Title

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

Authors

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]

Affiliations

[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, 1950 East-West Road, Honolulu, HI 96822.

[2]Department of Oceanography, School of Ocean and Earth Science and Technology (SOEST), University of Hawai‘i at Mānoa, 1000 Pope Road, Honolulu, HI 96822.

[3]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720.

*Corresponding author(s): pkbyl@hawaii.edu; grieg@hawaii.edu

Abstract

Chrysophytes are widespread nanoflagellate protists in aquatic ecosystems with diverse trophic strategies ranging from phototrophy to phagotrophy. Molecular evidence has suggested that giant viruses infect chrysophytes, and here, we report the first isolation of a chrysophyte-infecting virus, Chrysophyceae clade H virus SA1 (ChrysoHV). Both host and virus were isolated from surface waters in the subtropical North Pacific. The ChrysoHV capsid (291 ± 42 nm diameter) is associated with a loose, sac-like membrane that extends its effective diameter (716 ± 115 nm), and presents a long, thin tail extending 1,202 (± 243) 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, and the closest related rhodopsin sequences from cultivated organisms derive from bacteria in the genus *Pelagibacter*. Therefore, the rhodopsin homolog in ChrysoHV might be representative of a previously undescribed lineage of virus rhodopsin, laterally transferred from the chrysophyte's putative bacterial prey. In addition, the ChrysoHV genome contains nine genes, mostly related to nutrient uptake, whose closest known homologs come from *Prochlorococcus*. This highly abundant cyanobacterium is known to be consumed by the chrysophyte host, providing further evidence for prey-virus gene transfer in the evolution of ChrysoHV.

Key words

Nucleocytoviricota, horizontal gene transfer (HGT), phytoplankton, mixotrophy, SAR11, *Pelagibacter*, *Prochlorococcus*

Main Text

Introduction

Giant viruses are large double-stranded DNA viruses in the phylum *Nucleocytoviricota* [1], and while they are diverse and globally abundant [2–4], a relatively small fraction of their known diversity is represented by cultivated isolates [5]. Giant viruses display extensive genomic repertoires, including viral homologs of cellular genes [6], which may augment or compensate for host functions compromised during infection [7–9] or interfere with host defenses [10].

Chrysophytes are single-celled or colonial protists in the class Chrysophyceae, and they are globally abundant in freshwater and marine ecosystems [11]. Chrysophyte taxa include phago-mixotrophic forms that perform photosynthesis and also ingest prey, heterotrophic forms that do not perform photosynthesis, and potentially non-phagotrophic phototrophs [12]. Chrysophytes are known to be predators of bacteria [13, 14], and isolates from Chrysophyceae clade H have been shown to be voracious consumers of the cyanobacterium *Prochlorococcus*, the most abundant phototroph in the ocean and a dominant primary producer in oligotrophic waters [15]. Although molecular evidence suggests that chrysophytes are infected by viruses in the *Nucleocytoviricota* phylum [2, 16–18], to date there are no isolated chrysophyte viruses. Here we report the isolation of a giant virus infecting a clade H chrysophyte, ChrysoHV.

Microbial rhodopsins are light-activated ion pumps that may supplement a given organism's energetic demands for growth and survival [19], and they are among the most commonly found viral homologs of cellular genes [3, 4, 20]. Viral rhodopsin homologs have been experimentally confirmed to convert light energy into electrochemical potential across multiple clades: viral rhodopsin groups I and II [7, 21, 22], channelrhodopsin [8], and

70 heliorhodopsin [10]. Furthermore, these viral rhodopsin homologs are globally distributed across
71 aquatic habitats, with their relative abundance decreasing over depth with the attenuation of
72 sunlight [7, 23].

73 Although giant virus rhodopsin homologs are widespread and potentially functional, only
74 a few rhodopsin-encoding giant viruses have been isolated and remain in culture with their native
75 host [9, 10, 24]. In this study we identify a previously undescribed lineage of viral rhodopsin,
76 descended from a bacterial proteorhodopsin, encoded by ChrysoHV. The chrysophyte-virus pair
77 we characterize presents an ecologically relevant system for studying viral homologs of cellular
78 genes and other aspects of viral ecology and evolution.

79

80 **Materials and Methods**

81 *Isolation and cultivation*

82 The six mixotrophic protists used in this study include two chrysophytes (UHM3500 and
83 UHM3501), a bolidophyte (UHM3600), and three dictyochophytes (UHM3020, UHM3053, and
84 UHM3071), all of which were isolated from the epipelagic zone (5-100m) of Station ALOHA
85 through serial dilutions of whole seawater samples enriched with *Prochlorococcus*. Additional
86 details on this isolation method are described in Li *et al*. (2022) [15]. The mixotrophs were
87 cultivated as xenic, unialgal cultures on K medium minus dissolved, inorganic nitrogen sources
88 ($NaNO_3$ and NH_4Cl). Cultures were instead amended with cyanobacterial prey (*Prochlorococcus*
89 MIT9301) for their nitrogen source [15, 25]. In addition, there were two heterotrophic flagellates
90 used in this study, chrysophytes (UHM3530 and UHM3560) were isolated from seawater
91 collected from the mixed layer (5-50 m) at Station ALOHA in 2022 and 2024 through serial

92 dilutions of seawater samples. The heterotrophs were cultivated on filtered seawater amended
93 with quinoa to stimulate ambient bacteria as their food source.
94 The Chrysophyceae clade H virus SA1 (ChrysoHV) was isolated by challenging
95 *Chrysophyceae clade H* sp. UHM3500 with unfiltered seawater collected from 25 m depth at
96 Station ALOHA on 04 May 2019. This is the first chrysophyte-infecting virus isolate from
97 Station ALOHA, thus the strain designation SA1. The lytic agent detected in the challenge was
98 rendered clonal by repeated serial dilutions to extinction [26]. After discovering that ChrysoHV
99 could infect chrysophytes UHM3500 and UHM3501 (Supplementary Fig. S1), two
100 *Chrysophyceae clade H* representatives with distinct grazing capabilities and slightly different
101 18S rRNA identities [15], the virus was maintained through monthly challenges at a 300 µL:25
102 mL volumetric ratio of virus: chrysophyte on both UHM3500 and UHM3501, followed by
103 storage of whole lysate at 4°C, 3-4 days post-infection (DPI). Experiments were performed at
104 ~200-220 µE m^{-2} s^{-1} irradiance on a 12:12 cycle to mimic light conditions in the upper mixed
105 layer of the North Pacific Subtropical Gyre at Station ALOHA [27].
106 *Host range tests and flow cytometry*
107 We challenged in-house Ochrophyta protists, both mixotrophs and heterotrophs, with raw
108 ChrysoHV lysate, then tracked the population dynamics of the protists and viruses over time
109 using flow cytometry. Due to differences in growth patterns, the mixotrophic algal protists were
110 tracked every two days up to ten DPI, and the heterotrophic protists were tracked every day up to
111 five DPI. The experimental design included triplicate, 25 mL protist cultures for both negative
112 controls and infection treatments. To prepare the infection treatment, 500 µL of raw lysate (3-6 ×
113 10^5 virus-like particles mL^{-1} final concentration) was added to exponential or early stationary
114 phase protist cultures. Sample collection and storage for flow cytometry counts were based on

115 the protocol in Li *et al*. (2021) [28]. The samples were fixed with ice-cold glutaraldehyde at
116 0.5% final concentration, incubated for 15 min at 4°C, and then flash frozen in liquid nitrogen.
117 The samples were stored at -80°C up to four days, then thawed at room temperature and
118 processed within the same day on a CytoFlex-S (Beckman Coulter Life Sciences, Indianapolis,
119 IN, USA) in the Biological Electron Microscope Facility (BEMF) at the University of Hawai‘i at
120 Mānoa. The samples were split into two fractions, with 200 µL allocated towards cell
121 abundances and loaded directly into a 96-well plate, and 5 µL allocated toward virus abundances.
122 These latter 5 µL fractions were diluted into 195 µL solutions of TE (pH 8.0) with 2.5X SYBR
123 Green I nucleic acid stain (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) and
124 incubated at 80°C for 20 min in darkness based on the protocol of Brussaard (2004) [29], then
125 loaded onto a 96-well plate. The cell counts were run on auto-record for one minute at a flow
126 rate of 20 µL min^{-1} for the algae and 30 µL min^{-1} for the heterotrophs. The mixotroph
127 populations were distinguished through chlorophyll fluorescence and 488 nm, 90° side scatter
128 (SSC), whereas the heterotrophs were detected using SYBR Green fluorescence and 488 nm, 90°
129 SSC [30]. The viral-like particle (VLP) population was identified through SYBR Green
130 fluorescence and 405 nm, 90° violet side scatter (VSSC), enhancing marine virus detection [31].
131 The samples were run on auto-record for one minute at a flow rate of 10 µL min^{-1}. Using the
132 shapefile output, both algae and virus populations were delineated through a custom Python
133 script using the following packages: fcsparser (v0.2.8), pandas (v2.2.2), matplotlib (v3.7.1),
134 while the FlowJo (v.10.10.0) software was used to delineate the heterotrophic protist population.
135 *Electron microscopy*
136 Transmission electron microscope (TEM) micrographs of viral particles were prepared
137 from raw, unfixed lysate and negatively stained with 2% uranyl acetate for 45 sec on 200 mesh

138 copper grids coated with carbon-stabilized formvar support that had been rendered hydrophilic
139 by glow discharge [32]. Images were captured with a Hitachi HT7700 transmission electron
140 microscope (Tokyo, Japan) at the Biological Electron Microscope Facility (BEMF) at the
141 University of Hawai‘i at Mānoa. Particle dimensions were measured with ImageJ from Fiji
142 (v2.16.0/1.54p) [33], including: capsid diameter (n = 38), envelope (n = 34), tail (n = 17), and
143 tail fibers (n = 6).

144 *DNA extraction, sequencing, assembly, and annotation*

145 To prepare sufficient viral biomass for the DNA sequencing, we extracted DNA from 610
146 mL of raw lysate (6.4×10^6 VLP/mL), after allowing the infection (MOI: 1) of *Chrysophyceae*
147 *clade H* sp. UHM3501 to clear for four days. The lysate was concentrated and pooled through
148 serial centrifugations (6X) in a 100 kDA Centricon Plus centrifugal filter (MilliporeSigma,
149 Burlington, MA, USA) at 3,500 × *g* at 20°C for 15 minutes. Approximately 200 ng of DNA was
150 extracted from the concentrate using the ZymoBiomics Quick-DNA High Molecular Weight
151 MagBead Kit (Zymo Research, Irvine, CA, USA), flash frozen in liquid nitrogen, stored at -
152 20°C, then shipped on dry ice to the US Department of Energy Joint Genome Institute (JGI).
153 There, the DNA was sequenced using PacBio Multiplexed 6-10 kb Low Input parameters.

154 A PacBio Low Input (DNA) library was constructed and sequenced using the PacBio
155 SEQUELIIe platform (Pacific Biosciences, Menlo Park, CA USA), which generated 3,267,906
156 reads totaling 31,033,080,767 bp. Reads were deduplicated using pbmarkdup (v1.0.3,
157 pbdmarkup) [34]. BBDuk (v38.99) [35] was used to remove reads that contained PacBio control
158 sequences and PCR adapters, and IceCreamFinder (v38.99) [36] was used to remove reads
159 without SMRTbell sequences. The final filtered fastq contained 3,257,268 reads totaling
160 30,921,351,264 bp with mean read length of 9,493 bp.

161 The filtered reads (31 Gb) were assembled using Flye (v2.9) [37] with the "--meta --
162 pacbio-hifi" argument. From this initial assembly, viral contigs were identified and grouped
163 through a low-complexity binning workflow, utilizing CONCOCT (v1.1.0) [38], MaxBin
164 (v2.2.7) [39], MetaBAT2 (v2.15-5) [40], and SemiBin2 (v1.5.1) [41]. For each binning strategy,
165 the viral bin was identified by assessing the presence of giant virus core marker genes described
166 in Aylward *et al*. (2021) [42] through GVClass (v1.0) [43] and the MetaBAT2-generated viral
167 bin was used in downstream analysis. Subsequently, circular consensus sequencing reads were
168 recruited against these identified viral contigs, and a 100 Mb subsample of these reads was
169 assembled using Flye without the "--meta" argument (v2.9) [37]. This iterative assembly resulted
170 in one high-quality viral contig, which was polished using two rounds of Racon (v1.6.0) [44] to
171 ensure accuracy.

172 Gene calling was performed with GVClass (v1.0) [43], and the 1.196 Mb viral genome
173 had an estimated 93.77% coding density. Gene annotation was performed with KofamScan
174 (v1.3.0) [45] and eggnog mapper (v2.1.12) [46] with annotations of genes of interests confirmed
175 with blastp [47]. tRNAs were annotated with tRNAscan-SE (v2.0.12) [48].

176 *Phylogenetic analysis*

177 Phylogenetic analysis of ChrysoHV with IQ-TREE (v2.2.6) [49] was based on a
178 concatenated alignment of seven core proteins recommended in Aylward *et al*. (2021) [42].
179 Query sequences were retrieved from the GVMAGs [4] and GenBank Assembly [50] databases,
180 and quality filtered to contain at least four out of seven core proteins determined through
181 hmmsearch (v3.3.2, E-value $\leq 1 \times 10^{-10}$) [51]. Additionally, the query sequences were filtered to
182 remove genomes with more than 10 copies of any core protein and more than four core proteins
183 with multiple copies. The remaining 990 query sequences were aligned with MAFFT (v7.520)

184 [52] and trimmed with trimAl (-gt 0.1; v1.4.1) [53], then concatenated with a custom python
185 script [54]. The trimmed, concatenated alignment was used for phylogenetic analysis using the
186 mixture model LG+I+F+G4 and ultrafast bootstrap supports (option: -bb 1000) (v2.2.6) [49].
187 The tree was annotated with iTOL (v6) [55].

188 Gene phylogeny of microbial rhodopsins with a heliorhodopsin outgroup were built with
189 IQ-TREE (v2.2.6) [49] using model LG+F+R10 and ultrafast bootstrap supports (option: -bb
190 1000) and annotated with iTOL (v6) [55]. Representatives of rhodopsin homologs from
191 cultivated virus isolates include: EhV-202 (heliorhodopsin) [10], PgV-12T and -16T (viral
192 rhodopsin I) [24], FloV-SA2 (viral rhodopsin II) [9]. Additionally, the gene phylogeny includes
193 synthetic construct: vPyACR_21921 (channelrhodopsin) [8], and rhodopsin homologs from
194 single amplified genomes (SAGs) and metagenome assembled genomes (MAGS): ChoanoV1
195 (viral group II) [7], cPacV (viral group II) [56], OLPV 1 and 2 (viral groups I & II) [57]. Query
196 rhodopsin sequences were retrieved from NCBI and the supplementary materials of Olson *et al*.
197 (2018) [58]. The alignment was built with MAFFT (L-INS-i; v7.520) [52] and trimmed with
198 trimAl (-gt 0.1; v1.4.1) [53].

199 The phylogeny of near-complete 18S rRNA genes from select algae isolates were built
200 with IQ-TREE (v2.2.6) [49] using GTR+I+G model with ultrafast bootstrap support (option: -bb
201 1000) and annotated with iTOL (v6) [55]. The nucleotide sequences were retrieved from NCBI,
202 aligned with MAFFT (L-INS-i; v7.520) [52] and trimmed with trimAl (-gt 0.1; v1.4.1) [53].

203 *Gene distribution in the global ocean*

204 Comparisons to PolB amino acid sequences from the Tara Oceans dataset were
205 performed using blastp (v2.13.0, E-value $\leq 1 \times 10^{-120}$) [47], and sequences with 90% or greater
206 average amino acid identity (AAI) were plotted using a custom R script with the following

207 packages maps (v3.4.2), mapdata (v2.3.1), and ggplot2 (v.3.5.1). The relative abundance of the
208 PolB genes were normalized to total *Nucleocytoviricota* PolBs as in Endo *et al*. (2020) [2].
209 Likewise, ChysoHV proteorhodopsin-like protein sequences from the Tara Oceans dataset with a
210 minimum of 80% AAI were identified through Ocean Gene Atlas [59, 60]. The relative
211 abundance of ChysoHV proteorhodopsin-like protein sequences were normalized to total
212 metagenomic reads in the given sample [59, 60].

213 *Protein structure predictions*

214 The conserved domain of the ChrysoHV rhodopsin structure was predicted using
215 ColabFold (v1.5.5) [61] and identified using Foldseek [62] against the AlphaFold Protein
216 Structure Databases– SwissProt [63] and CATH (v4.4) [64]. The conserved domain of the
217 ChrysoHV rhodopsin and HOT 75m4 Gamma-proteobacterium proteorhodopsin structures were
218 rendered in ChimeraX (v1.8.dev202403010247) [65].

219

220 **Results and Discussion**

221 *Host range, infection dynamics, and morphology*

222 We tested the host range of ChrysoHV against eight protist isolates from Station
223 ALOHA, resulting in no lysis for six of the eight isolates (Fig. 1A, C). Representatives of
224 *Chrysophyceae clade H*, UHM3500 and UHM3501, lysed when challenged with ChrysoHV.
225 However, a greater proportion of UHM3501 cells lysed, resulting in the production of more
226 virions (Fig. 1B and Fig. S1). Despite this, the number of virions produced per cell (burst size)
227 was similar for each host strain, ranging from 22 to 25.5. The lower mortality rate of
228 *Chrysophyceae clade H* sp. UHM3500 may be due to the presence of resistant subpopulations or
229 lineage-specific host immune responses. Additionally, the activation of integrated virophages is a

230 hypothesis that should not be ruled out, as virophages are widespread in uncultivated members of
231 *Chrysophyceae clade H* [16, 66]. Beyond differences in mortality, the virion particles produced
232 from the two chrysophytes shared the same notable morphological features, *i.e.*, an icosahedral
233 capsid (291 ± 42 nm diameter [mean ± SD]) associated with a loose membrane (716 ± 115 nm),
234 and a long tail (1,202 ± 243 nm length) ending in tail fibers (206 ± 72 nm length) (Fig. 2A-C).
235 *Genomic characterization*
236 The ChrysoHV genome was assembled as a putatively linear DNA sequence of 1.19 Mbp
237 encoding 1,138 predicted genes with a G+C content of 23.42% (Fig. 2B). The ChrysoHV
238 genome encodes genes of diverse putative function, including the ribosomal proteins eL40 and
239 S27a, making it the second isolated eukaryote-infecting virus known to encode ribosomal
240 proteins, and the first to encode S27a (Supplementary Table S1) [9]. ChrysoHV also encodes
241 electron transport proteins such as cytochrome b5, widespread in prokaryotes, eukaryotes, and
242 giant viruses (Supplementary Table S1) [67], along with succinate dehydrogenase flavoprotein
243 and iron-sulfur subunits (Supplementary Table S1). Additionally, the virus encodes seven ABC
244 transporters— including one specific to phosphate transport (XVM28869.1) and two AmtB
245 ammonium transporters. Of these ammonium transporters, one (XVM28992.1) has 55% AAI to
246 *Prochlorococcus* and *Synechococcus* homologs (Supplementary Table S1). Likewise, ChrysoHV
247 encodes three porins with 44-66% AAI to *Prochlorococcus* iron and carbohydrate porins
248 (XVM28452.1, XVM28569.1, XVM28570.1), two urea ABC transporter substrate-binding
249 proteins with 81-90% AAI to a *Prochlorococcus* homolog (XVM29005.1, XVM29006.1), a
250 semiSWEET sugar transporter with 44% AAI to a *Prochlorococcus* homolog (XVM28124.1), a
251 sulfotransferase family protein with 55% AAI to a *Prochlorococcus* homolog (XVM28964.1),
252 and a N-acetylmuramoyl-L-alanine amidase for cell wall remodeling with 39% AAI to a

253 *Prochlorococcus* homolog (XVM29061.1) (Supplementary Table S1). The presence of a large
254 number of genes whose closest known relatives are *Prochlorococcus* homologs suggests
255 extensive horizontal gene transfer (HGT) from cyanobacterial prey of the chrysophyte to a giant
256 virus that infects it. In addition, the high frequency of nutrient transport proteins suggests that the
257 viral infection strategy involves modification of host nutrient uptake, or potentially viral mimicry
258 of the host's prey via cell surface proteins.

259 The ChrysoHV genome also includes a region of light-sensitive proteins, encompassing a
260 proteorhodopsin-like protein surrounded by two photolyases and followed by a heliorhodopsin
261 (Fig. 2D, Fig. 3C, Supplementary Table S1). The positioning and shared orientation of the
262 photolyases and heliorhodopsin around the proteorhodopsin-like protein implies that they may be
263 part of the same operon and function like the heliorhodopsin in *Trichococcus flocculiformis* [68].
264 It was proposed that this protein could bind to nearby photolyases, enhancing DNA repair [68].

265 *Comparison to Nucleocytovirica representatives*

266 ChrysoHV is the third cultivated isolate in the *Aliimimivirinae* subfamily of *Mimiviridae*
267 (Fig. 2E); other known isolates include CroV BV-PW1 [69], infecting a heterotrophic
268 nanoflagellate, and CV-XW01 [70], infecting a green alga. In this subfamily only two other
269 representatives, cPacV-1605 and GVMAG-M-3300013131-43, encode rhodopsins, which belong
270 to the viral rhodopsin group I clade (Fig. 2A). However, 40 other *Aliimimivirinae* representatives
271 encode at least one photolyase homolog (Fig. 2E), either the (6-4) photolyase or DASH-type
272 cryptochrome. Only ChrysoHV and cPacV-1605 encode one microbial rhodopsin with multiple
273 heliorhodopsins, and photolyase homologs, but cPacV's rhodopsin is placed within the Viral
274 Rhodopsin Group II lineage (Fig. 3A). Furthermore, instead of being nested within a light
275 sensitive gene neighborhood, the position of the viral rhodopsin in cPacV-1605 (locus tag:

276 CPAV1605_207, GenBank ID: VVU94485.1) is further away from the photolyases (locus tag:
277 CPAV1605_133, CPAV1605_136; GenBank ID: VVU94411.1, VVU94414.1) and
278 heliorhodopsins (locus tag: CPAV1605_533, CPAV1605_1309; GenBank ID: VVU95557.1,
279 VVU94808.1) [57]. ChrysoHV is the only cultivated giant virus to encode the proteorhodopsin
280 lineage of rhodopsin along with two heliorhodopsins (Fig 2E, Fig. 3A). Given the mismatched
281 distribution of rhodopsin and photolyase homologs within the *Aliimimivirinae* subfamily, the
282 presence of multiple, diverse rhodopsins in ChrysoHV may represent a unique adaptation to high
283 light in the surface waters of the subtropical ocean where it was isolated.

284 *Sequence characterization of a distinct lineage of proteorhodopsin*

285 Phylogenetic analysis of ChrysoHV's rhodopsin placed it within the proteorhodopsin
286 clade (Fig. 3A). These results suggest that ChrysoHV may have acquired the proteorhodopsin-
287 like rhodopsin from *Pelagibacter*, which is a member of the highly abundant and widespread
288 SAR11 clade of bacteria. Like *Prochlorococcus*, it is probably a prey organism of the
289 chrysophyte host. The sequence alignment of transmembrane domain helix-III (TM3) indicates
290 that the ChrysoHV rhodopsin is spectrally tuned to blue-light (glutamine, Q; Fig. 3A), which is
291 the dominant spectral tuning amino acid in proteorhodopsins from Station ALOHA, the isolation
292 site of ChrysoHV [58]. The two most closely related rhodopsin sequences were recovered from
293 metagenomes from the Black Sea at 50 m depth (SAMN27960419) and the North Pacific at 5 m
294 depth in the 0.22-3μm size fraction (SAMEA2623059). These rhodopsins may also absorb blue
295 light (Fig. 3A) and share a poly-N (asparagine) tract at the start of the rhodopsin coding
296 sequence.

297 While ChrysoHV lacks the conserved DTE proton pumping motif found in most
298 proteorhodopsins, it possesses an ETL motif (Fig. 3A), which has been detected in

299 proteorhodopsins from deep water at Station ALOHA (Fig. 3A) [58]. A comparison of the
300 predicted structure of the conserved domain of the ChrysoHV rhodopsin against existing
301 functional characterizations revealed that it was most similar to the uncultured gamma-
302 proteobacterium clone HOT 75m4 rhodopsin (Fig. 3B). This rhodopsin was confirmed to pump
303 protons (H^+) [71]. Although the AAI was not the highest (35.9%), the HOT 75m4
304 proteorhodopsin was also retrieved from Station ALOHA [71].

305 ChrysoHV does not encode β-carotene, a precursor to retinal, the light-absorbing
306 chromophore in proteorhodopsin, and the salt bridge between K102 and E228 most likely
307 inhibits retinal binding. Further investigation is needed to confirm whether the function of the
308 ChrysoHV proteorhodopsin-like protein is light-independent [72]. This loss of retinal-binding
309 could explain the long branch of the ChrysoHV PR-like rhodopsins (Fig. 3A). Understanding
310 how the rhodopsin may function without retinal-binding could be important to interpreting the
311 function of viral rhodopsins detected below the photic zone [73, 74].

312 *Global ocean distribution of closely related ChrysoHV phylotypes and proteorhodopsins*

313 Analysis on the distribution of ChrysoHV phylotypes (≥ 90% PolB AAI) from the TARA
314 Oceans dataset [2] revealed their prevalence in subtropical surface ocean sites in the Pacific and
315 Indian Ocean as well as the Mediterranean Sea (Fig. 4A). Notably, at TARA station 132, north of
316 the Hawaiian archipelago, ChrysoHV phylotypes were detected in both the mixed layer and
317 mesopelagic zone, suggesting connectivity between surface and deep ocean virus populations.
318 The abundance of the ChrysoHV proteorhodopsin-like protein (≥ 80% PolB AAI) was also
319 highest in the subtropical surface ocean (Fig. 4B), but detected at fewer overall sites than the
320 ChrysoHV PolB. Additionally, the ChrysoHV proteorhodopsin relative abundance was highest at

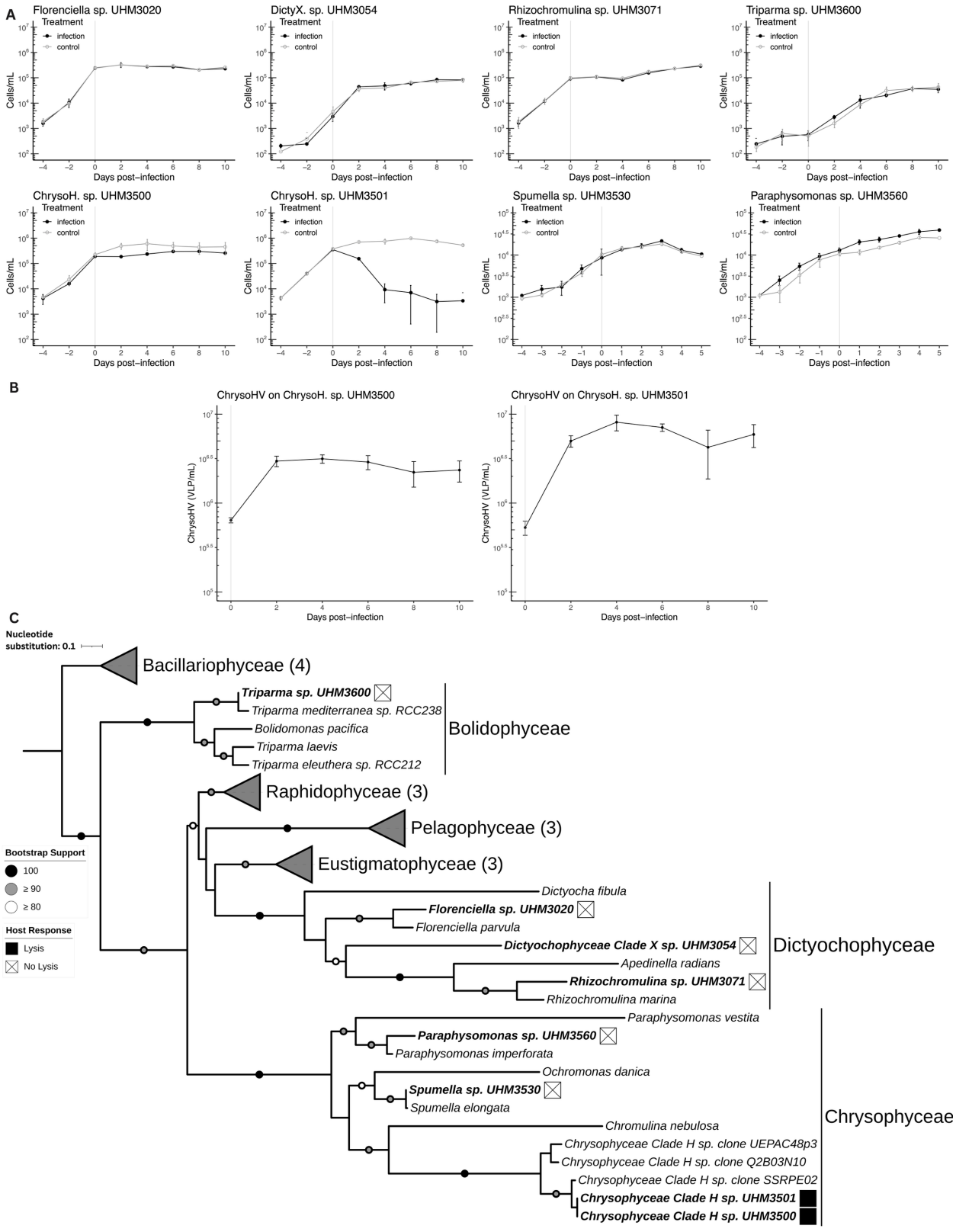
321 TARA station 132, indicating that these proteorhodopsin sequences may be derived from a
322 ChrysoHV phylotype.

323 *Conclusion*

324 In the marine environment, diverse microorganisms from proteobacteria to microalgae
325 utilize rhodopsins to supplement their energetic needs [58, 75]. SAR11 bacteria, of which
326 *Pelagibacter* is an abundant subclade, account for 25% of all plankton [76], and are a likely prey
327 source for mixotrophic algae such as *Chrysophyceae clade H* sp. UHM3500 and UHM3501. The
328 discovery of a proteorhodopsin-like protein encoded by a giant virus supports the conceptual
329 model that lateral gene transfer is driven by the co-mingling of infectious virions and ingested
330 prey within the same host cell [77]. The presence of nine genes whose closest known relatives
331 are *Prochlorococcus* homologs further suggests that prey-virus HGT has played an important
332 role in the evolution of ChrysoHV. By leveraging naturally co-occurring and experimentally
333 accessible virus-host pairs, this study lays the groundwork for future investigations into how the
334 virocell environment of predatory protists influences gene acquisition in giant viruses. As the
335 only cultivated isolate known to encode one microbial rhodopsin and two heliorhodopsins, this
336 system provides the opportunity for novel insights into the role of rhodopsins in giant virus
337 infection strategies.

338

339 **Figures**


A
Florenciella sp. UHM3020
DictyX. sp. UHM3054
Rhizochromulina sp. UHM3071
Triparma sp. UHM3600
ChrysoH. sp. UHM3500
ChrysoH. sp. UHM3501
Spumella sp. UHM3530
Paraphysomonas sp. UHM3560
Treatment
infection
control
Cells/mL
Days post−infection
B
ChrysoHV on ChrysoH. sp. UHM3500
ChrysoHV on ChrysoH. sp. UHM3501
ChrysoHV (VLP/mL)
Days post−infection
C
Nucleotide substitution: 0.1
Bacillariophyceae (4)
Triparma sp. UHM3600
Triparma mediterranea sp. RCC238
Bolidomonas pacifica
Triparma laevis
Triparma eleuthera sp. RCC212
Bolidophyceae
Raphidophyceae (3)
Pelagophyceae (3)
Eustigmatophyceae (3)
Dictyocha fibula
Florenciella sp. UHM3020
Florenciella parvula
Dictyochophyceae Clade X sp. UHM3054
Apedinella radians
Rhizochromulina sp. UHM3071
Rhizochromulina marina
Dictyochophyceae
Paraphysomonas vestita
Paraphysomonas sp. UHM3560
Paraphysomonas imperforata
Ochromonas danica
Spumella sp. UHM3530
Spumella elongata
Chromulina nebulosa
Chrysophyceae Clade H sp. clone UEPAC48p3
Chrysophyceae Clade H sp. clone Q2B03N10
Chrysophyceae Clade H sp. clone SSRPE02
Chrysophyceae Clade H sp. UHM3501
Chrysophyceae Clade H sp. UHM3500
Chrysophyceae
Bootstrap Support
100
≥ 90
≥ 80
Host Response
Lysis
No Lysis

340

341 **Figure 1**. (A) Time series of protist abundance (cells/mL) in response to the addition of raw
342 ChrysoHV lysate. Days pre- and post-infection are displayed on the x-axis. The algae were
343 isolated from Station ALOHA in the North Pacific Subtropical Gyre, and include: *Florenciella*
344 so. UHM3020, *Dictyochophyceae clade X* sp. UHM3054, *Rhizochromulina* sp. UHM3071,
345 *Chrysophyceae clade H* sp. UHM3500, *Chrysophyceae clade H* sp. UHM3501, *Triparma* sp.
346 UHM3600, *Spumella* sp. UHM3530, and *Paraphysomonas* sp. UHM3560. The control treatment
347 is shown in the gray line with white-fill points, and the infection treatment shown in the black
348 line with black-fill points. Error bars represent the standard deviation of the triplicate counts. (B)
349 Time series of ChrysoHV abundance (VLP/mL) propagated on permissive hosts. Error bars
350 represent the standard deviation of the triplicate counts. (C) ML 18S rRNA phylogeny
351 (GTR+I+R model) of the Ochrophyta phylum with isolates challenged in bold. The white or
352 black boxes by the bolded isolate names indicate whether the protist culture did not lyse (white
353 with "X") or lysed and produced virions (black) in response to the ChrysoHV challenge. The
354 node symbols indicate the bootstrap values equal to or greater than 80% (white), 90% (gray) or
355 100% (black). There is no node symbol for branches with less than 80% support.

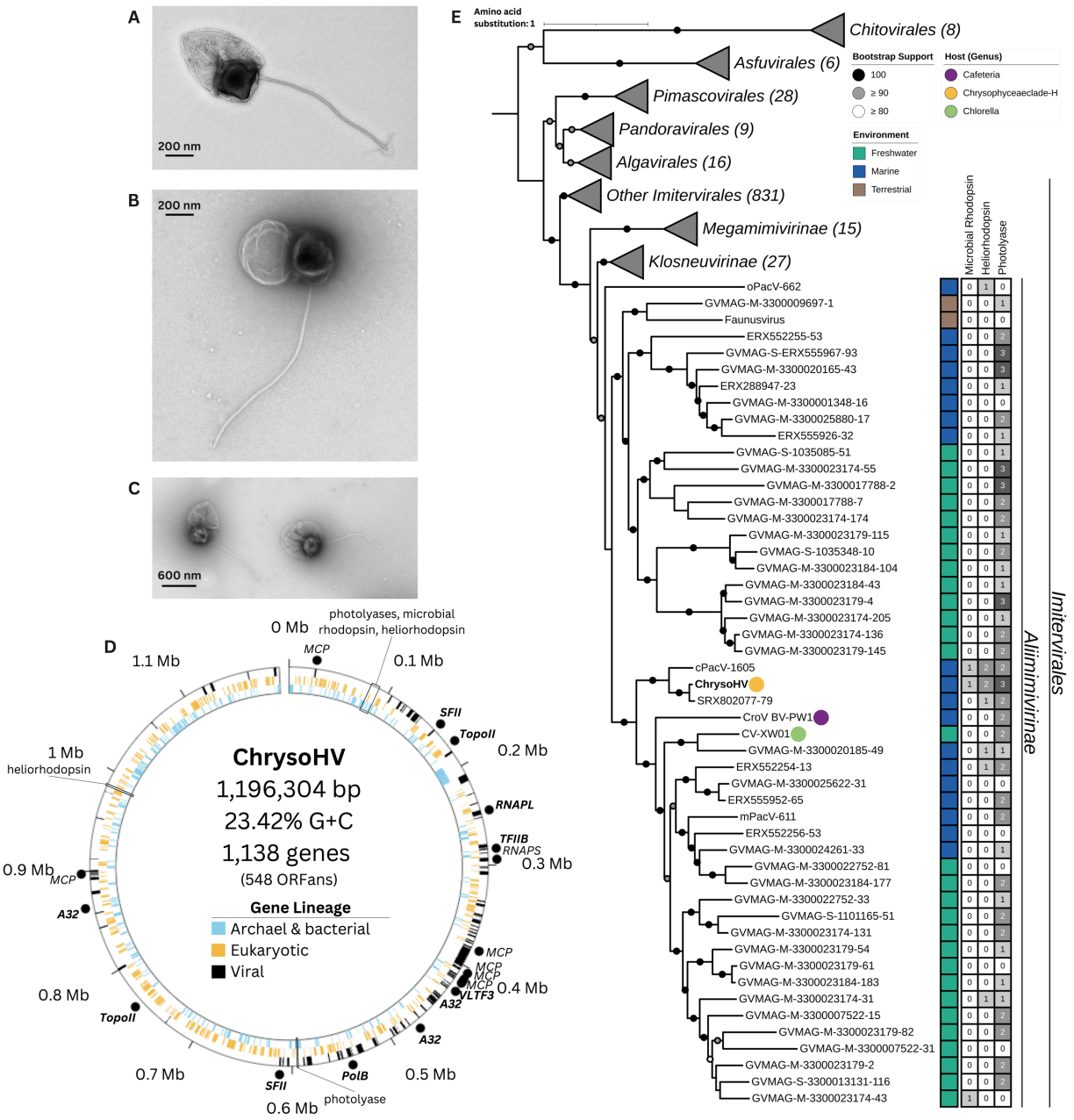


Figure 2. (A) Transmission electron microscopy (TEM) negative stain micrograph of ChrysoHV propagated on UHM3501. (B, C) TEM negative stain micrograph of ChrysoHV propagated on UHM3500. (D) Genome plot of ChrysoHV with track placement and color showing gene lineage annotated according to domain, inner to outer tracks showing: prokaryote (blue), eukaryote (yellow), and virus (black). Phylogenetically-informative giant virus marker genes are

362 represented as black dots on the outside track and show gene name. The genes used in the
363 *Nucleocytoviricota* phylogenetic analysis are bolded. (E) Maximum Likelihood (ML) phylogeny
364 of the *Nucleocytoviricota* phyluminferred from a concatenated alignment of amino acid
365 sequences of marker genes (TopoII, RNAPL, TFIIB, VLTFB, A32, PolB, SFII) using model
366 LG+I+F+G4 with ultrafast bootstrap supports (-bb 1000) shown by node symbol color– white (≥
367 80%), gray (≥ 90%), and black (100%) with no node symbol for supports < 80%. The tree is
368 rooted with *Chitovirales* as the outgroup, and the scale bar represents one amino acid substitution
369 per site. The frequency of light-sensitive proteins: microbial rhodopsins, heliorhodopsins, and
370 photolyases present in the *Aliimimivirinae* subfamily are annotated in the grayscale matrix and
371 the environment the virus was collected from is shown in jade (freshwater), blue (marine), and
372 brown (terrestrial).

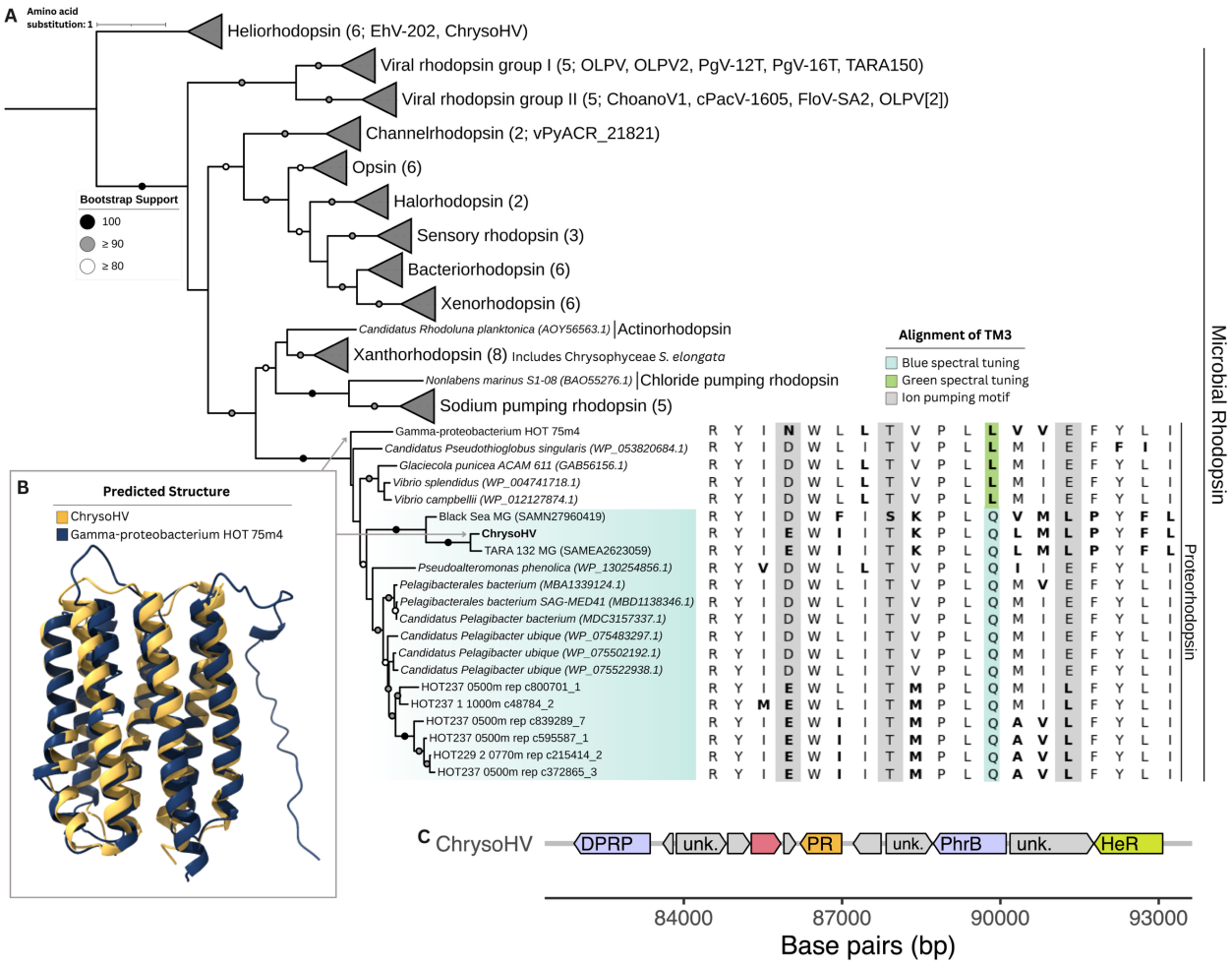
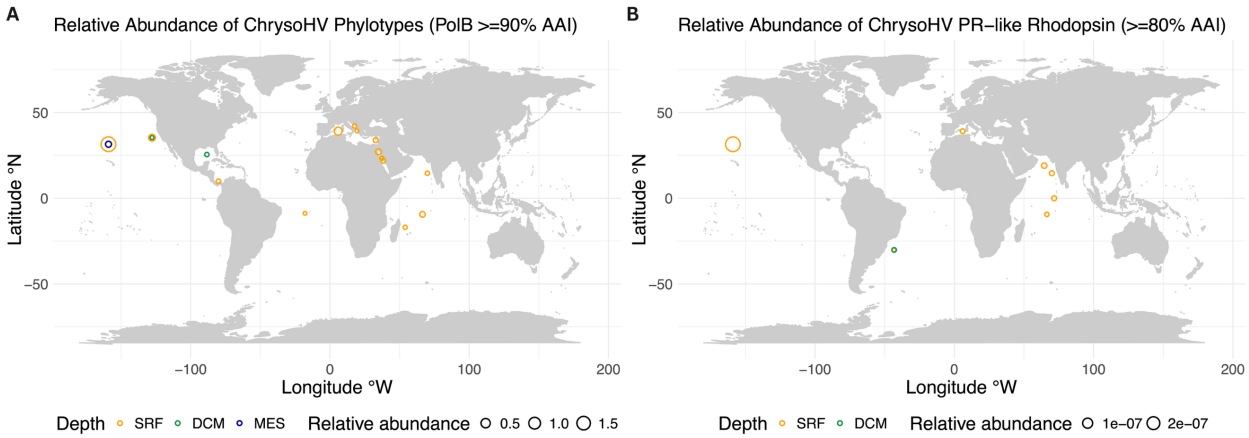


Figure 3. (A) Microbial rhodopsin ML phylogeny with heliorhodopsin as the outgroup, built with model LG+F+R10. Select representatives of viral rhodopsin homologs and putative chrysophyte rhodopsins are referred to in the parenthesis of their corresponding collapsed clade. The proteorhodopsin (PR) clade 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 conserved domain (yellow) and the HOT 75m4 Gamma-proteobacterium proteorhodopsin (blue). (C) Gene neighborhood of light-sensitive proteins in ChrysoHV, showing the microbial rhodopsin (locus tag: ChrysoHV_0101; GenBank ID: XVM28022.1) between two photolyases (locus tag: ChrysoHV_0095,

384 ChyrsoHV_0104; GenBank ID: XVM28016.1, XVM28025.1) and one heliorhodopsin (locus

385 tag: ChrysoHV_0106, GenBank ID: XVM28027.1).

386



387

388 **Figure 4**. (A) Global distribution of phylotypes with ≥ 90% PolB AAI to ChrysoHV from the

389 Tara Oceans dataset. Size of the circle indicates the relative abundance of the summed

390 phylotypes at each given sample site to total *Nucleocytoviricota* PolB sequences per site. (B)

391 Global distribution of homologous proteorhodopsins with ≥ 80% AAI to the ChrysoHV

392 proteorhodopsins with size of the circle indicating relative abundance normalized to total

393 metagenomic reads per site. Sample depth is shown through color: surface (SRF: 3-7 m, orange),

394 deep chlorophyll maximum (DCM: 7-200 m, green), and mesopelagic (MES: 200-1,000 m,

395 purple).

396

397 **Acknowledgements**

398 We thank the personnel of the Hawai‘i Ocean Time-series program (NSF award 12-60164) and

399 Research Vessel Kilo Moana for assistance with seawater collection. We thank the Biological

400 Electron Microscope Facility (BEMF) at the University of Hawai‘i at Mānoa, especially facility

401 supervisors, Tina Weatherby and Dr. Orion Rivers, for their assistance with electron microscopy

402 procedures. A special thanks to laboratory technician Kelsey McBeain for maintenance of the
403 UHM library culture collection, and US DOE Joint Genome Institute Scientific Project Manager,
404 Kerrie Barry, for overseeing DNA library preparation, quality control, and sequencing of
405 ChrysoHV.
406 We are grateful for the thoughtful feedback provided by Dr. Karen Selph (Department of
407 Oceanography, University of Hawai‘i at Mānoa), Dr. Alison Sherwood (School of Life Sciences,
408 University of Hawai‘i at Mānoa), Dr. Oded Béjà (Faculty of Biology, Technion-Israel Institute of
409 Technology), and Dr. Andrey Rozenberg (Faculty of Biology, Technion-Israel Institute of
410 Technology).
411
412 **Funding**
413 This work was supported by the US National Science Foundation Graduate Research Fellowship
414 Program (Grant No. 1842402 to PB), Division of Ocean Sciences (Grant No. 2129697 to GFS
415 and KFE), and RII Track-2 FEC (Grant No. 1736030 to GFS and KFE). The work conducted by
416 the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a Department
417 of Energy (DOE) Office of Science User Facility, was supported by the Office of Science of the
418 US DOE, operated under Contract No. DE-AC02-05CH11231.
419
420 **Author Contributions**
421 GFS, KFE, and PB developed, designed, and acquired funding support for the project. CRS
422 isolated the virus and algae strains. CW and QL assisted with the collection of virus host-range
423 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.

References

446 1. Aylward FO et al. Taxonomic update for giant viruses in the order Imitervirales (phylum
447 Nucleocytoviricota). *Arch Virol* 2023;**168**:283. https://doi.org/10.1007/s00705-023-05906-3
448 2. Endo H et al. Biogeography of marine giant viruses reveals their interplay with eukaryotes and
449 ecological functions. *Nat Ecol Evol* 2020;**4**:1639–1649. https://doi.org/10.1038/s41559-020-
450 01288-w
451 3. Moniruzzaman M et al. Dynamic genome evolution and complex virocell metabolism of
452 globally-distributed giant viruses. *Nat Commun* 2020;**11**:1710. https://doi.org/10.1038/s41467-
453 020-15507-2
454 4. Schulz F et al. Giant virus diversity and host interactions through global metagenomics. *Nature*
455 2020;**578**:432–436. https://doi.org/10.1038/s41586-020-1957-x
456 5. Bosmon T, Abergel C, Claverie J-M. 20 years of research on giant viruses. *Npj Viruses*
457 2025;**3**:9. https://doi.org/10.1038/s44298-025-00093-1
458 6. Moniruzzaman M et al. Virologs, viral mimicry, and virocell metabolism: the expanding scale
459 of cellular functions encoded in the complex genomes of giant viruses. *FEMS Microbiol Rev*
460 2023;**47**:fuad053. https://doi.org/10.1093/femsre/fuad053
461 7. Needham DM et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to
462 unicellular marine predators. *Proc Natl Acad Sci* 2019;**116**:20574–20583.
463 https://doi.org/10.1073/pnas.1907517116
464 8. Rozenberg A et al. Lateral gene transfer of anion-conducting channelrhodopsins between green
465 algae and giant viruses. *Curr Biol* 2020;**30**:4910-4920.e5.
466 https://doi.org/10.1016/j.cub.2020.09.056
467 9. Thomy J et al. Eukaryotic viruses encode the ribosomal protein eL40. *Npj Viruses* 2024;**2**:51.
468 https://doi.org/10.1038/s44298-024-00060-2

469 10. Hososhima S et al. Proton-transporting heliorhodopsins from marine giant viruses. *eLife*
470 2022;**11**:e78416. https://doi.org/10.7554/eLife.78416
471 11. Del Campo J, Massana R. Emerging diversity within chrysophytes, choanoflagellates and
472 bicosoecids based on molecular surveys. *Protist* 2011;**162**:435–448.
473 https://doi.org/10.1016/j.protis.2010.10.003
474 12. Olefeld JL et al. Genome size of chrysophytes varies with cell size and nutritional mode. *Org*
475 *Divers Evol* 2018;**18**:163–173. https://doi.org/10.1007/s13127-018-0365-7
476 13. Frias-Lopez J et al. Use of stable isotope-labelled cells to identify active grazers of
477 picocyanobacteria in ocean surface waters. *Environ Microbiol* 2009;**11**:512–525.
478 https://doi.org/10.1111/j.1462-2920.2008.01793.x
479 14. Wilken S et al. Choanoflagellates alongside diverse uncultured predatory protists consume the
480 abundant open-ocean cyanobacterium *Prochlorococcus*. *Proc Natl Acad Sci*
481 2023;**120**:e2302388120. https://doi.org/10.1073/pnas.2302388120
482 15. Li Q et al. Broad phylogenetic and functional diversity among mixotrophic consumers of
483 *Prochlorococcus*. *ISME J* 2022;**16**:1557–1569. https://doi.org/10.1038/s41396-022-01204-z
484 16. Castillo YM et al. Assessing the viral content of uncultured picoeukaryotes in the global-ocean
485 by single cell genomics. *Mol Ecol* 2019;**28**:4272–4289. https://doi.org/10.1111/mec.15210
486 17. Fromm A et al. Single-cell RNA-seq of the rare virosphere reveals the native hosts of giant
487 viruses in the marine environment. *Nat Microbiol* 2024;**9**:1619–1629.
488 https://doi.org/10.1038/s41564-024-01669-y
489 18. Michoud G et al. Mapping the metagenomic diversity of the multi-kingdom glacier-fed stream
490 microbiome. *Nat Microbiol* 2025;**10**:217–230. https://doi.org/10.1038/s41564-024-01874-9

19. Pinhassi J et al. Marine Bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiol Mol Biol Rev* 2016;**80**:929–954. https://doi.org/10.1128/MMBR.00003-16

20. Yutin N, Koonin EV. Proteorhodopsin genes in giant viruses. *Biol Direct* 2012;**7**:34. https://doi.org/10.1186/1745-6150-7-34

21. Bratanov D et al. Unique structure and function of viral rhodopsins. *Nat Commun* 2019;**10**:4939. https://doi.org/10.1038/s41467-019-12718-0

22. Zabelskii D et al. Viral rhodopsins 1 are an unique family of light-gated cation channels. *Nat Commun* 2020;**11**:5707. https://doi.org/10.1038/s41467-020-19457-7

23. Sheam MM, Luo E. Vertical transport and spatiotemporal dynamics of giant viruses in the North Pacific subtropical gyre. *ISME J* 2025;**19**:wraf094. https://doi.org/10.1093/ismejo/wraf094

24. Brussaard CPD et al. Phaeocystis and its interaction with viruses. *Biogeochemistry* 2007;**83**:201–215. https://doi.org/10.1007/s10533-007-9096-0

25. Keller MD et al. Media for the culture of oceanic ultraphytoplankton[1,2]. *J Phycol* 1987;**23**:633–638. https://doi.org/10.1111/j.1529-8817.1987.tb04217.x

26. Nagasaki K, Bratbak G. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists. In: Wilhelm S, Weinbauer M, Suttle C (eds), Manual of Aquatic Viral Ecology. American Society of Limnology and Oceanography, 2010, 92–101.

27. Letelier RM et al. Light absorption by phytoplankton in the North Pacific Subtropical Gyre. *Limnol Oceanogr* 2017;**62**:1526–1540. https://doi.org/10.1002/lno.10515

28. Li Q et al. Plasticity in the grazing ecophysiology of *Florenciella* (Dichtyochophyceae), a mixotrophic nanoflagellate that consumes *Prochlorococcus* and other bacteria. *Limnol Oceanogr* 2021;**66**:47–60. https://doi.org/10.1002/lno.11585

29. Brussaard CPD. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol* 2004;**70**:1506–1513. https://doi.org/10.1128/AEM.70.3.1506-1513.2004

30. Christaki U et al. Optimized routine flow cytometric enumeration of heterotrophic flagellates using SYBR Green I. *Limnol Oceanogr Methods* 2011;**9**:329–339. https://doi.org/10.4319/lom.2011.9.329

31. Zhao Y et al. Enhanced resolution of marine viruses with violet side scatter. *Cytometry A* 2023;**103**:260–268. https://doi.org/10.1002/cyto.a.24674

32. Doane FW, Anderson N. Electron Microscopy in Diagnostic Virology: A Practical Guide and Atlas. New York: Cambridge University Press, 1987.

33. Schindelin J et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;**9**:676–682. https://doi.org/10.1038/nmeth.2019

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

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

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

37. Kolmogorov M et al. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;**37**:540–546. https://doi.org/10.1038/s41587-019-0072-8

534 38. Alneberg J et al. Binning metagenomic contigs by coverage and composition. *Nat Methods*
535 2014;**11**:1144–1146. https://doi.org/10.1038/nmeth.3103
536 39. Wu Y-W et al. MaxBin: an automated binning method to recover individual genomes from
537 metagenomes using an expectation-maximization algorithm. *Microbiome* 2014;**2**:26.
538 https://doi.org/10.1186/2049-2618-2-26
539 40. Kang DD et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome
540 reconstruction from metagenome assemblies. *PeerJ* 2019;**7**:e7359.
541 https://doi.org/10.7717/peerj.7359
542 41. Pan S, Zhao X-M, Coelho LP. SemiBin2: self-supervised contrastive learning leads to better
543 MAGs for short- and long-read sequencing. *Bioinformatics* 2023;**39**:i21–i29.
544 https://doi.org/10.1093/bioinformatics/btad209
545 42. Aylward FO et al. A phylogenomic framework for charting the diversity and evolution of giant
546 viruses. *PLOS Biol* 2021;**19**:e3001430. https://doi.org/10.1371/journal.pbio.3001430
547 43. Pitot TM, Brůna T, Schulz F. Conservative taxonomy and quality assessment of giant virus
548 genomes with GVClass. *Npj Viruses* 2024;**2**:60. https://doi.org/10.1038/s44298-024-00069-7
549 44. Vaser R et al. Fast and accurate de novo genome assembly from long uncorrected reads.
550 *Genome Res* 2017;**27**:737–746. https://doi.org/10.1101/gr.214270.116
551 45. Aramaki T et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and
552 adaptive score threshold. *Bioinformatics* 2020;**36**:2251–2252.
553 https://doi.org/10.1093/bioinformatics/btz859
554 46. Cantalapiedra CP et al. eggNOG-mapper v2: functional annotation, orthology assignments, and
555 domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;**38**:5825–5829.
556 https://doi.org/10.1093/molbev/msab293

557 47. Camacho C et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;**10**:421.
558 https://doi.org/10.1186/1471-2105-10-421
559 48. Chan PP et al. tRNAscan-SE 2.0: improved detection and functional classification of transfer
560 RNA genes. *Nucleic Acids Res* 2021;**49**:9077–9096. https://doi.org/10.1093/nar/gkab688
561 49. Minh BQ et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in
562 the genomic era. *Mol Biol Evol* 2020;**37**:1530–1534. https://doi.org/10.1093/molbev/msaa015
563 50. Benson DA et al. GenBank. *Nucleic Acids Res* 2012;**41**:D36–D42.
564 https://doi.org/10.1093/nar/gks1195
565 51. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;**7**:e1002195.
566 https://doi.org/10.1371/journal.pcbi.1002195
567 52. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
568 improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–780.
569 https://doi.org/10.1093/molbev/mst010
570 53. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment
571 trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–1973.
572 https://doi.org/10.1093/bioinformatics/btp348
573 54. NeLli-team. nsgtree. https://github.com/NeLLi-team/nsgtree (03 March 2024, date last
574 accessed)
575 55. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree
576 display and annotation tool. *Nucleic Acids Res* 2024;**52**:W78–W82.
577 https://doi.org/10.1093/nar/gkae268

578 56. Needham DM et al. Targeted metagenomic recovery of four divergent viruses reveals shared
579 and distinctive characteristics of giant viruses of marine eukaryotes. *Philos Trans R Soc B Biol*
580 *Sci* 2019;**374**:20190086. https://doi.org/10.1098/rstb.2019.0086
581 57. Yau S et al. Virophage control of antarctic algal host–virus dynamics. *Proc Natl Acad Sci*
582 2011;**108**:6163–6168. https://doi.org/10.1073/pnas.1018221108
583 58. Olson DK et al. Proteorhodopsin variability and distribution in the North Pacific Subtropical
584 Gyre. *ISME J* 2018;**12**:1047–1060. https://doi.org/10.1038/s41396-018-0074-4
585 59. Villar E et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online.
586 *Nucleic Acids Res* 2018;**46**:W289–W295. https://doi.org/10.1093/nar/gky376
587 60. Vernette C et al. The Ocean Gene Atlas v2.0: online exploration of the biogeography and
588 phylogeny of plankton genes. *Nucleic Acids Res* 2022;**50**:W516–W526.
589 https://doi.org/10.1093/nar/gkac420
590 61. Mirdita M et al. ColabFold: making protein folding accessible to all. *Nat Methods* 2022;**19**:679–
591 682. https://doi.org/10.1038/s41592-022-01488-1
592 62. Van Kempen M et al. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*
593 2024;**42**:243–246. https://doi.org/10.1038/s41587-023-01773-0
594 63. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*
595 2017;**45**:D158–D169. https://doi.org/10.1093/nar/gkw1099
596 64. Waman VP et al. CATH v4.4: major expansion of CATH by experimental and predicted
597 structural data. *Nucleic Acids Res* 2025;**53**:D348–D355. https://doi.org/10.1093/nar/gkae1087
598 65. Meng EC et al. UCSF CHIMERAX: Tools for structure building and analysis. *Protein Sci*
599 2023;**32**. https://doi.org/10.1002/pro.4792

600 66. Bellas C et al. Large-scale invasion of unicellular eukaryotic genomes by integrating DNA
601 viruses. *Proc Natl Acad Sci* 2023;**120**:e2300465120. https://doi.org/10.1073/pnas.2300465120
602 67. Lamb DC et al. Cytochrome b5 occurrence in giant and other viruses belonging to the phylum
603 Nucleocytoviricota. *Npj Viruses* 2025;**3**:8. https://doi.org/10.1038/s44298-025-00091-3
604 68. Shim J et al. Heliorhodopsin helps photolyase to enhance the DNA repair capacity. *Microbiol*
605 *Spectr* 2022;**10**:e02215-22. https://doi.org/10.1128/spectrum.02215-22
606 69. Fischer MG et al. Giant virus with a remarkable complement of genes infects marine
607 zooplankton. *Proc Natl Acad Sci* 2010;**107**:19508–19513.
608 https://doi.org/10.1073/pnas.1007615107
609 70. Sheng Y et al. Isolation and identification of a large green alga virus (*Chlorella* Virus XW01) of
610 *Mimiviridae* and Its Virophage (*Chlorella* Virus Virophage SW01) by using unicellular green
611 algal cultures. *J Virol* 2022;**96**:e02114-21. https://doi.org/10.1128/jvi.02114-21
612 71. Ran T et al. Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin
613 complexes. *Acta Crystallogr D Biol Crystallogr* 2013;**69**:1965–1980.
614 https://doi.org/10.1107/S0907444913017575
615 72. Fang J et al. Scramblase activity of proteorhodopsin confers physiological advantages to
616 Escherichia coli in the absence of light. *iScience* 2023;**26**:108551.
617 https://doi.org/10.1016/j.isci.2023.108551
618 73. Zhang L et al. Spatiotemporal dynamics of giant viruses within a deep freshwater lake reveal a
619 distinct dark-water community. *ISME J* 2024;**18**:wrae182.
620 https://doi.org/10.1093/ismejo/wrae182

621 74. Pitot TM et al. Viral niche-partitioning: comparative genomics of giant viruses across
622 environmental gradients in a high Arctic freshwater-saltwater lake. *ISME Commun*
623 2025;**5**:ycae155. https://doi.org/10.1093/ismeco/ycae155
624 75. Strauss J et al. Plastid-localized xanthorhodopsin increases diatom biomass and ecosystem
625 productivity in iron-limited surface oceans. *Nat Microbiol* 2023;**8**:2050–2066.
626 https://doi.org/10.1038/s41564-023-01498-5
627 76. Giovannoni SJ. SAR11 bacteria: the most abundant plankton in the oceans. *Annu Rev Mar Sci*
628 2017;**9**:231–255. https://doi.org/10.1146/annurev-marine-010814-015934
629 77. Filée J, Siguier P, Chandler M. I am what I eat and I eat what I am: acquisition of bacterial
630 genes by giant viruses. *Trends Genet* 2007;**23**:10–15. https://doi.org/10.1016/j.tig.2006.11.002