A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators

David M. Needham[a,1], Susumu Yoshizawa[b,1], Toshiaki Hosaka[c,1], Camille Poirier[a,d], Chang Jae Choi[a,d], Elisabeth Hehenberger[a,d], Nicholas A. T. Irwin[e], Susanne Wilken[a,2], Cheuk-Man Yung[a,d], Charles Bachy[a,3], Rika Kurihara[f], Yu Nakajima[b], Keiichi Kojima[f], Tomomi Kimura-Someya[c], Guy Leonard[g], Rex R. Malmstrom[h], Daniel R. Mende[i], Daniel K. Olson[i], Yuki Sudo[f], Sebastian Sudek[a], Thomas A. Richards[g], Edward F. DeLong[i], Patrick J. Keeling[e], Alyson E. Santoro[j], Mikako Shirouzu[c], Wataru Iwasaki[b,k,4], and Alexandra Z. Worden[a,d,4]

[a]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039; [b]Atmosphere & Ocean Research Institute, University of Tokyo, Chiba 277-8564, Japan; [c]Laboratory for Protein Functional & Structural Biology, RIKEN Center for Biosystems Dynamics Research, Yokohama, Kanagawa 230-0045, Japan; [d]Ocean EcoSystems Biology Unit, GEOMAR Helmholtz Centre for Ocean Research, 24105 Kiel, Germany; [e]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; [f]Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama 700-8530, Japan; [g]Living Systems Institute, School of Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4SB, United Kingdom; [h]Department of Energy Joint Genome Institute, Walnut Creek, CA 94598; [i]Daniel K. Inouye Center for Microbial Oceanography, University of Hawaii, Manoa, Honolulu, HI 96822; [j]Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93106; and [k]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 113-0032, Japan

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved August 8, 2019 (received for review May 27, 2019)

Giant viruses are remarkable for their large genomes, often rivaling those of small bacteria, and for having genes thought exclusive to cellular life. Most isolated to date infect nonmarine protists, leaving their strategies and prevalence in marine environments largely unknown. Using eukaryotic single-cell metagenomics in the Pacific, we discovered a *Mimiviridae* lineage of giant viruses, which infects choanoflagellates, widespread protistan predators related to metazoans. The ChoanoVirus genomes are the largest yet from pelagic ecosystems, with 442 of 862 predicted proteins lacking known homologs. They are enriched in enzymes for modifying organic compounds, including degradation of chitin, an abundant polysaccharide in oceans, and they encode 3 divergent type-1 rhodopsins (VirR) with distinct evolutionary histories from those that capture sunlight in cellular organisms. One ($VirR_{DTS}$) is similar to the only other putative rhodopsin from a virus (PgV) with a known host (a marine alga). Unlike the algal virus, ChoanoViruses encode the entire pigment biosynthesis pathway and cleavage enzyme for producing the required chromophore, retinal. We demonstrate that the rhodopsin shared by ChoanoViruses and PgV binds retinal and pumps protons. Moreover, our 1.65-Å resolved $VirR_{DTS}$ crystal structure and mutational analyses exposed differences from previously characterized type-1 rhodopsins, all of which come from cellular organisms. Multiple VirR types are present in metagenomes from across surface oceans, where they are correlated with and nearly as abundant as a canonical marker gene from *Mimiviridae*. Our findings indicate that light-dependent energy transfer systems are likely common components of giant viruses of photosynthetic and phagotrophic unicellular marine eukaryotes.

giant viruses | viral evolution | marine carbon cycle | single-cell genomics | host–virus interactions

Viruses are increasingly recognized as key participants in the marine carbon cycle, short circuiting the classical flow of carbon through food chains to higher trophic levels (1–3). Much is known about how marine phages alter bacterial metabolism, such as supplementing photosynthetic machinery during infection (4, 5), and about viruses that infect protists (unicellular eukaryotes), especially photosynthetic taxa, and the auxiliary metabolic genes (AMGs) that they possess (6–8). Over the last 15 y, there has also been the remarkable discovery of viruses with large genomes (>300 Kb) that infect eukaryotes, the so-called giant viruses (9–13). Giant viruses encode numerous functions previously considered exclusive to cellular life, such as transfer RNA (tRNA) synthetases, translation initiation and elongation factors, and tRNAs. Those described so far primarily infect predatory protists that live in soils, wastewater, and freshwater, especially members of the Amoebozoa and Excavata eukaryotic supergroups, and have genomes that range up to 2.4 Mb (Fig. 1*A*) (9–13). The 6 isolated from the ocean water column, an environment where both viruses and protists have massive ecological importance (14–17), infect 3 haptophyte algal species (*Phaeocystis globosa*, *Emiliania huxleyi*, and *Chrysochromulina ericina*), 1 green alga (*Tetraselmis* sp.), 1 stramenopile alga (*Aureococcus anophagefferens*), and 1 non-photosynthetic predatory stramenopile (*Cafeteria roenbergensis*) (18–23). These marine viruses have smaller genomes, ranging from

Significance

Although viruses are well-characterized regulators of eukaryotic algae, little is known about those infecting unicellular predators in oceans. We report the largest marine virus genome yet discovered, found in a wild predatory choanoflagellate sorted away from other Pacific microbes and pursued using integration of cultivation-independent and laboratory methods. The giant virus encodes nearly 900 proteins, many unlike known proteins, others related to cellular metabolism and organic matter degradation, and 3 type-1 rhodopsins. The viral rhodopsin that is most abundant in ocean metagenomes, and also present in an algal virus, pumps protons when illuminated, akin to cellular rhodopsins that generate a proton-motive force. Giant viruses likely provision multiple host species with photoheterotrophic capacities, including predatory unicellular relatives of animals.

Author contributions: D.M.N., S.Y., E.F.D., M.S., W.I., and A.Z.W. designed research; D.M.N., S.Y., T.H., C.P., S.W., R.K., Y.N., K.K., T.K.-S., R.R.M., D.R.M., D.K.O., Y.S., S.S., T.A.R., E.F.D., P.J.K., A.E.S., W.I., and A.Z.W. performed research; D.M.N., S.Y., T.H., C.P., C.J.C., E.H., N.A.T.I., C.-M.Y., C.B., G.L., T.A.R., E.F.D., P.J.K., A.E.S., M.S., and A.Z.W. analyzed data; and D.M.N., S.Y., W.I., and A.Z.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Data accession numbers and databases in which they have been deposited are provided in Dataset S2 for all sequence data; the protein biochemical characterization/crystallography data have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 6JO0).

[1]D.M.N., S.Y., and T.H. contributed equally to this work.

[2]Present address: Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam 1090 GE, The Netherlands.

[3]Station Biologique de Roscoff, Sorbonne Université, CNRS, 29688 Roscoff, France.

[4]To whom correspondence may be addressed. Email: iwasaki@bs.s.u-tokyo.ac.jp or azworden@geomar.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1907517116/-/DCSupplemental.

First Published September 23, 2019.

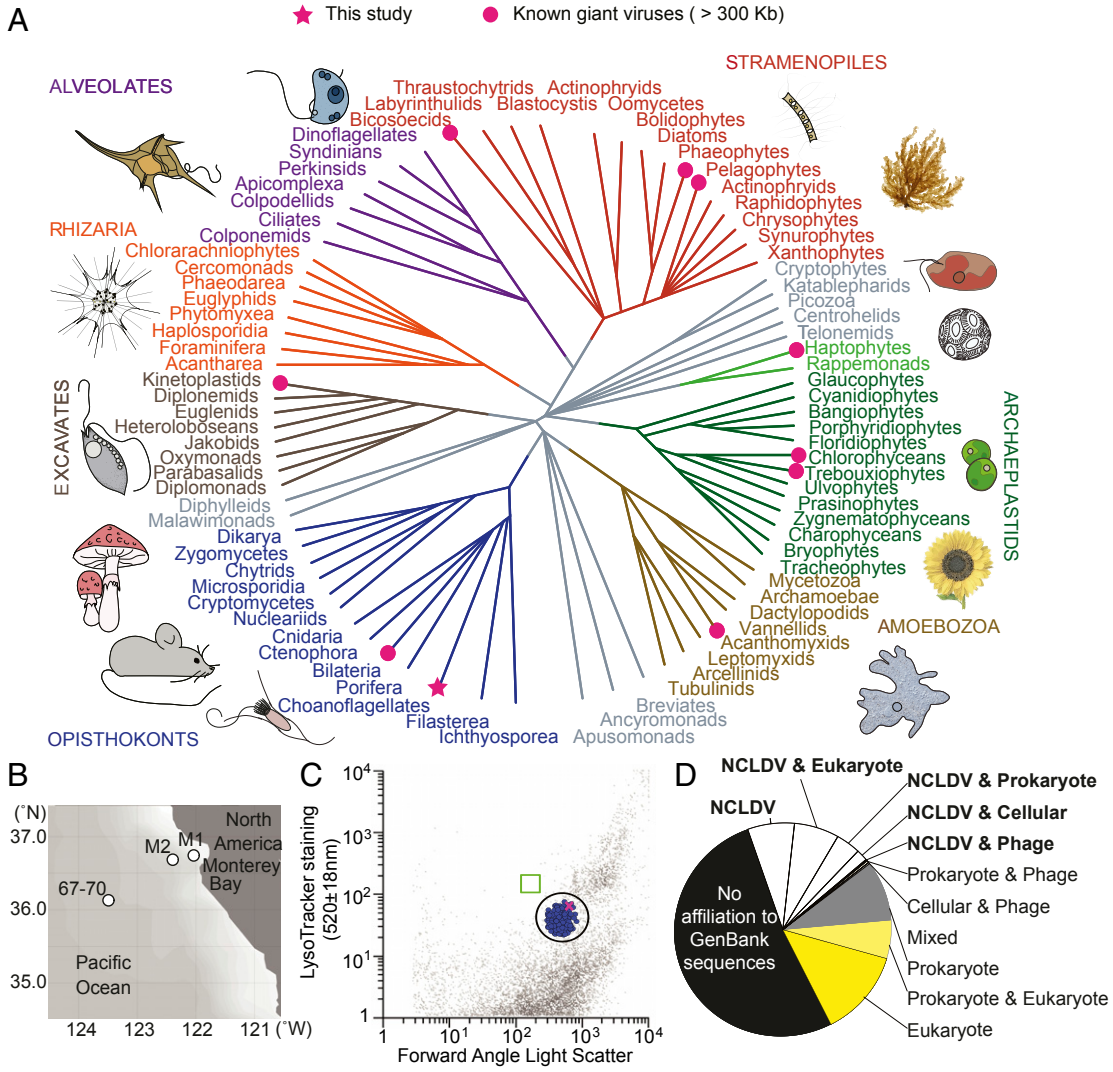

Fig. 1. A giant virus infects a predatory protist that is considered to be among the closest living unicellular relatives of metazoans. (*A*) Schematic tree of eukaryotes, with supergroups indicated by colors or gray branches if in contentious positions. Lineages with giant viruses (pink) known (circles) or discovered here (star) are indicated. (*B*) Locations of single-cell sorting where ChoanoV1 and its host, *B. minor*, were recovered (Station M2), where ChoanoV2 (Station 67-70) was found, and where metatranscriptomes were sequenced from unmanipulated seawater (M1, M2, 67-70; Station 67-155, 785 km from shore, not displayed on map for scale reasons). (*C*) Histogram showing the population (circled) of sorted choanoflagellate cells (blue dots), including the viral-infected cell (pink), based on index sorting and V4 18S rRNA gene amplicon sequencing. Other data points reflect unsorted particles in the stained seawater analyzed. The box (green) indicates the position of YG bead standards run before and after sorting at the same settings. (*D*) Categorized summary of the top 10 BLASTp matches for 862 ChoanoV1 proteins (e-value < 10^{-5}) in cellular organisms and NCLDV.

370 to 670 Kb, than many other giant viruses, and all belong to the nucleocytoplasmic large DNA viruses (NCLDV) family, which houses smaller eukaryotic marine viruses as well (24) (Dataset S1). Nevertheless, the marine giants encode a number of AMGs that connect to how they alter host metabolism during infection, such as fermentation-related genes (20) and sphingolipid-biosynthesis genes (6) in algal viruses, essential information for considering downstream biogeochemical processes and modeling the impacts of virus–host interactions on ecosystem processes.

The paucity of giant viruses isolated from marine ecosystems likely results from dependence of classical viral isolation methods on cultured hosts, such as the bacterivorous stramenopile *Cafeteria*, for recovering CroV (21). Unfortunately, many marine protists remain uncultured (15, 25) and hence, are not available for use as viral bait. This is especially so for predatory protists, in part because the natural consortia that constitute their food base are outcompeted by a few copiotrophic, relatively large bacterial taxa once in enriched medium in the laboratory (25). In some cases, metagenomics has been used to recover genome-level information while obviating cultivation. In particular, giant virus genomes have been assembled from metagenomic data acquired from low-diversity, simplified ecosystems [e.g., wastewater (12) and a hypersaline lake in Antarctica (26)]. However, these approaches are less successful in high-diversity environments, unless the biological entity has high abundance, and they fail to directly link virus to host (13), an important factor for understanding ecological impacts. To overcome these challenges, we integrated multiple culture-independent and laboratory methods to perform this cross-scale study, in which we first sorted individual wild predatory protists and used single-cell metagenomics to examine these eukaryotes and coassociated entities. With a resulting genome from an uncultured giant virus in hand, we asked how its predicted functional attributes differed from the marine giant virus genomes characterized previously, all of which come from cultivation-based isolation and sequencing, and from the plethora of giant viruses from nonmarine habitats. Furthermore,

we identified conserved attributes and established the distribution and biochemical function of a viral rhodopsin that thus far seems unique to giant viruses in the marine biosphere.

Results and Discussion

A Wild Predatory Protist in the North Pacific Ocean and Its Virus. To capture uncultured heterotrophic protists, we used high-purity fluorescence-activated cell sorting (FACS) of single cells with acidic vacuole staining to discriminate protists from prokaryotes and an additional exclusion gate against photosynthetic organisms to select heterotrophic protists only (*SI Appendix*). In a FACS survey in the eastern North Pacific, we recovered a coherent population of choanoflagellates (Fig. 1 *B* and *C*), heterotrophic predators belonging to the supergroup Opisthokonta that are considered to be among the closest living unicellular relatives of metazoans (27). Choanoflagellates comprised 99% of the 198 wells for which V4 18S ribosomal RNA (rRNA) gene amplicons were recovered after initial multiple displacement amplification of DNA from single cells, and the remaining 3 wells harbored amplicons with highest identity to uncultured syndiniales (putative parasites) and 2 different uncultured cercozoans (bacterivores), respectively. Choanoflagellates are widespread bacterivorous protists that we expected to be targeted by our staining protocol, because they contain an acidic food vacuole.

From one choanoflagellate cell, we assembled an 875-Kb viral genome after eukaryotic single-cell metagenomic sequencing (*SI Appendix*, Figs. S1 and S2). The virus, ChoanoV1, represents the largest pelagic marine giant virus genome sequenced yet; its genomic DNA base composition (GC content) was low (22%), rivaled only by nonmarine Hokoviruses (21%) and CroV (23%), whereas other giant viruses range to 64% GC (10, 12, 21) (*SI Appendix*, Fig. S2 and Dataset S1). The ChoanoV1 genome encoded 862 predicted proteins, and its gene content suggested that it belonged to the NCLDV (Fig. 1*D* and *SI Appendix*, Fig. S2), a diverse group of eukaryotic viruses (10, 11).

Presence of a eukaryotic virus coassociated with a single choanoflagellate cell could reflect several possible ecological interactions: first, that the virus had infected the choanoflagellate and replicated there; second, that the virus had been consumed by the predator as a prey item as reported in 2 prior culture-based studies on viral-feeding by predatory protists (28, 29); and third, that the virus had infected a prey item of the choanoflagellate (before that prey was consumed). Multiple lines of evidence support the first scenario. The average sequencing depth of the viral genome (215 ± 157×) and other assembly statistics (*SI Appendix*) suggested the virus was highly replicated (30) in the sorted choanoflagellate, implying there were many ChoanoV1 genomes present in the host cell. Among nonviral reads in the well, more than half belonged to the uncultured choanoflagellate *Bicosta minor*. This was determined by mapping reads against an 87-Mb partial *B. minor* genome that we generated from 4 other sort wells (*SI Appendix*, Fig. S1*B*), each containing single identical 18S rRNA gene sequences (assembled from metagenomic data and in V4 18S rRNA gene amplicons) (Dataset S2) that had 99% identity to *B. minor* as identified, handpicked, and sequenced in a prior field study (31). Contigs from bacterial prey (and phages) were also present in the choanoflagellate–virus-containing well but had a lower N50 (i.e., the minimum contig length needed to cover 50% of the genome; specifically, 13,326 vs. 86,624 for the virus genome), and none had genomes close to completion. These results suggest that the bacteria present were diverse and potentially in a degraded state as would occur in the choanoflagellate food vacuole. Additionally, the N50 of *B. minor* contigs (2,098) was lower than in wells where the virus was not detected (8,546), suggesting that it (as host) was also being degraded (*SI Appendix*, Fig. S1 *B* and *C*). While these statistics point to an active infection, it is hypothetically possible that many of the same virus had been ingested, leading to the high-coverage statistics for ChoanoV1. However, traditional metagenomic data from the same Pacific Ocean site and sort date showed that prokaryotes (prey) were >50,000 times more abundant than ChoanoV1 based on the relative numbers of bacterial 16S rRNA gene reads (a gene that is often single copy in marine bacteria) and ChoanoV1 DNA Polymerase B (PolB) reads (a single-copy gene in viral genomes). Hence, if choanoflagellates were to feed on giant viruses, the predator–prey encounter rate would strongly favor consumption of bacterial cells such that consumption of more than 1 ChoanoV1 virion is improbable. The other mechanism by which many of the virus could have been ingested is consumption of an infected small eukaryotic prey item. We did not detect sequences in the sort wells from any of the picoeukaryotes that are abundant in marine waters, including those in prior reports on this region (32, 33). Additionally, for the encounter rate of algal prey to be sufficiently higher than bacteria, one might presume that a bloom is necessary. However, Chlorophyll *a* concentrations at the depth sampled for sorting and others from the same water column and date were not indicative of a bloom; rather, the spring bloom seemed to be initiated later in the season (Dataset S3) as is typical for the region (34). Furthermore, the gene content of ChoanoV1 is highly distinct from the many available genome sequences from viruses of picoeukaryotes (35, 36) or other known algal viruses (18–20, 22, 23, 37) (*SI Appendix*, Fig. S3). Collectively, these results point to us having recovered an actively infected *B. minor* host cell in which ChoanoV1 had already replicated. After Canarypox virus, which infects birds (38), ChoanoV1 represents just the second giant virus identified with an opisthokont host (Dataset S1).

We next sought to recover a ChoanoVirus genome from another field site. Therefore, we exploited the low GC content observed in ChoanoV1 to sequence and assemble a related virus in an eastern North Pacific sample collected 200 km offshore 7 y before the *Bicosta* single-cell study (Fig. 1*B*). This sample was chosen for low %GC DNA enrichment on a density gradient, followed by deep sequencing, because environmental clone libraries showed that the *B. minor* 18S rRNA gene was present (100% identity) and vintage metagenomic data from the sample (7) contained ChoanoV1-like reads. The resulting ChoanoV2 assembly contained 89% of ChoanoV1 genes (average 94% amino acid identity), despite its fragmented nature resulting from traditional metagenome assembly limitations (*SI Appendix*, Fig. S4*A*). Our discovery poised us to investigate the evolution, function, and importance of specific metabolic traits in viruses of a key group of opisthokonts or more generally, heterotrophic marine protists and broader ecological implications.

Evolutionary Analyses Establish a Distinct NCLDV Giant Virus Lineage. Preliminary analyses suggested the ChoanoViruses were NCLDVs, with about 20% of the ChoanoV1 predicted proteins and 23% of the more fragmented ChoanoV2 proteins showing highest BLASTp affiliations to NCLDV proteins (Fig. 1*D* and *SI Appendix*, Fig. S4*B*). For proteins that had BLASTp affiliations primarily to cellular life, most of those closest to eukaryotic proteins seemed to be opisthokont derived, suggesting acquisition from hosts in past time (*SI Appendix*, Fig. S4*C*). Unfortunately, the paucity of genomic resources for marine eukaryotic viruses and marine protists themselves precludes statistically valid examination of potential horizontal or host-to-virus gene transfer (HGT) at a genome wide scale, and hence, we did not examine questions of origin globally. The other half of the ChoanoVirus proteins have not been seen in cellular organisms or viruses sequenced to date. Overall, these observations, including ~50% of proteins being unknown, are quite typical of newly sequenced NCLDV genomes (13, 39), at least at this stage in time, in which relatively few have been sequenced. Of these ChoanoVirus orphan genes, 70% were detected in metatranscriptomes that we sequenced from the eastern North Pacific, demonstrating expression (*SI Appendix*, Fig. S2).

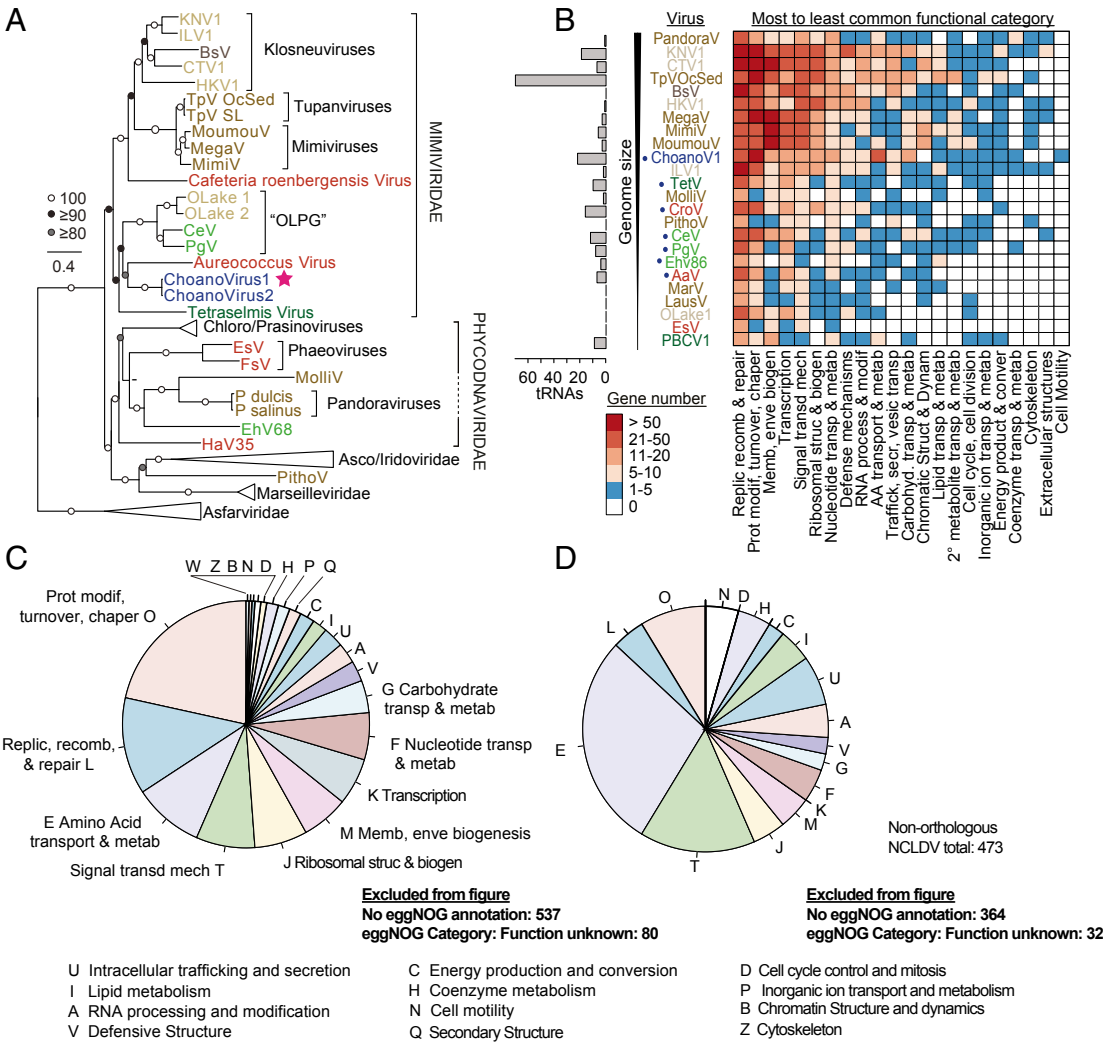

Fig. 2. Evolutionary relationships and functional aspects of the ChoanoVirus lineage. (*A*) Maximum likelihood phylogenomic reconstruction inferred from 10 proteins. Support >80% (500 bootstrap replicates) is indicated (LG + C20 + F + G-PMSF model) (*SI Appendix*, Fig. S5), and host group coloring is as in Fig. 1*A*. ChoanoV1 (star; from M2 single-cell sort) and ChoanoV2 (from Station 67-70; low %GC-selected DNA with metagenomics) branched together in all reconstructions adjacent to an algal stramenopile virus AaV (when included) (*SI Appendix*, Fig. S5), for which placement appears influenced by long-branch attraction. (*B*) Total number of tRNAs (*Left*) and orthogroup functional categorization (heat map; EggNOG categorization) of ChoanoV1 and representative giant NCLDV (Dataset S1). The frequency of each category across the viral genomes determines *x*-axis ordering. (*C*) Distribution of functional categories in ChoanoV1 (via EggNOG) for all annotated proteins. (*D*) ChoanoV1 proteins with no orthologs in the NCLDV representative genome set. Note that, in pies in *B* to *D*, we have omitted fractions representing the EggNOG functional category "Unknown function," but the values are shown as text on panels along with the total number of proteins with no significant database match.

To reconstruct evolutionary relationships, we used phylogenomic approaches to analyze proteins considered core to NCLDV genomes (40). We reexamined presence, absence, and copy number patterns for the 47 proteins previously proposed to be core (40). We next excluded, for example, fast-evolving proteins and proteins for which unclear paralogs existed within a single NCLDV genome, and thereby expanded the set of NCLDV proteins suitable for phylogenomics used in recent reconstructions (12) from 5 to 10 (Fig. 2*A* and Dataset S2). Phylogenomic reconstructions with the 2 protein sets provided similar topologies, with higher statistical node support in the 10 protein phylogeny (*SI Appendix*, Fig. S5). These reconstructions showed the ChoanoViruses belong to the extended *Mimiviridae*, comprising a divergent clade from those already established (12, 13). PolB reconstructions highlighted a large group of marine viral PolB, distinct from nonmarine *Mimiviridae* (Mimiviruses, Tupanviruses, Klosneuviruses) and CroV, when assembled metagenomic sequences from TARA Oceans (41) and Global Ocean Survey (GOS) (42) were searched and included (*SI Appendix*, Fig. S6). Within this broad marine group, the ChoanoViruses formed a supported clade that incorporated Pacific Ocean, Atlantic Ocean, and Southern Ocean sequences for which the viral hosts remain unknown. These analyses demonstrated the value of recovering viral genomes from uncultured hosts, which exposed here the unique ChoanoVirus lineage and its presence in multiple oceans.

ChoanoVirus Auxiliary Metabolic Genes and Biogeochemical Implications. AMGs are host-derived genes carried by viruses that are not directly involved in viral replication but rather supplement or augment cellular functions within infected cells (5, 6). An important example in marine bacteriophages is oxygenic photosynthesis proteins that augment cyanobacterial photosynthetic machinery during infection (4). Although oxygenic photosynthesis-related proteins have not been found in eukaryotic viruses sequenced to date, the giant viruses encode a plethora of AMGs that augment cellular processes. These include proteins involved in, for example,

translation, transcription, lipid biosynthesis, and transport of phosphate or ammonium (6–8, 43). Systematic analyses of ChoanoVirus metabolic potential revealed a broad repertoire of such proteins, several types being enriched or unique in ChoanoViruses relative to other NCLDVs (Fig. 2 *B–D* and *SI Appendix*, Figs. S3 and S7*A*). Like other giant viruses, the ChoanoVirus genomes encode proteins for augmenting host processes, including aminoacyl-tRNA synthetases, photolyases, and proteins involved in signal transduction, replication, recombination and repair, cell wall biogenesis, and posttranslational modifications (Fig. 2*B* and *SI Appendix*, Fig. S7*A*) (6–13, 18–23). The ChoanoViruses also encode 22 tRNAs (Fig. 2*B*) such that tRNA numbers seem to roughly scale with genome size, with more being found in the larger genome-sized Tupanvirus from deep sea sediment (43) and less in the smaller genome-sized pelagic marine giant viruses TetV, CroV, PgV, and CeV (18, 20–22). Furthermore, the ChoanoVirus tRNAs correspond to amino acid usage, suggesting preferential retention of those optimized for amino acid usage of virus over host, and 17 tRNAs are collocated in a single genomic region (*SI Appendix*, Fig. S7 *B* and *C*). Hence, the large ChoanoVirus genomes encoded many proteins once considered unique to cellular life, that now seem to be held in common across disparate giant viruses (10–13, 18–21).

Clustering based on presence and absence patterns of orthologous protein groups in NCLDV placed ChoanoV1 adjacent to the only other sequenced marine pelagic virus with a host that is a heterotrophic predator, CroV (*SI Appendix*, Fig. S3) (21). These 2 viruses were part of a broader cluster incorporating marine algal giant viruses, which appeared more similar to each other in their orthogroup presence and absence patterns than to nonmarine giant viruses or smaller viruses that infect marine algae. Many of the proteins making up these orthogroups lack characterized functions or have only broad functional classification. Combined with the limited overall representation of giant virus lineages, these findings call for a major initiative to expand viral taxonomic sampling so that the significance of the presence and absence pattern observations could be estimated.

Comparison of ChoanoV1 with other genome-sequenced viruses shows an enrichment in NCLDV orthologs involved in transport and metabolism of nucleotides, amino acids, and carbohydrates (Fig. 2*B* and *SI Appendix*, Fig. S7*A*). ChoanoV2 shows the same trend, although its more fragmented state precludes robust global ortholog comparisons. Even among ChoanoVirus proteins lacking orthologs in other NCLDV, these functional categories are prominent (Fig. 2*D* and *SI Appendix*, Fig. S7*D*) and include a chitinase new to marine viruses that is present in both ChoanoViruses (*SI Appendix*, Fig. S8). Chitinase degrades the polysaccharide chitin, a component of zooplankton, some algae, and many other organisms, to labile saccharides readily consumed by marine microbes (44). This enzyme has been reported in a virus of the freshwater alga *Chlorella* (45) and viruses that infect insects, specifically *Lepidoptera* (46). Our phylogenetic analyses placed moth virus chitinases in a clade with sequences from their *Lepidoptera* hosts within bacterial chitinases (potentially a complex series of transfer events), while Chlorella virus and fungal chitinases grouped together (*SI Appendix*, Fig. S8). The ChoanoVirus chitinase branched with opisthokont chitinases, suggesting potential acquisition from a host of an ancestral opisthokont virus. Collectively, these results suggest that acquisition by each of the 3 types of viruses occurred in independent events. From a functional perspective, release of viral chitinase in *Leptidoptera* larvae is necessary for liquefaction, but the mechanism and overall roles during infection are unclear (46). The Chlorella virus chitinase has hypothesized roles in degrading the chitin-rich host cell wall (45). However, in contrast to moths and *Chlorella*, which have chitin as an abundant structural component, choanoflagellates lack known chitin-based structures, although they possess chitin synthase (47). Thus, ChoanoVirus chitinase activity, potentially on prey material, alongside activities of viral carbohydrate metabolism proteins may supply hosts with nutrition when choanoflagellate feeding is impacted by the infection or other factors. Alternatively, a structural feature of choanoflagellate cells, such as the theca, may have an as yet unrecognized chitin-containing composition, in which case, the viral chitinase may operate in host degradation. Regardless, the organic matter released from the lysed host will provide more readily available carbon sources, such as labile saccharides, to marine microbes than will hosts infected and lysed by viruses that lack these enzymes or other forces of mortality. As such, in addition to release of cellular substrates on lysis, viral infection may "prime" substrates to be accessed more readily, potentially altering the microbial loop (48) in terms of rate and fate of the cellular material remineralization in the ocean.

Viral Rhodopsin Sequence Characterization. Strikingly, we also identified 3 distinct putative rhodopsins in each ChoanoVirus genome (Dataset S4). Rhodopsins are integral membrane proteins that capture or sense sunlight using a bound retinal chromophore in cellular organisms (49). Microbial (type-1) rhodopsins include a variety of light-driven ion pumps (including H^+, Cl^-, Na^-) (*SI Appendix*, Table S1) and sensory receptors involved in signal transduction (including Sensory Rhodopsins I and II, which have been shown to regulate phototaxis in some protists) (50–52). Additionally, heliorhodopsins are considered distantly related family members and are thought to have light-sensing activities (53). Type-1 proton-pumping rhodopsins are widespread in heterotrophic marine bacteria (54, 55), increasing survival during starvation when illuminated (50), and homology-based studies postulate that some eukaryotic algae have similar systems (56). Phylogenetic analyses show that the ChoanoVirus rhodopsins split into 2 type-1 groups composed primarily of metagenomic sequences, which collectively exhibit distinct phylogenetic histories from those in cellular organisms (Fig. 3*A*). Among viruses with known hosts, the only other rhodopsin reported is in the giant virus PgV, which infects the marine haptophyte alga *P. globosa* (18, 57), and belongs to a clade that includes 1 of the 3 ChoanoVirus rhodopsins (Fig. 3*A*). We term these 2 groups (clades) that have this distinct history from those of cellular organisms VirR Group-I and VirR Group-II. Importantly, all VirR are highly diverged from a microbial rhodopsin clade harboring the fusion protein Rho-PDE that is present in the genome-sequenced choanoflagellate *Salpingoeca rosetta*, wherein it exhibits light-dependent phosphodiesterase activity (58, 59). While we identified homologs of Rho-PDE in 2 transcriptome-sequenced choanoflagellate species (Fig. 3*A*), it is absent from genome-sequenced *Monosiga brevicollis* and is not found in transcriptome assemblies from 17 other choanoflagellate species or in the *Bicosta* 4-well partial genome assembly. Overall, the ChoanoVirus VirR proteins do not seem to be derived from extant opisthokonts. Indeed, the tree topology and additional testing (*SI Appendix*) suggest that rhodopsin may have been present in an ancestral virus before host-range expansion into disparate algae and heterotrophs (Fig. 3*A*).

Several marine studies have now reported putative viral rhodopsins in traditional metagenomic data—for which the viral hosts are by default unknown (57, 60–62). The function of these is not clear, since often, they lack the amino acid motifs that have been shown through biochemical characterization of various type-1 rhodopsins to generally confer functional differences. Indeed, the function of type-1 rhodopsins can sometimes be inferred from 3 key amino acid residues (referred to as motif sequences), such as the proton (DTD, DTE) and chloride (TSA, NTQ) pump motifs (49). In bacteriorhodopsin (BR), the residues that make up the motif are at positions 85, 89, and 96. BR has been biochemically characterized to function as a proton pump, wherein the D85 acts as a proton acceptor, T89 forms a

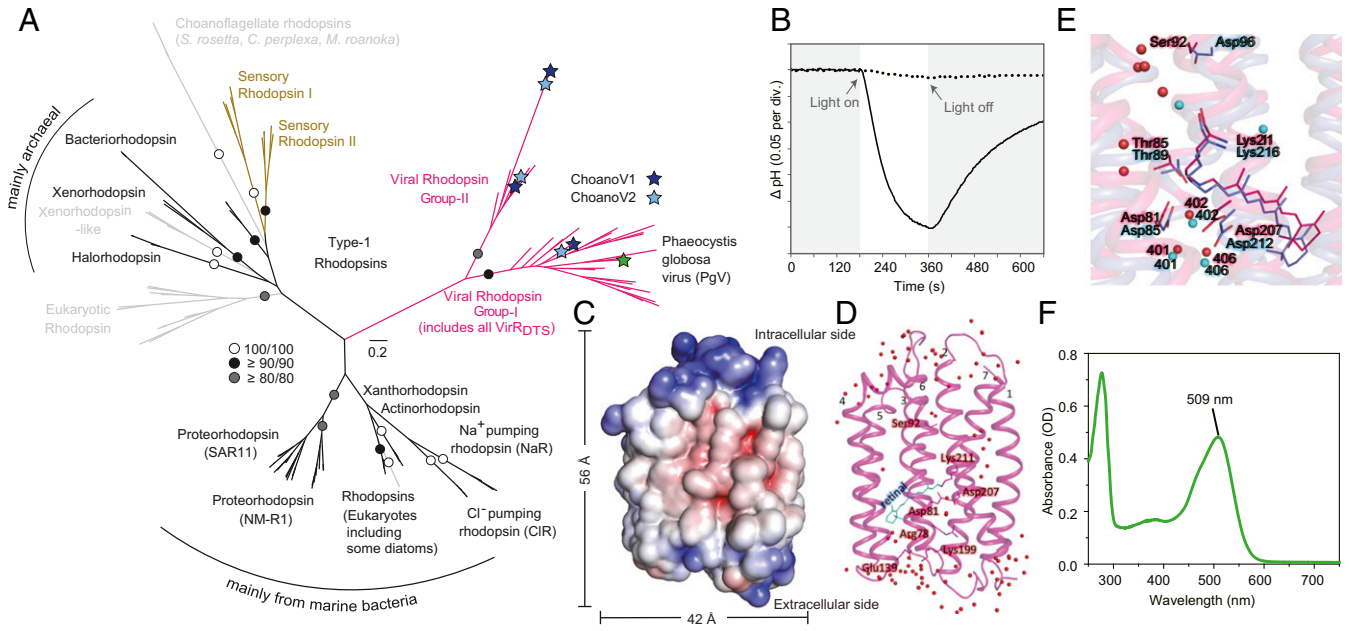

Fig. 3. Evolution, structure, and function of viral rhodopsins. (*A*) Maximum likelihood phylogenetic reconstruction of bacterial, archaeal, eukaryotic, and viral rhodopsins. Viral (pink), nonviral ion-pumping (black), nonviral sensory (brown) and limited information or unclear function nonviral (gray) rhodopsins are indicated and support >80% (1,000 bootstrap replicates). Sensory rhodopsins present in the choanoflagellate *S. rosetta* (59) and detected here in *Choanoeca perplexa* and *Microstomoeca roanoka*, but not other choanoflagellates (27) or *B. minor*, have a fused phosphodiesterase region and are distant from ChoanoVirus VirR proteins. Metagenomic sequences from a sediment study reporting 30 PgV and Organic Lake virus-like VirR (62) could not be included, because they are not in GenBank, were not recovered in the IMG (Integrated Microbial Genomes) database, and, based on statistics provided, were largely partial length. This prior study recovered different VirR topologies using maximum likelihood vs. trait-informed Bayesian reconstructions that also differed from our highly supported topology, indicating that conclusions (62) regarding identification of a putative ancestor of viral rhodopsins should be revisited. Channelrhodopsins (52) were not included due to high divergence that resulted in the loss of many positions for type-1 phylogenetic analysis. Additionally, heliorhodopsins were excluded, because they are too divergent from the microbial type-1 rhodopsins. (*B*) Light-induced acidification of medium containing *E. coli*-expressing $VirR_{DTS}$ in the presence of the chromophore retinal (solid line) and its abolishment by protonophore addition (i.e., carbonyl cyanide m-chlorophenyl hydrazone [[(3-chlorophenyl)hydrazono]malononitrile (CCCP)); dotted line). (*C*) Surface representation of the 1.65-Å resolution $VirR_{DTS}$ crystal structure with electrostatic potential indicated (red, negative; blue positive) as viewed parallel to the membrane. (*D*) Ribbon diagram showing the retinal (light blue lines), H_2O molecules (red spheres), and 7 TM α-helices connected by 3 cytoplasmic loops, 3 extracellular loops, and short helices between TM3 and TM4. Numbers denote TM domains. (*E*) $VirR_{DTS}$ (magenta) and *H. salinarum* proton-pumping BR (71) (purple; Protein Data Bank ID code 1C3W; 21% amino acid identity) structural comparison. Key residues (teal, BR; red, $VirR_{DTS}$) and H_2O molecules (spheres) are indicated. (*F*) $VirR_{DTS}$ absorption spectrum.

hydrogen bond with D85, and D96 acts as a proton donor in this DTD motif rhodopsin (49); other motifs have proton pumping or other functions (*SI Appendix*, Table S1). Previously detected VirR sequences in PgV and GOS were hypothesized to have sensory roles in host phototaxis (57) or to be involved in light sensing in the host (61), because some lack the retinylidence Schiff base proton donor carboxylate, which has been taken to be essential for proton transport, similar to sensory rhodopsins (63). However, recent work has shown that some rhodopsins lacking the proton donor carboxylate do pump protons (64). Based on in silico transmembrane predictions (TMHMM, a method for prediction of transmembrane domains based on hidden markov models), the 3 different rhodopsin proteins in the ChoanoViruses each have 7 transmembrane (TM) domains, as expected (49), and we detected transcripts for 2 of 3 in eastern North Pacific metatranscriptomes from Stations M1 and M2 (Fig. 1*B*), demonstrating their expression (*SI Appendix*, Fig. S2). The Viral Group-I rhodopsin present in each ChoanoVirus and in PgV has a DTS motif ($VirR_{DTS}$) (Fig. 3*A*). Prokaryotic DTS-motif rhodopsins have been reported in proton-pumping clades (e.g., the proteorhodopsin [PR] clade and DTG-motif clade) and the xenorhodopsin clade (e.g., *Anabaena* sensory rhodopsin, ASR) of sensory rhodopsins, indicating that information on the motif sequence alone is not enough to predict function (65, 66). The motifs of the ChoanoVirus Group-II rhodopsins, DTV and YML, are not present in functionally characterized rhodopsins (*SI Appendix*, Fig. S9). The bacterium *Thermochromatium tedium* has a YTM motif, with some similarity to YML, that is predicted to be sensor type but as yet not functionally characterized (67). Unlike the observed YML motif, the DTS and DTV motifs have been observed in environmental sequences inferred to come from viruses at Station ALOHA in the North Pacific Gyre (60), in the Red Sea (61), and in coastal sediments (62). Our results provided evidence for VirR proteins being in viruses of heterotrophic protists and for a single virus having both Group-I and Group-II viral rhodopsins. However, the amino acid differences for all VirR from biochemically characterized proteins alongside their long-branch lengths (Fig. 3*A*) left uncertainty regarding function, as is the case for many proteins identified in marine metagenomic studies.

Viral Rhodopsin Activity and Structure. Because of the presence of $VirR_{DTS}$ in the only pelagic marine giant viruses with known hosts (i.e., the uncultured ChoanoViruses and the cultured algal virus PgV), we next turned to laboratory experiments to examine the structure and function of this VirR protein. Heterologous expression in *Escherichia coli* of the homolog from PgV caused substantial light-induced acidification of retinal-amended medium up on illumination, demonstrating that it has proton-pumping capabilities (Fig. 3*B*). This clear pH change was abolished by protonophore addition. $VirR_{DTS}$ predominantly possessed all-*trans* retinal (*SI Appendix*, Fig. S10*A*). At neutral pH, the Schiff base linkage was protonated ($pK_a = 7.8$), and a counterion residue was deprotonated ($pK_a = 3.6$) (*SI Appendix*, Fig. S10 *B* and *C*). We analyzed the photocycle of $VirR_{DTS}$, demonstrating that time constant of

recovery from the O540 intermediate to the original state was 386 ms (*SI Appendix*, *SI Results and Discussion* and Fig. S10 *D* and *E*). This recovery time is longer than that of BR from *Halobacterium salinarum* (BR, t = 10 ms), an archaeal proton-pumping rhodopsin, but similar to proton-pumping rhodopsins from other taxa, such as BR from *Haloquadratum walsbyi* (~300 ms), thermophilic rhodopsin from *Thermus thermophilus* (277 ms), and PRs from a number of marine bacteria (PRs; ~250 ms) (*SI Appendix*, Table S2) (68–70).

Because $VirR_{DTS}$ is divergent from characterized light-driven proton-pumping rhodopsins and no viral rhodopsin structure is known, we next dissected how it pumps protons. The crystal structure of the cell-free synthesized $VirR_{DTS}$ was determined at 1.65-Å resolution, revealing broad-scale similarities to BR (Fig. 3 *C* and *D* and *SI Appendix*, *SI Results and Discussion* and Fig. S11 *A–E*) (71). The root-mean-square deviation (RMSD) was 1.83 Å, while adoption of a different structure from BR was observed in the loops, especially the TM3–TM4 short helix. The pentagonal cluster formed by 3 water molecules (Wat401, -402, and -406), Asp81, and Asp207, corresponding to the most important region for BR proton pumping, did have a similar structure to that of BR (Fig. 3*E*). Electron densities around the retinal showed that it is in all-*trans* conformation, covalently attached to Lys211. We then examined several residues that hold key positions in $VirR_{DTS}$ and other opsins (*SI Appendix*, Fig. S11 *B–I*), including Asp81 and Ser92, which are similarly positioned to Asp85 and Asp96 of the BR DTD-motif group (71) (Fig. 3*E*). Mutation analyses of these and other residues established their essentiality for proton-pumping activity, especially the proton acceptor residue Asp81 (*SI Appendix*, Fig. S11*J*). In addition, we showed that maximal $VirR_{DTS}$ absorption is in the green wavelengths (Fig. 3*F*).

Finally, we compared the $VirR_{DTS}$ structure with 2 typical structures of sensory rhodopsins: ASR (from Anabaena) and SRII, the *Natronomonas pharaonis* sensory rhodopsin II (*SI Appendix*, Fig. S11 *F* and *G*) (66, 72). Given our data, it seems that $VirR_{DTS}$ is a proton-pumping opsin; however, it is possible that it could have a sensory function as previously proposed based on sequence data (61). There is much debate about interpretation of sequence data alone as well as photocycle data and its comparability when conducted using different conditions. Hence, ultimately, in vivo manipulation in the proper cell biological context will be needed to determine overall function. Our in silico comparisons show that the overall structures of ASR and SRII have similarities to that of $VirR_{DTS}$, with RMSDs of 1.94 and 2.22 Å, respectively. While the positions of Ser92 (corresponding to Ser86 in ASR) are similar between $VirR_{DTS}$ and ASR, the water molecule and amino acid positions around the retinal adopt quite different structures (*SI Appendix*, Fig. S11*F*). Likewise, these aspects of ASR positions are different from BR (*SI Appendix*, Fig. S11*H*). However, the corresponding portion of SRII is similar to that of $VirR_{DTS}$ and BR (*SI Appendix*, Fig. S11 *G* and *I*). Our searches for the proteins required for signal transduction by sensory rhodopsins using queries known to fulfill this function (e.g., HtrI and HtrII [73]) did not recover related proteins in either the *Bicosta* 4-well assembly or the ChoanoVirus genomes. The viral rhodopsins also lack fusions of known transducer-related domains that occur in eukaryotic sensory rhodopsins (74), although notably, $VirR_{YML}$ has an N-terminal domain of unknown function. Furthermore, a fusion protein integrating a rhodopsin and phosphodiesterase (RhoPDE; also discussed above, Fig. 3*A*) was recently discovered in *S. rosetta*, which, like other choanoflagellates, lacks an eyespot or other known light-sensory structures (58, 59, 75). While we found phosphodiesterases in *Bicosta*, again, no rhodopsin (or related fusion protein) was recovered, and we did not find these proteins in *M. brevicollis* or 17 of 19 transcriptome-sequenced choanoflagellates (27). Thus, if the viral rhodopsin was a sensory rhodopsin, the potential mechanisms by which it operates remain elusive as are the biological implications. These observations indicate that motifs, monomeric structures, or photocycle data are individually not enough to determine whether a rhodopsin functions as a pump or sensor. Collectively, our results show that $VirR_{DTS}$ is a green light-absorbing proton pump that has a structure similar to that of BR and transfers light energy in a manner that substantially changes medium pH when expressed in a cell.

A Viral Chromophore Biosynthesis Pathway. Demonstration of $VirR_{DTS}$ proton-pumping activity on illumination raises questions regarding the natural source of the carotenoids needed to produce the light-harvesting chromophore, retinal (50, 51), especially in a nonphotosynthetic host, like *Bicosta*. Most algae, including PgV's host *Phaeocystis*, biosynthesize the required pigment, β-carotene (and related carotenoids), as well as the retinal-producing carotenoid cleavage oxygenase (Blh) (Fig. 4). However, most heterotrophic eukaryotes, including animals, do not biosynthesize β-carotene, instead acquiring carotenoids through diet. As expected, cultured genome-sequenced choanoflagellates encode only early steps that overlap between sterol and carotenoid biosynthesis and a final cleavage enzyme (Dataset S5). Likewise, BLASTx searches against the *Bicosta* 4-well partial genome assembly failed to recover carotenoid biosynthesis enzymes. Remarkably, the ChoanoVirus genome analyses exposed both the β-carotene biosynthesis pathway and Blh, with 4 proteins being adjacent to one another, similar to the pathway in bacteria (76) (Fig. 4, *SI Appendix*, Fig. S12*A*, and Dataset S5). Eastern North Pacific metatranscriptomes confirmed expression of all components (Fig. 4). Thus, while the algal virus relies on its host to biosynthesize the pigment used in light-energy transfer, ChoanoViruses encode the complete rhodopsin-based photosystem.

The evolutionary origins of the retinal biosynthesis proteins in the ChoanoViruses remain unclear. They seem to derive from

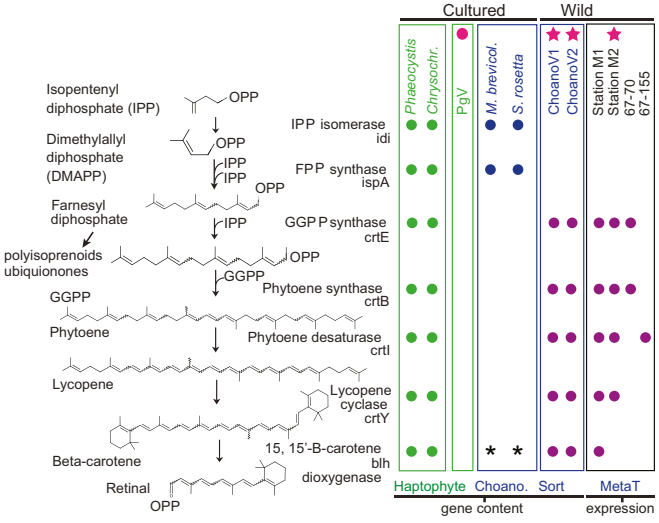

Fig. 4. Functional attributes of ChoanoViruses include chromophore biosynthesis. Shown are carotenoid pathway components and final retinal-forming cleavage step in genome data from haptophytes (*Phaeocystis antarctica* and *Chrysochromulina* representing *P. globosa*, which lacks genome data), choanoflagellates (*M. brevicollis* and *S. rosetta*), and relevant viruses and in metatranscriptomes. The stars indicate the two ChoanoVirus genomes and a metatranscriptome from the station where ChoanoV1 was recovered. The circle indicates the only cultured virus with a rhodopsin. *These taxa lack Blh but have RPE65 used for retinal production (e.g., in vertebrates and relatives). Detection in Pacific metatranscriptomes based on reads recruited to ChoanoV1 by BLASTx (e-value < 10^{-10}); those that mapped at >95% nucleotide identity are indicated in Dataset S5. OPP, pyrophosphate group; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate.

archaea (phytoene synthase) or marine bacteria (phytoene desaturase) or are too divergent for robust phylogenetic conclusions (lycopene cyclase, Blh) (*SI Appendix*, Fig. S12 *B–E*). In each case, the respective ChoanoV1 and ChoanoV2 proteins clustered together, indicating their common origin. Rhodopsin-bearing bacterial or archaeal lineages with retinal biosynthesis-related genes are each thought to have acquired them together as a unit by HGT (77). However, despite the 4 ChoanoVirus retinal biosynthesis genes being colocated in the genome, long branch lengths and incomplete taxonomic sampling make it unclear whether these proteins were accumulated over time or acquired in a single HGT event, although the latter scenario seems most likely.

Viral Rhodopsins in the Global Ocean. Our studies now provided the structure and function of $VirR_{DTS}$, but the frequency of VirR genes as a whole in nature remained unclear. Prior analyses of viral rhodopsins in traditional metagenomic data focused on individual locations, specifically the Red Sea (61) and Station ALOHA (60), or had relatively shallow sequencing depth, such as GOS (57). It should be noted that one other metagenomic study of coastal sediments reported 30 VirR (62) that were similar to PgV $VirR_{DTS}$ and to the VirR metagenomic sequences from Organic Lake that have been suggested to come from another (currently unknown) haptophyte algal virus. These partial metagenomic sequences (62) may well, therefore, represent remnants of a senesced (infected) haptophyte bloom exported to sediments at 11- to 50-m bottom depth. Our searches of TARA metagenomic assemblies greatly expanded the global VirR repertoire (Fig. 5*A* and Dataset S6). Assembled VirR proteins were recovered at 37 of 39 TARA photic-zone sampling sites examined, and only at photic-zone depths in Station ALOHA profiles that included deep ocean sampling (Fig. 5*B*), as expected for a sunlight-dependent energy transfer system. Motifs were diverse; however, the DTS motif was the most common vertically and globally (Fig. 5 *B* and *C*).

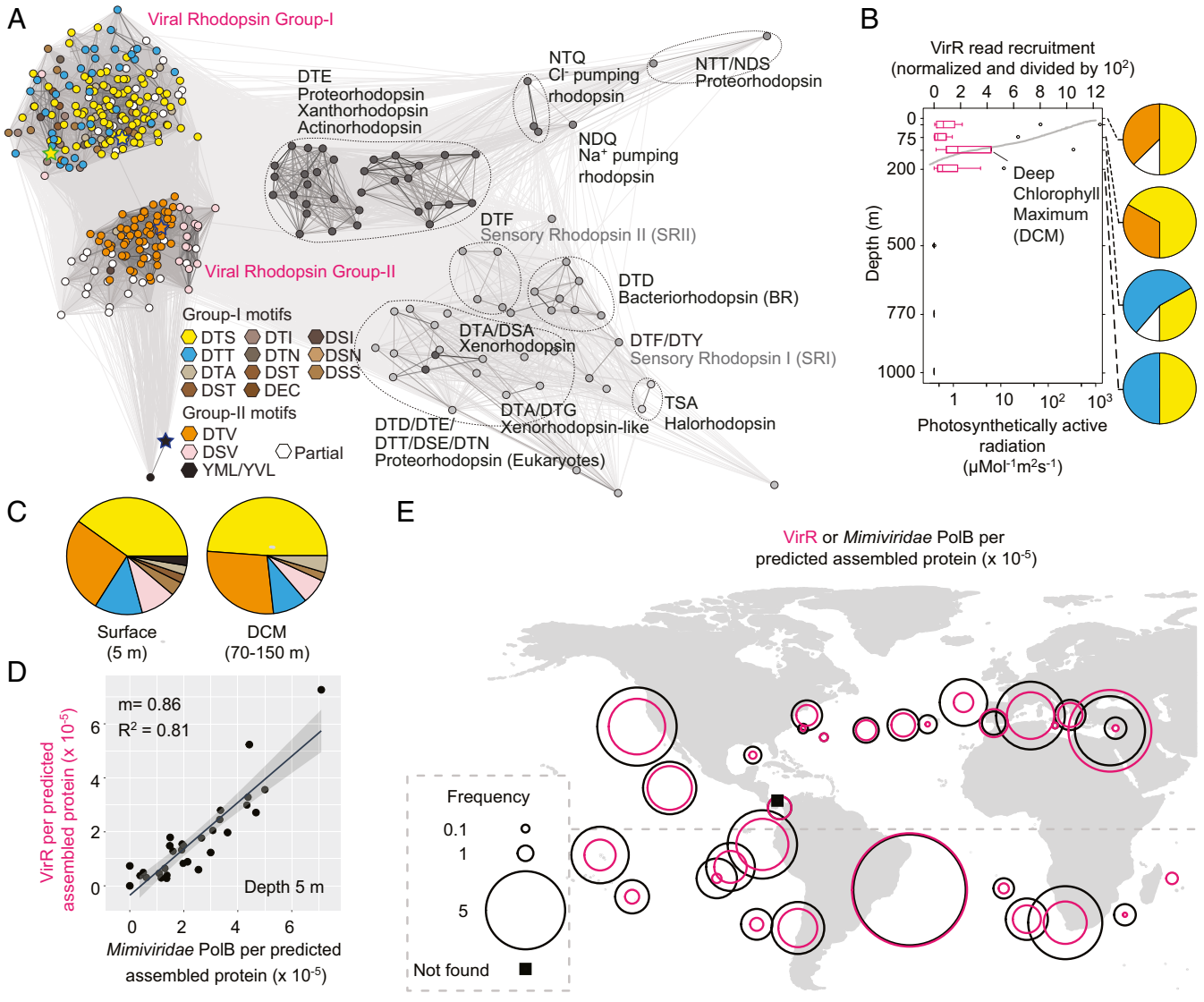

Fig. 5. Viral rhodopsins are distributed across the world oceans. (*A*) Environmental VirR motifs and cluster analysis of sequences (CLANS)-based relationships between full-length proteins recruited from TARA Oceans and Station ALOHA data. (*B*) Normalized VirR depth distributions in the North Pacific Gyre determined by mapping metagenomic reads to VirR gene assemblies from ALOHA (60) and VirR motif distributions (pies; colors as in *A*). (*C*) VirR motifs in TARA metagenome assemblies having >300,000 contigs from 5 m (304 full-length sequences in total) and samples reflecting a true deep chlorophyll maximum (43 full-length sequences in total), which typically occurs in stratified open ocean water columns between 75 and 130 m. (*D*) Correlation between *Mimiviridae* PolB and VirR across analyzed TARA samples. (*E*) Normalized VirR and *Mimiviridae* PolB frequencies in TARA assemblies (with >300,000 contigs).

Phylogenetic analyses of the assembled sequences that we recovered from deeply sequenced TARA samples and other metagenomic studies showed 3 statistically supported clades within Group-I VirR proteins: one harboring only DTT (64%) and DTS (36%) motifs, another dominated by DTS (72%) and DTT (17%) and having 4 other motifs represented, and a small clade composed of 5 different motifs (*SI Appendix*, Fig. S13). Likewise, the Group-II VirR proteins delineated into 3 clades, 2 of which are dominated by the DTV motif, with 100 and 73% DTV, respectively. The latter had 4 additional motifs, including 18% DSV. The smallest Group-II VirR clade had just 4 sequences and 3 motifs. In total, our survey of VirR revealed 8 previously unreported motifs, in addition to the observed DTS, DTV, and DTT, and indicated that motifs generally grouped in a manner connecting to the evolutionary history of these proteins. Functional characterization will be important for understanding the cell biological implications during infection as will identification of the corresponding natural hosts.

Finally, more than 99% of environmental VirR had 1 of 2 amino acids (M, L; Met89 in $VirR_{DTS}$) that confer increased green light absorption relative to blue, in contrast to the multiple wavelengths used by marine prokaryotic rhodopsins (78). With VirR proteins being in only 2 genome-sequenced viruses with known hosts, we could not parse the metagenomic sequences recovered into percentages coming from viruses of phytoplankton vs. heterotrophs. However, our recruitment of assembled PolB genes from TARA (12,684 in total) recovered 6 times more than vintage TARA 454-metagenome analyses (13, 79), and protein similarity networks identified 1,026 PolB as being from *Mimiviridae* (*SI Appendix*, Fig. S14). The computed ratio of VirR to *Mimiviridae* PolB was 0.7, suggesting that VirR is a common component of many giant viruses in sunlit ocean environments (Fig. 5 *D* and *E* and Dataset S6).

Conclusions

Predatory protists have important ecosystem roles in the transfer of organic carbon (prey-based) to higher trophic levels in addition to their top-down control of microbial cells (15). Here, using both cultivation-independent and laboratory methods, we performed cross-scale analyses—from the sequencing of giant virus genomes and their host to evolutionary relationships, functional attributes, and presence in broader ocean samples. Our studies reveal a virus of a widespread group of marine predatory protists related to metazoans, the choanoflagellates.

Alongside CroV, which infects a host from different eukaryotic supergroup, the ChoanoViruses bring the number of genomes available from giant viruses that infect known predatory hosts in pelagic marine environments to a total of 3. While the ChoanoViruses share cellular life-like proteins observed in nonmarine viruses, overall gene retention patterns seem tuned to habitat, akin to the marine nature of VirR and potentially, to host trophic mode. Additional studies that target uncultured eukaryotes and coassociated entities will more fully expose viral mechanisms for modulating the host environment and the relative strengths of evolutionary history vs. environment on viral gene content.

The AMGs identified in the ChoanoVirus genomes are of particular import for the intracellular replication environment (the host) and marine host–environment interactions. During infection, the host can be considered a "virocell," wherein the viruses modulate intracellular processes and their own replication (39). It is during this time that the ChoanoVirus has multifaceted roles in utilizing organic compounds in a manner that could facilitate rhodopsin-based photoheterotrophy, providing mechanisms for reshaping host nutrition, physiology, and the quality of remaindered organic matter. These findings raise questions on the dynamics of host–virus interactions and potential for transient mutualism. In addition to chitinase and enzymes for transport and metabolism of multiple organic molecules, the ChoanoViruses encode an unprecedented viral multigene pathway for a rhodopsin-based photosystem. Our characterization of a putative rhodopsin encoded by eukaryotic viruses, crystal structure, and biochemical assays identifies a mechanism for viral-induced light-driven energy transfer. While $VirR_{DTS}$ proton pumping may facilitate energy transfer in connection with host-derived adenosine triphosphate (ATP) synthases, the plethora of other VirR-motif types observed herein remains to be functionally and structurally characterized. This will involve identifying the host membrane in which the rhodopsins localize, be it mitochondrial, plasma, or endomembrane systems, as well as the duration, timing, and precise functional role during infection. Given the variety of viral rhodopsins, it seems that a diverse suite of roles in host manipulation awaits discovery, potentially involving host signaling, photomotility, and action potential for flagellar beating or lysis as well as photoheterotrophy. While type-1 rhodopsins are relatively widespread in marine bacteria (53, 78), they appear to be absent from marine phage genomes. In contrast, our studies, and their extension by analysis of global metagenomes, indicate that VirR is a common aspect of how giant viruses of eukaryotes reshape host physiology and potentially energy transfer in both heterotrophic and photosynthetic marine protists.

Materials and Methods

Detailed materials and methods are provided in the *SI Appendix*. This includes information on field work, flow sorting, whole-genome amplification, amplicon library construction and sequencing, data processing, assembly and analysis, including gene predictions, as well as phylogenetic, crystallization and phylogenomic analyses. The final ChoanoV1 genome assembly was assembled from 13,802,665 quality-filtered reads, and viral contigs were differentiated from the cellular assembly by tetranucleotide frequency and GC content. The ChoanoV1 assembly consisted of 11 contigs with average coverage of 215 ± 157×. For eastern North Pacific Ocean gene expression analyses, reads from metatranscriptomes were mapped to ChoanoVirus genomes at >95% nucleotide identity with bbmap. For 18S V4 amplicon sequencing, we had on average 131,385 ± 121,027 amplicons $well^{-1}$ (the lowest number being 1,037) clustered at 99% and classified via the Protistan Ribosomal Reference database. Rhodopsin functionality of viral $VirR_{DTS}$ was determined via heterologous expression in *E. coli*. The $VirR_{DTS}$ crystallization samples were produced by a cell-free system, and crystals were grown using the in meso approach. Accession numbers and DOI for alignment and tree files are available in Dataset S2.

ACKNOWLEDGMENTS. We thank the captain and crew of the R/V Western Flyer, and we thank M. Ares, L. Gómez-Consarnau, and K. Bergauer for discussions. We are grateful for the availability of Monterey Bay Time Series chlorophyll data through the Monterey Bay Aquarium Research Institute (MBARI). We thank a United States Department of Energy Joint Genome Institute Technology Development Program grant for some initial sequencing. Support came from the Canon Foundation (S.Y., T.H., T.K.-S., M.S., and W.I.), Japan Society for the Promotion of Science KAKENHI Grants 16H06279 (to S.Y. and W.I.) and 18H04136 (to S.Y. and W.I.), Gordon & Betty Moore Foundation Grants GBMF3307 (to T.A.R., P.J.K., A.E.S., and A.Z.W.) and GBMF3788 (to A.Z.W.), MBARI (A.Z.W.), and GEOMAR Helmholtz Centre for Ocean Research Kiel (A.Z.W.).

1. A. E. Zimmerman *et al.*, Closely related viruses of the marine picoeukaryotic alga *Ostreococcus lucimarinus* exhibit different ecological strategies. *Environ. Microbiol.* **21**, 2148–2170 (2019).
2. J. S. Weitz, S. W. Wilhelm, Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* **4**, 17 (2012).
3. N. Y. Ankrah *et al.*, Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J.* **8**, 1089–1100 (2014).
4. D. Lindell, J. D. Jaffe, Z. I. Johnson, G. M. Church, S. W. Chisholm, Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).
5. M. Breitbart, C. Bonnain, K. Malki, N. A. Sawaya, Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* **3**, 754–766 (2018).
6. S. Rosenwasser, C. Ziv, S. G. V. Creveld, A. Vardi, Virocell metabolism: Metabolic innovations during host-virus interactions in the ocean. *Trends Microbiol.* **24**, 821–832 (2016).
7. A. Monier *et al.*, Phosphate transporters in marine phytoplankton and their viruses: Cross-domain commonalities in viral-host gene exchanges. *Environ. Microbiol.* **14**, 162–176 (2012).
8. A. Monier *et al.*, Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7489–E7498 (2017).

9. D. Raoult *et al*., The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350 (2004).
10. C. Abergel, M. Legendre, J. M. Claverie, The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* **39**, 779–796 (2015).
11. P. Colson, B. La Scola, A. Levasseur, G. Caetano-Anollés, D. Raoult, Mimivirus: Leading the way in the discovery of giant viruses of amoebae. *Nat. Rev. Microbiol.* **15**, 243–254 (2017).
12. F. Schulz *et al*., Giant viruses with an expanded complement of translation system components. *Science* **356**, 82–85 (2017).
13. C. M. Deeg, C. T. Chow, C. A. Suttle, The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. *eLife* **7**, e33014 (2018).
14. C. A. Suttle, Marine viruses–Major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
15. A. Z. Worden *et al*., Environmental science. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**, 1257594 (2015).
16. C. P. Laber *et al*., Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat. Microbiol.* **3**, 537–547 (2018).
17. F. H. Coutinho *et al*., Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* **8**, 15955 (2017).
18. S. Santini *et al*., Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10800–10805 (2013).
19. M. Moniruzzaman *et al*., Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. *Virology* **466–467**, 60–70 (2014).
20. C. R. Schvarcz, G. F. Steward, A giant virus infecting green algae encodes key fermentation genes. *Virology* **518**, 423–433 (2018).
21. 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).
22. L. Gallot-Lavallée, G. Blanc, J. M. Claverie, Comparative genomics of Chrysochromulina ericina virus and other microalga-infecting large DNA viruses highlights their intricate evolutionary relationship with the established Mimiviridae family. *J. Virol.* **91**, e00230-17 (2017).
23. W. H. Wilson *et al*., Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* **309**, 1090–1092 (2005).
24. S. W. Wilhelm *et al*., A student's guide to giant viruses infecting small eukaryotes: From acanthamoeba to zooxanthellae. *Viruses* **9**, E46 (2017).
25. 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).
26. S. Yau *et al*., Virophage control of antarctic algal host-virus dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6163–6168 (2011).
27. D. J. Richter, P. Fozouni, M. B. Eisen, N. King, Gene family innovation, conservation and loss on the animal stem lineage. *eLife* **7**, e34226 (2018).
28. L. Deng *et al*., Grazing of heterotrophic flagellates on viruses is driven by feeding behaviour. *Environ. Microbiol. Rep.* **6**, 325–330 (2014).
29. J. M. Gonzalez, C. Suttle, Grazing by marine nanoflagellates on viruses and virus-sized particles: Ingestion and digestion. *Mar. Ecol. Prog. Ser.* **94**, 1–10 (1993).
30. J. M. Labonté *et al*., Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* **9**, 2386–2399 (2015).
31. N. Frank, T. Helge Abuldhauge, R. Daniel J, Bridging the gap between morphological species and molecular barcodes–Exemplified by loricate choanoflagellates. *Eur. J. Protistol.* **57**, 26–37 (2017).
32. A. J. Limardo *et al*., Quantitative biogeography of picoprasinophytes establishes ecotype distributions and significant contributions to marine phytoplankton. *Environ. Microbiol.* **19**, 3219–3234 (2017).
33. M. P. Simmons *et al*., Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the Eastern North Pacific Ocean. *Appl. Environ. Microbiol.* **82**, 1693–1705 (2016).
34. J. T. Pennington, F. P. Chavez, Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989-1996 in Monterey Bay, California. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **47**, 947–974 (2000).
35. E. Derelle *et al*., Diversity of viruses infecting the green microalga *Ostreococcus lucimarinus*. *J. Virol.* **89**, 5812–5821 (2015).
36. C. Bachy *et al*., Transcriptional responses of the marine green alga *Micromonas pusilla* and an infecting prasinovirus under different phosphate conditions. *Environ. Microbiol.* **20**, 2898–2912 (2018).
37. Y. Tomaru, K. Nagasaki, *Diatom Viruses* (Springer, 2011).
38. E. R. Tulman *et al*., The genome of canarypox virus. *J. Virol.* **78**, 353–366 (2004).
39. P. Forterre, The virocell concept and environmental microbiology. *ISME J.* **7**, 233–236 (2013).
40. N. Yutin, Y. I. Wolf, D. Raoult, E. V. Koonin, Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. *Vriol. J.* **6**, 223 (2009).
41. S. Sunagawa *et al*.; Tara Oceans coordinators, Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
42. S. Yooseph *et al*., The Sorcerer II Global Ocean sampling expedition: Expanding the universe of protein families. *PLoS Biol.* **5**, e16 (2007).
43. J. Abrahão *et al*., Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* **9**, 749 (2018).
44. D. De Corte *et al*., Metagenomic insights into zooplankton-associated bacterial communities. *Environ. Microbiol.* **20**, 492–505 (2018).
45. G. Blanc *et al*., The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22**, 2943–2955 (2010).
46. C. J. Thomas *et al*., Localization of a baculovirus-induced chitinase in the insect cell endoplasmic reticulum. *J. Virol.* **72**, 10207–10212 (1998).
47. G. Torruella *et al*., Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404–2410 (2015).
48. F. Azam *et al*., The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
49. O. P. Ernst *et al*., Microbial and animal rhodopsins: Structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014).
50. L. Gómez-Consarnau *et al*., Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol.* **8**, e1000358 (2010).
51. A. Martinez, A. S. Bradley, J. R. Waldbauer, R. E. Summons, E. F. DeLong, Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5590–5595 (2007).
52. 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).
53. A. Pushkarev *et al*., A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558**, 595–599 (2018).
54. O. Béjà, E. N. Spudich, J. L. Spudich, M. Leclerc, E. F. DeLong, Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
55. L. Gómez-Consarnau *et al*., Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci. Adv.* **5**, eaaw8855 (2019).
56. A. Marchetti, D. Catlett, B. M. Hopkinson, K. Ellis, N. Cassar, Marine diatom proteorhodopsins and their potential role in coping with low iron availability. *ISME J.* **9**, 2745–2748 (2015).
57. N. Yutin, E. V. Koonin, Proteorhodopsin genes in giant viruses. *Biol. Direct* **7**, 34 (2012).
58. L. B. Lamarche *et al*., Purification and characterization of RhoPDE, a retinylidene/phosphodiesterase fusion protein and potential optogenetic tool from the choanoflagellate Salpingoeca rosetta. *Biochemistry* **56**, 5812–5822 (2017).
59. K. Yoshida, S. P. Tsunoda, L. S. Brown, H. Kandori, A unique choanoflagellate enzyme rhodopsin exhibits light-dependent cyclic nucleotide phosphodiesterase activity. *J. Biol. Chem.* **292**, 7531–7541 (2017).
60. 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).
61. A. Philosof, O. Béjà, Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environ. Microbiol. Rep.* **5**, 475–482 (2013).
62. J. L. López *et al*., Microbial and viral-like rhodopsins present in coastal marine sediments from four polar and subpolar regions. *FEMS Microbiol. Ecol.* **93**, (2017).
63. J. L. Spudich, K.-H. Jung, "Microbial rhodopsins: Phylogenetic and functional diversity" in *Handbook of Photosensory Receptors*, W. R. Briggs, J. L. Spudich, Eds. (Wiley, 2005), pp. 1–23.
64. A. Harris *et al*., A new group of eubacterial light-driven retinal-binding proton pumps with an unusual cytoplasmic proton donor. *Biochim. Biophys. Acta* **1847**, 1518–1529 (2015).
65. P. A. Bulzu *et al*., Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nat. Microbiol.* **4**, 1129–1137 (2019).
66. L. Vogeley *et al*., Anabaena sensory rhodopsin: A photochromic color sensor at 2.0 A. *Science* **306**, 1390–1393 (2004).
67. J. A. Kyndt, T. E. Meyer, M. A. Cusanovich, Photoactive yellow protein, bacteriophytochrome, and sensory rhodopsin in purple phototrophic bacteria. *Photochem. Photobiol. Sci.* **3**, 519–530 (2004).
68. M. Mehler *et al*., The EF loop in green proteorhodopsin affects conformation and photocycle dynamics. *Biophys. J.* **105**, 385–397 (2013).
69. Y. Sudo *et al*., A microbial rhodopsin with a unique retinal composition shows both sensory rhodopsin II and bacteriorhodopsin-like properties. *J. Biol. Chem.* **286**, 5967–5976 (2011).
70. T. Tsukamoto, K. Inoue, H. Kandori, Y. Sudo, Thermal and spectroscopic characterization of a proton pumping rhodopsin from an extreme thermophile. *J. Biol. Chem.* **288**, 21581–21592 (2013).
71. H. Luecke, B. Schobert, H. T. Richter, J. P. Cartailler, J. K. Lanyi, Structure of bacteriorhodopsin at 1.55 A resolution. *J. Mol. Biol.* **291**, 899–911 (1999).
72. H. Luecke, B. Schobert, J. K. Lanyi, E. N. Spudich, J. L. Spudich, Crystal structure of sensory rhodopsin II at 2.4 angstroms: Insights into color tuning and transducer interaction. *Science* **293**, 1499–1503 (2001).
73. J. L. Spudich, The multitalented microbial sensory rhodopsins. *Trends Microbiol.* **14**, 480–487 (2006).
74. G. M. Avelar *et al*., A rhodopsin-guanylyl cyclase gene fusion functions in visual perception in a fungus. *Curr. Biol.* **24**, 1234–1240 (2014).
75. D. Laundon, B. T. Larson, K. McDonald, N. King, P. Burkhardt, The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLoS Biol.* **17**, e3000226 (2019).
76. G. Sabehi *et al*., New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol.* **3**, e273 (2005).
77. J. McCarren, E. F. DeLong, Proteorhodopsin photosystem gene clusters exhibit coevolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ. Microbiol.* **9**, 846–858 (2007).
78. J. Pinhassi, E. F. DeLong, O. Béjà, J. M. González, C. Pedrós-Alió, Marine bacterial and archaeal ion-pumping rhodopsins: Genetic diversity, physiology, and ecology. *Microbiol. Mol. Biol. Rev.* **80**, 929–954 (2016).
79. P. Hingamp *et al*., Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* **7**, 1678–1695 (2013).