1   <u>Title</u>: **Making sense of virus size and the tradeoffs shaping viral fitness**

2   <u>Running title</u>: Tradeoffs shaping virus size

3

4   <u>Authors</u>: Kyle F. Edwards[*1,2], Grieg F. Steward[1,3], Christopher R. Schvarcz[1,4]

5   [1]Department of Oceanography, University of Hawaiʻi at Mānoa, Honolulu, HI

6   [2]kfe@hawaii.edu, [3]grieg@hawaii.edu, [4]schvarcz@hawaii.edu

7   *corresponding author address: 1000 Pope Rd, Honolulu, HI 96822, phone: 808-956-6198, fax:

8   808-956-9225

9

10 <u>Keywords</u>: giant virus, allometric scaling, host range, burst size, diffusivity, decay rate,

11 competition, phytoplankton, bacteria, phage

12 <u>Article type</u>: Reviews and Syntheses

13 <u>Abstract word count</u>: 200

14 <u>Main text word count</u>: 6708

15 <u>Number of references</u>: 83

16 <u>Numbers of figures / tables / text boxes</u>: 4 / 0 / 0

17

18 <u>Authorship statement</u>: All authors contributed to the conception and outline of the work. K.F.E wrote

19 the initial manuscript, analyzed mathematical models, and contributed to data compilation and

20 analysis. G.F.S and C.R.S. both contributed to manuscript writing and to data compilation and

21 analysis.

22

23 <u>Data accessibility statement</u>: Should the manuscript be accepted, the data supporting the results will be

24 archived in an appropriate public repository and the data DOI will be included at the end of the article.

Abstract

Viruses span an impressive size range, with genome length varying a thousandfold and virion volume nearly a millionfold. For cellular organisms the scaling of traits with size is a pervasive influence on ecological processes, but whether size plays a central role in viral ecology is unknown. Here we focus on viruses of aquatic unicellular organisms, which exhibit the greatest known range of virus size. We develop and synthesize theory, and analyze data where available, to consider how size affects the primary components of viral fitness. We argue that larger viruses have fewer offspring per infection and slower contact rates with host cells, but a larger genome tends to increase infection efficiency, broaden host range, and potentially increase attachment success and decrease decay rate. These countervailing selective pressures may explain why a breadth of sizes exist and even coexist when infecting the same host populations. Oligotrophic ecosystems may be enriched in "giant" viruses, because environments with resource-limited phagotrophs at low concentrations may select for broader host range, better control of host metabolism, lower decay rate, and a physical size that mimics bacterial prey. Finally, we describe where further research is needed to understand the ecology and evolution of viral size diversity.

Introduction

Viruses are ubiquitous and abundant molecular symbionts that influence individual health, population and community dynamics, evolution, and biogeochemistry, across the tree of life. By nature, viruses are smaller than the cells they infect, but the range of virus sizes is nonetheless substantial, with lengths of viral particles (virions) varying from 17 nm to ~1.5 μm, and genome size varying from ~1 kb to 2.5 Mb (Campillo-Balderas et al. 2015). The largest 'giant' viruses have primarily been isolated from unicellular protists (Campillo-Balderas et al. 2015, Wilhelm et al. 2017), although there is metagenomic evidence for 'megaphages' of prokaryotes with genomes up to 716 kb (Devoto et al. 2019, Al-Shayeb et al. 2020), and a chaetognath appears to be infected by viruses 1.25 μm in length (Shinn and Bullard 2018). In contrast, the known viruses of plants and fungi have genomes < 30 kb (Campillo-Balderas et al. 2015). At a finer phylogenetic scale, particular species or strains of prokaryotes and eukaryotes can be infected by viruses of very different size. For example, the marine bacterium *Cellulophaga baltica* is infected by phages ranging from 6.5 to 242 kb (Holmfeldt et al. 2007) and the marine dinoflagellate *Heterocapsa circularisquama* is infected by a 4.4 kb ssRNA virus and a 365 kb dsDNA virus (Tomaru et al. 2009).

For cellular life, body size is a 'master trait' that influences numerous organismal properties, such as metabolic rate, nutrient uptake affinity, predator-prey linkages, and population growth rate (Finkel 2001, Brown et al. 2004, Fuchs and Franks 2010, Edwards et al. 2012). The substantial variation in virus size raises the question of whether size plays a similar central role in virus ecology and evolution (Record et al. 2016). For example, are there general relationships between virus size and key viral traits? Do size-related tradeoffs lead to selection for different sizes of viruses infecting different kinds of hosts, or under different environmental conditions?

68 How do viruses of different size coexist when infecting the same host population? There are

69 straightforward physical reasons why being smaller should be advantageous for a virus: smaller

70 particles should encounter hosts faster due to greater diffusivity, at least in aquatic systems or

71 aqueous microenvironments, and limited host resources during infection can be partitioned

72 among a greater number of 'offspring'. The existence of a spectrum of virus sizes implies that

73 the costs of increased size can be offset by countervailing benefits. Virion size and genome size

74 are tightly correlated (Cui et al. 2014), and benefits of increased size likely derive from the

75 functions encoded by viral genes, including better control of attachment to the host, replication,

76 transcription, translation, and host metabolism; strategies countering host antiviral defenses; and

77 repair of damaged viral nucleic acids (Sharon et al. 2011, Samson et al. 2013, Fischer et al.

78 2014, Koonin and Yutin 2019, Mendoza et al. 2019). These functions could increase the

79 probability of successful infection, the number of virions produced per infection, the range of

80 hosts than can be successfully infected, and/or the persistence of virions in the extracellular

81 environment.

82   In this study, we focus on the question of how virus size affects key viral traits, and how

83 these traits affect viral fitness. We develop and summarize relevant theory, synthesize and

84 analyze available data, and outline major knowledge gaps and future research directions. We

85 focus primarily on viruses that infect aquatic microbes (unicellular prokaryotes and eukaryotes),

86 as these viruses are known to vary greatly in size and have been relatively well-studied in culture.

87 However, most of the concepts we develop should be useful for understanding viruses in general.

88

Theory for how viral fitness is determined by key viral traits

90   In order to connect a metric of fitness to virus traits (burst size, latent period, contact rate,

91   decay rate, etc.) we imagine a lytic virus population that may compete for hosts with one or

92   more additional virus populations. Using a simple model of virus-microbe population dynamics,

93   at steady state the density of the limiting resource, uninfected cells, is $S^* = \frac{d}{k(be^{-mL}-1)}$ (eqn. 1;

94   Appendix S1). Here S is the density of uninfected cells, m is the host mortality rate from causes

95   other than viral infection, b is the burst size (new virions produced per infection), L is the latent

96   period, d is the viral decay rate, and k is the effective adsorption rate—the rate at which

97   successful new infections are formed. The parameter k can be decomposed into subprocesses,

98   and here we define $k = caw$, where c is the contact rate at which host and virus encounter each

99   other, a is the attachment efficiency (probability that encounter leads to successful attachment),

100   and w is the probability that attachment leads to a successful infection (eventual lysis of the host,

101   releasing new virions). The distinction between contact rate and attachment efficiency will be

102   important when considering the role of virus size.

103   The quantity S^* is the uninfected host density at which growth of the viral population

104   balances the decay rate. Therefore, this is also the host density threshold required for persistence

105   of a viral population, and if the host density is initially greater it will be cropped down to this

106   level at steady state. S^* can be used as a measure of viral fitness because if there are two viral

107   strains, and $S_1^* < S_2^*$, then strain 1 will drive the host to a lower density and competitively

108   exclude the other strain at equilibrium (Tilman 1982). It is more intuitive to consider the inverse

109   of S^* as a metric of fitness, because a smaller S* equates to greater competitive ability; the

110   inverse of S^* is $\frac{k(be^{-mL}-1)}{d}$. This quantity is the average net number of new virions produced per

111 infection ($be^{-mL} - 1$), scaled by the rate of successfully encountering a new host (k) relative to

112 the rate of 'dying' while waiting to encounter a new host (d). This analysis assumes that host-

113 virus dynamics reach a steady-state attractor, which may not be true, but the S^* quantity is still

114 a useful index of competitive ability. Furthermore, in a simpler model with no latent period it

115 can be shown that S^* predicts the winner in competition even if populations fluctuate (Appendix

116 S2). This analysis relates traits to fitness for lytic viruses, which we focus on in this study due to

117 a greater accumulation of relevant trait data, but potential effects of size on viruses with

118 temperate strategies will be discussed as well.

119   In Appendix S3 we extend this analysis to ask under what conditions a broader host range

120 is selected for. The main result is that fitness (measured in terms of competitive outcomes) is

121 proportional to host range (measured as the number of host strains that can be infected). This

122 means that a generalist virus and a specialist virus will have similar fitness when the cost of

123 generalism is directly proportional to host range breadth. For example, if a generalist virus can

124 infect twice as many strains as specialist viruses, but has an adsorption rate that is 50% lower

125 on each strain, it will be competitively equivalent to the specialists. If the cost of generalism is

126 lower then generalism will be favored, and vice versa. These results are derived from a simple

127 model but they allow us to quantify, as a first approximation, how tradeoffs involving host range

128 and other viral traits may affect selection on virus size.

129

130   **How does virus size affect contact rate and attachment to host cells?**

131   We now consider key viral traits individually, to outline expectations for how virus size

132 may affect each trait and analyze relevant data where it is available. Adsorption rate is important

133   for viral fitness (eqn. 1) because it determines the rate at which new infections can be established,

134   as well as the time a viral particle spends in the extracellular environment where it may be

135   exposed to UV radiation, adsorption to non-host material, ingestion, etc. (Suttle and Chen 1992,

136   Noble and Fuhrman 1997). As described above, it is useful to separate the adsorption rate into

137   the contact rate c (per capita rate at which hosts and viruses encounter each other) and the

138   attachment efficiency a (probability of successful attachment to the host). In theory the contact

139   rate will depend on physical processes of Brownian motion, advection, and turbulence, while

140   attachment efficiency will be a function of host receptor availability, affinity of the receptor for

141   the virus, and mechanisms such as reversible binding by viral fibers that keep the virus from

142   diffusing away from the host (Schwartz 1976, Wickham et al. 1990, Storms and Sauvageau

143   2015).

144   Physical theory for contact rate typically starts by asking what the rate would be if viruses

145   relied solely on molecular diffusion to encounter their hosts, and if all viruses that contact the

146   host are adsorbed. Under pure diffusion the contact rate is predicted to be:

147    $$c = 4\pi r_H D_V \tag{2}$$

148   where r_H is host radius and D_V is the diffusivity of the virus (Murray and Jackson 1992). The

149   diffusivity of a spherical virus is predicted to be:

150    $$D_V = \frac{k_B T}{6\pi \eta r_V} \tag{3}$$

151   where k_B is the Boltzmann constant, T is temperature, η is dynamic viscosity of water, and r_V is

152   the virion radius. Therefore, the diffusion-limited contact rate is predicted to be inversely

153   proportional to virus diameter, which is a substantial fitness cost of increasing size.

154   The diffusivity of viruses due to Brownian motion is low enough that contact rates could

155   be increased considerably by processes that create fluid motion relative to the host cell. Rates

156   of diffusion can be enhanced by advective flow arising from host motility, feeding currents, or

157   host sinking, and diffusion can also be enhanced by turbulence, which causes shear around the

158   host cell (Murray and Jackson 1992). Host motility, feeding currents, sinking, and turbulence

159   can also lead to the host cell encountering the virus by direct interception (i.e., without the aid

160   of Brownian motion), which is the mechanism by which small flagellates are thought to

161   encounter immotile prey (Shimeta 1993, Kiørboe 2008).

162   The results in Appendix S4 show that diffusion enhanced by advection is the primary

163   mechanism that could significantly increase virus contact rates beyond the 'pure diffusion'

164   scenario (eqn. 2), and so we consider here how that mechanism depends on virus size. An

165   approximate formula for contact rate when advection enhances diffusion is:

166
$$c = 4\pi r_H D_V 0.5 \left[1 + \left(1 + 2\frac{u_H r_H}{D_V} \right)^{1/3} \right] \tag{4}$$

167   where u_H is the velocity of the host relative to the surrounding water (Murray and Jackson 1992).

168   Because the formula includes the term $\left(1 + 2\frac{u_H r_H}{D_V} \right)^{1/3}$, the enhancement of contact rates due to

169   advection is greater as u_H and r_H increase, and also as r_V increases (because D_V is in the

170   denominator, and $D_V \sim 1/r_V$). In other words, host motility matters more for bigger hosts, for

171   hosts that swim faster, and for bigger, less diffusive viruses. These effects are visualized in Fig.

172   1A, which shows that contact rate always declines with virus size, but the penalty for large size

173   is slightly less when hosts are motile. The effect of host motility on contact rates ranges from

174   modest (~2-fold increase for a 20 nm virus infecting a 1μm host swimming at 30 μm s⁻¹) to large

175   (>10-fold increase for a 300 nm virus infecting a 20 μm host swimming at 250 μm s⁻¹). Langlois

176   et al. (2009) use numerical simulation to show that the effect of swimming on diffusion may be

177   underestimated by the formulas used here, but the effect is only ~2-fold for the relevant particle

178   sizes and swimming speeds.

179   Although Brownian motion, potentially enhanced by advection, is expected to drive

180   contact rates, it is possible that direct interception is important for particularly large viruses

181   encountering hosts that generate a strong current (Figs. S1-2). The simple model of interception

182   based on Stokes flow may underestimate particle contact rates (Langlois et al. 2009), and if

183   interception is great enough then 'giant' viruses could have greater contact rates than slightly

184   smaller viruses (Figs. S1-2). Therefore, a better understanding of the physics of particle encounter

185   will be important for understanding virus ecology and size evolution, in addition to

186   understanding predator-prey dynamics among microbes.

187   We have compiled published data on the adsorption rates of viruses of aquatic microbes

188   (cyanobacteria, heterotrophic bacteria, eukaryotic phytoplankton, heterotrophic protists; Table

189   S1; Methods S1) to ask whether virus size has any relation to contact rate or attachment

190   efficiency. There is a tendency for adsorption rate to be greater for larger viruses (Fig. 1B), but

191   this may be due to larger viruses having hosts that are larger, more motile, or both. We therefore

192   used eqns. 2-4 to ask how observed adsorption rates compare to the theoretical maximum (Fig.

193   1C). Predictions and observations are positively correlated (r = 0.68), but many of the

194   observations are 10-100x lower than predicted. This is consistent with a previous analysis of

195   adsorption rates of phages (Talmy et al. 2019) and could be due to sparse host receptors, a low

196   binding affinity of the virus ligand to the host receptor, or a lack of mechanisms for keeping the

197   virus from diffusing away from the host before irreversible binding to the receptor occurs (Storms

198   and Sauvageau 2015). Several of the large eukaryotic viruses have adsorption rates that are

199   higher than the prediction based on Brownian motion alone (eqn. 2), but accounting for host

200   swimming (eqn. 4) brings them closer to the 1:1 line (Fig. S3). However, a number of the phages

201   also have motile hosts, and including reasonable numbers for host swimming speed moves them

202   further below the 1:1 line (Fig. S3).

203   Fig. 1D shows the proximity of adsorption rate to the theoretical maximum as a function

204   of virus size. Although the sample size is limited, it is noteworthy that the five largest viruses are

205   all fairly close to the theoretical maximum. It is possible that some of the functions encoded in

206   larger genomes increase attachment efficiency, such as the synthesis of proteins that aid

207   attachment to host glycans (Rodrigues et al. 2015), or a greater diversity of proteins for binding

208   host receptors (Schwarzer et al. 2012). If attachment efficiency is promoted strongly by certain

209   genes, this could outweigh the reduction in diffusivity associated with larger size, increasing the

210   actual adsorption rate. If a virus is large enough to induce phagocytosis this could also

211   potentially increase encounter efficiency relative to other mechanisms of entry, and some of the

212   largest viruses have been shown to enter their amoeba hosts via phagocytosis (Rodrigues et al.

213   2016). If phagocytosis is in fact a more efficient entry mechanism then large size could be

214   selected for, in order to induce phagocytosis (Rodrigues et al. 2016). Protists that eat bacteria-

215   sized prey are known to ingest larger prey at higher rates, which could be due to differences in

216   contact rates or size preferences during ingestion (Chrzanowski and Šimek 1990, Holen and

217   Boraas 1991, Šimek and Chrzanowski 1992, Epstein and Shiaris 1992). Additional

218   measurements of adsorption rates for viruses across the full size spectrum will be needed to test

219   whether larger size is on average an advantage or disadvantage, and whether rates of successful

220   encounter and infection differ for viruses that enter by phagocytosis.

221

How does virus size affect viral production during infection?

Burst size and latent period of a lytic virus are determined by the rate at which new virions are created during an infection and the timing of cell lysis (You et al. 2002). The production of virions may decline as host resources are depleted, and the timing of cell lysis may evolve in response to intracellular conditions, host density, and other factors (Wang et al. 1996, Abedon et al. 2003). A previous analysis of phytoplankton viruses showed that burst size of dsDNA viruses may be limited by the host resources used in virus genome replication, with lysis occurring once those resources are exhausted (Edwards and Steward 2018). By contrast, small ssDNA and ssRNA viruses that infect large hosts may maximize fitness by lysing the host before those resources are exhausted (Edwards and Steward 2018).

In light of these results, we focus here on the role of viral genome size in constraining burst size and the rate of viral replication. Eqn. 1 shows that burst size and latent period are expected to play a large role in virus fitness, and therefore viral size evolution may be driven in part by its effects on these life history parameters. We previously showed that burst size is correlated with the host:virus genome size ratio (Edwards and Steward 2018), and this relationship can be decomposed into the effects of host genome size and virus genome size (Methods S1; Table S2). Of the total variation in burst size across phytoplankton viruses, 48% is explained by host and virus genome sizes in combination, with a partial R^2 of 30% for host genome size and a partial R^2 of 14% for virus genome size (4% of variation cannot be uniquely attributed to either predictor because host and virus genome sizes are partially correlated). Although burst size tends to decline for larger viruses, the effect of virus genome size is less than proportional, i.e., a tenfold increase in genome size leads to a less than tenfold decrease in burst size (Fig. 2A). The estimated slope for virus genome size is -0.52 (95% CI =[-0.225, -0.911]) when host and

245 virus taxonomy are included as random effects, or a slope of -0.3 (95% CI = [-0.11, -0.55]) when

246 host and virus taxonomy are not included. This means that a tenfold increase in virus genome

247 size would be expected to reduce burst size by a factor of $1/10^{-0.52} = 3.3$ or $1/10^{-0.3} = 2.0$.

248   A less-than-proportional relationship between virus genome size and burst size suggests

249 that larger viruses are producing more total viral material per infection, which could happen if

250 viruses with larger genomes are better at extracting resources from their hosts, better at

251 maintaining metabolic processes that fuel replication, more efficient at transcription or

252 translation, etc. For another perspective on the same processes we can consider how quickly

253 new virions are produced during an infection. All else equal, we would expect that larger virions

254 take longer to construct, due to rate limitation by protein elongation, supply of amino acids or

255 dNTPs, or other processes (You et al. 2002, Birch et al. 2012). The data for phytoplankton viruses

256 exhibit a weak trend of virion production rate declining with virus size ($F_{1,36} = 2.3$, $p = 0.14$; Fig

257 2B), which is consistent with a penalty for larger size that is not directly proportional to size.

258 Finally, rather than looking at production rate on a per virion basis we can consider production

259 rate on a per nucleotide basis, to quantify the total rate at which viral nucleotides are produced

260 during an infection. Nucleotide production rate increases strongly with viral size, with the

261 largest viruses on average producing viral nucleotides ~100x faster than the smallest viruses (Fig.

262 2C; $F_{1,11} = 6.5$, $p = 0.027$). In combination with Fig. 2A-B, this argues that larger size does incur

263 a cost of producing fewer offspring per infection, but that the cost is partially mitigated by a

264 more effective infection process facilitated by the functions encoded in larger genomes.

265 Repeating these analyses using virion volume instead of genome size produces similar results,

266 but the slope of burst size vs. virus genome size is shallower (a 10-fold increase in virion volume

267 leads to a 1.5-fold decrease in burst size; results not shown), and the increase in production rate

268   with virion volume is steeper (the largest viruses create virion volume ~700x faster than the

269   smallest viruses; results not shown). These difference in scaling when using virus genome size

270   vs. virion volume as the predictor are expected, because the genome is a smaller proportion of

271   total virion volume for larger viruses (discussed further in the section *Are larger viruses more*

272   *persistent in the environment?*). An important question for future research is how resource

273   limitation affects these relationships. Nutrient or light limitation have been shown to reduce

274   burst size and lengthen latent period (Wilson et al. 1996, Maat and Brussard 2016,

275   Thamatrakoln et al. 2019), but the data analyzed here are from experiments under resource-

276   replete conditions. It is possible that the advantage of large genome size is greater under

277   resource limitation, due to a greater control of rate-limiting metabolic reactions.

278

279   **Do larger viruses have broader host ranges?**

280   Relatively large viruses could be favored in competition with smaller viruses if a larger

281   genome size is associated with a broader host range (Appendix S3; Chow and Suttle 2015).

282   There is substantial evidence that the ability to attach to host cells plays a major role in defining

283   viral host range (Tétart et al. 1996, Tarutani et al. 2006, Stoddard et al. 2007, Lin et al. 2012, Le

284   et al. 2013), and laboratory experiments often find that hosts evolve resistance by limiting

285   attachment (Lenski 1988, Stoddard et al. 2007). Therefore, virus size may be correlated with

286   host range if having more genes facilitates attachment to a broader range of receptors. For

287   example, the large myovirus phi92 possesses a 'Swiss army knife' of multiple tail fibers and/or

288   spikes that appears to facilitate a relatively broad host range encompassing diverse *Escherichia*

289   *coli* and *Salmonella* strains (Schwarzer et al. 2012). In addition, if the largest viruses are typically

290   ingested by their hosts then they may have a broader host range than smaller non-ingested

291 viruses, due to the fact that ingestion of prey tends to be a less specific interaction than ligand-

292 receptor binding.

293 Although attachment is known to be important in defining host range, in some cases

294 viruses can attach with equal or lesser efficiency to related strains or taxa that do not yield

295 productive infections (Samimi and Drews 1978, Thomas et al. 2011, Yau et al. 2018).

296 Presumably the infections were not productive in these cases because of many processes

297 downstream of attachment that can limit infection success, such as successful entry into the host,

298 intracellular host defense mechanisms, and effective meshing with host replication and

299 translation machinery (Samson et al. 2013). Furthermore, attaching to a relatively narrow range

300 of cell types in the environment can be adaptive if broader attachment results in a loss of virions

301 to hosts that cannot be infected as productively (Heineman et al. 2008). Therefore, the ability

302 to create productive infections once attached may play a large role in defining host range, and

303 viruses with more genes that provide greater autonomy during infection, via control of

304 replication, transcription, translation, or metabolism, or by evading host defenses, may achieve

305 productive infections in a broader range of hosts.

306 To assess whether there is an association between genome size and host range we re-

307 analyzed data from two studies of marine phages. Holmfeldt and colleagues characterized a

308 taxonomically diverse collection of 40 phages isolated on 21 strains of the marine bacterium

309 *Cellulophaga baltica* (Holmfeldt et al. 2007, Holmfeldt et al. 2016, Sulcius and Holmfeldt 2016).

310 Across these phages there is an overall positive correlation between host range and genome

311 size, if host range is quantified as the proportion of 21 *Cellulophaga* strains infected (Fig. 3A; r

312 = 0.49, p = 0.001). We also quantified host range in a way that incorporates phylogenetic

313 relatedness of the host strains, using the summed branch lengths of the infected strains,

314 calculated from a phylogeny estimated from ribosomal internal transcribed spacer (ITS)

315 sequences; this yielded similar results (results not shown). A positive relationship also appears

316 to occur within phage families, with larger isolates of myovirus, siphovirus, and podovirus

317 having a greater host range than smaller isolates within the same groups (Fig. 3A). However,

318 when phage family and phage genus (as classified by Holmfeldt et al. [2016]) are both included

319 as a random effects in a mixed model, in order to account for phylogenetic non-independence

320 in host range (Felsenstein 1985), the effect of host genome size is less clear (χ_1 = 2.45, p = 0.12).

321 This suggests that an even greater phylogenetic diversity of viruses may be needed to robustly

322 test such relationships using a comparative approach. Wichels et al. (1998) characterized 22

323 phages from the North Sea that infect the bacteria *Pseudoalteromonas*. Across the phages in this

324 study there is also a positive correlation between genome size and host range (r = 0.79, p <

325 0.001), and evidence for such a relationship remains after taxonomic random effects are

326 included in a generalized additive mixed model ($\chi_{1.2}$ = 7.2, p = 0.007; taxonomic terms include

327 family, morphotype, and species as defined by the authors).

328   In sum, it may be the case that viruses with larger genomes tend to infect a broader range

329 of hosts, and future analyses from diverse host-virus systems would help test the generality of

330 this pattern. At the same time, it is noteworthy that among *Cellulophaga* phages the smallest

331 phage family, the Microviridae, exhibit relatively broad host range on average (Fig. 3A). A large

332 study of *Vibrio* phages also found that small phages, which the authors classified as

333 Autolykiviridae, had broader host ranges than larger Caudovirales (Kauffman et al. 2018). Future

334 work incorporating a quantitative metric of viral fitness on each host strain would help test

335 whether small, broad-range viruses suffer a cost of lower fitness on each individual host strain

336 (Jover et al. 2013, Record et al. 2016). Tradeoffs affecting viral traits, including those related to

337   size and those orthogonal to size, are likely multidimensional (Goldhill and Turner 2014), and

338   therefore it will be important to measure multiple traits on a diversity of viruses to better

339   understand the constraints on viral evolution.

340

341   **Are larger viruses more persistent in the environment?**

342   The rate at which free virions are lost from a viral population is as important for fitness as

343   adsorption rate, burst size, or host range (eqn. 1). However, the effects of virus size on loss rates

344   are poorly known. Decay of phage infectivity in marine systems has been shown to be

345   influenced by sunlight, adsorption to particles, high molecular weight dissolved material such

346   as enzymes, and ingestion by protists (Suttle and Chen 1992, Noble and Furhman 1997).

347   Although a number of studies have estimated decay rates and how they vary across

348   environmental gradients, we are not aware of studies that look at whether these rates vary

349   systematically with size. Heldal and Bratbaak (1991) noted that viruses > 60 nm disappeared

350   more slowly when viral production was halted with cyanide, but they presented no quantitative

351   data.

352   The physical forces that affect virion stability likely vary with size. For viruses with larger

353   double-stranded genomes the capsid can be highly pressurized due to the dense packaging of

354   negatively charged, dehydrated, curved nucleic acids (Purohit et al. 2003, Li et al. 2008,

355   Molineux and Panja 2013). In a comparative study of coliphages, De Paepe and Taddei (2006)

356   found that phages with a faster multiplication rate in culture had a faster decay rate as well.

357   Faster decay was also associated with a higher nucleotide packaging density and a lower

358   surfacic mass (capsid molecular weight per capsid surface area), suggesting that greater pressure

16

359   makes capsids less stable, and this can be partially mitigated by increased capsid thickness. A

360   re-analysis of their data shows that virus diameter is also negatively correlated with decay rate

361   (r = -0.64; mixed model with virus family random effect – $F_{1,13}$ = 9.1, p = 0.01; Fig. 4A), which

362   could be due to larger phages having lower packaging density, higher surfacic mass, or other

363   causes.

364   We have compiled measurements of genome length and virion dimensions for 193 viruses

365   (Table S3; Methods S1). Among dsDNA viruses infecting unicellular organisms, the fraction of

366   the virion volume occupied by the viral genome declines systematically with increasing virion

367   size (Fig. 4B), although there are notable differences among virus types. The tailless viruses

368   infecting prokaryotes tend to have a lower fractional volume than other dsDNA viruses of the

369   same size, and genome fractional volume declines steeply with increasing virion size for this

370   group and for the eukaryote-infecting dsDNA viruses, which vary more than tenfold in diameter.

371   In contrast, genome fractional volume of the tailed viruses infecting prokaryotes (members of

372   the family *Caudovirales*) is uniformly high and weakly correlated with virion size. Regression

373   analyses indicate that the slope of log(genome fractional volume) vs. log(equivalent spherical

374   diameter) is ~ -1.7 for dsDNA eukaryote viruses and tailless dsDNA viruses infecting prokaryotes,

375   while the slope for tailed phages is -0.24 (Fig. S4A). It is also noteworthy that the largest viruses

376   overlap with small prokaryote and eukaryote cells, both in diameter and genome fractional

377   volume, and that the largest viruses tend to infect phagotrophic eukaryotes.

378   A decline in genome fractional volume with virion size could be driven by selection for

379   virion stability, because the pressure at which a capsid bursts is expected to be inversely

380   proportional to capsid radius (Aznar et al. 2012). If dsDNA viruses generally evolve to have an

381   internal pressure near the burst limit, then the packaging density of the genome would have to

382   decline such that internal pressure is inversely proportional to capsid radius. However, without

383   direct measurements it is unclear whether the observed decline in density is sufficient to

384   equalize the stability of larger and smaller viruses, or whether larger viruses tend to be more or

385   less stable on average. In addition, for tailed dsDNA bacteriophages the injection of the viral

386   genome into the host cell may be driven by high genome packaging density, likely due to

387   hydrodynamic effects of the osmotic imbalance with the host cytoplasm (Molineux and Panja

388   2013). In contrast, eukaryote-infecting viruses and tailless prokaryote viruses often use

389   membrane fusion, endocytosis, or phagocytosis as an entry mechanism, although some have a

390   more phage-like strategy (Nurmemmedov et al. 2007, Wulfmeyer et al. 2012, Mäntynen et al.

391   2019). Therefore, while tailed phages may require dense packaging of nucleic acids, many of

392   the largest eukaryote-infecting viruses may reap little benefit from a pressurized capsid. The

393   presence of a lipid envelope around the capsid may contribute to a lower genome fractional

394   volume for many of the tailless dsDNA prokaryote viruses, but the large difference in fractional

395   volume between these viruses and the tailed phages indicates that the envelope itself is likely

396   not the primary cause (Fig. S4B). The dsRNA viruses may follow a scaling relationship similar

397   to the tailless dsDNA phages, but the number of representatives in the dataset is relatively small

398   (Fig. 4B).

399   Fractional genome volume also declines with virion size for ssRNA and ssDNA viruses,

400   and these two kinds of viruses appear to follow a similar scaling relationship (Fig. 4B). This

401   similarity is also apparent when viruses infecting multicellular organisms are included in the

402   comparison (Fig. S4C). The slope of the relationship is ~ -2.8 for the single-stranded viruses

403   infecting unicells (Fig. S4A). This systematic size scaling may also be due to selection equalizing

404   the stability of larger and smaller viruses, although the scaling relationship and mean fractional

405   volume likely differ between single-stranded viruses and double-stranded viruses due to

406   different physical processes underlying virion assembly and stability (Šiber et al. 2012).

407   Finally, physical instability may not be the primary cause of losses of infectious virions, at

408   least in environments where solar radiation and/or non-specific adsorption are high (Suttle and

409   Chen 1992, Noble and Furhman 1997). Viruses with larger genomes have the capacity to code

410   for and package protective enzymes such as photolyase (Fischer et al. 2014), which could lead

411   to slower decay rates for larger viruses. To understand the consequences of these patterns for

412   viral fitness, and to test whether decay rate generally changes with virus size, future work should

413   investigate decay rates in the laboratory and in natural systems for a broad size range of viruses.

414

415   **Synthesis and outlook**

416   Here we summarize our findings on the relationships between virus size and important

417   virus traits, and we discuss implications and future research directions.

418   The physics of Brownian motion predicts that smaller viruses should encounter their hosts

419   at a faster rate (Fig. 1A), but observed adsorption rates are often much lower than theoretical

420   contact rates (Fig. 1C), and it is possible that larger viruses have a greater attachment efficiency

421   (Fig. 1D). New measurements of adsorption rates of large viruses are needed to test this

422   possibility. Furthermore, entering the host cell via phagocytosis may be a particularly good

423   strategy for ensuring that encounter leads to infection, but demonstrating this quantitatively will

424   require studies on the mechanism of entry, and efficiency of attachment and entry, for diverse

425   viruses.

426   Burst size is lower for larger viruses (Fig. 2A), which is expected if host materials and

427   energy limit viral production, but the cost is less than would be expected if viral production was

428   inversely proportional to genome size. This is likely due to greater control of viral replication

429   and host physiology by viruses with larger genomes, as evidenced by their greater nucleotide

430   production rate (Fig. 2C). Host range may generally increase with virus size, as observed for two

431   diverse groups of phages infecting aquatic bacteria (Fig. 3), and this may be due to greater

432   autonomy during replication, a greater range of counter-defenses, or ability to attach to a greater

433   diversity of receptors. Finally, decay rate has not been widely studied for viruses with unicellular

434   hosts, but a comparison of coliphages suggests that larger viruses could have a lower decay rate

435   (Fig. 4A). In sum these observations suggest that larger viruses experience reductions in burst

436   size, a modest or negligible reduction in adsorption rate, an increase in host range, and

437   potentially a decline in decay rate. These countervailing selection pressures may explain how

438   viruses of very different size have evolved and can persist when infecting the same host

439   population.

440   Are there particular host traits or environmental conditions that could select for larger or

441   smaller viruses? To address this, we can ask whether particular contexts may change the

442   magnitude or direction of relationships between virus size and different virus traits. For example,

443   are there conditions under which the reduction in burst size with increased virus size is lessened?

444   The data synthesized in Figure 2 come from experiments with resource-replete host cultures,

445   and it is possible that under host resource limitation a greater virus size is more costly (due to

446   less energy and materials available for replication) or less costly (due to greater control of host

447   metabolism). Testing this possibility will require measurements of the infection cycle of diverse

448   viruses under different resource conditions. Although contact rates are predicted to decline for

449   larger viruses due to reduced diffusivity, the magnitude of the decline is somewhat less when

450   hosts are motile or generate feeding currents (Fig. 1A). Therefore, larger viruses may be more

451   common among hosts that are highly motile. Likewise, if phagotrophy is an effective means of

452   entering host cells, for viruses large enough to induce phagotrophy, then the largest viruses may

453   be particularly prevalent among phagotrophic hosts (Fig. 4B).

454   If larger viruses tend to have a broader host range, the benefits of broad host range may be

455   greatest when hosts are at low abundance. In our model of host range evolution, a virus will be

456   able to persist if the sum of its host populations, in the absence of viral mortality, exceeds the

457   minimum persistence threshold S^* (Appendix S3). Therefore, oligotrophic environments may be

458   enriched in larger viruses if smaller viruses with narrower host ranges cannot persist. Low host

459   density is also expected to select for lysogeny (Stewart and Levin 1984, Weitz et al. 2019), and

460   so environments with a greater proportion of lysogenic viruses may also tend to have larger

461   viruses in the lytic fraction. Finally, if larger size is associated with reduced decay rates then this

462   relationship may be steeper under conditions of rapid decay, such as exposure to high insolation,

463   which could select for larger viruses.

464   If we combine several of the conditions that could favor large viruses, it may be that motile,

465   phagotrophic protists with low population densities are particularly likely to host giant viruses.

466   Low population densities are characteristic of the oligotrophic open ocean and other

467   environments with low nutrient or energy supply. These environments are also relatively

468   enriched in motile, phagotrophic eukaryotes, including mixotrophic phytoplankton and

469   heterotrophic protists (Edwards 2019), compared to more productive environments where

470   immotile, non-phagotrophic diatoms often dominate microbial biomass. Therefore, the effects

471   of low resource supply on both population densities and community structure may cause larger

472   viruses to be favored in oligotrophic environments. Testing for patterns in virus size distributions

473   across environmental conditions or host types will require a suite of methods, including

474   substantial new isolation efforts, as well as metagenomic protocols that can capture the full size

475   spectrum of viruses in the environment while quantitatively comparing viruses with RNA,

476   ssDNA, and dsDNA genomes. For example, two studies compared RNA and DNA viral

477   metagenomes in coastal ocean environments in Hawai'i (Steward et al. 2013) and Antarctica

478   (Miranda et al. 2016) and found that the abundance of RNA viruses rivals that of DNA viruses.

479   The RNA viruses were essentially exclusively eukaryote-infecting, while most of the DNA

480   viruses were likely phages. This implies that RNA viruses, which tend to be smaller, are more

481   prevalent than larger DNA viruses among eukaryotic viruses in coastal environments. No

482   comparable studies have been performed in the open ocean, but we predict that in the pool of

483   eukaryote-infecting viruses, larger DNA viruses, and 'giant' *Mimiviridae* in particular, are more

484   prevalent than small ssDNA or ssRNA viruses in open ocean environments that tend to be more

485   oligotrophic. It is less clear whether one should expect the size structure of prokaryote-infecting

486   viruses to vary as much across environmental gradients in the ocean, because the abundance

487   of prokaryotes varies much less than the abundance and biomass of unicellular eukaryotes (Li

488   et al. 2004). A study of virus morphology across ocean regions found little variation in the

489   structure of the bulk viral community, which is thought to be numerically dominated by

490   bacteriophages (Brum et al. 2013). However, the locations compared in this study were open

491   ocean environments that varied little in chlorophyll-a, a proxy for community biomass;

492   comparing these environments to productive coastal locations may show that smaller phages

493   become more prevalent in coastal systems.

494 Compilations of virus isolates across the tree of life show that bacteria and archaea are

495 mainly infected by dsDNA viruses with a range of sizes, while eukaryote viruses primarily have

496 RNA genomes that tend to be small, although there is a substantial minority of eukaryote-

497 infecting DNA viruses that tend to be larger (Koonin et al. 2015, Campillo-Balderas et al. 2015).

498 The genome composition and size distribution of eukaryote viruses may reflect alternative

499 strategies responding to the barrier of the eukaryote nucleus, which restricts access to the host's

500 DNA replication and transcription machinery. Larger DNA viruses typically replicate partially

501 or entirely in the cytoplasm, producing their own replication 'factories', while positive-sense

502 single-stranded RNA viral genomes can immediately act as a template for translation (Schmid

503 et al. 2014, Koonin et al. 2015). In our analyses of burst size and decay rate we have treated

504 viruses with different genome types as comparable data points along the virus size spectrum

505 (Figs. 2,4), but future work may consider whether there are important trait or niche differences

506 between these groups that are not explained by size alone. It could also be the case that viruses

507 of different size or different genome type differ in evolutionary rates of 'speciation' or extinction,

508 for example if size or genome type affect the probability of host range shifts, or success in

509 coevolutionary arms races with hosts (Lenski and Levin 1985). Effects of size or genome type

510 on diversification rates could vary with host type or environmental conditions, leading to

511 patterns of virus diversity that vary with host type or environment. Future work may address

512 these possibilities by applying macroevolutionary phylogenetic models to diverse viral clades

513 (e.g., Caetano et al. 2018).

514 Finally, we have focused here on lytic viruses, because a substantial number of lytic viruses

515 infecting unicellular aquatic organisms have been isolated and characterized. Selection for a

516 temperate strategy of integrating into the host genome may also lead to important constraints on

517   virus size evolution. The temperate strategy is a form of vertical symbiont transmission and is

518   expected to be selected for when host densities are low (making horizontal transmission less

519   likely) and when the virus presents a low fitness cost or even a benefit (Weitz et al. 2019).

520   Considering the potential costs and benefits of prophage or provirus in the host genome, a larger

521   virus genome will incur greater material and energy costs but will also be able to encode more

522   functions that could benefit the host. Future work may consider how the size distribution of

523   temperate viruses compares to lytic viruses, across environmental gradients or different host

524   characteristics.

525

Acknowledgements

531

References

533   Abedon, S. T., Hyman, P., & Thomas, C. (2003). Experimental examination of bacteriophage latent-

534   period evolution as a response to bacterial availability. *Appl. Environ. Microbiol.*, *69*(12), 7499–7506.

535   Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R.,

536   Bouma-Gregson, K., Amano, Y. and He, C. (2020). Clades of huge phages from across Earth's

537   ecosystems. *Nature*, 1-7.

538   Aznar, M., Luque, A., & Reguera, D. (2012). Relevance of capsid structure in the buckling and maturation

539   of spherical viruses. *Physical biology*, *9*(3), 036003.

540   Birch, E. W., Ruggero, N. A., & Covert, M. W. (2012). Determining host metabolic limitations on viral

541   replication via integrated modeling and experimental perturbation. *PLoS Computational Biology*, *8*(10),

542   e1002746.

543   Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory

544   of ecology. *Ecology*, *85*(7), 1771–1789.

545   Brum, J. R., Schenck, R. O., & Sullivan, M. B. (2013). Global morphological analysis of marine viruses

546   shows minimal regional variation and dominance of non-tailed viruses. *The ISME journal*, *7*(9), 1738-

547   1751.

548   Caetano, D. S., O'Meara, B. C., & Beaulieu, J. M. (2018). Hidden state models improve state-dependent

549   diversification approaches, including biogeographical models. *Evolution*, *72*(11), 2308-2324.

550   Campillo-Balderas, J. A., Lazcano, A., & Becerra, A. (2015). Viral genome size distribution does not

551   correlate with the antiquity of the host lineages. *Frontiers in Ecology and Evolution*, *3*, 143.

552   Chow, C. E. T., & Suttle, C. A. (2015). Biogeography of viruses in the sea. *Annual review of virology*, *2*,

553   41-66.

554   Chrzanowski, T. H., & Šimek, K. (1990). Prey-size selection by freshwater flagellated protozoa. *Limnology*

555   *and Oceanography*, *35*(7), 1429–1436.

556   Cui, J., Schlub, T.E. and Holmes, E.C., 2014. An allometric relationship between the genome length and

557   virion volume of viruses. *Journal of Virology*, *88*(11), pp.6403-6410.

558   De Paepe, M., & Taddei, F. (2006). Viruses' life history: towards a mechanistic basis of a trade-off

559   between survival and reproduction among phages. *PLoS Biology*, *4*(7), e193.

560   Devoto, A.E., Santini, J.M., Olm, M.R., Anantharaman, K., Munk, P., Tung, J., Archie, E.A., Turnbaugh,

561   P.J., Seed, K.D., Blekhman, R. and Aarestrup, F.M. (2019). Megaphages infect Prevotella and variants

562   are widespread in gut microbiomes. *Nature Microbiology*, *4*(4), 693.

563   Edwards, K. F., & Steward, G. F. (2018). Host traits drive viral life histories across phytoplankton viruses.

564   *The American Naturalist*, *191*(5), 566–581.

565   Edwards, K. F., Thomas, M. K., Klausmeier, C. A., & Litchman, E. (2012). Allometric scaling and

566   taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton.

567   *Limnology and Oceanography*, *57*(2), 554–566.

568   Epstein, S. S., & Shiaris, M. P. (1992). Size-selective grazing of coastal bacterioplankton by natural

569   assemblages of pigmented flagellates, colorless flagellates, and ciliates. *Microbial Ecology*, *23*(3), 211–

570   225.

571   Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, *125*(1), 1–15.

572   Finkel, Z. V, & Irwin, A. J. (2001). Light absorption by phytoplankton and the filter amplification

573   correction: cell size and species effects. *Journal of Experimental Marine Biology and Ecology*, *259*(1),

574   51–61.

575   Fischer, M. G., Allen, M. J., Wilson, W. H., & Suttle, C. A. (2010). Giant virus with a remarkable

576   complement of genes infects marine zooplankton. *Proceedings of the National Academy of*

577   *Sciences*, *107*(45), 19508-19513.

578   Fischer, M. G., Kelly, I., Foster, L. J., & Suttle, C. A. (2014). The virion of Cafeteria roenbergensis virus

579   (CroV) contains a complex suite of proteins for transcription and DNA repair. *Virology*, *466*, 82-94.

580   Fuchs, H. L., & Franks, P. J. S. (2010). Plankton community properties determined by nutrients and size-

581   selective feeding. *Marine Ecology Progress Series*, *413*, 1–15.

582   Goldhill, D. H., & Turner, P. E. (2014). The evolution of life history trade-offs in viruses. *Current opinion*

583   *in virology*, *8*, 79-84.

584   Heineman, R. H., Springman, R., & Bull, J. J. (2008). Optimal foraging by bacteriophages through host

585   avoidance. *The American Naturalist*, *171*(4), E149–E157.

586   Heldal, M., & Bratbak, G. (1991). Production and decay of viruses in aquatic environments. *Mar. Ecol.*

587   *Prog. Ser*, *72*(3), 205–212.

588   Holen, D. A., & Boraas, M. E. (1991). The feeding behavior of *Spumella* sp. as a function of particle size:

589   Implications for bacterial size in pelagic systems. *Hydrobiologia*, *220*(1), 73–88.

590   Holmfeldt, K., Middelboe, M., Nybroe, O., & Riemann, L. (2007). Large variabilities in host strain

591   susceptibility and phage host range govern interactions between lytic marine phages and their

592   *Flavobacterium* hosts. *Appl. Environ. Microbiol.*, *73*(21), 6730–6739.

593   Holmfeldt, K., Solonenko, N., Howard-Varona, C., Moreno, M., Malmstrom, R. R., Blow, M. J., & Sullivan,

594   M. B. (2016). Large-scale maps of variable infection efficiencies in aquatic Bacteroidetes phage-host

595   model systems. *Environmental Microbiology*, *18*(11), 3949–3961.

596   Holmfeldt, K., Solonenko, N., Shah, M., Corrier, K., Riemann, L., VerBerkmoes, N. C., & Sullivan, M. B.

597   (2013). Twelve previously unknown phage genera are ubiquitous in global oceans. *Proceedings of*

598   *the National Academy of Sciences*, *110*(31), 12798–12803.

599   Jover, L. F., Cortez, M. H., & Weitz, J. S. (2013). Mechanisms of multi-strain coexistence in host–phage

600   systems with nested infection networks. *Journal of Theoretical Biology*, *332*, 65–77.

601   Karp-Boss, L., Boss, E., & Jumars, P. A. (1996). Nutrient fluxes to planktonic osmotrophs in the presence

602   of fluid motion. *Oceanography and Marine Biology*, *34*, 71–108.

603 Kauffman, K.M., Hussain, F.A., Yang, J., Arevalo, P., Brown, J.M., Chang, W.K., VanInsberghe, D.,
604   Elsherbini, J., Sharma, R.S., Cutler, M.B. and Kelly, L. (2018). A major lineage of non-tailed dsDNA
605   viruses as unrecognized killers of marine bacteria. *Nature*, *554*(7690), 118.

606 Kiørboe, T. (2008). *A mechanistic approach to plankton ecology*. Princeton University Press.

607 Koonin, E. V, & Yutin, N. (2019). Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes
608   and convergent origins of viral gigantism. *Adv. Virus Res*, *103*, 167–202.

609 Langlois, V. J., Andersen, A., Bohr, T., Visser, A. W., & Kiørboe, T. (2009). Significance of swimming and
610   feeding currents for nutrient uptake in osmotrophic and interception-feeding flagellates. *Aquatic*
611   *Microbial Ecology*, *54*(1), 35–44.

612 Le, S., He, X., Tan, Y., Huang, G., Zhang, L., Lux, R., Shi, W. and Hu, F. (2013). Mapping the tail fiber
613   as the receptor binding protein responsible for differential host specificity of *Pseudomonas aeruginosa*
614   bacteriophages PaP1 and JG004. *PloS One*, *8*(7), e68562.

615 Lenski, R. E., & Levin, B. R. (1985). Constraints on the coevolution of bacteria and virulent phage: a
616   model, some experiments, and predictions for natural communities. *The American Naturalist*, *125*(4),
617   585-602.

618 Lenski, R. E. (1988). Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in
619   competitive fitness among mutants resistant to virus T4. *Evolution*, *42*(3), 425–432.

620 Levin, B. R., Stewart, F. M., & Chao, L. (1977). Resource-limited growth, competition, and predation: a
621   model and experimental studies with bacteria and bacteriophage. *The American Naturalist*, *111*(977),
622   3–24.

623 Li, W. K., Head, E. J., & Harrison, W. G. (2004). Macroecological limits of heterotrophic bacterial
624   abundance in the ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*(11), 1529-
625   1540.

626   Li, Z., Wu, J., & Wang, Z. G. (2008). Osmotic pressure and packaging structure of caged

627   DNA. *Biophysical journal*, *94*(3), 737-746.

628   Lin, T.-Y., Lo, Y.-H., Tseng, P.-W., Chang, S.-F., Lin, Y.-T., & Chen, T.-S. (2012). A T3 and T7 recombinant

629   phage acquires efficient adsorption and a broader host range. *PloS One*, *7*(2), e30954.

630   Maat, D. S., & Brussaard, C. P. D. (2016). Both phosphorus-and nitrogen limitation constrain viral

631   proliferation in marine phytoplankton. *Aquatic Microbial Ecology*, *77*(2), 87–97.

632   Mäntynen, S., Sundberg, L. R., Oksanen, H. M., & Poranen, M. M. (2019). Half a century of research on

633   Membrane-Containing bacteriophages: bringing new concepts to modern virology. *Viruses*, *11*(1), 76.

634   Miranda, J. A., Culley, A. I., Schvarcz, C. R., & Steward, G. F. (2016). RNA viruses as major contributors

635   to Antarctic virioplankton. *Environmental microbiology*, *18*(11), 3714-3727.

636   Molineux, I. J., & Panja, D. (2013). Popping the cork: mechanisms of phage genome ejection. *Nature*

637   *Reviews Microbiology*, *11*(3), 194-204.

638   Murray, A. G., & Jackson, G. A. (1992). Viral dynamics: a model of the effects of size, shape, motion and

639   abundance of single-celled planktonic organisms and other particles. *Marine Ecology Progress Series*,

640   103–116.

641   Noble, R. T., & Fuhrman, J. A. (1997). Virus decay and its causes in coastal waters. *Appl. Environ.*

642   *Microbiol.*, *63*(1), 77–83.

643   Nurmemmedov E, Castelnovo M, Catalano CE, Evilevitch A. (2007). Biophysics of viral infectivity:

644   matching genome length with capsid size. *Q Rev Biophys,* 40, 327–356.

645   Purohit, P. K., Kondev, J., & Phillips, R. (2003). Mechanics of DNA packaging in viruses. *Proceedings of*

646   *the National Academy of Sciences*, *100*(6), 3173-3178.

647   Rodrigues, R. A. L., Abrahão, J. S., Drumond, B. P., & Kroon, E. G. (2016). Giants among larges: how

648   gigantism impacts giant virus entry into amoebae. *Current Opinion in Microbiology*, *31*, 88–93.

649 Samimi, B., & Drews, G. (1978). Adsorption of cyanophage AS-1 to unicellular cyanobacteria and
650   isolation of receptor material from *Anacystis nidulans*. *Journal of Virology*, *25*(1), 164–174.

651 Samson, J. E., Magadán, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating
652   bacterial defences. *Nature Reviews Microbiology*, *11*(10), 675.

653 Schmid, M., Speiseder, T., Dobner, T., & Gonzalez, R. A. (2014). DNA virus replication
654   compartments. *Journal of virology*, *88*(3), 1404-1420.

655 Schwartz, M. (1976). The adsorption of coliphage lambda to its host: effect of variations in the surface
656   density of receptor and in phage-receptor affinity. *Journal of Molecular Biology*, *103*(3), 521–536.

657 Schwarzer, D., Buettner, F.F., Browning, C., Nazarov, S., Rabsch, W., Bethe, A., Oberbeck, A., Bowman,
658   V.D., Stummeyer, K., Mühlenhoff, M. and Leiman, P.G. (2012). A multivalent adsorption apparatus
659   explains the broad host range of phage phi92: a comprehensive genomic and structural analysis.
660   *Journal of Virology*, *86*(19), 10384–10398.

661 Sharon, I., Battchikova, N., Aro, E.M., Giglione, C., Meinnel, T., Glaser, F., Pinter, R.Y., Breitbart, M.,
662   Rohwer, F. and Béja, O. (2011). Comparative metagenomics of microbial traits within oceanic viral
663   communities. *The ISME Journal*, *5*(7), 1178.

664 Shimeta, J. (1993). Diffusional encounter of submicrometer particles and small cells by suspension
665   feeders. *Limnol. Oceanogr.*, *38*, 456–465.

666 Shinn, G. L., & Bullard, B. L. (2018). Ultrastructure of Meelsvirus: A nuclear virus of arrow worms (phylum
667   Chaetognatha) producing giant "tailed" virions. *PloS One*, *13*(9), e0203282.

668 Šiber, A., Božič, A. L., & Podgornik, R. (2012). Energies and pressures in viruses: contribution of
669   nonspecific electrostatic interactions. *Physical chemistry chemical physics*, *14*(11), 3746-3765.

670 Šimek, K., & Chrzanowski, T. H. (1992). Direct and indirect evidence of size-selective grazing on pelagic
671   bacteria by freshwater nanoflagellates. *Appl. Environ. Microbiol.*, *58*(11), 3715–3720.

672   Steward, G. F., Culley, A. I., Mueller, J. A., Wood-Charlson, E. M., Belcaid, M., & Poisson, G. (2013).

673   Are we missing half of the viruses in the ocean?. *The ISME journal*, *7*(3), 672-679.

674   Stoddard, L. I., Martiny, J. B. H., & Marston, M. F. (2007). Selection and characterization of cyanophage

675   resistance in marine Synechococcus strains. *Appl. Environ. Microbiol.*, *73*(17), 5516–5522.

676   Storms, Z. J., & Sauvageau, D. (2015). Modeling tailed bacteriophage adsorption: Insight into

677   mechanisms. *Virology*, *485*, 355–362.

678   Šulčius, S., & Holmfeldt, K. (2016). Viruses of microorganisms in the Baltic Sea: current state of research

679   and perspectives. *Marine Biology Research*, *12*(2), 115–124.

680   Suttle, C. A., & Chen, F. (1992). Mechanisms and rates of decay of marine viruses in seawater. *Appl.*

681   *Environ. Microbiol.*, *58*(11), 3721–3729.

682   Talmy, D., Beckett, S. J., Zhang, A. B., Taniguchi, D. A. A., Weitz, J. S., & Follows, M. J. (2019).

683   Contrasting controls on microzooplankton grazing and viral infection of microbial prey. Front. *Mar.*

684   *Sci*, *6*, 182.

685   Tarutani, K., Nagasaki, K., & Yamaguchi, M. (2006). Virus adsorption process determines virus

686   susceptibility in *Heterosigma akashiwo* (Raphidophyceae). *Aquatic Microbial Ecology*, *42*(3), 209–

687   213.

688   Tétart, F., Repoila, F., Monod, C., & Krisch, H. M. (1996). Bacteriophage T4 host range is expanded by

689   duplications of a small domain of the tail fiber adhesin. Elsevier.

690   Thamatrakoln, K., Talmy, D., Haramaty, L., Maniscalco, C., Latham, J.R., Knowles, B., Natale, F., Coolen,

691   M.J., Follows, M.J. and Bidle, K.D. (2019). Light regulation of coccolithophore host–virus interactions.

692   *New Phytologist*, *221*(3), 1289–1302.

693   Thingstad, T. F., Våge, S., Storesund, J. E., Sandaa, R.-A., & Giske, J. (2014). A theoretical analysis of how

694   strain-specific viruses can control microbial species diversity. *Proceedings of the National Academy*

695   *of Sciences*, *111*(21), 7813–7818.

696   Thomas, R., Grimsley, N., Escande, M., Subirana, L., Derelle, E., & Moreau, H. (2011). Acquisition and

697   maintenance of resistance to viruses in eukaryotic phytoplankton populations. *Environmental*

698   *Microbiology*, *13*(6), 1412–1420.

699   Tilman, D. (1982). *Resource competition and community structure*. Princeton university press.

700   Tomaru, Y., Mizumoto, H., Takao, Y., & Nagasaki, K. (2009). Co-occurrence of DNA-and RNA-viruses

701   infecting the bloom-forming dinoflagellate, *Heterocapsa circularisquama*, on the Japan coast.

702   *Plankton and Benthos Research*, *4*(4), 129–134.

703   Wang, N., Dykhuizen, D. E., & Slobodkin, L. B. (1996). The evolution of phage lysis timing. *Evolutionary*

704   *Ecology*, *10*(5), 545–558.

705   Weitz, J. S., Li, G., Gulbudak, H., Cortez, M. H., & Whitaker, R. J. (2019). Viral invasion fitness across a

706   continuum from lysis to latency. *Virus evolution*, *5*(1), vez006.

707   Wichels, A., Biel, S. S., Gelderblom, H. R., Brinkhoff, T., Muyzer, G., & Schütt, C. (1998). Bacteriophage

708   diversity in the North Sea. *Appl. Environ. Microbiol.*, *64*(11), 4128–4133.

709   Wickham, T. J., Granados, R. R., Wood, H. A., Hammer, D. A., & Shuler, M. L. (1990). General analysis

710   of receptor-mediated viral attachment to cell surfaces. *Biophysical Journal*, *58*(6), 1501–1516.
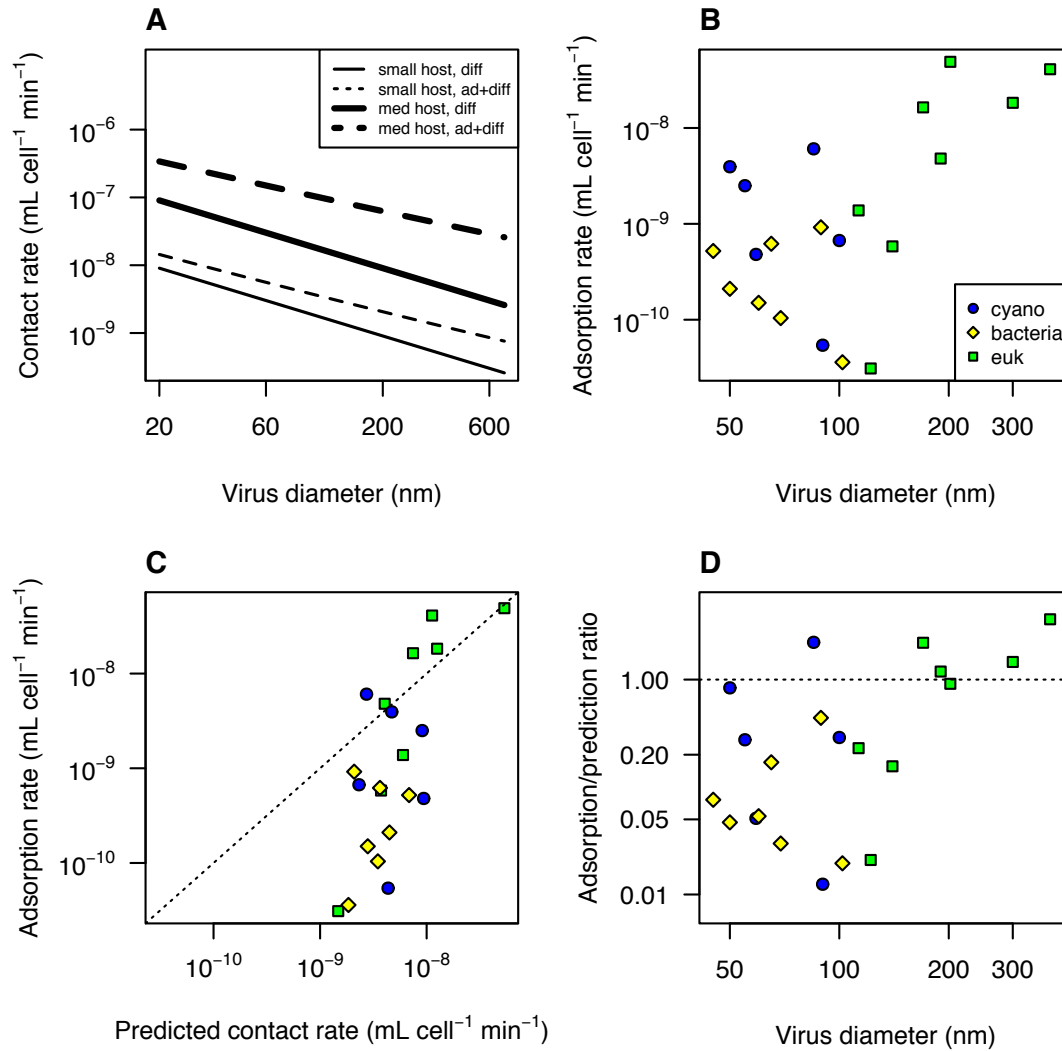
711   Wilhelm, S.W., Bird, J.T., Bonifer, K.S., Calfee, B.C., Chen, T., Coy, S.R., Gainer, P.J., Gann, E.R.,

712   Heatherly, H.T., Lee, J. and Liang, X., 2017. A student's guide to giant viruses infecting small

713   eukaryotes: from Acanthamoeba to Zooxanthellae. *Viruses*, *9*(3), 46.

714   Wilson, W. H., Carr, N. G., & Mann, N. H. (1996). The effect of phosphate status on the kinetics of

715   cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH7803. *Journal of*

716   *Phycology*, *32*(4), 506–516.

717   Wulfmeyer, T., Polzer, C., Hiepler, G., Hamacher, K., Shoeman, R., Dunigan, D.D., Van Etten, J.L.,

718   Lolicato, M., Moroni, A., Thiel, G. and Meckel, T. (2012). Structural organization of DNA in chlorella

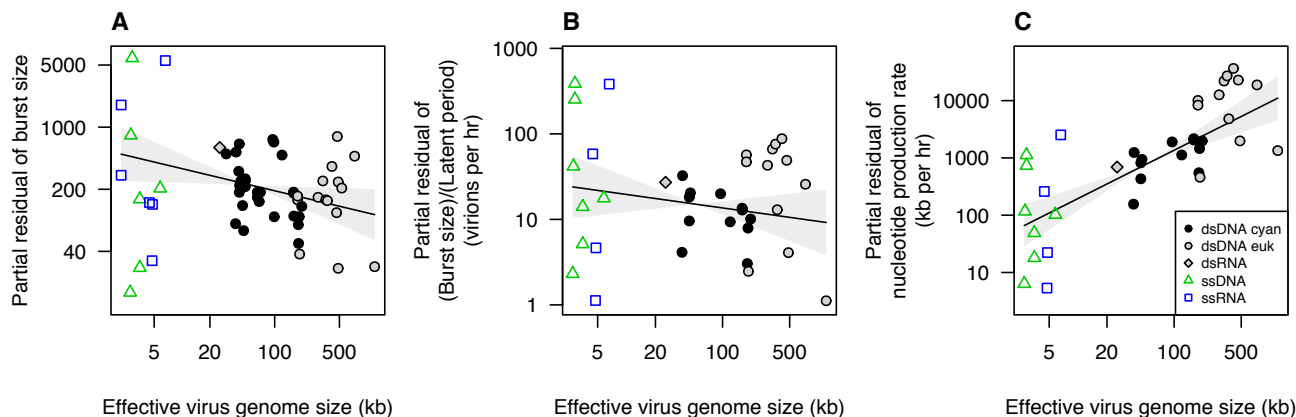719   viruses. *PLoS One*, *7*(2).

720   Yau, S., Caravello, G., Fonvieille, N., Desgranges, É., Moreau, H., & Grimsley, N. (2018). Rapidity of

721   genomic adaptations to prasinovirus infection in a marine microalga. *Viruses*, *10*(8), 441.

722   You, L., Suthers, P. F., & Yin, J. (2002). Effects of *Escherichia coli* physiology on growth of phage T7 in

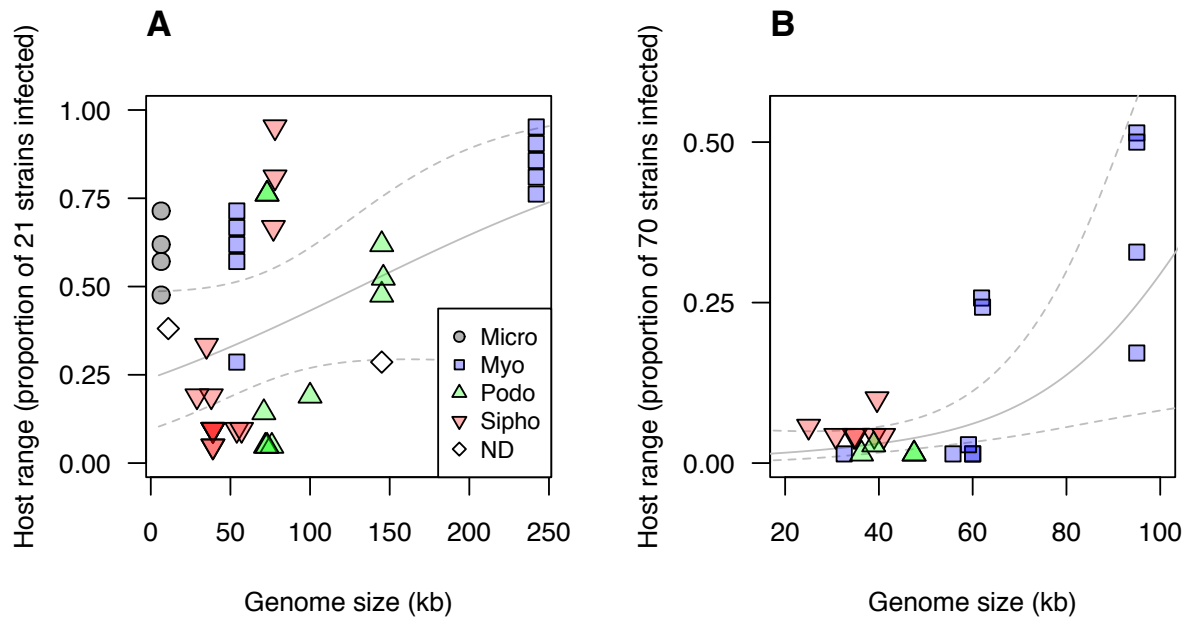723   vivo and in silico. *Journal of Bacteriology*, *184*(7), 1888–1894.

Figure 1. Model predictions and observations of contact rate and adsorption rate. (A) Predicted contact rate as a function of virus size. The 'small host' lines correspond to a host with diameter 1 μm and swimming speed 10 μm s^{-1}, and the 'medium host' lines correspond to a host with diameter 10 μm and swimming speed 100 μm s^{-1}. The solid lines are the pure diffusion prediction (no advection), and the dashed lines are for swimming hosts (advection+diffusion). (B) Observed adsorption rates for viruses of aquatic bacteria, phytoplankton, and the protist *Cafeteria roenbergensis* (Table S1; Methods S1). (C) Observed rates vs. theoretical predictions described in Methods S1. (D) Observed adsorption rate, relative to the theoretical maximum contact rate, as a function of virus size. 'cyano' = virus infecting
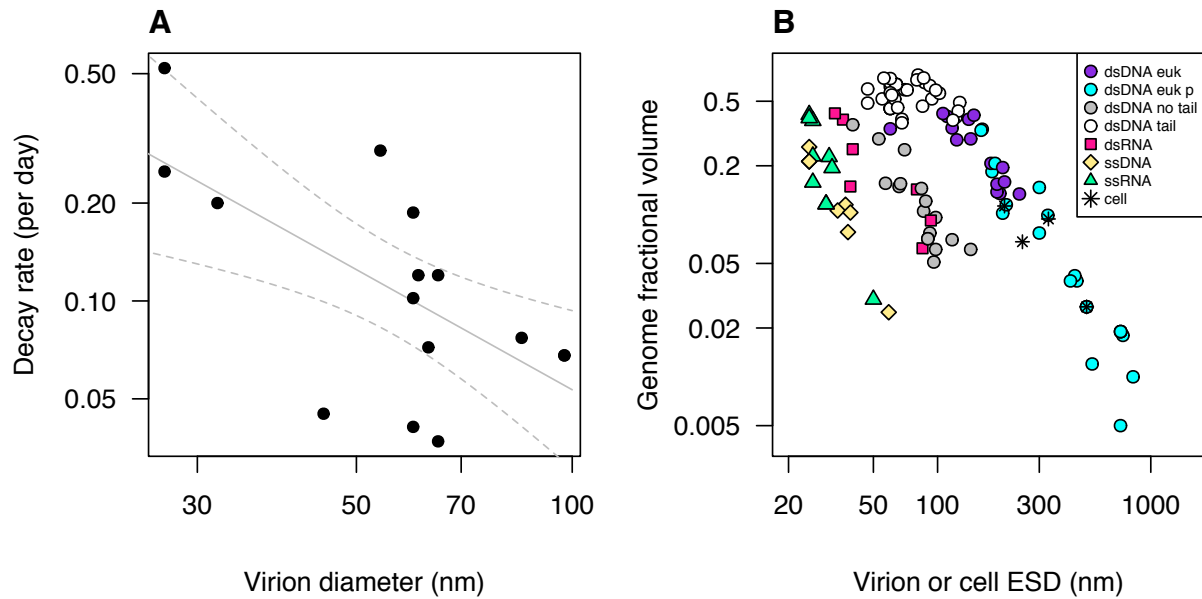
cyanobacterium, 'bacteria' = virus infecting heterotrophic bacterium, 'euk' = virus infecting unicellular

733 eukaryote.

734 **Figure 2**. Burst size and production rates as a function of virus genome size for viruses of phytoplankton

735 and the protist *Cafeteria roenbergensis*. (A) Burst size vs. virus genome size. Partial residual burst size is

736 plotted, which removes variation in burst size explained by host genome size. Effective virus genome

737 size is plotted, which divides genome size by 2 for single-stranded viruses. (B) Virion production rate,

738 (burst size)/(latent period), vs. virus genome size. Partial residual production rate is plotted, which

739 removes variation explained by host growth rate. (C) Nucleotide production rate (kb per hr) vs. virus

740 genome size. Partial residual production rate is plotted, which removes variation explained by host

741 growth rate. Plotted lines are fitted smoothers ± 95% CI from generalized additive mixed models. 'dsDNA

742 cyan' = dsDNA viruses infecting cyanobacteria; 'dsDNA euk' = dsDNA viruses infecting unicellular

743 eukaryotes. dsRNA, ssDNA, and ssRNA viruses all infect eukaryotes in this dataset. Data and data sources

744 are presented in Table S2 and described in Methods S1.

Figure 3. Host range vs. genome size in marine bacteriophages. (A) Host range (proportion of 21 strains infected) vs. genome size for *Cellulophaga baltica* phages (Sulcius and Holmfeldt 2016). (B) Host range (proportion of 70 strains infected) vs. genome size for *Pseudoalteromonas* phages (Wichels et al. 1998). Lines depict fitted smoothers and 95% CI from generalized additive mixed models that include taxonomic random effects to account for related viruses having similar host ranges. 'Micro' = Microviridae, 'Myo' = Myoviridae, 'Podo', = Podoviridae, 'Sipho' = Siphoviridae, 'ND' = taxonomy not determined. Statistical methods are described further in Methods S1.

A

Decay rate (per day) vs Virion diameter (nm)

B

Genome fractional volume vs Virion or cell ESD (nm)

Legend:
- ● dsDNA euk
- ● dsDNA euk p
- ● dsDNA no tail
- ○ dsDNA tail
- ■ dsRNA
- ◇ ssDNA
- ▲ ssRNA
- ✳ cell

752 **Figure 4**. (A) Decay rate vs. virion diameter, for phages that infect *E. coli* (data from Table 1 in De Paepe

753 and Taddei 2006). Lines are linear regression fit ± 95% CI. (B) Genome fractional volume vs. equivalent

754 spherical diameter, for viruses that infect unicellular organisms, and for four representative cellular

755 organisms. Genome fractional volume is the estimated nucleic acid volume divided by total virion

756 volume. Cell or virion volumes were estimated from reported outer dimensions (including outer envelope,

757 if present, but excluding tails or fibrils) using volume formulae for simplified approximate shapes as

758 indicated in Table S3. Nucleic acid volume is estimated as volume of a cylinder with diameter being

759 2.37 nm for double-stranded nucleic acid and 1.19 nm for single-stranded and 0.34 nm per nt or bp.

760 'dsDNA euk' = dsDNA viruses infecting non-phagotrophic eukaryotes, 'dsDNA euk p' = dsDNA viruses

761 infecting phagotrophic eukaryotes, 'dsDNA no tail' = tailless dsDNA viruses infecting prokaryotes,

762 'dsDNA tail' = tailed dsDNA viruses infecting prokaryotes, 'dsRNA' = dsRNA viruses, 'ssDNA' = ssDNA

763 viruses, 'ssRNA' = ssRNA viruses, 'cell' = cellular organisms (one archaeon, two bacteria, one eukaryote).

764 Data and data sources are presented in Table S3 and described in Methods S1.

37