential roles of these proteins, the timing of peak transcript abundance (transcripts per liter seawater) was estimated for those genes that displayed significant rhythmicity (RAIN analysis; $P < 0.001$). Transcript abundance of homologs to well-known photoreceptor and other light-sensitive proteins peaked either at or just before dawn, illustrating a high level of synchrony across taxonomies (Fig. 4 *A* and *B* and Dataset S5*I*) and suggesting the anticipation of the dawn light signal by these organisms. This includes transcript abundance of cryptochrome/photolyase genes (Fig. 4 *A* and *B*, blue hues), the LOV-domain genes associated with a pKinase domain (e.g., phototropin; Fig. 4 *A* and *B*, brown hue), and most rhodopsin genes except one ion-pump variant from the Dinophyceae and transcript abundance of heterotrophic protists belonging to Animalia, Bigyra, Ciliphora, and Rhizaria (Cercozoa) (Fig. 4 *A* and *B*, green hues). This is in

accordance with the daily steady-state transcript peak levels of homolog photoreceptors recorded for a diversity of terrestrial model organisms (6, 7), for marine algal model organisms such as the diatoms *P. tricornutum* and *Thalassiosira pseudonana* (32, 91), the green algae *Ostreococcus tauri* (38) and the Dinophyceae *Prorocentrum donghaiense* (92), as well as for a variety of eukaryotic taxonomies, as measured in situ in the California current (20).

Phototaxis-related genes displayed a different transcript abundance pattern. Transcripts associated with helmchrome from motile photosynthetic stramenopiles were transcribed throughout the light phase (Fig. 4 *A* and *B*, purple hue), possibly reflecting rapid protein turnover times common for flagellar proteins (93, 94). In contrast, helmchrome-like transcripts in Haptophyta peaked at dawn, suggesting a potentially divergent role for these proteins in Haptophyta. Diel periodicity was not detected for transcripts associated with channel and sensory rhodopsins (*SI Appendix*, Fig. S10).

The greatest diversity in the timing of peak transcript abundance occurred with the LOV domain-containing transcription factors present in the Haptophyta and photosynthetic stramenopiles (Fig. 4 *A* and *B*). The light-responsive transcription factors aureochrome 1A and 1C were transcribed during the day, with peak levels shifted earlier for 1C, as previously demonstrated in *P. tricornutum* (95). Interestingly, while Aureochrome 2 transcripts in the model diatom did not oscillate over the diel cycle, the environmental Bacillariophyceae homologs showed a sharp peak in transcript abundance at dawn. Aureochrome 2 transcripts transcribed by the kleptoplastic Peridiniales (Fig. 2*B*) also peaked at dawn, suggesting their enslaved diatom maintained diel rhythmicity, a remarkable example of symbiotic intracellular regulation. Within the Haptophyta, transcript abundance of the putative transcription factors peaked at different times depending on the particular type or the taxonomic order. The latter may reflect different metabolic lifestyles of the underlying species (photoautotrophic versus mixotrophic), a possibility that cannot be resolved at the taxonomic-order level within this study. Homeobox-associated LOV domains peaked in the morning in Prymnesiales and Zygodiscales (Fig. 4 *A* and *B*, pink hue). Haptophyta and stramenopiles lack canonical clock components (25), and thus it remains unclear whether or how transcription factors may interact with clock components. The recent identification of a novel clock component RITMO1 (bHLH-PAS) in diatoms (23) should provide new avenues for exploring potential interactions.

The HSF-associated LOV domains peaked at dusk in Isochrysidales and Pavlovales and later at night in Coccolithales, Phaeocystales, and Prymnesiales (Fig. 4 *A* and *B*, gray hue). Transcripts for the HSF, bZIP, and the EF-hand calcium binding-containing LOV-domain proteins also peaked at night within the Dictyochophyceae and other photosynthetic stramenopiles (Fig. 4 *A* and *B*). Nighttime peaks in photoreceptor transcript abundances are not uncommon in photosynthetic organisms and have been previously detected for *P. tricornutum* phytochrome (57, 91) and for plant Cry in *C. reinhardtii* (96). In higher plants, night-time transcribed phytochrome B regulates flowering time (97) in a process sensitive to night-time light pollution (98). However, it must be noted that a disconnect between transcript abundance and cellular protein levels can arise, as the lifetime of a protein is dependent on both the rate of synthesis and the turnover. This is exemplified by the *P. tricornutum* phytochrome transcript levels that peak at dusk, while phytochrome protein levels peak at dawn (57), and by the light-dependent degradation of the *C. reinhardtii* plant Cry (99).

To further evaluate the transcriptional sensitivity of the different photoreceptor and other light-sensitive proteins to changes in light quantity and quality across the euphotic zone, we collected samples on day 4 of the study from depths of 5, 119, 150 m across the euphotic zone. Logistical constraints limited sampling to dusk at 1800 hours. A majority of the 25 most abundant transcripts were differentially detected with depth and either contained a blue light-sensitive LOV domain or corresponded to the blue light-sensitive
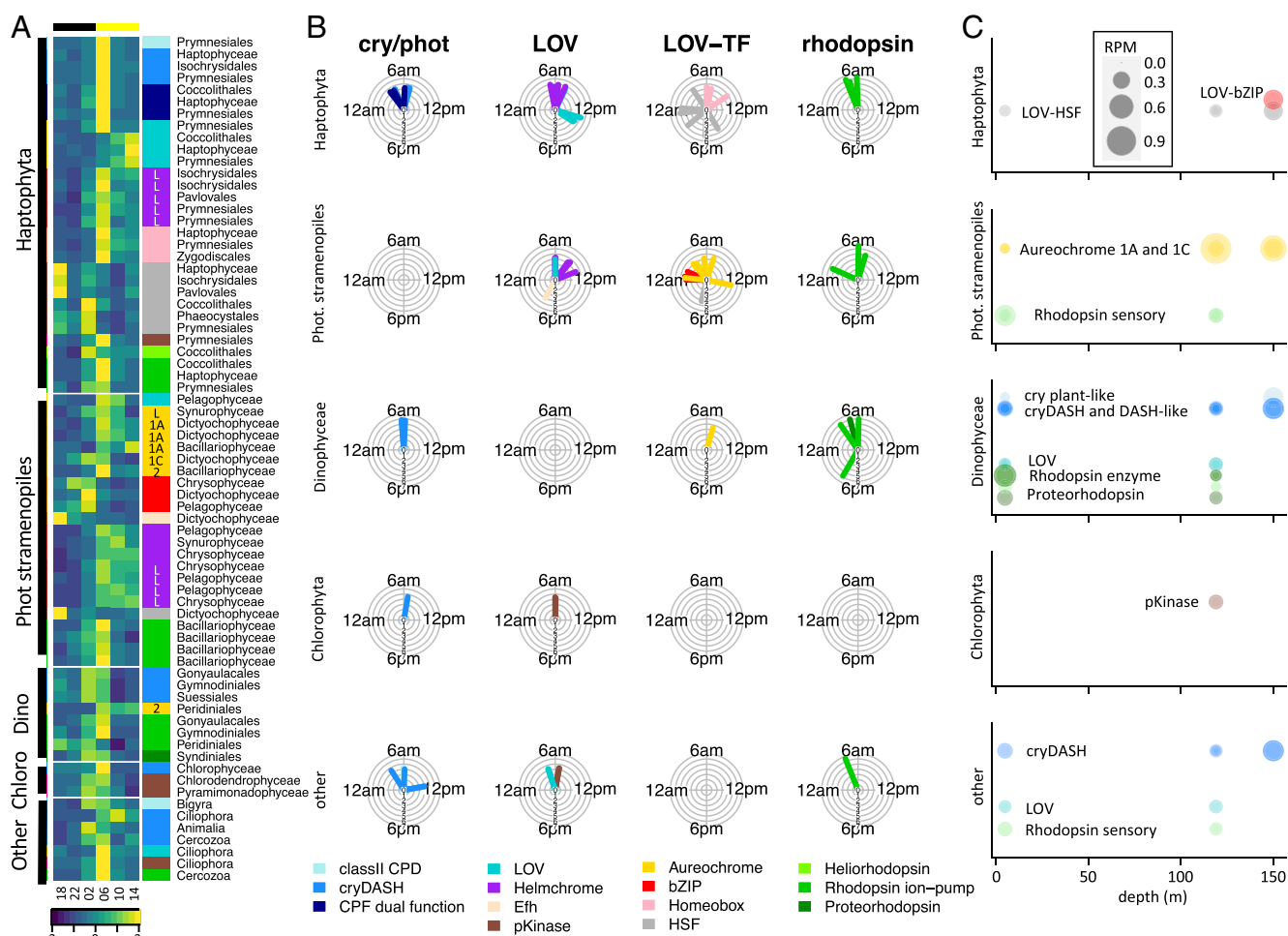
Fig. 4. Diel and depth signatures of environmental photoreceptor and other light-sensitive protein-encoding transcripts. (*A*) Abundance (*z* score-normalized) of transcripts that displayed a 24-h diel rhythmicity (RAIN; *P* < 0.001), clustered at the phylogenetic order level for the three dominant eukaryotic phyla: Haptophyta, photosynthetic stramenopiles, and Dinophyceae. The color strip next to the heat map marks the different protein types. Subtypes are indicated for aureochrome 1A, 1C, and 2 and for aureochrome- and helmchrome-like (L). Taxonomies are indicated. (*B*) Modeled peak times (mFourfit) for each class of protein with a predicted period between 23 and 25 h are indicated in the radial plots. The length of the pointer corresponds to the mean transcript abundance in log scale, with the inner ring corresponding to 10^1 reads per liter and the outer ring corresponding to 10^7 reads per liter. The pointers are color-coded by protein type as in *A*. (*C*) The top 25 most abundant light-sensitive protein-encoding transcripts at three depths (5, 119, 150 m) in metatranscriptomic datasets collected at 1800 hours on day 4 of the cruise. The protein classes are color-coded as in *A* and also written within the charts. The size of the circles corresponds to the RPM values.

cryptochrome cryDASH (Fig. 4*C* and Dataset S5*J*). Transcripts for putative transcription factors were detected for both Haptophyta and photosynthetic stramenopiles that combine the LOV domain with an HSF domain (Haptophyta) or bZIP domain (Haptophyta and photosynthetic stramenopiles). CryDASH and DASH-like transcripts were detected for Dinophyceae, as well as Rhodophyta. Low levels of Chlorophyceae transcripts associated with the LOV-containing phototropin were detected only at 150 m and were present at the limit of detection in most samples. A minority of transcripts were associated with sensory rhodopsins detected in flagellated cells (members of Pedinellales, Pinguiochrysidales, and Chlorarachnea), lending support for a role in motility and other light-dependent processes. Enzyme rhodopsin and proteorhodopsin transcripts were detected for Dinophyceae only.

Detection of putative photoreceptors and other light-sensitive protein-encoding transcripts at and below the deep chlorophyll maximum (DCM), where light is less than 1% of surface illumination, illustrates the extreme sensitivity of the open-ocean light-sensitive proteins to photon flux and spectral quality. Members of the CPF such as dual-function CPF from diatoms have previously

been shown to be responsive to short exposures of light fluence rates as low as 3.3 µmol m^{-2} s^{-1} blue light (31), comparable to available light at depth. The results presented here suggest that protists that thrive in low-light environments near the DCM are able to maintain an internal estimate of time to coordinate activities across the daily light/dark cycle.

Concluding Remarks

Our results provide an entrée into determining the molecular regulation of the observed synchronous metabolism, growth, and division of open-ocean marine protists over the daily cycle (18, 20, 21, 39, 100). We described a variety of light-sensitive proteins and LOV-domain proteins that may function in light-dependent transcriptional regulation in Haptophyta and photosynthetic stramenopiles, in phototaxis in motile photosynthetic protists, and in light-dependent processes in heterotrophic protists. The variety of photoreceptors and other light-sensitive proteins detected in plastidic protists of the subtropical open-ocean and the freshwater/wet soil alga *C. reinhardtii* (101) contrasts dramatically with that of heterotrophic eukaryotes in both marine and

Coesel et al.
Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities

PNAS | 9 of 12
https://doi.org/10.1073/pnas.2011038118

terrestrial environments, reiterating the diversity within protist communities with trophic mode as a potentially important evolutionary driving force. Our phylogenetic approaches allowed us to identify LOV-domain regulators without representatives in current reference databases. A compelling next step will be to evaluate the transcriptional patterns of photoreceptors and other light-sensitive proteins in higher-latitude regions where variations in the seasonal cycle may select planktonic organisms with different strategies to maintain diel rhythms. Additionally, the diversified protein sequences described here may provide avenues toward optogenetic approaches that use light to control the activation and deactivation of protein function.

Materials and Methods

Fieldwork Design and Sampling. Samples were collected from July 26 to 30, 2015, during the Research Vessel (R/V) *Kilo Moana* cruise KM1513 at 100 km northeast of Station ALOHA in the North Pacific Subtropical Gyre (39). A Lagrangian sampling strategy was implemented to sample the same water mass during the observational period, using free-drifting drogues centered at a depth of 15 m, as described in ref. 102. Seawater sampling was conducted using a 24 × 12 L Niskin bottle rosette attached to a conductivity–temperature–depth package (SBE 911Plus; SeaBird). The mixed-layer depth was defined based on a seawater density offset of 0.03 kg m^{-3} from 10 db (45). Incident irradiance (400- to 700-nm wavelength band) at the sea surface was measured using a LI-COR LI-1000 data logger and cosine collector. Vertical irradiance profiles were obtained on July 30, 2015, at 1211 hours with a Free-falling Optical Profiler (Hyperpro; Satlantic). SeaFlow (103) was used to make continuous measurements of small phytoplankton abundance (<5 μm in equivalent spherical diameter) at a depth of 7 m via the ship's seawater intake system. Cell diameters of individual cells were estimated from forward-angle light scatter by the application of Mie theory for spherical particles using an index of refraction of 1.032, which were in good agreement with cell diameters measured independently in phytoplankton cultures (42). Duplicate samples for metatranscriptome analysis were taken every 4 h for a period of 4 d at 15 m depth, prefiltered through 100-μm Nitex mesh and collected on a 0.2-μm polycarbonate filter (Sterlitech), as described in refs. 41 and 102. Total RNA was extracted using the ToTALLY RNA Kit (Invitrogen) spiked with a set of 14 internal RNA standards (104) that included 8 standards synthesized with poly(A) tails to mimic eukaryotic messenger RNAs (mRNAs). Poly(A)-selected mRNAs were used for Illumina NextSeq 500 sequencing, and raw sequences were quality-controlled and processed as in described in ref. 41 and 102, resulting in 2,426,923,906 merged sequence fragments with a median length of ~240 bp. Data are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject ID PRJNA492142 (105). Sequences were translated into six-frame amino acid peptides with transeq version EMBOSS:6.6.0.059 (106) using Standard Genetic Code. Peptide sequences with open reading frames greater than 40 amino acid residues were retained for downstream analyses (107). Duplicate samples for metatranscriptome analysis at 5, 119, and 150 m depth were taken on July 30, 2015, at 1800 hours. Cells were filtered through an 80-μm Nitex mesh and collected on a 0.7-μm GF/F filter. RNA extraction (Qiagen) and sequencing [Illumina HiSeq High Output 125-bp paired-end sequencing] were performed as described (108). Synthetic mRNA standards were not added to these samples and data are presented in reads per million (RPM). Data are publicly available: SRA BioProj PRJNA406025; BioSamples: SAMN07647714 to SAMN07647718. Sequence overlap was not sufficient to merge forward and reversed paired-end Illumina reads. Therefore, the forward reads only were translated into amino acid as described above and used for subsequent analysis.

Bioinformatics Pipeline. To classify environmental short reads with homology to cryptochrome/photolyase, phytochrome, rhodopsin, and LOV-domain proteins, the sequences were placed on fixed reference trees via a reference alignment using the phylogenetic placement approach (pplacer) (47). This method is particularly suited to analyze large volumes of metatranscriptome data. An overview of the analytical pipeline is as follows, with specific steps detailed in subsequent sections. First, reference sequences are retrieved from established phylogenetic studies and used to generate an hmm-profile spanning the homologous region of the protein of interest. Second, reference sequences within relevant marine databases are identified with the hmm-profile and used to generate a phylogenetic tree of reference sequences. Third, environmental short reads with homology to the hmm-profile are placed on the fixed reference tree using the maximum-likelihood approach. The reference trees provide the phylogenetic framework for both taxonomic and functional annotation (when available) of the

environmental reads based on their respective placement within the framework. The use of relatively lenient stringencies in homology searches enables identification of homologous sequences in genetically divergent marine protist species, as well as the identification of potentially novel elements not present in available model organisms. Placements of sequences of interest can be inspected at the amino acid level within the phylogenetic framework.

Marine-Relevant Photoreceptor and Other Light-Sensitive Protein-Related Reference Trees. Cryptochrome/photolyase, phytochrome, rhodopsin, and LOV protein sequences described in the literature (32, 57, 71, 72, 109) were aligned with Multiple Alignment using Fast Fourier Transform (MAFFT) version 7.313 (parameters: –localpair–maxiterate 100–reorder–leavegappyregion) (110) and used to generate hmm-profiles (Dataset S2). Hmmsearches (HMMER version 3.1b2; parameters: -E 0.001) (46) on a reference database containing 907 marine-relevant genomes and transcriptomes obtained through the Joint Genome Institute, NCBI, the Marine Microbial Eukaryote Transcriptome Sequence Project (111), and Roseobase.org, representing a total of 557 unique taxonomic reference organisms (Dataset S1). hmm-identified reference sequences were clustered at 80% identity for cryptochrome/photolyase and rhodopsin, 90% for LOV and 99% for phytochrome (Dataset S3) using usearch version 10.0.240_i86osx32 (112) and aligned with MAFFT using the same parameters as above. Clustering levels were custom set for each alignment to minimize redundancy at the taxonomic species level. Gaps in the sequence alignments were trimmed using trimAl version 1.4.rev15 (parameters: -gt 0.1) (113), and the best-fit amino acid substitution matrix for each alignment was determined using Prottest version 3.4.2 (114). hmm-identified marine reference sequences with a length shorter than the shortest sequence found within the respective hmm-profile were removed. Approximate maximum-likelihood phylogenetic reference trees were generated using FastTree version 2.1.9 (parameters: -wag for cryptochrome/photolyase and phytochrome, -lg for rhodopsin and LOV) (115). Trees were taxonomically visualized and explored with Archaeopteryx version 1.0 (116). Full-length reference sequences were queried against NCBI's conserved domain database (CDD) (73) for functional domain annotation using NCBI's Batch Web CD-search tool (Dataset S6; parameters: e-value < 0.01). The phylogenetic reference trees were functionally annotated based on CD annotations as well as homology to experimentally characterized reference sequences from the literature.

Phylogenetic Placement Analysis of Environmental Reads. Environmental metatranscriptome reads with homology to one of the four light-sensitive protein types were recruited to the hmm-profiles described in *Marine-Relevant Photoreceptor and Other Light-Sensitive Protein-Related Reference Trees.* (Dataset S2) using hmmsearch (parameters: -E 0.001) (46) and hmm-aligned to their respective reference alignment. Phylogenetic placement analysis (pplacer) (47) was used to assign NCBI taxonomy identification numbers to each environmental sequence using pplacer version 1.1.alpha17-6-g5cecf99 based on the read placement with the best maximum-likelihood score to the reference tree (parameters: –keep-at-most 1,–max-pend 0.7), as in refs. 41 and 102. Functional assignments were based on tree-edge annotations (Dataset S3). A synthetic metatranscriptome dataset of genome-derived gene models of *T. pseudonana*, *Thalassiosira oceanica*, *Emiliania huxleyi*, *C. reinhardtii*, and *P. tricornutum* reads was generated by Grinder (117) (parameters: -coverage_fold 10, -read_dist 80 normal 7, -fastq_output 0, -qual_levels 30 25,-unidirectional 1, -mutation_dist poly4 3e-3 3.3e-8) and used for independent assessment of phylogenetic placement of reads with known taxonomic origin (*SI Appendix*, Fig. S5 and Dataset S4). Read counts for each edge (maximum pendant length, <0.7; e.g., the branch length for the placement edge of the short reads) were normalized based on the internal mRNA standards to estimate environmental transcript abundance per liter seawater (Dataset S5 A–D) (104). Statistical significance for diel periodicity of transcripts aggregated on type and taxonomic-order level was determined using the RAIN package in R (48), and *P* values of <0.001 were considered significantly diel (Dataset S5 E–H). The curve-fitting method MFourfit was applied to model peak phasing, period, and amplitude using Biodare2 (biodare2.ed.ac.uk) (118) (Dataset S5I)

Maximum-Likelihood Phylogenetic Trees. A maximum-likelihood (ML) phylogenetic 18S rDNA tree representing 117 marine relevant eukaryotic order levels was built using RAxML version 8.2.8 (119) (parameters: -f a -m GTRGAMMA -p 12345 -x 12345 -# 100). 18S rDNA input sequences (Dataset S7A; one representative taxon per order level) with a minimum length of 1,543 bp were aligned with MAFFT (parameters: –localpair–maxiterate 100–reorder–leavegappyregion) (110), and gaps were removed with trimAl version 1.4.rev15 (parameters: -gt 0.5) (113). Taxonomies that assumed ambiguous positions in the ML tree were removed using RogueNaRok version 1.0 (120).

Environmental contigs were assembled with Trinity (121) version 2.3.2 on the Pittsburgh Supercomputing Center's Bridges Large Memory system

(parameters: –normalize_reads–min_kmer_cov 2–min_contig_length 300) (The Extreme Science and Engineering Discovery Environment; XSEDE) (122). All assemblies were subjected to quality-control analysis via Transrate (123) version 1.0.3 using their paired-end assembly method. Quality-controlled contigs were translated in six frames with transeq (106) version EMBOSS:6.6.0.059 using Standard Genetic Code. The longest open reading frame from each contig was retained and clustered at the 99% identity threshold level with linclust (124). Full-length amino acid sequences were queried against NCBI's CDD (73) for functional domain annotation using NCBI's Batch Web CD-search tool (Dataset S6C; parameters: e-value < 0.01). A maximum-likelihood phylogenetic tree was built using RAxML version 8.2.8 (119) (parameters: -f a -m PROTGAMMAILG -p 12345 -x 12345 -# 100) for diel-targeted LOV-domain reference sequences ($P < 0.001$) and environmental assembled contigs that mapped to those reference sequences with pplacer analysis (Dataset S7B). Also included are the LOV-domain homologs used to generate the LOV hmm-profile of this work (Dataset S2A) and the PAS-domain seed sequences (PF00989) derived from the Pfam database (pfam.xfam.org) (87). Tree visualizations were performed in the Interactive Tree of Life version 5 (https://itol.embl.de/) (125). WRKY-LOV homologs were identified in the Ocean Gene Atlas **Marine Atlas of Tara Ocean** Unigenes_version 1_metaT database (88) by protein basic local alignment search tool with an Expect threshold of 1E-10. Only contigs spanning both the LOV and WRKY domains were retained (Dataset S7C).

1. T. Roenneberg, M. Merrow, Circadian clocks - The fall and rise of physiology. *Nat. Rev. Mol. Cell Biol.* **6**, 965–971 (2005).
2. H. Wijnen, M. W. Young, Interplay of circadian clocks and metabolic rhythms. *Annu. Rev. Genet.* **40**, 409–448 (2006).
3. Z. B. Noordally, A. J. Millar, Clocks in algae. *Biochemistry* **54**, 171–183 (2015).
4. D. E. Somers, Phytochromes and cryptochromes in the entrainment of the Arabidopsis circadian clock. *Science* **282**, 1488–1490 (1998).
5. A. R. Cashmore, Cryptochromes: Enabling plants and animals to determine circadian time. *Cell* **114**, 537–543 (2003).
6. R. Tóth *et al.*, Circadian clock-regulated expression of phytochrome and cryptochrome genes in Arabidopsis. *Plant Physiol.* **127**, 1607–1616 (2001).
7. P. Emery, W. V. So, M. Kaneko, J. C. Hall, M. Rosbash, CRY, a Drosophila clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* **95**, 669–679 (1998).
8. P. Facella *et al.*, CRY-DASH gene expression is under the control of the circadian clock machinery in tomato. *FEBS Lett.* **580**, 4618–4624 (2006).
9. A. Falciatore, C. Bowler, The evolution and function of blue and red light photoreceptors. *Curr. Top. Dev. Biol.* **68**, 317–350 (2005).
10. W. Y. Kim *et al.*, ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nature* **449**, 356–360 (2007).
11. C. Lin, T. Todo, The cryptochromes. *Genome Biol.* **6**, 220 (2005).
12. S. T. Glantz *et al.*, Functional and topological diversity of LOV domain photoreceptors. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1442–E1451 (2016).
13. J. Li, G. Li, H. Wang, X. Wang Deng, Phytochrome signaling mechanisms. *Arabidopsis Book* **9**, e0148 (2011).
14. W. R. Briggs, J. M. Christie, Phototropins 1 and 2: Versatile plant blue-light receptors. *Trends Plant Sci.* **7**, 204–210 (2002).
15. A. E. Fortunato, R. Annunziata, M. Jaubert, J.-P. Bouly, A. Falciatore, Dealing with light: The widespread and multitasking cryptochrome/photolyase family in photosynthetic organisms. *J. Plant Physiol.* **172**, 42–54 (2015).
16. J. Marra, Phytoplankton photosynthetic response to vertical movement in a mixed layer. *Mar. Biol.* **46**, 203–208 (1978).
17. R. R. Bidigare *et al.*, Evaluation of the utility of xanthophyll cycle pigment dynamics for assessing upper ocean mixing processes at Station ALOHA. *J. Plankton Res.* **36**, 1423–1433 (2014).
18. E. A. Ottesen *et al.*, Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E488–E497 (2013).
19. F. O. Aylward *et al.*, Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5443–5448 (2015).
20. B. C. Kolody *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).
21. S. K. Hu, P. E. Connell, L. Y. Mesrop, D. A. Caron, A hard day's night: Diel shifts in microbial eukaryotic activity in the North Pacific subtropical gyre. *Front. Mar. Sci.* **5**, 205–217 (2018).
22. K. W. Becker *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.*, 10.1038/s41396-020-00793-x (2020).
23. R. Annunziata *et al.*, bHLH-PAS protein RITMO1 regulates diel biological rhythms in the marine diatom *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13137–13142 (2019).
24. B. Djouani-Tahri *et al.*, A eukaryotic LOV-histidine kinase with circadian clock function in the picoalga Ostreococcus. *Plant J.* **65**, 578–588 (2011).
25. E. M. Farré, The brown clock: Circadian rhythms in stramenopiles. *Physiol. Plant.* **169**, 430–441 (2020).
26. Q. Thommen *et al.*, Probing entrainment of Ostreococcus tauri circadian clock by green and blue light through a mathematical modeling approach. *Front. Genet.* **6**, 65 (2015).
27. D. A. Caron *et al.*, Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* **15**, 6–20 (2017).
28. F. Takahashi *et al.*, AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19625–19630 (2007).
29. M. J. J. Huysman *et al.*, AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (Phaeodactylum tricornutum). *Plant Cell* **25**, 215–228 (2013).
30. B. Schellenberger Costa *et al.*, Aureochrome 1a is involved in the photoacclimation of the diatom Phaeodactylum tricornutum. *PLoS One* **8**, e74451 (2013).
31. S. Coesel *et al.*, Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. *EMBO Rep.* **10**, 655–661 (2009).
32. P. Oliveri *et al.*, The Cryptochrome/Photolyase Family in aquatic organisms. *Mar. Genomics* **14**, 23–37 (2014).
33. E. G. Govorunova, O. A. Sineshchekov, H. Li, J. L. Spudich, Microbial rhodopsins: Diversity, mechanisms, and optogenetic applications. *Annu. Rev. Biochem.* **86**, 845–872 (2017).
34. N. J. Colley, D.-E. Nilsson, Photoreception in phytoplankton. *Integr. Comp. Biol.* **56**, 764–775 (2016).
35. N. C. Rockwell *et al.*, Eukaryotic algal phytochromes span the visible spectrum. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3871–3876 (2014).
36. D. K. Olson, S. Yoshizawa, D. Boeuf, W. Iwasaki, E. F. DeLong, Proteorhodopsin variability and distribution in the North Pacific subtropical gyre. *ISME J.* **12**, 1047–1060 (2018).
37. D. Man *et al.*, Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
38. M. Heijde *et al.*, Characterization of two members of the cryptochrome/photolyase family from Ostreococcus tauri provides insights into the origin and evolution of cryptochromes. *Plant Cell Environ.* **33**, 1614–1626 (2010).
39. S. T. Wilson *et al.*, Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium Crocosphaera. *Nat. Microbiol.* **2**, 17118 (2017).
40. H. Alexander *et al.*, Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5972–E5979 (2015).
41. K. W. Becker *et al.*, Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean. *Nat. Commun.* **9**, 5179 (2018).
42. F. Ribalet *et al.*, SeaFlow data v1, high-resolution abundance, size and biomass of small phytoplankton in the North Pacific. *Sci. Data* **6**, 277 (2019).
43. F. Henderikx Freitas *et al.*, Diel variability of bulk optical properties associated with the growth and division of small phytoplankton in the North Pacific Subtropical Gyre. *Appl. Opt.* **59**, 6702–6716 (2020).
44. E. Marañón, Cell size as a key determinant of phytoplankton metabolism and community structure. *Annu. Rev. Mar. Sci.* **7**, 241–264 (2015).
45. C. de Boyer Montégut, Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *J. Geophys. Res.* **109**, C12003 (2004).
46. S. R. Eddy, Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
47. F. A. Matsen, R. B. Kodner, E. V. Armbrust, pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**, 538 (2010).
48. P. F. Thaben, P. O. Westermark, Detecting rhythms in time series with RAIN. *J. Biol. Rhythms* **29**, 391–400 (2014).
49. S. Kojima, D. L. Shingle, C. B. Green, Post-transcriptional control of circadian rhythms. *J. Cell Sci.* **124**, 311–320 (2011).
50. J. W. Hastings, The gonyaulax clock at 50: Translational control of circadian expression. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 141–144 (2007).
51. H. Zhang *et al.*, Spliced leader RNA trans-splicing in dinoflagellates. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4618–4623 (2007).
52. S. Roy, R. Jagus, D. Morse, Translation and translational control in dinoflagellates. *Microorganisms* **6**, 30 (2018).

Coesel et al.
Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities

53. M. G. Fischer, M. J. Allen, W. H. Wilson, C. A. Suttle, Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19508–19513 (2010).

54. M. Furuta *et al.*, Chlorella virus PBCV-1 encodes a homolog of the bacteriophage T4 UV damage repair gene denV. *Appl. Environ. Microbiol.* **63**, 1551–1556 (1997).

55. V. Srinivasan, W. M. Schnitzlein, D. N. Tripathy, Fowlpox virus encodes a novel DNA repair enzyme, CPD-photolyase, that restores infectivity of UV light-damaged virus. *J. Virol.* **75**, 1681–1688 (2001).

56. N. Yutin, E. V. Koonin, Proteorhodopsin genes in giant viruses. *Biol. Direct* **7**, 34 (2012).

57. A. E. Fortunato *et al.*, Diatom phytochromes reveal the existence of far-red-light-based sensing in the ocean. *Plant Cell* **28**, 616–628 (2016).

58. W. Gehring, M. Rosbash, The coevolution of blue-light photoreception and circadian rhythms. *J. Mol. Evol.* **57** (suppl. 1), S286–S289 (2003).

59. N. Müller *et al.*, A plant cryptochrome controls key features of the *Chlamydomonas* circadian clock and its life cycle. *Plant Physiol.* **174**, 185–201 (2017).

60. S. König *et al.*, The influence of a cryptochrome on the gene expression profile in the diatom Phaeodactylum tricornutum under blue light and in darkness. *Plant Cell Physiol.* **58**, 1914–1923 (2017).

61. M. Castrillo, J. García-Martínez, J. Avalos, Light-dependent functions of the Fusarium fujikuroi CryD DASH cryptochrome in development and secondary metabolism. *Appl. Environ. Microbiol.* **79**, 2777–2788 (2013).

62. S. Franz *et al.*, Structure of the bifunctional cryptochrome aCRY from Chlamydomonas reinhardtii. *Nucleic Acids Res.* **46**, 8010–8022 (2018).

63. S. Oldemeyer, A. Z. Haddad, G. R. Fleming, Interconnection of the antenna pigment 8-HDF and flavin facilitates red-light reception in a bifunctional animal-like cryptochrome. *Biochemistry* **59**, 594–604 (2020).

64. C. H. Slamovits, N. Okamoto, L. Burri, E. R. James, P. J. Keeling, A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat. Commun.* **2**, 183–186 (2011).

65. D. M. Needham *et al.*, A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20574–20583 (2019).

66. O. A. Sineshchekov, K.-H. Jung, J. L. Spudich, Two rhodopsins mediate phototaxis to low- and high-intensity light in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8689–8694 (2002).

67. J. L. Spudich, The multitalented microbial sensory rhodopsins. *Trends Microbiol.* **14**, 480–487 (2006).

68. A. K. Sharma, J. L. Spudich, W. F. Doolittle, Microbial rhodopsins: Functional versatility and genetic mobility. *Trends Microbiol.* **14**, 463–469 (2006).

69. A. Kianianmomeni, K. Stehfest, G. Nematollahi, P. Hegemann, A. Hallmann, Channelrhodopsins of Volvox carteri are photochromic proteins that are specifically expressed in somatic cells under control of light, temperature, and the sex inducer. *Plant Physiol.* **151**, 347–366 (2009).

70. O. A. Sineshchekov *et al.*, Rhodopsin-mediated photoreception in cryptophyte flagellates. *Biophys. J.* **89**, 4310–4319 (2005).

71. A. Pushkarev *et al.*, A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558**, 595–599 (2018).

72. U. Krauss *et al.*, Distribution and phylogeny of light-oxygen-voltage-blue-light-signaling proteins in the three kingdoms of life. *J. Bacteriol.* **191**, 7234–7242 (2009).

73. A. Marchler-Bauer *et al.*, CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).

74. J. Collén *et al.*, Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5247–5252 (2013).

75. A. Ishikawa *et al.*, Distribution and phylogeny of the blue light receptors aureochromes in eukaryotes. *Planta* **230**, 543–552 (2009).

76. B. Imanian, P. J. Keeling, The dinoflagellates Durinskia baltica and Kryptoperidinium foliaceum retain functionally overlapping mitochondria from two evolutionarily distinct lineages. *BMC Evol. Biol.* **7**, 172 (2007).

77. N. Yamada *et al.*, Discovery of a kleptoplastic 'dinotom' dinoflagellate and the unique nuclear dynamics of converting kleptoplastids to permanent plastids. *Sci. Rep.* **9**, 10474 (2019).

78. P. G. Falkowski, The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).

79. E. V. Armbrust, The life of diatoms in the world's oceans. *Nature* **459**, 185–192 (2009).

80. G. Fu, C. Nagasato, S. Oka, J. M. Cock, T. Motomura, Proteomics analysis of heterogeneous flagella in brown algae (stramenopiles). *Protist* **165**, 662–675 (2014).

81. G. Fu *et al.*, Ubiquitous distribution of helmchrome in phototactic swarmers of the stramenopiles. *Protoplasma* **253**, 929–941 (2016).

82. M. Latasa, A. M. Cabello, X. A. G. Morán, R. Massana, R. Scharek, Distribution of phytoplankton groups within the deep chlorophyll maximum. *Limnol. Oceanogr.* **62**, 665–685 (2017).

83. A. M. Cabello, M. Latasa, I. Forn, X. A. G. Morán, R. Massana, Vertical distribution of major photosynthetic picoeukaryotic groups in stratified marine waters. *Environ. Microbiol.* **18**, 1578–1590 (2016).

84. S. T. Glantz *et al.*, Directly light-regulated binding of RGS-LOV photoreceptors to anionic membrane phospholipids. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7720–E7727 (2018).

85. S. Nakayama, N. D. Moncrief, R. H. Kretsinger, Evolution of EF-hand calcium-modulated proteins. II. Domains of several subfamilies have diverse evolutionary histories. *J. Mol. Evol.* **34**, 416–448 (1992).

86. S. Liu, G. L. Boulianne, The NHR domains of Neuralized and related proteins: Beyond Notch signalling. *Cell. Signal.* **29**, 62–68 (2017).

87. S. El-Gebali *et al.*, The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

88. Q. Carradec *et al.*, Data from "A global ocean atlas of eukaryotic genes." Ocean Gene Atlas. http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/. Deposited 25 January 2018.

89. M. H. Liang, J. G. Jiang, Analysis of carotenogenic genes promoters and WRKY transcription factors in response to salt stress in Dunaliella bardawil. *Sci. Rep.* **7**, 37025 (2017).

90. E. F. Yee *et al.*, Signal transduction in light-oxygen-voltage receptors lacking the adduct-forming cysteine residue. *Nat. Commun.* **6**, 10079 (2015).

91. J. Ashworth *et al.*, Genome-wide diel growth state transitions in the diatom Thalassiosira pseudonana. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7518–7523 (2013).

92. X. Shi *et al.*, Rhodopsin gene expression regulated by the light dark cycle, light spectrum and light intensity in the dinoflagellate Prorocentrum. *Front. Microbiol.* **6**, 555 (2015).

93. R. A. Bloodgood, Preferential turnover of membrane proteins in the intact Chlamydomonas flagellum. *Exp. Cell Res.* **150**, 488–493 (1984).

94. L. Song, W. L. Dentler, Flagellar protein dynamics in Chlamydomonas. *J. Biol. Chem.* **276**, 29754–29763 (2001).

95. A. Banerjee *et al.*, Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes. *Nucleic Acids Res.* **44**, 5957–5970 (2016).

96. T. Kottke, S. Oldemeyer, S. Wenzel, Y. Zou, M. Mittag, Cryptochrome photoreceptors in green algae: Unexpected versatility of mechanisms and functions. *J. Plant Physiol.* **217**, 4–14 (2017).

97. A. Hajdu *et al.*, High-level expression and phosphorylation of phytochrome B modulates flowering time in Arabidopsis. *Plant J.* **83**, 794–805 (2015).

98. R. Ishikawa, T. Shinomura, M. Takano, K. Shimamoto, Phytochrome dependent quantitative control of Hd3a transcription is the basis of the night break effect in rice flowering. *Genes Genet. Syst.* **84**, 179–184 (2009).

99. N. A. Reisdorph, G. D. Small, The CPH1 gene of Chlamydomonas reinhardtii encodes two forms of cryptochrome whose levels are controlled by light-induced proteolysis. *Plant Physiol.* **134**, 1546–1554 (2004).

100. F. O. Aylward *et al.*, Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11446–11451 (2017).

101. A. Greiner *et al.*, Targeting of photoreceptor genes in *Chlamydomonas reinhardtii* via zinc-finger nucleases and CRISPR/Cas9. *Plant Cell* **29**, 2498–2518 (2017).

102. B. P. Durham *et al.*, Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat. Microbiol.* **4**, 1706–1715 (2019).

103. J. E. Swalwell, F. Ribalet, E. V. Armbrust, SeaFlow: A novel underway flow-cytometer for continuous observations of phytoplankton in the Ocean. *Limnol. Oceanogr. Methods* **9**, 466–477 (2011).

104. B. M. Satinsky, S. M. Gifford, B. C. Crump, M. A. Moran, Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods Enzymol.* **531**, 237–250, (2013).

105. Simons Collaboration on Ocean Processes and Ecology (SCOPE) - University of Washington, Data from "Diel Eukaryotic Metatranscriptomes from the North Pacific Subtropical Gyre." NCBI Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA492142. Deposited 19 September 2018.

106. P. Rice, I. Longden, A. Bleasby, EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).

107. I. Wagner *et al.*, morFeus: A web-based program to detect remotely conserved orthologs using symmetrical best hits and orthology network scoring. *BMC Bioinformatics* **15**, 263 (2014).

108. S. K. Hu *et al.*, Shifting metabolic priorities among key protistan taxa within and below the euphotic zone. *Environ. Microbiol.* **20**, 2865–2879 (2018).

109. O. P. Ernst *et al.*, Microbial and animal rhodopsins: Structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014).

110. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

111. P. J. Keeling *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).

112. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

113. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

114. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).

115. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2–Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

116. M. V. Han, C. M. Zmasek, X. M. L. Phylo, phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**, 356 (2009).

117. F. E. Angly, D. Willner, F. Rohwer, P. Hugenholtz, G. W. Tyson, Grinder: A versatile amplicon and shotgun sequence simulator. *Nucleic Acids Res.* **40**, e94 (2012).

118. T. Zielinski, A. M. Moore, E. Troup, K. J. Halliday, A. J. Millar, Strengths and limitations of period estimation methods for circadian data. *PLoS One* **9**, e96462 (2014).

119. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

120. A. J. Aberer, D. Krompass, A. Stamatakis, Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. *Syst. Biol.* **62**, 162–166 (2013).

121. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

122. J. Towns *et al.*, XSEDE: Accelerating scientific discovery. *Comput. Sci. Eng.* **16**, 62–74 (2014).

123. R. Smith-Unna, C. Boursnell, R. Patro, J. M. Hibberd, S. Kelly, TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 1134–1144 (2016).

124. M. Steinegger, J. Söding, Clustering huge protein sequence sets in linear time. *Nat. Commun.* **9**, 2542 (2018).

125. I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).