Evidence for the acquisition of a proteorhodopsin-like rhodopsin by a chrysophyte-infecting giant virus

Petra Byl[#, 1, 2, 3, *], Christopher R. Schvarcz[1, 2], Julie Thomy[1, 2], Qian Li[1, 2], Cori B. Williams[1], Kurt LaButti[3], Frederik Schulz[3], Kyle F. Edwards[2], Grieg F. Steward[#, 1, 2]

[1]Daniel K. Inouye Center for Microbial Oceanography: Research and Education, School of Ocean and Earth Science and Technology (SOEST), University of Hawai‘i at Mānoa, Honolulu, HI, USA.

[2]Department of Oceanography, School of Ocean and Earth Science and Technology (SOEST), University of Hawai‘i at Mānoa, Honolulu, HI, USA.

[3]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA. USA.

Running Head: Proteorhodopsin encoded by chrysophyte virus

#Address correspondence to Petra Byl pbyl@umces.edu & Grieg Steward grieg@hawaii.edu

*Present address: University of Maryland Center for Environmental Science, Horn Point Laboratory, Cambridge, MD, USA.

Abstract word count: 239

Importance word count: 144

Main text word count: 3,619

Abstract

Chrysophytes are widespread nanoflagellate protists in aquatic ecosystems with diverse trophic roles as primary producers and bacterivores. While molecular evidence suggests that chrysophytes are commonly infected by giant viruses, no previous isolates have been reported. Here, we describe the first isolated chrysophyte-infecting virus, Chrysophyceae Clade H virus SA1 (ChrysoHV). ChrysoHV and its mixotrophic host were isolated from surface waters in the tropical North Pacific. The ChrysoHV capsid (290 ± 40 nm diameter) is associated with a loose, sac-like membrane that extends its effective diameter (720 ± 120 nm), and presents a long, thin tail extending 1,200 (± 240) nm in length. This morphology has not been previously observed in other virus isolates or environmental surveys. The sequenced, assembled genome is 1.19 Mbp, and phylogenetic analysis places ChrysoHV as the third cultivated member of the *Aliimimivirinae* subfamily in the *Mimiviridae* family of giant viruses. ChrysoHV is the first cultivated member of *Aliimimivirinae* to encode a rhodopsin. Moreover, ChrysoHV encodes three rhodopsins total, two heliorhodopsins and one proteorhodopsin. No other virus-derived rhodopsin has been found in the proteorhodopsin clade, and the loss of retinal binding hints at a light-independent function. ChrysoHV also encodes nine genes with closest known homologs from marine cyanobacteria, mostly related to nutrient uptake. This suggests that lateral gene transfer occurs between abundant marine bacteria and giant viruses, supporting the conceptual model that genes are horizontally transferred between ingested prey and infectious virions within predatory protists.

Importance

Chrysophytes are abundant eukaryotic phytoplankton with diverse trophic strategies ranging from photosynthesis to phagotrophy, and have been studied as models of mixotrophy among marine protists. However, no chrysophyte-infecting viruses had been cultivated, leaving a major gap in our understanding of viral diversity. This study reports the functional and genomic characterization of the first isolated chrysophyte-infecting virus, the giant virus ChrysoHV. The virion morphology is unusual, with a loosely associated outer membrane and a long thin tail. The viral genome encodes three rhodopsins, one of which is affiliated with a bacterial proteorhodopsin lineage not previously identified in viruses. ChrysoHV also encodes genes with closest homologs in marine cyanobacteria, suggesting that phagocytic protists may provide a conduit for horizontal gene transfer from bacterial prey to viral pathogens. This isolate offers a useful model for exploring giant virus genomics and the function of a new viral rhodopsin.

Main text

Introduction

Chrysophytes are single-celled or colonial protists in the class Chrysophyceae, and they are globally abundant in freshwater and marine ecosystems (1, 2). Chrysophytes play multiple ecosystem roles as primary producers and bacterivores, and include phago-mixotrophic forms that perform photosynthesis and ingest prey, strictly heterotrophic phagotrophs, and potentially non-phagotrophic phototrophs (3). In subtropical ocean gyre ecosystems, which span approximately a quarter of the global surface ocean (4), chrysophytes play an important role as predators of marine cyanobacteria (5, 6), influencing the community structure at the base of the food web. Isolates from Chrysophyceae Clade H have been shown to be voracious consumers of

70 the cyanobacterium *Prochlorococcus* (7), the dominant primary producer in subtropical gyres
71 and most abundant phototroph in the ocean (8). Additionally, single cell amplified genomes of
72 Chrysophyceae Clade H have been shown to encode endogenous viral elements from the
73 *Nucleocytoviricota* phylum (9), referred to as giant viruses (10). Genetic marker surveys of
74 protist and giant virus distributions from the *Tara* Oceans global dataset also indicate that
75 chrysophytes commonly co-occur with the giant virus family *Mimiviridae* (11), and further
76 RNA-seq studies revealed that chrysophytes are in fact native hosts of mimivirids (12).
77 However, chrysophyte-infecting giant virus isolates have not been previously described. In this
78 study we detail the isolation and initial characterization of a giant virus that infects a Clade H
79 chrysophyte, ChrysoHV.

80 Giant viruses encode vast genomic repertoires that include numerous homologs of
81 cellular genes (13). These cellular homologs may augment or compensate for host functions
82 compromised during infection (14–16), or interfere with host defenses (17). Among the most
83 commonly found viral homologs of cellular genes are microbial rhodopsins (18–20), which are
84 photoreceptive proteins that may aid in the generation of ion gradients or intracellular signaling
85 (21, 22). Viral rhodopsin homologs have been experimentally confirmed to convert light energy
86 into electrochemical potential across multiple clades: viral rhodopsin groups I and II (14, 23, 24),
87 channelrhodopsin (15), and heliorhodopsin (17). Furthermore, these viral rhodopsin homologs
88 are globally distributed across aquatic habitats, with their relative abundance decreasing over
89 depth with the attenuation of sunlight (14, 25).

90 Although giant viruses are diverse and globally abundant (11, 19, 20), a relatively small
91 fraction of their known diversity is represented by cultivated isolates (26). Likewise, giant virus
92 rhodopsin homologs are widespread and potentially functional, but only a few rhodopsin-

93 encoding giant viruses have been isolated and remain in culture with their native host (16, 17,
94 27). Cultivated isolates are important for establishing virus-host linkages, and furthermore, can
95 reveal novel gene lineages that may go undetected in environmental datasets and clarify the
96 evolutionary history of highly divergent sequences (16). Through the isolation and
97 characterization of ChrysoHV, we identify a previously undescribed lineage of viral rhodopsin
98 that appears to derive from bacterial proteorhodopsin.
99

100 **Results & Discussion**

101 *Host range, infection dynamics, and morphology*

102 We tested the host range of ChrysoHV against seven protist isolates from Station
103 ALOHA, spanning three classes: three Chrysophyceae, three Dictyochophyceae, and one
104 Bolidophyceae representative. This resulted in no lysis for six of the seven isolates (Fig. 1B)
105 (Supplementary Fig. S1). Only the representative of chrysophyte Clade H, lysed when
106 challenged with ChrysoHV (Fig. 1). The apparent burst size, *i.e.*, estimated number of virions
107 produced per cell, was 24 ± 2 ([mean ± SD]) (Fig. 1A). The virion particles produced from the
108 infection of chrysophyte strain UHM3501 displayed the notable morphological features of an
109 icosahedral capsid (290 ± 40 nm diameter [mean ± SD]) associated with a loose membrane (720
110 ± 120 nm), and a long tail (1,200 ± 240 nm length) terminating in tail fibers (210 ± 700 nm
111 length) (Fig. 2A-C).
112

113 *Genomic characterization*

114 The ChrysoHV genome was assembled as a putatively linear DNA sequence of 1.19 Mbp
115 encoding 1,138 predicted genes with a G+C content of 23.42% (Fig. 2B). The encoded genes

116 have diverse putative functions, including the ribosomal proteins eL40 and S27a, making it the
117 second isolated eukaryote-infecting virus known to encode ribosomal proteins (21), and the first
118 to encode S27a (Supplementary Table S1). ChrysoHV also encodes electron transport proteins
119 such as cytochrome b5, widespread in prokaryotes, eukaryotes, and giant viruses (Supplementary
120 Table S1) (28), along with succinate dehydrogenase flavoprotein and iron-sulfur subunits
121 (Supplementary Table S1). Additionally, the virus encodes seven ABC transporters including
122 one specific to phosphate transport (XVM28869.1) and two AmtB ammonium transporters. Of
123 these ammonium transporters, one (XVM28992.1) has 55% amino acid sequence identity (aaID)
124 to *Prochlorococcus* and *Synechococcus* homologs (Supplementary Table S1). Likewise,
125 ChrysoHV encodes three porins with 44-66% aaID to *Prochlorococcus* iron and carbohydrate
126 porins (XVM28452.1, XVM28569.1, XVM28570.1), two urea ABC transporter substrate-
127 binding proteins with 81-90% aaID to a *Prochlorococcus* homolog (XVM29005.1,
128 XVM29006.1), a semiSWEET sugar transporter with 44% aaID to a *Prochlorococcus* homolog
129 (XVM28124.1), a sulfotransferase family protein with 55% aaID to a *Prochlorococcus* homolog
130 (XVM28964.1), and a N-acetylmuramoyl-L-alanine amidase for cell wall remodeling with 39%
131 aaID to a *Prochlorococcus* homolog (XVM29061.1) (Supplementary Table S1). The presence of
132 genes whose closest known relatives are *Prochlorococcus* homologs suggests extensive
133 horizontal gene transfer from cyanobacterial prey of the chrysophyte to a giant virus that infects
134 it.

135 The ChrysoHV genome also includes a region of putatively light-sensitive proteins,
136 encompassing a proteorhodopsin-like protein surrounded by two photolyases and followed by a
137 heliorhodopsin (Fig. 2D, Fig. 3C, Supplementary Table S1). The positioning and shared

138 orientation of the photolyases and heliorhodopsin around the proteorhodopsin-like protein
139 implies that they may be part of the same operon.
140
141 *Sequence characterization of a distinct lineage of proteorhodopsin*
142 Phylogenetic analysis of ChrysoHV's rhodopsin placed it within the proteorhodopsin
143 clade (Fig. 3A). The sequence alignment of transmembrane domain helix-III (TM3) indicates
144 that the ChrysoHV rhodopsin is spectrally tuned to blue light (glutamine, Q; Fig. 3A), which is
145 the dominant spectral tuning residue in proteorhodopsins from Station ALOHA, the isolation site
146 of ChrysoHV (29). The two most closely related rhodopsin sequences were recovered from
147 metagenomes from the Black Sea at 50 m depth (SAMN27960419) and the North Pacific at 5 m
148 depth in the 0.22-3μm size fraction (SAMEA2623059). These rhodopsins may also absorb blue
149 light (Fig. 3A) and share a poly-N (asparagine) tract at the start of the rhodopsin coding
150 sequence.
151 ChrysoHV does not encode β-carotene, a precursor to retinal, the light-absorbing
152 chromophore in proteorhodopsin, and the salt bridge between K102 and E228 most likely
153 inhibits retinal binding. This loss of retinal binding, if associated with a change in rhodopsin
154 function, could explain the high rate of sequence divergence of ChrysoHV's proteorhodopsin
155 compared to the rest of the proteorhodopsin clade (Fig. 3A). Understanding how the rhodopsin
156 may function without retinal binding could be important to interpreting the function of viral
157 rhodopsins detected below the photic zone (30, 31). Further investigation is needed to confirm
158 whether the ChrysoHV proteorhodopsin can confer other physiological benefits in the absence of
159 light absorption, similar to the *Psychroflexus torquis* proteorhodopsin ptqPR, which acts as a
160 phospholipid scramblase to preserve membrane potential, fluidity, and integrity (32).

161 ChrysoHV also lacks the conserved DTE proton pumping motif found in most
162 proteorhodopsins, and instead, possesses an ETL motif (Fig. 3A), which has been detected in
163 proteorhodopsins from deep water at Station ALOHA (Fig. 3A) (29). A comparison of the
164 predicted structure of the conserved domain of the ChrysoHV rhodopsin against existing
165 functional characterizations revealed that it was most similar (35.9% aaID) to the uncultured
166 gamma-proteobacterium clone HOT 75m4 rhodopsin (Fig. 3B). This rhodopsin was confirmed to
167 pump protons (H^+) (33), and, like ChrysoHV, was collected from Station ALOHA (33). These
168 results suggest that ChrysoHV may have acquired the proteorhodopsin from a marine bacterium,
169 which, like *Prochlorococcus*, may be prey for chrysophyte Clade H.
170
171 *Comparison to photoreceptive proteins across Nucleocytoviricota*
172 ChrysoHV is the third cultivated isolate in the *Aliimimivirinae* subfamily of *Mimiviridae*
173 (Fig. 2E); other known isolates include CroV BV-PW1 (34), infecting a heterotrophic
174 nanoflagellate, and CV-XW01 (35), infecting a green alga. In this subfamily, only two other
175 representatives, cPacV-1605 and GVMAG-M-3300023174-43, encode rhodopsins, which belong
176 to viral rhodopsin group II (Fig. 2A). However, 40 other *Aliimimivirinae* representatives encode
177 at least one photolyase homolog (Fig. 2E), either the (6-4) photolyase or DASH-type
178 cryptochrome. Only ChrysoHV and cPacV-1605 encode multiple rhodopsin and photolyase
179 types, including two heliorhodopsins (Fig. 3A). However, cPacV-1605 encodes a viral rhodopsin
180 (group II) and ChrysoHV encodes a proteorhodopsin (Fig. 3A). Furthermore, instead of being
181 nested within a photoreceptive gene neighborhood, the position of the viral rhodopsin in cPacV-
182 1605 (locus tag: CPAV1605_207, GenBank ID: VVU94485.1) is further away from the
183 photolyases (locus tag: CPAV1605_133, CPAV1605_136; GenBank ID: VVU94411.1,

184 VVU94414.1) and heliorhodopsins (locus tag: CPAV1605_533, CPAV1605_1309; GenBank ID:
185 VVU95557.1, VVU94808.1) (36). ChrysoHV is the only cultivated giant virus to encode the
186 proteorhodopsin lineage of rhodopsin along with two heliorhodopsins (Fig 2E, Fig. 3A).
187
188 *Global ocean distribution of ChrysoHV phylotypes and proteorhodopsins*
189 Analysis of the distribution of ChrysoHV phylotypes (≥ 90% PolB aaID) from the *Tara*
190 Oceans dataset (11) revealed their prevalence in subtropical surface ocean sites in the Pacific and
191 Indian Ocean as well as the Mediterranean Sea (Fig. 4A). Notably, at station 132, north of the
192 Hawaiian archipelago, ChrysoHV phylotypes were detected in both the mixed layer and
193 mesopelagic zone, suggesting connectivity between surface and deep ocean virus populations,
194 which has previously been observed in the tropical North Pacific (25). The abundance of the
195 ChrysoHV proteorhodopsin-like protein (≥ 80% PolB aaID) was also highest in the subtropical
196 surface ocean (Fig. 4B), but detected at fewer overall sites than the ChrysoHV PolB.
197 Additionally, the ChrysoHV proteorhodopsin relative abundance was highest at station 132,
198 indicating that these proteorhodopsin sequences may be derived from a ChrysoHV phylotype.
199
200 **Conclusion**
201 In the marine environment, diverse microorganisms from proteobacteria to microalgae
202 utilize rhodopsins to supplement their energetic needs (29, 37). SAR11 bacteria, of which
203 *Pelagibacter* is an abundant subclade, account for 25% of all plankton (38), and are a likely prey
204 source for mixotrophic algae such as Clade H chrysophytes. The discovery of a proteorhodopsin-
205 like protein encoded by a giant virus supports the conceptual model that lateral gene transfer is
206 driven by the co-mingling of infectious virions and ingested prey within the same host cell (39).

207 The presence of nine genes whose closest known relatives are *Prochlorococcus* homologs further
208 suggests that prey-virus HGT has played an important role in the evolution of ChrysoHV. By
209 leveraging naturally co-occurring and experimentally accessible virus-host pairs, this study lays
210 the groundwork for future investigations into how the virocell environment of predatory protists
211 influences gene acquisition in giant viruses. As the only cultivated isolate known to encode one
212 microbial rhodopsin and two heliorhodopsins, this system provides the opportunity for novel
213 insights into the role of rhodopsins in giant virus infection strategies.
214

215 **Materials and Methods**

216 *Isolation and cultivation*

217 The five mixotrophic protists used in this study include one chrysophyte (UHM3501),
218 one bolidophyte (UHM3600), and three dictyochophytes (UHM3020, UHM3053, UHM3071),
219 all of which were isolated from the epipelagic zone (5–100m) of Station ALOHA through serial
220 dilutions of whole seawater samples enriched with *Prochlorococcus* (7). The mixotrophs were
221 cultivated at 24°C at approximately 200–220 µmol photons m^{-2} s^{-1} irradiance to mimic the high
222 light conditions at station ALOHA (40). The unialgal cultures were grown on K medium without
223 dissolved, inorganic nitrogen sources ($NaNO_3$ and NH_4Cl). Cultures were instead amended with
224 cyanobacteria prey (*Prochlorococcus* MIT9301) for their nitrogen source (7, 41). In addition,
225 two heterotrophic chrysophytes (UHM3530 and UHM3560) were used in this study. These
226 strains were isolated from seawater collected from the mixed layer (5–50 m) at Station ALOHA
227 in 2022 and 2024 through serial dilutions of seawater samples. The heterotrophs were cultivated
228 on filtered seawater amended with sterile quinoa grains to stimulate growth of ambient bacteria
229 as their food source.

230 The Chrysophyceae Clade H virus SA1 (ChrysoHV) was isolated by challenging the
231 chrysophyte isolate UHM3500 with unfiltered seawater collected from 25 m depth at Station
232 ALOHA on 04 May 2019. The lytic agent detected in the challenge was rendered clonal by serial
233 dilutions to extinction (42). This is the first chrysophyte-infecting virus isolated from Station
234 ALOHA, thus the strain designation SA1. In addition to the host used for isolation, the virus was
235 found to infect a related chrysophyte strain UHM3501 (Supplementary Fig. S1). Both strains are
236 members of Chrysophyceae Clade H, but differ in their grazing kinetics and 18S rRNA gene
237 sequences (7). The virus was maintained on both chrysophytes through serial, monthly
238 challenges of exponential phase culture with lysate from the preceding challenge until
239 chrysophyte strain UHM3500 was lost. Approximately three to four days post-infection (DPI),
240 infected cultures were stored at 4°C to serve as lysate for the next round of infection.
241
242 *Host range tests and flow cytometry*
243 We challenged isolates of Ochrophyta protists, both mixotrophs and heterotrophs, with
244 raw ChrysoHV lysate, then tracked the population dynamics of the protists and viruses over time
245 using flow cytometry. The mixotrophic algal protists were tracked every two days up to ten DPI,
246 and the heterotrophic protists were tracked every day up to five DPI. The experimental design
247 included triplicate, 25 mL protist cultures for both negative controls and infection treatments. To
248 prepare the infection treatment, 500 µL of unfiltered lysate (3-6 × 10^5 virus-like particles mL^{-1}
249 final concentration) was added to exponential or early stationary phase protist cultures. The
250 samples were fixed with ice-cold glutaraldehyde at 0.5% final concentration, incubated for 15
251 min at 4°C, and then flash frozen in liquid nitrogen (43).

252 The samples were stored at -80°C up to four days, then thawed at room temperature and
253 processed within the same day on a CytoFlex-S (Beckman Coulter Life Sciences, Indianapolis,
254 IN, USA) in the Biological Electron Microscope Facility (BEMF) at the University of Hawai‘i at
255 Mānoa. The samples were split into two fractions, with 200 µL allocated towards quantifying
256 cell abundances and loaded directly onto a 96-well plate, and 5 µL allocated toward quantifying
257 virus abundances. The latter fractions were diluted into 195 µL solutions of TE (pH 8.0) with
258 2.5X SYBR Green I nucleic acid stain (Invitrogen, Thermo Fisher Scientific, Waltham, MA,
259 USA) and incubated at 80°C for 20 min in darkness (44), then loaded onto a 96-well plate. The
260 cell counts were run on auto-record for one minute at a flow rate of 20 µL min^{-1} for the algae and
261 30 µL min^{-1} for the heterotrophs. The mixotroph populations were distinguished through
262 chlorophyll fluorescence and 488 nm, 90° side scatter (SSC), whereas the heterotrophs were
263 detected using SYBR Green fluorescence and 488 nm, 90° SSC (45). The viral-like particle
264 (VLP) population was identified through SYBR Green fluorescence and 405 nm, 90° violet side
265 scatter (VSSC), enhancing marine virus detection (46). The samples were run on auto-record for
266 one minute at a flow rate of 10 µL min^{-1}. Both algae and virus populations were delineated
267 through a custom Python script using the following packages: fcsparser (v0.2.8), pandas (v2.2.2),
268 matplotlib (v3.7.1), while the FlowJo (v.10.10.0) software was used to delineate the
269 heterotrophic protist population. Apparent burst size was determined by dividing the maximum
270 concentration of VLPs post-infection by the inferred number of lysed cells, calculated as the
271 difference between the cell concentration on the day of infection minus the minimum
272 concentration post-infection.
273
274 *Electron microscopy*

275 Transmission electron microscope (TEM) micrographs of viral particles were prepared
276 from raw, unfixed lysate and negatively stained with 2% uranyl acetate for 45 sec on 200-mesh
277 copper grids with carbon-stabilized formvar support that had been rendered hydrophilic by glow
278 discharge (47). Images were captured with a Hitachi HT7700 transmission electron microscope
279 (Tokyo, Japan) at the Biological Electron Microscope Facility (BEMF) at the University of
280 Hawai‘i at Mānoa. Particle dimensions were measured with ImageJ from Fiji (v2.16.0/1.54p)
281 (48), including: capsid diameter (n = 38), envelope (n = 34), tail (n = 17), and tail fibers (n = 6).
282
283 *DNA extraction, sequencing, assembly, and annotation*
284 To prepare sufficient viral biomass for DNA sequencing, DNA was extracted from 610
285 mL of lysate (6.4×10^6 VLP/mL), after allowing the infection of chrysophyte Clade H sp.
286 UHM3501 to clear four DPI. The unfiltered lysate was concentrated and pooled through serial
287 centrifugations (6X) in a 100 kDA Centricon Plus centrifugal filter (MilliporeSigma, Burlington,
288 MA, USA) at 3,500 × *g* at 20°C for 15 minutes. Approximately 200 ng of DNA was extracted
289 from the concentrate using the ZymoBiomics Quick-DNA High Molecular Weight MagBead Kit
290 (Zymo Research, Irvine, CA, USA), flash frozen in liquid nitrogen, stored at -20°C, then shipped
291 on dry ice to the US Department of Energy Joint Genome Institute (JGI). There, the DNA was
292 sequenced using PacBio Multiplexed 6–10 kb Low Input parameters.
293 A PacBio Low Input (DNA) library was constructed and sequenced using the PacBio
294 SEQUELIIe platform (Pacific Biosciences, Menlo Park, CA USA), which generated 3,267,906
295 reads totaling 31,033,080,767 bp. Reads were deduplicated using pbmarkdup (v1.0.3) (49).
296 BBDuk (v38.99) (50) was used to remove reads that contained PacBio control sequences and
297 PCR adapters, and IceCreamFinder (v38.99) (51) was used to remove reads without SMRTbell

298 sequences. The final filtered fastq contained 3,257,268 reads totaling 30,921,351,264 bp with
299 mean read length of 9,493 bp.
300 The filtered reads (31 Gb) were assembled using Flye (v2.9) (52) with the "--meta --
301 pacbio-hifi" argument. From this initial assembly, viral contigs were identified and grouped
302 through a low-complexity binning workflow, utilizing CONCOCT (v1.1.0) (53), MaxBin
303 (v2.2.7) (54), MetaBAT2 (v2.15-5) (55), and SemiBin2 (v1.5.1) (56). For each binning strategy,
304 the viral bin was identified by assessing the presence of giant virus marker genes (10) with
305 GVClass (v1.0) (57), and the MetaBAT2-generated viral bin was used in downstream analysis.
306 Subsequently, circular consensus sequencing reads were recruited against these identified viral
307 contigs, and a 100 Mb subsample of these reads was assembled using Flye without the "--meta"
308 argument (v2.9) (52). This iterative assembly resulted in one high-quality viral contig, which was
309 polished using two rounds of Racon (v1.6.0) (58) to ensure accuracy.
310 Gene calling was performed with GVClass (v1.0) (57), and the 1.196 Mb viral genome
311 had an estimated 93.77% coding density. Gene annotation was performed with KofamScan
312 (v1.3.0) (59) and eggnog mapper (v2.1.12) (60) with annotations of genes of interests confirmed
313 with BLASTp (61). tRNAs were annotated with tRNAscan-SE (v2.0.12) (62).
314
315 *Phylogenetic analysis*
316 Phylogenetic analysis of ChrysoHV with IQ-TREE (v2.2.6) (63) was based on a
317 concatenated alignment of seven core proteins: SFII, RNAPL, PolB, TFIIB, TopoII, A32, and
318 VLTF3 (10). Query sequences were retrieved from the GVMAGs (20) and GenBank Assembly
319 (64) databases, and quality filtered to contain at least four out of seven core proteins determined
320 through hmmsearch (v3.3.2, E-value $\leq 1 \times 10^{-10}$) (65). Additionally, the query sequences were

filtered to remove genomes with more than 10 copies of any core protein and more than four core proteins with multiple copies. The remaining 990 query sequences were aligned with MAFFT (v7.520) (66) and trimmed with trimAl (-gt 0.1; v1.4.1) (67), then concatenated with a custom python script (68). The trimmed, concatenated alignment was used for phylogenetic analysis using the mixture model LG+I+F+G4 and ultrafast bootstrap supports (option: -bb 1000) (v2.2.6) (63). The tree was annotated with iTOL (v6) (69).

Gene phylogeny of microbial rhodopsins with a heliorhodopsin outgroup were built with IQ-TREE (v2.2.6) (63) using model LG+F+R10 and ultrafast bootstrap supports (option: -bb 1000) and annotated with iTOL (v6) (69). Representatives of rhodopsin homologs from cultivated virus isolates include: EhV-202 (heliorhodopsin) (17), PgV-12T and -16T (viral rhodopsin I) (27), and FloV-SA2 (viral rhodopsin II) (16). Additionally, the gene phylogeny includes synthetic construct: vPyACR_21921 (channelrhodopsin) (15), and rhodopsin homologs from single amplified genomes (SAGs) and metagenome assembled genomes (MAGS): ChoanoV1 (viral group II) (14), cPacV-1605 (viral group II) (70), and OLPV 1 and 2 (viral groups I & II) (36). Query rhodopsin sequences were retrieved from NCBI and the supplementary materials of Olson *et al*. (2018), which included proteorhodopsin sequences recovered from metagenomic sequences at Station ALOHA (29). The alignment was built with MAFFT (L-INS-i; v7.520) (66) and trimmed with trimAl (-gt 0.1; v1.4.1) (67).

The phylogeny of near-complete 18S rRNA genes from select algal isolates were built with IQ-TREE (v2.2.6) (63) using GTR+I+G model with ultrafast bootstrap support (option: -bb 1000) and annotated with iTOL (v6) (69). The nucleotide sequences were retrieved from NCBI, aligned with MAFFT (L-INS-i; v7.520) (66) and trimmed with trimAl (-gt 0.1; v1.4.1) (67).

Gene distribution in the global ocean

Comparisons to *Nucleocytoviricota* PolB amino acid sequences from the *Tara* Oceans dataset (11) were performed using BLASTp (v2.13.0, E-value $\leq 1 \times 10^{-120}$) (61), and sequences with 90% or greater aaID were plotted using a custom R script with the following packages: maps (v3.4.2), mapdata (v2.3.1), and ggplot2 (v.3.5.1). The relative abundance of the PolB genes were normalized to the total number of *Nucleocytoviricota* PolBs (11). Likewise, ChysoHV proteorhodopsin-like protein sequences from the *Tara* Oceans dataset with a minimum of 80% aaID were identified through Ocean Gene Atlas (71, 72). The relative abundance of ChysoHV proteorhodopsin-like protein sequences were normalized to total metagenomic reads in the given sample (71, 72).

Protein structure predictions

The conserved domain of the ChrysoHV rhodopsin structure was predicted using ColabFold (v1.5.5) (73) and compared against the AlphaFold Protein Structure Databases–SwissProt (74) and CATH (v4.4) (75) using Foldseek (76). The conserved domain of the ChrysoHV rhodopsin and HOT 75m4 Gamma-proteobacterium proteorhodopsin structures were rendered in ChimeraX (v1.8.dev202403010247) (77).

Acknowledgments

We thank the personnel of the Hawai‘i Ocean Time-series program (NSF award 12-60164) and Research Vessel Kilo Moana for assistance with seawater collection. We thank the Biological Electron Microscope Facility (BEMF) at the University of Hawai‘i at Mānoa, especially facility supervisors, Tina Weatherby and Dr. Orion Rivers, for their assistance with electron microscopy

procedures. A special thanks to laboratory technician Kelsey McBeain for maintenance of the UHM library culture collection, and US DOE Joint Genome Institute Scientific Project Manager, Kerrie Barry, for overseeing DNA library preparation, quality control, and sequencing of ChrysoHV.

We are grateful for thoughtful feedback provided by Dr. Karen Selph (Department of Oceanography, University of Hawai‘i at Mānoa) and Dr. Alison Sherwood (School of Life Sciences, University of Hawai‘i at Mānoa) on manuscript preparation, and insights into rhodopsin structure and rhodopsin sequences (SAMN27960419 and SAMEA2623059), provided by Dr. Oded Béjà (Faculty of Biology, Technion-Israel Institute of Technology) and Dr. Andrey Rozenberg (Faculty of Biology, Technion-Israel Institute of Technology).

Funding

This work was supported by the US National Science Foundation Graduate Research Fellowship Program (Grant No. 1842402 to PB), Division of Ocean Sciences (Grant No. 2129697 to GFS and KFE), and RII Track-2 FEC (Grant No. 1736030 to GFS and KFE). The work conducted by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a Department of Energy (DOE) Office of Science User Facility, was supported by the Office of Science of the US DOE, operated under Contract No. DE-AC02-05CH11231.

Author Contributions

GFS, KFE, and PB developed, designed, and acquired funding support for the project. CRS isolated the virus and algae strains. CW and QL assisted with the collection of virus host-range data. FS provided sequencing support and supervised the bioinformatics analyses. KL assisted

with the virus genome assembly. JT assisted with gene phylogeny workflows. CRS and PB imaged the virus. PB wrote the manuscript with input from all authors.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability

ChrysoHV sequencing data is available through the NCBI sequence read archive SRR34002084, and the genome assembly has been deposited to NCBI GenBank under accession ID PV764857. The mixotrophic protists used in the study, GenBank IDs: MZ611734.1 (Chrysophyceae Clade H sp. UHM3500), MZ611720.1 (Chrysophyceae Clade H sp. UHM3501), MN615710.1 (*Florenciella* sp. UHM3020), MZ611733.1 (Dictyochophyceae clade X sp. UHM3020), MZ611705.1 (*Rhizochromulina* sp. UHM3071), MZ611730.1 (*Triparma* sp. UHM3600), are available through NCMA. The virus and heterotrophic protists used in this study, GenBank IDs: PV717318.1 (*Paraphysomonas* sp. UHM3560) and PV717319.1 (*Spumella* sp. UHM3530), are available through the Marine Viral Ecology Laboratories at the University of Hawai‘i at Mānoa upon request to KFE and GFS. Additional data and scripts on gene alignments, phylogenies, genome annotation, protein structural predictions, host range challenges, and biogeography are available on the Dryad repository: http://datadryad.org/share/8duyiXJ3riDLgLbseVaWvUBnbOPOM4SkBw9WNBaRYA4.

Figures

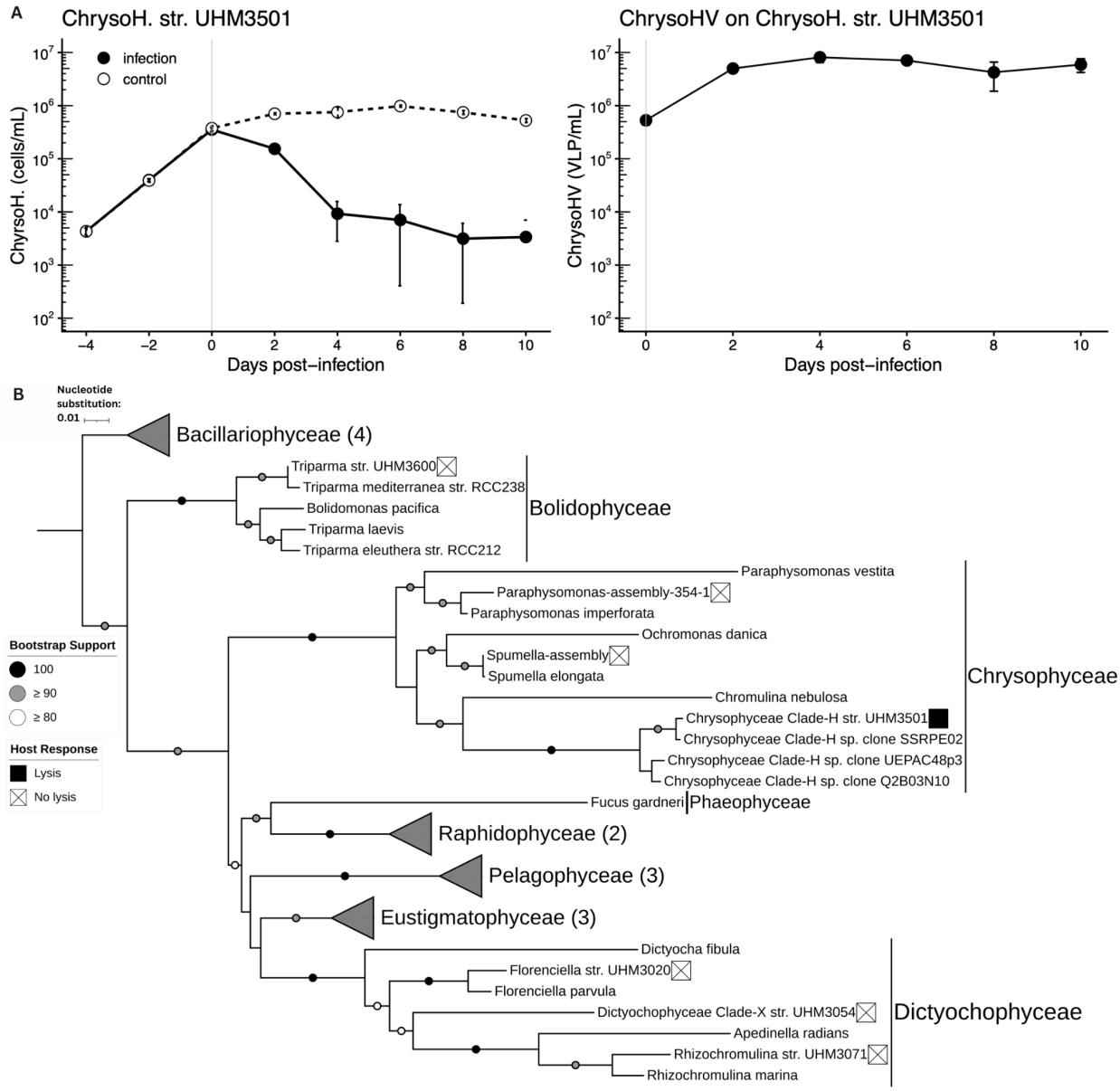


Figure 1. **ChrysoHV infection dynamics and host range tests.** (A) The left panel shows a time series of Chrysophyceae Clade H sp. UHM3501 abundance (cells/mL) in response to the addition of raw ChrysoHV lysate. The control treatment is shown in the black dashed line with white-fill points, and the infection treatment shown in the solid black line with black-fill points. The right panel shows the abundance of ChrysoHV (VLP/mL) during the same time series. Days pre- and post-infection are displayed on the x-axis. Error bars represent the standard deviation of the triplicate counts. (B) Maximum Likelihood (ML) 18S rRNA phylogeny (GTR+I+R model)

420 of the Ochrophyta phylum with isolates challenged in bold. The white or black boxes by the
421 bolded isolate names indicate whether the protist culture did not lyse (white with “X”) or lysed
422 and produced virions (black) in response to the ChrysoHV challenge. The node symbols indicate
423 the bootstrap values equal to or greater than 80% (white), 90% (gray) or 100% (black). There is
424 no node symbol for branches with less than 80% support.
425

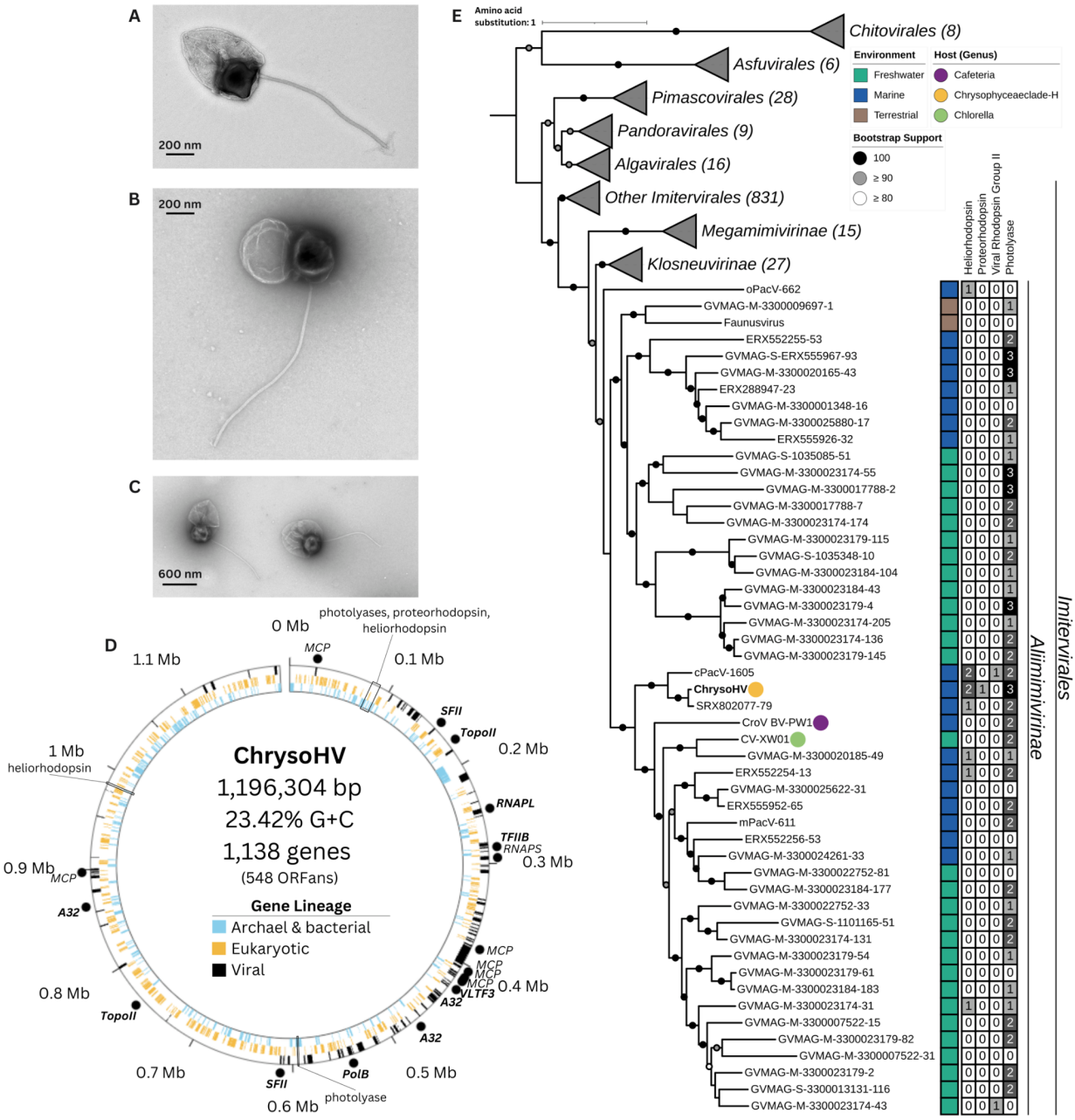


Figure 2. **ChrysoHV ultrastructure, genome, and phylogenetic analysis.** (A) Transmission electron microscopy (TEM) micrograph of ChrysoHV propagated on UHM3501. (B, C) TEM micrograph of negatively stained ChrysoHV propagated on UHM3500. (D) Genome plot of ChrysoHV with track placement and color showing gene lineage annotated according to domain, inner to outer tracks showing: archaeal and bacterial (blue), eukaryotic (yellow), and viral

432 (black) lineages. Phylogenetically informative giant virus marker genes are represented as black
433 dots on the outside track and show gene name. The genes used in the *Nucleocytoviricota*
434 phylogenetic analysis are bolded. (E) Maximum Likelihood (ML) phylogeny of the
435 *Nucleocytoviricota* phylum inferred from a concatenated alignment of amino acid sequences of
436 marker genes (SFII, RNAPL, PolB, TFIIB, TopoII, A32, VLTF3) using model LG+I+F+G4 with
437 ultrafast bootstrap supports (-bb 1000) shown by node symbol color– white (≥ 80%), gray (≥
438 90%), and black (100%) with no node symbol for supports < 80%. The tree is rooted with
439 *Chitovirales* as the outgroup, and the scale bar represents one amino acid substitution per site.
440 The frequency of microbial rhodopsins and photolyases present in *Aliimimivirinae*
441 representatives are annotated in the grayscale matrix, and the environment the representative was
442 collected from is shown in jade (freshwater), blue (marine), and brown (terrestrial).
443

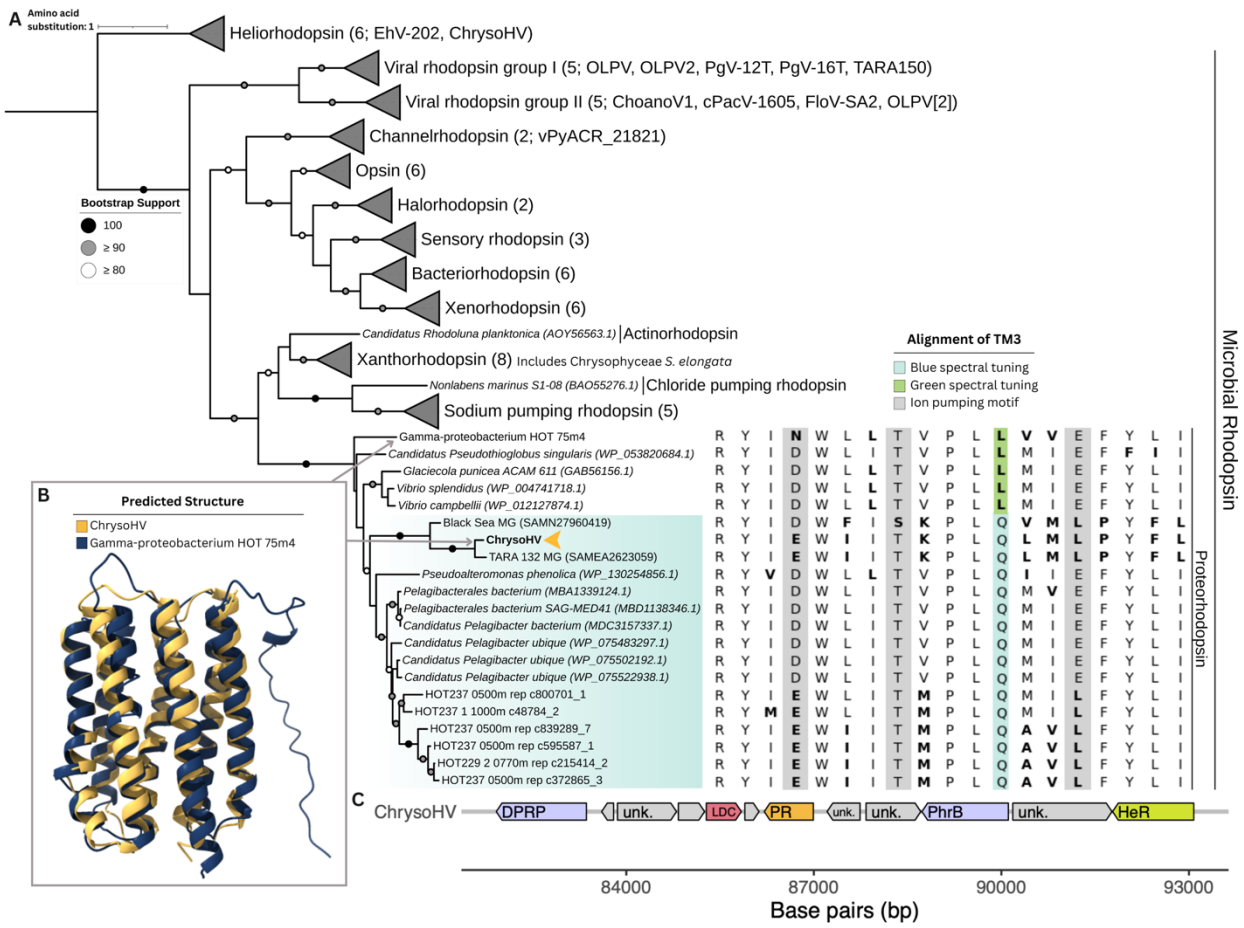
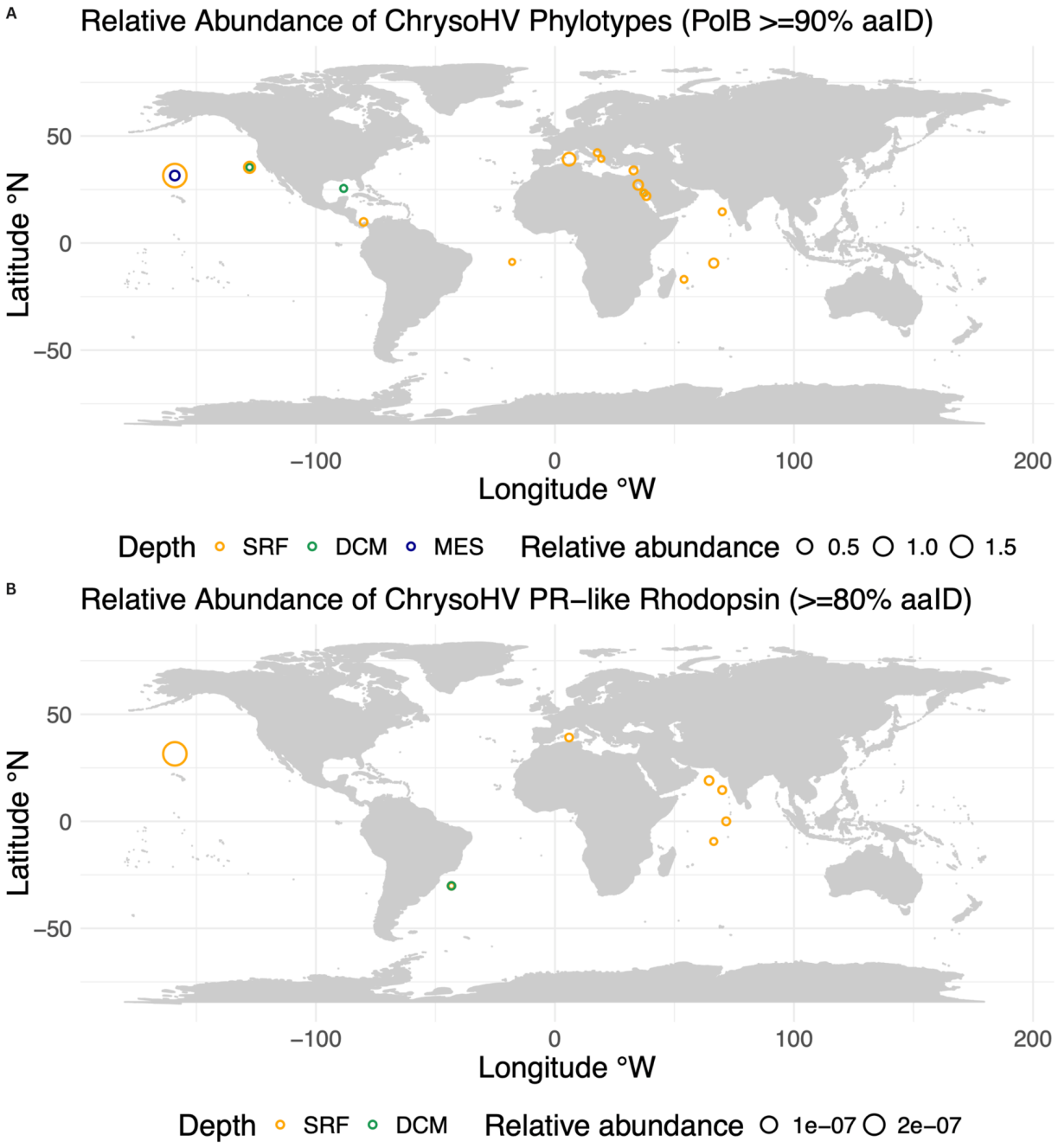


Figure 3. **Rhodopsin gene phylogeny, predicted structure, and gene neighborhood.** (A) Microbial rhodopsin ML phylogeny with heliorhodopsin as the outgroup, built with model LG+F+R10. Select representatives of viral rhodopsin homologs and putative chrysophyte rhodopsins are referred to in the parenthesis of their corresponding collapsed clade. The proteorhodopsin (PR) clade is annotated with the alignment of transmembrane helix 3 (TM3) outlining the ion-pumping motif (gray) and the spectral tuning amino acid (green or blue). Residues that differ from the most common residue at that locus are in bold (B). The predicted structure of ChrysoHV's proteorhodopsin-like protein's conserved domain (yellow) and the HOT 75m4 Gamma-proteobacterium proteorhodopsin (blue). (C) Gene neighborhood of photoreceptive proteins in ChrysoHV, showing the proteorhodopsin (locus tag: ChrysoHV_0101;

GenBank ID: XVM28022.1) next to the enzyme lysine decarboxylase (locus tag: ChrysoHV_0099, GenBank ID: XVM28020.1), between two photolyases (locus tag: ChrysoHV_0095, ChyrsoHV_0104; GenBank ID: XVM28016.1, XVM28025.1), and one heliorhodopsin (locus tag: ChrysoHV_0106, GenBank ID: XVM28027.1).



461 **Figure 4**. **Distribution and relative abundance of ChrysoHV PolB and PR**
462 **(proteorhodopsin)**. (A) Global distribution of phylotypes with ≥ 90% PolB aaID to ChrysoHV
463 from the *Tara* Oceans dataset. Size of the circle indicates the relative abundance of the summed
464 phylotypes at each given sample site to total *Nucleocytoviricota* PolB sequences per site. (B)
465 Global distribution of homologous PR with ≥ 80% aaID to the ChrysoHV proteorhodopsins with
466 size of the circle indicating relative abundance normalized to total metagenomic reads per site.
467 Sample depth is shown through color: surface (SRF: 3-7 m, orange), deep chlorophyll maximum
468 (DCM: 7-200 m, green), and mesopelagic (MES: 200-1,000 m, purple).
469

470 **References**

471 1. Del Campo J, Massana R. 2011. Emerging diversity within chrysophytes, choanoflagellates
472 and bicosoecids based on molecular surveys. Protist 162:435–448.

473 2. Seeleuthner Y, Mondy S, Lombard V, Carradec Q, Pelletier E, Wessner M, Leconte J,
474 Mangot J-F, Poulain J, Labadie K, Logares R, Sunagawa S, De Berardinis V, Salanoubat
475 M, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators, Acinas
476 SG, Boss E, Follows M, Gorsky G, Grimsley N, Karp-Boss L, Krzic U, Not F, Ogata H,
477 Raes J, Reynaud EG, Sardet C, Speich S, Stemmann L, Velayoudon D, Weissenbach J,
478 Pesant S, Poulton N, Stepanauskas R, Bork P, Bowler C, Hingamp P, Sullivan MB,
479 Iudicone D, Massana R, Aury J-M, Henrissat B, Karsenti E, Jaillon O, Sieracki M, De
480 Vargas C, Wincker P. 2018. Single-cell genomics of multiple uncultured stramenopiles
481 reveals underestimated functional diversity across oceans. Nat Commun 9:310.

482 3. Olefeld JL, Majda S, Albach DC, Marks S, Boenigk J. 2018. Genome size of chrysophytes
483 varies with cell size and nutritional mode. Org Divers Evol 18:163–173.

4. Dai M, Luo Y, Achterberg EP, Browning TJ, Cai Y, Cao Z, Chai F, Chen B, Church MJ, Ci D, Du C, Gao K, Guo X, Hu Z, Kao S, Laws EA, Lee Z, Lin H, Liu Q, Liu X, Luo W, Meng F, Shang S, Shi D, Saito H, Song L, Wan XS, Wang Y, Wang W, Wen Z, Xiu P, Zhang J, Zhang R, Zhou K. 2023. Upper ocean biogeochemistry of the oligotrophic North Pacific Subtropical Gyre: from nutrient sources to carbon export. Reviews of Geophysics 61:e2022RG000800.

5. Frias-Lopez J, Thompson A, Waldbauer J, Chisholm SW. 2009. Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. Environmental Microbiology 11:512–525.

6. Wilken S, Yung CCM, Poirier C, Massana R, Jimenez V, Worden AZ. 2023. Choanoflagellates alongside diverse uncultured predatory protists consume the abundant open-ocean cyanobacterium *Prochlorococcus*. Proc Natl Acad Sci USA 120:e2302388120.

7. Li Q, Edwards KF, Schvarcz CR, Steward GF. 2022. Broad phylogenetic and functional diversity among mixotrophic consumers of *Prochlorococcus*. The ISME Journal 16:1557–1569.

8. Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, Karl DM, Li WKW, Lomas MW, Veneziano D, Vera CS, Vrugt JA, Martiny AC. 2013. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. Proc Natl Acad Sci USA 110:9824–9829.

503 9. Castillo YM, Mangot J, Benites LF, Logares R, Kuronishi M, Ogata H, Jaillon O, Massana
504 R, Sebastián M, Vaqué D. 2019. Assessing the viral content of uncultured picoeukaryotes
505 in the global-ocean by single cell genomics. Molecular Ecology 28:4272–4289.

506 10. Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. 2021. A phylogenomic framework for
507 charting the diversity and evolution of giant viruses. PLoS Biol 19:e3001430.

508 11. Endo H, Blanc-Mathieu R, Li Y, Salazar G, Henry N, Labadie K, De Vargas C, Sullivan
509 MB, Bowler C, Wincker P, Karp-Boss L, Sunagawa S, Ogata H. 2020. Biogeography of
510 marine giant viruses reveals their interplay with eukaryotes and ecological functions. Nat
511 Ecol Evol 4:1639–1649.

512 12. Fromm A, Hevroni G, Vincent F, Schatz D, Martinez-Gutierrez CA, Aylward FO, Vardi A.
513 2024. Single-cell RNA-seq of the rare virosphere reveals the native hosts of giant viruses in
514 the marine environment. Nat Microbiol 9:1619–1629.

515 13. Moniruzzaman M, Erazo Garcia MP, Farzad R, Ha AD, Jivaji A, Karki S, Sheyn U, Stanton
516 J, Minch B, Stephens D, Hancks DC, Rodrigues RAL, Abrahao JS, Vardi A, Aylward FO.
517 2023. Virologs, viral mimicry, and virocell metabolism: the expanding scale of cellular
518 functions encoded in the complex genomes of giant viruses. FEMS Microbiology Reviews
519 47:fuad053.

520 14. Needham DM, Yoshizawa S, Hosaka T, Poirier C, Choi CJ, Hehenberger E, Irwin NAT,
521 Wilken S, Yung C-M, Bachy C, Kurihara R, Nakajima Y, Kojima K, Kimura-Someya T,
522 Leonard G, Malmstrom RR, Mende DR, Olson DK, Sudo Y, Sudek S, Richards TA,
523 DeLong EF, Keeling PJ, Santoro AE, Shirouzu M, Iwasaki W, Worden AZ. 2019. A

524 distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine
525 predators. Proc Natl Acad Sci USA 116:20574–20583.

526 15. Rozenberg A, Oppermann J, Wietek J, Fernandez Lahore RG, Sandaa R-A, Bratbak G,
527 Hegemann P, Béjà O. 2020. Lateral gene transfer of anion-conducting channelrhodopsins
528 between green algae and giant viruses. Current Biology 30:4910-4920.e5.

529 16. Thomy J, Schvarcz CR, McBeain KA, Edwards KF, Steward GF. 2024. Eukaryotic viruses
530 encode the ribosomal protein eL40. npj Viruses 2:51.

531 17. Hososhima S, Mizutori R, Abe-Yoshizumi R, Rozenberg A, Shigemura S, Pushkarev A,
532 Konno M, Katayama K, Inoue K, Tsunoda SP, Béjà O, Kandori H. 2022. Proton-
533 transporting heliorhodopsins from marine giant viruses. eLife 11:e78416.

534 18. Yutin N, Koonin EV. 2012. Proteorhodopsin genes in giant viruses. Biol Direct 7:34.

535 19. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. 2020. Dynamic
536 genome evolution and complex virocell metabolism of globally-distributed giant viruses.
537 Nat Commun 11:1710.

538 20. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denef VJ, McMahon KD,
539 Konstantinidis KT, Eloe-Fadrosh EA, Kyrpides NC, Woyke T. 2020. Giant virus diversity
540 and host interactions through global metagenomics. Nature 578:432–436.

541 21. Pinhassi J, DeLong EF, Béjà O, González JM, Pedrós-Alió C. 2016. Marine bacterial and
542 archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. Microbiol
543 Mol Biol Rev 80:929–954.

544 22. Rozenberg, A Inoue, K, Kandori, H, Béjà O. Microbial rhodopsins: the last two decades.
545 Annu Rev Microbiol 75:427-447.

546 23. Bratanov D, Kovalev K, Machtens J-P, Astashkin R, Chizhov I, Soloviov D, Volkov D,
547 Polovinkin V, Zabelskii D, Mager T, Gushchin I, Rokitskaya T, Antonenko Y, Alekseev A,
548 Shevchenko V, Yutin N, Rosselli R, Baeken C, Borshchevskiy V, Bourenkov G, Popov A,
549 Balandin T, Büldt G, Manstein DJ, Rodriguez-Valera F, Fahlke C, Bamberg E, Koonin E,
550 Gordeliy V. 2019. Unique structure and function of viral rhodopsins. Nat Commun
551 10:4939.

552 24. Zabelskii D, Alekseev A, Kovalev K, Rankovic V, Balandin T, Soloviov D, Bratanov D,
553 Savelyeva E, Podolyak E, Volkov D, Vaganova S, Astashkin R, Chizhov I, Yutin N, Rulev
554 M, Popov A, Eria-Oliveira A-S, Rokitskaya T, Mager T, Antonenko Y, Rosselli R, Armeev
555 G, Shaitan K, Vivaudou M, Büldt G, Rogachev A, Rodriguez-Valera F, Kirpichnikov M,
556 Moser T, Offenhäusser A, Willbold D, Koonin E, Bamberg E, Gordeliy V. 2020. Viral
557 rhodopsins 1 are an unique family of light-gated cation channels. Nat Commun 11:5707.

558 25. Sheam MM, Luo E. 2025. Vertical transport and spatiotemporal dynamics of giant viruses
559 in the North Pacific subtropical gyre. The ISME Journal 19:wraf094.

560 26. Bosmon T, Abergel C, Claverie J-M. 2025. 20 years of research on giant viruses. npj
561 Viruses 3:9.

562 27. Brussaard CPD, Bratbak G, Baudoux A-C, Ruardij P. 2007. Phaeocystis and its interaction
563 with viruses. Biogeochemistry 83:201–215.

564 28. Lamb DC, Goldstone JV, Belhaouari DB, Andréani J, Farooqi A, Allen MJ, Kelly SL, La
565 Scola B, Stegeman JJ. 2025. Cytochrome b5 occurrence in giant and other viruses
566 belonging to the phylum Nucleocytoviricota. npj Viruses 3:8.

567 29. Olson DK, Yoshizawa S, Boeuf D, Iwasaki W, DeLong EF. 2018. Proteorhodopsin
568 variability and distribution in the North Pacific Subtropical Gyre. The ISME Journal
569 12:1047–1060.

570 30. Zhang L, Meng L, Fang Y, Ogata H, Okazaki Y. 2024. Spatiotemporal dynamics of giant
571 viruses within a deep freshwater lake reveal a distinct dark-water community. The ISME
572 Journal 18:wrae182.

573 31. Pitot TM, Girard C, Rapp JZ, Somerville V, Culley AI, Vincent WF, Moineau S, Roux S.
574 2025. Viral niche-partitioning: comparative genomics of giant viruses across environmental
575 gradients in a high Arctic freshwater-saltwater lake. ISME Communications 5:ycae155.

576 32. Fang J, Zhang Y, Zhu T, Li Y. 2023. Scramblase activity of proteorhodopsin confers
577 physiological advantages to Escherichia coli in the absence of light. iScience 26:108551.

578 33. Ran T, Ozorowski G, Gao Y, Sineshchekov OA, Wang W, Spudich JL, Luecke H. 2013.
579 Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin
580 complexes. Acta Crystallogr D Biol Crystallogr 69:1965–1980.

581 34. Fischer MG, Allen MJ, Wilson WH, Suttle CA. 2010. Giant virus with a remarkable
582 complement of genes infects marine zooplankton. Proc Natl Acad Sci USA 107:19508–
583 19513.

584 35. Sheng Y, Wu Z, Xu S, Wang Y. 2022. Isolation and identification of a large green alga
585 virus (*Chlorella* Virus XW01) of *mimiviridae* and its virophage (*Chlorella* Virus Virophage
586 SW01) by using unicellular green algal cultures. J Virol 96:e02114-21.

587 36. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch
588 C, Lewis M, Hoffman JM, Gibson JA, Cavicchioli R. 2011. Virophage control of antarctic
589 algal host–virus dynamics. Proc Natl Acad Sci USA 108:6163–6168.

590 37. Strauss J, Deng L, Gao S, Toseland A, Bachy C, Zhang C, Kirkham A, Hopes A, Utting R,
591 Joest EF, Tagliabue A, Löw C, Worden AZ, Nagel G, Mock T. 2023. Plastid-localized
592 xanthorhodopsin increases diatom biomass and ecosystem productivity in iron-limited
593 surface oceans. Nat Microbiol 8:2050–2066.

594 38. Giovannoni SJ. 2017. SAR11 Bacteria: The most abundant plankton in the oceans. Annu
595 Rev Mar Sci 9:231–255.

596 39. Filée J. 2018. Giant viruses and their mobile genetic elements: the molecular symbiosis
597 hypothesis. Current Opinion in Virology 33:81–88.

598 40. Letelier RM, White AE, Bidigare RR, Barone B, Church MJ, Karl DM. 2017. Light
599 absorption by phytoplankton in the North Pacific Subtropical Gyre. Limnology &
600 Oceanography 62:1526–1540.

601 41. Keller MD, Selvin RC, Claus W, Guillard RRL. 1987. Media for the culture of oceanic
602 ultraphytoplankton. Journal of Phycology 23:633–638.

603 42. Bratbak G, Jacobsen A, Heldal M, Nagasaki K, Thingstad F. 1998. Virus production in
604 Phaeocystis pouchetii and its relation to host cell growth and nutrition. Aquat Microb Ecol
605 16:1–9.

606 43. Li Q, Edwards KF, Schvarcz CR, Selph KE, Steward GF. 2021. Plasticity in the grazing
607 ecophysiology of *Florenciella* (Dichtyochophyceae), a mixotrophic nanoflagellate that
608 consumes *Prochlorococcus* and other bacteria. Limnology & Oceanography 66:47–60.

609 44. Brussaard CPD. 2004. Optimization of Procedures for Counting Viruses by Flow
610 Cytometry. Appl Environ Microbiol 70:1506–1513.

611 45. Christaki U, Courties C, Massana R, Catala P, Lebaron P, Gasol JM, Zubkov MV. 2011.
612 Optimized routine flow cytometric enumeration of heterotrophic flagellates using SYBR
613 Green I. Limnology & Ocean Methods 9:329–339.

614 46. Zhao Y, Zhao Y, Zheng S, Zhao L, Zhang W, Xiao T, Grégori G. 2023. Enhanced
615 resolution of marine viruses with violet side scatter. Cytometry Pt A 103:260–268.

616 47. Doane FW, Anderson N. 1987. Electron microscopy in diagnostic virology: a practical
617 guide and atlas. Cambridge University Press, New York.

618 48. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S,
619 Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K,
620 Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis.
621 Nat Methods 9:676–682.

622 49. PacificBiosciences. pbdmarkdup. https://github.com/PacificBiosciences/pbmarkdup (25
623 February 2025, date last accessed)

624 50. BioInfoTools. BBMap. https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh
625 (21 February 2021, date last accessed)

626 51. Kbaseapps. BBTools. https://github.com/kbaseapps/BBTools?tab=readme-ov-file (21
627 February 2021, date last accessed)

628 52. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads
629 using repeat graphs. Nat Biotechnol 37:540–546.

630 53. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ,
631 Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and
632 composition. Nat Methods 11:1144–1146.

633 54. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated
634 binning method to recover individual genomes from metagenomes using an expectation-
635 maximization algorithm. Microbiome 2:26.

636 55. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an
637 adaptive binning algorithm for robust and efficient genome reconstruction from
638 metagenome assemblies. PeerJ 7:e7359.

639 56. Pan S, Zhao X-M, Coelho LP. 2023. SemiBin2: self-supervised contrastive learning leads
640 to better MAGs for short- and long-read sequencing. Bioinformatics 39:i21–i29.

641 57. Pitot TM, Brůna T, Schulz F. 2024. Conservative taxonomy and quality assessment of giant
642 virus genomes with GVClass. npj Viruses 2:60.

643 58. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly
644 from long uncorrected reads. Genome Res 27:737–746.

645 59. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020.
646 KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score
647 threshold. Bioinformatics 36:2251–2252.

648 60. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-
649 mapper v2: functional annotation, orthology assignments, and domain prediction at the
650 metagenomic scale. Molecular Biology and Evolution 38:5825–5829.

651 61. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
652 BLAST+: architecture and applications. BMC Bioinformatics 10:421.

653 62. Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and
654 functional classification of transfer RNA genes. Nucleic Acids Research 49:9077–9096.

655 63. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A,
656 Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic
657 inference in the genomic era. Molecular Biology and Evolution 37:1530–1534.

658 64. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.
659 2012. GenBank. Nucleic Acids Research 41:D36–D42.

660 65. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195.

661 66. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
662 improvements in performance and usability. Molecular Biology and Evolution 30:772–780.

663 67. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated
664 alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

665 68. NeLli-team. nsgtree. https://github.com/NeLLi-team/nsgtree (03 March 2024, date last
666 accessed)

667 69. Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the
668 phylogenetic tree display and annotation tool. Nucleic Acids Research 52:W78–W82.

669 70. Needham DM, Poirier C, Hehenberger E, Jiménez V, Swalwell JE, Santoro AE, Worden
670 AZ. 2019. Targeted metagenomic recovery of four divergent viruses reveals shared and
671 distinctive characteristics of giant viruses of marine eukaryotes. Phil Trans R Soc B
672 374:20190086.

673 71. Villar E, Vannier T, Vernette C, Lescot M, Cuenca M, Alexandre A, Bachelerie P, Rosnet
674 T, Pelletier E, Sunagawa S, Hingamp P. 2018. The Ocean Gene Atlas: exploring the
675 biogeography of plankton genes online. Nucleic Acids Research 46:W289–W295.

676 72. Vernette C, Lecubin J, Sánchez P, Tara Oceans Coordinators, Acinas SG, Babin M, Bork P,
677 Boss E, Bowler C, Cochrane G, De Vargas C, Gorsky G, Guidi L, Grimsley N, Hingamp1
678 P, Iudicone D, Jaillon O, Kandels-Lewis S, Karp-Boss L, Karsenti E, Not F, Ogata H,
679 Poulton N, Pesant S, Sardet C, Speich S, Stemmann L, Sullivan MB, Sunagawa S, Wincker
680 P, Sunagawa S, Delmont TO, Acinas SG, Pelletier E, Hingamp P, Lescot M. 2022. The

681 Ocean Gene Atlas v2.0: online exploration of the biogeography and phylogeny of plankton
682 genes. Nucleic Acids Research 50:W516–W526.

683 73. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022.
684 ColabFold: making protein folding accessible to all. Nat Methods 19:679–682.

685 74. The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic
686 Acids Res 45:D158–D169.

687 75. Waman VP, Bordin N, Lau A, Kandathil S, Wells J, Miller D, Velankar S, Jones DT,
688 Sillitoe I, Orengo C. 2025. CATH v4.4: major expansion of CATH by experimental and
689 predicted structural data. Nucleic Acids Research 53:D348–D355.

690 76. Van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J,
691 Steinegger M. 2024. Fast and accurate protein structure search with Foldseek. Nat
692 Biotechnol 42:243–246.

693 77. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. 2023.
694 UCSF CHIMERAX: Tools for structure building and analysis. Protein Science 32.

695