**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 chrysophytes are commonly infected by giant viruses, but no isolates of such viruses have been reported. Here, we describe the chrysophyte-infecting virus Chrysophyceae Clade H virus SA1 (ChrysoHV). Both mixotrophic host and virus were isolated from surface waters in the subtropical North Pacific Ocean. 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, and the most similar rhodopsin sequences from cultivated organisms derive from marine bacteria in the genus *Pelagibacter*. The rhodopsin homolog in ChrysoHV may thus represent a new lineage of viral rhodopsins. The ChrysoHV genome also contains nine genes, mostly related to nutrient uptake, with the closest known homologs in the marine cyanobacterium *Prochlorococcus*. Therefore, lateral gene transfer may be occurring between abundant marine bacterial clades and giant viruses, a process that may be promoted by phagotrophic mixotrophs as both giant viruses and bacterial prey are taken into the cell.

**Key words**

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

## 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]. Chrysophyte taxa include phago-mixotrophic forms that perform photosynthesis and ingest prey as well as strictly heterotrophic phagotrophs and potentially non-phagotrophic phototrophs [3]. Many chrysophytes are known to be predators of bacteria [4, 5], 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 [6]. Although molecular evidence suggests that chrysophytes are infected by viruses in the *Nucleocytoviricota* phylum [7–10], to date no chrysophyte virus isolates have been described in the literature. Here we report on the isolation and initial characterization of a giant virus that infects a Clade H chrysophyte, ChrysoHV.

Giant viruses are large double-stranded DNA viruses in the phylum *Nucleocytoviricota* [11], and they display extensive genomic repertoires, including viral homologs of cellular genes [12], which may augment or compensate for host functions compromised during infection [13–15] or interfere with host defenses [16]. Among the most commonly found viral homologs of cellular genes are microbial rhodopsins [17-19], which are light-activated ion pumps that may supplement a given organism's energetic demands for growth and survival [20]. Viral rhodopsin homologs have been experimentally confirmed to convert light energy into electrochemical potential across multiple clades: viral rhodopsin groups I and II [13, 21, 22], channelrhodopsin [14], and heliorhodopsin [16]. Furthermore, these viral rhodopsin homologs are globally

69 distributed across aquatic habitats, with their relative abundance decreasing over depth with the
70 attenuation of sunlight [13, 23].
71 Although giant viruses are diverse and globally abundant [7, 17, 18], a relatively small
72 fraction of their known diversity is represented by cultivated isolates [24]. Likewise, giant virus
73 rhodopsin homologs are widespread and potentially functional, but only a few rhodopsin-
74 encoding giant viruses have been isolated and remain in culture with their native host [15, 16,
75 25]. Cultivated isolates are important for establishing virus-host linkages, and can reveal novel
76 gene lineages that may go undetected in environmental datasets and clarify the evolutionary
77 history of highly divergent sequences [15]. Through the isolation and characterization of
78 ChrysoHV, we identify a previously undescribed lineage of viral rhodopsin that appears to derive
79 from bacterial proteorhodopsin.
80
81 **Materials and Methods**
82
83 *Isolation and cultivation*
84 The six mixotrophic protists used in this study include two chrysophytes (UHM3500 and
85 UHM3501), a bolidophyte (UHM3600), and three dictyochophytes (UHM3020, UHM3053, and
86 UHM3071), all of which were isolated from the epipelagic zone (5–100m) of Station ALOHA
87 through serial dilutions of whole seawater samples enriched with *Prochlorococcus*. Additional
88 details on this isolation method are described in Li *et al*. (2022) [6]. The mixotrophs were
89 cultivated at 24°C at approximately 200–220 µmol quanta $m^{-2}$ $s^{-1}$ to mimic the high light
90 conditions at station ALOHA [26]. The xenic, unialgal cultures were grown on K medium minus
91 dissolved, inorganic nitrogen sources ($NaNO_3$ and $NH_4Cl$). Cultures were instead amended with

92 cyanobacteria prey (*Prochlorococcus* MIT9301) for their nitrogen source [6, 27]. In addition,
93 two heterotrophic chrysophytes (UHM3530 and UHM3560) were used in this study. These
94 strains were isolated from seawater collected from the mixed layer (5–50 m) at Station ALOHA
95 in 2022 and 2024 through serial dilutions of seawater samples. The heterotrophs were cultivated
96 on filtered seawater amended with sterile quinoa grains to stimulate growth of ambient bacteria
97 as their food source.

98 The Chrysophyceae Clade H virus SA1 (ChrysoHV) was isolated by challenging the
99 chrysophyte isolate UHM3500 with unfiltered seawater collected from 25 m depth at Station
100 ALOHA on 04 May 2019. This is the first chrysophyte-infecting virus isolate from Station
101 ALOHA, thus the strain designation SA1. The lytic agent detected in the challenge was rendered
102 clonal by repeated serial dilutions to extinction [28]. In addition to the original isolation host, the
103 virus was found to infect a related chrysophyte strain UHM3501 (Supplementary Fig. S1). Both
104 strains are members of Chrysophyceae Clade H, but they differ in their grazing kinetics and 18S
105 rRNA gene sequences [6]. The virus was maintained on both strains through serial, monthly
106 challenges of 25 mL of exponential phase chrysophyte culture with 0.3 mL of lysate from the
107 preceding challenge. Approximately 3–4 days post-infection (DPI), infected cultures were stored
108 at 4°C to serve as lysate for the next round of infection.

109

110 *Host range tests and flow cytometry*

111 We challenged in-house Ochrophyta protists, both mixotrophs and heterotrophs, with raw
112 ChrysoHV lysate, then tracked the population dynamics of the protists and viruses over time
113 using flow cytometry. The mixotrophic algal protists were tracked every two days up to ten DPI,
114 and the heterotrophic protists were tracked every day up to five DPI. The experimental design

115 included triplicate, 25 mL protist cultures for both negative controls and infection treatments. To
116 prepare the infection treatment, 500 µL of unfiltered lysate (3-6 × $10^5$ virus-like particles $mL^{-1}$
117 final concentration) was added to exponential or early stationary phase protist cultures. Sample
118 collection and storage for flow cytometry counts were based on the protocol in Li *et al*. (2021)
119 [29]. The samples were fixed with ice-cold glutaraldehyde at 0.5% final concentration, incubated
120 for 15 min at 4°C, and then flash frozen in liquid nitrogen.

121 The samples were stored at -80°C up to four days, then thawed at room temperature and
122 processed within the same day on a CytoFlex-S (Beckman Coulter Life Sciences, Indianapolis,
123 IN, USA) in the Biological Electron Microscope Facility (BEMF) at the University of Hawai‘i at
124 Mānoa. The samples were split into two fractions, with 200 µL allocated towards cell
125 abundances and loaded directly into a 96-well plate, and 5 µL allocated toward virus abundances.
126 The latter fractions were diluted into 195 µL solutions of TE (pH 8.0) with 2.5X SYBR Green I
127 nucleic acid stain (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) and incubated at
128 80°C for 20 min in darkness based on the protocol of Brussaard (2004) [30], then loaded onto a
129 96-well plate. The cell counts were run on auto-record for one minute at a flow rate of 20 µL
130 $min^{-1}$ for the algae and 30 µL $min^{-1}$ for the heterotrophs. The mixotroph populations were
131 distinguished through chlorophyll fluorescence and 488 nm, 90° side scatter (SSC), whereas the
132 heterotrophs were detected using SYBR Green fluorescence and 488 nm, 90° SSC [31]. The
133 viral-like particle (VLP) population was identified through SYBR Green fluorescence and
134 405 nm, 90° violet side scatter (VSSC), enhancing marine virus detection [32]. The samples were
135 run on auto-record for one minute at a flow rate of 10 µL $min^{-1}$. Using the shapefile output, both
136 algae and virus populations were delineated through a custom Python script using the following
137 packages: fcsparser (v0.2.8), pandas (v2.2.2), matplotlib (v3.7.1), while the FlowJo (v.10.10.0)

138 software was used to delineate the heterotrophic protist population. Apparent burst size was
139 determined by dividing the maximum concentration of VLPs post-infection by the inferred
140 number of lysed cells, calculated as the difference between the cell concentration on the day of
141 infection minus the minimum concentration post-infection.
142
143 *Electron microscopy*
144 Transmission electron microscope (TEM) micrographs of viral particles were prepared
145 from raw, unfixed lysate and negatively stained with 2% uranyl acetate for 45 sec on 200-mesh
146 copper grids with carbon-stabilized formvar support that had been rendered hydrophilic by glow
147 discharge [33]. Images were captured with a Hitachi HT7700 transmission electron microscope
148 (Tokyo, Japan) at the Biological Electron Microscope Facility (BEMF) at the University of
149 Hawai‘i at Mānoa. Particle dimensions were measured with ImageJ from Fiji (v2.16.0/1.54p)
150 [34], including: capsid diameter (n = 38), envelope (n = 34), tail (n = 17), and tail fibers (n = 6).
151 *DNA extraction, sequencing, assembly, and annotation*
152 To prepare sufficient viral biomass for the DNA sequencing, we extracted DNA from 610
153 mL of lysate ($6.4 \times 10^6$ VLP/mL), after allowing the infection (multiplicity of infection: 1) of
154 chrysophyte Clade H sp. UHM3501 to clear for four days. The unfiltered lysate was concentrated
155 and pooled through serial centrifugations (6X) in a 100 kDA Centricon Plus centrifugal filter
156 (MilliporeSigma, Burlington, MA, USA) at 3,500 × *g* at 20°C for 15 minutes. Approximately
157 200 ng of DNA was extracted from the concentrate using the ZymoBiomics Quick-DNA High
158 Molecular Weight MagBead Kit (Zymo Research, Irvine, CA, USA), flash frozen in liquid
159 nitrogen, stored at -20°C, then shipped on dry ice to the US Department of Energy Joint Genome

160 Institute (JGI). There, the DNA was sequenced using PacBio Multiplexed 6–10 kb Low Input
161 parameters.
162 A PacBio Low Input (DNA) library was constructed and sequenced using the PacBio
163 SEQUELIIe platform (Pacific Biosciences, Menlo Park, CA USA), which generated 3,267,906
164 reads totaling 31,033,080,767 bp. Reads were deduplicated using pbmarkdup (v1.0.3,
165 pbdmarkup) [35]. BBDuk (v38.99) [36] was used to remove reads that contained PacBio control
166 sequences and PCR adapters, and IceCreamFinder (v38.99) [37] was used to remove reads
167 without SMRTbell sequences. The final filtered fastq contained 3,257,268 reads totaling
168 30,921,351,264 bp with mean read length of 9,493 bp.
169 The filtered reads (31 Gb) were assembled using Flye (v2.9) [38] with the "--meta --
170 pacbio-hifi" argument. From this initial assembly, viral contigs were identified and grouped
171 through a low-complexity binning workflow, utilizing CONCOCT (v1.1.0) [39], MaxBin
172 (v2.2.7) [40], MetaBAT2 (v2.15-5) [41], and SemiBin2 (v1.5.1) [42]. For each binning strategy,
173 the viral bin was identified by assessing the presence of giant virus core marker genes described
174 in Aylward *et al*. (2021) [43] through GVClass (v1.0) [44] and the MetaBAT2-generated viral
175 bin was used in downstream analysis. Subsequently, circular consensus sequencing reads were
176 recruited against these identified viral contigs, and a 100 Mb subsample of these reads was
177 assembled using Flye without the "--meta" argument (v2.9) [38]. This iterative assembly resulted
178 in one high-quality viral contig, which was polished using two rounds of Racon (v1.6.0) [45] to
179 ensure accuracy.
180 Gene calling was performed with GVClass (v1.0) [44], and the 1.196 Mb viral genome
181 had an estimated 93.77% coding density. Gene annotation was performed with KofamScan

182 (v1.3.0) [46] and eggnog mapper (v2.1.12) [47] with annotations of genes of interests confirmed
183 with BLASTp [48]. tRNAs were annotated with tRNAscan-SE (v2.0.12) [49].
184
185 *Phylogenetic analysis*
186 Phylogenetic analysis of ChrysoHV with IQ-TREE (v2.2.6) [50] was based on a
187 concatenated alignment of seven core proteins recommended in Aylward *et al*. (2021) [43].
188 Query sequences were retrieved from the GVMAGs [13] and GenBank Assembly [51]
189 databases, and quality filtered to contain at least four out of seven core proteins determined
190 through hmmsearch (v3.3.2, E-value $\leq 1 \times 10^{-10}$) [52]. Additionally, the query sequences were
191 filtered to remove genomes with more than 10 copies of any core protein and more than four
192 core proteins with multiple copies. The remaining 990 query sequences were aligned with
193 MAFFT (v7.520) [53] and trimmed with trimAl (-gt 0.1; v1.4.1) [54], then concatenated with a
194 custom python script [55]. The trimmed, concatenated alignment was used for phylogenetic
195 analysis using the mixture model LG+I+F+G4 and ultrafast bootstrap supports (option: -bb
196 1000) (v2.2.6) [50]. The tree was annotated with iTOL (v6) [56].
197 Gene phylogeny of microbial rhodopsins with a heliorhodopsin outgroup were built with
198 IQ-TREE (v2.2.6) [50] using model LG+F+R10 and ultrafast bootstrap supports (option: -bb
199 1000) and annotated with iTOL (v6) [56]. Representatives of rhodopsin homologs from
200 cultivated virus isolates include: EhV-202 (heliorhodopsin) [16], PgV-12T and -16T (viral
201 rhodopsin I) [25], FloV-SA2 (viral rhodopsin II) [15]. Additionally, the gene phylogeny includes
202 synthetic construct: vPyACR_21921 (channelrhodopsin) [14], and rhodopsin homologs from
203 single amplified genomes (SAGs) and metagenome assembled genomes (MAGS): ChoanoV1
204 (viral group II) [13], cPacV-1605 (viral group II) [57], OLPV 1 and 2 (viral groups I & II) [58].

205 Query rhodopsin sequences were retrieved from NCBI and the supplementary materials of Olson
206 *et al*. (2018) [59]. The alignment was built with MAFFT (L-INS-i; v7.520) [53] and trimmed
207 with trimAl (-gt 0.1; v1.4.1) [54].
208 The phylogeny of near-complete 18S rRNA genes from select algal isolates were built
209 with IQ-TREE (v2.2.6) [50] using GTR+I+G model with ultrafast bootstrap support (option: -bb
210 1000) and annotated with iTOL (v6) [56]. The nucleotide sequences were retrieved from NCBI,
211 aligned with MAFFT (L-INS-i; v7.520) [53] and trimmed with trimAl (-gt 0.1; v1.4.1) [54].
212
213 *Gene distribution in the global ocean*
214 Comparisons to PolB amino acid sequences from the Tara Oceans dataset were
215 performed using BLASTp (v2.13.0, E-value $\leq 1 \times 10^{-120}$) [48], and sequences with 90% or
216 greater average amino acid identity (AAI) were plotted using a custom R script with the
217 following packages maps (v3.4.2), mapdata (v2.3.1), and ggplot2 (v.3.5.1). The relative
218 abundance of the PolB genes were normalized to total *Nucleocytoviricota* PolBs as in Endo *et al*.
219 (2020) [8]. Likewise, ChysoHV proteorhodopsin-like protein sequences from the Tara Oceans
220 dataset with a minimum of 80% AAI were identified through Ocean Gene Atlas [60, 61]. The
221 relative abundance of ChysoHV proteorhodopsin-like protein sequences were normalized to total
222 metagenomic reads in the given sample [60, 61].
223
224 *Protein structure predictions*
225 The conserved domain of the ChrysoHV rhodopsin structure was predicted using
226 ColabFold (v1.5.5) [62] and identified using Foldseek [63] against the AlphaFold Protein
227 Structure Databases– SwissProt [64] and CATH (v4.4) [65]. The conserved domain of the

228 ChrysoHV rhodopsin and HOT 75m4 Gamma-proteobacterium proteorhodopsin structures were
229 rendered in ChimeraX (v1.8.dev202403010247) [66].
230
231 **Results and Discussion**
232
233 *Host range, infection dynamics, and morphology*
234 We tested the host range of ChrysoHV against eight protist isolates from Station
235 ALOHA, resulting in no lysis for six of the eight isolates (Fig. 1A, C). Representatives of
236 chrysophyte Clade H, UHM3500 and UHM3501, lysed when challenged with ChrysoHV.
237 However, a greater proportion of UHM3501 cells lysed, resulting in the production of more
238 virions (Fig. 1AB and Fig. S1). Despite this, the apparent burst size, *i.e.*, estimated number of
239 virions produced per cell, was similar for each host strain, 27 ± 11 for UHM3500 and 26 ± 5 for
240 UHM3501 ([mean ± SD]) (Fig. 1AB, Fig. S1). The lower mortality rate of chrysophyte
241 UHM3500 may be due to the presence of resistant subpopulations or lineage-specific host
242 immune responses, potentially involving the activation virophages, which are widespread in
243 uncultivated members of Chrysophyceae Clade H [7, 67]. The virion particles produced from the
244 two chrysophytes shared the same notable morphological features, *i.e.*, an icosahedral capsid
245 (290 ± 40 nm diameter [mean ± SD]) associated with a loose membrane (720 ± 120 nm), and a
246 long tail (1,200 ± 240 nm length) ending in tail fibers (210 ± 700 nm length) (Fig. 2A-C).
247
248 *Genomic characterization*
249 The ChrysoHV genome was assembled as a putatively linear DNA sequence of 1.19 Mbp
250 encoding 1,138 predicted genes with a G+C content of 23.42% (Fig. 2B). The encoded genes

251 have diverse putative functions, including the ribosomal proteins eL40 and S27a, making it the
252 second isolated eukaryote-infecting virus known to encode ribosomal proteins [18], and the first
253 to encode S27a (Supplementary Table S1). ChrysoHV also encodes electron transport proteins
254 such as cytochrome b5, widespread in prokaryotes, eukaryotes, and giant viruses (Supplementary
255 Table S1) [68], along with succinate dehydrogenase flavoprotein and iron-sulfur subunits
256 (Supplementary Table S1). Additionally, the virus encodes seven ABC transporters including
257 one specific to phosphate transport (XVM28869.1) and two AmtB ammonium transporters. Of
258 these ammonium transporters, one (XVM28992.1) has 55% AAI to *Prochlorococcus* and
259 *Synechococcus* homologs (Supplementary Table S1). Likewise, ChrysoHV encodes three porins
260 with 44-66% AAI to *Prochlorococcus* iron and carbohydrate porins (XVM28452.1,
261 XVM28569.1, XVM28570.1), two urea ABC transporter substrate-binding proteins with 81-90%
262 AAI to a *Prochlorococcus* homolog (XVM29005.1, XVM29006.1), a semiSWEET sugar
263 transporter with 44% AAI to a *Prochlorococcus* homolog (XVM28124.1), a sulfotransferase
264 family protein with 55% AAI to a *Prochlorococcus* homolog (XVM28964.1), and a N-
265 acetylmuramoyl-L-alanine amidase for cell wall remodeling with 39% AAI to a
266 *Prochlorococcus* homolog (XVM29061.1) (Supplementary Table S1). The presence of a large
267 number of genes whose closest known relatives are *Prochlorococcus* homologs suggests
268 extensive horizontal gene transfer from cyanobacterial prey of the chrysophyte to a giant virus
269 that infects it.

270 The ChrysoHV genome also includes a region of light-sensitive proteins, encompassing a
271 proteorhodopsin-like protein surrounded by two photolyases and followed by a heliorhodopsin
272 (Fig. 2D, Fig. 3C, Supplementary Table S1). The positioning and shared orientation of the
273 photolyases and heliorhodopsin around the proteorhodopsin-like protein implies that they may be

274 part of the same operon and function like the heliorhodopsin in *Trichococcus flocculiformis* [69].
275 It was proposed that this protein could bind to nearby photolyases, enhancing DNA repair [69].
276
277 *Comparison to Nucleocytoviricota representatives*
278 ChrysoHV is the third cultivated isolate in the *Aliimimivirinae* subfamily of *Mimiviridae*
279 (Fig. 2E); other known isolates include CroV BV-PW1 [70], infecting a heterotrophic
280 nanoflagellate, and CV-XW01 [71], infecting a green alga. In this subfamily, only two other
281 representatives, cPacV-1605 and GVMAG-M-3300013131-43, encode rhodopsins, which belong
282 to the viral rhodopsin group I clade (Fig. 2A). However, 40 other *Aliimimivirinae* representatives
283 encode at least one photolyase homolog (Fig. 2E), either the (6-4) photolyase or DASH-type
284 cryptochrome. Only ChrysoHV and cPacV-1605 encode one microbial rhodopsin with multiple
285 heliorhodopsins, and photolyase homologs, but cPacV-1605's rhodopsin is placed within the
286 Viral Rhodopsin Group II lineage (Fig. 3A). Furthermore, instead of being nested within a light
287 sensitive gene neighborhood, the position of the viral rhodopsin in cPacV-1605 (locus tag:
288 CPAV1605_207, GenBank ID: VVU94485.1) is further away from the photolyases (locus tag:
289 CPAV1605_133, CPAV1605_136; GenBank ID: VVU94411.1, VVU94414.1) and
290 heliorhodopsins (locus tag: CPAV1605_533, CPAV1605_1309; GenBank ID: VVU95557.1,
291 VVU94808.1) [58]. ChrysoHV is the only cultivated giant virus to encode the proteorhodopsin
292 lineage of rhodopsin along with two heliorhodopsins (Fig 2E, Fig. 3A). The presence of multiple,
293 diverse rhodopsins in ChrysoHV may represent a unique adaptation to high light in the surface
294 waters of the subtropical ocean where it was isolated.
295
296 *Sequence characterization of a distinct lineage of proteorhodopsin*

297 Phylogenetic analysis of ChrysoHV's rhodopsin placed it within the proteorhodopsin
298 clade (Fig. 3A). These results suggest that ChrysoHV may have acquired the proteorhodopsin-
299 like rhodopsin from *Pelagibacter*, which is a member of the highly abundant and widespread
300 SAR11 clade of bacteria. These bacteria are probably, like *Prochlorococcus*, prey for
301 chrysophytes. The sequence alignment of transmembrane domain helix-III (TM3) indicates that
302 the ChrysoHV rhodopsin is spectrally tuned to blue-light (glutamine, Q; Fig. 3A), which is the
303 dominant spectral tuning amino acid in proteorhodopsins from Station ALOHA, the isolation site
304 of ChrysoHV [59]. The two most closely related rhodopsin sequences were recovered from
305 metagenomes from the Black Sea at 50 m depth (SAMN27960419) and the North Pacific at 5 m
306 depth in the 0.22-3μm size fraction (SAMEA2623059). These rhodopsins may also absorb blue
307 light (Fig. 3A) and share a poly-N (asparagine) tract at the start of the rhodopsin coding
308 sequence.
309 While ChrysoHV lacks the conserved DTE proton pumping motif found in most
310 proteorhodopsins, it possesses an ETL motif (Fig. 3A), which has been detected in
311 proteorhodopsins from deep water at Station ALOHA (Fig. 3A) [59]. A comparison of the
312 predicted structure of the conserved domain of the ChrysoHV rhodopsin against existing
313 functional characterizations revealed that it was most similar to the uncultured gamma-
314 proteobacterium clone HOT 75m4 rhodopsin (Fig. 3B). This rhodopsin was confirmed to pump
315 protons ($H^+$) [72]. Although the AAI was not the highest (35.9%), the HOT 75m4
316 proteorhodopsin was also retrieved from Station ALOHA [72].
317 ChrysoHV does not encode β-carotene, a precursor to retinal, the light-absorbing
318 chromophore in proteorhodopsin, and the salt bridge between K102 and E228 most likely
319 inhibits retinal binding. This loss of retinal binding could explain the long branch of the

320 ChrysoHV proteorhodopsin-like rhodopsin (Fig. 3A). Understanding how the rhodopsin may
321 function without retinal binding could be important to interpreting the function of viral
322 rhodopsins detected below the photic zone [73, 74]. Further investigation is needed to confirm
323 whether the ChrysoHV proteorhodopsin can confer other physiological benefits in the absence of
324 light, similar to the *Psychroflexus torquis* proteorhodopsin ptqPR, which acts as a phospholipid
325 scramblase to preserve membrane potential, fluidity, and integrity [75].
326
327 *Global ocean distribution of closely related ChrysoHV phylotypes and proteorhodopsins*
328 Analysis on the distribution of ChrysoHV phylotypes (≥ 90% PolB AAI) from the TARA
329 Oceans dataset [8] revealed their prevalence in subtropical surface ocean sites in the Pacific and
330 Indian Ocean as well as the Mediterranean Sea (Fig. 4A). Notably, at TARA station 132, north of
331 the Hawaiian archipelago, ChrysoHV phylotypes were detected in both the mixed layer and
332 mesopelagic zone, suggesting connectivity between surface and deep ocean virus populations.
333 The abundance of the ChrysoHV proteorhodopsin-like protein (≥ 80% PolB AAI) was also
334 highest in the subtropical surface ocean (Fig. 4B), but detected at fewer overall sites than the
335 ChrysoHV PolB. Additionally, the ChrysoHV proteorhodopsin relative abundance was highest at
336 TARA station 132, indicating that these proteorhodopsin sequences may be derived from a
337 ChrysoHV phylotype.
338
339 *Conclusion*
340 In the marine environment, diverse microorganisms from proteobacteria to microalgae
341 utilize rhodopsins to supplement their energetic needs [59, 76]. SAR11 bacteria, of which
342 *Pelagibacter* is an abundant subclade, account for 25% of all plankton [77], and are a likely prey

source for mixotrophic algae such as Clade H chrysophytes UHM3500 and UHM3501. The discovery of a proteorhodopsin-like protein encoded by a giant virus supports the conceptual model that lateral gene transfer is driven by the co-mingling of infectious virions and ingested prey within the same host cell [78]. The presence of nine genes whose closest known relatives are *Prochlorococcus* homologs further suggests that prey-virus HGT has played an important role in the evolution of ChrysoHV. By leveraging naturally co-occurring and experimentally accessible virus-host pairs, this study lays the groundwork for future investigations into how the virocell environment of predatory protists influences gene acquisition in giant viruses. As the only cultivated isolate known to encode one microbial rhodopsin and two heliorhodopsins, this system provides the opportunity for novel insights into the role of rhodopsins in giant virus infection strategies.

**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 the 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 sharing 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.

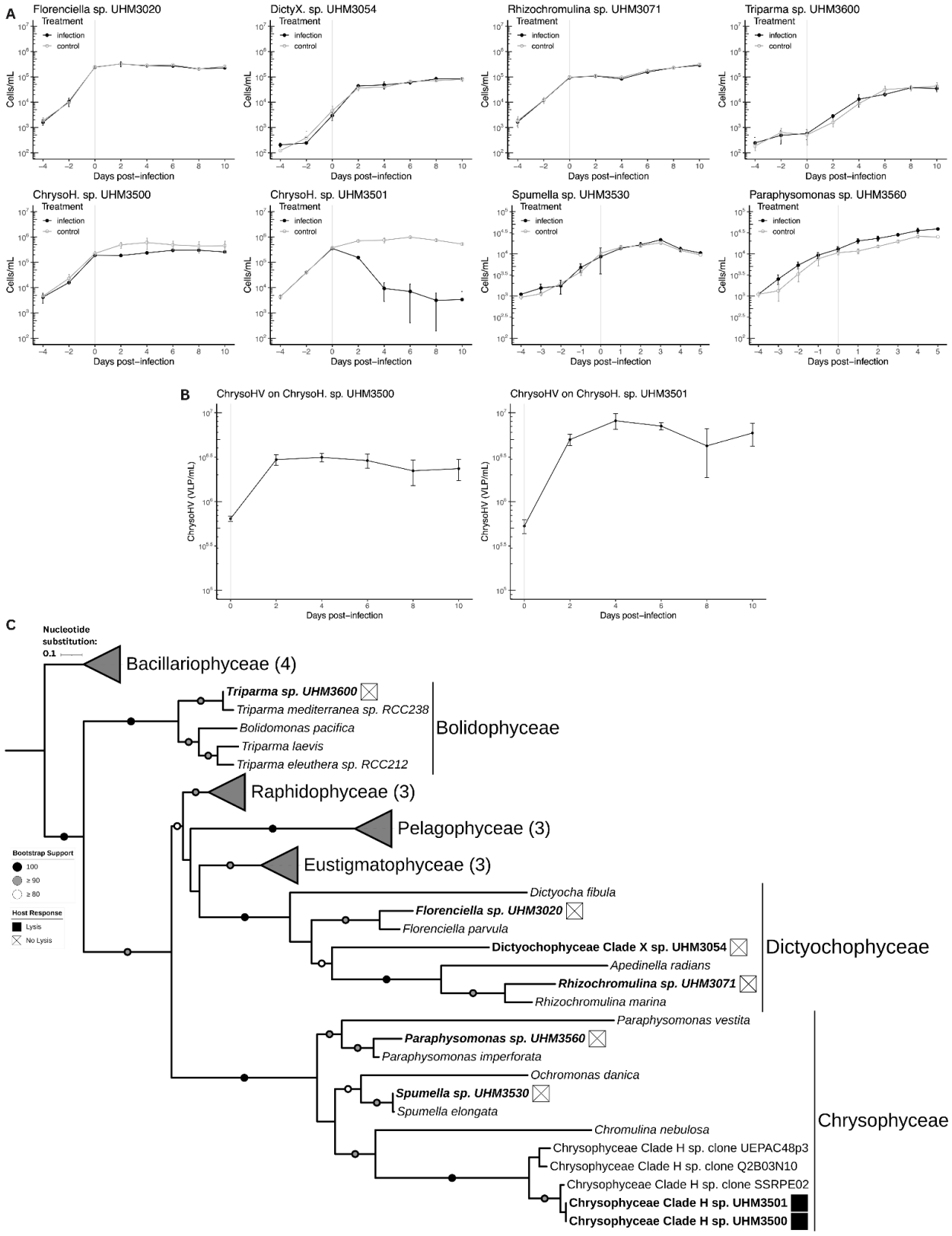
**Conflicts of Interest**

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


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)
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)
Bootstrap Support
100
≥ 90
≥ 80
Host Response
Lysis
No Lysis
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

405

406 **Figure 1**. (A) Time series of protist abundance (cells/mL) in response to the addition of raw
407 ChrysoHV lysate. Days pre- and post-infection are displayed on the x-axis. The algae were
408 isolated from Station ALOHA in the North Pacific Subtropical Gyre, and include: *Florenciella*
409 so. UHM3020, Dictyochophyceae Clade X sp. UHM3054, *Rhizochromulina* sp. UHM3071,
410 Chrysophyceae Clade H sp. UHM3500, Chrysophyceae Clade H sp. UHM3501, *Triparma* sp.
411 UHM3600, *Spumella* sp. UHM3530, and *Paraphysomonas* sp. UHM3560. The control treatment
412 is shown in the gray line with white-fill points, and the infection treatment shown in the black
413 line with black-fill points. Error bars represent the standard deviation of the triplicate counts. (B)
414 Time series of ChrysoHV abundance (VLP/mL) propagated on permissive hosts. Error bars
415 represent the standard deviation of the triplicate counts. (C) ML 18S rRNA phylogeny
416 (GTR+I+R model) of the Ochrophyta phylum with isolates challenged in bold. The white or
417 black boxes by the bolded isolate names indicate whether the protist culture did not lyse (white
418 with "X") or lysed and produced virions (black) in response to the ChrysoHV challenge. The
419 node symbols indicate the bootstrap values equal to or greater than 80% (white), 90% (gray) or
420 100% (black). There is no node symbol for branches with less than 80% support.

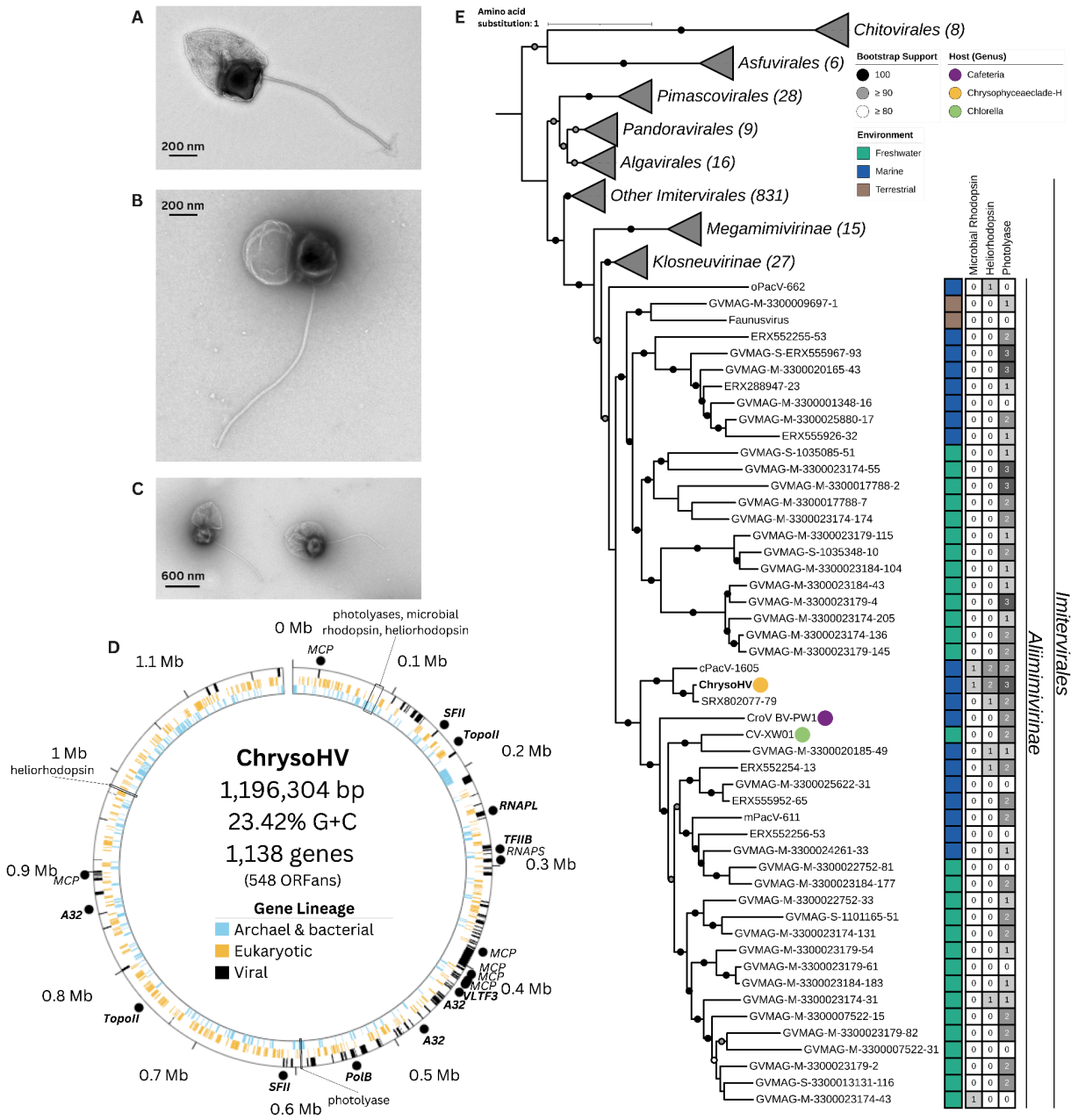



**Figure 2**. (A) Transmission electron microscopy (TEM) negative stain micrograph of ChrysoHV propagated on UHM3501. (B, C) TEM negative stain micrograph of ChrysoHV propagated on UHM3500. (D) Genome plot of ChrysoHV with track placement and color showing gene lineage annotated according to domain,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

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

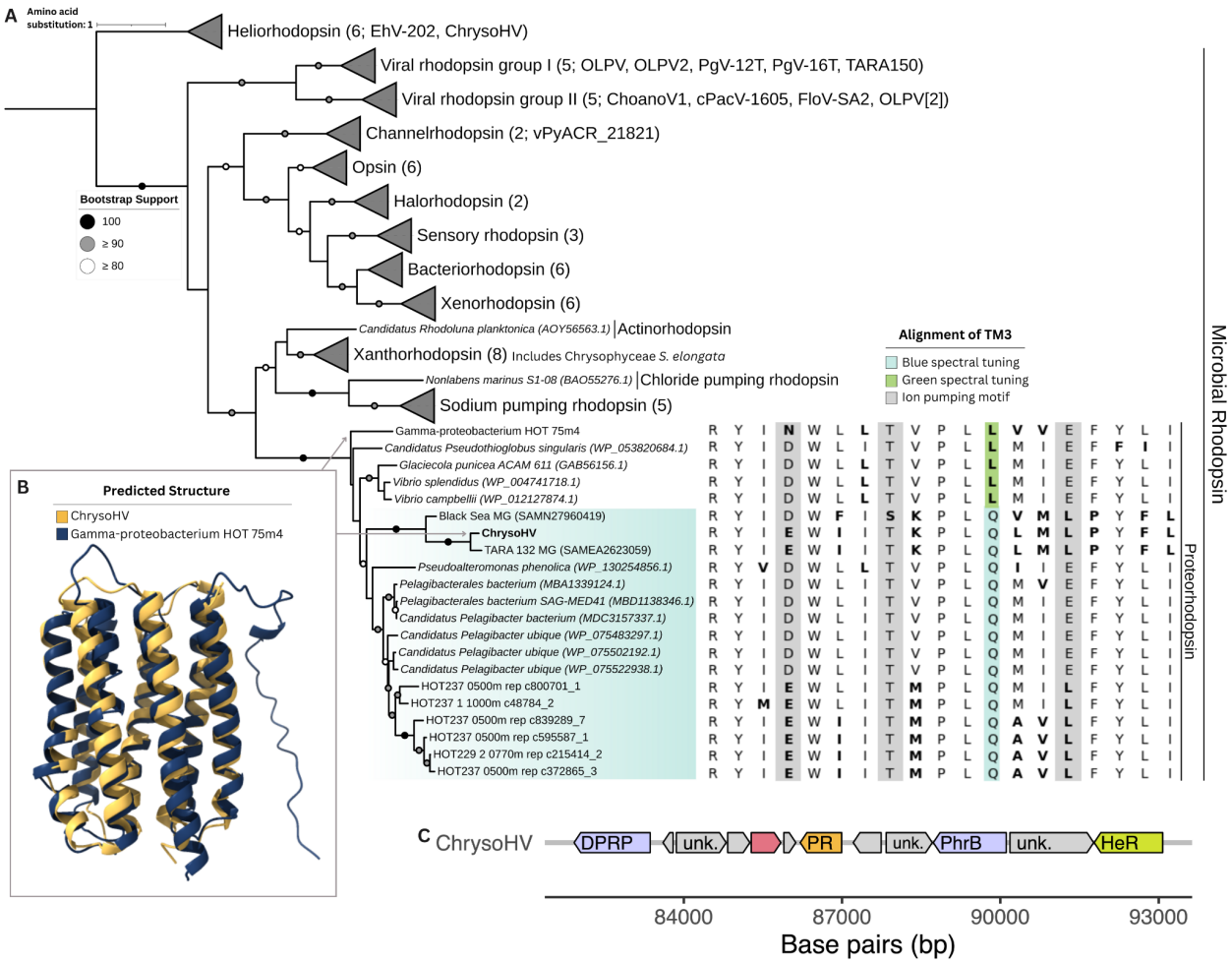



**Figure 3**. (A) Microbial rhodopsin ML phylogeny with heliorhodopsin as the outgroup, built with model LG+F+R10. Select representatives of viral rhodopsin homologs and putative chrysophyte rhodopsins are referred to in the parenthesis of their corresponding collapsed clade. The proteorhodopsin (PR) 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,

ChyrsoHV_0104; GenBank ID: XVM28016.1, XVM28025.1) and one heliorhodopsin (locus tag: ChrysoHV_0106, GenBank ID: XVM28027.1).

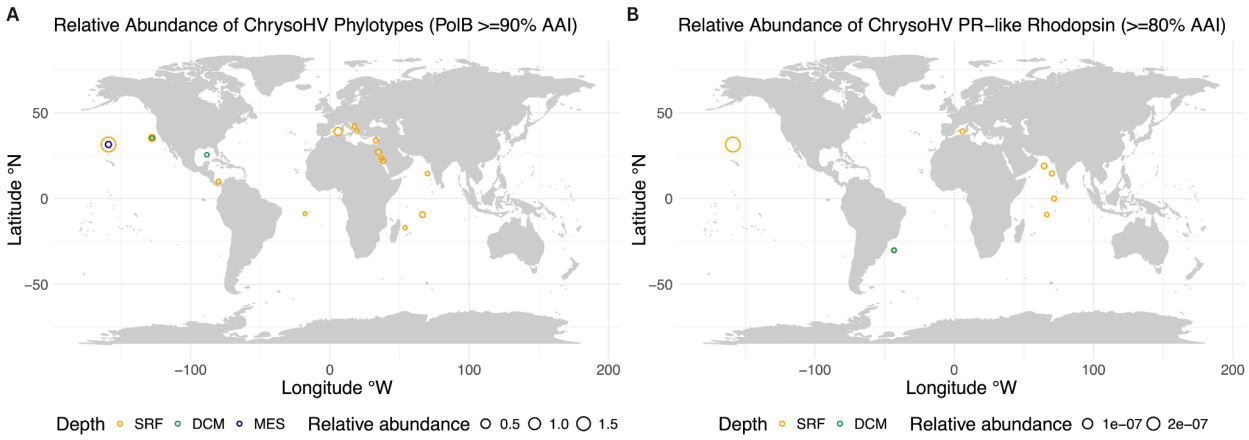



**Figure 4**. (A) Global distribution of phylotypes with ≥ 90% PolB AAI to ChrysoHV from the Tara Oceans dataset. Size of the circle indicates the relative abundance of the summed phylotypes at each given sample site to total *Nucleocytoviricota* PolB sequences per site. (B) Global distribution of homologous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

## References

1. Del Campo J, Massana R. Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist* 2011;**162**:435–448. https://doi.org/10.1016/j.protis.2010.10.003

468 2. Seeleuthner Y et al. Single-cell genomics of multiple uncultured stramenopiles reveals
469 underestimated functional diversity across oceans. *Nat Commun* 2018;**9**:310.
470 https://doi.org/10.1038/s41467-017-02235-3
471
472 3. Olefeld JL et al. Genome size of chrysophytes varies with cell size and nutritional mode. *Org*
473 *Divers Evol* 2018;**18**:163–173. https://doi.org/10.1007/s13127-018-0365-7
474
475 4. Frias-Lopez J et al. Use of stable isotope-labelled cells to identify active grazers of
476 picocyanobacteria in ocean surface waters. *Environ Microbiol* 2009;**11**:512–525.
477 https://doi.org/10.1111/j.1462-2920.2008.01793.x
478
479 5. 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
483 6. Li Q et al. Broad phylogenetic and functional diversity among mixotrophic consumers of
484 *Prochlorococcus*. *ISME J* 2022;**16**:1557–1569. https://doi.org/10.1038/s41396-022-01204-z
485
486 7. Castillo YM et al. Assessing the viral content of uncultured picoeukaryotes in the global-ocean
487 by single cell genomics. *Mol Ecol* 2019;**28**:4272–4289. https://doi.org/10.1111/mec.15210
488

489 8. Endo H et al. Biogeography of marine giant viruses reveals their interplay with eukaryotes and
490 ecological functions. *Nat Ecol Evol* 2020;**4**:1639–1649. https://doi.org/10.1038/s41559-020-
491 01288-w
492
493 9. Fromm A et al. Single-cell RNA-seq of the rare virosphere reveals the native hosts of giant
494 viruses in the marine environment. *Nat Microbiol* 2024;**9**:1619–1629.
495 https://doi.org/10.1038/s41564-024-01669-y
496
497 10. Michoud G et al. Mapping the metagenomic diversity of the multi-kingdom glacier-fed
498 stream microbiome. *Nat Microbiol* 2025;**10**:217–230. https://doi.org/10.1038/s41564-024-
499 01874-9
500
501 11. Aylward FO et al. Taxonomic update for giant viruses in the order Imitervirales (phylum
502 Nucleocytoviricota). *Arch Virol* 2023;**168**:283. https://doi.org/10.1007/s00705-023-05906-3
503
504 12. Moniruzzaman M et al. Virologs, viral mimicry, and virocell metabolism: the expanding
505 scale of cellular functions encoded in the complex genomes of giant viruses. FEMS Microbiol
506 Rev 2023;47:fuad053. https://doi.org/10.1093/femsre/fuad053
507
508 13. Needham DM et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to
509 unicellular marine predators. Proc Natl Acad Sci 2019;116:20574–20583.
510 https://doi.org/10.1073/pnas.1907517116
511

512 14. Rozenberg A et al. Lateral gene transfer of anion-conducting channelrhodopsins between
513 green algae and giant viruses. Curr Biol 2020;30:4910-4920.e5.
514 https://doi.org/10.1016/j.cub.2020.09.056
515
516 15. Thomy J et al. Eukaryotic viruses encode the ribosomal protein eL40. Npj Viruses 2024;2:51.
517 https://doi.org/10.1038/s44298-024-00060-2
518
519 16. Hososhima S et al. Proton-transporting heliorhodopsins from marine giant viruses. eLife
520 2022;11:e78416. https://doi.org/10.7554/eLife.78416
521
522 17. Moniruzzaman M et al. Dynamic genome evolution and complex virocell metabolism of
523 globally-distributed giant viruses. Nat Commun 2020;11:1710. https://doi.org/10.1038/s41467-
524 020-15507-2
525
526 18. Schulz F et al. Giant virus diversity and host interactions through global metagenomics.
527 Nature 2020;578:432–436. https://doi.org/10.1038/s41586-020-1957-x
528
529 19. Yutin N, Koonin EV. Proteorhodopsin genes in giant viruses. Biol Direct 2012;7:34.
530 https://doi.org/10.1186/1745-6150-7-34
531
532 20. Pinhassi J et al. Marine Bacterial and archaeal ion-pumping rhodopsins: genetic diversity,
533 physiology, and ecology. Microbiol Mol Biol Rev 2016;80:929–954.
534 https://doi.org/10.1128/MMBR.00003-16

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. Bosmon T, Abergel C, Claverie J-M. 20 years of research on giant viruses. Npj Viruses 2025;3:9. https://doi.org/10.1038/s44298-025-00093-1

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

26. 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

27. 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

28. 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.

29. 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

30. 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

31. 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

32. 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

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

34. 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

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

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

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

38. 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

39. Alneberg J et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;**11**:1144–1146. https://doi.org/10.1038/nmeth.3103

40. Wu Y-W et al. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2014;**2**:26. https://doi.org/10.1186/2049-2618-2-26

41. Kang DD et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;**7**:e7359. https://doi.org/10.7717/peerj.7359

42. Pan S, Zhao X-M, Coelho LP. SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. *Bioinformatics* 2023;**39**:i21–i29. https://doi.org/10.1093/bioinformatics/btad209

43. Aylward FO et al. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLOS Biol* 2021;**19**:e3001430. https://doi.org/10.1371/journal.pbio.3001430

44. Pitot TM, Brůna T, Schulz F. Conservative taxonomy and quality assessment of giant virus genomes with GVClass. *Npj Viruses* 2024;**2**:60. https://doi.org/10.1038/s44298-024-00069-7

45. Vaser R et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;**27**:737–746. https://doi.org/10.1101/gr.214270.116

46. Aramaki T et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 2020;**36**:2251–2252. https://doi.org/10.1093/bioinformatics/btz859

47. Cantalapiedra CP et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;**38**:5825–5829. https://doi.org/10.1093/molbev/msab293

625 48. Camacho C et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;**10**:421.
626 https://doi.org/10.1186/1471-2105-10-421
627
628 49. Chan PP et al. tRNAscan-SE 2.0: improved detection and functional classification of transfer
629 RNA genes. *Nucleic Acids Res* 2021;**49**:9077–9096. https://doi.org/10.1093/nar/gkab688
630
631 50. Minh BQ et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in
632 the genomic era. *Mol Biol Evol* 2020;**37**:1530–1534. https://doi.org/10.1093/molbev/msaa015
633
634 51. Benson DA et al. GenBank. *Nucleic Acids Res* 2012;**41**:D36–D42.
635 https://doi.org/10.1093/nar/gks1195
636
637 52. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;**7**:e1002195.
638 https://doi.org/10.1371/journal.pcbi.1002195
639
640 53. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
641 improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–780.
642 https://doi.org/10.1093/molbev/mst010
643
644 54. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment
645 trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–1973.
646 https://doi.org/10.1093/bioinformatics/btp348
647

648 55. NeLli-team. nsgtree. https://github.com/NeLLi-team/nsgtree (03 March 2024, date last
649 accessed)
650
651 56. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree
652 display and annotation tool. *Nucleic Acids Res* 2024;**52**:W78–W82.
653 https://doi.org/10.1093/nar/gkae268
654
655 57. Needham DM et al. Targeted metagenomic recovery of four divergent viruses reveals shared
656 and distinctive characteristics of giant viruses of marine eukaryotes. *Philos Trans R Soc B Biol*
657 *Sci* 2019;**374**:20190086. https://doi.org/10.1098/rstb.2019.0086
658
659 58. Yau S et al. Virophage control of antarctic algal host–virus dynamics. *Proc Natl Acad Sci*
660 2011;**108**:6163–6168. https://doi.org/10.1073/pnas.1018221108
661
662 59. Olson DK et al. Proteorhodopsin variability and distribution in the North Pacific Subtropical
663 Gyre. *ISME J* 2018;**12**:1047–1060. https://doi.org/10.1038/s41396-018-0074-4
664
665 60. Villar E et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online.
666 *Nucleic Acids Res* 2018;**46**:W289–W295. https://doi.org/10.1093/nar/gky376
667
668 61. Vernette C et al. The Ocean Gene Atlas v2.0: online exploration of the biogeography and
669 phylogeny of plankton genes. *Nucleic Acids Res* 2022;**50**:W516–W526.
670 https://doi.org/10.1093/nar/gkac420

671

672 62. Mirdita M et al. ColabFold: making protein folding accessible to all. *Nat Methods*

673 2022;**19**:679–682. https://doi.org/10.1038/s41592-022-01488-1

674

675 63. Van Kempen M et al. Fast and accurate protein structure search with Foldseek. *Nat*

676 *Biotechnol* 2024;**42**:243–246. https://doi.org/10.1038/s41587-023-01773-0

677

678 64. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*

679 2017;**45**:D158–D169. https://doi.org/10.1093/nar/gkw1099

680

681 65. Waman VP et al. CATH v4.4: major expansion of CATH by experimental and predicted

682 structural data. *Nucleic Acids Res* 2025;**53**:D348–D355. https://doi.org/10.1093/nar/gkae1087

683

684 66. Meng EC et al. UCSF CHIMERAX: Tools for structure building and analysis. *Protein Sci*

685 2023;**32**. https://doi.org/10.1002/pro.4792

686

687 67. Bellas C et al. Large-scale invasion of unicellular eukaryotic genomes by integrating DNA

688 viruses. *Proc Natl Acad Sci* 2023;**120**:e2300465120. https://doi.org/10.1073/pnas.2300465120

689

690 68. Lamb DC et al. Cytochrome b5 occurrence in giant and other viruses belonging to the

691 phylum Nucleocytoviricota. *Npj Viruses* 2025;**3**:8. https://doi.org/10.1038/s44298-025-00091-3

692

693 69. Shim J et al. Heliorhodopsin helps photolyase to enhance the DNA repair capacity. *Microbiol*
694 *Spectr* 2022;**10**:e02215-22. https://doi.org/10.1128/spectrum.02215-22
695
696 70. Fischer MG et al. Giant virus with a remarkable complement of genes infects marine
697 zooplankton. *Proc Natl Acad Sci* 2010;**107**:19508–19513.
698 https://doi.org/10.1073/pnas.1007615107
699
700 71. Sheng Y et al. Isolation and identification of a large green alga virus (*Chlorella* Virus XW01)
701 of *Mimiviridae* and Its Virophage (*Chlorella* Virus Virophage SW01) by using unicellular green
702 algal cultures. *J Virol* 2022;**96**:e02114-21. https://doi.org/10.1128/jvi.02114-21
703
704 72. Ran T et al. Cross-protomer interaction with the photoactive site in oligomeric
705 proteorhodopsin complexes. *Acta Crystallogr D Biol Crystallogr* 2013;**69**:1965–1980.
706 https://doi.org/10.1107/S0907444913017575
707
708 73. Zhang L et al. Spatiotemporal dynamics of giant viruses within a deep freshwater lake reveal
709 a distinct dark-water community. *ISME J* 2024;**18**:wrae182.
710 https://doi.org/10.1093/ismejo/wrae182
711
712 74. Pitot TM et al. Viral niche-partitioning: comparative genomics of giant viruses across
713 environmental gradients in a high Arctic freshwater-saltwater lake. *ISME Commun*
714 2025;**5**:ycae155. https://doi.org/10.1093/ismeco/ycae155
715

716 75. Fang J et al. Scramblase activity of proteorhodopsin confers physiological advantages to
717 Escherichia coli in the absence of light. *iScience* 2023;**26**:108551.
718 https://doi.org/10.1016/j.isci.2023.108551
719
720 76. Strauss J et al. Plastid-localized xanthorhodopsin increases diatom biomass and ecosystem
721 productivity in iron-limited surface oceans. *Nat Microbiol* 2023;**8**:2050–2066.
722 https://doi.org/10.1038/s41564-023-01498-5
723
724 77. Giovannoni SJ. SAR11 bacteria: the most abundant plankton in the oceans. *Annu Rev Mar*
725 *Sci* 2017;**9**:231–255. https://doi.org/10.1146/annurev-marine-010814-015934
726
727 78. Filée J, Siguier P, Chandler M. I am what I eat and I eat what I am: acquisition of bacterial
728 genes by giant viruses. *Trends Genet* 2007;**23**:10–15. https://doi.org/10.1016/j.tig.2006.11.002