

1 *Review*

2 **Giant Viruses – Big Surprises**

3 **Nadav Brandes¹ and Michal Linial^{2,*}**

4

5 ¹ The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of
6 Jerusalem, Israel; nadav.brandes@mail.huji.ac.il

7 ² Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew
8 University of Jerusalem, Israel; michall@cc.huji.ac.il

9 * Correspondence: michall@cc.huji.ac.il; Tel.: +972-02-6585425

10 Received: date; Accepted: date; Published: date

11 **Abstract:** Viruses are the most prevalent infectious agents, populating almost every ecosystem on
12 earth. Most viruses carry only a handful of genes supporting their replication and the production of
13 capsids. It came as a great surprise in 2003 when the first giant virus was discovered and found to
14 have a >1Mbp genome encoding almost a thousand proteins. Following this first discovery, dozens
15 of giant virus strains across several viral families have been reported. Here, we provide an updated
16 quantitative and qualitative view on giant viruses and elaborate on their shared and variable
17 features. We review the complexity of giant virus proteomes, which include functions traditionally
18 associated only with cellular organisms. These unprecedented functions include components of the
19 translation machinery, DNA maintenance, and metabolic enzymes. We discuss the possible
20 underlying evolutionary processes and mechanisms that might have shaped the diversity of giant
21 viruses and their genomes, highlighting their remarkable capacity to hijack genes and genomic
22 sequences from their hosts and environments. This leads us to examine prominent theories
23 regarding the origin of giant viruses. Finally, we present the emerging ecological view of giant
24 viruses, found across widespread habitats and ecological systems, with respect to the environment
25 and human health.

26 **Keywords:** Amebae viruses; Viral evolution; Protein domains; Mimivirus; dsDNA viruses;
27 Translation machinery; Pandoravirus; NCLDV.

28

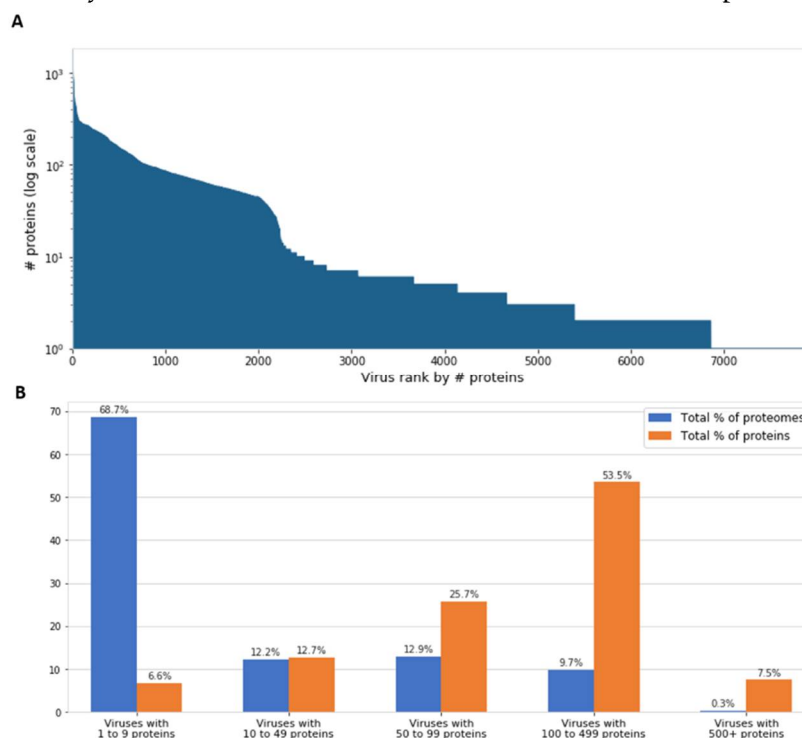
29 **1. Giant viruses and the viral world**

30 Viruses are cell infecting agents present in almost every ecosystem. Questions regarding viral
31 origin and early evolution alongside all living organisms (Bacteria, Archaea and Eukarya) are still
32 widely open, and relevant theories remain speculative [1-4]. As viruses are exceptionally diverse and
33 undergo rapid changes, it is impossible to construct an ancestral lineage tree for the viral world [5-9].
34 Instead, virus families are categorized according to the nature of their genetic material, mode of
35 replication, pathogenicity, and structural properties [10].

36 At present, the viral world is represented by over 8,000 reference genomes [11]. The International
37 Committee on Taxonomy of Viruses (ICTV) provides a universal virus taxonomical classification
38 proposal that covers ~150 families and ~850 genera, with many viruses yet unclassified [12]. This
39 collection provides a comprehensive, compact set of virus representatives.

40 Inspection of viral genomes reveals that most known viruses have genomes encoding only a few
41 proteins. Actually, 69% of all known viruses have less than 10 proteins encoded in their genomes
42 (Figure 1). It is a common assumption that viruses demonstrate near-optimal genome packing and
43 information compression, presumably in order to maximize their replication rate, number of
44 progenies, and other parameters that increase infectivity [13,14]. However, a debate is still ongoing
45 over the generality of these phenomena [15], and there is a non-negligible percentage of larger viruses
46 (Figure 1). On the far end of the distribution, there are the viruses with hundreds of genes, most of

47 them belong to giant viruses. Specifically, while only 0.3% of currently known viral proteomes have
 48 500 or more proteins, they encode as much as 7.5% of the total number of viral proteins (Figure 1B).



49
 50
 51 **Figure 1.** Number of proteins encoded by viruses. (A) The number of encoded proteins
 52 (y-axis) in all 7,959 viral representatives, ranked in descending order. (B) Partitioning
 53 of the 7,959 viral proteomes by the number of encoded proteins. The 0.3% viral
 54 proteomes with the highest number of proteins (over 500) encode 7.5% of the total
 55 number of viral proteins.
 56

57 2. The discovery of giant viruses

58 The first giant virus, *Acanthamoeba polyphaga mimivirus* (APMV), was discovered in 2003 [16].
 59 Its size was unprecedented, being on the scale of small bacteria or archaea cells [17]. Unlike any
 60 previously identified virus, APMV could be seen with a light microscope [18,19]. Initially it was
 61 mistaken for a bacterium, and recognized as a virus only ten years after its isolation [20]. Up to this
 62 day, most of its proteins remain uncharacterized [21,22]. Notably, even more than a decade after the
 63 discovery of APMV, the identification of giant viruses is sometimes still involved with confusion, as
 64 illustrated in the discovery of the *Pandoravirus inopinatum* [23] that was initially described as an
 65 endoparasitic organism, and *Pithovirus sibericum* [24] that was also misinterpreted as an archaeal
 66 endocytobiont (see discussion in [20,25]).

67 In the following years from the initial discovery of APMV (2003), many additional giant viral
 68 species have been identified and their genomes fully sequenced. Most giant viral genomes have been
 69 obtained from large-scale metagenomic sequencing projects covering aquatic ecosystems (e.g.,
 70 oceans, pools, lakes and cooling wastewater units) [26,27]; others sequenced from samples extracted
 71 from underexplored geographical and ecological niches (e.g., the Amazon River, deep seas and forest
 72 soils) [28-31]. Despite the accumulation of many more giant virus representatives, the fraction of
 73 uncharacterized proteins remains exceptionally high [32]. Many of these uncharacterized proteins
 74 were also considered ORFans (i.e., no significant match to any other sequence). However, with
 75 proteomes of closely related species, the fraction of ORFans obviously drops. For example, 93% of
 76 the *Pandoravirus salinus* proteins, the first representative of this family [33] were reported as
 77 ORFans. However, with the complete proteomes of 5 additional *Pandoravirus* species (*inopinatum*,

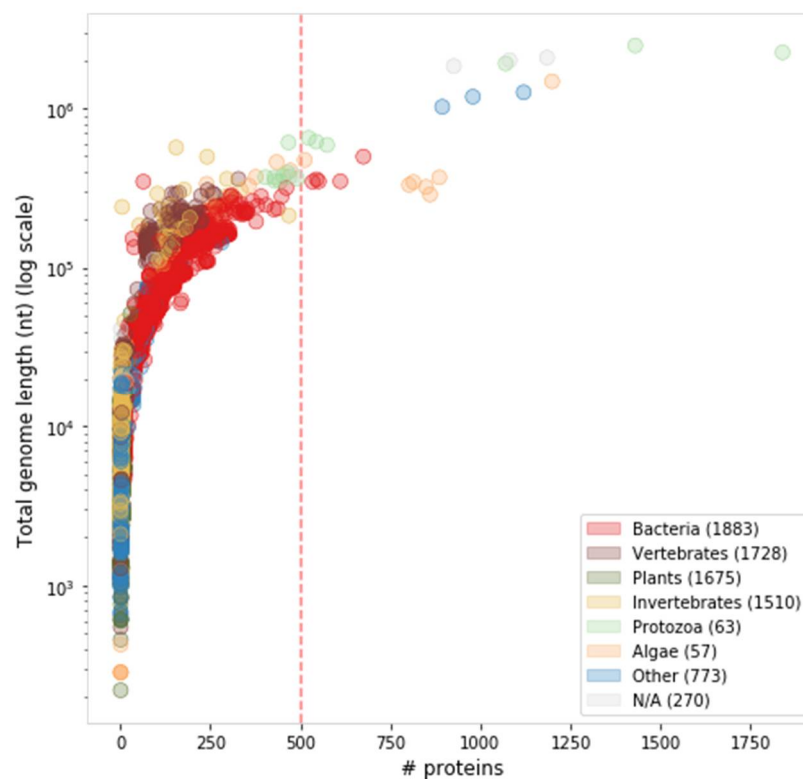
78 macleodensis, neocaledonia, dulcis, and quercus) the number of ORFans dropped to 29% (i.e., with a
 79 substantial similarity to at least another Pandoravirus protein sequence). Still, the vast majority of
 80 Pandoravirus proteins remain uncharacterized.

81 At present, there are over a hundred giant virus isolates which reveal fascinating and
 82 unexpected characteristics. These extreme instances on the viral landscape challenge the current
 83 theories on genome size and compactness in viruses, and provide a new perspective on the very
 84 concept of a virus and viral origin [4].

86 3. Definition of giant viruses

87 Attempts to distinguish giant viruses from other large viruses remain somewhat fuzzy [34,35].
 88 Any definition for giant viruses would necessarily involve some arbitrary threshold, as virus size,
 89 whether physical, genomic or proteomic, is clearly a continuum (Figure 2). Giant viruses were
 90 initially defined by their physical size as allowing visibility by a light microscope [32]. In this report,
 91 we prefer a proteomic definition, even if somewhat arbitrary. We consider giant viruses as Eukaryote-
 92 infecting viruses with at least 500 protein-coding genes (Figure 2). Of the 7,959 curated viral genomes
 93 (extracted from NCBI Taxonomy complete genomes), 24 represented genomes meet the threshold,
 94 among them 5 bacteria-infecting will not be further discussed. The 19 eukaryote-infecting viruses are
 95 the genuine giant viruses (Table 1).

96 Recall that reported proteome sizes are primarily based on automatic bioinformatics tools, which
 97 may differ from the experimental expression measurements (e.g., Mimivirus (APMV) [36]).
 98 Moreover, physical dimension is not in perfect correlation with the number of proteins or genome
 99 size. For example, Pithovirus sibericum, which was recovered from a 30,000-year-old permafrost
 100 sample [24], is one of the largest viruses by its physical dimensions (1.5 μm in length and 0.5 μm in
 101 diameter). However, it is excluded from this report, as its genome encodes only 467 proteins.
 102



103
 104
 105
 106
 107
 108

Figure 2. Distribution of viral proteome and genome sizes, colored by host taxonomy. There are 24 represented genomes that meet the threshold of ≥ 500 proteins among them 5 bacteria-infecting and 19 eukaryote-infecting viruses (dashed red line).

109 4. Classification of giant viruses and the question of origin

110 All giant viruses belong to the superfamily of nucleocytoplasmic large DNA viruses (NCLDV),
 111 which was substantially expanded following the discoveries of giant viruses [37,38]. The NCLDV
 112 superfamily had traditionally been comprised of the following families: Phycodnaviridae,
 113 Iridoviridae, Poxviridae, Asfarviridae, and Ascoviridae [39,40], for which a common ancestor had
 114 been proposed [41,42]. Following the inclusion of additional giant virus taxonomy groups
 115 (Mimiviridae, Pandoravirus and Marseillevirus) into the NCLDV superfamily, there remained only
 116 a handful of genes shared by the entire superfamily. Additional disparities in virion shapes and
 117 replication modes among NCLDV has led to the conclusion that the superfamily is not necessarily a
 118 taxonomic group, and that NCLDV families are more likely to have evolved separately [43-45].

119 **Table 1. Giant viruses**

Genome ^a	Accession	Genome			Host ^b	Year ^c
		length (kb)	# of proteins			
Mi-Acanthamoeba polyphaga mimivirus	NC_014649	1181.5	979	Pz, Ver	2010	
Mi-Acanthamoeba polyphaga moulouvirus	NC_020104	1021.3	894	Pz, Ver	2013	
Ph-Acanthocystis turfacea chlorella virus 1	NC_008724	288.0	860	Algae	2006	
Mi-Cafeteria roenbergensis virus BV-PW1	NC_014637	617.5	544	Pz	2010	
Pi-Cedratvirus A11	NC_032108	589.1	574	Pz	2016	
Ph-Chrysochromulina ericina virus	NC_028094	473.6	512	Algae	2015	
Mi-Megavirus chiliensis	NC_016072	1259.2	1120	Pz, Ver	2011	
UC-Mollivirus sibericum	NC_027867	651.5	523	Pz	2015	
Ph-Orpheovirus IHUMI-LCC2	NC_036594	1473.6	1199	Algae	2017	
Pa-Pandoravirus dulcis	NC_021858	1908.5	1070	Pz	2013	
Pa-Pandoravirus inopinatum	NC_026440	2243.1	1839	Pz	2015	
Pa-Pandoravirus macleodensis	NC_037665	1838.3	926	Pz	2018	
Pa-Pandoravirus neocaledonia	NC_037666	2003.2	1081	Pz	2018	
Pa-Pandoravirus quercus	NC_037667	2077.3	1185	Pz	2018	
Pa-Pandoravirus salinus	NC_022098	2473.9	1430	Pz	2013	
Ph-Paramecium bursaria Chlorella virus 1	NC_000852	330.6	802	Algae	1995	
Ph-Paramecium bursaria Chlorella virus AR158	NC_009899	344.7	814	Algae	2007	
Ph-Paramecium bursaria Chlorella virus FR483	NC_008603	321.2	849	Algae	2006	
Ph-Paramecium bursaria Chlorella virus NY2A	NC_009898	368.7	886	Algae	2007	

120 ^aFamilies: Mi, Mimiviridae; Ph, Phycodnaviridae; Pi, Pithoviridae; Pa, Pandoraviridae; UC, uncharacterized. ^bPz,
 121 protozoa; Ver, vertebrates. ^cYear of genome submission to NCBI.

122 Two models have been proposed for the evolution of giant viruses. According to the reductive
 123 model, an ancestral cellular genome has reduced in size, leading to dependence of the resulted
 124 genome on host cells. The presence of genes carrying cellular functions in almost any giant virus (e.g.
 125 translation components) [46] is consistent with this model. An alternative and more accepted theory
 126 argues for an expansion model. According to this model, current giant viruses have originated from

127 smaller ancestral viruses carrying only a few dozens of genes, and through gene duplications and
128 horizontal gene transfer (HGT), have rapidly expanded and diversified [44,47-49]. This model agrees
129 with metagenomic studies and the wave of giant virus discoveries in recent years (e.g. [31]).

130 To account for the limited number of homologous genes among giant viruses, different HGT
131 mechanisms have been proposed. The amoebae host in particular is often described as a melting pot
132 for DNA exchange [50] that leads to chimeric genomes.

133 Additional important players are virophages, small double-stranded DNA viruses that
134 hitchhike the replication system of giant viruses following coinfection of the host, and are considered
135 parasites of the coinfecting giant viruses [51]. A rich network of mobile genetic elements contributes
136 to the host-virus coevolution [52]. Virophages and other mobile elements could facilitate HGT
137 process, including interviral gene transfer, thereby have the potential of shaping the genomes of giant
138 viruses and impact their diversity [20,53,54]. Additional agents that play a role in the rapid dynamics
139 of giant viral genomes are a specific class of canonical transposable elements, which normally act in
140 cellular organisms. The discovery of transpovirons with sequences that are reminiscent to a CRISPR-
141 Cas system propose their contribution to host antiviral [55].

142 The majority of genes in giant viruses and specifically Mimiviridae have originated from the
143 cells they parasitize mostly amoebal and bacteria. Based on Phylogenetic trees, it is likely that
144 extensive HGT events led to their chimeric genomes. It was also suggested that the spectrum of hosts
145 may be larger than anticipated [56]. Therefore, comparative genomics over giant viruses which infect
146 the same host is unlikely to unambiguously resolve questions of gene origin, namely, whether shared
147 genes have originated from a common viral ancestor or the host. Thus, the degree of similarity among
148 giant viruses infecting different hosts is of a special interest. For example, the phyletic relationship
149 between Mimiviridae (which infect *Acanthamoeba*) and Phycodnaviridae (infecting algae) was
150 investigated, and it was found that the algae-infecting *Chrysochromulina ericina* virus (CeV, Table
151 1) showed moderate resemblance to the amoebae-infecting Mimivirus [57]. As a result, it was
152 suggested to reclassify CeV as a new clade of Mimiviridae rather than Phycodnaviridae. However, a
153 later discovery of another algae-infecting Phycodnaviridae virus (*Heterosigma akashiwo* virus,
154 HaV53) has provided a coherent phyletic relationship among Phycodnaviridae, thereby questioning
155 this reclassification [48].

156 In summary, the taxonomy of giant viruses, as all viruses, is still very unstable, and rapidly
157 updated with new discoveries [30]. The origin and ancestrality of giant viruses have remained
158 controversial with questions of origin are also unresolved [35]. Many newly discovered giant viruses
159 are not compatible with the notion of a single common ancestor, and some giant viruses remain
160 completely undetermined [4].

161

162 5. Common features

163 Despite the ongoing debate on their origin, giant viruses still share some important features. All
164 giant viruses belong to the double-stranded DNA (dsDNA) group, as all NCLDV families. The total
165 genome size of all the giant viruses listed in Table 1 is at least 288 Kbp (Figure 2). These giant viruses
166 are classified into several families: Mimiviridae, Pithoviridae, Pandoraviridae, Phycodnaviridae and
167 the Mollivirus genus [20,24,58].

168 All amoebae-infecting giant viruses rely on the non-specific phagocytosis by the amoebae host
169 [56]. Interestingly, a necessary condition for phagocytosis is a minimal particle size (~0.6 μm , [59]).
170 As amoebae (and related protozoa) are naturally fed on bacteria, it is likely that this minimal size for
171 inducing phagocytosis has become an evolutionary driving force for giant viruses. This fact, together
172 with the largely uncharacterized genomic content of giant viruses, may suggest that much of the
173 content in the genomes of giant viruses serves only for volume filling to increase their physical size.

174 Giant viruses share not only the cell entry process. When they exist the host cells during lysis,
175 as many as 1,000 virions are released from each lysed host via membrane fusion and active exocytosis
176 [60], which are relatively rare exit mechanisms in viruses.

177 Other than these genome and cell-biology similarities, other features of giant viruses are mostly
178 family-specific. For example, virion shapes and symmetries, nuclear involvement, duration for the

179 infection cycle and the steps in virion assembly, substantially vary among viruses from different
180 families [20,61,62].

181

182

183 **6. Proteome complexity and functional diversity**

184 The majority of the giant virus proteomes remain with no known function (Figure 3). Actually,
185 the fraction of uncharacterized proteins reaches 65%-85% of all reported proteins in giant viral
186 proteomes, many of them are ORFans. The most striking finding regarding the proteomes of giant
187 viruses is the presence of protein functions that are among the trademarks of cellular organisms, and
188 are never detected among other viruses. To exemplify the complexity of proteome functions in giant
189 viruses, we examine the proteome of the Cafeteria roenbergensis virus (CroV), which infects the
190 marine microzooplankton community in the Gulf of Mexico.

191 CroV was sequenced in 2010 as the first representative of an algae-infecting virus in the
192 Mimiviridae family. Unexpectedly, despite its affiliation with a recognized viral family, the majority
193 of its proteins show no significant similarity to any other known protein sequence. Of the remaining
194 proteins that show significant BLAST hits to other proteins from all domains of life, 45% are
195 eukaryotic sequences, 22% are from bacteria, and the rest are mostly from other viruses, including
196 other Mimivirus strains. A similar partition of protein origin applies across other members of the
197 Mimiviridae family.

198 The CroV proteome includes a rich set of genes involved in protein translation [63]. These genes
199 include multiple translation factors, a dozen of ribosomal proteins, tRNA synthetases, and 22
200 sequences encoding 5 different tRNAs [63]. As the lack of translation potential is considered a
201 hallmark of the virosphere, the presence of translational machinery components raised a debate on
202 the very definition of viruses [64,65]. Similar findings were extended to other giant virus strains of
203 the Tupanvirus genus in the same Mimiviridae family, which were recently isolated in Brazil [66].
204 The two viruses have 20 ORFs related to tRNA aminoacylation (aaRS), ~70 tRNA sequences decoding
205 the majority of the codons, 8 translation initiation factors, and elongation and release factors. The
206 theory that translation optimization is an evolutionary driving force in viruses [67] may in part
207 explain the curious presence of translation machinery in giant viruses.

208 In addition to translation, numerous CroV proteins are associated with the transcription
209 machinery. Specifically, The CroV proteome contains several subunits of the DNA-dependent RNA
210 polymerase II, initiation, elongation, and termination factors, the mRNA capping enzyme, and a
211 poly(A) polymerase. Presumably, the virus can activate its own transcription in the viral factory foci
212 in the cytoplasm of its host cell [43].

213 Another unexpected function detected in CroV is the DNA repair system, specifically of UV
214 radiation damage and base-excision repair. Other DNA-maintenance functions found in CroV
215 include helicase and topoisomerases (type I and II), suggesting a regulation on DNA replication,
216 recombination and chromatin remodeling.

217 Other rich set of functions related to protein maintenance include chaperons [69] and the
218 ubiquitin-proteasome system [70]. Interestingly, some of these genes seem to be acquired from
219 bacteria (e.g. a homolog of the E. coli heat-shock chaperon). Another rich collection of sugar-, lipid-
220 and amino acid-related metabolic enzymes were also found [17,71] which occupy 13% of the CroV
221 proteome (Figure 3).

222 It appears that the CroV proteome covers most functions traditionally attributed to cellular
223 organisms, including: protein translation, RNA maturation, DNA maintenance, proteostasis and
224 metabolism. Although CroV exemplifies many widespread functions in giant viruses, each strain has
225 its own unique functional composition. For example, the most abundant group of giant viruses in
226 ocean metagenomes, the Bodo saltans virus (BsV), was recently identified and classified into the same
227 microzooplankton-infecting Mimiviridae family [72]. Unlike the other family members, BsV does not
228 have an elaborate translation apparatus or tRNA genes, but it carries proteins active in cell membrane
229 trafficking and phagocytosis, yet more unprecedented functions discovered in viruses.

230

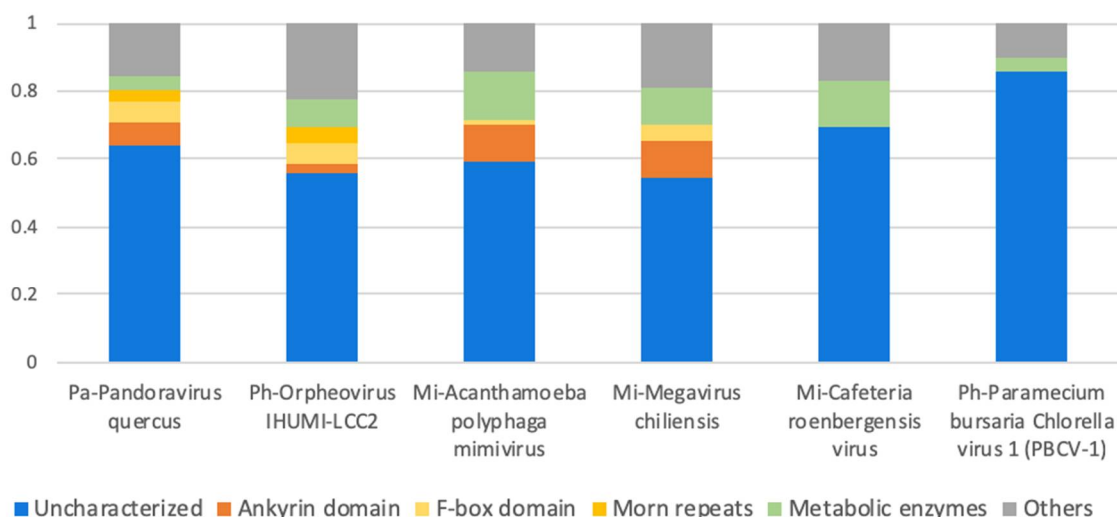


Figure 3. Protein function categories in 6 giant virus representatives from three families: Mimiviridae (Mi), Pandoviridae (Pa) and Phycodnaviridae (Ph). In all the proteomes, the majority of proteins are uncharacterized. Short repeated domains are abundant in the proteomes of amoebae-infecting giant viruses [71].

7. The emerging ecological view

Viruses are the most abundant entities in nature. In marine and fresh water habitats, there are millions of viruses in each milliliter of water [73]. However, the collection of virus isolates is often sporadic, especially for those without clinical or agricultural relevance. The accelerated pace in the discovery of giant viruses reflects the increasing number of sequencing projects of exotic environmental, including metagenomic projects [31,74].

Giant viruses have been isolated from various environmental niches and distant geographic places, revealing their global distribution and diversity. Current evidence suggests that the representation of giant viruses is underexplored, especially in soil ecosystems [30] and unique ecological niches [75,76]. In fact, ~60% of the giant viral genomes were completed after 2013 (Table 1). Many more virus–host systems that were reported for the last 5 years, are still await isolation and characterization [77].

The hosts of contemporary isolates include mainly protozoa, specifically amoeba (Table 1). However, the prevalence of amoeba as hosts may in part be attributed to sampling bias, specifically to the widespread use of amoebal coculture methods for testing ecological environments [27] (Table 1).

Despite their prevalence, the impact of giant viruses on human health deserve further investigation [21]. A cross-talks between giant viruses and activation of the innate immune cell system in human was reported [78]. Many viruses, including giant viruses were sequenced as part of the large-scale gut microbiome sequencing projects [79] but their composition and dynamic are yet to be determined [80]. Reports on the presence of sequences of giant viruses in the blood, the presence of antibodies against viral proteins, and their association with a broad collection of diseases (e.g. rheumatoid arthritis, unexplained pneumonia, lymphoma) are accumulating [81].

The presence of giant viruses in almost any environment, including extreme niches and manmade sites (e.g., sewage and wastewater plants) suggests that the ecological role of these fascinating viruses and their impact on human health are yet to be determined.

Author Contributions: The authors contributed to writing and manuscript design.

Funding: This research received no external funding.

Acknowledgments: The authors wish to thank the support to M.L. from Yad Hanadiv. We are grateful to Elixir infrastructure tools and resources that were used throughout.

268 **Conflicts of Interest:** The authors declare no conflict of interest.

269

270 **References**

271

272 1. Holmes, E.C. What does virus evolution tell us about virus origins? *J Virol* **2011**, *85*, 5247-5251,
273 doi:10.1128/JVI.02203-10.

274 2. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus*
275 *Res* **2006**, *117*, 5-16, doi:10.1016/j.virusres.2006.01.010.

276 3. Yutin, N.; Koonin, E.V. Evolution of DNA ligases of nucleo-cytoplasmic large DNA viruses of
277 eukaryotes: a case of hidden complexity. *Biol Direct* **2009**, *4*, 51, doi:10.1186/1745-6150-4-51.

278 4. Forterre, P. Viruses in the 21st Century: From the Curiosity-Driven Discovery of Giant Viruses to
279 New Concepts and Definition of Life. *Clin Infect Dis* **2017**, *65*, S74-S79, doi:10.1093/cid/cix349.

280 5. Moreira, D.; Lopez-Garcia, P. Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol*
281 **2009**, *7*, 306-311, doi:10.1038/nrmicro2108.

282 6. Brüssow, H. The not so universal tree of life or the place of viruses in the living world. *Philosophical*
283 *Transactions of the Royal Society B: Biological Sciences* **2009**, *364*, 2263-2274.

284 7. Bamford, D.H.; Grimes, J.M.; Stuart, D.I. What does structure tell us about virus evolution? *Current*
285 *opinion in structural biology* **2005**, *15*, 655-663.

286 8. Domingo, E.; Escarmis, C.; Sevilla, N.; Moya, A.; Elena, S.; Quer, J.; Novella, I.; Holland, J. Basic
287 concepts in RNA virus evolution. *The FASEB Journal* **1996**, *10*, 859-864.

288 9. Koonin, E.V.; Yutin, N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses.
289 *Intervirology* **2010**, *53*, 284-292, doi:10.1159/000312913.

290 10. Murphy, F.A.; Fauquet, C.M.; Bishop, D.H.; Ghabrial, S.A.; Jarvis, A.W.; Martelli, G.P.; Mayo,
291 M.A.; Summers, M.D. *Virus taxonomy: classification and nomenclature of viruses*; Springer Science &
292 Business Media: 2012; Vol. 10.

293 11. Stano, M.; Beke, G.; Klucar, L. viruSITE-integrated database for viral genomics. *Database (Oxford)*
294 **2016**, *2016*, doi:10.1093/database/baw162.

295 12. Lefkowitz, E.J.; Dempsey, D.M.; Hendrickson, R.C.; Orton, R.J.; Siddell, S.G.; Smith, D.B. Virus
296 taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic*
297 *Acids Res* **2018**, *46*, D708-D717, doi:10.1093/nar/gkx932.

298 13. Pybus, O.G.; Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat*
299 *Rev Genet* **2009**, *10*, 540-550, doi:10.1038/nrg2583.

300 14. Pérez-Losada, M.; Arenas, M.; Galán, J.C.; Palero, F.; González-Candelas, F. Recombination in
301 viruses: mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and*
302 *Evolution* **2015**, *30*, 296-307.

303 15. Brandes, N.; Linial, M. Gene overlapping and size constraints in the viral world. *Biol Direct* **2016**,
304 *11*, 26, doi:10.1186/s13062-016-0128-3.

305 16. La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; de Lamballerie, X.; Drancourt, M.; Birtles, R.;
306 Claverie, J.M.; Raoult, D. A giant virus in amoebae. *Science* **2003**, *299*, 2033,
307 doi:10.1126/science.1081867.

308 17. Raoult, D.; Audic, S.; Robert, C.; Abergel, C.; Renesto, P.; Ogata, H.; La Scola, B.; Suzan, M.;
309 Claverie, J.M. The 1.2-megabase genome sequence of Mimivirus. *Science* **2004**, *306*, 1344-1350,
310 doi:10.1126/science.1101485.

- 311 18. Xiao, C.; Chipman, P.R.; Battisti, A.J.; Bowman, V.D.; Renesto, P.; Raoult, D.; Rossmann, M.G.
312 Cryo-electron microscopy of the giant Mimivirus. *J Mol Biol* **2005**, *353*, 493-496,
313 doi:10.1016/j.jmb.2005.08.060.
- 314 19. Van Etten, J.L.; Meints, R.H. Giant viruses infecting algae. *Annu Rev Microbiol* **1999**, *53*, 447-494,
315 doi:10.1146/annurev.micro.53.1.447.
- 316 20. Abergel, C.; Legendre, M.; Claverie, J.M. The rapidly expanding universe of giant viruses:
317 Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol Rev* **2015**, *39*, 779-796,
318 doi:10.1093/femsre/fuv037.
- 319 21. Abrahão, J.S.; Dornas, F.P.; Silva, L.C.; Almeida, G.M.; Boratto, P.V.; Colson, P.; La Scola, B.; Kroon,
320 E.G. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding
321 discoveries. *Virology journal* **2014**, *11*, 120.
- 322 22. Campos, R.K.; Boratto, P.V.; Assis, F.L.; Aguiar, E.R.; Silva, L.C.; Albarnaz, J.D.; Dornas, F.P.;
323 Trindade, G.S.; Ferreira, P.P.; Marques, J.T. Samba virus: a novel mimivirus from a giant rain
324 forest, the Brazilian Amazon. *Virology journal* **2014**, *11*, 95.
- 325 23. Antwerpen, M.H.; Georgi, E.; Zoeller, L.; Woelfel, R.; Stoecker, K.; Scheid, P. Whole-genome
326 sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba. *Genome Announc*
327 **2015**, *3*, doi:10.1128/genomeA.00136-15.
- 328 24. Legendre, M.; Bartoli, J.; Shmakova, L.; Jeudy, S.; Labadie, K.; Adrait, A.; Lescot, M.; Poirot, O.;
329 Bertaux, L.; Bruley, C., et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA
330 viruses with a pandoravirus morphology. *Proc Natl Acad Sci U S A* **2014**, *111*, 4274-4279,
331 doi:10.1073/pnas.1320670111.
- 332 25. Claverie, J.M.; Abergel, C. From extraordinary endocytobionts to pandoraviruses. Comment on
333 Scheid et al.: Some secrets are revealed: parasitic keratitis amoebae as vectors of the scarcely
334 described pandoraviruses to humans. *Parasitol Res* **2015**, *114*, 1625-1627, doi:10.1007/s00436-014-
335 4079-2.
- 336 26. Andreani, J.; Verneau, J.; Raoult, D.; Levasseur, A.; La Scola, B. Deciphering viral presences: two
337 novel partial giant viruses detected in marine metagenome and in a mine drainage metagenome.
338 *Virol J* **2018**, *15*, 66, doi:10.1186/s12985-018-0976-9.
- 339 27. Aherfi, S.; Colson, P.; La Scola, B.; Raoult, D. Giant Viruses of Amoebas: An Update. *Front Microbiol*
340 **2016**, *7*, 349, doi:10.3389/fmicb.2016.00349.
- 341 28. Andrade, A.; Arantes, T.S.; Rodrigues, R.A.L.; Machado, T.B.; Dornas, F.P.; Landell, M.F.; Furst,
342 C.; Borges, L.G.A.; Dutra, L.A.L.; Almeida, G., et al. Ubiquitous giants: a plethora of giant viruses
343 found in Brazil and Antarctica. *Virol J* **2018**, *15*, 22, doi:10.1186/s12985-018-0930-x.
- 344 29. Dornas, F.P.; Assis, F.L.; Aherfi, S.; Arantes, T.; Abrahao, J.S.; Colson, P.; La Scola, B. A Brazilian
345 Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. *Viruses* **2016**, *8*,
346 76, doi:10.3390/v8030076.
- 347 30. Schulz, F.; Alteio, L.; Goudeau, D.; Ryan, E.M.; Yu, F.B.; Malmstrom, R.R.; Blanchard, J.; Woyke,
348 T. Hidden diversity of soil giant viruses. *Nat Commun* **2018**, *9*, 4881, doi:10.1038/s41467-018-07335-
349 2.
- 350 31. Backstrom, D.; Yutin, N.; Jorgensen, S.L.; Dharamshi, J.; Homa, F.; Zaremba-Niedwiedzka, K.;
351 Spang, A.; Wolf, Y.I.; Koonin, E.V.; Ettema, T.J.G. Virus Genomes from Deep Sea Sediments
352 Expand the Ocean Megavirome and Support Independent Origins of Viral Gigantism. *MBio* **2019**,
353 *10*, doi:10.1128/mBio.02497-18.

- 354 32. Colson, P.; La Scola, B.; Levasseur, A.; Caetano-Anolles, G.; Raoult, D. Mimivirus: leading the way
355 in the discovery of giant viruses of amoebae. *Nat Rev Microbiol* **2017**, *15*, 243-254,
356 doi:10.1038/nrmicro.2016.197.
- 357 33. Philippe, N.; Legendre, M.; Doutre, G.; Coute, Y.; Poirot, O.; Lescot, M.; Arslan, D.; Seltzer, V.;
358 Bertaux, L.; Bruley, C., et al. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching
359 that of parasitic eukaryotes. *Science* **2013**, *341*, 281-286, doi:10.1126/science.1239181.
- 360 34. Claverie, J.M.; Abergel, C. Mimiviridae: An Expanding Family of Highly Diverse Large dsDNA
361 Viruses Infecting a Wide Phylogenetic Range of Aquatic Eukaryotes. *Viruses* **2018**, *10*,
362 doi:10.3390/v10090506.
- 363 35. Colson, P.; Levasseur, A.; La Scola, B.; Sharma, V.; Nasir, A.; Pontarotti, P.; Caetano-Anolles, G.;
364 Raoult, D. Ancestrality and Mosaicism of Giant Viruses Supporting the Definition of the Fourth
365 TRUC of Microbes. *Front Microbiol* **2018**, *9*, 2668, doi:10.3389/fmicb.2018.02668.
- 366 36. Legendre, M.; Santini, S.; Rico, A.; Abergel, C.; Claverie, J.M. Breaking the 1000-gene barrier for
367 Mimivirus using ultra-deep genome and transcriptome sequencing. *Virol J* **2011**, *8*, 99,
368 doi:10.1186/1743-422X-8-99.
- 369 37. Yutin, N.; Wolf, Y.I.; Raoult, D.; Koonin, E.V. Eukaryotic large nucleo-cytoplasmic DNA viruses:
370 clusters of orthologous genes and reconstruction of viral genome evolution. *Virol J* **2009**, *6*, 223,
371 doi:10.1186/1743-422X-6-223.
- 372 38. Koonin, E.V.; Yutin, N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes
373 and Convergent Origins of Viral Gigantism. *Adv Virus Res* **2019**, *103*, 167-202,
374 doi:10.1016/bs.aivir.2018.09.002.
- 375 39. Andreani, J.; Aherfi, S.; Bou Khalil, J.Y.; Di Pinto, F.; Bitam, I.; Raoult, D.; Colson, P.; La Scola, B.
376 Cedratvirus, a Double-Cork Structured Giant Virus, is a Distant Relative of Pithoviruses. *Viruses*
377 **2016**, *8*, doi:10.3390/v8110300.
- 378 40. Wilhelm, S.W.; Coy, S.R.; Gann, E.R.; Moniruzzaman, M.; Stough, J.M. Standing on the Shoulders
379 of Giant Viruses: Five Lessons Learned about Large Viruses Infecting Small Eukaryotes and the
380 Opportunities They Create. *PLoS Pathog* **2016**, *12*, e1005752, doi:10.1371/journal.ppat.1005752.
- 381 41. Iyer, L.M.; Aravind, L.; Koonin, E.V. Common origin of four diverse families of large eukaryotic
382 DNA viruses. *J Virol* **2001**, *75*, 11720-11734, doi:10.1128/JVI.75.23.11720-11734.2001.
- 383 42. Nasir, A.; Kim, K.M.; Caetano-Anolles, G. Giant viruses coexisted with the cellular ancestors and
384 represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC*
385 *Evol Biol* **2012**, *12*, 156, doi:10.1186/1471-2148-12-156.
- 386 43. Fischer, M.G.; Allen, M.J.; Wilson, W.H.; Suttle, C.A. Giant virus with a remarkable complement
387 of genes infects marine zooplankton. *Proc Natl Acad Sci U S A* **2010**, *107*, 19508-19513,
388 doi:10.1073/pnas.1007615107.
- 389 44. Yutin, N.; Wolf, Y.I.; Koonin, E.V. Origin of giant viruses from smaller DNA viruses not from a
390 fourth domain of cellular life. *Virology* **2014**, *466-467*, 38-52, doi:10.1016/j.virol.2014.06.032.
- 391 45. Koonin, E.V.; Yutin, N. Multiple evolutionary origins of giant viruses. *F1000Res* **2018**, *7*,
392 doi:10.12688/f1000research.16248.1.
- 393 46. Arslan, D.; Legendre, M.; Seltzer, V.; Abergel, C.; Claverie, J.M. Distant Mimivirus relative with a
394 larger genome highlights the fundamental features of Megaviridae. *Proc Natl Acad Sci U S A* **2011**,
395 *108*, 17486-17491, doi:10.1073/pnas.1110889108.

- 396 47. Iyer, L.M.; Balaji, S.; Koonin, E.V.; Aravind, L. Evolutionary genomics of nucleocytoplasmic large
397 DNA viruses. *Virus Res* **2006**, *117*, 156-184, doi:10.1016/j.virusres.2006.01.009.
- 398 48. Maruyama, F.; Ueki, S. Evolution and Phylogeny of Large DNA Viruses, Mimiviridae and
399 Phycodnaviridae Including Newly Characterized Heterosigma akashiwo Virus. *Front Microbiol*
400 **2016**, *7*, 1942, doi:10.3389/fmicb.2016.01942.
- 401 49. Filee, J.; Chandler, M. Gene exchange and the origin of giant viruses. *Intervirology* **2010**, *53*, 354-
402 361, doi:10.1159/000312920.
- 403 50. Boyer, M.; Yutin, N.; Pagnier, I.; Barrassi, L.; Fournous, G.; Espinosa, L.; Robert, C.; Azza, S.; Sun,
404 S.; Rossmann, M.G., et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in
405 emergence of chimeric microorganisms. *Proc Natl Acad Sci U S A* **2009**, *106*, 21848-21853,
406 doi:10.1073/pnas.0911354106.
- 407 51. Desnues, C.; La Scola, B.; Yutin, N.; Fournous, G.; Robert, C.; Azza, S.; Jardot, P.; Monteil, S.;
408 Campocasso, A.; Koonin, E.V., et al. Provirophages and transpovirons as the diverse mobilome of
409 giant viruses. *Proc Natl Acad Sci U S A* **2012**, *109*, 18078-18083, doi:10.1073/pnas.1208835109.
- 410 52. Koonin, E.V.; Krupovic, M. Polintons, virophages and transpovirons: a tangled web linking
411 viruses, transposons and immunity. *Curr Opin Virol* **2017**, *25*, 7-15, doi:10.1016/j.coviro.2017.06.008.
- 412 53. Campbell, S.; Aswad, A.; Katzourakis, A. Disentangling the origins of virophages and polintons.
413 *Curr Opin Virol* **2017**, *25*, 59-65, doi:10.1016/j.coviro.2017.07.011.
- 414 54. La Scola, B.; Desnues, C.; Pagnier, I.; Robert, C.; Barrassi, L.; Fournous, G.; Merchat, M.; Suzan-
415 Monti, M.; Forterre, P.; Koonin, E., et al. The virophage as a unique parasite of the giant mimivirus.
416 *Nature* **2008**, *455*, 100-104, doi:10.1038/nature07218.
- 417 55. Levasseur, A.; Bekliz, M.; Chabriere, E.; Pontarotti, P.; La Scola, B.; Raoult, D. MIMIVIRE is a
418 defence system in mimivirus that confers resistance to virophage. *Nature* **2016**, *531*, 249-252,
419 doi:10.1038/nature17146.
- 420 56. Moreira, D.; Brochier-Armanet, C. Giant viruses, giant chimeras: the multiple evolutionary
421 histories of Mimivirus genes. *BMC Evol Biol* **2008**, *8*, 12, doi:10.1186/1471-2148-8-12.
- 422 57. Gallot-Lavallee, L.; Blanc, G.; Claverie, J.M. Comparative Genomics of Chrysochromulina Ericina
423 Virus and Other Microalga-Infecting Large DNA Viruses Highlights Their Intricate Evolutionary
424 Relationship with the Established Mimiviridae Family. *J Virol* **2017**, *91*, doi:10.1128/JVI.00230-17.
- 425 58. Yoosuf, N.; Pagnier, I.; Fournous, G.; Robert, C.; La Scola, B.; Raoult, D.; Colson, P. Complete
426 genome sequence of Courdo11 virus, a member of the family Mimiviridae. *Virus Genes* **2014**, *48*,
427 218-223, doi:10.1007/s11262-013-1016-x.
- 428 59. Rodrigues, R.A.L.; Abrahao, J.S.; Drumond, B.P.; Kroon, E.G. Giants among larges: how gigantism
429 impacts giant virus entry into amoebae. *Curr Opin Microbiol* **2016**, *31*, 88-93,
430 doi:10.1016/j.mib.2016.03.009.
- 431 60. Silva, L.; Andrade, A.; Domas, F.P.; Rodrigues, R.A.L.; Arantes, T.; Kroon, E.G.; Bonjardim, C.A.;
432 Abrahao, J.S. Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. *Sci Rep*
433 **2018**, *8*, 4000, doi:10.1038/s41598-018-22398-3.
- 434 61. Suzan-Monti, M.; La Scola, B.; Barrassi, L.; Espinosa, L.; Raoult, D. Ultrastructural characterization
435 of the giant volcano-like virus factory of *Acanthamoeba polyphaga* Mimivirus. *PLoS One* **2007**, *2*,
436 e328, doi:10.1371/journal.pone.0000328.

- 437 62. Diesend, J.; Kruse, J.; Hagedorn, M.; Hammann, C. Amoebae, Giant Viruses, and Virophages
438 Make Up a Complex, Multilayered Threesome. *Front Cell Infect Microbiol* **2017**, *7*, 527,
439 doi:10.3389/fcimb.2017.00527.
- 440 63. Schulz, F.; Yutin, N.; Ivanova, N.N.; Ortega, D.R.; Lee, T.K.; Vierheilig, J.; Daims, H.; Horn, M.;
441 Wagner, M.; Jensen, G.J., et al. Giant viruses with an expanded complement of translation system
442 components. *Science* **2017**, *356*, 82-85, doi:10.1126/science.aal4657.
- 443 64. Abrahao, J.S.; Araujo, R.; Colson, P.; La Scola, B. The analysis of translation-related gene set boosts
444 debates around origin and evolution of mimiviruses. *PLoS Genet* **2017**, *13*, e1006532,
445 doi:10.1371/journal.pgen.1006532.
- 446 65. Raoult, D.; Forterre, P. Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol* **2008**, *6*, 315-
447 319, doi:10.1038/nrmicro1858.
- 448 66. Abrahao, J.; Silva, L.; Silva, L.S.; Khalil, J.Y.B.; Rodrigues, R.; Arantes, T.; Assis, F.; Boratto, P.;
449 Andrade, M.; Kroon, E.G., et al. Tailed giant Tupanvirus possesses the most complete translational
450 apparatus of the known virosphere. *Nat Commun* **2018**, *9*, 749, doi:10.1038/s41467-018-03168-1.
- 451 67. Bahir, I.; Fromer, M.; Prat, Y.; Linial, M. Viral adaptation to host: a proteome-based analysis of
452 codon usage and amino acid preferences. *Mol Syst Biol* **2009**, *5*, 311, doi:10.1038/msb.2009.71.
- 453 68. Barik, S. A Family of Novel Cyclophilins, Conserved in the Mimivirus Genus of the Giant DNA
454 Viruses. *Comput Struct Biotechnol J* **2018**, *16*, 231-236, doi:10.1016/j.csbj.2018.07.001.
- 455 69. Yutin, N.; Colson, P.; Raoult, D.; Koonin, E.V. Mimiviridae: clusters of orthologous genes,
456 reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. *Virol*
457 *J* **2013**, *10*, 106, doi:10.1186/1743-422X-10-106.
- 458 70. Van Etten, J.L. Unusual life style of giant chlorella viruses. *Annu Rev Genet* **2003**, *37*, 153-195,
459 doi:10.1146/annurev.genet.37.110801.143915.
- 460 71. Deeg, C.M.; Chow, C.T.; Suttle, C.A. The kinetoplastid-infecting Bodo saltans virus (BsV), a
461 window into the most abundant giant viruses in the sea. *Elife* **2018**, *7*, doi:10.7554/eLife.33014.
- 462 72. Shukla, A.; Chatterjee, A.; Kondabagil, K. The number of genes encoding repeat domain-
463 containing proteins positively correlates with genome size in amoebal giant viruses. *Virus Evol*
464 **2018**, *4*, vex039, doi:10.1093/ve/vex039.
- 465 73. Suttle, C.A. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **2007**, *5*, 801-
466 812, doi:10.1038/nrmicro1750.
- 467 74. Chatterjee, A.; Sicheritz-Ponten, T.; Yadav, R.; Kondabagil, K. Genomic and metagenomic
468 signatures of giant viruses are ubiquitous in water samples from sewage, inland lake, waste water
469 treatment plant, and municipal water supply in Mumbai, India. *Sci Rep* **2019**, *9*, 3690,
470 doi:10.1038/s41598-019-40171-y.
- 471 75. Campos, R.K.; Boratto, P.V.; Assis, F.L.; Aguiar, E.R.; Silva, L.C.; Albarnaz, J.D.; Dornas, F.P.;
472 Trindade, G.S.; Ferreira, P.P.; Marques, J.T., et al. Samba virus: a novel mimivirus from a giant
473 rain forest, the Brazilian Amazon. *Virol J* **2014**, *11*, 95, doi:10.1186/1743-422X-11-95.
- 474 76. Yoshikawa, G.; Blanc-Mathieu, R.; Song, C.; Kayama, Y.; Mochizuki, T.; Murata, K.; Ogata, H.;
475 Takemura, M. Medusavirus, a novel large DNA virus discovered from hot spring water. *J Virol*
476 **2019**, 10.1128/JVI.02130-18, doi:10.1128/JVI.02130-18.
- 477 77. Wilhelm, S.W.; Bird, J.T.; Bonifer, K.S.; Calfee, B.C.; Chen, T.; Coy, S.R.; Gainer, P.J.; Gann, E.R.;
478 Heatherly, H.T.; Lee, J., et al. A Student's Guide to Giant Viruses Infecting Small Eukaryotes: From
479 *Acanthamoeba* to *Zooxanthellae*. *Viruses* **2017**, *9*, doi:10.3390/v9030046.

- 480 78. Silva, L.C.; Almeida, G.M.; Oliveira, D.B.; Dornas, F.P.; Campos, R.K.; La Scola, B.; Ferreira, P.C.;
481 Kroon, E.G.; Abrahao, J.S. A resourceful giant: APMV is able to interfere with the human type I
482 interferon system. *Microbes Infect* **2014**, *16*, 187-195, doi:10.1016/j.micinf.2013.11.011.
- 483 79. Scarpellini, E.; Ianiro, G.; Attili, F.; Bassanelli, C.; De Santis, A.; Gasbarrini, A. The human gut
484 microbiota and virome: Potential therapeutic implications. *Dig Liver Dis* **2015**, *47*, 1007-1012,
485 doi:10.1016/j.dld.2015.07.008.
- 486 80. Popgeorgiev, N.; Temmam, S.; Raoult, D.; Desnues, C. Describing the silent human virome with
487 an emphasis on giant viruses. *Intervirology* **2013**, *56*, 395-412, doi:10.1159/000354561.
- 488 81. Colson, P.; La Scola, B.; Raoult, D. Giant Viruses of Amoebae: A Journey Through Innovative
489 Research and Paradigm Changes. *Annu Rev Virol* **2017**, *4*, 61-85, doi:10.1146/annurev-virology-
490 101416-041816.
- 491

492